Probabilistic-based approach to optimal filtering

A. Hannachi
Atmospheric, Oceanic and Planetary Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, United Kingdom
(Received 25 June 1999)

The signal-to-noise ratio maximizing approach in optimal filtering provides a robust tool to detect signals in the presence of colored noise. The method fails, however, when the data present a regimelike behavior. An approach is developed in this manuscript to recover local (in phase space) behavior in an intermittent regimelike behaving system. The method is first formulated in its general form within a Gaussian framework, given an estimate of the noise covariance, and demands that the signal corresponds to minimizing the noise probability distribution for any given value, i.e., on isosurfaces, of the data probability distribution. The extension to the non-Gaussian case is provided through the use of finite mixture models for data that show regimelike behavior. The method yields the correct signal when applied in a simplified manner to synthetic time series with and without regimes, compared to the signal-to-noise ratio approach, and helps identify the right frequency of the oscillation spells in the classical and variants of the Lorenz system.

PACS number(s): 05.45.−a, 02.50.−r

I. INTRODUCTION

It is a fact that physical and natural phenomena, such as geophysical systems, have a complicated nonlinear character involving, most of the time, high- and intermittent low-frequency behavior. The question of identifying regular signals from such systems is therefore an important and challenging one given that (*i*) in most, if not all, physical systems only a subset of observables are generally accessible, and (*ii*) even if we gain much information about the system, the quality of the observations, which is at least as crucial as the observables themselves, can never be perfect. Hence the need of tools to unveil information regarding signal detection from such systems.

Most techniques of signal detection, such as singular spectrum analysis, (SSA) [1,2], and signal-to-noise ratio maximizing approach [3], all attempt to identify oscillations embedded in a given data set that behave ''nicely.'' To the best of our knowledge no such attempt has been made to genuinely investigate the question of a true high-frequency signal in an intermittent low-frequency behavior. More precisely, although the problem of signal detection has gone through a wide range of application, the question of identifying oscillations in an intermittent regimelike behaving system does not seem, to our knowledge, to have been truly addressed.

The latter question is of crucial importance in geophysical sciences and particularly in analyzing climate variability given the need to understand the climate system in order to be able to predict it for long lead time. Because of its complicated nonlinear structure, the climate system can often exhibit a regimelike behavior [4,5]. From a probabilistic framework, the latter feature implies that the climate system can often display a significant departure from Gaussian behavior. Therefore, detection approaches should take this fact into consideration if one wants to identify reliable climate signals.

In an attempt to identify climate signals, Ghil and Vautard [6] applied SSA to the Intergovernmental Panel on Climate Change (IPCC) 136-year record of global annual-mean near-surface temperature to look for the signature of an interdecadal oscillation by identifying an eigenorthogonal function (EOF), i.e., covariance matrix eigenvector, pairs of sine-cosine waves corresponding to (nearly) equal eigenvalues of the covariance matrix. However, Allen and Smith [3] have shown that the IPCC series is consistent with a nonlinear trend plus a first-order autoregressive process [AR(1)] noise, and therefore does not indicate either interannual or interdecadal oscillations. In fact, SSA is justified only if the hypothetical noise in the data set is white, but that it can fail to identify genuine and detectable oscillations in the presence of colored noise. Reference [3] proposed a refinement based on maximizing a signal-to-noise (S/N) ratio, similar to the metric-based principal component analysis [7]. The S/N ratio technique fails, however, when, for example, the data present a non-Gaussian behavior, which is the main concern here.

In this paper we attempt to include both factors, the noise and regimelike behavior, in order to recover a true high-frequency signal in an intermittent low-frequency behaving system. Because this study was initially driven by some climate variability investigations, the method assumes the knowledge of the noise. Most climate studies use general circulation models (GCMs) to analyze climate variability. For example, to analyze climate change, GCMs provide a good tool to estimate and therefore separate the internal variability noise from the response to external forcing by performing an ensemble of integrations corresponding to the same forcing but with different starting conditions. Also, in many geophysical examples, the noise can be well approximated by a colored noise with parameters calculated from the data. Nevertheless, the problem of estimating the noise in general remains a difficult one.

Since the present approach includes and also generalizes the S/N maximizing ratio technique [3] we briefly review the background of the technique in Sec. II and introduce some notations that will be used later. The formulation of the approach within the Gaussian framework is presented in Sec. III, where both probability distribution functions (PDFs) of the data and the noise are Gaussian. Extension to non-Gaussian cases is presented in Sec. IV. Application to syn-

thetic univariate time series as well as to variants of the classical Lorenz system [4] is presented in Sec. V and finally a summary and conclusion are given in Sec. VI.

II. BACKGROUND AND THE *S/N* RATIO

A natural way to estimate a signal from (noisy) data, given an estimate of the noise, is to look for patterns in the phase space that maximize the S/N ratio. The latter technique was used in [3] as an extension to the SSA to deal with colored noise and we give here a brief description of the method along with useful notations. A prewhitening transformation [8] was used by [3] to maximize the S/N ratio

$$\rho = \frac{\mathbf{u}^T \mathbf{C}_D \mathbf{u}}{\mathbf{u}^T \mathbf{C}_N \mathbf{u}} = \frac{\boldsymbol{\zeta}^T \mathbf{C}_D^* \boldsymbol{\zeta}}{\boldsymbol{\zeta}^T \boldsymbol{\zeta}}, \tag{1}$$

where

$$\boldsymbol{\zeta} = \mathbf{C}_N^{1/2} \mathbf{u}, \quad \mathbf{u} = \mathbf{C}_N^{-1/2} \boldsymbol{\zeta}, \text{ and } \mathbf{C}_D^* = \mathbf{C}_N^{-1/2^T} \mathbf{C}_D \mathbf{C}_N^{-1/2}. \tag{2}$$

In Eqs. (1) and (2), $\mathbf{C}_D$ and $\mathbf{C}_N$ are the covariance matrices of the data and the noise, respectively. The vectors $\mathbf{u} = \mathbf{C}_N^{-1/2} \boldsymbol{\zeta}$, where $\boldsymbol{\zeta}$ are the eigenvectors of $\mathbf{C}_D^*$ with the largest eigenvalues, thus provide an optimal set of signal-to-noise ratio (S/N) maximizing patterns.

The notation in Eq. (2), which will be used throughout, is derived from the singular value decomposition (SVD) procedure [9,10]. Accordingly, if $\mathbf{C}$ is a covariance matrix we have

$$\mathbf{C} = \mathbf{E} \boldsymbol{\Lambda} \mathbf{E}^T = \mathbf{C}^{1/2^T} \mathbf{C}^{1/2} \text{ and } \mathbf{C}^{-1} = \mathbf{E} \boldsymbol{\Lambda}^{-1} \mathbf{E}^T = \mathbf{C}^{-1/2} \mathbf{C}^{-1/2^T}, \tag{3}$$

where $\mathbf{C}^{1/2} = \Lambda^{1/2} \mathbf{E}^T$ and $\mathbf{C}^{-1/2} = \mathbf{E} \boldsymbol{\Lambda}^{-1/2}$ (note that the ''square root'' of a covariance matrix is not unique). The diagonal matrix $\boldsymbol{\Lambda}$ contains the eigenvalues of $\mathbf{C}$ and the columns of $\mathbf{E}$ constitute its eigenvectors. Note that for univariate data, for example, an oscillatory signal corresponds to the case of two leading eigenvalues being (nearly) equal and separated from the rest of the spectrum. In this case the two vectors $\mathbf{u}_0 = \mathbf{C}_N^{-1/2} \boldsymbol{\zeta}_0$ and $\mathbf{u}_1 = \mathbf{C}_N^{-1/2} \boldsymbol{\zeta}_1$, where $\boldsymbol{\zeta}_0$ and $\boldsymbol{\zeta}_1$ are the leading eigenvectors of $\mathbf{C}_D^*$, constitute two oscillatory waves in quadrature.

