

Reduction of noise of large amplitude through adaptive neighbourhoods

M. Eugenia Mera and Manuel Morán

Dpto. de Fundamentos del Análisis Económico I

*Universidad Complutense, 28223 Madrid. Spain**

(Dated: March 27, 2009)

We propose a noise reduction algorithm based on adaptive neighbourhood selection able to obtain high levels of noise reduction for chaotic vector time series corrupted by observational noises with a noise to signal ratio of up to 300%.

PACS numbers: 05.45.-a, 05.45.Tp, 05.10-a

I. INTRODUCTION

This research treats the noise reduction of vector chaotic time series corrupted by observational noise of large amplitude. The aim is to isolate the deterministic component of the signal. The goal is more than merely detecting chaos. A totally or partially cleaned signal can be used for accurate short term forecasting of its deterministic component, and to improve the forecasting of its probable long term behaviour.

Early research on noise reduction [1–8] was based on local linear estimations of the unknown dynamics using least squares methods. The algorithms proposed in these papers showed the possibility of separating the signal from the noise for time series of large lengths and low or moderate noise amplitudes. These algorithms have two intrinsic limitations that can be insubstantial for noises of small amplitude but which may be crucial for the analysis of chaotic time series corrupted by noises of moderate or large amplitudes:

- (i) The inaccuracy of the least squares method in the local estimation of the dynamics was recognized already in the early literature on noise reduction [7, 9]. All the nonparametric algorithms were, however, based on this method.
- (ii) All noise reduction algorithms use an iterative procedure that takes the output of the algorithm as the input in the next iteration. If the method is to be efficient there must be a significant difference between the noise levels of the outputs of the first and the final iterations. For time series corrupted by noises of large amplitude the neighbourhoods

*Electronic address: mmoranca@ccee.ucm.es

used in estimating the local dynamics in the early iterations must contain a significant fraction of the data points, whereas in the final iterations the use of big neighbourhoods is clearly inappropriate, due to the errors caused by nonlinearities. In order to design an efficient algorithm it is crucial to have efficient criteria for determining the optimal sizes of the neighbourhoods to be used in the local estimations of the dynamics at each iteration.

Issue (i) above was addressed in our previous paper [10] using the theory of *measurement error models* [11]. This theory gives unbiased and consistent estimators which, for Gaussian errors, are those of maximum likelihood. Such estimators can be regarded as a local linear projection onto an optimal linear subspace, with respect to a metric that incorporates the information contained in the error covariance matrix. K. Judd [12] has pointed out the failure of the maximum likelihood methods to identify the true trajectory of a chaotic dynamical system specially in the case of large noise. Judd has also observed that the residuals present too many outliers, and in general their distribution is not consistent with the original distribution of the noise. Our results confirm such observation, the kurtosis of the residuals are considerably higher than the given by the Gaussian distribution. We overcome this difficulty through a smoothing device based on averaging multiple estimations of the value of each clean data point.

In this paper we address the issue of overcoming the basic limitation (ii) pointed out above. Our adaptive neighbourhoods technique is based on a statistical test with null computational cost. At a given iteration it determines from the properties and level of the remaining noise a single uniform size and shape for the neighbourhoods of all data points. The criterion used for determining the neighbourhoods is to keep their sizes as small as possible while guaranteeing, with a given confidence level, that the relevant information is contained in the neighbourhoods. As the number of iterations increases, the noise level decreases, and the sizes of the neighbourhoods reduce accordingly, thus minimizing the errors due to nonlinearity. Another issue addressed when constructing the neighbourhoods is that of determining their optimal shapes. Our algorithm is designed for vector time series, and if the degree of uncertainty in the observed state variables is different then the appropriate neighbourhoods will be nonspherical. They should be instead ellipsoids for which the semiaxial lengths are related to the uncertainties in the corresponding coordinates.

The results that we obtain with our algorithm are good for noises of small and medium amplitudes. Furthermore they are remarkable for short length time series corrupted by noises of large amplitude: we obtain a noise reduction of up to 84% for a Henon time series of only 500 data points with a 300% noise to signal ratio. Such a result is achieved because the algorithm takes full advantage of the heteroskedasticity of the noise (the noise to signal ratio is 10% in the first component and 300% in the second component of the time series). The lower level of uncertainty in the first coordinate allows noise reduction in the second coordinate out of reach for existing algorithms, which have until now

not been used to analyse chaotic noisy time series with noise to signal ratios in such a large range.

Kern, Steeb and Stoop [13] have also proposed an adaptive neighbourhood criterion. At each iteration and for each point they take neighbourhoods of different sizes and choose that which optimizes the quality of the best linear fit to the local dynamics. They showed their method to be advantageous only in the case of noises of medium amplitudes, but for noises of small or large amplitudes they do not find justified the high computational cost required.

Efficient noise reduction algorithms based also on local linear projections into optimal linear subspaces are those developed for speech enhancement. The method consists on minimizing the signal distortion while the residual noise energy is maintained below some given threshold [14–16]. For this type of time series, almost periodic, a method for the automatic detection of the best neighbourhood size, based on recurrence plots, is the given in [17].

The paper is organized as follows. In Section II we describe our algorithm, focusing attention on the local linear estimation of the dynamics using the theory of measurement error models, and on the adaptive neighbourhoods method. Section III presents some results obtained by our algorithm for Henon and Lorenz time series corrupted by noises of large noise to signal ratio.

II. THE ALGORITHM

We assume that the observed time series $\{\mathbf{X}_i, i = 1, \dots, N\} \subset \mathbb{R}^{de}$ is the sum of an unknown deterministic time series $\{\mathbf{s}_i, i = 1, \dots, N\}$ and an unknown *i.i.d.* stochastic time series $\{\mathbf{e}_i, i = 1, \dots, N\}$ with null mean. Then $\mathbf{X}_i = \mathbf{s}_i + \mathbf{e}_i, 1 \leq i \leq N$, and $\mathbf{s}_{i+1} = f(\mathbf{s}_i)$, where $f : M \subset \mathbb{R}^{de} \rightarrow M$ is an unknown smooth chaotic dynamics. The aim of noise reduction algorithms is the separation of the deterministic component of the time series from the observational noise.

The noise level in the j th component of the observed time series is

$$(NL)_j := \sqrt{\frac{\sum_{i=1}^N \left((\mathbf{e}_i)_j \right)^2}{\sum_{i=1}^N \left((\mathbf{s}_i)_j \right)^2}}, \quad j = 1, \dots, de \quad (1)$$

where $(\mathbf{s}_i)_j$ and $(\mathbf{e}_i)_j$ denote the j th components of \mathbf{s}_i and \mathbf{e}_i respectively. We denote by NL the vector whose components are $(NL)_j, j = 1, \dots, de$.

The algorithm we design takes as input data the vector time series $\{\mathbf{X}_i, i = 1, \dots, N\}$ instead of a scalar time series as do most existing noise reduction algorithms.

A. Local estimations of the dynamics and noise reduction

The principal technique used in noise reduction algorithms is local orthogonal projection of the data on optimal linear subspaces, which is equivalent to finding the best local linear estimations of the unknown dynamics. This technique is also used in the forecasting problem, for the estimation of the Lyapunov spectrum [18, 19], and for the estimation from a time series of the degrees of freedom of a dynamical system [20].

