Wavelet filtering of chaotic data

M. Grzesiak

Nonlinear Processes in Geophysics (2000) 7: 111–116

Nonlinear Processes in Geophysics

Wavelet filtering of chaotic data

M. Grzesiak

Space Research Center of Polish Academy of Sciences, Warsaw, Poland

Received: 8 December 1997 – Accepted: 23 December 1999

Abstract. Satisfactory method of removing noise from experimental chaotic data is still an open problem. Normally it is necessary to assume certain properties of the noise and dynamics, which one wants to extract from time series. The wavelet based method of denoising of time series originating from low-dimensional dynamical systems and polluted by the Gaussian white noise is considered. Its efficiency is investigated by comparing the correlation dimension of clean and noisy data generated for some well-known dynamical systems. The wavelet method is contrasted with the singular value decomposition (SVD) and finite impulse response (FIR) filter methods.

1 Introduction

Owing to a continuous character of the power spectra filtering of chaotic data requires specific methods. The traditional, Fourier domain methods such as band pass filtering fail, thus several other methods, especially dedicated to analysis of chaotic time series, were proposed (Grassbeger et al., 1993; Kostelich and Yorke, 1988). In this paper a new technique of filtering, based on the wavelet transform, is presented and compared with the singular value decomposition (SVD) and FIR filter methods.

The basic idea is to distinguish noise from data by means of their local regularity, thus it is assumed that the regularity of the noise is different from that of data. One can show that the local regularity of a function can be effectively studied with the continuous wavelet transform and the notion of the Hölder exponent. These concepts are briefly introduced in Sect. 2 and a scaling behavior of wavelet coefficients of the Gaussian colored noise is determined. For more information on this subject see Daubechies (1992); Mallat and Hwang (1992); Muzy et al. (1994).

In Sect. 3 the wavelet filtering method is applied to time series associated with several dynamical systems of known properties. Results are compared with the widely used SVD filtering rooted in the idea of Karhunen-Loève representation for the multichannel data, and the finite impulse response (FIR) filter. The Karhunen-Loève basis (Albano et al., 1988; Broomhead and King, 1986) is the basis of eigenvectors of autocovariance matrix $A = E[x \bigotimes x]$, where x denotes the vector in the reconstructed phase-space (Takens, 1981). Since the autocovariance matrix is symmetric, SVD representation acts like rotation in phase space.

2 Background

The continuous wavelet transform of a function f is defined by:

$$W_\psi[f](a,b) = \tilde{f}(a,b) =< f \mid \psi_{a,b} >, \tag{1}$$

where $< f \mid \psi_{a,b} >= \int_{-\infty}^{+\infty} dx f(x)\overline{g(x)}$ is the standard inner product in $L^2(R)$. The family of functions $\psi_{a,b}$ is constructed by the recipe: $\psi_{a,b}(x) = a^{-1/r}\psi\left(\frac{x-b}{a}\right)$, where a and b are dilation and translation parameters, respectively. The factor $a^{-1/r}$ ensures that a norm of $\psi_{a,b}$ in $L^2(R)$ is constant. From now on we choose $r = 1$. A function ψ is called "mother wavelet".

To be able to reconstruct the function from its wavelet representation ψ must satisfy so called admissibility condition:

$$\int \frac{d\omega}{|\omega|}\left|\hat{\psi}(\omega)\right|^2 < \infty, \tag{2}$$

where $\hat{\psi}$ is the Fourier transform of ψ (Daubechies, 1992).

The wavelet transform is said to have n_ψ vanishing moments if $\int dx\psi(x)x^k = 0$ for $k = 0, 1, ..., n_\psi - 1$.

By definition, the function f has the Hölder exponent h at point b if there exists a constant C and a polynomial $P_n(x)$ of the order n such that:

$$|f(x) - P_n(x-b)| \leq C|x-b|^h. \tag{3}$$

Thus the behavior of $f(x)$ around b may be expressed in terms of the Taylor expansion:

$$f(x) = f(b) + \sum_{k-1}^{n-1} \frac{1}{k!}(x-b)^k f^{(k)}(b) + C|x-b|^{h(b)}. \quad (4)$$

Applying the wavelet transform to this expansion and assuming that the mother wavelet function has at least $n_\psi > h(b)$ vanishing moments we obtain:

$$\begin{aligned} W_\psi[f](a,b) &= Ca^{-1} \int dx |x-b|^{h(b)} \overline{\psi\left(\frac{x-b}{a}\right)} \\ &= Ca^{h(b)} \int d\zeta |\zeta|^{h(b)} \overline{\psi(\zeta)} \propto a^{h(b)}. \end{aligned} \quad (5)$$

The last two formulas demonstrate the role of wavelets in studying the local regularity of a function (Muzy et al., 1994). In order to retrieve correctly the Hölder exponent the mother wavelet must have $n_\psi > h$ vanishing moments.