III. PDF-BASED SIGNAL DETECTION APPROACH: GAUSSIAN CASE

A. Introduction and objectives

In the real world, nonlinearities can be important in introducing non-Gaussian behavior, which can ''frustrate'' such linear diagnostics in characterising climate signals as pointed out in [11]. A probabilistic approach based on the PDF is therefore appropriate to the problem and is easier to generalize to non-Gaussian systems. Our objective here is to identify regular signals (oscillation in this case as presented in Sec. V) embedded in a noisy data set based on the assumption that we know the noise probability distribution. We suppose here that the data and the noise are both Gaussian, centered at the phase space origin, and with respective covariances $\mathbf{C}_D$ and $\mathbf{C}_N$. The generalization to non-Gaussian systems follows in Sec. IV.

It is the case that the knowledge of the noise probability distribution is a strong hypothesis in the real world. However, as pointed out in the Introduction, in many examples of climate analyses this hypothesis can be met. For instance, climate models can provide the means to estimate the internal variability noise of the climate system. The use of the Gaussian distribution function, on the other hand, is based on the completeness of the set of such distributions that will be exploited in Sec. IV to extend this technique to include the case where the PDF of the data is not (approximately) Gaussian but rather a mixture of Gaussians.

B. Description of the approach

To formulate the approach, we denote first by p_D and p_N the Gaussian probability distributions of the data and the noise, respectively. Then the respective probability densities of a particular pattern $\mathbf{v}$ are

$$p_D(\mathbf{v}) = \frac{1}{(2\pi)^{m/2} |\mathbf{C}_D|^{1/2}} e^{-1/2 \mathbf{v}^T \mathbf{C}_D^{-1} \mathbf{v}}, \tag{4}$$

and

$$p_N(\mathbf{v}) = \frac{1}{(2\pi)^{m/2} |\mathbf{C}_N|^{1/2}} e^{-1/2 \mathbf{v}^T \mathbf{C}_N^{-1} \mathbf{v}}. \tag{5}$$

The notation $|\mathbf{C}_D|$ in Eq. (4) stands for the determinant of $\mathbf{C}_D$. Note that $p_D(\mathbf{v}) dV [p_N(\mathbf{v}) dV]$ is the probability of the data [noise] being within the phase space volume dV near to the position or pattern $\mathbf{v}$. Note also that the independence between the signal and the noise is implicitely implied by Eqs. (4) and (5). In theory it is possible to drop this assumption provided we know the joint probability distribution. In practice, however, such information is very seldom available.

Ideally, we would be interested in locations within the phase space with simultaneous high data and low noise PDFs, respectively. If both data and noise PDFs are adequately defined, these positions within the phase space yield the most likely signal location. Note that problems may arise with the estimation of these PDFs from limited samples but the principle should be clear.

Since, in particular, we are seeking locations with minimum noise PDF, we adopt the following natural procedure. Within the phase space, we form connected subsets S_d corresponding to isopleths of p_D, obtained by cutting through the data PDF, p_D, at different constant values δ_d, i.e., $S_d = p_D^{-1}(\delta_d)$, and we then minimize the density p_N over each of these connected subsets. To see this we note first that each S_d is given by $d^2 = \mathbf{v}^T C_D^{-1} \mathbf{v}$. The PDF of the data p_D can then be transformed using the preceding metric, i.e., the Mahalanobis metric based on $\mathbf{C}_D^{-1}$, to yield an isotropic distribution whereby all directions are equally probable. We then look for regions or directions within the phase space with minimum noise PDF for any given data PDF value, given by $(2\pi)^{-m/2} |\mathbf{C}_D|^{-1/2} \exp(-\frac{1}{2} d^2)$. Such minimum exists and, as shown below, it turns out that when

varying the data PDF value, i.e., varying δ_d, all the minima remain in the same direction, which is by definition the required signal pattern.

To proceed we then minimize

$$\min p_N = \frac{1}{(2\pi)^{m/2}|\mathbf{C}_N|^{1/2}} e^{-1/2\mathbf{v}^T\mathbf{C}_N^{-1}\mathbf{v}}, \tag{6}$$

or, equivalently (the exponential is monotonically increasing),

$$\max \mathbf{v}^T\mathbf{C}_N^{-1}\mathbf{v}, \tag{7}$$

for any given p_D value or, similarly, any given

$$d^2 = \mathbf{v}^T\mathbf{C}_D^{-1}\mathbf{v}. \tag{8}$$

We find the patterns $\mathbf{v}$ that minimize p_N over isosurfaces of p_D by finding the stationary points of the Lagrangian

$$\mathcal{L} = \mathbf{v}^T\mathbf{C}_N^{-1}\mathbf{v} - \lambda(d^2 - \mathbf{v}^T\mathbf{C}_D^{-1}\mathbf{v}), \tag{9}$$

where λ is a Lagrange multiplier. This yields the generalized eigenproblem $\mathbf{C}_N^{-1}\mathbf{v} = \lambda\mathbf{C}_D^{-1}\mathbf{v}$, which can be written in the more compact form

$$C_D\mathbf{C}_N^{-1}\mathbf{v} = \lambda\mathbf{v}. \tag{10}$$

It is worth noting that the same solution patterns are obtained by maximizing p_D on isosurfaces of p_N instead, but we follow the first choice, i.e., minimizing p_N, because it is simpler if the data distribution is more complicated than the noise distribution as is often the case (see Sec. IV).

The other way to solve the problem is via the SVD procedure. Using notation from Eq. (3) and letting $\boldsymbol{\xi} = \mathbf{C}_D^{-1/2^T}\mathbf{v}$, then Eq. (7) becomes equivalent to

$$\max \boldsymbol{\xi}^T(\mathbf{C}_D^{1/2}\mathbf{C}_N^{-1}\mathbf{C}_D^{1/2^T})\boldsymbol{\xi}, \tag{11}$$

along with the constraint (8), $|\boldsymbol{\xi}|^2 = d^2$. Combined with a Lagrange multiplier, Eq. (11) results in the eigenvector $\boldsymbol{\xi}_0$ corresponding to the largest eigenvalue λ_0 of the symmetric operator $\mathbf{C}_D' = \mathbf{C}_D^{1/2}\mathbf{C}_N^{-1}\mathbf{C}_D^{1/2^T}$ whose decomposition is

$$\mathbf{C}_D' = \mathbf{C}_D^{1/2}\mathbf{C}_N^{-1}\mathbf{C}_D^{1/2^T} = \mathbf{E}_D'\boldsymbol{\Lambda}_D'\mathbf{E}'^T_D. \tag{12}$$

The diagonal matrix Λ_D' contains the eigenvalues of $\mathbf{C}_D'$ whose eigenvectors form the columns of $\mathbf{E}_D'$. The signal pattern $\mathbf{v}$ solution to the original problem is then given by the transformation

$$\mathbf{v} = \mathbf{C}_D^{1/2^T}\boldsymbol{\xi}_0 = \mathbf{E}_D\boldsymbol{\Lambda}_D^{1/2}\boldsymbol{\xi}_0. \tag{13}$$

The required patterns [Eqs. (10) or (13)] are therefore eigenvectors of the matrix $\mathbf{C}_D\mathbf{C}_N^{-1}$ that we label *detection* matrix for obvious reasons and whose spectrum is given by the diagonal matrix Λ_D'. Hence, provided $\mathbf{C}_N$ represents the true covariance of the noise, the eigenvector of the detection matrix corresponding to the maximum eigenvalue represents an estimate of the signal orientation. Application to synthetic test series has been performed in Sec. V A and the true oscillation in the data was identified (see Fig. 2).