Let \mathbf{s}_i be any point of the clean time series, and let U_i be a small neighbourhood of \mathbf{s}_i . Since f is a differentiable dynamics,

$$f(\mathbf{s}) - f(\mathbf{s}_i) \approx Df(\mathbf{s}_i)(\mathbf{s} - \mathbf{s}_i), \text{ for } \mathbf{s} \in U_i$$

where $Df(\mathbf{s}_i)$ is the tangent map of f at \mathbf{s}_i . The tangent space $\mathbf{y} = f(\mathbf{s}_i) + Df(\mathbf{s}_i)(\mathbf{s} - \mathbf{s}_i)$ is the best de -dimensional linear subspace of \mathbb{R}^{2de} through $(\mathbf{s}_i, f(\mathbf{s}_i))$ for the data points $\{(\mathbf{s}, f(\mathbf{s})), \text{ for } \mathbf{s} \in U_i\}$. Such data points are arbitrarily close to this tangent space when the size of U_i is sufficiently small. In [21] it is shown how to estimate $Df(\mathbf{s}_i)$ from a time series and what are the conditions guaranteeing that such estimations converge to the tangent map.

The effect of the noise is to separate the observed data points from the tangent spaces. Then the noise can be partially removed by projecting the observed data points on the best de -dimensional linear subspaces through $(\langle \mathbf{X}_i \rangle, \langle \mathbf{X}_{i+1} \rangle)$ where $\langle \mathbf{X}_i \rangle = \frac{1}{\#U_i} \sum_{j: \mathbf{X}_j \in U_i} \mathbf{X}_j$, $\#U_i$ denotes the number of points within U_i , and U_i is now a small neighbourhood of \mathbf{X}_i . Since the errors have null mean, the point $(\langle \mathbf{X}_i \rangle, \langle \mathbf{X}_{i+1} \rangle)$ is likely to be closer to $(\mathbf{s}_i, \mathbf{s}_{i+1})$ than the point $(\mathbf{X}_i, \mathbf{X}_{i+1})$ is, and it is better to seek the best linear subspace through $(\langle \mathbf{X}_i \rangle, \langle \mathbf{X}_{i+1} \rangle)$ rather than that through $(\mathbf{X}_i, \mathbf{X}_{i+1})$ (see Fig. 1).

A fact to be considered is that the unknown dynamics is chaotic. Then a small ball B_i centred at \mathbf{s}_i is mapped by f onto an ellipsoid centred at \mathbf{s}_{i+1} with the principal axes oriented in the directions of the unstable manifold at \mathbf{s}_i . Therefore a local estimation of f based on the data points $(\mathbf{X}_j - \langle \mathbf{X}_i \rangle, \mathbf{X}_{j+1} - \langle \mathbf{X}_{i+1} \rangle)$, for $\mathbf{X}_j \in U_i$ is likely to be good on the unstable manifold and poor on the stable manifold. Reciprocally the ball B_i is mapped by f^{-1} onto an ellipsoid centred at \mathbf{s}_{i-1} and with the principal axes oriented in the directions of the stable manifold at \mathbf{s}_i . This indicates that in order to obtain good estimations of the clean values of the time series using local estimations of the dynamics we should incorporate information about the past, the present, and the future as was already pointed out by Grassberger et al [3].

This is done by taking the best de -dimensional linear subspace \mathbf{T}_i (according to a criterion described below) for the data points $\{\mathbf{Z}_j - \langle \mathbf{Z}_i \rangle : \mathbf{Z}_j := (\mathbf{X}_{j-1}, \mathbf{X}_j, \mathbf{X}_{j+1}), \mathbf{X}_j \in U_i\}$ and then projecting the points $\mathbf{Z}_j - \langle \mathbf{Z}_i \rangle$ onto \mathbf{T}_i . Then

the estimation of

$$\mathbf{z}_j := (\mathbf{s}_{j-1}, \mathbf{s}_j, \mathbf{s}_{j+1}) \quad (2)$$

obtained from the neighbourhood U_i is

$$\widehat{\mathbf{z}}_j^i = \langle \mathbf{z}_i \rangle + \mathbf{P}_{\mathbf{T}_i}(\mathbf{z}_j - \langle \mathbf{z}_i \rangle) \quad (3)$$

where $\mathbf{P}_{\mathbf{T}_i} \mathbf{z}$ denotes the orthogonal projection, with respect to the metric used to obtain \mathbf{T}_i , of the vector \mathbf{z} onto \mathbf{T}_i . The optimization criterion is the minimization of the mean square distance of the points $\mathbf{z}_j - \langle \mathbf{z}_i \rangle$ to the subspace, taking as the metric that induced by $\widehat{\Sigma}_3^{-1}$ where $\widehat{\Sigma}_3^{-1}$ is the empirical covariance matrix of the errors $(\mathbf{e}_{j-1}, \mathbf{e}_j, \mathbf{e}_{j+1})$ contained in \mathbf{z}_j , and then

$$\mathbf{T}_i := \arg \min_{\mathbf{T}} \sum_{j: \mathbf{x}_j \in U_i} \mathbf{v}_j^t \widehat{\Sigma}_3^{-1} \mathbf{v}_j \quad (4)$$

where $\mathbf{v}_j := \mathbf{z}_j - \langle \mathbf{z}_i \rangle - \mathbf{P}_{\mathbf{T}}(\mathbf{z}_j - \langle \mathbf{z}_i \rangle)$ and $\mathbf{P}_{\mathbf{T}} \mathbf{z} := \arg \min_{\mathbf{u} \in \mathbf{T}} (\mathbf{u} - \mathbf{z})^t \Sigma_3^{-1} (\mathbf{u} - \mathbf{z})$ (see Appendix A). The distance induced by Σ_3^{-1} , called the Mahalanobis distance in the statistical literature, takes into account that the independent variables \mathbf{s}_j in the underlying linear model in the variables $(\mathbf{s}_{j-1}, \mathbf{s}_j, \mathbf{s}_{j+1})$ are also measured with error, and exploits in the optimal way the information about the structure of the error, in particular the degree of uncertainty in each of the coordinates of the time series. The solution of (4) gives unbiased and consistent estimators of the parameters of the model [11], which are also those of maximum likelihood if the errors are Gaussian.

If Σ_3 is a multiple of the identity matrix, (4) is called the orthogonal least squares solution, and the Mahalanobis distance coincides with the Euclidean distance. Hegger and Schreiber [4] have also proposed a noise reduction algorithm for vector time series based on local orthogonal projections onto optimal linear subspaces, but taking as the metric on \mathbb{R}^{3de} one that gives almost all the weight to the de -central coordinates of the data points. This weight is the same for the de -coordinates, so all the individual coordinates are treated as they have the same degree of uncertainty. In the case of scalar time series, Grassberger et al [3] have proposed a metric which is designed to exploit the characteristics of the embedded time series, and Cawley and Hsu [1] use the Euclidean metric.

B. The adaptive neighbourhood construction

In the case of a clean time series, \mathbf{T}_i is obtained as in (4) using the neighbourhood $\{\mathbf{s}_j : (\mathbf{s}_j - \mathbf{s}_i)^t (\mathbf{s}_j - \mathbf{s}_i) \leq r_0^2\}$, where r_0 is fixed (and should tend to zero as N tends to infinity). For the estimation of a de -dimensional linear subspace there are needed at least $de + 1$ points \mathbf{s}_j such that the corresponding vectors $\mathbf{z}_j - \langle \mathbf{z}_i \rangle$ (see (2)) are linearly

independent. The value of r_0 must guarantee this condition. Furthermore, since f is not a linear function the clean data do not belong to the tangent spaces, and the estimates of \mathbf{T}_i have more statistical robustness if the number of points used is greater than $de + 1$.

