Reducing noise in discretized time series

Milena C. Cuéllar and P.-M. Binder
Departamento de Física, Universidad de Los Andes, Apartado Aéreo 4976, Bogotá, Colombia
(Received 28 February 2001; revised manuscript received 11 June 2001; published 21 September 2001)

We show that applying a noise-reduction algorithm to a discretized time series increases its average error, compared to the original series. We find that adding external noise comparable to the discretization step before noise reduction limits the increase of the average error and improves the estimation of Lyapunov exponents.

DOI: 10.1103/PhysRevE.64.046211 PACS number(s): 05.45.Tp, 06.20.Dk, 05.40.Ca, 42.55.−f

I. INTRODUCTION

Time series analysis has been one of the most successful areas of nonlinear science [1–4]. There have been considerable advances in the areas of attractor reconstruction, noise reduction, model inference, control, synchronization, and prediction. This paper addresses discretized time series, that arise from finite resolution in the measurement itself, the channel through which the measured signal is transmitted, or the recording device. In a previous paper (Ref. [5], Fig. 5) we showed that the portion of a signal lost in a finite-resolution measurement is highly correlated, and apparently low dimensional, contrary to previous expectations [4] of an uncorrelated, uniform distribution. This in principle should make noise reduction (also called cleaning or filtering) in such series impossible.

In the present work we study a two-dimensional map as a controlled example [6], and confirm our earlier conjecture that the apparently cleaned data obtained from applying a noise-reduction algorithm to discretized data, in fact has *higher* average error, although it may look smoother to the naked eye. This finding is relevant to discretized time series from lasers [7,8], population biology [9], the social sciences [10], or the stock market, among others. We also find that, contrary to intuition, adding external noise with width comparable to the discretization step to the discretized series *before* cleaning, tends to reduce the increase in average error and improves the estimation of Lyapunov exponents.

II. METHOD

In our work we have used a well-known dissipative system, the Hénon map [6], with the standard parameters $a=1.4$, $b=0.3$. For this map the embedding dimension is $d_E=2$ and the optimal time delay for reconstruction is $\tau=1$ timestep. A zoom of a reconstruction of the x coordinate, i.e., an x_{n+1} vs x_n plot, is shown in Fig. 1 for comparison with noisy and cleaned versions of the same map. As in previous work [5], we have normalized the x variable between 0 and 1. We have chosen the simple, yet well-tested nonlinear cleaning algorithm introduced in Refs. [4,11], which we will describe below in more detail. This algorithm has been implemented in the time series analysis program TISEAN [12], which we have used.

We discretize the normalized map by rounding up, i.e., by defining a discretization step $\delta_m=2^{-m}$ and replacing each x value by the nearest higher integer multiple of δ_m. The data of Fig. 1, discretized with $m=7$ is shown in Fig. 2(a). Part (b) of the same figure shows the reconstruction of the discretized data, to which we have added uniform, random noise of width $\delta_l \sim \delta_m$. We refer to these modifications of the original data as the discretized (D) and smoothed (S) series, respectively. The corresponding data sets are denoted by d_n and s_n. Support for the choice of noise width comparable to the discretization step will be given below.

Next, we have applied the simple noise-reduction algorithm of Refs. [4,11] to the discretized and smoothed data, with $5 \leqslant m \leqslant 8$, which correspond to noise widths between 0.4% and 3%. These are within the range of performance of the simple algorithm, and have been combined with different widths of external noise.

The algorithm works as follows. One constructs vectors with $m=d_E$ components, through standard phase-space reconstruction [3]. We denote these by $\mathbf{x}_i$. To clean each m-dimensional vector $\mathbf{x}_i=(x_{i-m-1},\ldots,x_i)$ one identifies vectors close to $\mathbf{x}_i$ such that each component is within a neighborhood of radius r of the corresponding component of the vector to be cleaned. After this, one calculates the cleaned value of $x_{i+m/2}$, the central component of the vector, by averaging over the central component of the close vectors, with the hope that the noise has expectation value of zero, and will approximately cancel out. In the worst case, if $\mathbf{x}_i$ has no neighbors, the cleaned value will be the unchanged value of the central component of $\mathbf{x}_i$.

