Adaptive recognition and filtering of noise using wavelets

S. Boccaletti, A. Giaquinta,* and F. T. Arecchi†
Istituto Nazionale di Ottica, Largo E. Fermi, 6, I50125 Florence, Italy
(Received 7 October 1996; revised manuscript received 20 December 1996)

We combine wavelet transform and adaptive recognition techniques to introduce a filtering process able to analyze, categorize, and remove additive noise from experimental time series, without previous information either on the correlation properties of noise or on the dimension of the deterministic signal. The method is applied to a high dimensional delayed chaotic time series affected by additive white and colored noises. The obtained results show that the reconstruction of the signal both in real and in Fourier space is effective through the discrimination of noise from the deterministic part. [S1063-651X(97)15005-X]

PACS number(s): 05.45.+b, 05.40.+j

The wavelet transform (WT) is a mathematical microscope [1], whereby a time signal can be decomposed in a representation into both time and frequency. With respect to the Fourier transform (FT), WT has the advantage of localizing the appearance of transient events.

Its application to experimental time series has become widespread [2] in seismic signal analysis [3], image processing [4], music [5], magnetic resonance imaging [6], image compression [7], optics [8], turbulence [9], neurophysiology [10], speech discrimination [11], fractal analysis [12], DNA sequence analysis [13], galaxies [14], and asteroids [15] discrimination in observational cosmology.

Here the one-dimensional DAUB20 [16] version of WT is combined with an adaptive strategy which recognizes chaotic features in experimental time series [17], to provide an efficient filtering process, whereby one can detect and remove additive noise, with no previous knowledge either of the noise correlation properties, or of the dimension of the noiseless data set. The method recognizes and removes frequency by frequency the amount of energy coming from noise, thus reconstructing the deterministic part of the signal.

The adaptive algorithm performs an observation task with a sequence of nonfixed observation time intervals (OTI's) which minimize the second variations of the signal. This way, whenever a fixed OTI yields a geometrically irregular data series (as in the case of a chaotic signal), the adaptive algorithm provides a geometrically regular one at irregular OTI's. Given an experimental chaotic signal $x(t)$, for each embedding component i $(i=1,\ldots,m$; m being the selected embedding dimension [18]) and at the time $t_{n+1}=t_n+\tau_n$ (τ_n being the actual OTI to be later updated), one defines the variations

$$\delta x_i(t_{n+1}) \equiv x_i(t_{n+1}) - x_i(t_n), \tag{1}$$

and the variation rates

$$\lambda_i(t_{n+1}) \equiv \frac{1}{\tau_n} \log \left| \frac{\delta x_i(t_{n+1})}{\delta x_i(t_n)} \right|. \tag{2}$$

Then, the minimum of all

$$\tau_{n+1}^{(i)} = \tau_n^{(i)} [1 - \tanh(g \lambda_i(t_{n+1}))] \tag{3}$$

$(i=i,\ldots,m)$ is chosen as τ_{n+1} and a new observation is performed at $t_{n+2}=t_{n+1}+\tau_{n+1}$. The procedure described above is repeated for any embedding dimension m up to a maximum embedding dimension M.

Equation (3) arises from the following considerations. In order to obtain a sequence of geometrically regular δx_i, we contract (expand) the observation time interval when the actual value of δx_i is bigger (smaller) than the previously observed one. The hyperbolic tangent function maps the whole range of $g\lambda_i$ into the interval $(-1,+1)$. The constant g, strictly positive, is chosen in such a way as to forbid $\tau_n^{(i)}$ from going to zero. It may be taken as an *a priori* sensitivity. A more sensible assignment would consist in looking at the unbiased dynamical evolution for a while and then taking a g value smaller than the reciprocal of the maximal λ recorded in that time span. Choosing a fixed g is like adjusting the connectivities of a neural network by a preliminary learning session, while adjusting g upon the information accumulated over a given number of previous time steps corresponds to considering g as a kind of long-term memory, as opposed to the short-term memory represented by the sequence of τ_n [19].