For noisy time series, the optimal subspace obtained using the neighbourhood $\{\mathbf{X}_j : (\mathbf{X}_j - \mathbf{X}_i)^t(\mathbf{X}_j - \mathbf{X}_i) \leq r_0^2\}$ may be far from the optimal linear subspace for the clean data. This is due to the noise, which introduces false neighbours in the neighbourhoods, and it separates points which are close in the clean time series. The neighbourhoods must be sufficiently large as to guarantee that a significant portion of the data within them corresponds to close neighbours for the clean time series. It is desirable to eliminate from the neighbourhoods as many false neighbours as possible. Furthermore, the sizes of such neighbourhoods should be reduced in accordance with the noise reduction occurring as the iterative process progresses.

On the other hand, since the uncertainties of the individual coordinates of the time series may be different, the Euclidean distance is not the most appropriate for the construction of the neighbourhoods. We use the distance induced by $\widehat{\Sigma}_1^{-1}$ where $\widehat{\Sigma}_1$ is the estimate of the $de \times de$ covariance matrix of the error in the time series \mathbf{X} . This gives a neighbourhood of the form $\{\mathbf{X} : (\mathbf{X} - \mathbf{X}_i)^t \widehat{\Sigma}_1^{-1} (\mathbf{X} - \mathbf{X}_i) \leq r\}$ where r is a value which depends on the noise level and on the value r_0 we consider suitable in the case of N noise-free data points. The value of r is fixed as small as possible while guaranteeing, with a given confidence level $1 - \alpha_0$, that the relevant information is contained in the neighbourhoods (see in Appendix B the details). Fig. 1 shows these neighbourhoods at two points of a noisy Henon time series.

C. Further details of the algorithm

Estimation of Σ_1 and Σ_3 .

An estimate $\widehat{\Sigma}_3$ of the sample covariance matrix of the errors in $\{\mathbf{Z}_j, j = 2, \dots, N - 1\}$ is needed for the estimation of the subspaces \mathbf{T}_i . An estimate $\widehat{\Sigma}_1$ of the sample covariance matrix of the errors in $\{\mathbf{X}_j, j = 1, \dots, N\}$ is needed in constructing the neighbourhoods U_i . We estimate the errors at iteration $k > 1$ from the corrections made by the algorithm at such an iteration.

Let $\{\widehat{\mathbf{s}}_i^k, i = 1, \dots, N\}$ be the cleaned time series at iteration k , which is also the input time series at iteration $k + 1$, and let $\widehat{\mathbf{s}}_i^0 = \mathbf{X}_i, i = 1, \dots, N$. At the first iteration we take $\widehat{\Sigma}_1 = \mathbf{I}$ where \mathbf{I} is the $de \times de$ identity matrix, and as U_i the neighbourhoods which contain a fixed percentage NP of the data points, corresponding to the points closest to \mathbf{X}_i in the Euclidean distance. The value of NP we choose depends on an a priori estimation of the noise level. At

iteration $k + 1$, $k > 1$, we take as $\widehat{\Sigma}_1$ the sample covariance matrix of the estimated errors

$$\widehat{\mathbf{e}}_j^k := \widehat{\mathbf{s}}_j^k - \widehat{\mathbf{s}}_j^{k-1}, \quad j = 1, \dots, N, \quad (5)$$

at iteration k , and as $\widehat{\Sigma}_3$ the sample covariance matrix of its 3-embedding.

Minimum and maximum number of neighbours

We check that all the neighbourhoods U_i contain at least a given number $N_{\min} \geq de + 1$ of neighbours, where N_{\min} is a parameter of the algorithm. It applies to the final iterations mainly to compensate the reduced sizes of the confidence ellipsoids that for large noise amplitudes and large k result from the underestimation of the real errors by the estimated errors given by (5). This is because expression (5) reflects the convergence of the outputs of the algorithm, but does not incorporate the remaining errors contained in such outputs.

In the results of the next section the value of N_{\min} taken at a given iteration is $N_{\min} := \max\{0.014N, N_{mean} - s_N\}$ where N_{mean} and s_N are the sample mean and the sample standard deviation of the distribution of the number of points within each U_i at that iteration.

Another parameter of the algorithm is the maximum N_{\max} of the number of points permitted to be contained in a neighbourhood U_i . It serves to ensure that at initial iterations and for noise levels larger than 100%, the neighbourhoods do not contain too large a fraction of the data points. In the results of Section III we take $N_{\max} = 0.25N$.

Embedding space

We have assumed that the dimension d of the space on which the unknown dynamics f is defined coincides with the number de of coordinates of the observed noisy time series (see in [20, 22] how to estimate d when it is unknown). However our algorithm can be adapted to the case $d \neq de$. Assume that the observed time series $\{\mathbf{X}_i, i = 1, \dots, N\} \subset \mathbb{R}^{de}$ satisfies $\mathbf{X}_i = g(\mathbf{s}_i) + \mathbf{e}_i$ where $\{\mathbf{e}_i, i = 1, \dots, N\} \subset \mathbb{R}^{de}$ is an unknown *i.i.d.* stochastic time series, $\mathbf{s}_{i+1} = f(\mathbf{s}_i)$, $i = 1, \dots, N - 1$, where $f : M \subset \mathbb{R}^d \rightarrow M$ is the unknown smooth dynamics, and $g : M \subset \mathbb{R}^d \rightarrow \mathbb{R}^{de}$ is an unknown smooth vector function or observable. Then \mathbf{T}_i will be the best d -dimensional subspace instead of the best de -dimensional subspace (i.e. the matrix \mathbf{B} in expression (A1) in Appendix A has as columns the vectors $\{\mathbf{w}_{3de-d+1}, \dots, \mathbf{w}_{3de}\}$). Furthermore if $d > de$ and $3de < 2d + 1$ then the working space \mathbb{R}^{3de} (i.e. the 3-embedding of the time series) must be replaced by $\mathbb{R}^{(2m+1)de}$ with m such that $(2m + 1)de > 2d + 1$ in order to guarantee we have a good reconstruction [23] of the unknown dynamics.

III. NUMERICAL RESULTS

In this section we show the results obtained applying our algorithm to time series generated by Hénon and Lorenz dynamics corrupted by noise. The Hénon map is given by the equations

$$x_1(k+1) = 1 - ax_1(k)^2 + x_2(k), \quad x_2(k+1) = bx_1(k), \quad (6)$$

and we take the parameter values $a = 1.4$ and $b = 0.3$. The Lorenz dynamics is defined by

$$x'_1 = \sigma(x_2 - x_1), \quad x'_2 = x_1(R - x_3) - x_2, \quad x'_3 = x_1x_2 - bx_3,$$

The equations were integrated using a fourth order Runge-Kutta algorithm with an integration step of $\Delta t = 0.001$.

The values used for the sampling time τ and the parameters σ , R , and b are given in Table II.

In all the examples below, we use uncorrelated and heteroskedastic errors with Gaussian distributions. For the construction of the neighbourhoods we take $\alpha_0 = 0.01$ (see Appendix B).

