Robust method for periodicity detection and characterization of irregular cyclical series in terms of embedded periodic components

P. P. Kanjilal,* J. Bhattacharya,† and G. Saha
Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur 721302, India
(Received 8 May 1998; revised manuscript received 26 October 1998)

A method for periodicity detection is proposed where unlike available methods a periodic component is characterized in terms of three basic *periodicity attributes*: the periodicity (or period length), the periodic pattern, and the scaling factors associated with the successive nearly repetitive segments. A scheme is proposed for subsequent successive detection and extraction of such (hidden) periodic or nearly periodic components constituting an irregular cyclical series. To our knowledge, the proposed decomposition is much more powerful in terms of information content and robustness than the presently available tools based on Fourier decomposition. Through the analysis of a variety of natural, experimental, and simulated data series, it is shown that the features of the periodicity attributes of the embedded periodic components can lead to a meaningful characterization of an irregular series in a new perspective. [S1063-651X(99)00404-3]

PACS number(s): 05.45.−a

I. INTRODUCTION

Any periodic or nearly periodic series [like electrocardiogram (ECG) [1,2]] can be precisely defined in terms of three basic features or *periodicity attributes*, namely, *the periodicity* (or period length), *the pattern* over the successive repetitive segments, and *the scaling factors* associated with the repetitive pattern segments. A real life irregular series may comprise a number of such components. Conventional analytical tools [3] based on the Fourier model can provide decomposition of an irregular series into sinusoidal components with constant scaling only for each repetitive segment, and hence lack in meaningfulness as far as individual nonsinusoidal components of a real life series are concerned. This paper seeks to make a contribution by proposing a method for the detection and subsequent separation of periodic components embedded in an irregular series, where the individual components *may not be sinusoidal* and the repetitive successive segments may be scaled differently; the characteristic features or periodicity attributes of the components are shown to provide a platform for the characterization of irregular cyclical series.

The motivation for the proposed paper is basic and vast as follows. Bounded yet cyclical processes are ubiquitous in nature as well as in man-made systems, e.g., ECG [1,2] and electroencephalogram (EEG) [4] signals, the white blood cell count in a patient of leukaemia [5], the childhood epidemic phenomenon [6], the oscillation in global temperature time series [7], the solar activity as reflected in sunspot numbers [8], the mass extinction activity [9], the intensity of a variable dwarf star [10], the light intensity pulsations of laser [11], the electrical power load pattern in urban areas [12], etc. In many cases (like ECG), the signals are nearly periodic with characteristic nearly repetitive patterns, whereas in others (e.g., EEG) there is usually no repetitive pattern. Analysts have shown that some irregular cyclical series are composed of a number of components (e.g., the yearly sunspot number series has been shown to comprise two components or three components [13–14]). In some cases, the decomposition of a series into components can be directly meaningful. For example, from the composite ECG signal obtained from the abdominal lead of an expectant mother both fetal ECG and maternal ECG components can be obtained [2]; here, since the two components have overlapping frequency bands, conventional Fourier decomposition cannot be applied for separation of two nearly periodic components from a single channel signal.

Thus for a cyclical series, the basic characteristics that stand out are that the signals are bounded in magnitude; the cyclicity shown may be regular, or may be irregular where the degree of irregularity can vary. For a *strictly periodic* series (which may or may not be sinusoidal), all three periodicity attributes remain unchanged. In case of a real life cyclical series (e.g., ECG, far infrared ammonia laser intensity), all three attributes may vary to some extent, where the nature of variation may relate to the characteristics of the underlying process [15]); we use the broad term ''nearly periodic'' for such processes. Thus, an irregular cyclical series can belong to three broad categories: (i) series which can be decomposed into a number of sinusoidal or nonsinusoidal constituents (e.g., the sunspot series), (ii) series that show only one periodic component, whose periodicity attributes may remain stationary only locally but may vary globally (e.g., the far infrared (FIR) laser series), (iii) series that contain no periodic components that are at least locally stable (e.g., heart rate variability series for healthy subjects [16]).

The constraint in analyzing such irregular cyclical series through conventional techniques [3] (such as power spectral density, periodogram etc.) is that it is assumed that the series can be decomposed into multiple components, where only sinusoidal pattern is permissible for each component and the successive periodic segments of each component cannot be scaled differently. As a result, the decomposed components lack physical significance. Again, the cyclic theory of chaos

*Electronic address: ppk@ece.iitkgp.ernet.in
†Present address: Max Planck Inst. Fur Komplex System, Dresden D01187, Germany.

[17] offers different approaches for the detection of periodic orbits, where periodicity in terms of repeating occurrences of points are detected; in the case of chaotic processes, such periodic orbits are found to be unstable. Here also the available insight into the nature of the individual components constituting the irregular cyclical series is limited.

The basic questions that need addressing are: (a) can unknown periodic or nearly periodic components constituting an irregular series be detected and extracted, where the components can be nonsinusoidal and may have overlapping frequency bands, and (b) can the periodicity attributes of the single or multiple components detected offer newer understanding in terms of the characterization of the underlying process? This paper attempts to address these issues.

The proposed method of periodic decomposition is detailed in Sec. II. The concept of periodicity spectrum (or *p* spectrum) is introduced, and it is shown how *p* spectrum can be used to detect the presence of a periodic component hidden in any data series. The procedure for successive detection and extraction of the constituent periodic components is discussed. A corollary to the above procedure is the *long-term prediction* of the composite series through the periodic prediction of individual component series. The proposed decomposition scheme is based on singular value decomposition (SVD), a robust algebraic tool, which has been widely used for solving least-squares estimation problems [18,19], for modeling and prediction [20–22]. The results are detailed in Secs. III and IV. Section III demonstrates the superiority of the proposed method against Fourier decomposition. Since Fourier decomposition provides only a partial picture of the underlying process because of the inherent limitation of each component being sinusoidal with constant scaling throughout, a direct comparison is not possible; however, it is shown how Fourier decomposition can be misleading even for periodicity detection, when noise and signal bandwidths overlap or when the constituent periodic components have overlapping frequency bands, whereas the *p* spectrum remains relatively unaffected. In Sec. IV, the potential in the proposed method for providing a platform for characterizing irregular cyclical series is demonstrated through four examples: (1) the natural series of sunspot numbers, (2) the experimental series of FIR-ammonia laser intensity pulsations, and (3) the simulated chaotic series of the Mackey-Glass equation and (4) the logistic map process.

II. THE PROPOSED METHOD

A. Data configuration and singular value decomposition

Let the series $\{x(k)\}$ be configured into an $m\times n$ matrix $\mathbf{A}_n$:

$$\mathbf{A}_n=\begin{bmatrix} x(1) & x(2) & \cdots & x(n) \\ x(n+1) & x(n+2) & \cdots & x(2n) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ x((m-1)n+1) & . & \cdots & x(mn) \end{bmatrix}. \quad (1)$$

SVD [18–21] of $\mathbf{A}_n$ produces $\mathbf{A}_n=\mathbf{USV}^T$, where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices: $\mathbf{UU}^T=\mathbf{U}^T\mathbf{U}=\mathbf{I}$, $\mathbf{VV}^T=\mathbf{V}^T\mathbf{V}=\mathbf{I}$; $\mathbf{S}=\mathrm{diag}(s_1,s_2,...,s_r:\mathbf{0})$, $r=\min(m,n)$ and $s_1\geqslant s_2\geqslant\ldots s_r\geqslant 0$. The SVD provides the orthonormal basis of the range and the null space of the matrix $\mathbf{A}_n$. The columns of $\mathbf{U}$ $(=[\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_m])$ (i.e., the left singular vectors) corresponding to nonzero diagonal elements of $\mathbf{S}$ span the range, and the columns of $\mathbf{V}$ $(=[\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n])$ (i.e., the right singular vectors) corresponding to zero diagonal elements of $\mathbf{S}$ are an orthonormal basis of the null space of $\mathbf{A}_n$. The number of nonzero singular values (s_i) gives the rank of $\mathbf{A}_n$.

In the present context, SVD offers some unique advantages in connection with the assessment of embedded periodicity in $\{x(k)\}$, as follows.

(i) For a *strictly periodic* $\{x(k)\}$ with period length N, i.e., $x(k)=x(k+N)$, $\mathbf{A}_n$ will be a strictly rank one matrix if row length $n=N$. Here, s_1 is nonzero but $s_2=\ldots=s_r=0$; $s_1/s_2=\infty$. The vector $\mathbf{v}_1$ represents the *periodic pattern* of the signal normalized to a unit vector. The successive elements of the vector $\mathbf{u}_1 s_1$ represent the *amplitude scaling factors* of successive pattern segments; $\{x(k)\}$ being perfectly periodic, the elements of $\mathbf{u}_1 s_1$ will all be the same.

(ii) If $\{x(k)\}$ is *nearly periodic* with fixed period length N but $x(k)\neq x(k+N)$, two possibilities may arise: (a) $\{x(k)\}$ has same repeating pattern of length N but with different scaling over different periods. Still $\mathrm{rank}(\mathbf{A}_n)=1$ and $s_1/s_2=\infty$; $\mathbf{v}_1$ represents the pattern but the elements of $\mathbf{u}_1 s_1$ now vary according to the scaling associated with the rows of $\mathbf{A}_n$. We use the term ''periodic'' to include such phenomenon also. (b) $\{x(k)\}$ has nearly repeating patterns with different scaling factors over successive segments. The matrix $\mathbf{A}_n$ can now be a full rank matrix but with s_1 much larger compared to the rest of the singular values, i.e., $s_1/s_2\gg 1$.