The algorithm has several parameters: one is the embedding dimension d_E, already mentioned. While d_E for the Hénon map is 2, Refs. [4,11] recommend a higher number

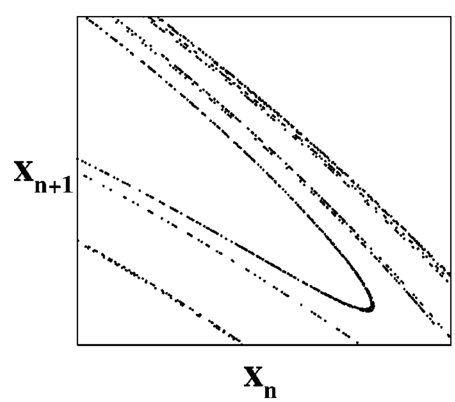

FIG. 1. Enlargement of the original Hénon map in time-delayed x coordinates, normalized between 0 and 1. We show the region $0.6 \leqslant x_n \leqslant 0.8$ and $0.7 \leqslant x_{n+1} \leqslant 0.8$ in dimensionless units.

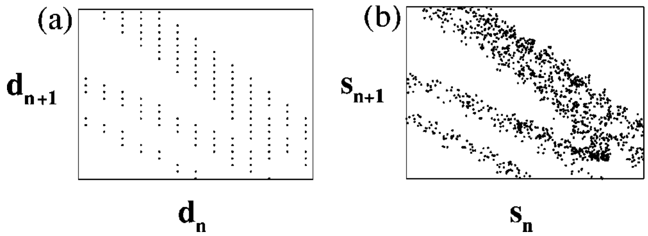

FIG. 2. (a) Zoom of the Hénon map, with discretization $m=7$ (steps of 1/128 of the variable range). (b) Zoom of the Hénon map, with discretization $m=7$ plus external noise of the same width. We show the same region as the previous figure.

(they use 7). We use 5, which we found to be slightly more convenient [13]. Another parameter is the radius of the neighborhood, r, that defines close points or trajectories. The same references suggest about three times the noise amplitude. We have swept radius values between 1 and m times the noise width. Finally, the algorithm can be iterated successively, i.e., a cleaned set of points can be used as the starting point for a new cleaning. The same references suggest 2–6 iterations. We have monitored several iterations, and show the fourth in what follows.

III. RESULTS

We show typical results, that refer to the original, D, and S series shown in the previous two figures. We denote the noise-reduced points of the D and S series with d_n^r and s_n^r, respectively. Figure 3 shows (a) the cleaned result of the D series with a neighborhood radius $r=2\delta$, and (b) the cleaned result of the S series, with the same radius. In both cases there is an apparent sharpening of lines in the attractor, while in the latter case there appears to be more similarity with the original signal (Fig. 1) than in the former case. To make the comparison more quantitative, consider an altered series, $n_1, n_2, \ldots, n_N$ and the original, noise-free signal, $x_1, x_2, \ldots, x_N$, both consisting of N points. We define the average percentage error as

$$e=\frac{100}{N}\sum_{i=1}^{N}|n_i-x_i|. \tag{1}$$

We calculated e for the D and S series before and after noise reduction, using the original Hénon series as the noise-free signal. A representative case of these results is shown in Fig. 4 for $m=6$. While the error plots for other values of m are qualitatively similar, we have used the case $m=6$ be-

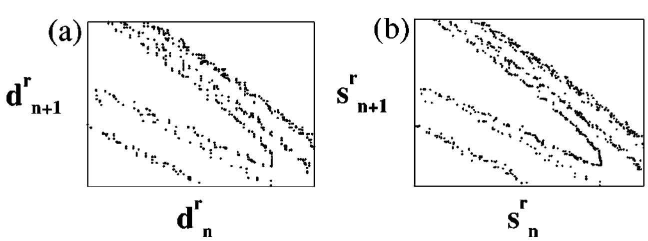

FIG. 3. Same data as Fig. 2, after noise reduction with $r=2\delta$ and four iterations of the algorithm.