One of the important characteristics of the wavelet transform of colored noise is the scaling behavior of the coefficients. Power spectra of colored noise are of the form:

$$P(\omega) \sim |\omega|^{-p}, \quad (6)$$

where p is the spectral index. The scaling relation for the wavelet coefficients computed in L^1 norm may be derived as follows:

$$\begin{aligned} & E\left[\left|\tilde{f}(a,b)\right|^2\right] \\ = \; & E\left[\int dt_1 f(t_1)\overline{\psi_{a,b}(t_1)} \overline{\int dt_2 f(t_2)\overline{\psi_{a,b}(t_2)}}\right] \\ = \; & \int\int dt_1 dt_2 E[f(t_1)f(t_2)]\overline{\psi_{a,b}(t_1)}\psi_{a,b}(t_2) \\ = \; & a^{-2}\int\int dt d\tau g(\tau) \overline{\psi\left(\frac{t-b}{a}\right)}\psi\left(\frac{t-(b-\tau)}{a}\right) \\ = \; & a^{-2}\int d\tau g(\tau)\int d\omega a^2 |\hat{\psi}(a\omega)|^2 e^{i\omega\tau} \\ = \; & \int d\omega a^2 |\hat{\psi}(a\omega)|^2 P(\omega) \\ = \; & \int \frac{d\zeta}{a}|\hat{\psi}(\zeta)|^2 \left|\frac{\zeta}{a}\right|^{-p} \propto a^{1-p}. \end{aligned} \quad (7)$$

Here $g(\tau)$ is the autocorrelation function, provided that the process is stationary.

These results suggest a method of distinguishing colored noise with small p from a smooth signal (we assume that signal is of C^s class, with s sufficiently large). One can prove that, in contrast to colored noise, the wavelet coefficients for smooth signal have an asymptotic behavior a^{n_ψ}, in the limit $a \to 0$. Thus the simplest way to eliminate noise from smooth data is to remove the power law behavior controlled by the index p. We propose subtracting at any given point b the noise contribution of the form:

$$n(a,b) = C(b)a^{\frac{p-1}{2}}. \quad (8)$$

Here $C(b)$ is a constant computed in the region of (a,b) plane where noise prevails over a smooth part of signal. For a white Gaussian noise this obviously takes place at small scales a. Thus the cleaned signal would be:

$$\tilde{f}_c(a,b) = \tilde{f}(a,b)\left[1 - \frac{n(a,b)}{|\tilde{f}(a,b)|}\right]. \quad (9)$$

As in any application of the wavelet transform, the problem of correct choice of the mother wavelet arises. As has been shown earlier, for the singularity detection one should use the wavelet, which cancels moments of the order greater than the order of the local, at a given position b, Hölder regularity. Equation 7 suggests that a Gaussian white noise can be removed with the first derivative of the Gaussian function as mother wavelet.

3 Numerical Experiments

The purpose of this paper is to gauge the wavelet filtering method. When testing the method we have to use a certain reference quantity, which in our case is the correlation dimension D_2. For real data the correlation dimension is not known a priori. That is why we will only use data corresponding to known dynamical systems.

From a great number of algorithms for estimating D_2 we choose the most popular Grassberger-Procaccia algorithm.

To estimate some dynamical quantities we have to be able to reconstruct a d-dimensional phase space in Euclidean space (d is called embedding dimension) from a single time-dependent function $f(t)$. It was proven by Takens (1981) that a possible reconstruction formula is:

$$x_i = [f(t_i), f(t_i+\tau), ..., f(t_i+(m-1)\tau)], \quad (10)$$

with $m \geq 2d+1$.

The choice of the time delay τ is the crucial point in application of the embedding procedure. It does not need to be too short to eliminate linear correlation between vectors and too long because of finiteness of data series. We have taken τ to satisfy the condition $g(\tau) = \frac{1}{2}g(0)$, where g is the autocorrelation function of $f(t)$ (Shuster, 1988).