We update g every L OTI, through the rule

$$g = g_0 + m \Big/ \left(M \sum_{i=1}^{m} \sum_{k=1}^{L} |\lambda_i(t_k)| \right). \tag{4}$$

Here the sum runs over all the actual embedding dimensions $(i=1,\ldots,m)$ and over all the previous L OTI's $(k=1,\ldots,L)$. $g_0>0$ is a safety term providing a minimum sensitivity in case $1/[\Sigma_{i=1,\ldots,m}\Sigma_{k=1,\ldots,L}|\lambda_i(t_k)|]$ gets close to zero [17]. The factor m/M has been introduced to homogenize the calculation of g at any embedding dimension. Notice that, when the maximum embedding dimension M is reached, Eq. (4) reduces to the definition introduced for the case of a model problem with preassigned dimensions [17].

*Also at Dept. of Sistemi e Informatica, University of Florence, Italy.

†Also at Dept. of Physics, University of Florence, Italy.

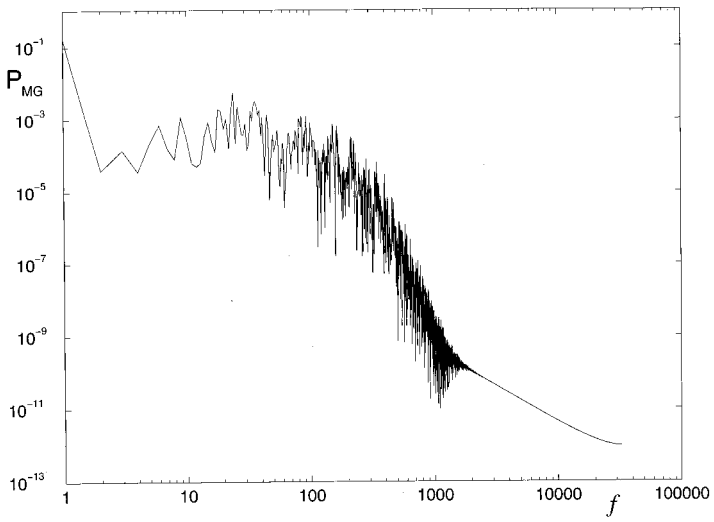

FIG. 1. Power spectrum of the solution of the Mackey-Glass (MG) delayed equation (see text). The x axis is labeled with the dimensionless frequency channel number f, the y axis with the values of the power spectrum P_{MG} (in arbitrary units).

We thus obtain a sequence of observation times (starting from t_0 and $\tilde{\tau}\equiv\tau_0$) given by

$$t_0, t_1 = t_0 + \tilde{\tau}, \quad t_2 = t_1 + \tau_1, \ldots, \quad t_{n+1} = t_n + \tau_n, \ldots$$

corresponding to which the variations of $\delta x_i(t_n)$ can be reduced below a preassigned value. The observations performed at those times provide a ''regularized'' geometry, however the above time sequence is irregular and now includes the chaotic information that was in the original geometric sequence $\mathbf{x}(t)$. From the obtained irregular OTI sequence one can extract the correlation properties of the signal [17] as the periods of the unstable periodic orbits embedded in the chaotic attractor [20]. It is important to notice that in Eqs. (1) and (2), differences and variation rates are evaluated over the adaptive τ, which has to be much larger than the experimental sampling time τ_s [17], but much smaller than the characteristic period τ_{UPO} of the unstable periodic orbits embedded within the chaotic attractor [20]. As a consequence, our method requires τ_s to be about two orders of magnitude smaller than τ_{UPO}. This way, the method introduces a natural adaptation time scale intermediate between the minimum resolution time and the time scale of the periodic orbits.

For an empirical series of data, giving rise to a sequence of N OTI's, a discrimination between determinism and noise is provided by the following indicator:

$$\beta(m) = \frac{1}{N}\sum_{n=1}^{N}\left|\prod_{i=1}^{N}\lambda_i(t_n)\right|. \tag{5}$$

Its heuristic meaning emerges from the following considerations. Expanding λ_i to first order and referring to the unit time step $\tau_n = 1$, we can write $\lambda_i(t_{n+1}) = [\delta x_i(t_{n+1}) - \delta x_i(t_n)]/\delta x_i(t_n)$. We now evaluate the variation over the unit time of the volume $V_n = \Pi_{i=1}^{m}\delta x_i(t_n)$ made by all m measured variations at time t_n. The relative variation rate $r_n = (V_{n+1} - V_n)/V_n$ is given by

$$r_n = \sum_i \lambda_i + \sum_{i\neq j}\lambda_i\lambda_j + \cdots + \prod_i \lambda_i, \tag{6}$$