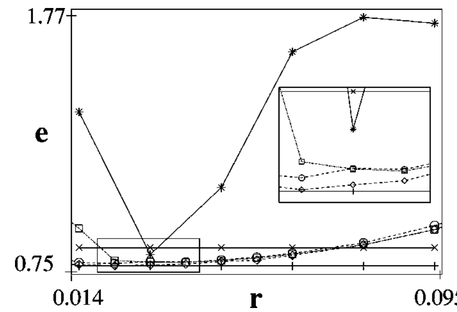

FIG. 4. Average percentage error e vs neighborhood radius r for the Hénon map, with discretization $m=6$. Units are dimensionless. Symbols in the figure refer to the following series: $\times$ for S with added noise of width 0.8 % and $+$ for D, both before cleaning. $\square$, $\bigcirc$, and $\diamond$ for S with noise widths of 1.5%, 0.8%, and 0.4%, respectively: D after noise reduction.

cause it is the clearest. This figure shows the average error e of the D and S series with noise of width $\delta_l=\delta_m$ (lower and higher horizontal lines, respectively), the cleaned D series, and the cleaned S-data generated with several noise widths ($\delta_m/2$, δ_m, and $2\delta_m$.) We interpret this figure as follows. (1) The noise level of the S series is higher than that of the D series, although one might expect that adding a signal with expected value of zero should not change e. In fact, $e_S \sim \frac{13}{12}e_D$, in agreement with a simple calculation [14]. (2) The cleaned D series shows a higher error level than the D series, although one might expect otherwise from visual inspection of the noise-reduced data. Our results, however, are consistent with previous findings [5] that the portion of a signal lost under discretization is low dimensional, and essentially not recoverable. A cleaning algorithm should not be able to improve the quality of the signal, and we find that indeed it does not. (3) Over a large range of neighborhood radii, $2\delta \leqslant r \leqslant 7\delta$, we observe that the cleaning algorithm lowers the error of the S series, without ever reaching the lower limit of the D series. This suggests that the smoothing effect of the additional noise improves the averaging effects of the cleaning algorithm. Again, we stress that the results are similar for all values of m between 5 and 8.

In addition, we have explored the effects of noise smoothing and noise reduction on the calculation of dynamical invariants. Adding noise to the discretized series allows the exploration of smaller scales in the $\ln N(\epsilon)$ vs $\ln(1/\epsilon)$ estimation of the capacity dimension. However, due to the randomness of the added (uniform or Gaussian) noise, no additional information about the small-scale structure of the attractor is expected. This argument applies to the generalized dimensions D_q.

In Table I we show the estimation of the largest Lyapunov exponent using the discretized Hénon series (column 4) and the same data smoothed with uniform noise of the same width (column 5), for two levels of discretization (column 2) and several radii of the noise-reduction algorithm (column 3.) The results of the last two columns should be compared with those of the original map (column 1.) Just adding noise brings the Lyapunov exponent from a factor of about 30 off to just a factor of 5 off (rows with a dash in column 3.) This improvement can be understood by looking at part (a) of

TABLE I. Original Lyapunov exponent, discretization level, radius of noise-reduction algorithm (''-'' indicates no reduction, and integer indicates multiple of discretization level), and estimation of Lyapunov exponent after noise reduction for discretized series (D) and series smoothed with uniform noise comparable to the discretization step (S).

λ_1	m	r	λ_{D1}	λ_{S1}
0.42	6		16.0603	1.9703
		1	15.9955	2.4302
		2	5.7284	2.3613
		3	4.0996	2.3236
		4	2.8981	2.3116
		5	2.8575	2.2991
0.42	7		12.1003	2.0431
		1	11.5739	2.1392
		2	3.1412	2.1830
		3	2.6110	2.1680
		4	2.4811	2.1947
		5	2.3091	2.2082

Figs. 2 and 3: in the first, two nearby points that appear identical under discretization can be mapped to points one or several discretization steps away, resulting in serious overestimation of Lyapunov exponents. It seems that just adding noise reduces this effect, as the second figure shows. The improvement in the Lyapunov exponent obtained from noise reduction of the discretized series seems to tend asymptotically to that obtained just by noise smoothing, strongly suggesting that the improvements in column 4 are just due to smoothing of the data, and not to actual recovery of information about the original signal. These results are consistent with [5], in the sense that the original value of the Lyapunov exponent is never recovered.

