Noise-Filtering by Multichannel Singular System Analysis

Francesco Lisi[*] Alfredo Medio[†] Marco Sandri[‡]

5 Dicembre 1994

Riassunto

In questo articolo viene discusso un metodo di filtraggio per ridurre il rumore (bianco) per sistemi dinamici a tempo continuo e a tempo discreto, basato sull'Analisi dei Sistemi Singolari (SSA). In particolare viene proposta un'estensione multivariata (M-SSA) della usuale Analisi dei Sistemi Singolari univariata (S-SSA).

Il metodo è stato applicato a dati generati da alcuni dei più noti modelli caotici (Hénon, Ikeda, Rössler and Lorenz) per evidenziare la effettiva capacità del metodo di ridurre il rumore, ciò in particolar modo quando le serie sono brevi (< 1000 osservazioni) e il rapporto rumore/segnale è elevato.

Abstract

In this paper, we discuss a method of filtering (white) noise in continuous- and discrete-time dynamical systems, based on Singular System Analysis (SSA). In particular, we extend the most common single-channel analysis to the multichannel case (M-SSA). We apply M-SSA to a number of well-known "toy models", (Hénon, Ikeda, Rössler and Lorenz), showing that it is an effective filtering method, especially so when the time series are short (< 1000 observations) and highly contaminated.

[*]Dipartimento di Statistica, Università di Padova, Italia. E-Mail: lisi@gauss.stat.unipd.it

[†]Dipartimento di Economia, Univerità di Venezia, Italia. E-Mail: medio@unive.it

[‡]Istituto di Scienze Economiche, Univerità di Verona, Italia. E-Mail: msandri@chiostro.univr.it

1. Introduction

The contemporary literature on nonlinear dynamics has until recently been mostly concentrated on "direct" problems, i.e., problems arising from the analysis of a given mathematical system - typically a system of differential or difference equations. In this context, the main object of the investigation is that of characterizing the dynamics of the system in relation to different types of (nonlinear) functional relationships and different parameter constellations. For example, one might wish to establish whether a certain class of systems, for certain values of the relevant parameters, exhibits complex, or chaotic behavior, the latter being defined, say, by the existence of one or more positive Lyapunov exponent.

In the 1980's, however, researchers' attention has been increasingly diverted to the complementary, "inverse" problem of investigating experimental data - either actually observed or generated *ad hoc* - in order to derive certain properties of an hypothetical generating system. The works of Packard *et al.* (1980), Mañé (1981) and Takens (1981) provided a firm theoretical ground for techniques aimed at extracting qualitative information from experimental data. In particular, various embedding results (besides the works just quoted, see Sauer, Yorke and Casdagli (1991)) rigorously proved the possibility to reconstruct the n-dimensional asymptotic dynamics (i.e., the dynamics on an attractor) of a system from the observation of an m-dimensional $(1 \leq m \leq n)$ time series.

Important though the abovementioned inverse results may be, they leave many fundamental questions open, whenever the idea of reconstructing the state space of a dynamical system and characterizing the attractor "living" in it is applied to real data. Embedding results generally hold only for "clean", noiseless data. In this ideal case, all reconstruction methods work well and most choices of the relevant parameters (in particular of the sampling time) are equivalently good. However, real data are imprecise and noise-contaminated and time series are often too short and/or too sparsely recorded to be suitable for the evaluation of asymptotic properties of dynamical systems, such as fractal dimensions, Lyapunov exponents, entropy, etc. Moreover, algorithms for estimation of these and other significant invariants are typically complex and time-consuming. Therefore, considerations of cost efficiency also come into consideration.

In this paper, we shall discuss the method called "singular system analysis" (SSA). The origin of SSA in mathematics goes as far back as 1873[1] and it generated ramifications in various fields of analysis, including statistics[2] and information theory[3]. In the dynamical system community, SSA was first suggested by Broomhead and King (1986) and it has been widely used since, but its merits and efficiency in solving inverse problems has been rather controversial[4].

In what follows, we shall argue that singular system analysis (SSA) is an effective noise filtering method and shall extend the more common "single-channel singular system analysis" (S-SSA) to "multichannel singular system analysis" (M-SSA). This is a recent and relatively unchartered area of research and we shall prove that M-SSA enhanches effectiveness of the method, by helping to remove some of the commonly encountered difficulties.

In order to make it possible to compare our results with those of other similarly oriented papers, the method in question will be applied to some of the most widely studied "toy-models". It will be seen that it is especially effective and likely to beat the competition when time series are short and higly contaminated by noise. Since this is usually the case for social sciences in general and economics in particular, we hope that our approach will be especially helpful to researchers in those areas.

Most works which discuss state space re-costruction from the point of view of nonlinear dynamical system theory concentrate their attention on the ability of the suggested methods to capture the main qualitative features of the attractor of the hypothetical generating system. For this purpose, the fact that dynamics in the re-structed state space is described in terms of a set of variables different from the observed one (or ones) is most often essentially irrelevant. However, when re-construction and noise filtering is performed as a preliminary step towards forecasting, it becoms essential to re-convert the (filtered) new variables into the original one (or ones). But, owing to filtering and the consequent loss of information, this re-conversion is not trivial nor it is generally unique. To overcome this difficulty, we shall discuss a simple, although possibly time-consuming technique, generalizing to a method first developed by Vautard *et al.* (1992) to the multidimensional case.

2. Theoretical backgroud

We shall now recall some basic theoretical results which will provide a background for the rest of the paper and will help to putt our numerical results into focus. Since there already exists an enormous literature on the subject, we shall limit ourselves to a concise and non-rigorous description of few selected points of interest.

Let us first consider a m-dimensional map

$$\mathbf{x}_{t+\tau} = \mathbf{F}(\mathbf{x}_t) \qquad (2.1)$$

where $\mathbf{x} \in \Re^m$; $\mathbf{F}: U \rightarrow \Re^m$ is a continuous and sufficiently derivable function; U is an open subset of $\Re^m$; $t \in \Re$ denotes time; $\tau \in \Re_+$ is a time interval, which can always be put equal to one by appropriately choosing the arbitrary unit of measure of time. Equation 2.1 can represents either a direct, discrete-time

[1]The first recorded publication on singular value decomposition was authored by the Italian mathematician Eugenio Beltrami (1873). For an interesting historical survey of the early results in the field, see Stewart (1993).

[2]In this area, we shall mention factor and principal component analysis and the classical contribution of Pearson's (1901) and Hotelling's (1933).

[3]See, for example, Pike *et al.* (1984).

[4]The reader can appreciate the issues at stake by consulting Mees, Rapp and Jennings (1987) - with related comment by Broomhead, Jones and King (1988) and reply by Mees and Rapp (1988) - and Paluš and Dvořák (1992).

formulation of the dynamic problem under scrutiny, or the so-called "flow map", derived from the solution of a continuous-time system of differential equations.

The m-dimensional real vector $\mathbf{x}$ can be looked at as a set of observations, describing the state of the system at a certain instant of time. The space $\Re^m$ (or a sub-space of it) is called *state space*, or *phase space*, i.e., the space where the system evolves in time, while $\mathbf{F}$ represents the deterministic "law of motion" of the system. According to this representation, knowledge of the state of the system at a given time and of the function $\mathbf{F}$ permits one to determine its future.

In reality, however, we do not necessarily have direct and complete access to all the state variables of the system. Leaving aside for the moment noise and errors of observation, we typically observe only one or few variables through a certain (deterministic) "viewer". Formally, what we observe is a (possibly vector-valued) variable $\mathbf{y}$, such that

$$\mathbf{y}_t = \mathbf{h}(\mathbf{x}_t) \tag{2.2}$$

where $\mathbf{y}_t \in \Re^l$ and the function $\mathbf{h}: \Re^m \to \Re^l$, $l \leq m$, represents the "viewer".