Remarks. (1) A dominant first singular value for any $m\times n$ matrix $\mathbf{A}_n$ is indicative of the presence of a strong periodic component (of period length n) in $\{x(k)\}$, given by the rows of $\mathbf{u}_1 s_1 \mathbf{v}_1^T$. (2) SVD is the most robust null-space detector of a matrix compared to other eigen decompositions [22]; it is numerically well conditioned and can be computed in a numerically stable way. The efficiency of SVD in noise separation and in estimating embedding dimension is well established [23]. (3) The present configuration of the data matrix (1) is different from the conventional form of the trajectory matrix formed of lag vectors or states [23], where all the states (in sufficient embedding dimensional space) are considered, whereas only the states, which are n sequences apart from each other, are considered in Eq. (1), with no overlapping of data elements across the rows of the matrix.

B. The proposed method of periodicity detection

$\{x(k)\}$ is configured into $m\times n$ matrix $\mathbf{A}_n$ with varying row length n and SVD of $\mathbf{A}_n$ is performed. The spectrum of the ratio of first two singular values s_1/s_2 vs. row length (n) is called the *periodicity spectrum* or ''*p* spectrum'' [24], which will show repetitive peaks at $n=iN$ (where i is a positive integer), if there is any embedded periodic component of periodicity N in $\{x(k)\}$, and this serves as periodicity detection.

Here, the repetitive peaks in the *p* spectrum are formed at $n=iN$ as the presence of a periodic component of periodicity N tends to increase the closeness to rank oneness of $\mathbf{A}_n$ (when configured as $n=iN$), which is reflected in the increase in the value of s_1 and decrease in the value of s_2. The *p* spectrum is detrended for improving the readability of the peaks [25].

Remarks. (i) $\{x(k)\}$ has to be at least $4N$ long for the periodicity of length N to be detectable using the p spectrum, as at least two peaks (i.e., at $n=N$ and $n=2N$) are necessary. (ii) The absence of a set of repetitive peaks in the p spectrum of $\{x(k)\}$ confirms the absence of a periodic component in the series. (iii) Many ramifications of the proposed periodic detection scheme are possible, for example, the data may be nonlinearly transformed prior to periodicity detection; the appropriateness of the transformation may be assessed through the improvements in the sharpness of the peaks in the p spectrum. (iv) For automated detection of periodicity, any suitable index $I(n)$ may be defined expressed as a function of the strengths of s_1/s_2 at jn row lengths, where j is a positive integer (see, for example, [26]). (v) The computational load for the proposed method is expected to be moderate. It is not necessary to compute the complete SVD. Computation of singular values for an $m \times n$ matrix involves $4mn^2$-$4m^3/3$ flop counts [18].

C. Successive extraction of periodic components

Following the detection of a periodic or nearly periodic component (of periodicity N), $\{x(k)\}$ is configured into $\mathbf{A}_n$, (with $n=N$), for extraction of the concerned period. The best rank-1 approximation of $\mathbf{A}_N$ in least squares (LS) sense is given by $\mathbf{u}_1 s_1 \mathbf{v}_1^T$ [18–19]. Here [27], $\mathbf{v}_1$ represents the pattern over the periodic segments of the extracted component of periodicity N; the successive elements of the vector $\mathbf{u}_1 s_1$ (which is modeled as the series $\{g(k)\}$ as in Sec. II E) will give the scaling factors for the successive periodic segments. The time series defined by $\mathbf{u}_1 s_1 \mathbf{v}_1^T$, being periodic with period length N, is the LS estimation of the periodic component present in $\{x(k)\}$ having maximum energy $(=s_1^2)$.

The matrix $[\mathbf{A}_N - (\mathbf{u}_1 s_1 \mathbf{v}_1^T)]$ can be made into a residual series $\{x_r(k)\}$. The p spectrum of $\{x_r(k)\}$ will show the presence of additional dominant periodic component (if any), which can be extracted the same way as above, and the process is repeated. The extraction stops when the p spectrum does not show any repetitive peaks.

Thus the method leads to periodic decomposition through the successive extraction of periodic components from the original series; an individual component has fixed period length and periodic pattern, which is not necessarily sinusoidal.

For an irregular series, if no globally stable periodicity is detectable through the p spectrum, presence of periodicity over shorter local data segments may be detected using a moving data window as discussed in Sec II F.

D. The final model

A data series $\{x(k)\}$ with multiple components can be modeled as

$$\{x(k)\}=\{g_{1\mathbf{i}}\nu_1(k_1)\}+\{g_{2\mathbf{j}}\nu_2(k_2)\}+\cdots, \qquad (2)$$

where i is the period index of first component, j is the period index of second component, $k_1=1,...,N_1$, where N_1 is the periodicity of first component, $k_2=1,...,N_2$, where N_2 is the periodicity of second component, g_{1i} is the scaling factor for 1st component in ith periodic segment, g_{2j} is the scaling factor for 2nd component in jth periodic segment, $\nu_1(k_1)$ are elements of the pattern of the first component, $\nu_2(k_2)$ are elements of the pattern of the second component, $[\nu_1(1),\ldots,\nu_1(N_1)]=\mathbf{v}_1^T$, the first right singular vector of data matrix $\mathbf{A}_{N_1}$, $[\nu_2(1),\ldots,\nu_2(N_2)]=\mathbf{v}_2^T$, the first right singular vector of data matrix $\mathbf{A}_{N_2}$ generated from $\{x_1(k)\}$ and $\{g_{1i}\nu_1(k_1)\}$ is a component series with periodicity N_1; the successive segments of the series have the N_1-long pattern $\mathbf{v}_1^T$, which are scaled by g_{11}, g_{12}, etc.

E. Long-term prediction

The long-term prediction of each of the extracted periodic components is performed. The overall prediction is obtained by adding up the predicted components. For any periodic series, $g(k+i|k)$, the i step ahead prediction of $\{g(k)\}$, will lead to i period ahead prediction $g(k+i|k)\mathbf{v}_1^T$, with the assumption that pattern $\mathbf{v}_1$ remains unchanged over the predicted horizon (the suffixes as in Eq. (2) are omitted for clarity). The modeling is performed as follows: (i) First consider the linear model for the scaling factor series for one component:

$$g(k)=a_1 g(k-1)+a_2 g(k-2)+\cdots+a_r g(k-r). \qquad (3)$$

The best set of r_1 $(\leqslant r)$ lagged variables is selected using modified QR with column pivoting factorization (m-$\mathrm{QR}_{\mathrm{cp}}$ factorization [28]), with minimum C_p statistic [29], as explained in the Appendix. (ii) From the selected r_1 variables, square and bilinear variables are generated [30]; it is expected that incorporation of these additional variables in the description of variables along with the original r lagged variables $g(k)$ in Eq. (3) will improve the model. (iii) The final set of variables are determined using m-$\mathrm{QR}_{\mathrm{cp}}$ factorization, with minimum C_p statistic, and thus an optimal linear in the parameter but nonlinear in the variables model is developed. (iv) The parameters of the model are estimated using the LS estimation, and $g(k+i|k)$ predictions are produced. (v) The individual predicted components are added up to give the overall prediction [31].

Remark. The unique advantage of the proposed approach of modeling and prediction is that the conventional i step ahead prediction $\{g(k+i|k)$, etc.$\}$ is rendered iN step ahead prediction.

F. Detection and extraction of periodic component using moving data window

In Secs. II B and II C the detection and the extraction of the periodic component(s) was assumed to be over the entire data set. But for nonstationary data series, the periodicity attributes are dynamic in nature, and all three periodicity attributes may vary throughout the process. To accommodate such nonstationarity, a moving data window is considered as follows. The data series $\{x(k)\}$ is divided into overlapping data segments, referred to as data windows, as shown in Fig. 1. Here, $m=$ a parameter determining the length of one window, $N_1=$ the periodicity in the first data window, and $N_2=$ the periodicity in the second data window, etc.

In general, if $N_i=$ the periodicity in the ith data window, then the length of the $(i+1)$th data window is mN_i. Two data windows (say, ith and $(i+1)$th thus overlap over

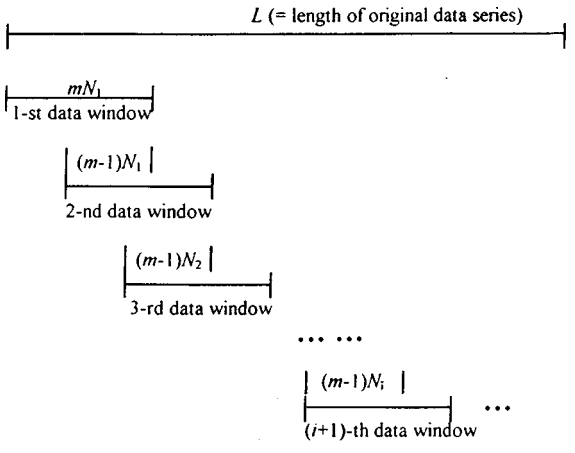

FIG. 1. The schematic diagram of the moving data window. There are two variations of this algorithm: (i) moving data window with fixed period, and (ii) moving data window with varying periods. In the former case, the periodicity is globally stable but the pattern and the scaling factors associated with each window can vary. In the latter case, p spectrum is performed for each data window; all three periodicity attributes can vary over the total data set.