We quantify the noise removed by the algorithm using the pointwise distance between the clean \mathbf{s} and the cleaned $\widehat{\mathbf{s}}$ time series

$$d_P(\mathbf{s}, \widehat{\mathbf{s}}) := \left(\frac{1}{N} \sum_{i=1}^N (\|\widehat{\mathbf{s}}_i - \mathbf{s}_i\|_2)^2 \right)^{1/2}.$$

If $d_P(\mathbf{s}, \widehat{\mathbf{s}}) < d_P(\mathbf{s}, \mathbf{X})$ then the noise level in $\widehat{\mathbf{s}}$ is less than the noise level in the input time series \mathbf{X} . The percentage of global pointwise noise reduction is $R_p := 100 \left(1 - \frac{d_P(\mathbf{s}, \widehat{\mathbf{s}})}{d_P(\mathbf{s}, \mathbf{X})} \right)$ and that corresponding to the j th coordinate is $(R_p)_j := 100 \left(1 - \frac{d_P((\mathbf{s})_j, (\widehat{\mathbf{s}})_j)}{d_P((\mathbf{s})_j, (\mathbf{X})_j)} \right)$ where

$$d_P((\mathbf{s})_j, (\widehat{\mathbf{s}})_j) := \left(\frac{1}{N} \sum_{i=1}^N ((\widehat{\mathbf{s}}_i)_j - (\mathbf{s}_i)_j)^2 \right)^{1/2}, \quad j = 1, \dots, de.$$

Notice that these measures require knowledge of the clean time series. We use such knowledge only to quantify the noise level reduction. This information is used neither in the noise reduction scheme nor in deciding when the algorithm must stop. Thus the algorithm may work on data generated by an unknown process. The stopping criterion we use is related to our adaptive neighbourhood construction. As the iterations increase, the remaining noise level and the size of the neighbourhood U_i must decrease. A clear indication that further iterations of the algorithm will not reduce the remaining noise is the stabilization of the mean number of points in the neighbourhoods.

We focus our attention on time series of short lengths corrupted by noises with large amplitudes. We intend to show through numerical experiments that the combination of the measurement error theory and the adaptive

neighbourhood construction allows us to reduce noises having large amplitudes in one of the components especially if the other components of the time series are corrupted by noises of smaller amplitudes.

The results for a Henon time series corrupted by noise of large amplitude (up to 300%) in the second component and by noise of small or moderate amplitude in the first component are shown in Table I. For short length time series and high noise levels the results of the algorithm depend strongly on the realization of the error term, and also on the clean time series considered. Thus the noise level reduction achieved by any noise reduction algorithm must be understood in terms of the distribution of the output of the algorithm. For this reason we give as results the sample mean $\langle (R_p)_2 \rangle$ and the sample standard deviation $s_{(R_p)_2}$ of $(R_p)_2$ obtained from 50 noisy time series.

The results show values of $\langle (R_p)_2 \rangle$ higher than 80% in all the cases. These values increase with N whereas $s_{(R_p)_2}$ decreases with N . For noise levels $(NL)_2$ up to 200% the $\langle (R_p)_2 \rangle$ achieves values greater than 90% for time series with $N = 5000$. The results are very good even for time series of only 500 data points, although they present greater variability. For instance, for $N = 500$ and $NL = (5\%, 200\%)$ we obtain $\langle (R_p)_2 \rangle = 85.00\%$ and $s_{(R_p)_2} = 3.14$, and the 50 values of $(R_p)_2$ belong to the interval $(72.2\%, 89.62\%)$.

Fig. 2 show the output of the algorithm for one of the 50 noisy time series corresponding to a noise level $NL = (10\%, 300\%)$ and $N = 500$. The figure shows that even for a time series having such a large noise level and such a short length the algorithm is able to recover a significant part of the geometric structure of the clean time series.

The use of adaptive neighbourhoods is more relevant for noises of large amplitude. Then the mean number of neighbours in the initial iterations may be close to N_{\max} whereas in the final iterations it is stabilized and close to N_{\min} . For instance, for a Henon time series with $N = 1000$ and $NL = (10\%, 300\%)$, the mean number of neighbours in U_i at the first iteration is $N_{\max} = 250$ and at iteration 20 this mean number is only 25.

Another measure of noise reduction standard in the noise reduction literature is based on the deviation of the time series from the deterministic behaviour. It assumes that the dynamics f is known and is defined through

$$d_{dyn}(\hat{\mathbf{s}}) := \left(\frac{1}{N-1} \sum_{i=1}^{N-1} (\|\hat{\mathbf{s}}_{i+1} - f(\hat{\mathbf{s}}_i)\|_2)^2 \right)^{1/2}. \quad (7)$$

If $d_{dyn}(\hat{\mathbf{s}}) < d_{dyn}(\mathbf{X})$ then the cleaned time series emulates better the dynamics than does the input time series.

The percentage of global dynamical noise reduction is $R_{dyn} := 100 \left(1 - \frac{d_{dyn}(\hat{\mathbf{s}})}{d_{dyn}(\mathbf{X})} \right)$ and that corresponding to the j th coordinate is $(R_{dyn})_j := 100 \left(1 - \frac{(d_{dyn}(\hat{\mathbf{s}}))_j}{(d_{dyn}(\mathbf{X}))_j} \right)$ where

$$(d_{dyn}(\hat{\mathbf{s}}))_j := \left(\frac{1}{N-1} \sum_{i=1}^{N-1} \left((\hat{\mathbf{s}}_{i+1})_j - (f(\hat{\mathbf{s}}_i))_j \right)^2 \right)^{1/2}, j = 1, \dots, de.$$

If f is unknown a measure \hat{R}_{dyn} of the dynamical noise reduction is obtained by replacing f in (7) by its local linear

estimations.

The values of $\langle R_p \rangle$, $\langle R_{dyn} \rangle$, and $\langle (R_{dyn})_2 \rangle$ for the time series of Table I differ slightly from $\langle (R_p)_2 \rangle$ and are omitted for the sake of clarity. The only significant difference is for $NL = (50\%, 300\%)$, where the values of $\langle R_p \rangle$ and $\langle R_{dyn} \rangle$ are around 60% and 70% respectively, which are values significantly smaller than the values of $\langle (R_p)_2 \rangle$ in Table I.

We have tested in all these examples that the stopping criterion based on the stabilization of the mean number of points within the neighbourhoods gives results analogous to those obtained using as the stopping criterion the stabilization of R_p or R_{dyn} , which criteria require knowledge of the clean time series or f . A stopping criterion based on the stabilization of \widehat{R}_{dyn} , which does not require additional information, gives poorer results because \widehat{R}_{dyn} stabilizes at high values even when R_p and R_{dyn} indicate that the algorithm is still able to reduce a significant part of the remaining noise.

Our algorithm estimates very effectively the initial noise levels in each of the components of a noisy time series. Such estimations are given by $(\widehat{NL})_j := \sqrt{\frac{\sum_{i=1}^N ((\mathbf{x}_i - \widehat{\mathbf{s}}_i)_j)^2}{\sum_{i=1}^N ((\widehat{\mathbf{s}}_i)_j)^2}}$, $j = 1, \dots, de$. For instance, the mean of the values of $(\widehat{NL})_1, (\widehat{NL})_2$ obtained from 50 noisy time series of 1000 data points with an initial noise levels $NL = (2\%, 100\%)$ are $\widehat{NL} = (2.012\%, 100.803\%)$ and the corresponding standard deviations are $(0.074, 2.89)$.