Let us now consider the simpler unidimensional case, in which $l = 1$, $y \in \Re$, and let us suppose we have a series of observations $\{y_t\}_{t=1}^N$. In order to construct a dynamical model capable of producing the observed series (or a good approximation of it), we have, first of all, to re-construct a state space. For this purpose, the basic theoretical result is the theorem simultaneously and independently produced by Takens (1981) and Mañè (1981). The former's slightly modified (continuous-time, one-dimensional) version reads as follows.

Theorem. Let M be a compact manifold of dimension m. For a given pair $(\mathbf{F}, h)$, where $\mathbf{F}$ is a smooth (at least C^2) vectorfield and $h: M \to \Re$ is a smooth function, the map $\Phi_{\mathbf{F},h}: M \to \Re^{2m+1}$, defined by

$$\Phi_{\mathbf{F},h}(x) = [h(x), h(\phi_\tau(x)), h(\phi_{2\tau}(x)), \ldots, h(\phi_{2m\tau}(x))]^T \tag{2.3}$$

is generally an embedding[5], where ϕ is the flow map associated with $\mathbf{F}$ for a given time interval τ.□

Heuristically speaking, this result means that, for a suitable (twice differentiable) observation function and for a sufficiently large positive integer d, there corresponds to the time evolution of of the true system - of which we only observe one variable - a time evolution of a d-dimensional system which is diffeomorphically equivalent to the former's. The elements of the vector-valued variable of the re-costructed system, denoted by $\mathbf{Y}_t$, are in this case the delayed values of the one-dimensional observed variable. Putting $\tau = 1$, we have

$$\mathbf{Y}_t = [y_t, y_{t-\tau}, \ldots, y_{t-d\tau+1}]. \tag{2.4}$$

The theorem guarantees that, for a given "law of motion" $\mathbf{F}$ and a (deterministic, smooth) "viewer" h, the re-construction in question preserves the basic characterizations of the attractors of the system. For example, stable fixed points or limit cycle, attracting quasi-periodic, or chaotic orbits of the re-constructed system will correspond to similar configurations of the "true", unknown system and the same will be true for the fundamental statistical invariants, such as fractal dimensions, positive Lyapunov characteristic exponents, entropy.

Implicitly, the theorem tells us that there exists a d-dimensional map $\mathbf{G}^{(d)}: \Re^d \to \Re^d$, which in general depends on d, such that

$$\mathbf{Y}_t = \mathbf{G}^{(d)}(\mathbf{Y}_{t-1}) \tag{2.5}$$

or, equivalently, that for any t, there exist d maps $G_i^{(d)}: \Re^d \to \Re$ $(i = 1, \ldots, d)$ such that

$$y_t = G_i^{(d)}(y_{t-1}, \ldots, y_{t-d}). \tag{2.6}$$

Since $G_i^{(d)}$ is the i.th element of $\mathbf{G}^{(d)}$, equations (2.5) and (2.6) simply say that there exists an equivalence between the (re-constructed) dynamics of $\mathbf{Y}$ in the $\Re^d$ space and the dynamics of any of its components in $\Re$. The embedding properties of the re-construction of the state space based on delayed variables also holds when other auxiliary variables are used, e.g., the first $d \geq (2m+1)$ derivatives of the unique observed variable. We shall see later that certain linear transformations of the original variables can also be used, still obtaining a "correct" re-construction.

Whatever method we use, however, in applications we face rather formidable problems. First of all, the choice of the parameter d, crucial for the validity of the Takens-Mañé theorem, must be made in the ignorance of the "true" dimension m of the invisible manifold M. Typically, one makes a guess and then increase the initial value of d, hoping for some "saturation" of the results. Secondly, the theorem in question is only a preliminary result toward the more ambitious goal of predicting future values of the observed variable. But one thing is to know that a map $\mathbf{G}^{(d)}$ exists and quite another thing is to find or approximate it. Finally, the condition that the observation function h be smooth, implicitly says that the result of the theorem does not necessarily hold even approximately when the data are contaminated by noise. This is the problem on which we shall concentrate in the sequel of this paper. To be more specific (and more modest) we shall try to provide a contribution to the problem of filtering the simpler form of noise which is the observational, additive, white noise.

3. Singular System Analysis. An Overview.

Formally, we shall consider the case in which observations consist of a series $\{z_t\}_{t=1}^{N^*}$, such that $z_t = h(x_t) + w_t = y_t + w_t$, where w_t is the error in observation.

If we perform re-costruction as indicated above, we shall end up with vector-valued variables $\mathbf{Z}_t = \mathbf{Y}_t + \mathbf{W}_t$. A succession of values of $\mathbf{Z}_t$ in time will again produce an orbit in $\Re^d$, which will describe some geometrical object in that

[5] Strictly speaking "embedding" can be defined as follows. Let M be a compact set in a Banach space B, and E a subspace of finite dimension. Then an "embedding" is a smooth map $\Phi: B \to E$, such that $\Phi(M) \subset E$ is a smooth submanifold and Φ is a diffeomorphism between M and $\Phi(M)$.

space. However, in this case, the overall "viewer" can no longer be represented by a smooth function, the embedding theorem does not hold and consequently we can no longer be sure that the properties of the orbit and the geometrical object associated with it have a relation of diffeomorphic equivalence with the unknown true orbit and the system attractor. Nor can we assume that there exists a map $\mathbf{G}^{(d)}$ characterizing the time evolution of the system.

Even though, in the presence of noise, analysis cannot be solidly based on a neat, rigorous result such as the embedding theorem, numerical investigations of "toy models" indicate that re-construction even from heavily contaminated series can be performed quite successfully. In this case, however, filtering is an essential preliminary step.

For this purpose several methods exist and have extensively been tried both on toy-models and real experimental series. All of these methods have strengths and shortcomings and we cannot expect to find a method which is superior for all kinds of series and all kinds of noise.

In the dynamical system community, SSA was first suggested by Broomhead and King (1986a, 1986b) and, as we have mentioned before, has stimulated the production of a vast literature not always enthusiastic about that method.

In what follows we shall make use of a variation of singular channel SSA first applied (to paleoclimatic series) by Vautard and Ghil (1989), shall extend it to the multidimensional case and shall finally apply it to series generated by some of the best known toy-models.

The first step in the implementation of S-SSA is to construct the so-called "trajectory matrix". If $\{z_t\}$ is the available series, that matrix can be written as follows

$$\mathbf{Z} = (N^*)^{-1/2} \begin{bmatrix} z_1, \ldots\ldots, z_m \\ z_2, \ldots\ldots, z_{m+1} \\ \vdots \\ z_{N^*-m+1}, \ldots, z_{N^*} \end{bmatrix} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_{N^*-m+1} \end{bmatrix} \tag{3.1}$$

The vectors $\mathbf{z}_i$ are called "augmented series"; $(N^*)^{-1/2}$ is a normalizing term and ^{T} denotes transpose.

Now let be $N = N^* - m + 1$. As is well-known, the rectangular $(N \times m)$ matrix $\mathbf{Z}$ can be de-composed as

$$\mathbf{Z} = \mathbf{S}\mathbf{\Sigma}\mathbf{C}^T \tag{3.2}$$

where $\mathbf{S}$ is a $(N \times m)$ matrix whose columns $\mathbf{s}_i$, $(i = 1, \ldots, m)$ are the eigenvectors of the symmetric $(N \times N)$ matrix $\mathbf{Z}\mathbf{Z}^T$; $\mathbf{C}$ is a $(m \times m)$ matrix whose columns $\mathbf{c}_i$, $(i = 1, \ldots, m)$ are the eigenvectors of the symmetric $(m \times m)$ matrix $\mathbf{Z}^T\mathbf{Z}$; $\mathbf{\Sigma}$ is an $(m \times m)$ diagonal matrix whose elements are the positive square roots of the eigenvalues of $\mathbf{Z}^T\mathbf{Z}$. The vectors $\mathbf{s}_i$ and $\mathbf{c}_i$ are also called "singular vectors" and the elements of $\mathbf{\Sigma}$ are called "singular values" of $\mathbf{Z}$. A moment's reflection will show that the matrix $\mathbf{Z}^T\mathbf{Z}$ is the covariance matrix of the series $\{z_t\}$.