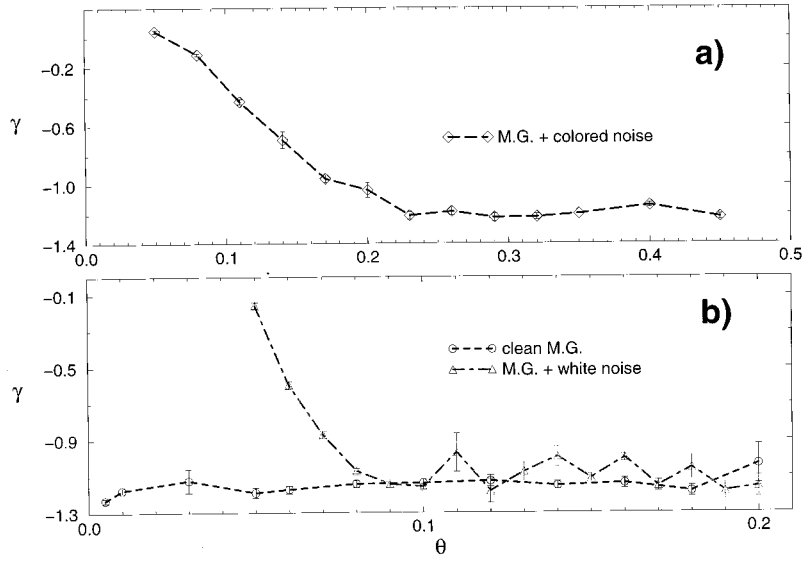

FIG. 2. γ vs θ plots (see text for definitions). (a) MG plus colored noise. (b) Pure MG (circles and dashed line) and MG plus white noise (triangles and dot-dashed line). The procedure for the construction of white and colored noises is described in the text. $L = 50$, $g_0 = 5\times10^{-6}$, $M = 15$. It is evident that the plots of the noisy data define an optimal filtering threshold $\bar{\theta} = 0.23$ for colored noise and $\bar{\theta} = 0.1$ for white noise. γ and θ are dimensionless quantities.

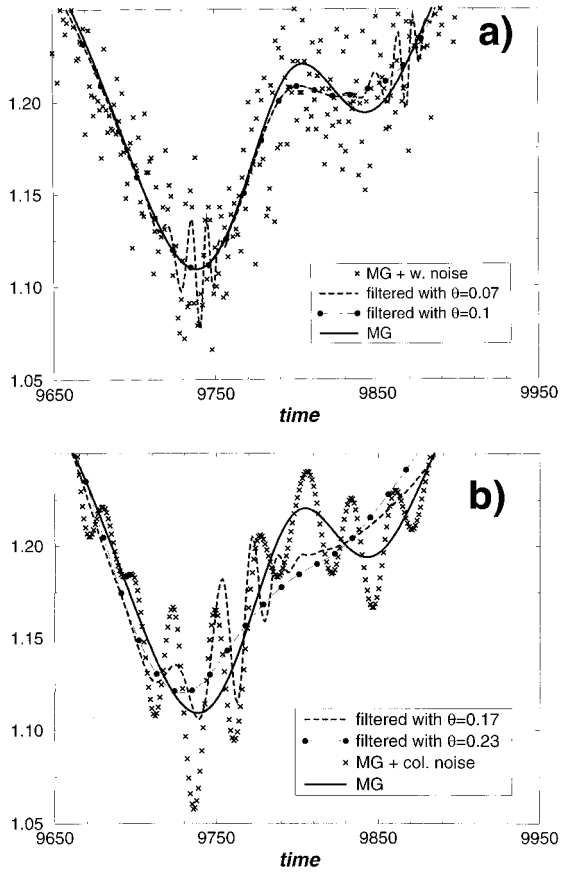

FIG. 3. Time domain reconstruction of the signal for (a) MG plus white noise and (b) MG plus colored noise. Symbols and lines are in the legends. Same parameters and noises as in Fig. 2. The maximum accuracy in the reconstruction is obtained in both cases for $\theta=\bar{\theta}$. Time is in sampling units, the y axis is labeled in arbitrary units.