C. Discussion

The choice of labelling the operator $\mathbf{C}_D\mathbf{C}_N^{-1}$ detection matrix is based on various reasons. First, the matrix is used to detect signal patterns as presented here, and also to find coefficient vectors of the metric-based principal components presented in [7]. It also appears explicitly in other related subjects such as the study of information content of observed atmospheric data [12] and numerical weather prediction [13]. Moreover, in the field of classification, one also encounters the same kind of matrix. In fact, when looking for classification rules to detect or identify clusters, a similar detection matrix emerges explicitly when Fisher's linear discriminant analysis is applied [14] or when a simple optimization criterion of the trace of the class covariance matrix is involved [15]. For numerical computation of its spectrum it may be preferable to proceed via Eqs. (11) and (12) since the detection matrix is in general nonsymmetric. Other iterative methods exist to solve the generalized eigenproblem (10) based on the so-called Jacobi-Davidson method when the matrices are large and we refer the reader, for details, to [16].

Now, to make clear the link to the S/N ratio approach, we differentiate Eq. (1) with respect to $\mathbf{u}$ to yield the (generalized) eigenproblem

$$\mathbf{C}_N^{-1}\mathbf{C}_D\mathbf{u} = \rho\mathbf{u}, \tag{14}$$

whose solutions are the eigenvectors of the detection matrix adjoint. Therefore, the S/N maximizing patterns and the PDF-based patterns form, respectively, the left and right eigenvectors of the detection matrix $\mathbf{C}_D\mathbf{C}_N^{-1}$, i.e., they are in duality of each other. In the single channel, i.e., the univariate case, these two patterns generally coincide with the eigenvector of the signal covariance when this latter is low rank. However, in multichannel problems the S/N maximizing patterns and the PDF patterns are generally different for the simple reason that the detection matrix is not symmetric, so the distinction between them is important. Furthermore, in addition to being a S/N ratio, an eigenvalue λ of the detection matrix can now be interpreted in such a way that $\exp(-\lambda)$ represents the noise probability around the corresponding pattern within a unit element of volume of the phase space.

To complement our analysis, the metric-based principal components presented in [7] can be discussed. Although the original problem addressed in this paper is different from that of [7], several common points emerge. The noise covariance matrix $\mathbf{C}_N$ is the equivalent of the error covariance matrix of [7], while the S/N ratio is the equivalent of the variability of the index (generalized principal component) relative to its uncertainty. Equations (10) and (14) lead, after simple algebra, to

$$\mathbf{v} = \mathbf{C}_N\mathbf{u}, \tag{15}$$

so that $\mathbf{u}$ and $\mathbf{v}$ are, respectively, the equivalent of the coefficient or weight vectors and state space patterns of [7].

Note that owing to the symmetric, but not identical, roles of $\mathbf{u}$ and $\mathbf{v}$ [Eqs. (10) and (14)] a similar equation to (15), involving $\mathbf{C}_D$ instead, exists. Also, Thacker and Lewandowicz [17] emphasized the importance of the Gauss-Markov theorem in determining the Gauss-Markov weights for indi-

ces to predict local variables. Although stated in a different context, it is worth pointing out the connection between the probabilistic interpretation of the patterns and the Gauss-Markov weights.

IV. EXTENSION TO THE NON-GAUSSIAN CASE

A. Introduction

Although the real world may not be Gaussian, it can, however, exhibit regimelike behavior, where each regime is nearly Gaussian. This regime behavior is observed almost everywhere, from dynamo theory in geophysics, related to the reversals of the Earth's magnetic field [18], to geophysical fluid dynamics [4,5]. For instance, it has been observed that the atmosphere is not a mere ''red noise'' of day-to-day fluctuations but displays, rather, a kind of regimelike behavior. As Leith [19] quoted: ''One suggested source of such slow changes is the possible existence of separate regimes of internal behavior such that the atmosphere, having entered one regime, tends to remain there for extended periods of time. The long internal time scales induced by such almost-intransitive behavior [20] give a possibility of long-range forecasting skill, especially if the transitions between regimes are predictable.'' Unfortunately, these transitions are generally unpredictable, as can be modeled by the simplified Lorenz model [4].

Gaussian mixture models are a good candidate for modelling such regimelike behavior where each regime cluster can be well approximated by a Gaussian and the overall data set becomes a weighted sum of these individual Gaussian PDFs. The cornerstone of the mixture model is the well-known elementary, but important, analytical result summarized in the following lemma [21].

Lemma. Any probability density $p(\mathbf{x})$ can be approximated as closely as desired [in $L^1(\mathbf{R}^m)$ norm, i.e., $\|f\| = \int_{\mathbf{R}^m} |f(\mathbf{x})| d\mathbf{x}$] by a finite mixture (weighted sum) of Gaussians as; $p(\mathbf{x}) = \sum_{i=1}^{k} \alpha_i \gamma_i(\mathbf{x})$ for some integer k, positive scalars α_i with $\sum_{i=1}^{k} \alpha_i = 1$, and Gaussian PDFs γ_i.

The mixture model technique is widely used in applied statistics [22] and also in geophysics and other fields such as the application to atmospheric-model data clustering [23]. It is very convenient for PDF estimation and cluster identification [24] especially given the existence of efficient iterative approaches, such as the expectation-maximization algorithm [25]. Everitt and Hand [26] also provide a good review and cover most algorithms used in mixture analysis.

B. The non-Gaussian case

We now consider the extension of the approach described in Sec. III B to deal with the non-Gaussian case for the data and based, as before, on the assumption that we have an estimate of the noise probability distribution. Precisely, we consider, here and in the corresponding application in Sec. V B, the noise to be Gaussian and the data to be, for simplicity, a mixture of two Gaussians. The case of more than two can be treated similarly. The expression of the noise probability distribution p_N is given by Eq. (5), while for the data we now have

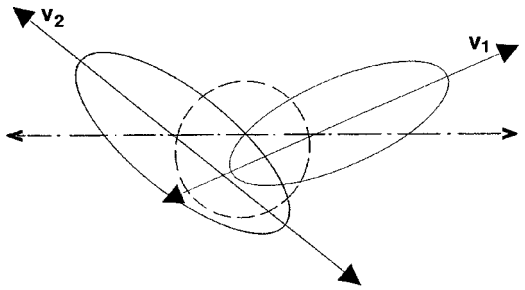

FIG. 1. Schematic representation of a 2D case with two regimes shown by ellipses. The dashed-dotted line is what a conventional, like the S/N ratio, approach would give while the present approach would produce the signals $\mathbf{v}_1$ and $\mathbf{v}_2$. The dashed circle represents the prewhitened noise.

$$(2\pi)^{m/2} p_D(\mathbf{x}) = \frac{\alpha}{|\mathbf{C}_{D_1}|^{1/2}} e^{-1/2(\mathbf{x}-\boldsymbol{\mu}_1)^T \mathbf{C}_{D_1}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)} + \frac{1-\alpha}{|\mathbf{C}_{D_2}|^{1/2}} e^{-1/2(\mathbf{x}-\boldsymbol{\mu}_2)^T \mathbf{C}_{D_2}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)}, \quad (16)$$

where $\boldsymbol{\mu}_i$ and $\mathbf{C}_{D_i}$ $(i=1,2)$ are the center and the covariance matrix of the ith cluster, respectively. In Eq. (16), α is the proportion of the data in the first cluster.