The singular vectors are orthogonal, i.e. $\mathbf{c}_i^T\mathbf{c}_j = \mathbf{s}_i^T\mathbf{s}_j = \delta_{ij}$. Thus, the vectors $\mathbf{c}_i$ can be used as an orthonormal basis of the space $\Re^m$ on which m-dimensional points $\mathbf{z}_i$ can be projected. Those vectors are sometimes called "empirical orthogonal functions (EOF)". The columns of the $(N \times m)$ matrix $\mathbf{A} \equiv \mathbf{ZC}$ are called "principal components" and represent the coefficients of projection of vectors $\mathbf{z}_i$ onto the EOFs. Finally, the singular values σ_i can be ordered as $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_m \geq 0$.

As we have just said, the matrix $\mathbf{ZC}$ represents the trajectory matrix projected onto the basis $\mathbf{c}_i$. If we think of a trajectory as exploring on average an ellipsoid of dimension m, the vectors $\mathbf{c}_i$ $(i = 1, \ldots, m)$ correspond to the directions of the principal axes of the ellipsoid, and the values σ_i^2 associated with them correspond to the lengths of those axes. Each singular value σ_i can be looked at as the variance of the i.th principal component.

From the definition of $\mathbf{A}$, we have that $\mathbf{Z} = \mathbf{AC}^T$ and therefore that

$$z_{ij} = \sum_{r=1}^{m} a_{ir} c_{jr} \tag{3.3}$$

In words, this means that we can expand the original series in terms of EOFs and "principal components". We shall call the process of re-covering values of z from EOFs and PCs "re-conversion".

The projection of the trajectory matrix onto the space spanned by the eigenvectors of the covariance matrix of the time series is optimal in the sense that it makes the columns of the (modified) trajectory matrix $\mathbf{ZC}$ linearly independent[6] and minimizes the mean square error of projection. It also provides a first, rough filtering of the series as the projection of the matrix $\mathbf{Z}$ onto the basis $\mathbf{c}_i$ can be viewed as a weighted moving average of the original data.

Suppose now that the series in question is contaminated by noise and let us restrict ourselves to the simpler case in which there is an observational white noise. In this case, if the series is stationary and sufficiently long, the perturbed covariance matrix $\mathbf{Z}^T\mathbf{Z}$ can be approximated as follows

$$\mathbf{Z}^T\mathbf{Z} = \mathbf{Y}^T\mathbf{Y} + (\sigma_w^2/m)\mathbf{I} \tag{3.4}$$

where σ_w^2 denotes the variance due to noise and $\mathbf{I}$ is the $(m \times m)$ identity matrix. Clearly, in this case the singular vectors $\mathbf{c}_i$ of the perturbed trajectory matrix will be the same as those of the unperturbed one, whereas, owing to (white) noise, the singular values will be uniformly increased by an amount σ_w^2/m.

For any given value of the window length m and the level of noise σ_w^2, the signal-to-noise ratio (s.n.r.) associated with each direction - measured by the quantity $m\sigma_i^2/\sigma_w^2$ - will decrease as the order of the corresponding singular value increases and clearly noise can entirely dominate signal for "higher order directions". While, in general, the singular values of a noise-free series will be uniformely declining with the order i, in the presence of sufficiently strong white noise we should observe a plateau - the "noise floor" - in the spectrum. If this is the case, only $k \leq m$ singular values will be above the "noise floor", i.e., we shall have here $\sigma_i^2 > \sigma_w^2$ for $i \leq k$ and $\sigma_i^2 < \sigma_w^2$ for $i > k$.

[6] This can be seen by considering that, from equation (4) we have $(\mathbf{ZC})^T(\mathbf{ZC}) = \mathbf{\Sigma}^2$.

We shall follow Vautard and Ghil and call k "statistical dimension".

It has sometimes been suggested that the rank $n \leq m$ of the covariance matrix $\mathbf{Z}^T\mathbf{Z} = \mathbf{Y}^T\mathbf{Y}$ (in the noise-free case), or the statistical dimension (in the noisy case) can be used to identify the deterministic (or mainly deterministic) subspace, where the hypothetical attractor of the unperturbed system "lives". It has consequently be suggested that n or k can be taken as an estimate of the dimension of the deterministic dynamical system which has supposedly generated the observed data.

Unfortunately, in general this suggestion cannot be accepted, as neither n nor k are invariant with regards to the quality and quantity of data. In particular they are not invariant with respect to the length of the window (or, when the time lag is equal to one, with respect to the embedding dimension).

On the one hand, whenever the time series has been generated by a *nonlinear* dynamical system, the rank of the covariance matrix $\mathbf{Z}^T\mathbf{Z}$ will typically increase *pari passu* with the length of the window (the embedding dimension). This is due to the fact that covariance represents a *linear* relation among variables and it takes an infinite number of linear co-ordinates to represent a nonlinear system, even though the system evolves in a space of small dimension. On the other hand, from our discussion above we gather that, other things being equal, when we increase the window length m, the noise floor σ_w^2/m will be reduced. Therefore, however large the value of m, increasing it further will decrease the level of noise and correspondingly increase the number k of singular values above the noise floor.

4. Re-construction, Filtering and Re-conversion

Suppose now that for a given one-dimensional series $\{z_t\}$, we fix the time lag equal to one (i.e., we utilize all the available data) and we fix somehow the window length (=embedding dimension) m. We then form a trajectory matrix $\mathbf{Z}$, evaluate the singular value spectrum and identify the statistical dimension $k < m$. Finally we "filter" the noise in the series by truncating the expansion (3.3) at the k.th term, projecting the m-dimensional vectors of the tajectory matrix onto k "deterministic" EOFs.

In so doing, we shall of course discard some of the available information on the hypothesis that it is dominated by noise. But, if our final goal is forecasting, we want to predict future (or out-of-sample) values of the original variable z. In the assumed case that contamination is an additive white noise, the best prediction of z is equal, on average, to the prediction of the "clean" variable y.

However, "re-conversion" from a k-truncated expansion (3.3) is, of course, not unique.

The reader can in fact easily verify that in the trajectory matrix $\mathbf{Z}$ there is only one element z_1, two elements $z_2, \ldots$ In general, for each time index t, the number of elements z_t are:

(i) t, when $1 \leq t \leq m-1$,
(ii) m, when $m \leq t \leq N-m+1$,
(iii) $N-t+1$, when $N-m+2 \leq t \leq N$.

In the process of re-conversion, the values assigned to elements of the series $\{z_t\}$ with the same index will generally be different. In order to overcome this problem, we shall follow and extend a procedure suggested by Vautard *et al.* (1992, p. 105).

Let us perform the re-conversion making use of $1 \leq k \leq m$ singular elements, and let us call $\mathcal{A}$ the set of indices $\{1, \ldots, k\}$. We want to find a new series $\{u_t\}$, such that

$$H_{\mathcal{A}}(u) = \sum_{i=1}^{N-m+1} \sum_{j=1}^{m} \left(u_{ij} - \sum_{k \in \mathcal{A}} a_{ik} c_{jk} \right)^2 \tag{4.1}$$

is minimized.

For this purpose, we shall seek a series $\{u_t\}$ whose corresponding "augmented series" are closest in the sense of least squares to the projections of the "augmented series" of $\{z_t\}$ on the selected EOFs (with index $k \in \mathcal{A}$).
The solution of this problem can be written as follows

$$u_t = \frac{1}{t} \sum_{j=1}^{t} \sum_{k \in \mathcal{A}} a_{t-j,k} c_{jk} \qquad \text{per } 1 \leq t \leq m-1 \tag{4.2}$$

$$u_t = \frac{1}{m} \sum_{j=1}^{m} \sum_{k \in \mathcal{A}} a_{t-j,k} c_{jk} \qquad \text{per } m \leq t \leq N-m+1 \tag{4.3}$$

$$u_t = \frac{1}{N-t+1} \sum_{j=1-N+m}^{m} \sum_{k \in \mathcal{A}} a_{t-j,k} c_{j,k} \qquad \text{per } N-m+2 \leq t \leq N \tag{4.4}$$

In words, this cumbersome-looking procedure simply amounts to making averages among different values of z_t with the same index.