where the indices i and j runs from 1 to m. Summing up over all directions of phase space, we introduce the directional averages

$$\langle\lambda\rangle=\frac{1}{m}\sum_i \lambda_i, \quad \langle\lambda^2\rangle=\frac{2}{m(m-1)}\sum_{i\neq j}\lambda_i\lambda_j, \text{ etc.} \quad (7)$$

As we further sum over all n OTI's up to N, the twisting along the chaotic trajectory makes all directions statistically equivalent, thus in $\Sigma_n r_n$ we can replace $\langle\lambda^s\rangle$ by $\langle\lambda\rangle^s$ for $2\leqslant s\leqslant m$. In the case of stochastic noise, since variations over successive time steps are uncorrelated, $\delta x(t_{n+1})-\delta x(t_n)\simeq\delta x(t_n)$ and $\langle\lambda\rangle$ is close to 1, so that $\langle\lambda\rangle^m=O(1)$. Instead, for a deterministic dynamics two successive steps are strongly correlated, hence $\langle\lambda\rangle<1$, and the last term of Eq. (6), that is, $\langle\lambda^m\rangle\sim\langle\lambda\rangle^m=\exp(-m\log(1/\langle\lambda\rangle))$, yields the most sensitive test.

Based on these considerations, we take the sum over the N trajectory points of the last term of Eq. (6) as the β indicator. In view of what was said above, β scales as e^{-m} for a deterministic signal [beside a factor $O(1)$ in the exponent] whereas it scales as e^0 for noise.

Given an experimental signal $x(t)$ which is the sum of a deterministic and of a noisy part, we act on the coefficients of its WT (obtained with the DAUB20 basis [16]) by eliminating those that are smaller than a given threshold θ. The new signal $x'(t)$, generated by inverse WT of the filtered set of coefficients, undergoes the β text. Increasing θ decreases the slope γ of the $\log\beta(m)$ plot, down to a saturation value of $\bar{\gamma}$ for the corresponding pure deterministic dynamics. Thus, the γ vs θ plot results in a monotonically decreasing function up to $\theta=\bar{\theta}$ where saturation is reached. This way, the minimal threshold $\bar{\theta}$, which extracts from the data as much determinism as possible, is adaptively chosen by the algorithm which drives the filter in wavelet space.