Since, in general, the solution to the equivalent of Eqs. (6)–(8) when we have a mixture PDF for the data such as Eq. (16) can only be obtained numerically, and in order to get a simplified analytical solution we suppose that the two Gaussians are not very close to each other, i.e., the data PDF is bimodal, so that optimization can be done locally around each maximum of p_D (see the Appendix). Note that bimodality does not necessarily mean that the Gaussians are well separated; it is only a simplified way to solve the problem and the final result is independent of this hypothesis. It should be noted in this context that not all mixture distributions of type (16) show bimodality [26].

Subject to this bimodality assumption, the original problem is transformed into a signal detection problem over each data cluster (see the Appendix) and consequently there is a solution $\mathbf{v}$ for each cluster. More precisely, the solution $\mathbf{v}_i$, to the above problem, for the ith cluster $(i=1,2)$ is given by

$$\mathbf{v}_i = \mathbf{C}_{D_i}^{1/2^T} \boldsymbol{\xi}_0^i, \quad (17)$$

where ξ_0^i is the eigenvector of the matrix $\mathbf{C}'_{D_i} = \mathbf{C}_{D_i}^{1/2} \mathbf{C}_N^{-1} \mathbf{C}_{D_i}^{1/2^T}$ $(i=1,2$ for clusters 1 and 2, respectively) corresponding to its leading eigenvalue λ_0^i. Said in a simplified manner, the signals can be identified by reapplying the analysis of Sec. III but over each cluster of the data separately, having first subtracted the mean evaluated over that cluster. Figure 1 shows a schematic representation of a two-dimensional (2D) case with two Gaussians (regimes). The dashed-dotted line (major axis of the data covariance matrix) is what conventional approaches, such as the S/N ratio technique, would give, while the actual approach would identify the right signals $\mathbf{v}_1$ and $\mathbf{v}_2$.

V. APPLICATION TO A GENERATED TIME SERIES

The approach developed so far is general and can be used with embedded and nonembedded state vectors. Since, however, we are dealing with univariate signals in this section it is therefore helpful to embed the signal in a higher dimensional phase space, using the delay coordinates method [27] as usually performed in SSA, although SSA is not the key issue here. In this case, the time series $(x_i)_{1\leqslant i\leqslant N}=[x(t_i)]_{1\leqslant i\leqslant N}$ of a variable x at times (t_i) is transformed into a higher M-dimensional phase space by sampling the data using a sliding window of chosen length M to yield the new M-dimensional time series $(\mathbf{X}_i)_{1\leqslant i\leqslant N-M+1}$,

$$\mathbf{X}_i=(x_{i,\ldots,}x_{i+M-1}) \quad \text{for} \quad 1\leqslant i\leqslant N-M+1. \tag{18}$$

The covariance matrix $\mathbf{C}_D$ of the new time series is calculated based on the ''nonbiased'' Yule-Walker rule for the autocovariance function, $c_j=1/(N-j)\Sigma_{i=1}^{N-j}x_ix_{i+j}$, of the original time series (x_i) and is given by

$$(\mathbf{C}_D)_{ij}=c_{i-j-1}. \tag{19}$$

A. Case of an approximate Gaussian distribution: No regime behavior

We first construct the signal $S(t)$ by randomly generating damped sinusoidal bursts, with period $T=5$ units and an e-folding time τ of 30 units, initiated with 0.5% chance at each time step. Analytically, the signal can be written as

$$S(t)=\sum_{n=0}^{n=\infty} 1_{[T_n,T_{n+1}]}(t)\exp\left(-\frac{t-T_n}{\tau}\right)\sin\left(\frac{t-T_n}{T}\right), \tag{20}$$

where the process (T_n) is defined by its initial value $T_0=0$, and a recurrent relationship $T_{n+1}=\inf\{m>T_n,\Sigma_{k=T_n}^{m}V_k=1\}$, where (V_k) is a sequence of independent random variables, each taking zero and one with respective probabilities 99.5% and 0.5%. The noise η is an AR(1) centered red noise

$$\eta_{t+1}=\cos(\theta)\,\eta_t+\sin(\theta)z_{t+1}, \tag{21}$$

with $\theta=0.77$ and z_t is a unit variance white noise. In Eq. (20) the function $1_{[a,b]}$ equals one inside the interval $[a, b]$ and zero elsewhere. In this example the dataset $D(t)$ is obtained by

$$D(t)=(1+\varepsilon\,\eta_t)S(t)+\eta_t, \tag{22}$$

where ε is a small parameter, normally $o(1)$, that reflects the weak nonlinear interaction between the signal and the noise. The case where the data is simply the sum of the signal and the noise corresponds to $\varepsilon=0$. In Fig. 2(a) is shown the signal $S(t)$ (dotted) and the data time series $D(t)$ (solid line) for $\varepsilon=0.5$. Note in particular the strong amplitude of the noise compared to that of the signal.

The covariances $\mathbf{C}_D$ and $\mathbf{C}_N$ of the data and the noise, respectively, are computed as in Eq. (19). The eigenspectrum of the matrix $\mathbf{C}_D'$ [Eq. (12)], which is identical to that of the detection matrix $\mathbf{C}_D\mathbf{C}_N^{-1}$, is shown in Fig. 2(b), which indicates a leading pair of ''equal'' eigenvalues with a break in the spectrum, the signature of the existence of an oscillation in the data. The first two signal patterns [Eq. (13)] corresponding to the leading pair of eigenvalues are displayed in Fig. 2(c), where the two sine waves in quadrature clearly show the periodic signal contained in the data. Finally, to verify the Gaussian behavior of the distribution of the test series, we show in Fig. 2(d) the histograms of the noise and the data. The normal distribution is also shown for comparison.

B. Case with non-Gaussian data: Regimelike behavior

As an application to the theory presented in Sec. IV, two examples are considered. The first one deals with a generated signal with regimelike behavior, while in the second example the classical Lorenz model [4] is considered. In the first example, the signal $S(t)$ has two periods T_1 and T_2 and switches randomly from one regime center to the other. At each fifth time step there is a probability of 20% that the system changes regime. In each regime the signal is an amplified sine wave with an e-folding time τ of 30 units. The regime centers $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are chosen to be symmetric with respect to the origin rather similar to the original Lorenz system [4] and the distance $2\delta=|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2|$ is a parameter that one can modify. The parameter δ is used for analyzing the sensitivity of the validity of the Gaussian model. The test series $D(t)$ is obtained using Eq. (22) but for the results we show the case for $\varepsilon=0$.

In Fig. 3 we show both the signal $S(t)$ [Fig. 3(a)] and the data $D(t)$ [Fig. 3(b)] when $T_1=5$, $T_2=7$ units, and $\delta=1$. For small values of δ, i.e., δ varying between 0 and 0.32, both techniques, i.e., the S/N ratio method and the Gaussian model of Sec. III, give similar (correct) results. This is particularly true when $T_1=T_2$, since there is only one frequency to detect. This means that for this range of values of the parameter δ, $0\leqslant\delta=\frac{1}{2}|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2|\leqslant 0.32$, there is no need to separate the data into clusters and the case can still be treated using the Gaussian model of Sec. III. In this case the frequencies in the first and the second pairs of eigenvectors, corresponding, respectively, to the first and the second pairs of eigenvalues, represent, respectively, the first and the second oscillations in the data.