The correlation dimension may be defined by expression:

$$\lim_{r\to 0} P[\delta(x,y) < r] = Cr^{-D_2}, \quad (11)$$

which states that the probability that a distance δ between two points on the attractor is less than r scales as r^{-D_2}, for a properly chosen embedding dimension m.

In its simplest form the probability (in Eq. 11) may be approximated by the correlation integral $C(r)$:

$$P[\delta(x,y) < r] \approx C(r) = \frac{1}{N} \sum_{i \neq j} \Theta\left[r - ||x_i - x_j||\right], \quad (12)$$

where Θ is the Heaviside function.

Since the dimension of the phase space d is not known a priori the Grassberger-Procaccia algorithm is applied several times for various values of m_i to obtain a set of correlation integrals $C(m_i, r)$. Existence of the attractor results in the saturation of the slopes of the functions $C(m_i, r)$ on a log-log plot. It is convenient to plot the derivative of $C(m_i, r)$ with respect to r (local slope) vs. $\ln(r)$. If the noise level is not too high then over certain range of r, a scaling region, the slope remains approximately constant (see Fig. 2). If the plateaus coincide for several values of m_i, this limiting value of slope is considered as an estimate of the correlation dimension D_2.

In practice, for each embedding dimension we estimate the correlation dimension $D_2(m_i)$ as the average of local slopes over the region of the smallest standard deviation $\sigma_{D_2}(m_i)$ and the final value of D_2 is computed:

$$D_2 = \frac{\sum_{i-6}^{10} w(m_i) D_2(m_i)}{\sum_{i-6}^{10} w(m_i)}, \quad (13)$$

where weights are defined by:

$$w(m_i) = [\sigma_{D_2}(m_i)]^{-2}, \quad (14)$$

and the final error is:

$$\sigma_{D_2} = \left[\sum_{i=6}^{10} w(m_i)\right]^{-\frac{1}{2}}. \quad (15)$$

Thus we take into account only the last five embedding dimensions in the estimation of D_2. For improvements and limitations of the Grassberger-Procaccia algorithm see Ruelle (1990) and Teiler (1986).

The denoising method was tested on three models: Lorenz model, Mackey-Glass model, and non-smooth chaotic data.

3.1 The Lorentz model is described by the equations:

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x) \\ \frac{dy}{dt} &= rx - y - xz \\ \frac{dz}{dt} &= xy - bz, \end{aligned} \quad (16)$$

with $\sigma = 10$, $b = 8/3$, $r = 28$. For those parameters the system has an attractor with the correlation dimension $D_2 = 2.07$ (Shuster, 1988).

3.2 The Mackey-Glass model

$$\frac{dx}{dt} = \frac{ax(t-\tau)}{1 + [x(t-\tau)]^{10}} - bx(t), \quad (17)$$

with $a = 0.2$, $b = 0.1$, $\tau = 50$. For this set of parameters the correlation dimension $D_2 \approx 4$.

Data sets used in the analysis have been 4096 points long. The signal was polluted by an additive white Gaussian noise. To quantify noise level a parameter $R = \sigma_n/\sigma_s$was chosen, where σ_n and σ_s are standard deviations of the noise and signal, respectively. As a mother wavelet the 4th derivative of the Gaussian function was taken. The constant $C(b)$ is a modulus of the wavelet coefficient at the finest scale at a point b.

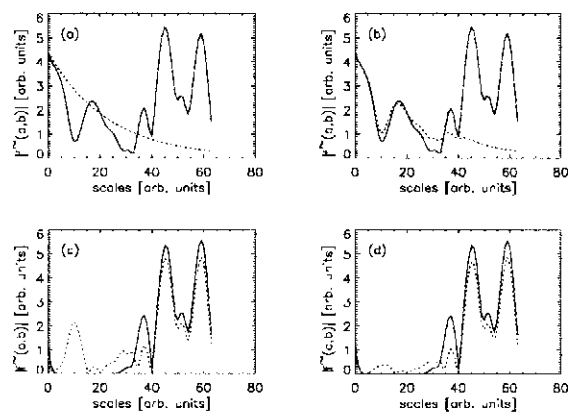

Fig. 1. Illustration of the application of formulas 9 and 18. Panels are the sections of modulus of the wavelet transform across the scales at certain point, (a) - noisy signal (solid line) and noise contribution (dotted line) as in formula 9, (b) - noisy signal (solid line) and noise contribution (dotted line) as in formula 18, (c) - clean signal (solid line) and after denoising (dotted line) as in formula 9, at small scales there are strong noise residuals, (d) - clean signal (solid line) and after denoising (dotted line) as in formula 18, filtering is much better.