$(m\text{-}1)N_i$ data points. In each data window (say, jth window), dominant periodicity (say, N_i) is detected through the p spectrum and the periodic component associated with the detected periodicity is determined; only the first period (of length N_i) from this periodic series is extracted, and this part of the series is truncated. The whole procedure is repeated for the next data window. The obtained extracted periods from the successive data windows are stitched together to form the entire estimated component. This estimated component is actually composed of different segments having different periodicity, as well as different pattern and scaling factors associated with the pattern. The estimated series can resemble one being generated by an oscillator (or periodicity generator) with time-varying periodicity attributes.

For some processes, the periodicity (N) may be globally stable, while the associated pattern can vary. This is a special case, where $N_1=N_2=\ldots=N_i=\ldots=N$, and may be termed as *moving window with fixed periodicity*.

In this paper, we use $m=5$, unless otherwise stated, which is not a limitation.

III. COMPARATIVE PERFORMANCE ASSESSMENT AGAINST FOURIER DECOMPOSITION

As argued in Sec. I, the fundamental difference between the proposed method of periodicity detection and the Fourier decomposition based approaches is that while the latter is confined to components with sinusoidal patterns and constant scaling for the successive repetitive segments, the former does not have any such restriction, and hence may be considered to provide a generalized decomposition. In this section, first the results with p spectrum are discussed. Next, through some simulation studies, it is shown how Fourier decomposition may lead to erroneous conclusions.

A. The *p* spectrum of some processes

The yearly averaged sunspot series is widely known to contain a dominant periodic component of periodicity of 11 years [13,14]. The features of this series are studied in detail in Sec. IV A. The p spectrum for the sunspot series (over 1700 to 1938) in Fig. 4(a) shows repetitive peaks at $(n=)$ integer multiples of 11, validating the presence of the most dominant component of periodicity of 11 years.

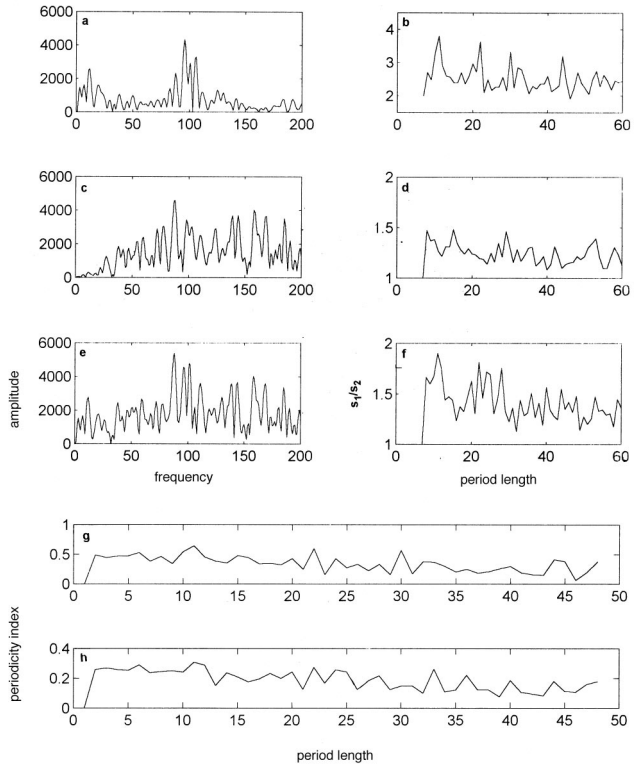

FIG. 2. (a) FFT of the sunspot series (1700–1988), (b) p spectrum of the same series, (c) FFT of the colored noise generated, (d) p spectrum of the colored noise signal, (e) FFT of the contaminated sunspot series where the strongest peak is different from that of Fig. 2(a), (f) p spectrum of the contaminated sunspot series, showing the prime periodicity of 11 years, (g) the periodicity index profile $I(n)$ for the original sunspot series (derived from the p spectrum [26]), and (h) for the contaminated sunspot series; in both cases, the strongest peak is at 11.

Again, as detailed in Sec. IV C 2, the one-dimensional logistic map [32] $x(k+1)=rx(k)(1-x(k))$, which is usually a chaotic process, exhibits stable periodicity for specific values of r within 3 and 3.57. The p spectrum [Figs. 13(a) and 13(b)] show distinct periodicity of 32 for $r=3.5687$, and periodicity of 2 for $r=3.6786$; the series of distinctly repetitive peaks in p spectrum (unlike the sunspot case) implies the presence of only *one* prime periodic component.

B. Study with white noise contamination

The sunspot series is reconsidered here with noise contamination. The strongest peak in fast Fourier transform (FFT) also shows the most dominant peak corresponding to the prime periodicity [Fig. 2(a)]. However, when contaminated by white noise, beyond 220% noise contamination (with respect to the signal in terms of energy), FFT fails to correctly detect this periodicity, whereas the p spectrum can withstand up to 300% contamination and yet detect 11 years as the prime periodicity.

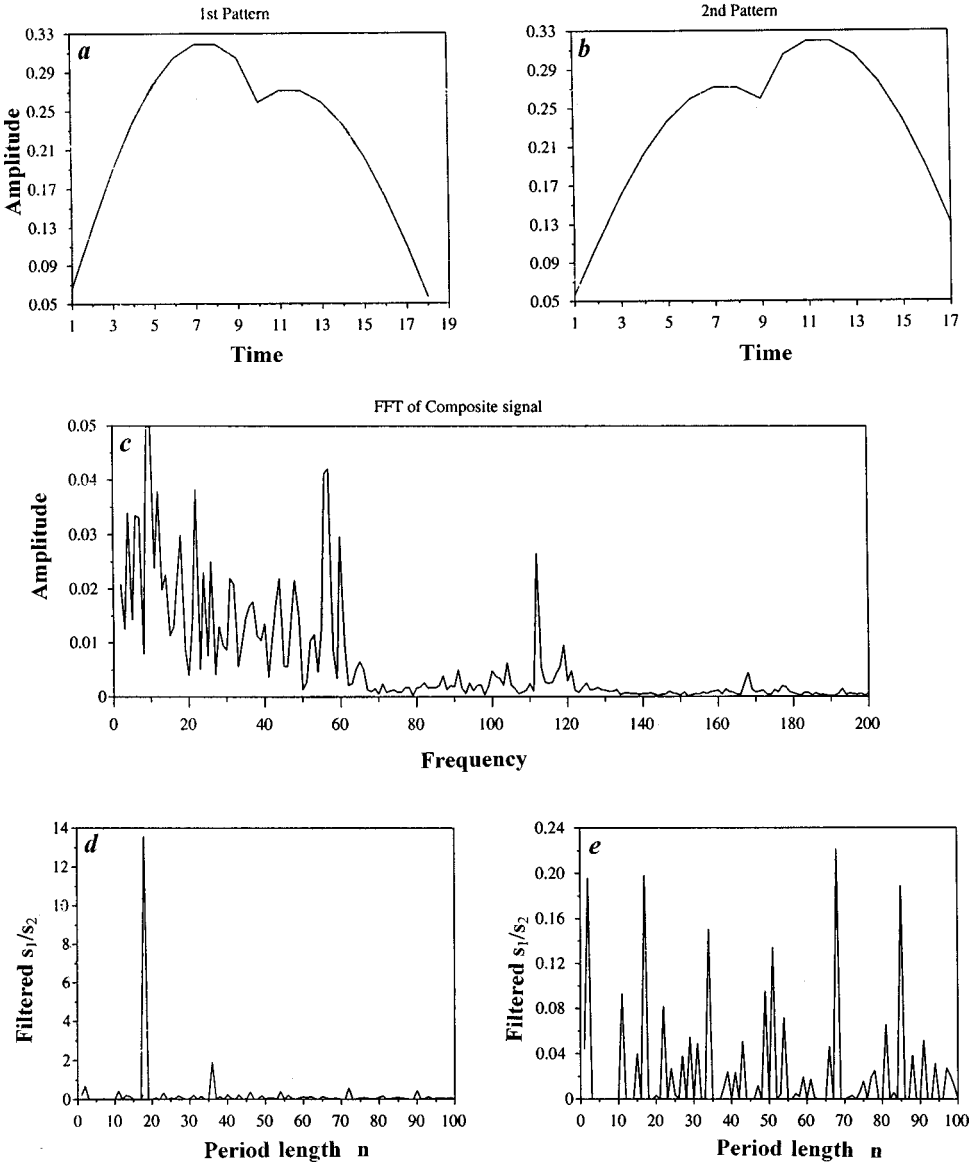

FIG. 3. (a) and (b) The pattern profiles of the periodic components of periodicity 18 and 17, respectively, (c) FFT of the noise contaminated composite series, which does not reveal the periodicity of 18 and 17 clearly, (d) the p spectrum of the same composite series, showing repetitive peaks corresponding to periodicity of 18, and (e) the p spectrum of the residual series (after extraction of component of periodicity 18) showing peaks corresponding to periodicity of 17.