Before moving on to a practical example, we discuss why noise width δ_l comparable to the discretization level δ_m seems to work best. We have found that using $\delta_l \gg \delta_m$ alters the apparent dynamics of the system, resulting in spurious crossings of the stable and unstable manifolds of the dynamical system. We see this disturbing effect, in which the noise in effect drowns the signal, even with $\delta_l = 4\delta_m$. On the other hand, if $\delta_m \gg \delta_l$, an effect which is just noticeable in Fig. 2 becomes far more serious: the smoothed series remains confined to a grid defined by the discretization step itself, a spurious structure that for small δ_l survives noise reduction. More details are given in [13]. Because of these results, we recommend smoothing with $\delta_l \sim \delta_m$.

IV. A PRATICAL EXAMPLE

We have extended our study to the more realistic case of the time series of the power output of an NH_3 laser [4,7,8]. We used 9896 points from this series, which was measured with eight-bit precision, or discretization $m=8$. We have added noise with widths between $m=6$ and $m=10$ to generate the corresponding S series. In this case, the original series is not available, and we only have qualitative tools to compare the results. Figure 5 shows reconstructed trajectories for D and S series before and after noise reduction with $r=2\delta$. We will focus the description in the denser region of the reconstructions, corresponding to values in the interval [20,80]. The protocol used is the same that we have used

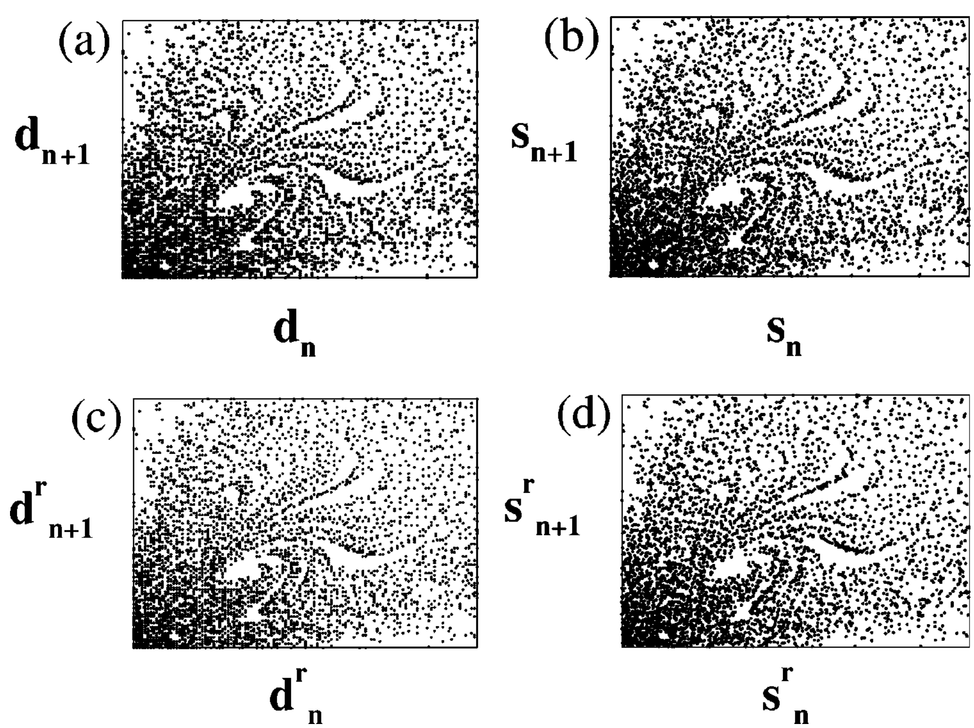

FIG. 5. Laser data in time-delayed coordinates, $20 \leqslant l_n, l_{n+1} \leqslant 140$. (a) Zoom of the original, eight-bit data. (b) Zoom of smoothed data with uniform noise of the same width. (c) and (d), Same data as in parts (a) and (b), after noise reduction using $r=2\delta$ and four iterations.

before, with $\tau=1$ timestep and $d_E=2$.