We have to keep in mind that Eq. 7 was derived as an ensemble average thus for a single realization the modulus of wavelet coefficients will vary around this average. To take this into account we propose to modify Eq. 9:

$$\tilde{f}_c(a,b) = \tilde{f}(a,b)\left[1 - \frac{n'(a,b)}{|\tilde{f}(a,b)|}\right], \quad (18)$$

where:

$$\begin{aligned} n'(a,b) &= n(a,b)B(a)^{1-a}, \\ B(a) &= \frac{|\tilde{f}(a,b)|}{n(a,b)}. \end{aligned} \quad (19)$$

The denoising algorithm was applied three times. Figure 1 illustrates the effect of application of Eq. 9 and 18 on the wavelet coefficients of the noisy Lorentz time series. Scale variation of wavelet coefficients of noise causes that (9) does not remove correctly the noise contribution (panel (c) dotted line). One can see that Eq.

18 performs denoising much more effectively (panel (d) dotted line).

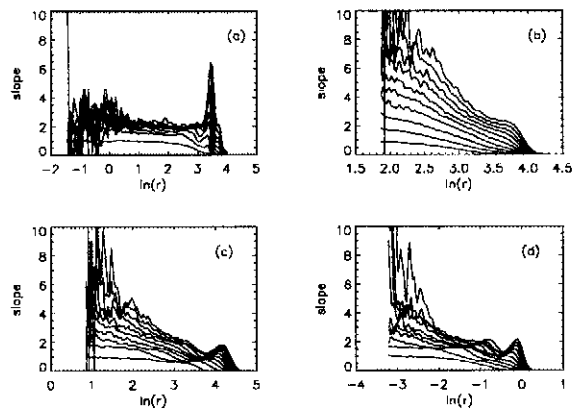

Fig. 2. Results of filtering for two kinds of filters applied to the Lorenz system. The panels are local slope vs. $\ln(r)$ (each curve is for different value of embedding dimension ranging from 1 to 10 with step 1),(a) - clean data, existence of scaling region is clearly visible, (b) - noisy data (R=50%), the scaling region does not exist, (c) - after SVD filtering, its hard to find the scaling region but effect of noise is reduced, (d) - after wavelet filtering, scaling region exists but is much narrower than for the clean data (a).

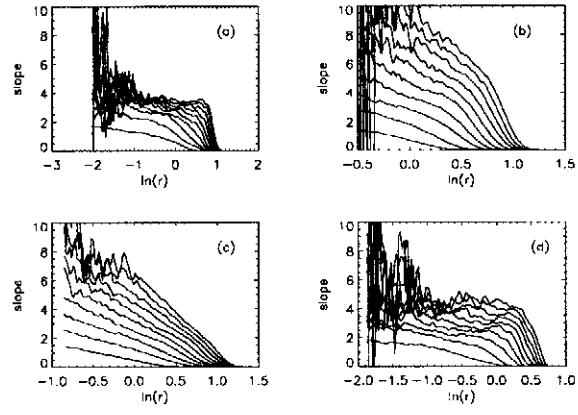

Fig. 3. Results of filtering for two kinds filters in the case of Mackey-Glass system. The panels are local slope vs. $\ln(r)$ (each curve is for different value of embedding dimension ranging from 2 to 20 with step 2), (a) - clean data, existence of scaling region is clearly visible, (b) - noisy data (R=50%), the scaling region does not exist; (c) - after SVD filtering, its hard to find the scaling region but impact of noise was decreased, (d) - after wavelet filtering, scaling region exists again but not so evident as in (a).

Figures 2 and 3 show the plots of local slope vs. $\ln(r)$ (for the Lorenz and Mackey-Glass system, respectively). As was mentioned earlier a constant value of local slope over a certain region of r suggests a finite dimension of the attractor. One can see that the wavelet filter extends the scaling region and the correlation dimension is close to the correct value, especially in the case of Mackey-Glass system, confirming the advantage of using the wavelet filtering over SVD filter. Note that the scaling region is shifted along $\ln(r)$ as a result of a reduction of the signal power in the wavelet denoising procedure.