Now, as the distance δ gets larger, $\delta>0.32$, the S/N algorithm fails (as shown below) to identify the correct frequency even when the two frequencies are identical, i.e., the signal contains only one frequency. This failure is easily understood since now the covariance matrix $\mathbf{C}_D$ no longer reflects the true structure of the data (refer to Fig. 1). In fact, Fig. 4 shows the noise and the data of Fig. 3 within the first two delay coordinates. In Fig. 4(b) the clusters are shown for $\delta=1.5$. Note that δ has been increased to 1.5 in Fig. 4(b) for the purpose of emphasizing the bimodality (non-Gaussian character) of the data. Figure 5(a) shows the two leading EOFs maximizing the S/N ratio. To show the sensitivity of the Gaussian model to changes in δ, we display in Fig. 5(b) the ratios (λ_0/λ_1) and (λ_1/λ_2) as a function of the parameter 2δ, where λ_0, λ_1, and λ_2 are the first three leading eigenvalues of the matrix $\mathbf{C}_D'$ [Eq. (12)]. Note that an oscillation corresponds approximately to values near to and (much) larger than 1 for the first and second ratios, respectively. Clearly this figure shows that for small δ ($\delta\leqslant 0.32$)

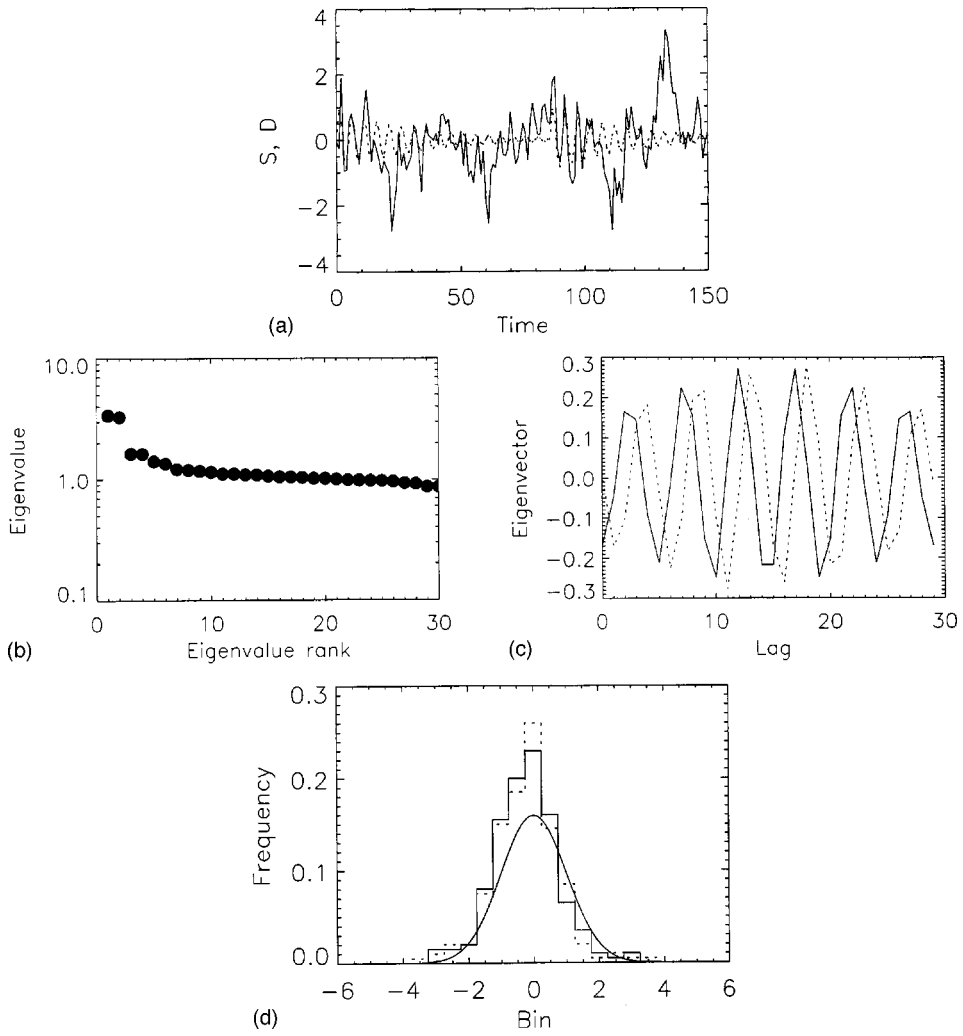

FIG. 2. (a) The signal $S(t)$ obtained by randomly generating damped sinusoidal bursts (dotted) and the test series $D(t)$ (solid) given by Eq. (22) with $\varepsilon = 0.5$; (b) the eigenspectrum of the transformed matrix $\mathbf{C}'_D$ [Eq. (12)] in decreasing order showing the leading pair corresponding to the oscillation in the data; (c) the first two signal patterns [Eq. (13)] corresponding to the leading pair of eigenvalues of $\mathbf{C}'_D$ and showing the correct oscillation; and (d) the histograms of the noise (solid) and the data $D(t)$ (dotted), respectively, along with the normal distribution for comparison.

the case can still be included in the Gaussian model, but as δ gets larger, the model is no longer valid and one has to resort to the non-Gaussian analysis. Note that this simple adequacy criterion, based on the generated univariate test series, cannot be applied to the complicated, multivariate general case and it is recommended, in general, to apply the mixture analysis to check whether the data at hand support multiregime behavior and then use the non-Gaussian approach. This is particularly important for atmospheric low-frequency variability given the regimelike behavior of the atmosphere on those scales.

Since the PDF-based algorithm takes into account the distribution of the data and the noise in phase space, we apply Eq. (17) to the same example with $\mathbf{C}_{D_1}$ ($\mathbf{C}_{D_2}$) being the

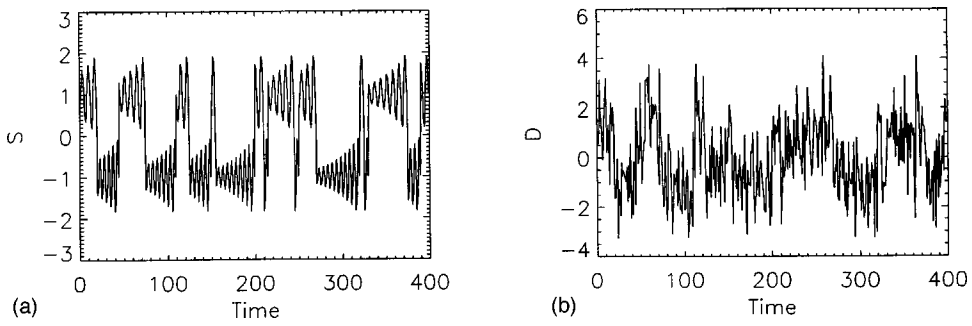

FIG. 3. (a) The signal $S(t)$ generated by initiating an amplified periodic signal switching between two regime centers, separated by $2\delta = 2$, with respective periods 5 and 7 units and (b) the time series $D(t)$ obtained by adding a red noise to the signal.