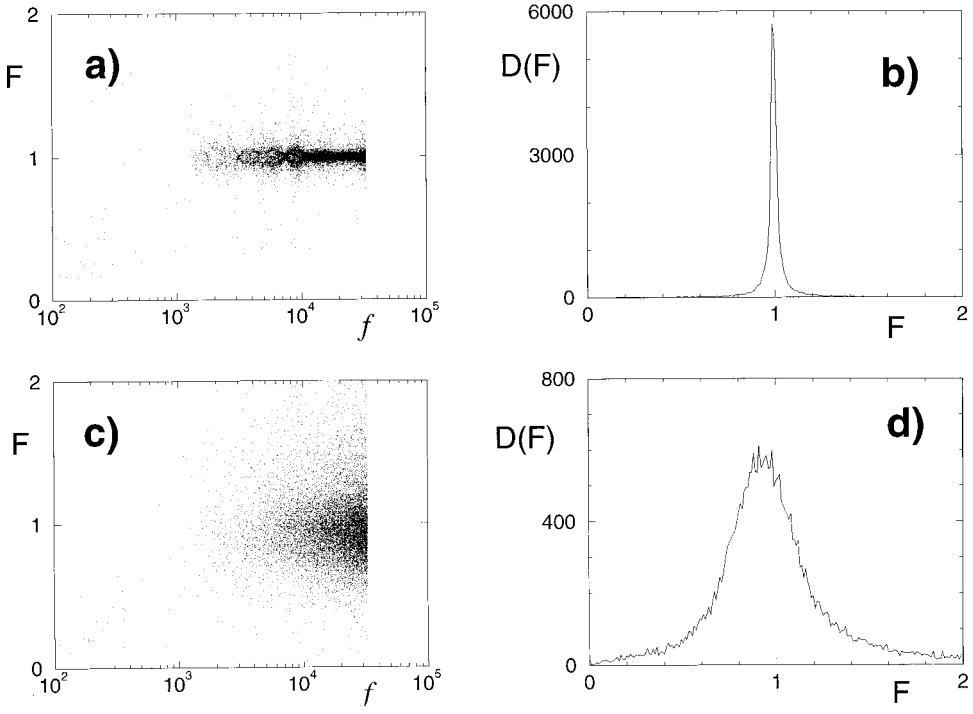

FIG. 4. Accuracy in the Fourier space reconstruction for MG plus white noise. (a) and (c) report the dimensionless quantity F (ratio of the amount P_{filtered} of filtered energy to the noise power spectrum P_{noise}) vs the frequency channel number f for $\theta=\bar{\theta}=0.1$ and $\theta\neq\bar{\theta}=0.07$, respectively. (b) and (d) show the corresponding histograms $D(F)$ vs F. Same parameters as in Fig. 2. The best reconstruction is again confirmed at $\theta=\bar{\theta}$.

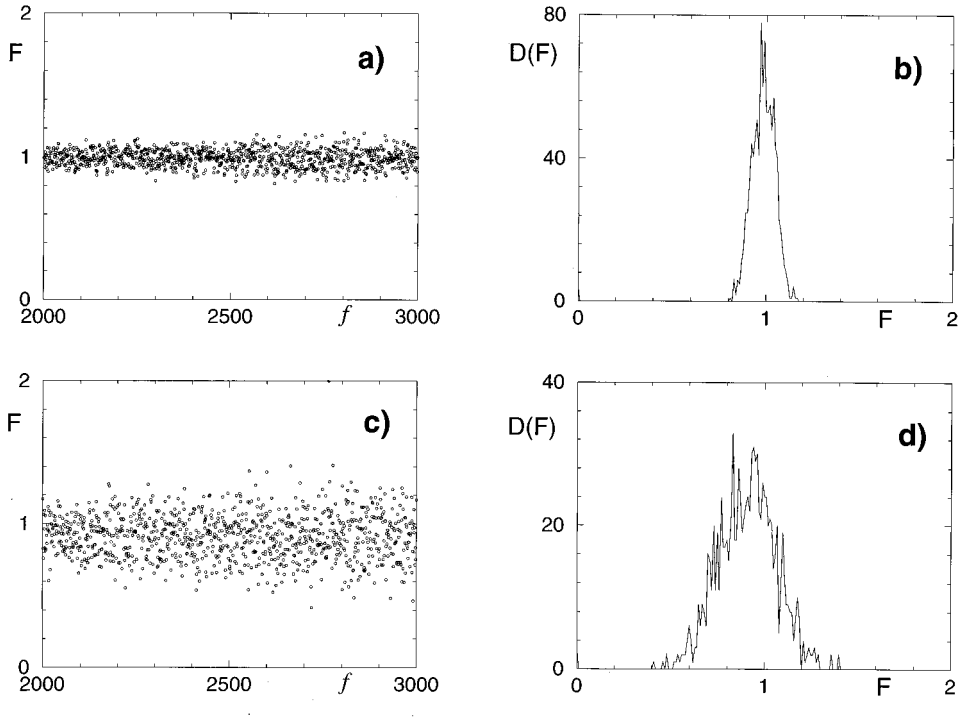

FIG. 5. Accuracy in the Fourier space reconstruction for MG plus colored noise (all plots limited to the frequency range covered by noise). (a) and (c) report F vs f for $\theta=\theta=0.23$ and $\theta\neq\theta=0.17$, respectively. (b) and (d) show the corresponding $D(F)$ vs F histograms. Same parameters as in Fig. 2. The best reconstruction is again confirmed at $\theta=\theta$.

A preliminary hint on the possible cleaning of a signal from noise by coupling WT with our β indicator was given in Ref. [21]. However, in that proposal g was taken as a fixed parameter, and the problem of the optimal threshold was not even tackled.