Some results obtained for Lorenz time series are shown in Table II. These two examples differ in the values of the parameters in the Lorenz dynamics, and in the sampling times. In both cases $NL = (100\%, 10\%, 5\%)$ so the noise level is large in the first component and moderate in the other two components. In fact, since the mean value of the z component of the clean time series is not zero, $(NL)_3 = 5\%$ is equivalent to 15% noise level in terms of the standard deviation of z in the clean time series. In this case f is unknown and the values of $\langle R_{dyn} \rangle$ and $\langle (R_{dyn})_1 \rangle$ of the table are obtained by approximating f using the fourth order Runge-Kutta method used to generate the clean time series. The values of the dynamical noise reduction are significant greater than the values of pointwise noise reduction, achieving values of $\langle (R_{dyn})_1 \rangle$ close to 95% for the second set of parameters, for all the lengths considered. Fig. 3 show the output of one of the noisy time series of Table II. It can be seen that the algorithm is able to recover a significant part of the geometric structure and the dynamics of the clean time series.

IV. CONCLUSIONS

We have proposed a noise reduction algorithm that combines the measurement error models theory with an adaptive neighbourhoods selection based on a statistical test. The algorithm works efficiently for short time series corrupted by noises of large amplitude. The results could be improved were consideration of nonlinearity effects added to our

adaptive neighbourhood method. The size of the neighbourhoods should be fixed as a compromise between the effects of nonlinearity (see [3]) and the amount of information in the neighbourhoods.

Our results show the possibility of detection and forecasting of deterministic scalar time series of small amplitudes buried in observational noises of amplitudes much larger than that of the signal if they are coupled through a smooth dynamical system to other state variables which can be observed with lower uncertainties. The high performance of our algorithm for short length time series gives hope that the scope of application of this method might extend to research in the social sciences.

APPENDIX A: TECHNICAL DETAILS

(i) Estimation of the best subspaces and the projection matrix

The solution of (4) is (see Appendix in Ref. [10]) $\mathbf{T}_i := \text{span}\{\widehat{\Sigma}_3 \mathbf{w}_1, \dots, \widehat{\Sigma}_3 \mathbf{w}_{3de}\}$ where $\{\mathbf{w}_1, \dots, \mathbf{w}_{3de}\}$ is an orthonormal basis of eigenvectors of the matrix $M_{\mathbf{Z}\mathbf{Z}} := \frac{1}{\#U_i} \sum_{j: \mathbf{X}_j \in U_i} (\mathbf{Z}_j - \langle \mathbf{Z}_i \rangle)(\mathbf{Z}_j - \langle \mathbf{Z}_i \rangle)^t$ in the metric of $\widehat{\Sigma}_3$, i.e.

$$M_{\mathbf{Z}\mathbf{Z}} \mathbf{w}_j = \lambda_j \widehat{\Sigma}_3 \mathbf{w}_j, \quad j = 1, \dots, 3de \quad \text{and} \quad \mathbf{w}_j \widehat{\Sigma}_3 \mathbf{w}_k = \delta_{jk}, \quad j, k = 1, \dots, 3de$$

where $\lambda_1 \geq \dots \geq \lambda_{3de}$ are the corresponding eigenvalues. Such a basis of eigenvectors is given by the columns of the matrix $Q\Lambda^{-1/2}H$, where Q is a matrix whose columns are an orthonormal basis of eigenvectors of $\widehat{\Sigma}_3$ (in the Euclidean metric), Λ is the diagonal matrix of eigenvalues of $\widehat{\Sigma}_3$, and the columns of H are an orthonormal basis of eigenvectors of the matrix $\Lambda^{-1/2}Q^t M_{\mathbf{Z}\mathbf{Z}} Q \Lambda^{-1/2}$.

The matrix of the projection onto the subspace \mathbf{T}_i is $\mathbf{I} - \widehat{\Sigma}_3 \mathbf{B}\mathbf{B}^t$ where \mathbf{I} is the $3de \times 3de$ identity matrix and \mathbf{B} is the matrix whose columns are $\{\mathbf{w}_{2de+1}, \dots, \mathbf{w}_{3de}\}$. Then (3) can be written as

$$\widehat{\mathbf{z}}_j^i = \mathbf{Z}_j - \widehat{\Sigma}_3 \mathbf{B}\mathbf{B}^t (\mathbf{Z}_j - \langle \mathbf{Z}_i \rangle) \quad \text{for } \mathbf{X}_j \in U_i \quad (\text{A1})$$

(ii) The problem of multiple $3de$ -dimensional estimations

Since any point \mathbf{X}_j belongs to many neighbourhoods U_i , we take the average of all the values $\{\widehat{\mathbf{z}}_j^i : \mathbf{X}_j \in U_i\}$ (see (3)) as the final estimation of \mathbf{z}_j , i.e.

$$\widehat{\mathbf{z}}_j := \frac{1}{\#\{i : \mathbf{X}_j \in U_i\}} \sum_{\{i: \mathbf{X}_j \in U_i\}} \widehat{\mathbf{z}}_j^i, \quad j = 2, \dots, N-1. \quad (\text{A2})$$

This is a $3de$ -dimensional time series that is not the 3-embedding of a de -dimensional time series, so there remains the problem of estimating the deterministic part of the observed time series.

(iii) The estimation of the deterministic component

Let $(\widehat{\mathbf{z}}_j)_i$ denote the i th coordinate of $\widehat{\mathbf{z}}_j$, $j = 2, \dots, N - 1$. Since $\widehat{\mathbf{z}}_j = (\widehat{\mathbf{s}}_{j-1}, \widehat{\mathbf{s}}_j, \widehat{\mathbf{s}}_{j+1})$ this time series gives as estimations of $(\mathbf{s}_j)_i$, $j = 3, \dots, N - 2$, the three values $(\widehat{\mathbf{z}}_{j-1})_{2de+i}$, $(\widehat{\mathbf{z}}_j)_{de+i}$, and $(\widehat{\mathbf{z}}_{j+1})_i$. These are the estimations of \mathbf{s}_j obtained from the past, the present, and the future. We take a weighted average of these three values as the estimation of $(\widehat{\mathbf{s}}_j)_i$, $i = 1, \dots, de$, with weights w_l^i , $l = 1, \dots, 3$ related with the sample variances of the coordinates of the corrections $\epsilon := \mathbf{Z} - \widehat{\mathbf{z}}$ and then

$$(\widehat{\mathbf{s}}_j)_i = \sum_{l=1}^3 w_l^i (\widehat{\mathbf{z}}_{j-1+(l-1)})_{(3-l)de+i} \quad \text{for } j = 3, \dots, N - 2; \quad i = 1, \dots, de. \quad (\text{A3})$$

This cleaned time series $\{\widehat{\mathbf{s}}_j, j = 1, \dots, N\}$ will be the input time series in the next iteration of the algorithm.

The combination of the averages (A2) and (A3) increases the statistical robustness of the estimations, without requiring controls for avoiding large corrections as do many noise reduction algorithms. These smoothing devices partially overcome the problems of the maximum likelihood estimates pointed out by K. Judd [12].