C. Study with colored noise contamination

If colored noise has a strong peak at a frequency different than that due to the embedded periodic component in the signal, the strongest colored noise frequency is picked up by FFT as the prime component, whereas the p spectrum remains unaffected as follows: (a) The sunspot series is contaminated by band-pass filtered white noise. First, the white noise series is passed through a low-pass filter $[(1-p_1)/(1-p_1q^{-1})]$ with single pole p_1 at 0.8; the output of the filter is subtracted from the original series. The resulting series is passed through a filter with a pole p_1 at 0.2. FFT of this colored noise shows a peak at frequency of 88 [Fig. 2(c)], while the sunspot series has the prime peak at 96 [Fig. 2(a)] (i.e., corresponding to a periodicity of $1024/96=10.91$). With 239% of this colored noise, FFT [Fig. 2(e)] of the composite signal shows 12 (years) as the most dominant periodicity (as the peak is at 88) in the contaminated sunspot series. The p spectrum of the sunspot series, the noise signal, and the contaminated sunspot series are shown in Figs. 2(b), 2(d), and 2(f), respectively, where p spectrum still shows 11 years as the prime periodicity irrespective of the contamination, which is also confirmed by the strongest peak at 11 in the periodicity index [26] profile shown in Figs. 2(g)–2(h). (b) The sunspot series is contaminated with a noise signal given by [33]

$$\eta(k)=870(\alpha/\pi)^{0.25}\exp(\alpha k^2/2+j\beta k^2/2+j\omega k),$$

where $\alpha=0.000005$, $\beta=0.00035$, $\omega=101$, and $j=\sqrt{(-1)}$. This series has an overlapping frequency band with that of the sunspot series. FFT fails to show 11 years as the prime periodicity in the contaminated data for $>11.45\%$ noise, whereas the p spectrum can still show 11 years as the prime periodicity.

D. Study with (multiplicative) chaotic noise contamination

Two periodic components of periodicity 18 and 17 having close patterns are generated [Figs. 3(a)–3(b)]; the scaling of the repetitive segments of the former is contaminated by noise elements derived from the Henon map process [34], and the two series of periodicities 18 and 17 are added together, where the former is 4 times stronger than the other. The FFT of the composite series [Fig. 3(c)] is misleading, and no clear picture emerges from the FFT about the actual

periodicities present. On the other hand, the *p* spectrum of the composite series detects 18 to be the most dominant periodic component present [Fig. 3(d)] and when it is extracted, the *p* spectrum [Fig. 3(e)] of the residual series shows 17 to be the next periodic component present.

E. Justification for the superiority of the *p* spectrum in periodicity assessment

The *p* spectrum involves LS estimation (in rank-1 sense) of the most dominant periodic pattern in the signal or data sequence, where the successive segments can be scaled differently. In a noise-free case, the strongest sinusoidal component will correspond to the strongest periodic component present in the signal. But when there is significant variation in either or both, (i) the optimal pattern from a sinusoidal pattern of the same period length and (ii) the scaling over successive repetitive segments, FFT is likely to miss the prime periodicity, because another sinusoidal component with constant scaling may appear to be stronger than the one corresponding to the periodicity of interest. Thus FFT can be misleading. In wide ranging trials, we found no occasion where FFT detects the periodicity correctly, whereas *p* spectrum fails.

Remark 1. *p* spectrum is conceptually, numerically, and computationally perhaps the most robust method for periodicity detection. The justification is as follows: ''Nearness to rank-oneness assessment'' is implicit with any method of periodicity detection (including Fourier decomposition and autocorrelation-function-based approach). The *p* spectrum does it more formally than the existing methods. Since *matrix rank* is the most basic property of any data set, and since SVD is the most robust tool for rank assessment (both numerically and computationally), for a given data set, there can be no better way of detecting embedded periodicity than through nearness to rank-oneness assessment through *p* spectrum, particularly when there are multiple periodic components present and/or when there is significant noise present. Even in the case of a single time-varying component [e.g., laser intensity (see Sec. IV B)], sinusoidal decomposition cannot provide as complete insight as obtained by the proposed method.

Remark 2. Although in the present studies, different types of extraneous noise contaminations are considered, it is understood that additional periodic components within the signal may act as contamination for the component(s) of interest.

IV. MISCELLANEOUS RESULTS

A. The natural process of sunspot numbers

Data type. The series of sunspot numbers is widely researched [7,13,14,35–37], as it is an indicator of the general solar activity. Here, the yearly averaged data over 1700–1938 are used for modeling, and the data over 1939–1988 are used for validation.

Periodic decomposition and prediction. The *p* spectrum of yearly averaged data over 1700–1938 shows repetitive peaks at $(n=)$ integer multiples of $N_1=11$ [Fig. 4(a)]; the extracted periodic component of periodicity 11 years is shown in Fig. 5(a). Proceeding the same way, we successively detect and extract two additional periodic components [Figs. 5(b) and 5(c)] of periodicity 10, and 12 years (in decreasing order of energy), respectively. The estimation using three stable periodic components is shown in Fig. 6. The multistep prediction was performed on each of the periodic components, which added together produce the overall 1 to 50 step ahead prediction [Fig. 7(a)] over 1939–1988, where the predicted peaks are found to be correct to ± 1 year; the corresponding correlation coefficient (ρ) vs prediction time (T_ρ) profile is shown in Fig. 7(b). Note that the correlation coefficient remains high over much longer prediction horizon (in contradiction to [37]).

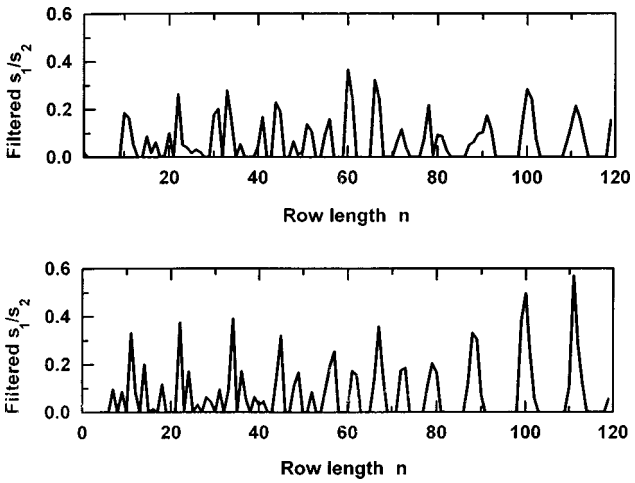

FIG. 4. (a) The *p* spectrum of the sunspot series showing repeating peaks at $n=11$ and its multiples, which is due to the most dominant embedded periodicity of 11 years. (b) The *p* spectrum of the surrogate series, showing much pronounced peaks at $n=11$ and its integer multiples. Such a feature is also observed in the *p* spectrum of logarithmically transformed sunspot series. The relative dominance of the peaks is due to the underlying nonlinearity associated with the series.

Study of noise sensitivity. To the original series, four different types of additive disturbances, namely, white noise, correlated noise, chaotic disturbances implemented through M-G equation, and Henon map process, are added. All the three periodic components could be detected for 74%, 58%, 75%, 35% contamination, respectively; the original series and the corresponding noise contaminated series are shown in Figs. 8(a)–8(e). For each contaminated series, multistep prediction was performed with the possible number of extracted periodic components; the corresponding residual sum of squares $\Sigma\varepsilon^2/\sigma^2$ profile is shown in Fig. 9. The results demonstrate the robustness of the proposed method against disturbances.

Analysis of p spectrum. In the *p* spectrum of the original series [Fig. 4(a)], some peaks appear at row lengths, which are not integer multiples of the periodicity of 11; the possible reasons are (a) the presence of additional periodicities (where the individual periodic components may have overlapping frequency bands), (b) the noninteger period lengths of periodic components, and (c) the nonlinearity in the underlying dynamics. It is already shown that the series is composed of three periodic components (Fig. 5); the two remaining issues are addressed as follows. To accommodate noninteger periodicity, the original data are interpolated tenfold, when three effective periodicities detected are 11.1, 10.0 (as in [13,14]), and 11.9. To study the nonlinearity in the series, a phase-

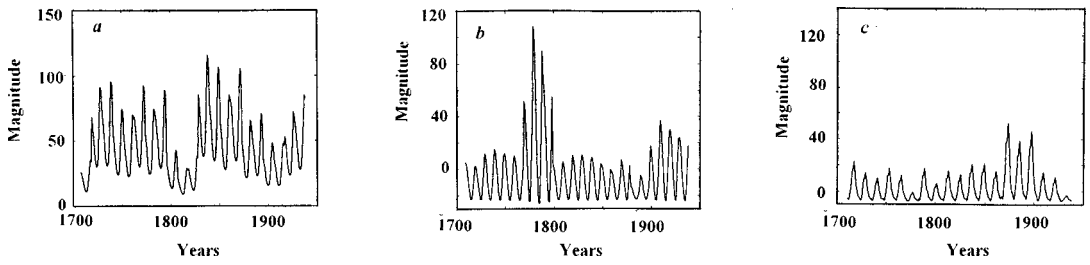

FIG. 5. (a)–(c) The individual components of periodicities ($N_i=$) 11, 10, and 12 years, respectively, successively extracted from the sunspot series. The components have been successively extracted in order of the decreasing energy content.

randomized surrogate series is generated from the original series [38] having the same mean and similar power spectrum as the original series. The *p*-spectrum peaks become significantly pronounced for the surrogate series [Fig. 4(b)], confirming nonlinearity in the data (as in [35–37]).

Observations. (i) The sunspot series is composed of three stable (nearly) periodic components (where the residual energy is less than 8% of the signal energy). (ii) Long term (with respect to the length of the data series concerned) prediction of the series is possible (in contradiction to [37]). This demonstrates (a) the ''richness of the information content'' of the periodic components obtained by the proposed method and (b) the inherent strength of the proposed approach for ''long-term periodic prediction.'' [See also observation (i) of Sec. IV C 1.] (ii) Although analysts have reported the sunspot series as chaotic [14,35] with reservation due to the insufficiency of data [13], according to the present results the *concerned length of the data series* does not distinctly reveal chaoticity, although we confirm inherent nonlinearity and determinism.