To show the reliability of this method, we apply it to a set of $l=2^{16}$ data extracted from the solution of the delayed Mackey-Glass (MG) equation [22] (dot denotes temporal derivative),

$$\dot{x}(t)=-0.1x(t)+\frac{0.2x(t-T)}{1+x(t-T)^{10}}.$$

For $T=100$, MG produces a 7.5-dimensional dynamics [23]. We affect each of the $l/2$ Fourier components of the above time data with different classes of zero average noise. Precisely, we generate $l/2$ random Fourier components (amplitudes homogeneously distributed between 0 and 1, phases homogeneously distributed between 0 and 2π), then the inverse FT is normalized in such a way as to produce a noise to signal ratio of 0.02 with respect to MG. In a standard way, the noise to signal ratio is the ratio between the rms noise variance and the rms of the average of the square signal. This way, we have produced white noise [24]. Limiting the same procedure to a narrower band of Fourier components from the frequency channel f_1 to the frequency channel f_2, we obtain a colored noise. The role of any narrower band noise with the same spectral density as for the white noise will of course be less crucial, as will be discussed later. However, to work in the most disadvantageous case, we consider a colored noise with the total spectral power as for the white noise (which of course is band limited by the measuring procedure between $f_{\min}=1/2T$ and $f_{\max}=1/2\tau_s$, where T is the total measuring interval and τ_s is the sampling time).

Our choice of f_1 and f_2 arises from the following considerations. On one side, a full overlap between the spectra of noise and signal would imply a very difficult filtering task; on the other side, some overlap is needed, otherwise a simple band-cut filter would do the job. In our case, choosing $f_1=2000$ and $f_2=3000$ means working with a nontrivial overlap with the power spectrum of MG, as it results from Fig. 1. In these conditions, any bandpass filter would cause quite big distortions in the reconstruction, insofar as some of the suppressed frequencies would contain relevant information on the deterministic dynamics.

Both white and colored noise are separately added to MG and the resulting signals have been analyzed with our filter. Figure 2 shows the selection mechanism for the filtering threshold. In both cases, a clear plateau ($\gamma\sim-1.2$) is reached in the corresponding γ vs θ plots, and this plateau corresponds to the γ value for pure MG. The plateau clearly defines the optimal filtering threshold θ. Figure 3 gives the reconstructions in the time domain operated by the filter at $\theta=\theta$ and $\theta\neq\theta$ for the two cases. The difference in the reconstructions at $\theta=\theta$ and $\theta\neq\theta$ highlights the effectiveness of the method.

In order to quantify the robustness of the method in analyzing and removing the noise from the data, it is worthwhile to refer to the reconstruction of the signal in the frequency domain (Fourier space). A stringent test consists in comparing the amount of energy that the filter eliminates from a given Fourier frequency channel f with the same component of the noise spectrum. We consider the ratio of the amount P_{filtered} of filtered energy to the noise power spectrum P_{noise}. $F(f)=P_{\text{filtered}}(f)/P_{\text{noise}}(f)$ thus measures the accuracy of the filter in recognizing and eliminating the noise component in the channel f. $F(f)=1$ for any channel f for which the noise has been successfully disentangled from the signal. Figures 4 and 5 report the F vs f plots for MG plus white noise and MG plus coloured noise respectively, calculated at $\theta=\theta$ and at $\theta\neq\theta$ together with the histograms of the

distribution around unity. Notice the worsening in F as we go from Fig. 4 to 5. This is due to having concentrated the total noise power within a narrow frequency range, thus Fig. 5 reports the disentanglement for the case of noise to signal ratio which is of order of 1 (the ratio between the bandwidth of white and colored noises is 32.5/2, thus the spectral density of the colored noise is about 16 times higher than 2%). This is the reason why the threshold required by the algorithm has increased from $\theta=0.1$ to $\theta=0.23$. On the contrary, if we work with unperturbed noise spectral power, the role of any narrow frequency window can be extracted directly from the corresponding section of Fig. 4(a). It is evident from Figs. 4 and 5 that in both cases the best recognition and filtering is obtained for $\theta=\overline{\theta}$.