The new series $\{u_t\}$ is the filtered series. If $\mathcal{A}$ consists of only one index k, the resulting series $u^{(k)}$ is called "k.th re-converted component (RC)".

It is easy to verify that the original series can be expanded as a sum of RCs, i.e., we have

$$z = \sum_{k=1}^{m} u^{(k)}. \tag{4.5}$$

We have so far assumed that the series to be investigated was one-dimensional. We would now like to extend the analysis to the case in which $l > 1$, i.e., the data is multivariate. In a perfectly noiseless world and with unlimited supply of data, this should not make any difference. In practice, however, availability of multivariate observations can significantly improve the quality of the results. The reason why this is so can be understood by relizing that the ability to probe the output of the system over (variable) space as well as over time, permits one to have a larger number of observations without increasing the window length or decreasing the time lag. However, increasing the window length, if possible at all, can reduce the quality of observation especially when we deal with series characterized by a chaotic deterministic component. Moreover, when covering a very long time interval, the implicit assumption that we are observing the same dynamical mechanism may become untenable. On the other hand decreasing the time lag is often impossible and, when possible, may not improve information significantly, due to too high correlation among observations very near in time. Also, when the lag becomes very short, the assumption that noise is uncorrelated may not be reasonable.

The procedure we have described above can be easily extended to the multivariate case. For simplicity's sake we shall describe in detail only the simpler case $l = 2$. In this case, the set of data will consist of two vectors (assumed of equal lengths) $\{z_t\}_{t=1}^{N^*}$ and $\{v_t\}_{t=1}^{N^*}$. The validity of the basic embedding theorem in the multivariate case is guaranteed by a corollary of the Takens theorem due to Broomhead and King (1986, p. 128)

We can now construct a trajectory matrix with partial window lengths m_1 and m_2 and total window length $m_1 + m_2 = m$. Suppose that $m_1 = m_2$. We can write then[7]:

$$\mathbf{D} = [\mathbf{D}_z \mid \mathbf{D}_v] = \begin{bmatrix} z_1 \ldots\ldots z_{m_1}, v_1 \ldots\ldots v_{m_2} \\ z_2 \ldots z_{m_1+1}, v_1 \ldots\ldots v_{m_2+1} \\ \vdots \qquad\quad \vdots \end{bmatrix} = \begin{bmatrix} d_1 \ldots\ldots d_m \\ d_2 \ldots d_{m+1} \\ \vdots \quad \vdots \end{bmatrix} \tag{4.6}$$

At this point, we can apply the same procedure discussed for the univariate one to the multivariate case . In particular, in the bi-variate case, the covariance matrix can written as

$$\mathbf{T}_D = \begin{bmatrix} \mathrm{Cov}(\mathbf{D}_Z\mathbf{D}_Z) & \mathrm{Cov}(\mathbf{D}_Z\mathbf{D}_V) \\ \mathrm{Cov}(\mathbf{D}_V\mathbf{D}_Z) & \mathrm{Cov}(\mathbf{D}_V\mathbf{D}_V) \end{bmatrix} \tag{4.7}$$

where the matrix blocks on the main diagonal denote the covariance matrices for the variables z and v, respectively, and the the other two blocks denote the cross covariance matrices.

From now on the procedure is the same as for the univariate case. Notice that, when the series are contaminated by (white) noise, singular value analysis automatically discriminates in favour of cleaner series.

Some modifications are necessary in the calculation of "reconstructed components". In particular, when we deal with two variables, $H_{\mathcal{A}}(u)$ can be written as follows

$$H_{\mathcal{A}}(u) = \sum_{i=1}^{N-m+1} \sum_{j=1}^{m} \left(u_{ij} - \sum_{k\in\mathcal{A}} a_{ik}c_{jk} \right)^2 = \tag{4.8}$$

[7]We must notice at this point that if variables are measured in the same unit of measure, their variances may be considerably different from one another. In this case, when evaluating the principal components, variables with larger variance will be given relatively greater weight and this can lead to distortions. Before constructing a multivariate trajectory matrix, we must therefore normalize the series so that all of them have zero mean and variance equal to one.

$$= \sum_{i=1}^{N-m+1} \sum_{j=1}^{m_1} \left(u_{ij} - \sum_{k\in\mathcal{A}} a_{ik}c_{jk} \right)^2 + \sum_{i=1}^{N-m+1} \sum_{j=m_1+1}^{m_1+m_2} \left(u_{ij} - \sum_{k\in\mathcal{A}} a_{ik}c_{jk} \right)^2$$

Naturally, (4.8) will be minimized when both terms on the R.H.S. of the last expression are. On the other hand, each of these terms is analogous to that on the R.H.S. of equation (4.1). Therefore, the solution that minimizes (4.8) can be written as

$$u_t = \left(\frac{1}{t} \sum_{j=1}^{t} \sum_{k\in\mathcal{A}} a_{t-j,k}c_{jk} \;,\; \frac{1}{t} \sum_{j=m_1+1}^{m_1+m_2} \sum_{k\in\mathcal{A}} a_{t-j,k}c_{jk} \right) \tag{4.9}$$

for $1 \le t \le (m/2) - 1$

$$u_t = \left(\frac{1}{m_1} \sum_{j=1}^{m_1} \sum_{k\in\mathcal{A}} a_{t-j,k}c_{jk} \;,\; \frac{1}{m_2} \sum_{j=m_1+1}^{m_1+m_2} \sum_{k\in\mathcal{A}} a_{t-j,k}c_{jk} \right) \tag{4.10}$$

for $(m/2) \le t \le N - (m/2) + 1$

$$u_t = \left(\frac{1}{N-t+1} \sum_{j=1-N+m_1}^{m_1} \sum_{k\in\mathcal{A}} a_{t-j,k}c_{jk} \;,\; \frac{1}{N-t+1} \sum_{j=1-N-(m_1+m_2)}^{m_1+m_2} \sum_{k\in\mathcal{A}} a_{t-j,k}c_{jk} \right) \tag{4.11}$$

for $N - (m/2) + 2 \le t \le N$

Notice that, since principal components (PCs) are linear combinations of the original variables and the re-converted components (RCs) are in their turn linear combinations of PCs, re-construction of the state space by means of RC in general is an embedding and preserves differential information[8]. As in the previous case, we can directly consider the sum of the first k components, or we can consider each of them separately and them sum them up.

The procedure just described can be extended without difficulties (but much longer computations) to the general n-dimensional case.

5. Numerical results

Having thus sketched the theoretical background of our method, we can proceed to analyse the numerical results. We have applied the method in question to the well-known models of Lorenz and Rössler (continuous-time) and Ikeda and Hénon (discrete-time). From these models we have generated series of data, integrating or iterating the relevant differential or difference equations. Notice that we have kept the length of series within 1000 observations, i.e. our series are shorter that those commonly available in physics and other natural sciences

[8]The theoretical justification for this statement can be found in Sauer, Yorke and Casdagli (1991, pp. 596-598).

by one or two orders of magnitude. Since the performance of our method is especially competitive for short series, it seems to suit the needs of economics and social sciences particularly well.

We have assumed that the noise inherent in these numerical computations is neglible *vis-à-vis* the exogenous one. The "clean" series have then been contaminated by adding various amounts (from 10% to 100%) of white noise, measured in terms of the ratio between noise and signal variance.

In order to evaluate the performances of the method in question and to compare them with those of other competing filtering techniques, we have adopted the "performance index" suggested by Grassberger *et al.* (1993, p. 135), namely

$$r_0 = \sqrt{\frac{\sum_{t=1}^{N}(z_t - y_t)^2}{\sum_{t=1}^{N}(u_t - y_t)^2}} \tag{5.1}$$

where z_t denotes a noise-contaminated observation, y_t is the corresponding "clean" observation, u_t is the "re-converted" observation. The index r_0 goes from 1 (no filtering) to ∞ (perfect filtering).