APPENDIX B: THE ADAPTIVE NEIGHBOURHOOD CONSTRUCTION

If we assume that the errors $\{\mathbf{e}_i\}$ have a $N(0, \Sigma_1)$ distribution then $(\mathbf{X}_i - \mathbf{s}_i)^t \Sigma_1^{-1} (\mathbf{X}_i - \mathbf{s}_i)$ has a χ^2 -distribution with de degrees of freedom, and

$$\{\mathbf{X} : (\mathbf{X} - \mathbf{X}_i)^t \Sigma_1^{-1} (\mathbf{X} - \mathbf{X}_i) \leq \chi_{de}^2(\alpha_0)\}$$

provides a $100(1 - \alpha_0)\%$ confidence ellipsoid for \mathbf{s}_i , where $\chi_{de}^2(\alpha_0)$ is the number such that $\Pr\{\chi_{de}^2 > \chi_{de}^2(\alpha_0)\} = \alpha_0$.

Since in general Σ_1 is unknown, it is replaced by the sample covariance matrix $\widehat{\Sigma}_1$. Let $T^2 := (\mathbf{X}_i - \mathbf{s}_i)^t \widehat{\Sigma}_1^{-1} (\mathbf{X}_i - \mathbf{s}_i)$. Then $\frac{(N-de)}{de(N-1)} T^2$ is distributed (see [24] page 176) as an F distribution with degrees of freedom de and $N - de$.

A $100(1 - \alpha_0)\%$ confidence ellipsoid for \mathbf{s}_i is given by

$$\left\{ \mathbf{X} : (\mathbf{X} - \mathbf{X}_i)^t \widehat{\Sigma}_1^{-1} (\mathbf{X} - \mathbf{X}_i) \leq \frac{de(N-1)}{N-de} F_{de; N-de}(\alpha_0) \right\} = \left\{ \mathbf{X} : \sum_{j=1}^{de} \left(\frac{(\mathbf{Y} - \mathbf{Y}_i)_j}{\sqrt{f_0 d_j}} \right)^2 \leq 1 \right\}$$

where $\mathbf{Y} := \mathbf{M}^t \mathbf{X}$, $(\mathbf{Y})_j$ is the j th coordinate of \mathbf{Y} , \mathbf{M} is a matrix whose columns are an orthonormal basis of eigenvectors of $\widehat{\Sigma}_1$, $\{d_j, j = 1, \dots, de\}$ are the eigenvalues of $\widehat{\Sigma}_1$, and $f_0 := \frac{de(N-1)}{N-de} F_{de; N-de}(\alpha_0)$. It is an ellipsoid centred at \mathbf{X}_i , whose axes are the eigenvectors of $\widehat{\Sigma}_1$ and with j th semi-axis of length $\sqrt{f_0 d_j}$ (see Fig. 4(a)). Since we want a confidence ellipsoid for the points $\{\mathbf{s}_j : (\mathbf{s}_j - \mathbf{s}_i)^t (\mathbf{s}_j - \mathbf{s}_i) \leq r_0^2\}$ we take $2\sqrt{f_0 d_j} + r_0$ as the length of the

j th semi-axis instead of $\sqrt{f_0 d_j}$ (see Fig. 4(b)), and therefore

$$U_i := \left\{ \mathbf{X} : \sum_{j=1}^{de} \left(\frac{(\mathbf{Y} - \mathbf{Y}_i)_j}{2\sqrt{f_0 d_j} + r_0} \right)^2 \leq 1 \right\} \subset \mathbb{R}^{de} \quad (\text{B1})$$

is the neighbourhood of \mathbf{X}_i that we consider.

We refine this neighbourhood slightly by removing from U_i those points which are neighbours only due to the noise. We construct a neighbourhood $V_i \subset \mathbb{R}^{3de}$ of \mathbf{Z}_i using the same procedure we use to build the neighbourhood U_i of \mathbf{X}_i , and we remove from U_i the points \mathbf{X}_j for which $\mathbf{Z}_j = (\mathbf{X}_{j-1}, \mathbf{X}_j, \mathbf{X}_{j+1}) \notin V_i$.

A parameter needed for the construction of the neighbourhoods U_i at each iteration is r_0 . In the results of Section III we take $r_0 = r_{mean} + s_r$ where r_{mean} and s_r are the sample mean and the sample standard deviation of the distribution of radii corresponding to the minimum radius r needed to get $0.014N$ neighbours in the balls centred at \mathbf{X}_i , $i = 1, \dots, N$, and with radius r with respect to the Euclidean metric. Since the noise is reduced as the iterations increase, the value of r_0 decreases.

ACKNOWLEDGMENTS

This research has been supported by the Ministerio de Ciencia y Tecnología, research project MTM2006-023721.

-
- [1] R. Cawley and G-H Hsu, Phys. Rev. A **46**, 3057 (1992).
 - [2] M. Davies, Physica D **79**, 174 (1994).
 - [3] P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath and T. Schreiber, Chaos **3**, 127 (1993).
 - [4] R. Hegger and T. Schreiber, Phys. Lett. A **170**, 305 (1992).
 - [5] E. J. Kostelich and T. Schreiber, Phys. Rev. E **48**, 1752 (1993).
 - [6] E. J. Kostelich and J. A. Yorke, Phys. Rev. A **38**, 1649 (1988).
 - [7] E. J. Kostelich and J. A. Yorke, Physica D **41**, 183 (1990).
 - [8] T. Sauer, Physica D **58**, 193 (1992).
 - [9] L. Jaeger and H. Kantz, Chaos **6**, 449 (1996).
 - [10] M. E. Mera and M. Morán, Chaos **16**, 013116 (2006).
 - [11] W. A. Fuller, *Measurement error models*. Wiley Series in Probability and Statistics. (John Wiley & Sons, 2006).
 - [12] K. Judd, Phys. Rev. E **75**, 036210 (2007).
 - [13] A. Kern, W-H Steeb and R. Stoop, Int. J. Modern Phys. C **11**, 125 (2000).

- [14] Y. Ephraim, IEEE Trans. Speech Audio Process. **3**, 251(1995).
- [15] M.T. Johnson and R.J. Povinelli, Physica D **201**, 306 (2005).
- [16] X. Luo, J. Zang and M. Small, Phys. Rev. E **72**, 046710 (2005).
- [17] L. Matassini, H. Kantz, J. Holyst and R. Hegger Phys. Rev. E **65**, 021102 (2002).
- [18] J-P. Eckmann, S. O. Kamphorst, D. Ruelle and S. Ciliberto, Phys. Rev. A **34**, 4971 (1986).
- [19] M. E. Mera and M. Morán, Erg. Theory Dyn. Syst. **20**, 531 (2000).
- [20] M. E. Mera and M. Morán, J. Stat. Phys. **106**, 125 (2002).
- [21] M. E. Mera and M. Morán, J. Math. Anal. Appl. **235**, 454 (1999).
- [22] A. Leontitsis, T. Bountis and J. Pagge, Chaos **14**, 106 (2004).
- [23] F. Takens in *Dynamical Systems and turbulence*, edited by D. A. Rang and L. -S. Young, Lecture Notes in Mathematics Vol. 898 (Springer, Berlin, 1981), pp. 366-381.
- [24] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis* (John Wiley & Sons, 2003).