Besides the simple bandpass filters discussed above, other nonlinear noise-reduction methods have been proposed for chaotic time series [25]. Most of them are based on a preliminary reconstruction of the attractor in the embedding space and a successive filtering process in this space by various techniques. While our method provides a signal reconstruction as faithful as that provided by the methods of Ref. [25], it requires a much smaller computational effort. Indeed, if a number σ of β applications are needed to extract the optimal filtering threshold $\overline{\theta}$, our strategy requires a number of operations equal to $(\sigma+1)\log(N)$ (one forward WT of the original signal and σ backward WT of the filtered set of coefficients) plus σN (σ successive applications of β to each reconstructed signal). For large N, the whole computational effort is linear in N, whereas any technique on the embedding space which exploits the local dynamical properties of the attractor would imply a number of operations scaling at least as $N\ \log(N)$ [25].

The above results show that the method here introduced is robust and effective in detecting, recognizing, and filtering additive noise also in case of rather high dimensional systems; it also succeeds in recognizing and removing frequency by frequency the exact amount of energy introduced by noise. It is important to notice that the described filtering procedure does not require information on the correlation properties of the additive noise or on the properties of the noiseless data, thus the method appears easily implementable in experimental situations.

The authors acknowledge G. Basti and A.L. Perrone for their contribution in the introduction of the adaptive filtering algorithm.

[1] M. Holschneider, J. Stat. Phys. **50**, 963 (1988).
[2] I. Daubechies, IEEE Trans. Inf. Theor. **36**, 961 (1990).
[3] P. Goupillaud, A. Grossmann, and J. Morlet, Geoexploration **23**, 85 (1984).
[4] S. G. Mallat, IEEE Trans. Acoust. Speech. Sig. Proc. **37**, 2091 (1989).
[5] R. Kronland-Martinet, J. Morlet, and A. Grossmann, Int. J. Pat. Recog. Artif. Intell. **1**, 97 (1987).
[6] J. B. Weaver, Y. Xu, D. M. Healy, and J. R. Driscoll, Magn. Reson. Med. **24**, 275 (1992).
[7] R. A. De Vore, B. Jawerth, and B. J. Lucier, IEEE Trans. Inf. Theor. **38**, 719 (1992).
[8] B. Pouligny, G. Gabriel, J. F. Muzy, A. Arneodo, and F. Argoul, J. Appl. Cryst. **24**, 526 (1991).
[9] F. Argoul, A. Arneodo, G. Grasseau, Y. Gagne, E. J. Hopfinger, and U. Frisch, Nature **338**, 51 (1989).
[10] A. W. Przybyszewski, J. Neurosci. Meth. **38**, 247 (1991).
[11] S. Kadambe and G. F. Boudreaux-Bartels, IEEE Trans. Inf. Theor. **38**, 917 (1992).
[12] A. Arneodo, G. Grasseau, and M. Holdschneider, Phys. Rev. Lett. **61**, 2281 (1988).
[13] A. Arneodo, E. Bacry, P. V. Graves, and J. F. Muzy, Phys. Rev. Lett. **74**, 3293 (1995).
[14] E. Slezak, A. Bijaoui, and G. Mars, Astron. Astrophys. **227**, 301 (1990).
[15] Ph. Bendjoya, E. Slezak, and CI. Froeschle, Astron. Astrophys. **251**, 312 (1991).
[16] I. Daubechies, *Wavelets* (SIAM, Philadelphia, 1992).
[17] F. T. Arecchi, G. Basti, S. Boccaletti, and A. L. Perrone, Europhys. Lett. **26**, 327 (1994).
[18] F. Takens, *Detecting Strange Attractors in Turbulence*, Lecture Notes in Math. Vol. 898 (Springer-Verlag, Berlin, 1981).
[19] S. Grossberg, Neural Networks **1**, 17 (1988).
[20] S. Boccaletti and F. T. Arecchi, Europhys. Lett. **31**, 127 (1995).
[21] A. L. Perrone, S. Boccaletti, G. Basti, and F. T. Arecchi, Proc. SPIE **2242**, 130 (1994).
[22] M. C. Mackey and L. Glass, Science **197**, 287 (1977).
[23] J. D. Farmer, Physica D **4**, 366 (1982); P. Grassberger and I. Procaccia, *ibid*. **9**, 189 (1983).
[24] S. O. Rice, Bell Syst. Tech. J. **24**, 46 (1945).
[25] A detailed review of common methods for noise reduction in chaotic time series is available in E. J. Kostelich and T. Schreiber, Phys. Rev. E **48**, 1752 (1993), and references therein.