Figures 4 and 5 illustrate the effectiveness of filtering as function of the noise level. The wavelet filter has better efficiency for both systems, especially for the Mackey-Glass system where the correlation dimension is nearly constant across the noise levels. The large errors of estimated dimensions for SVD result from the difficulty in defining the scaling region for large noise to signal ratio.

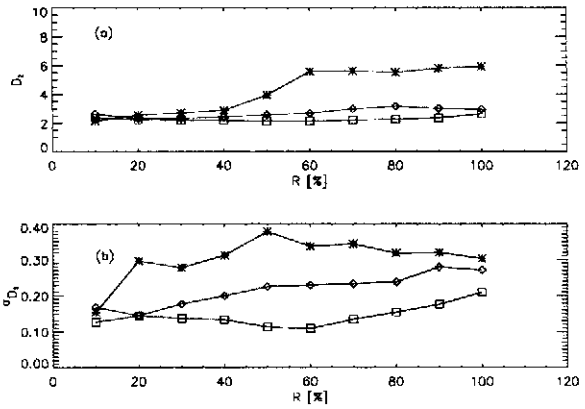

Fig. 4. Comparison of effectiveness of two filters for the Lorenz system. Top panel (a) shows correlation dimension vs. noise level (stars-noisy data, diamonds-SVD filter, squares-wavelet filter), bottom panel displays errors of estimated dimension.

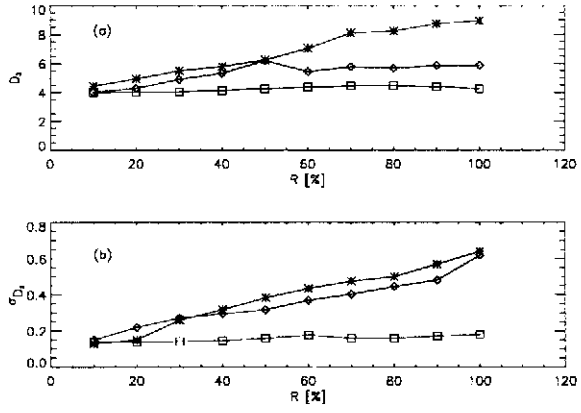

Fig. 5. Comparison of effectiveness of two filters for the Mackey-Glass system. Top panel (a) shows correlation dimension vs. noise level (stars-noisy data, diamonds-SVD filter, squares-wavelet filter), bottom panel displays errors of estimated dimension.

3.3 Non-smooth chaotic data

Now we present more complicated example of filtering of non-smooth chaotic data. Time series are obtained from a model describing the four- waves interaction in

plasma as derived by Krasnosel'skikh et al. (1998) using Zakharov equation (KLDM model). In Fig. 6 its power spectrum is compared with the spectra of previously discussed data sets. The power law behavior indicates an existence of singularities in the KLDM time series. Lorenz and Mackey-Glass spectra (Fig. 6 (a) and (b)) are typical of smooth signals, which decay very fast, faster than the power-law.

We added 50% of the white noise and used three filters: FIR (finite impulse response), SVD, and the wavelet filter. Filtering was accomplished using B spline function of 9th order i.e. by 9th-touple convolution of the characteristic function of interval [0,1].

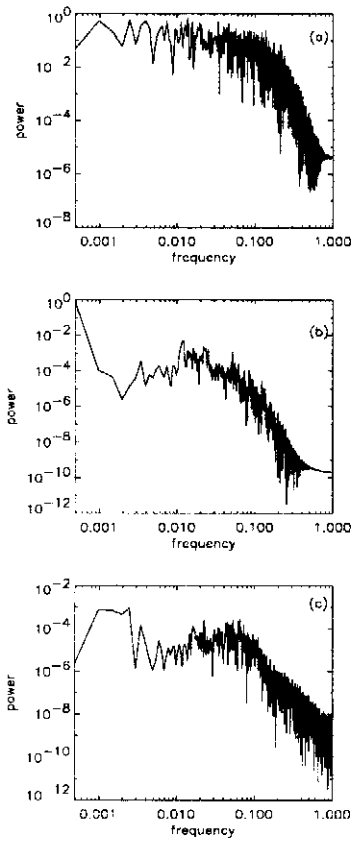

Fig. 6. Power spectra for: (a) - Lorenz system, (b) - Mackey-Glass system, (c) - KLDM system.