TABLE I: Sample mean $\langle(RP)_2\rangle$ and sample standard deviation $s_{(RP)_2}$ (in brackets) of $(RP)_2$ for Henon time series corrupted by Gaussian noise with the noise to signal levels in each component given in the first row of the table.

	$NL = (2\%, 100\%)$	$NL = (5\%, 200\%)$	$NL = (10\%, 300\%)$	$NL = (50\%, 300\%)$
$N = 500$	83.28 (2.38)	85.00 (3.14)	84.21 (3.96)	80.66 (2.96)
$N = 1000$	87.34 (2.06)	87.83 (2.35)	86.68 (2.91)	81.82 (2.57)
$N = 5000$	92.46 (1.22)	90.73 (1.04)	88.31 (1.50)	82.92 (1.51)

TABLE II: Sample means and sample standard deviations (in brackets) of $(RP)_1$, $(R_{dyn})_1$, (RP) , and (R_{dyn}) in Lorenz time series corrupted by Gaussian noises with the noise to signal levels in each component given in the first row of the table.

N	$NL = (100\%, 10\%, 5\%)$ $\sigma = 16, R = 45.92, b = 4, \tau = 0.03$				$NL = (100\%, 10\%, 5\%)$ $\sigma = 10, R = 28, b = 8/3, \tau = 0.02$			
	$\langle(RP)_1\rangle$	$\langle(R_{dyn})_1\rangle$	$\langle RP \rangle$	$\langle R_{dyn} \rangle$	$\langle(RP)_1\rangle$	$\langle(R_{dyn})_1\rangle$	$\langle RP \rangle$	$\langle R_{dyn} \rangle$
1500	76.73 (7.54)	86.65 (7.43)	69.70 (13.9)	83.94 (8.13)	79.80 (3.05)	94.03 (1.15)	73.79 (4.3)	86.47 (1.32)
3000	78.83 (5.49)	88.21 (3.63)	71.86 (11.78)	74.47 (2.77)	81.79 (2.34)	94.74 (0.53)	76.23 (3.85)	87.13 (0.75)
5000	81.10 (1.08)	89.79 (0.70)	74.78 (1.12)	86.91 (0.72)	82.32 (1.73)	94.90 (0.81)	77.01 (1.99)	87.27 (0.74)

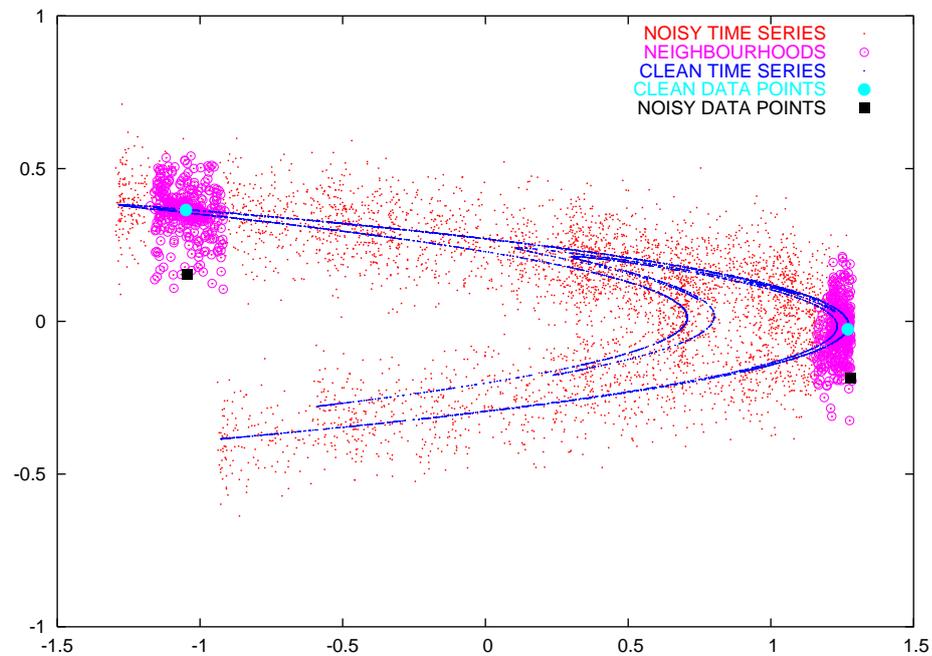


FIG. 1: Confidence ellipsoids for $\alpha_0 = 0.1$ at two points of a Henon time series with $N = 5000$ data points corrupted by Gaussian noise with $NL = (1\%, 40\%)$. Observe the difference between the clean (blue circles) and the noisy data (black squares) at which the neighbourhood is constructed and the convenience of taking the best linear subspaces through the mean points of the neighbourhoods instead of taking them through the observed noisy points.