Our results are illustrated in Tables I-IV and in Figures 1-5 and can be summarized as follows:

(i) Multichannel singular value analysis, being able to exploit both temporal and spacial information, performs systematically better than singular value analysis, for given amount of data and the same level of noise (see, e.g., Tables I-IV). This is true even in cases (e.g. the Ikeda model, see. Table II) in which the filtering power of SSA is virtually nil.

Tables I-IV approximatley here

(ii) When the level of noise contamination is different for different variables, M-SSA permits one to exploit information efficiently (i.e., automatically assigning greater weight to "cleaner" variables). For this see Table I-IV and Fig.1.

Fig.1 approximatley here

(iii) M-SSA works particularly well for short, highly contaminated data. In this case, the filtering power of M-SSA compares favourably not only with single channel SSA, but also with truly nonlinear methods such as those discussed in Grassberger *et al.*'s paper (1993).

Fig.2,3,4,5 approximatley here

There are, of course, open questions with our method. M-SSA works much better with continuous-time than with discrete-time models. More specifically, the maximum values of the performance index r_0 are much lower and the level of k corresponding to it much higher in the discrete case. We do not have a clear cut explanation of this fact but its thorough investigation would be of great interest.

A crucial problem with both single and multichannel singular value analysis is the choice of the number of components. Several methods have been suggested

in the literature - e.g. noise floor (Broomhead *et al.*, 1987) or Montecarlo-like methods (Vautard *et al.*, 1992) - each of which has advantages and disadvantages in terms of accuracy and computing costs. We shall defer a thorough discussion of this point to another paper. In our present investigation, where we used pseudo-experimental data and therefore we knew the correct, (almost) noise-free data, we have applied a pragmatical, non-rigorous strategy and selected the values of both m and k which appeared to give the best performance in terms of the index r_0. In so doing, we noticed that the "optimal" number of components is quite stable in the sense that approximately the same value of r_0 can be obtained over a set of neighboring values of k.

Of course, when operating with truly experimental data, there is no way of measuring errors in the same way we did here. In this case, performances of the filtering techniques could be evaluated by methods based on *cross-validation* or out-of-sample predictions (not included in this paper). On the other hand, in those cases in which, for some of the toy models under scrutiny, we performed both out of sample prediction and re-conversion evaluated in terms of the index r_0, the "best" choice of the truncation parameter k resulted the same. Our investigation of this point is preliminary but we signal it for the interested reader.

Remark. The computations in this paper have been performed by a software program written in Mathematica language and nicknamed "MELISSA". The program is available to interested readers who should contact any of the authors by e.mail.

References

[1] Broomhead D.S., King G.P. (1986 a). Extracting qualitative dynamics from experimental data, *Physica D* 20, pp.217-236.

[2] Broomhead D.S., King G.P. (1986 b). On the Qualitative Analysis of Experimental Dynamical Systems, in: Srakar, S. (ed.) *Nonlinear Phenomena and Chaos*, Adam Hilger Ltd, Bristol and Boston.

[3] Broomhead D.S., Jones R., King G.P. (1988). Comment on "Singular-value decomposition and embedding dimension", *Physical Review A*, 37(12), pp.5004-5005.

[4] Grassberger P., Hegger R., Holger K., Carsten S., Schreiber T. (1993). On noise reduction methods for chaotic data, *Chaos*, 3, pp. 127-141.

[5] Mañé (1981). On the dimension of the compact invariant sets of certain nonlinear maps. In Rand, D.A. and Young, L.S. (eds.), *Dynamical Systems and Turbulence*. (Lecture Notes in Mathematics, 898), 230–242. New York: Springer Verlag.

[6] Mees A.I., Rapp P.E. (1988). Reply to "Comment on "Singular-value decomposition and embedding dimension", *Physical Review A*, 37(12), pp.5006-5007.

[7] Mees A.I., Rapp P.E., Jennings L.S. (1987). Singular-value decomposition and embedding dimension, *Physical Review A*, 36 (1), pp.340-346.

[8] N. Packard, J.P. Crutchfield, J.D. Farmer, R.S. Shaw, Phys. Rev. Lett., 45, (1980), p.712.

[9] H. Hotelling (1933). Analysis of a complex of statistical variables in to principal components, *Journal of Educational Phychology*, 24, p.717 and 498.

[10] K. Pearson (1901). On lines and planes of closest fit to systems of points in space, *Phil. Mag.*, 6, 2, p.559.

[11] Paluš M., Dvořák I. (1992). Singular-value decomposition in attractor reconstruction: pitfalls and precautions, *Physica D*, 55, pp.221-234.

[12] Pike E.R., McWhriter J.G., Bertero M., de Mol C. (1984). Generalized information theory for inverse problems in signal processing, *IEEE Proceedings*, v.131.F, n.6, pp.660-667.

[13] Sauer T., Yorke J.A., Casdagli M. (1991). Embedology, *Journal of Statistical Physics*, 65, pp. 576-616.

[14] Stewart G.W. (1993). On the early history of the singular value decomposition, *SIAM Review*, v.35, n.4., pp.551-566.

[15] Takens F. (1981). Detecting strange attractors in turbulence. In Rand, D.A. and Young, L.S. (eds.), *Dynamical Systems and Turbulence*, 366–381. New York: Springer Verlag.

[16] Vautard R., Ghil M. (1989). Singular Spectrum Analysis in Nonlinear Dynamics, with Applications to Paleoclimatic Time Series, *Physica*, D 35, pp. 395-424.

[17] Vautard R., You, Ghil M. (1992). Singular-spectrum analysis: A toolkit for short, noisy chaotic signals, *Physica*, D 58, pp. 95-126.

Fig.1 - M-SSA *vs* S-SSA for the Lorenz system evaluated in terms of the index r_o. *a)* M-SSA for a bi-variate series $\{x, y\}_t$, t=1,...,1000. Noise: 25% noise-to-signal ratio (n.s.r.) for both variables. *b)* same with 25% n.s.r. for x and 10% for y.

Fig.2 - Lorenz attractor: *a*) clean orbit. *b*) noisy orbit with 25% n.s.r.. *c*) Filtered orbit. M-SSA on multivariate $\{x, y, z\}_t$ series with uniform noise (n.s.r. equal to, respectively, 25%, 10%, 10%).

Fig.3 - Lorenz system: *a*) clean and noisy trajectories for 300 observations in time space (25% n.s.r.). *b*) Clean and filtered trajectories for the same values. Filtered multivariate $\{x, y, z\}_t$ series (n.s.r. equal to, respectively 25%, 10%, 10%).

Fig.4 - Rössler attractor: *a*) clean orbit. *b*) noisy orbit (25% n.s.r.). *c*) Filtered orbit. M-SSA on multivariate $\{x, y, z\}_t$ series (n.s.r. equal to, respectively, 25%, 10%, 10%).

Fig.5 - Rössler system: *a*) clean and noisy trajectories for 400 observations in time space (25% n.s.r.). *b*) Clean and filtered trajectories for the same values. Filtered multivariate $\{x, y, z\}_t$ series (n.s.r. equal to, respectively, 25%, 10%, 10%).

HENON n=900; m_i=30		
Noise on the variables X,Y	*Max r_0 for M-SSA (in brackets the value of k)*	*Max r_0 for S-SSA (in brackets the value of k)*
100,100 %	1.643 (18)	1.30 (12)
50,50 %	1.40 (26-29)	1.08 (22-23)
25, 25 %	1.40 (31)	1.01 (28-29)
10, 10 %	1.38 (31)	1 (30)
100, 50 %	1.875 (19)	1.29 (12)
50,25 %	1.60 (25)	1.065 (22)
25, 10 %	1.65 (31)	1.006 (28)

Table I. M-SSA vs. S-SSA for the Hénon system

IKEDA n=900; m_i=30		
Noise on the variables X,Y	*Max r_0 for M-SSA (in brackets the value of k)*	*Max r_0 for S-SSA (in brackets the value of k)*
100,100 %	1.427 (24)	1.19 (12)
50,50 %	2.33 (30)	1.02 (29)
25, 25 %	1.16 (40)	1 (30)
100, 50 %	1.58 (23)	1.22 (13)
25, 10 %	1.19 (39)	1 (30)

Table II. M-SSA vs. S-SSA for the Ikeda system.