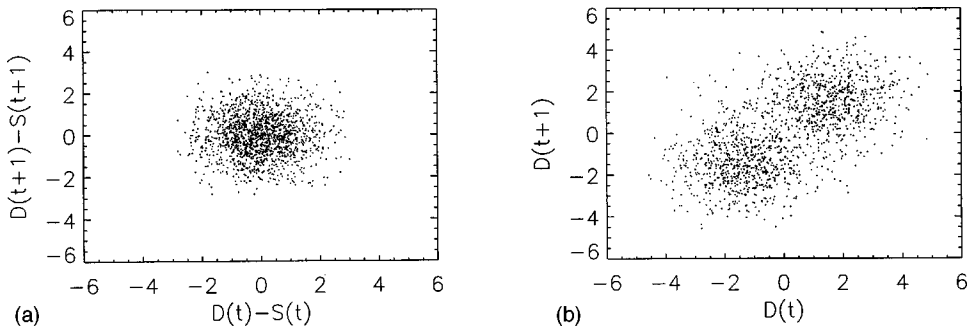

FIG. 4. Projection of the noise η_t (a) and the test series $D(t)$ (b) of Fig. 3 onto the first two delay coordinates. 1500 points have been used (b) we have chosen $\delta=1.5$ to emphasize the bimodality in the data.

covariance matrix [Eq. (19)] of the time series obtained simply by cutting through the original time series [Fig. 3(b)] by a horizontal line $h=h_0$, and analyzing the upper (lower) part separately. Several experiments have been carried out with varying h_0 and it is found that the method is quite robust to changes in h_0 over a wide range. As an example, for $-0.5 \leqslant h_0 \leqslant 0.5$, we always identify the correct signals. In Fig. 6 we show the eigenspectrum of $\mathbf{C}'_{D_1}$ [Fig. 6(a)] along with the eigenvectors given by Eq. (17) and corresponding to the two leading eigenvalues for the first regime [Fig. 6(b)]. A similar correct result is obtained for the second part of the time series (not shown). They both accurately identify the oscillations in the data.

The previous example was based on an artificial, purely probabilistic process. In order to involve a more realistic dynamical model with geophysical connection and with chaos and regime behavior, variants of the Lorenz model [4] are considered next. More precisely, the ''forced'' Lorenz

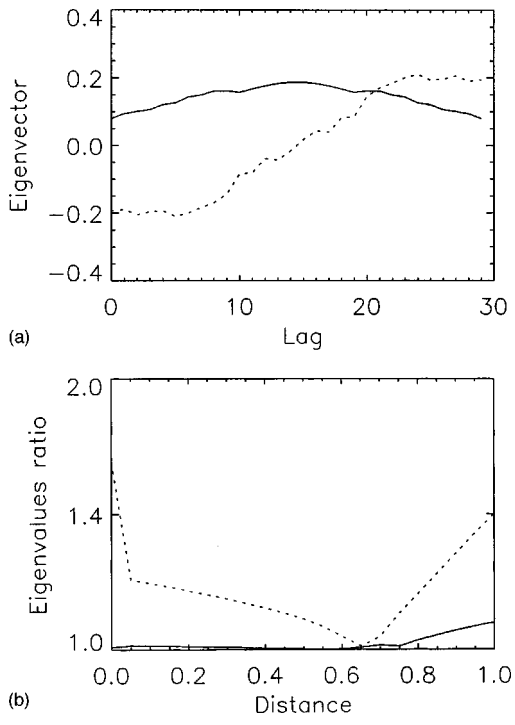

FIG. 5. (a) The two leading solutions maximizing the S/N ratio [Eq. (1)] applied to the global test series of Fig. 3 and (b) the ratios (λ_0/λ_1) (solid) and (λ_1/λ_2) (dotted) as a function of $2\delta=|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2|$. The vertical scale is logarithmic.

model [11] is the focus here. Compared to the original system, the forced version can introduce asymmetry into the regimes. The model equations are

$$\dot{x}=-\sigma(x-y)+f_0\cos\theta,$$

$$\dot{y}=-xz+rx-y+f_0\sin\theta, \tag{23}$$

$$\dot{z}=xy-bz.$$

In [11], this model has been used to interpret the climate in terms of a nonlinear dynamical system attractor with distinct regimes. The parameters σ, r, and b in Eq. (23) are the same as in the original Lorenz system with $(\sigma,r,b)=(10,30,8/3)$ and with $f_0=2.5$ as in [11]. The additional forcing $(f_0\cos\theta, f_0\sin\theta)$ can bias the model regimes by increasing the stability of one of the original model attractors, which therefore becomes occupied more often than the other regime rather similar to the ''extended'' version [28]. In fact, it has been shown in [11] that the angle θ of the forcing can modify the PDFs associated with the two regimes. It is also noted in [11] that the phase space position of the regime centroids does not change as the forcing rotates from one direction to another. In this paper we are interested in detecting the frequency of the oscillation spells in the system and we choose $\theta=\pi/4$.

System (23) is integrated using a Euler scheme with time step $dt=10^{-2}$. The system attractor within the x-z plane is shown in Fig. 7(a). For the analysis we choose to focus on the variable x since it displays regime behavior. The corresponding time series $x(t)$ is resampled every $\Delta t=10dt$ [Fig. 7(b)] in order to bring to a reasonable size the embedding window length. The resampled and rescaled (scaling factor of 0.2) time series is contaminated with red noise [Fig. 7(c)] and the filtering procedure as in the first example is performed.

As in the previous example, the result is found to be robust to changes in h_0. For example, for $-1.1 \leqslant h_0 \leqslant 1.1$, we always get the correct oscillation within the first regime, i.e., we do not need a precise boundary between the regimes. In fact, as in Fig. 5(b), Fig. 8(a) shows the ratios (λ_0/λ_1) and (λ_1/λ_2) as a function of h_0 in regime 1 (upper part) of the model. In particular, the figure illustrates that in order to estimate the signal within the first regime, the separation line can cross the domain of the second regime. This remains valid as long as the line does not get too close to the regime center. By looking towards the left side of Fig. 8(a), att he

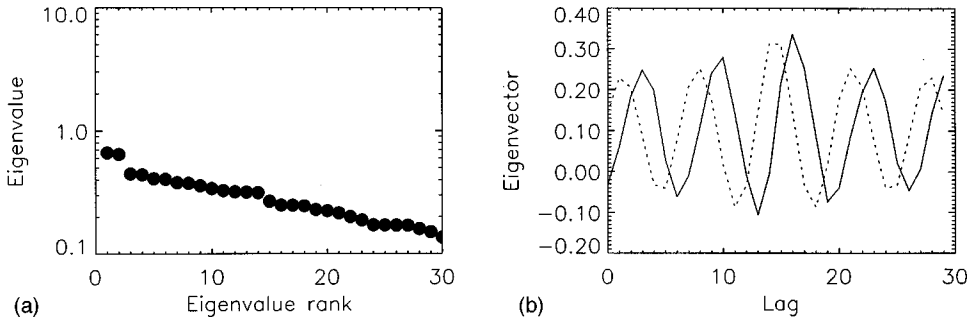

FIG. 6. (a) The eigenspectrum of $\mathbf{C}'_{D_1}$ and (b) the signal patterns [Eq. (17)], corresponding to the leading pair of eigenvalues and showing the oscillation in the first regime with $h_0 = 0.3$.

smallest (negative) values of h_0, i.e., considering the global time series, one can tell that the Gaussian model (Sec. III) fails to identify the signal. In fact, Figs. 8(b) and 8(c) show, respectively, the spectrum of $\mathbf{C}'_D$. [Eq. (12)] and the two leading patterns [Eq. (13)], i.e., when the whole test series is used at once. In Fig. 9, we show the spectrum of $\mathbf{C}'_{D_1}$ along with the first pair of eigenvector patterns [Eq. (17)] in the first regime ($h_0 = -0.5$). A similar oscillation is obtained in the second regime (not shown).