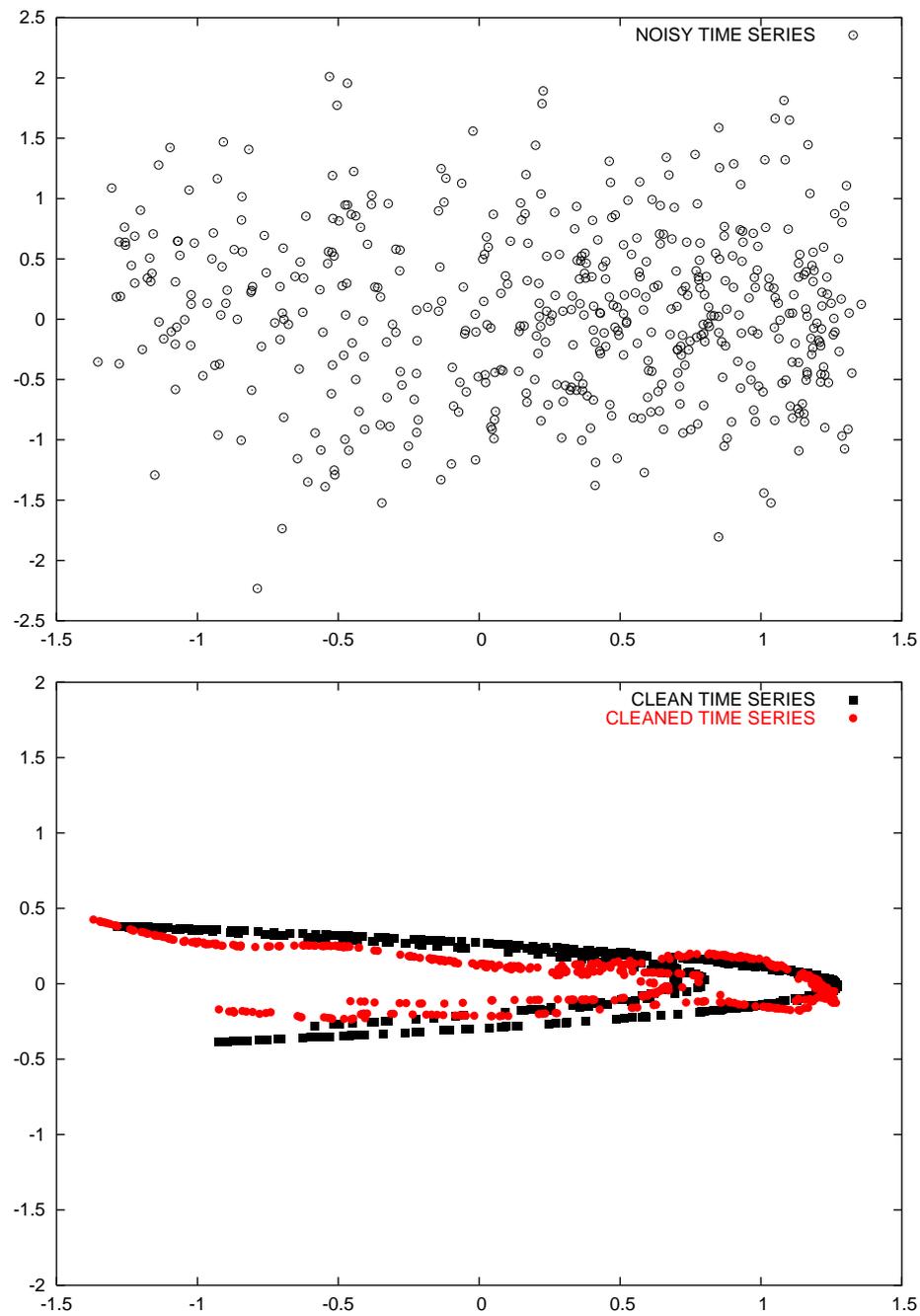


FIG. 2: (a) Noisy and (b) Clean and cleaned time series of a Henon time series with $N = 500$ corrupted by Gaussian noise with $NL = (10\%, 300\%)$.

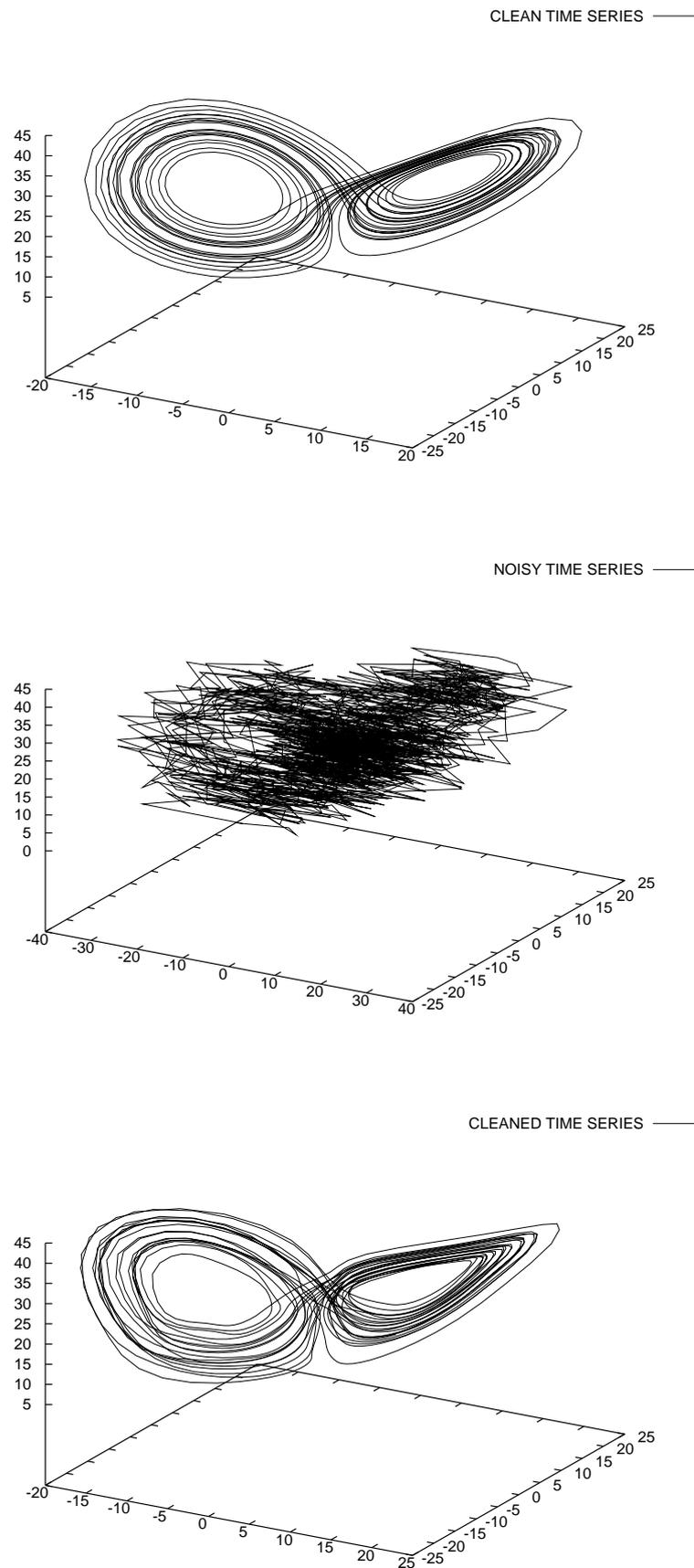


FIG. 3: (a) Clean, (b) noisy, and (c) cleaned time series of 1500 data points generated by Lorenz dynamics ($\sigma = 10$, $R = 28$, $b = 8/3$, and sampling time $\tau = 0.02$) corrupted by Gaussian noise with $NL = (100\%, 10\%, 5\%)$.

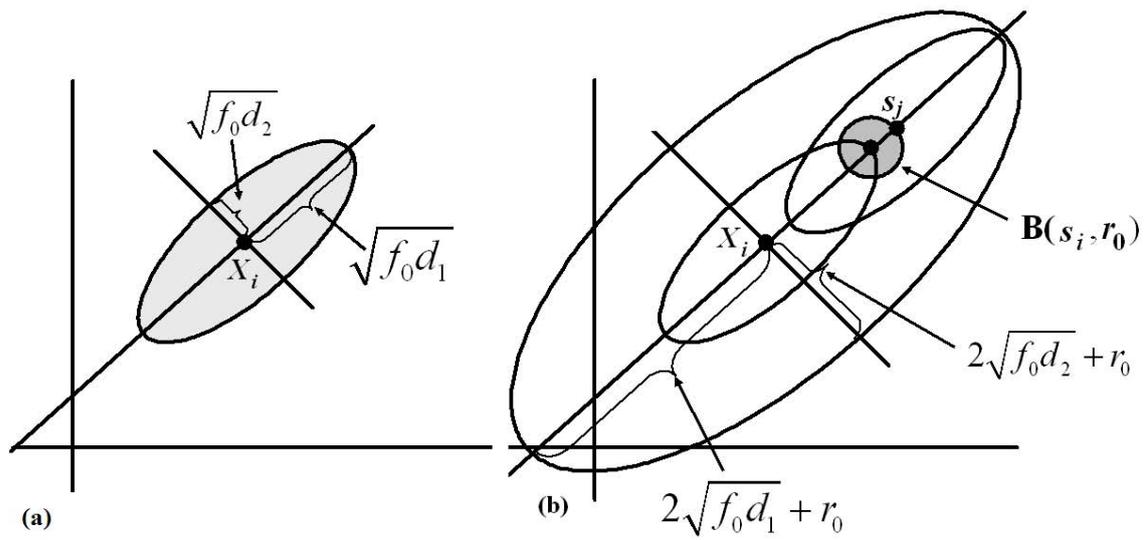


FIG. 4: (a) Confidence ellipsoid for \mathbf{s}_i , and (b) confidence ellipsoid for $B(\mathbf{s}_i, r_0)$.