LORENZ n=900; m_i=30 step size=0.02		
Noise on the variables X,Y,Z	*Max r_0 for M-SSA (in brackets the value of k)*	*Max r_0 for S-SSA (in brackets the value of k)*
100,100,100 %	4.55 (6)	3.56 (3)
50,50,50 %	4.10 (9)	2.8 (4)
25, 25,25 %	3.67 (11)	2.79 (5)
10, 10,10 %	3.44 (15)	2.20 (6)
25, 10,10 %	4.6 (11)	2.77 (5)
25, 25, - %	3.73 (6-7)	2.5 (5)
25,10,- %	4.95 (6)	2.77 (5)

Table III. M-SSA vs. S-SSA for the Lorenz system.

ROSSLER n=1000; m_i=30 step size=0.08		
Noise on the variables X,Y,Z	*Max r_0 for M-SSA (in brackets the value of k)*	*Max r_0 for S-SSA (in brackets the value of k)*
100,100,100 %	5.37 (8)	3.9 (2)
50,50,50 %	6.33 (7)	3.3 (2)
25, 25,25 %	5.47 (9)	2.95 (3)
10, 10,10 %	4.78 (10)	2.38 (5)
25, 10,10 %	7.22 (8)	3.11 (3)
25,10,- %	3.45 (4)	3.10 (3)
25, - , 10 %	4.59 (7)	3.10 (3)

Table IV. M-SSA vs. S-SSA for the Rössler system.

Fig. 1a

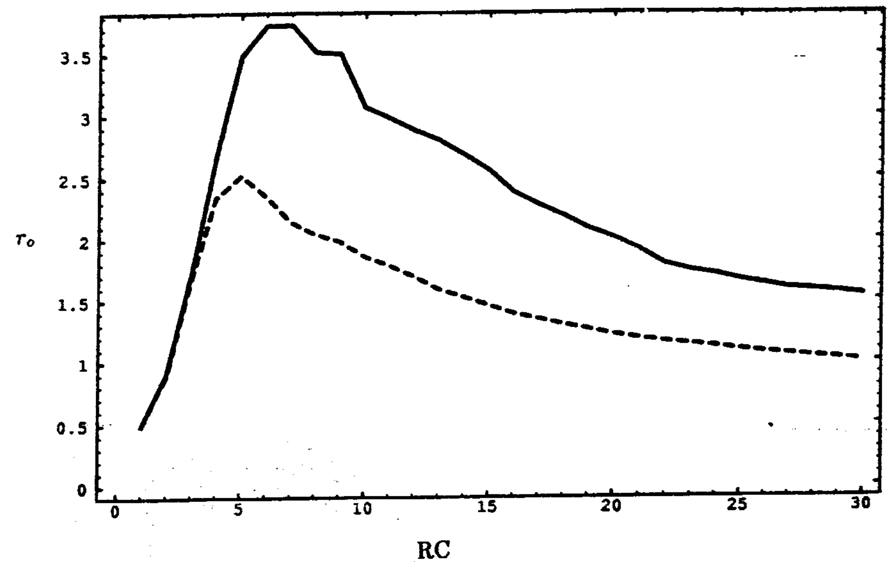


Fig. 1b

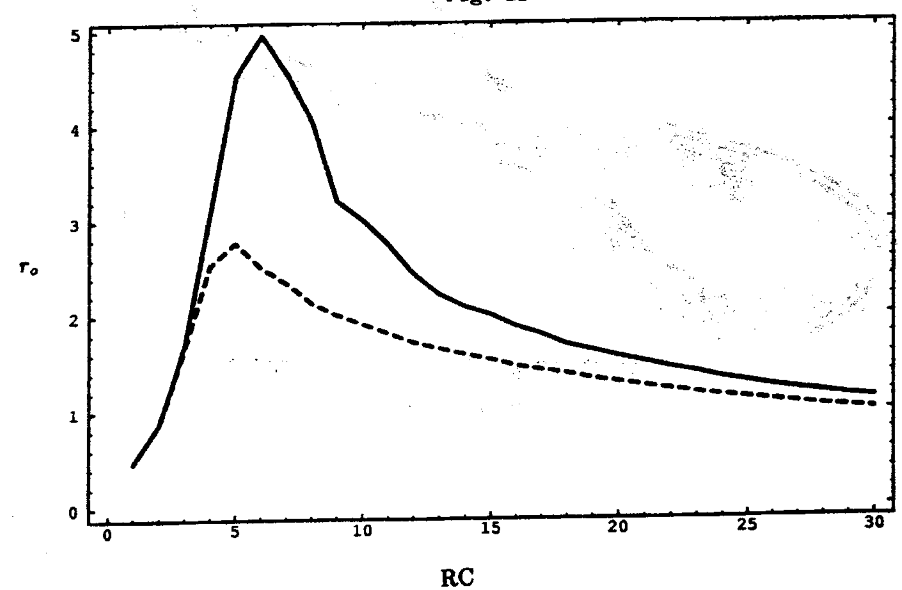
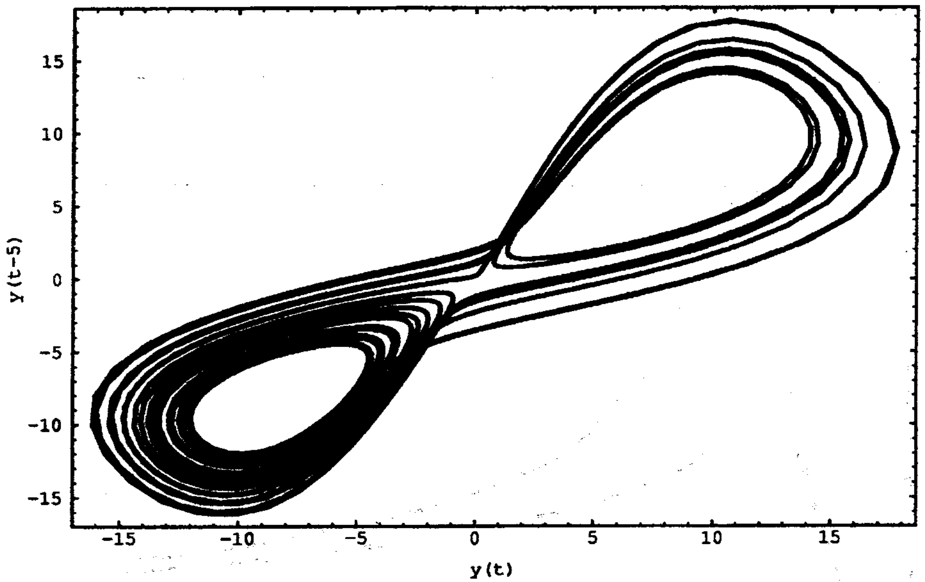
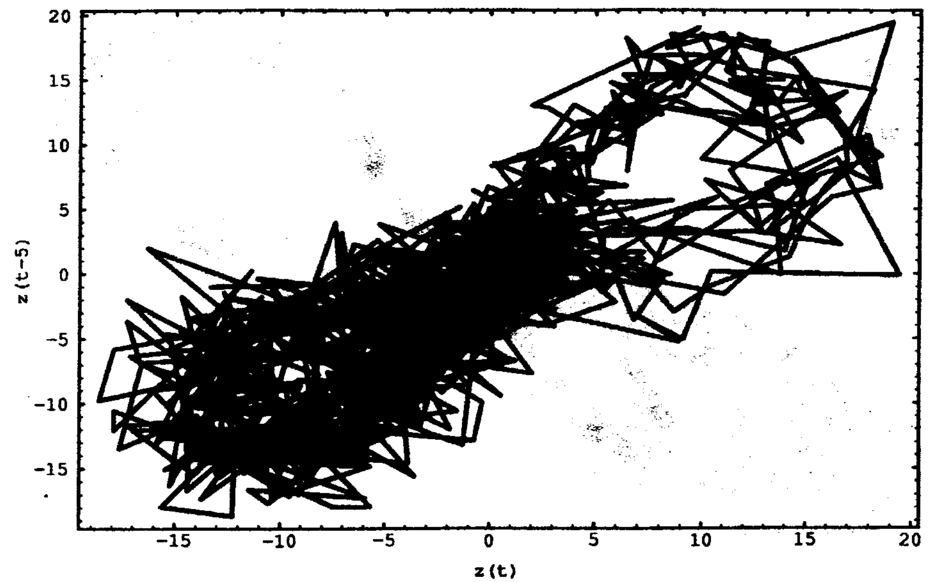