B. An experimental process: FIR NH_3 laser intensity

Data type. The FIR ammonia laser data series [11] [Fig. 10(a)], is believed to represent Lorenz-like chaoticity [39]. The series is made up of a series of segments, where within a segment the amplitude gradually tends to decrease. The length of each segment is about 500 data points and a spiral type transition occurs at the end of a segment leading to a subsequent segment, which again starts from a large amplitude. The *p* spectrum of the global data set (~5000 points) is shown in Fig. 11(a), the absence of repetitive distinguishable peaks reveal absence of any globally stable periodicity. So, the search for periodicity is extended to the local data segments.

Analysis of local segments. The first 270 points from one segment (~500 points) is taken for analysis. The *p* spectrum [Fig. 11(b)] shows nonuniformly sized peaks at (*n*) 7,14, . . . 56, 63, 69, 76/77, 84, 91, **97**, 104, . . . (the largest peak at 97 is 3.27 times the fundamental peak in terms of magnitude); note that the peaks in the *p* spectrum *drift* to row lengths, which are noninteger multiples of the fundamental row length ($n=7$). To verify the presence of noninteger periodicity, the data are interpolated 100-fold; the *p* spectrum [Fig. 11(c)] now shows effective periodicity of 6.88 (with relatively uniformly sized peaks at 688, 1376, 2064, 2752, **3440**, 4128, . . . where the strongest peak at 3440 is 1.7 times that at 688). The extraction based on stable periodicity (of $n=688$) and pattern on the interpolated data was found to be poor because of the instability of the pattern. So, the periodic extraction is performed with moving data windows having fixed periodicity (as in Sec. II F) as it permits variations in pattern over different windows, enhancing the extraction performance. Figure 10(b) shows the gradual variation of the patterns of the periodic segments of the extracted periodic component. No subsequent periodicity is detectable in the residual series through *p* spectrum.

Stability of periodicity attributes. In the subsequent analysis the periodicity and pattern are both permitted to vary with the moving data window. The dynamics primarily show oscillations in periodicity [Fig. 10(c)] with relatively similar periodic pattern [Fig. 10(c)] within one data segment, while over different data segments the patterns tend to vary. Figures 10(e) and 10(f) show the closeness of the state-space diagrams for the original and the extracted global series, where the residual energy is less than 0.2% of the signal energy.

Observations. Thus the present study reveals the following interesting (hitherto unavailable) insights into the laser intensity dynamics. The laser series exhibits (i) stable local periodicity but unstable pattern within a part of a segment, (ii) stable pattern with normalized periodicity within a segment, and (iii) unstable periodicity and pattern in global context, confirming inherent chaoticity.

C. Simulated chaotic processes

1. The Mackey-Glass series

Data type. Consider the delay differential Mackey-Glass (MG) equation [5,40]:

$$dx(t)/dt = 0.2x(t-\tau)/[1-x^{10}(1-\tau)] - 0.1x(t).$$

It has been introduced as a model for the regeneration of blood cells in leukaemia patients. $x(t)$ represents the density of circulating cells at time t, when it is produced, and $x(t-\tau)$ is the density when the request for more blood cells was

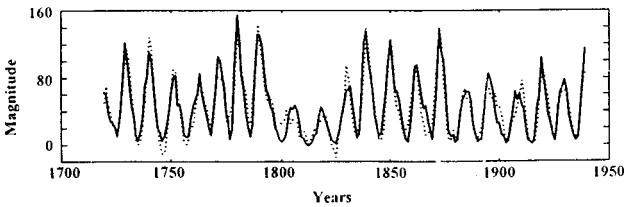

FIG. 6. Estimation of the sunspot series with three additive extracted components having periodicities of 11, 10, and 12 years (original series ——, estimated series .....); the residual energy is 7.52% of the original series. The *p* spectrum of the residual series does not show presence of any more periodic component.

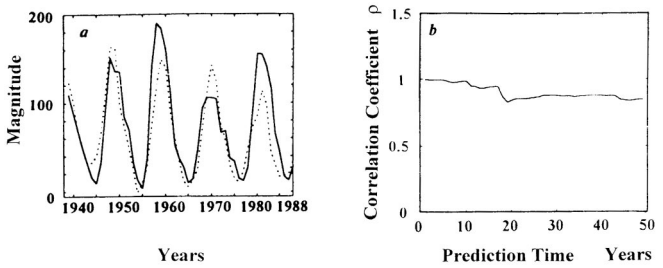

FIG. 7. (a) 1 to 50 year ahead prediction (over 1939–1988) of the sunspot series using three separately predicted components (Actual series ——, predicted series .....). The scaling factors for 11, 10, and 12 yearly components are modeled as: $g(k)=f(g(k-1),g^2(k-2))$, $g(k)=f(g(k-10))$, and $g(k)=f(g(k-6),g^2(k-6))$, respectively; f is a nonlinear function [30] formed from the past value of the $\{g(k)\}$ series, where the best set of variables has been selected through minimization of the C_p statistic [29]. (b) Correlation coefficient (ρ) between the predicted and the actual values, shown as a function of prediction horizon (T_p), remains at a high value signifying much improved long-term predictability than others [35–37] and inherent determinism.

made. For $4.53<\tau<13.3$, there is a stable limit cycle attractor. A period doubling bifurcation sequence is observed at $1.33<\tau<16.8$. For $\tau>16.8$, numerical simulations exhibit chaotic attractors [40].

Analysis. The present study is confined to two broad chaotic zones of the series. (a) For $\tau=17$ to 23, two dominant periodicities close to 3τ and 2τ are detectable. The periodicity attributes are dynamically varying; modeling and prediction are possible for short ranges (with respect to the periodicity) only. For $\tau=23$, with the first 3000 points discarded, two components of periodicities of 72 and 46 (over ~500 points) are found [Fig. 12(a)]. The long term prediction [Fig. 12(b)] produces $\rho=0.729$ for prediction horizon (T_p) $=200$, while ρ drops gradually beyond 200, e.g., $\rho=0.552$ for $T_p=450$. (b) For $\tau>23$, (i) the estimation through periodic decomposition is locally valid only and hence meaningful prediction is not possible, (ii) over larger data spans periodicity varies widely. For example, for $\tau=30$, the prime periodicity varies irregularly (unlike in the case of the global laser data series) between 85 to 109 detected over ~1000 points, using moving data windows.

Observations. (i) Once again high prediction correlation is obtained for long prediction horizons [Fig. 12(b)] irrespective of the underlying process being chaotic (contradicting conventional understanding of poor predictability for chaotic processes) [41]. The following conjectures are in order: (a) the components obtained though the proposed periodic decomposition contain rich information about the underlying process, and (b) long-term predictability of a chaotic process may be possible, if the inherent order (if any) can be determined and can be properly modeled by a prediction algorithm, as proposed. (ii) For $\tau=17$ to 23, there are two embedded periodicities of 3τ and 2τ, where ''chaoticity is revealed through the variations in the periodicity attributes.'' (iii) For $\tau>23$, with increasing chaoticity (i.e., with increasing τ [41]), the zone of relative regularity (in terms of stable periodicity attributes) contracts. This is a special case where the ''variability of the periodicity surpasses the concerned period lengths,'' a typical characteristic of strong chaoticity.

2. The logistic map process

The one-dimensional logistic map [32] $x(k+1)=rx(k)[1-x(k)]$ exhibits stable periodicity for specific values of r within 3 and 3.57; beyond which ($r>3.57$) the map exhibits chaoticity mixed with order. We analyze the map for different values for r. The p spectrum [Figs. 13(a) and 13(b)] show distinct periodicity of 32 for $r=3.5687$, and periodicity of 2 for $r=3.6786$, both being globally valid. The periodic extraction with fixed period and pattern is performed. In the

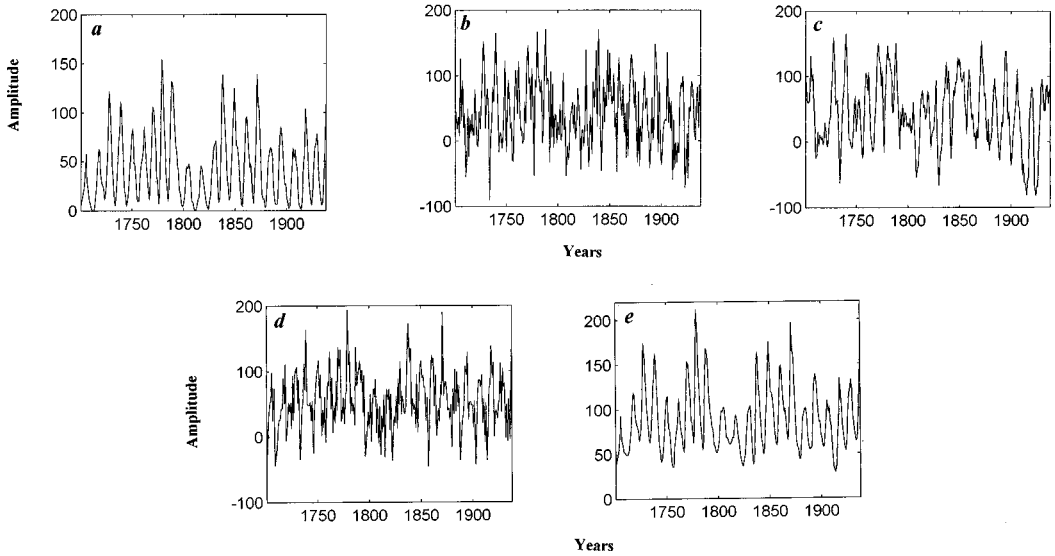

FIG. 8. (a)–(e) The original sunspot data series and the series contaminated with four types of additive measurement disturbances such as white Gaussian noise (74%), correlated noise (58%), chaotic contamination through Henon map (35%), and Mackey-Glass process (75%) with $\tau=100$, respectively. The bracketed term indicates the percentage of noise energy compared to signal energy. We have shown only those noisy data series with a maximum amount of disturbances for which the proposed algorithm successfully detects *all the three periodic components* with periodicity 11, 10, and 12 years.