The power spectra for noisy and cleaned data are compared in Fig. 7. With the FIR filter one can not recover the original power law behavior of a spectrum, and the SVD filter seems to be inefficient at all. Only the wavelet filter retrieves the original power spectrum shape (Fig. 7 (d)). Figure 8 illustrates the effect of filtering on the correlation dimension. Only the FIR and the wavelet filters (Fig. 8 (d) and (e)) give reasonable evaluation of D_2: 4.650 ± 0.42 and 4.070 ± 0.28, for the FIR and wavelet filtering, respectively. The correlation dimension estimated in Krasnosel'skikh et al. (1998) is 3.24.

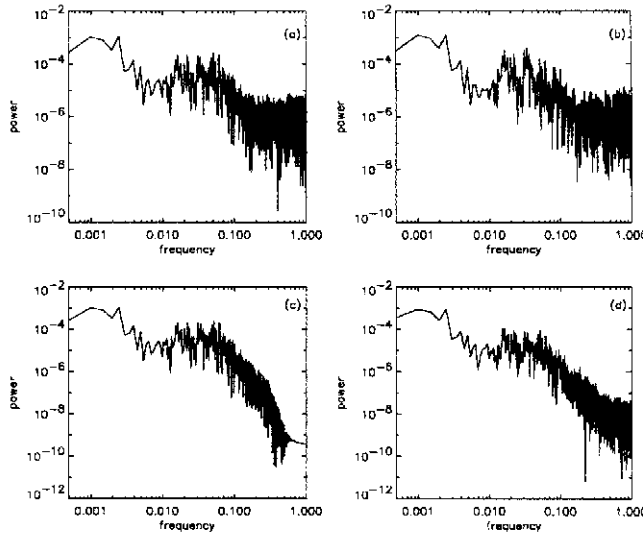

Fig. 7. Power spectra for noisy (a) (R=50%) and cleaned KLDM data with: (b) - SVD filter, (c) - FIR filter, (d) - wavelet filter.

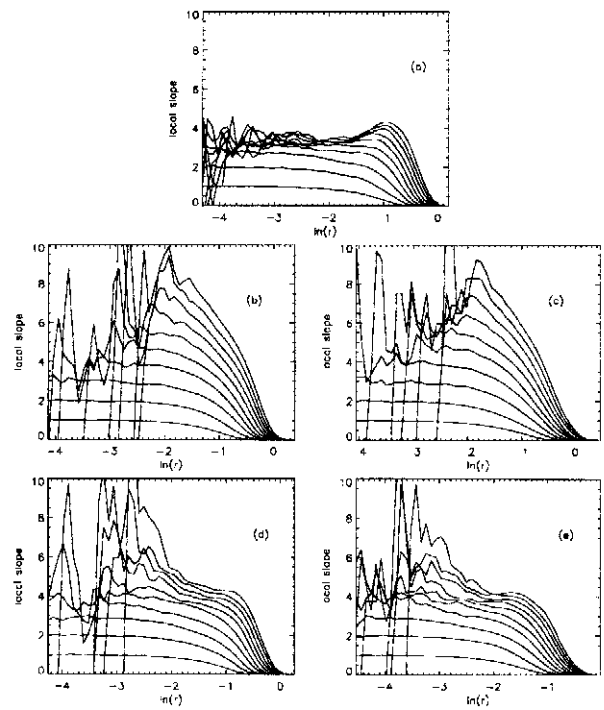

Fig. 8. Results of filtering for KLDM system. The panels are local slope vs. ln(r) (each curve is for different value of embedding dimension ranging from 1 to 10 with step 1), (a) - clean data, existence of scaling region is clearly visible, (b) - noisy data (R=50%), the scaling region does not exist; (c) - after SVD filtering, there is no evidence of scaling region; (d) - after FIR filtering; (e) - after wavelet filtering.

4 Conclusions

The numerical tests presented in this paper have shown that the wavelet transform filtering can compete with

the SVD and FIR filters.