Fig. 2a



15
10
5
0
-5
-10
-15
y(t-5)
-15
-10
-5
0
5
10
15
y(t)

Fig. 2b



20
15
10
5
0
-5
-10
-15
z(t-5)
-15
-10
-5
0
5
10
15
20
z(t)

Fig. 2c

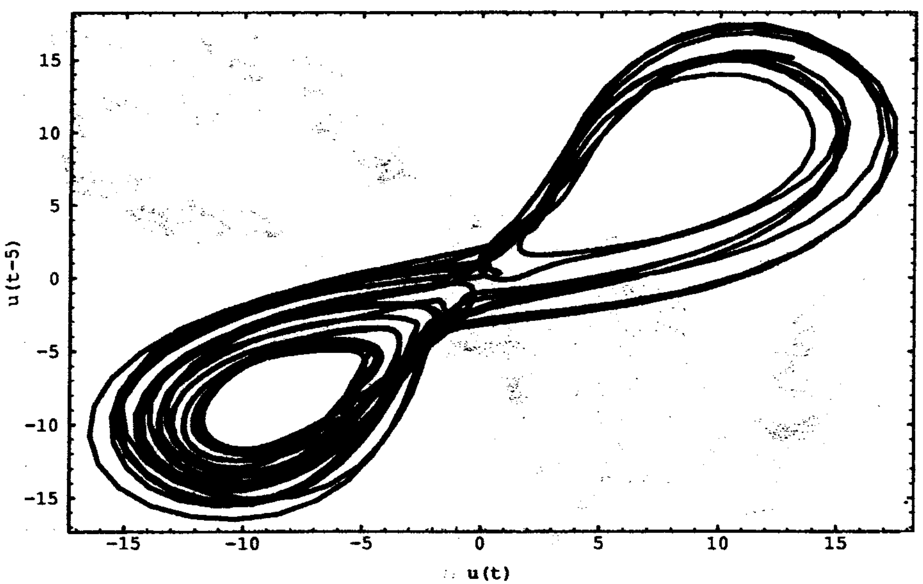


Fig. 3a

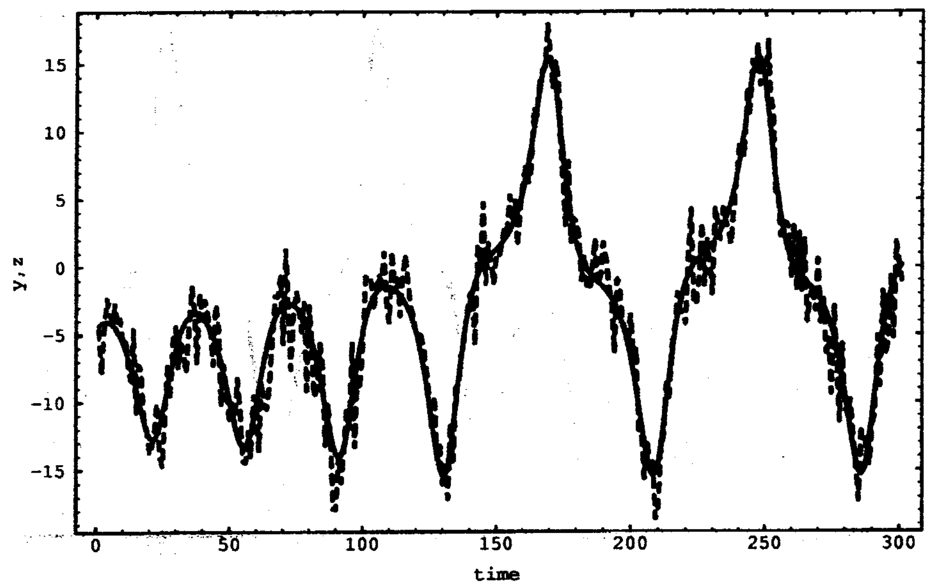
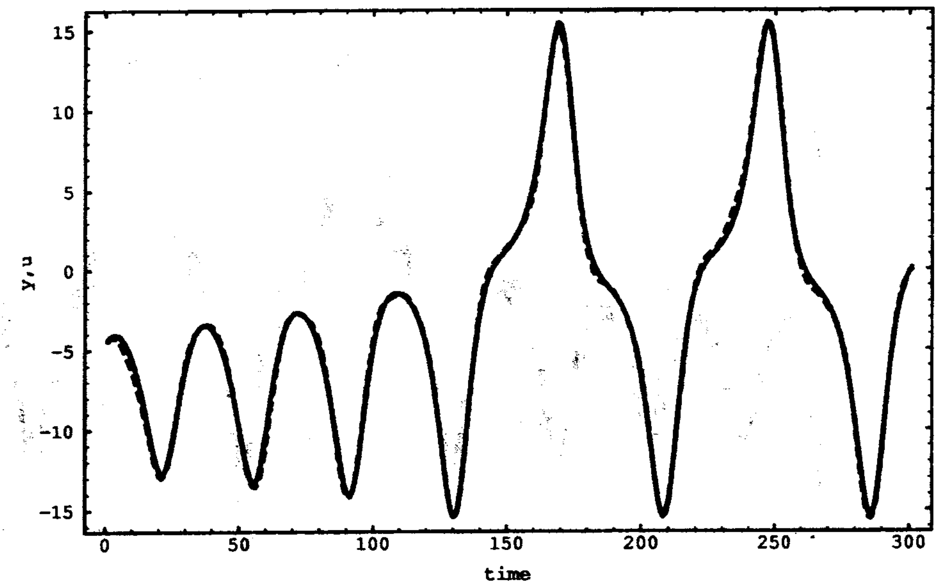
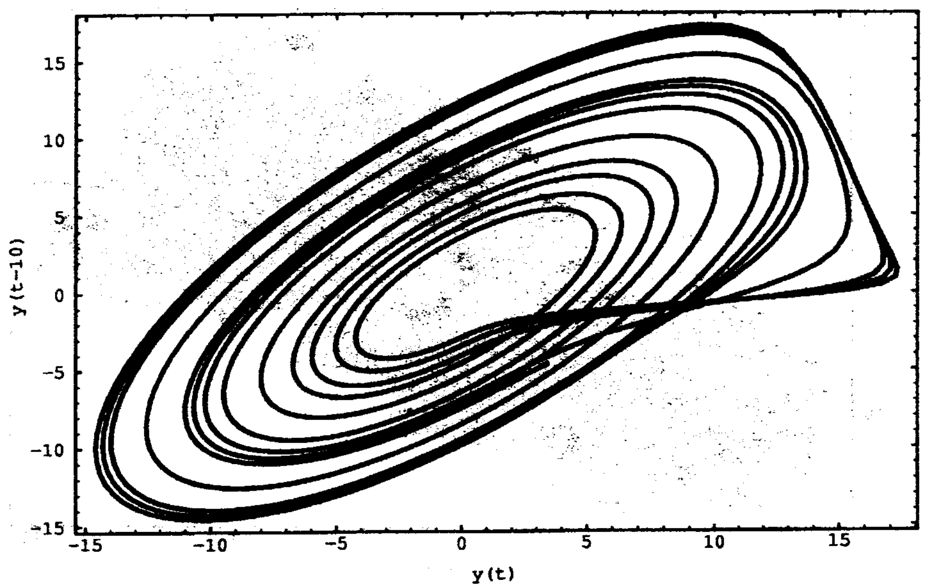