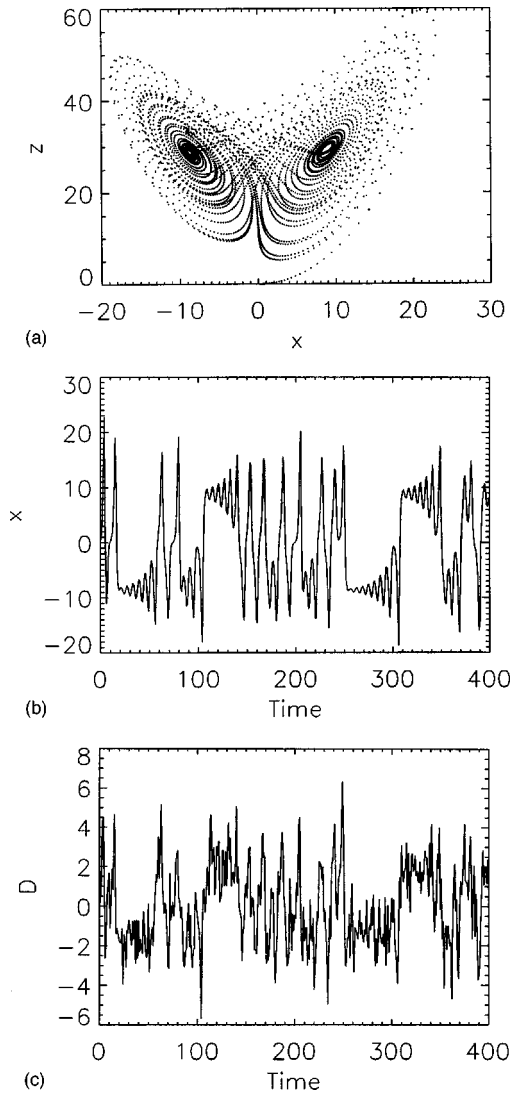

FIG. 7. (a) Trajectory of the ''forced'' Lorenz model within the *xz*-plane, (b) the time evolution of x sampled every tenth time step, and (c) the contaminated test series obtained by adding a red noise to the (resampled and rescaled) trajectory $x(t)$. A Euler numerical scheme with time step $dt = 10^{-2}$ is used for time integration.

Other experiments are performed where, for example, the data projected onto the delay phase space are split into clusters using the ''*k*-means'' technique, where the clusters become separated by a hyperplane [15]. In this latter case, and because we are dealing with embedded state vectors, only a subset of each cluster is used in signal detection. Again similar results are found with similar robustness to changes in the position of the hyperplane. Experiments with small nonzero ε and also using the original Lorenz model as well as the extended version [28] were also carried out, and it is found that the method successfully detects the correct signal.

VI. SUMMARY AND CONCLUSION

An approach has been developed to deal with the identification of regular signals contained in an intermittent regimelike behaving (noisy) system. The approach is based on the assumption that we are provided with an estimate of the noise. This assumption has been made because in climate analysis the internal variability noise of the climate system can be estimated using climate models. Also, in many other geophysical examples the noise can be reasonably represented by a colored noise with characteristics computed from the data. The approach is, of course, straightforward, applicable to noise-free or white-noise-contaminated data.

The approach has been first formulated within a Gaussian framework for the data and the noise, and aims to identify locations (signals) within the phase space that minimize the noise probability distribution on isosurfaces of the data probability distribution. The (PDF) signals are simply the eigenvectors of the detection matrix $\mathbf{C}_D\mathbf{C}_N^{-1}$ corresponding to the leading eigenvalues and are therefore dual of the signal-to-noise ratio maximizing vectors. The method is then extended to the non-Gaussian case by expressing the data by a mixture of Gaussians. The signal is then identified for each regime using the Gaussian approach.

The method has been successfully applied to synthetic univariate test series with and without regimes as well as to variants of the Lorenz system. In the single-channel case with regime behavior, the data is easily split into regimes as performed in Sec. V B, and it is shown that the approach is quite robust to reasonable (small) changes in the position of the line separating the regimes. In the multichannel case,

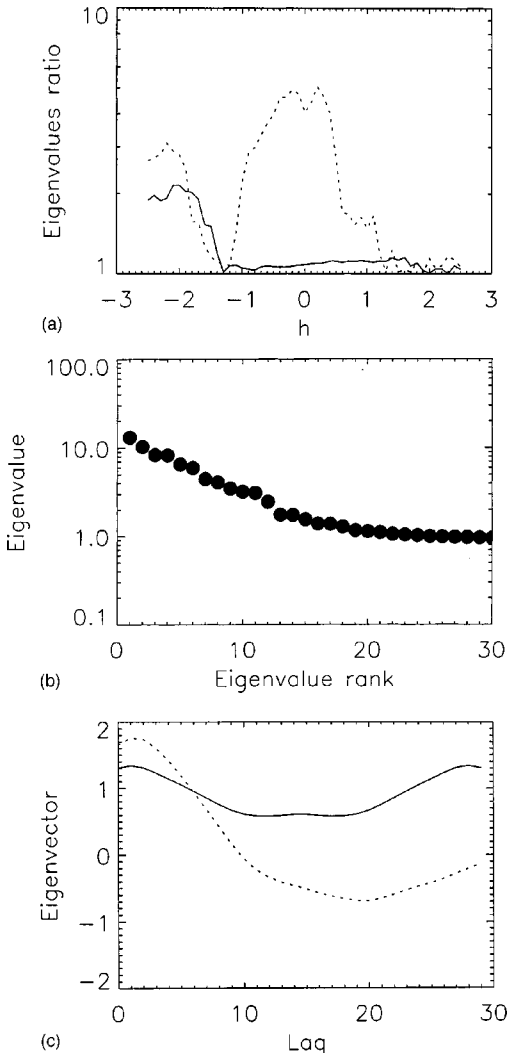

FIG. 8. (a) Ratios (λ_0/λ_1) (solid) and (λ_1/λ_2) (dotted) as a function of h_0 for the upper part (regime 1) of the forced Lorenz model, the vertical scale is logarithmic. (b) The spectrum of the matrix $\mathbf{C}_D'$ [Eq. (12)] and (c) the two leading patterns [Eq. (13)] of the global test series $D(t)$, i.e., corresponding to $h_0 \leqslant \min\{D(t)\}$.

however, the data have to be separated into (Gaussian) clusters and the expectation-maximization algorithm provides a powerful tool to diagnose the regimes and therefore identify the signals.

The method can be applied to atmospheric low-frequency variability to identify, for example, atmospheric signals (oscillations). A separation of the data into regimes is then required prior to the application of the Gaussian approach. In climate studies where the internal variability noise can be estimated from general circulation models, the method can be applied to diagnose, for example, the effect of external forcing on the climate system. This point is under investigation and will be presented elsewhere.

ACKNOWLEDGMENTS

I would like to thank Dr. M. R. Allen for his constructive comments on an earlier version of the manuscript. This work was funded by the UK University Global Atmospheric Modelling Project (UGAMP). Computations were performed on local Sun stations at AOPP.

APPENDIX

In this short appendix we consider the case where the noise is Gaussian with PDF p_N, and the data a mixture of two Gaussians with PDF p_D given by Eq. (16). We suppose in this Appendix that p_D is a bimodal distribution, i.e., p_D has two local maxima at, say, $\boldsymbol{\mu}_1^0$ and $\boldsymbol{\mu}_2^0$. Note that bimodality does not necessarily mean that the Gaussians are well separated, but we adopt this assumption just because it allows us to get to the heart of the problem, i.e., to work over each Gaussian component forming the data, and the result, as illustrated schematically in Fig. 1, is independent of this assumption.