The finite impulse response (FIR) filters are frequently used for cleaning chaotic data (Macek, 1998; Badii et al., 1988). It performs sufficiently well when the signal and noise are easily distinguishable in the Fourier domain, i.e. broad-band chaotic signal is concentrated at low frequencies while the noisy part dominates at high frequencies. Then one can choose FIR filter with spectral characteristics near to the optimal Wiener filter (see Kantz and Schreiber (1997)). But it is always difficult to select the optimal cutoff frequency, which would depend on the noise level. The choice of cutoff frequency is especially difficult when the signal or noise are nonstationary. Another problem with FIR filters is that they can increase the correlation dimension (Badii et al., 1988).

In the wavelet method only a noise contribution is subtracted. The wavelet method can be seen as a self-adjusting. Equation 18 proves that we are able to clean out a signal of a nonstationary noise provided the noise autocorrelation function is the Dirac delta function. In that case application of the FIR filters is not recommended. For nonstationary chaotic signals the nonlinear filtering in the phase-space (see for instance: Kantz and Schreiber (1997); Grassbeger et al. (1993)) is also of little use. Here we cannot obtain the proper embedding and thus extract the signal.

There is yet another advantage of using the wavelet method for cleaning a signal of the white noise contribution. Assuming that a chaotic signal is smooth up to a certain order, the wavelet removal of the noise contribution should not distort the original signal while, as was shown in the third test, FIR filter changes the original spectrum.

The assumption that chaotic signal is smooth may be viewed as too narrow, but in most cases the smoothness of dynamical systems comes from the smoothness of the right-hand side of differential equations defining the system. Besides, as has been shown, the wavelet filter can be efficient even for non-smooth signals provided its spectral index is significantly greater than the spectral index of the noise. Note however, that the wavelet filter, in this form, will not work on chaotic maps, such as the Hénon map.

The method was tested on signals with superimposed colored noise. Although the results are not as consistent as in the case of a white Gaussian noise yet the wavelet filter performs better than the SVD filter, which seems to amplify noise behavior rather than a signal.

Our tests also show that the efficiency of filtering is higher for high dimensional systems often encountered in geophysics.

Acknowledgements. The author would like to thank A.W. Wernik, M.V. Wickerhauser for stimulating discussions and V. Krasnoselskikh for making available the simulation data. This work was supported by the State Committee for Scientific Research through the grant 6P04D 04310.

References

Albano, A. M., Muench, J., Schwartz, C., Mees, A. I., Rapp, P. E., Singular-value decomposition and Grassberger-Procaccia algorithm, *Phys. Rev. A, 38*, 3017, 1988.

Badii, R., Broggi, G., Derighetti, B., Ravani, M., Ciliberto, S., Politi, A., Rubino, M. A., Dimesion increase in filtered chaotic signals, *Phys. Rev. Lett., 60*, 979, 1988.

Broomhead, D. S. and King, P., Extracting qualitative dynamics from experimental data, *Physica D, 20*, 217, 1986.

Daubechies, I., Ten lectures on wavelets, *S.I.A.M., Philadelphia*, 1992.

Grassberger, P., Hegger, R., Kantz, H., Schaffarath, C., Schreiber, T., On noise reduction for chaotic data, *Chaos, 3*, 127, 1993.

Kantz, H., Schreiber, T., Nonlinear time series analysis, *Cambridge Univ. Press*, 1997.

Kostelich, E. J. and Yorke, J. A., Noise reduction in dynamical systems, *Phys. Rev. A, 38*, 1649, 1988.

Krasnosel'skikh, V., Lefebvre, B., Dudok de Wit, T., Mourenas, D., Chaotic synchronization in three- and four-wave processes, *Physica Scripta, 57*, 246, 1998.

Macek, W. M., Testing for an attractor in the solar wind flow, *Physica D, 122*, 254, 1998.

Mallat, S. and Hwang, L. W., Singularity detection and processing with wavelets, *IEEE Trsns. on Information Theory, 38*, 1992.

Muzy, F. J., Bacry, E., Arneodo, A., The multifractal formalism revisited with wavelets, *International J. of Bifurcation and Chaos, 4(2)*, 245, 1994.

Ruelle, D., Deterministic chaos: the science and fiction, *Proc. Roy. Soc. London A, 427*, 241, 1990.

Shuster, H. G., Deterministic chaos, 2nd ed., *VCH, New York*, 1988.

Takens, F., Detecting strange attractors in turbulence, *Lecture Notes in Mathematics, 898*, 366, 1981.

Teiler, J., Spurious dimension from correlation algorithms applied to limited time-serie data, *Phys. Rev. A, 34(3)*, 2427, 1986.