Fig. 3b



15
10
5
0
-5
-10
-15
y,u
0
50
100
150
200
250
300
time

Fig. 4a



15
10
5
0
-5
-10
-15
y(t-10)
-15
-10
-5
0
5
10
15
y(t)

Fig. 4b

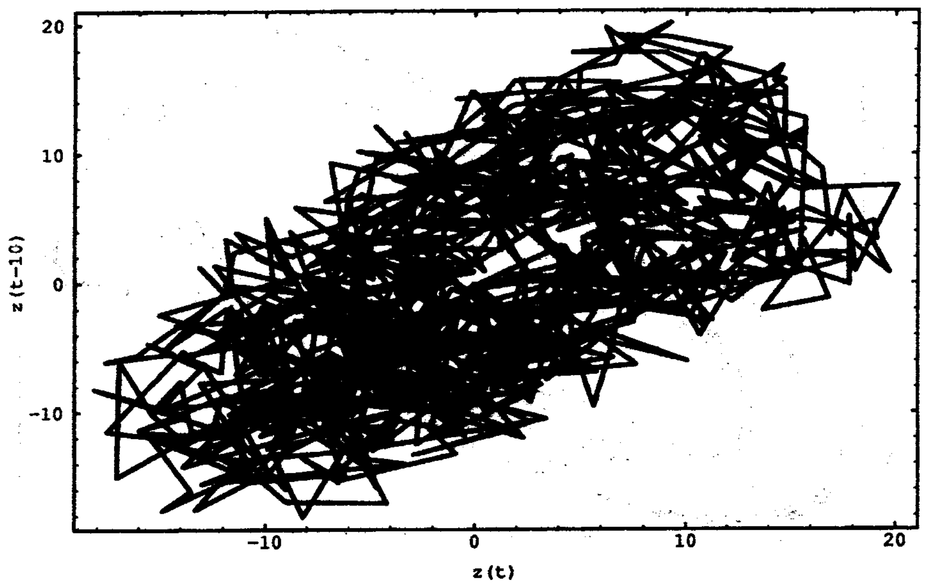


Fig. 4c

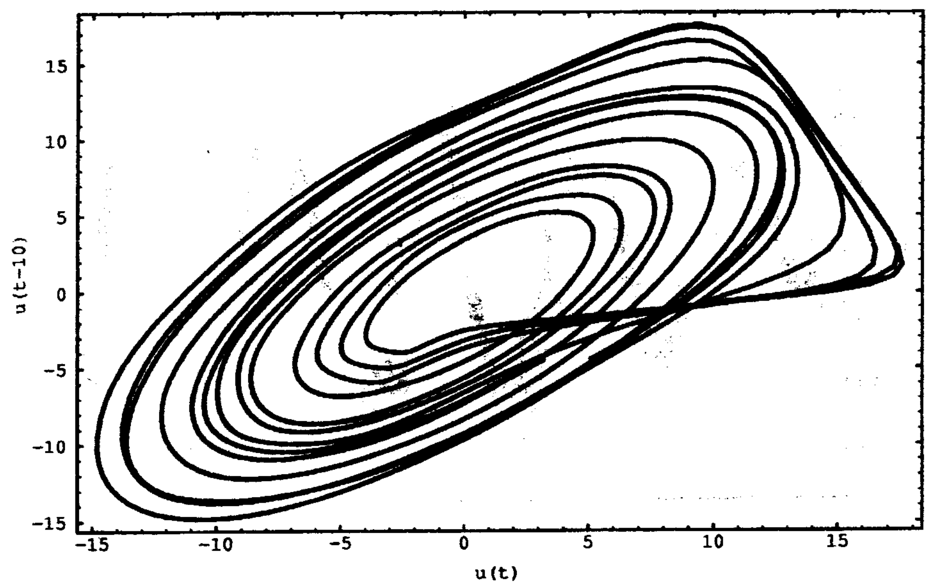
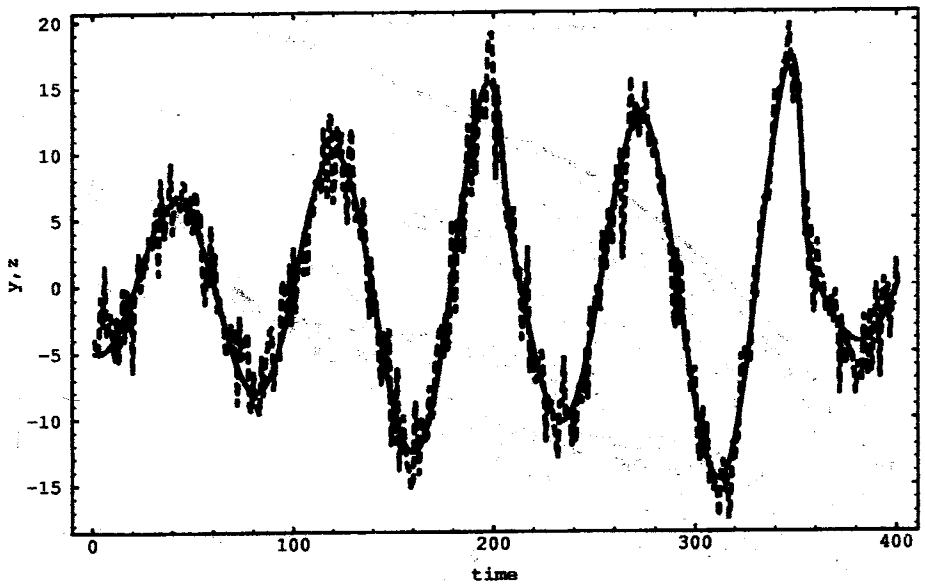
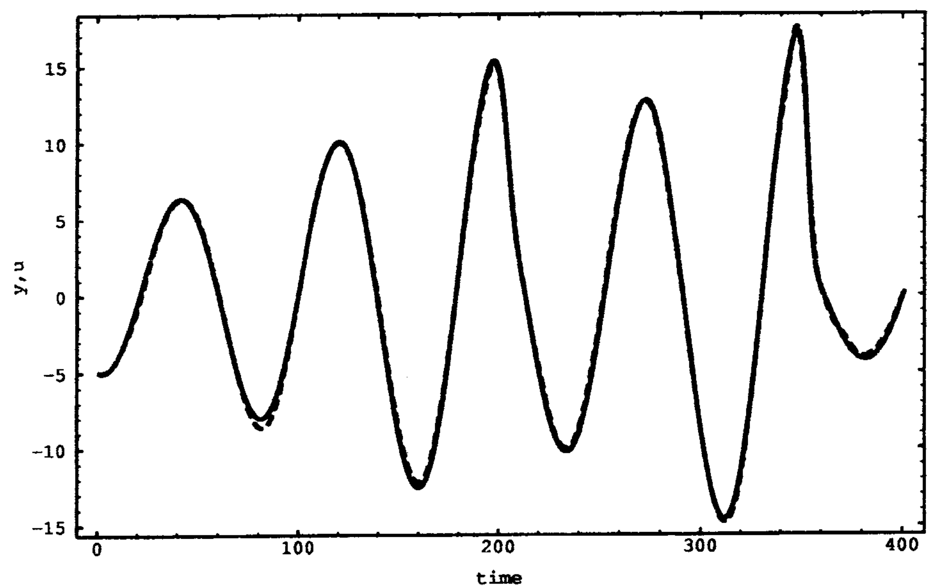


Fig. 5a



20
15
10
5
0
-5
-10
-15
y,z
0
100
200
300
400
time

Fig. 5b



15
10
5
0
-5
-10
-15
y,u
0
100
200
300
400
time