As can be inferred from Fig. 4(b), for example, the region between the two maxima is not the right location to look at, since at that point p_N is at a maximum while p_D is at a ''local'' minimum. Instead, we have to look for signal patterns in the neighborhood of the two bumps of p_D by following the same analysis of Sec. III. In fact, the only difference is the structure of $S_d = p_D^{-1}(\delta_d)$, which is now nonconnected and therefore the minimization should naturally be carried out over each connected part of S_d, i.e., around each bump of p_D. Now it can be seen that

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_1^0 + O(|\boldsymbol{\varepsilon}| e^{-1/2\boldsymbol{\varepsilon}^T \mathbf{C}_{D_2}^{-1} \boldsymbol{\varepsilon}}), \tag{A1}$$

where $\boldsymbol{\varepsilon} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. Therefore at a precision

$$O(|\boldsymbol{\varepsilon}| \exp[-1/2\boldsymbol{\varepsilon}^T \mathbf{C}_{D_2}^{-1} \boldsymbol{\varepsilon}]),$$

one can optimize around the centres $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ respectively, which yields, as in Sec. III, the minimization of the noise probability distribution p_N for any given data probability dis-

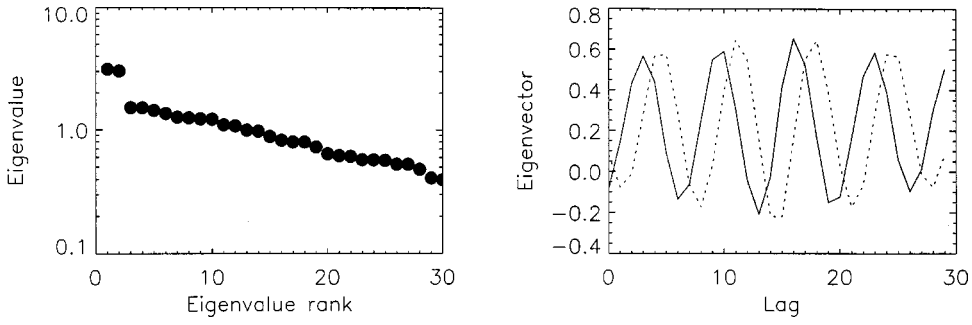

FIG. 9. As in Fig. 6 but for regime 1 of the forced Lorenz model ($h_0 = -0.5$).

tribution value $p=(2\pi)^{-m/2}|\mathbf{C}_{D_i}|^{-1/2}\exp[-1/2(\mathbf{v}-\boldsymbol{\mu}_i)^T\mathbf{C}_{D_i}^{-1}(\mathbf{v}-\boldsymbol{\mu}_i)]$, or similarly $d^2=(\mathbf{v}-\boldsymbol{\mu}_i)^T\mathbf{C}_{D_i}^{-1}(\mathbf{v}-\boldsymbol{\mu}_i)$, i.e.,

$$\min(2\pi)^{m/2}p_N=\frac{1}{|\mathbf{C}_N|^{1/2}}e^{-1/2}(\mathbf{v}-\boldsymbol{\mu}_i)^T\mathbf{C}_N^{-1}(\mathbf{v}-\boldsymbol{\mu}_i),$$

$$d^2=(\mathbf{v}-\boldsymbol{\mu}_i)^T\mathbf{C}_{D_i}^{-1}(\mathbf{v}-\boldsymbol{\mu}_i) \tag{A2}$$

for $i=1$ or 2.

The problem is therefore transformed into a minimization over each cluster (or Gaussian) separately and consequently there is a solution pattern for each cluster. In (A2) the noise density has been translated to the regime center for consistency with Sec. III. The problem is then identical to Eqs. (6), (8), and (11) and the solution is simply given by

$$\mathbf{v}_i=\mathbf{C}_{D_i}^{1/2^T}\boldsymbol{\xi}_0^{\,i}, \tag{A3}$$

where $\boldsymbol{\xi}_0^{\,i}$ is the leading eigenvector of $\mathbf{C}'_{D_i}=\mathbf{C}_{D_i}\mathbf{C}_N^{-1}\mathbf{C}_{D_i}^{1/2^T}$ corresponding to the ith cluster. The conclusion is that once the data have been separated into clusters, the signal detection procedure can be carried out as in Sec. III but over each individual cluster.

[1] K. Fraedrich, J. Atmos. Sci. **43**, 419 (1986).

[2] D. S. Broomhead and G. King, Physica D **20**, 217 (1986); *Nonlinear Phenomena and Chaos*, edited by S. Sarkar (Hilger, London, 1986), p. 113; R. Vautard, P. Yiou, and M. Ghil, Physica D **58**, 95 (1992).

[3] M. R. Allen and L. A. Smith, Phys. Lett. A **234**, 419 (1997).

[4] E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).

[5] J. G. Charney and J. G. DeVore, J. Atmos. Sci. **36**, 1205 (1979).

[6] M. Ghil and R. Vautard, Nature (London) **350**, 324 (1991).

[7] W. C. Thacker, Tellus, Ser. A **48A**, 584 (1996).

[8] K. Fukunaga, *Introduction to Statistical Pattern Recognition* (Academic, New York, 1972).

[9] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, England, 1988).

[10] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes* (Cambridge University Press, Cambridge, England, 1992).

[11] T. N. Palmer, J. Clim. **12**, 575 (1999).

[12] P. Prunet, J. Thiépaut, and V. Casset, Q. J. R. Meteorol. Soc. **124**, 211 (1998).

[13] J. Barkmeijer, M. V. Gijzen, and F. Bouttier, Q. J. R. Meteorol. Soc. **124**, 1695 (1998).

[14] D. J. Hand, *Construction and Assessment of Classification Rules* (Wiley, New York, 1997).

[15] B. S. Everitt, *Cluster Analysis* (Arnold, London, 1993); J. MacQueen, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, 1965* (University of California Press, Berkeley, 1967), Vol. 1.

[16] G. L. G. Sleijpen, J. G. L. Booten, D. R. Fokkema, and H. A. van der Vorst, BIT **36**, 595 (1996).

[17] W. C. Thacker and R. Lewandowicz, J. Clim. **9**, 1942 (1996).

[18] J. R. Heirtzler, G. O. Dickson, E. M. Herron, W. C. Pittman II, and X. LePichon, J. Geophys. Res. **73**, 2119 (1968); T.J . Ulrych, Nature (London) **235**, 218 (1972).

[19] C. E. Leith, J. Appl. Meteorol. **12**, 1066 (1973).

[20] E. N. Lorenz, Meteorol. Monogr. **8**, 1 (1968).

[21] B. D. Anderson and J. B. Moore, *Optimal Filtering* (Prentice-Hall, Englewood Cliffs, NJ, 1979).

[22] G. J. McLachlan and K. E. Basford, *Mixture Models: Inference and Applications to Clustering* (Dekker, New York, 1988); D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions* (Wiley, New York, 1985).

[23] K. Haines and A. Hannachi, J. Atmos. Sci. **52**, 2444 (1995).

[24] J. H. Wolfe, Multiv. Behav. Res. **5**, 329 (1970).

[25] A. P. Dempster, N. M. Laird, and D. B. Rubin, J. Roy. Stat. Soc. B **39**, 1 (1977).

[26] B. S. Everitt and D. J. Hand, *Finite Mixture Distributions* (Chapmann and Hall, London, 1981).

[27] F. Takens, in *Dynamical Systems and Turbulence*, edited by D. A. Rahd and L.-S. Young, Lecture Notes in Mathematics, Vol. 898 (Springer-Verlag, Berlin, 1981).

[28] F. Molteni, L. Ferranti, T. N. Palmer, and P. Viterbo, J. Clim. **6**, 777 (1993).