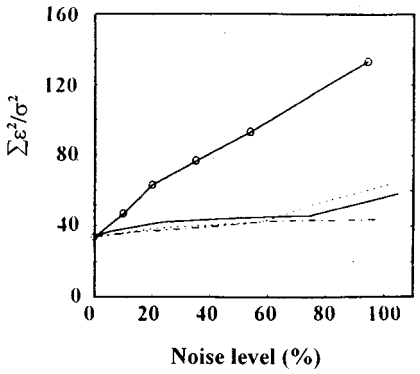

FIG. 9. Residual sum of square (error)/variance (i.e., $\Sigma\varepsilon^2/\sigma^2$) is shown as a function of measurement disturbances (as % of the signal energy) in terms of white noise (——), colored noise (.....), Henon map [34] (—⊖—), and Mackey-Glass [40] equation ($\tau = 100$) (—·—). We could detect all the periodic components for the worst type of additive disturbances (up to 35% of Henon map); the prime periodicity could be detected for all types of disturbances up to 99%. Similarly, effects on ρ were assessed (e.g., for worst cases, $\rho = 0.498$ for 99% disturbances).

former case [Fig. 14(a)], there is perfect match (RSS=0%) between the extracted component and the original series. For the latter ($r = 3.6786$) a relatively larger error is observed (RSS varying between 3.0% to 13.5% over different data segments). To accommodate the variation in the pattern in the latter case, the periodic component is extracted with a moving window with fixed period (=2) and is shown in Fig. 14(b); the error reduces significantly (RSS varies from 1.5% to 2.5%), confirming instability in pattern in global sense. For $r = 3.92$, no globally stable periodicity is observed through p spectrum [Fig. 13(c)] except for short data spans. For data segments of 50, periodicity varies between ($N=$) 5 to 8; the periodic extraction over a segment for $N = 5$ is shown in Fig. 14(c). Here larger RSS variations (from 10.5% to 17.2%) are observed over different data segments confirming the absence of any globally stable periodicity.

Observations. The study reveals that the logistic map process possesses (i) globally stable periodicity and pattern before entering into the chaotic regime, (ii) globally stable periodicity but unstable pattern for certain specific values of r, e.g., $r = 3.6786$ (expectedly due to inherent chaoticity), and (iii) globally unstable period and pattern well into the chaotic zone.

V. CONCLUSIONS

The proposed p spectrum based method is possibly the most robust method for the detection of periodicity of sinusoidal or nonsinusoidal periodic components contained in an irregular data sequence. The proposed is a generalized concept (for the detection of embedded periodicity and the decomposition of a data series into multiple periodic components, which are not necessarily sinusoidal), of which the Fourier decomposition can be considered to be a special subset.

The proposed characterization in terms of the three detected ''orthogonal'' *periodicity attributes*, namely, the periodicity (or period length), periodic pattern, and scaling factor for the successive periodic or nearly periodic segments of a periodic series, provide complete description of a periodic

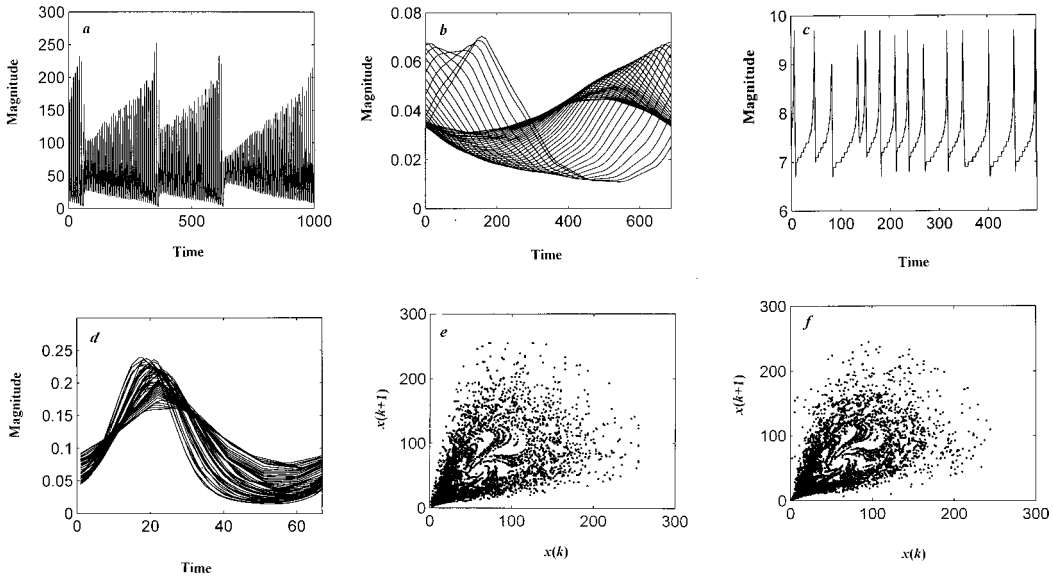

FIG. 10. (a) Chaotic pulsations of FIR-NH_3 laser. (b) The local variations of pattern with effective periodicity 6.88 showing gradual drifts of the patterns. The extraction based on stable periodicity and pattern was found to be poor because of the instability of the pattern with fixed period. So for extraction from the 100-fold interpolated data series, a moving window length (with fixed periodicity 688) has been used as it permits variations in pattern, enhancing the extraction performance. (c) Variations of the periodicity over the global data set; within a data segment, periodicity varies approximately between 6.7 to 9.7. The period usually starts from the lowest value, and then increases gradually; on reaching the upper limit, the periodicity abruptly drops to the lowest value, and so on. (d) The normalized (with respect to the period length) patterns within one segment showing nearly repetitive profiles. (e) and (f) State-space plots of the original series (~5000 points) and the extracted series, respectively, show that the extracted component carries almost the same dynamical information as in the original series; the residual energy is as low as 0.46% of the original series.

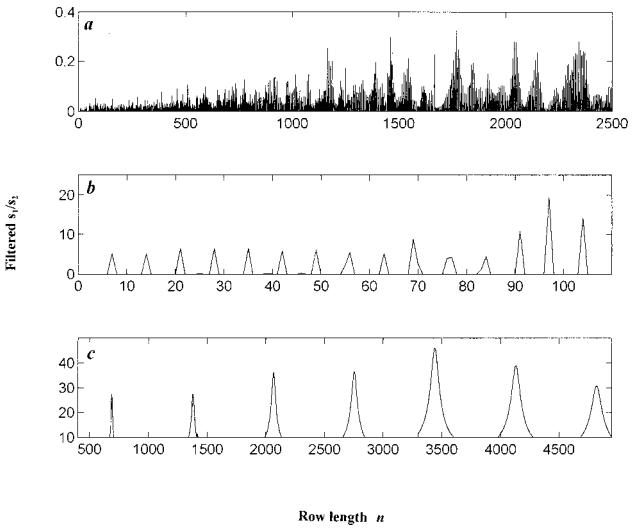

FIG. 11. (a) The p spectrum of the global FIR-NH_3 data series ($\sim$5000 points), which shows the absence of any stable periodicity. (b) The p spectrum of a local data span ($\sim$270 data points) within one data segment. The peaks appear at n $=7,14,...56,63,69,76/77,84,91,\mathbf{97},104,..$; the drifting of the peaks to noninteger multiples of the fundamental period ($=7$) indicates the presence of fractional periodicity. (c) The p spectrum of 100-fold interpolated data clearly indicates the presence of nearly stable periodicity of 688 (i.e., effective periodicity$=6.88$).

series. It has been demonstrated that the proposed approach can lead to the unearthing of rich information about the underlying process in terms of the characteristic periodicity attributes, and thereby extract the orderly part hidden within any irregular series, which can make long-term prediction possible even in the presence of chaoticity. The proposed is fundamentally different from the existing Fourier-model-based approaches, and has been shown to be more noise immune and to be able to provide more meaningful analysis than the existing analytical tools.

Periodic component(s) in a chaotic process are unstable in nature in terms of periodicity or pattern or both; the detectability of periodic components will depend on the locality (with respect to the periodicity) of the stability of periodicity. The proposed characterization of the underlying process through the periodicity attributes can be summarized by the following observations.

(a) If a globally or locally stable periodicity is not detectable through the p spectrum, the series must be stochastic or strongly chaotic. It cannot have any component that is strong enough and yet predictable to render the overall series to be locally or globally predictable.

(b) If globally stable periodicity is not detectable but locally stable periodicity is detectable through p spectrum, the series will be low dimensionally chaotic. Here the periodicity attribute(s) will vary. For segments over which p spectrum shows relatively steady periodicity, the pattern and/or the scaling factors will vary to limited extents, making predictability possible over ranges commensurate with the zone of stable periodicity.

(c) If globally stable periodicity is detectable, long-term predictability may be possible in spite of variability in other periodicity attributes, which (as is demonstrated) does not necessarily imply lack of chaoticity.

It is demonstrated that the characterization of a process in terms of its embedded periodic component(s) and the stability of the periodicity attributes of the individual component(s) can provide a new insight into the understanding of a large class of irregular cyclical processes.

ACKNOWLEDGMENTS

The financial support for the research by DOE, Government of India (J.B.) and CSIR, India (G.S.) is thankfully acknowledged.