The lower left corner of Fig. 5(a) shows a lower density of points than the equivalent region of Fig. 5(b), because the uniform noise in *S* spreads the points that coincide in a single value in *D*; with this, the discretized structure disappears in *S*. Figures 5(c) and (d), show that the discretized structure in the *D* series survives the noise reduction, but greatly diminished. Moreover, regions with high density of points can be seen throughout the phase space of the cleaned *S* series that do not appear in the phase space of the cleaned *D* series. While we have chosen parameter values that yield good results, in all fairness we must say that the best noise reduction we have seen [4] of this particular data set gets rid of the grid structure in *D*, but does not show as many high-density regions of points as we do.

V. CONCLUSIONS

To summarize, in this paper we corroborate the conjecture stated in Ref. [5], that discretization of a time series causes an unrecoverable information loss. Indeed, we have not seen a single instance of noise reduction of a discretized series that has lower average error than the discretized signal itself, despite an apparent sharpening of the reconstructed maps. The apparent improvement on Lyapunov exponent estimation shown in Table I is no better than the addition of a random signal. We expect these results to hold for other noise-reduction algorithms as well.

We also find that adding uniform noise to a discretized series previous to the application of a standard noise-reduction algorithm minimizes the increase in average error; it appears that the noise spreads the points concentrated at discrete intervals, and helps the averaging action of the cleaning algorithm. We show an example of this in Fig. 4. This spreading also helps improve the estimation of Lyapunov exponents, but not of generalized dimensions. After reporting the effects of adding noise levels much smaller or much greater than the discretization step, we recommend adding uniform noise of width comparable to the discretization step itself for good results. A final point is that this work provides an example of how noise makes it difficult to infer the model equation in a chaotic system [15], since in this case part of the original deterministic signal, and hence information about the governing equation, is lost during the measurement process.

ACKNOWLEDGMENTS

We thank H. Kantz for providing the TISEAN package. We acknowledge support from Colciencias (Grant RC 425-00).

[1] H.D.I. Abarbanel, R. Brown, J.J. Sidorowich, and L.Sh. Tsimring, Rev. Mod. Phys. **65**, 1331 (1993).

[2] *Coping with Chaos*, edited by E. Ott, T. Sauer, and J. A. Yorke (Wiley, New York, 1994).

[3] H.D.I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer, New York, 1996).

[4] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, UK, 1999).

[5] P.-M. Binder and M.C. Cuéllar, Phys. Rev. E **61**, 3685 (2000).

[6] M. Hénon, Commun. Math. Phys. **50**, 69 (1976).

[7] *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld (Addison-Wesley, Reading, MA, 1993).

[8] U. Hübner, N.B. Abraham, and C.O. Weiss, Phys. Rev. A **40**, 6354 (1989).

[9] S. Ellner and P. Turchin, Am. Nat. **145**, 343 (1995); R.F. Costantino, R.A. Desharnais, and B. Dennis, Science **275**, 389 (1997); C. Zimmer, *ibid.* **284**, 83 (1999).

[10] R. Savit, R. Manuca, and R. Riolo, Phys. Rev. Lett. **82**, 2203 (1999); W.B. Arthur, Science **284**, 107 (1999); N.F. Johnson, P.M. Hui, and T.S. Lo, Phys. Rev. Lett. **82**, 3360 (1999).

[11] T. Schreiber, Phys. Rev. E **47**, 2401 (1993).

[12] R. Hegger, H. Kantz, and T. Schreiber, Chaos **9**, 413 (1999).

[13] P.-M. Binder and M. C. Cuéllar (unpublished); M. C. Cuéllar, M.Sc. thesis, Universidad de Los Andes, Bogota, 2001.

[14] Under the rounding up scheme we have used, the average amount by which a data point is moved up is $\delta/2$, which is also the average error *e* of the discretized series. If one further adds uniformly distributed noise of width δ to the discretized series, one can estimate the average distance between points in the smoothed data, uniformly distributed in $(\delta/2, 3\delta/2)$ and points in the original data, uniformly distributed in $(0,\delta)$. This average is $13\delta/24$, and hence the result cited in the text. This footnote shows that when the discretization and external noise are added there is a slight ratcheting effect.

[15] L. Smith, in *Nonlinear Dynamics and Statistics*, edited by A. I. Mees (Birkhauser, Boston, 2001), pp. 31-64.