APPENDIX: OPTIMAL MODELING USING m-QR_{cp} FACTORIZATION AND C_p STATISTIC

Consider the modeling problem

$$\mathbf{y}=\mathbf{G}\boldsymbol{\theta}, \tag{A1}$$

where the $m\times n$ matrix $\mathbf{G}=[\mathbf{g}_1,\mathbf{g}_2,\ldots,\mathbf{g}_i,\ldots,\mathbf{g}_n]$ contains n ($<m$) m- long regressor vectors $\mathbf{g}_i$, $\mathbf{y}$ is the output vector, and $\boldsymbol{\theta}$ is the parameter vector. For optimal modeling, the p ($<n$ and unknown) most significant variables are selected leading to the model $\mathbf{y}=\mathbf{G}^*\boldsymbol{\theta}_e+\mathbf{e}$, where $\mathbf{G}^*\in\mathbf{G}$, $\boldsymbol{\theta}_e$ is the corresponding LS estimated p parameter vector, and $\mathbf{e}$ stands for the uncertainty. m-QRcp factorization can lead to ''optimal successive selection'' of the p ($\leqslant n$) regressors in $\mathbf{G}$ in Eq. (A1). The algorithm can be explained as follows.

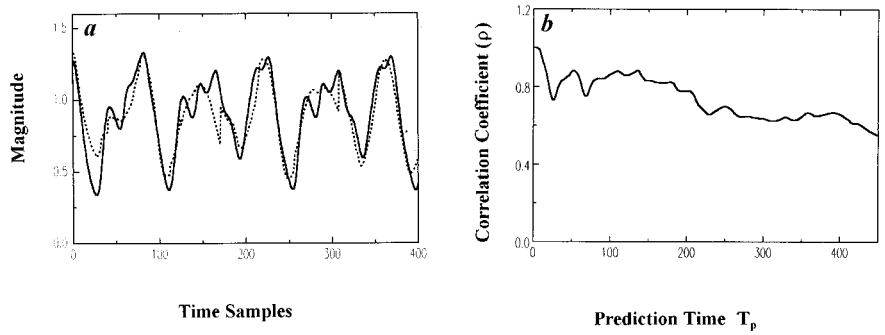

FIG. 12. (a) The estimation with two periodic components of periodicity 72, and 46 for Mackey-Glass process with $\tau=23$ (original series ——, estimated series .....) over $\sim$450 data points. The figure shows that although the series is known to be chaotic, it can be approximated by two additive periodic components. (b) The prediction performance in terms of correlation coefficient ρ vs prediction horizon T_p over 450 data points.

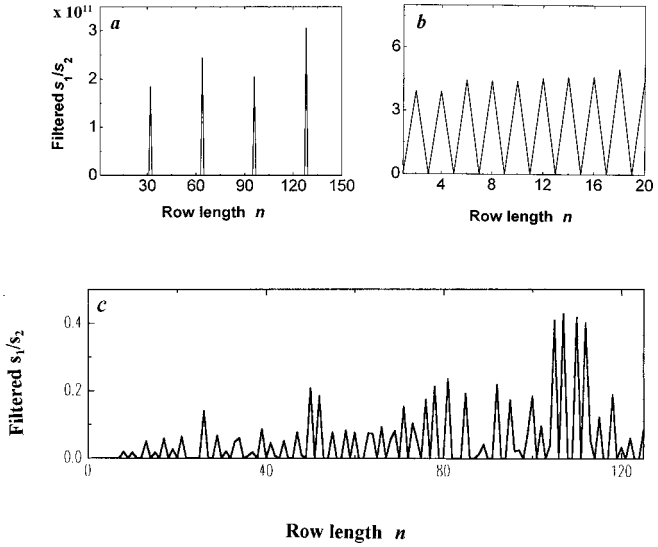

FIG. 13. (a)–(c) The p spectrum of the logistic map for $r=3.56876$, $r=3.6786$, and $r=3.92$, respectively. For the first two cases, the p spectrums show the presence of distinct globally stable periodicity of length 32 and 2, respectively, but in the last case (which is well into the chaotic regime [32]) no repetitive peaks are found, demonstrating the absence of any globally stable periodic component.

First, the column vector of $\mathbf{G}$ producing $\max|\mathbf{g}_i^T\mathbf{y}|$, $i=1$ to n, is selected and is swapped with $\mathbf{g}_1$. The subsequent columns of $\mathbf{G}$ are pivoted as follows. Using the Gram-Schmidt orthogonalization [18], if $\mathbf{q}_1$ be the unit vector in the direction of $\mathbf{g}_1$, the portion of $\mathbf{g}_j$ ($j=2$ to n) and $\mathbf{y}$ in a direction orthogonal to $\mathbf{g}_1$ will be given by $(\mathbf{g}_j-\mathbf{q}_1^T\mathbf{g}_j\mathbf{q}_1)$ and $(\mathbf{y}-\mathbf{q}_1^T\mathbf{y}\mathbf{q}_1)$, respectively.

At the ith stage of selection, the rotated variables vectors $(\mathbf{g}_i^*)$ and the rotated output $(\mathbf{y}^*)$ vector are

$$\mathbf{g}_i^*=\mathbf{g}_j-(\mathbf{q}_1^T\mathbf{g}_j\mathbf{q}_1+\cdots+\mathbf{q}_{i-1}^T\mathbf{g}_j\mathbf{q}_{i-1}),$$

$$i=2 \text{ to } n, \quad j=i \text{ to } n$$

$$\mathbf{y}^*=\mathbf{y}-(\mathbf{q}_1^T\mathbf{y}\mathbf{q}_1+\cdots+\mathbf{q}_{i-1}^T\mathbf{y}\mathbf{q}_{i-1}),$$

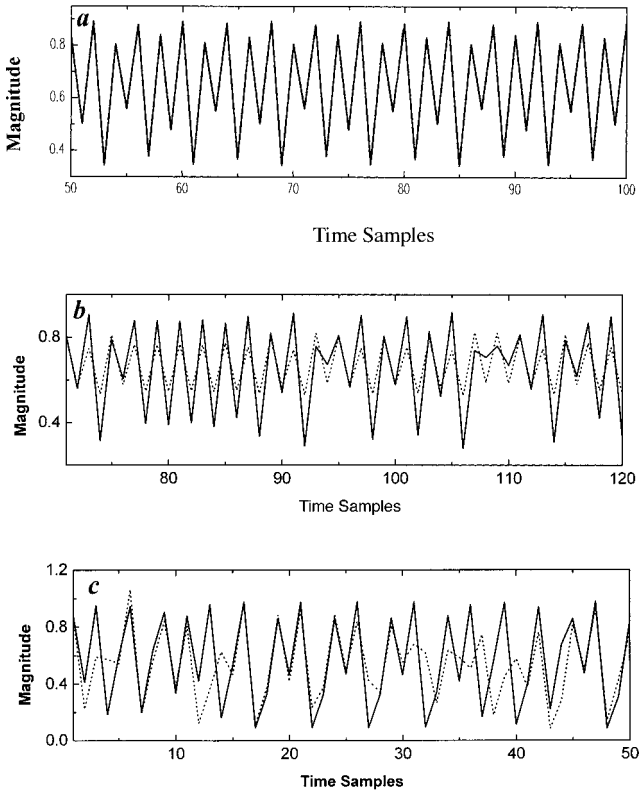

FIG. 14. (a) The estimation (original series ——, estimated series .....) for $r=3.567\,86$ shows that there is a perfect match with just one periodic component, which has globally stable periodicity and pattern; the residual energy is 0%. (b) The estimation for $r=3.6786$ (original series ——, estimated series .....) using the moving data window. The requirement for the moving data window has been explained in Sec. II F and Sec. IV C 2). (c) The estimation over a local segment having periodicity $N=5$ for $r=3.92$ (original series ——, estimated series .....). The estimation error is large but in some sections there is a nearly perfect fit by the estimated periodic component, which is due to the recurrence of the periodicity and pattern in the chaotic process.

and ith selected vector is the one maximizing $|\mathbf{g}_i^{*T}\mathbf{y}^*|$; the selection procedure is repeated until p regressors are selected. Minimum C_p statistic [29] leads to the selection of the optimal model order (p).

[1] T. A. Denton *et al.*, Am. Heart J. **120**, 1419 (1990); A. L. Goldberger, D. R. Rigney, and B. J. West, Sci. Am. **262**, 42 (1990); A. Babloyantz and A. Destexhe, Biol. Cybern. **58**, 203 (1988); L. Glass *et al.*, Proc. R. Soc. London, Ser. A **413**, 9 (1987).

[2] P. P. Kanjilal, S. Palit, and G. Saha, IEEE Trans. Biomed. Eng. **44**, 51 (1997).

[3] A. Schuster, Terr. Magn. Atmos. Electr. **3**, 13 (1898); H. O. A. Wold, *A Study in the Analysis of Stationary Time Series* (Almqvist and Wiksells, Uppsala, Sweden, 1954); B. P. Lathi, *Signals, Systems and Controls* (Harper & Row, New York, 1974); S. M. Kay, *Modern Spectral Estimation* (Prentice Hall, Englewood Cliffs, NJ, 1988); P. S. Naidu, *Modern Spectrum Analysis of Time Series* (CRC, Boca Raton, FL, 1996).

[4] W. S. Pritchard and D. W. Duke, Int. J. Neurosci. **67**, 31 (1992); *Machinery of the Mind*, edited by E. R. John (Birkhauser, Boston, 1990); A. C. K. Soong and I. J. M. Stuart, Biol. Cybern. **62**, 55 (1988).

[5] A. L. Goldberger, K. Kobalter, and V. Bhargava, IEEE Trans. Biomed. Eng. **33**, 874 (1986); M. C. Mackey and L. Glass, Science **197**, 287 (1977).

[6] B. T. Grenfell, J. R. Statist. Soc. B **54**, 383 (1992); L. F. Olsen and W. Schaffer, Science **249**, 499 (1990); W. P. London and J. A. Yorke, Am. J. Epidemiol. **98**, 453 (1973).

[7] R. Vautard, P. Yiou, and M. Ghil, Physica D **58**, 95 (1992); M. Ghil and R. Vautard, Nature (London) **350**, 324 (1991).

[8] J. Kurths and A. A. Ruzmaikin, Sol. Phys. **126**, 407 (1990); G. U. Yule, Philos. Trans. R. Soc. London, Ser. A **226**, 267 (1927).

[9] D. M. Raup, Nature (London) **314**, 341 (1985); M. R. Rampino and R. B. Stothers, Science **226**, 1427 (1984); D. M. Raup and J. J. Sepkoski, Jr., Proc. Natl. Acad. Sci. USA **81**, 801 (1984).

[10] T. Serre, Z. Kollath, and J. R. Buchler, Astron. Astrophys. **311**, 833 (1996).

[11] C. O. Weiss and J. Brock, Phys. Rev. Lett. **57**, 2804 (1986).

[12] D. C. Park *et al.*, IEEE Trans. Inf. Theory **6**, 442 (1991).

[13] N. O. Weiss, Philos. Trans. R. Soc. London, Ser. A **330**, 617 (1990).

[14] A. Berger, J. L. Mélice, and I. van der Mersch, Philos. Trans. R. Soc. London, Ser. A **330**, 529 (1990).

[15] C.-S. Poon and C. K. Merrill, Nature (London) **389**, 492 (1997).

[16] P. C. Ivanov *et al.*, Nature (London) **383**, 323 (1996).

[17] P. So *et al.*, Phys. Rev. Lett. **76**, 4705 (1996); R. Badii *et al.*, Rev. Mod. Phys. **66**, 1389 (1994); O. Binham and W. Wenzel, Phys. Rev. Lett. **63**, 819 (1989); D. P. Lathrop and E. J. Kostelich, Phys. Rev. A **40**, 4028 (1989); O. Binham and W. Wenzel, Phys. Rev. Lett. **63**, 819 (1989); P. Cvitanovic, *ibid.* **61**, 2729 (1988).

[18] G. H. Golub and C. F. Van Loan, *Matrix Computations* (Johns Hopkins University Press, Baltimore, 1989).

[19] R. A. Thisted, *Elements of Statistical Computing: Numerical Computations* (Chapman and Hall, New York, 1988); C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* (Prentice-Hall, Englewood Cliffs, NJ, 1974).

[20] P. P. Kanjilal, *Adaptive Prediction and Predictive Control* (Peter Peregrinus, Stevenage, UK, 1995).

[21] *SVD and Signal Processing: Algorithms, Applications, and Architectures*, edited by F. Deprettere (North-Holland, Amsterdam, 1988).

[22] V. C. Klema and A. J. Laub, IEEE Trans. Autom. Control. **AC-25**, 164 (1980); R. V. Patel, A. J. Laub, and P. M. Van Dooren, *Numerical Linear Algebra Techniques for Systems and Control* (IEEE, New York, 1994), pp. 1–35. There are many efficient and numerically stable algorithms [20] available for the computation of SVD of any matrix.

[23] D. S. Broomhead and G. P. King, Physica D **20**, 217 (1986); A. I. Mees, P. E. Rapp, and L. S. Jennings, Phys. Rev. A **36**, 340 (1987); A. M. Albano *et al.*, *ibid.* **38**, 3017 (1988); E. J. Kostelich and T. Schreiber, Phys. Rev. E **48**, 1752 (1993).

[24] The ''period-length spectrum or p spectrum'' can also be called ''singular value ratio'' spectrum of $\{x(k)\}$ [P. P. Kanjilal and S. Palit, IEEE Trans. Signal Process. **43**, 1536 (1995), [20]]. Since, in place of s_1/s_2, any other measure of closeness to rank oneness can serve, we use the generic term ''p spectrum.''

[25] The p spectrum is detrended using *bidirectional low-pass filtering*, [R. L. Longini *et al.*, IEEE Trans. Biomed. Eng. **22**, 432 (1975), [20]], and the information in the positive direction only is retained. For bidirectional filtering, the data set is first passed through a linear filter, the filtered output is time reversed and passed through the same filter again; the time reversed second filter output subtracted from the raw signal forms the desired signal. Here a linear filter with a pole at 0.2 is used. This is a general signal processing exercise, as otherwise the spectrum rides on a trend; since the prime information is in the amplitude of the peaks (with respect to the neighborhood), the spectrum is clearer with detrending and low-pass bidirectional filtering. p-spectrum may be used without detrending and filtering also [e.g., Figs. 2(b), 2(d), and 2(f)], although the peaks will be less pronounced.

[26] Here the periodicity index is derived as follows (which is not the only possible approach). The algorithm moves a data of window of size 4 over the p spectrum and identifies a peak [say $I'(n)$] to be present if its magnitude is at least 20% more than the moving average. The mean of $I'(jn)$ occurring at integral multiples (j) of the period length n serves as the estimated periodicity index $I(n)$.

[27] SVD is a case of joint orthogonal transformation between **U** and **V**. **V** rotates the points in scatterplot space (of $\mathbf{A}_n$), while **U** rotates the points in variable space. The decomposed vectors $\mathbf{u_i}$ and $\mathbf{v_i}$ are considered ''conjugate'' [I. J. Good, Technometrics **11**, 823 (1969)], and are expected to have related physical interpretations. For the present rank-one approximation, the perturbation bounds on $\mathbf{u}_1$, $\mathbf{s}_1$, and $\mathbf{v}_1$ are discussed in Ref. [2].

[28] P. P. Kanjilal, G. Saha, and T. J. Koickal, IEEE Trans. Syst. Man Cybern. Pt. B **29**, 1 (1999).

[29] C. L. Mallows, Technometrics **15**, 661 (1973). Minimization of $C_p\ [=\Sigma\varepsilon_p^2/\Sigma\varepsilon_n^2+(m-2p)]$ ensures optimal modeling (where with m data sets when p is the optimal model order leading to $\Sigma\varepsilon_p^2$ residual sum of square error and n is the total no. of candidate regressors) as in C. Daniel and F. S. Wood, *Fitting Equations to Data* (Wiley, New York, 1971).

[30] In the present case only up to quadratic terms have been used (which is not a limitation) subject to the availability of data. In place of the present model for $\{g(k)\}$, alternative nonlinear modeling methods [as in S. Haykin, *Neural Networks: A Comprehensive Foundation* (MacMillan, New York, 1994), [20] etc.] may be used.

[31] In case of a dynamic series, all three periodicity attributes of the component(s) may vary with time. The present prediction scheme assumes constant periodicity; no restrictions are imposed on the variations of the scaling factors.

[32] R. M. May, Nature (London) **261**, 459 (1976); Proc. R. Soc. London, Ser. A **413**, 217 (1987); S. H. Strogatz, *Nonlinear Dynamics and Chaos* (Addison-Wesley, Reading, MA, 1994).

[33] L. Cohen, *Time Frequency Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1995).

[34] M. Henon, Commun. Math. Phys. **50**, 69 (1976).

[35] A. Weigend, D. Rumelhart, and B. Huberman, Int. J. Neural Syst. **1**, 193 (1990); M. Casdagli, J. R. Statist. Soc. B **54**, 303 (1992); Weigend *et al.* (1990) and Berger *et al.* [14] considered the sunspot series to be low-dimensional chaotic while Casdagli (1992) concluded it to possess high-dimensional nonlinearity.

[36] H. Tong, *Nonlinear Time Series: A Dynamical System Approach* (Clarendon, Oxford, 1996). Yule [8] (1927) developed linear stochastic models whereas stochastic nonlinear models have been proposed by Tong (1996).

[37] G. Sugihara and R. M. May, Nature (London) **344**, 734 (1990). It is said that chaoticity reflects in drooping correlation coefficient (e.g., for sunspot series) for long prediction horizons.

[38] J. Theiler *et al.*, Physica D **58**, 77 (1992). We used the amplitude adjusted Fourier transform method as surrogate generator; an ensemble of 16 implementations was used.

[39] U. Huebner, N. B. Abraham, and C. O. Weiss, Phys. Rev. A

40, 6354 (1989). The laser sequence has an underlying attractor dimension of 2.0–2.3 and an entropy rate $(0.2-0.7)T^{-1}$ where T is the average intensity pulsing period. Our analysis unearths new structural features of the laser series in terms of period, and pattern variations (Fig. 5).

[40] J. D. Farmer, Physica D **4**, 366 (1982). For $\tau=100$, correlation dimension=7.5.

[41] On similar analysis [J. Bhattacharya and P. P. Kanjilal (unpublished)] on x variable of Rossler series [N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980)], two periodic components are obtained, which lead to good predictability (e.g., prediction correlation coefficient as high as 0.9 for $T_p=2200$). Fourier domain analysis shows four dominant Fourier components of frequency 0.25, 0.22, 0.2, 0.167 (where a periodicity of 1000≡normalized frequency 1); the corresponding strengths are LS estimated. Multistep prediction with these components produce poor validation results, unlike the proposed method.