Nonlinear noise reduction for electrocardiograms

Thomas Schreiber
Physics Department, University of Wuppertal,
D–42097 Wuppertal, Germany

Daniel T. Kaplan
Centre for Nonlinear Dynamics in Physiology and Medicine,
McGill University, 3655 Drummond Street,
Montréal, Québec H3G 1Y6, Canada

July 18, 1996

Abstract

The electrical activity of the heart usually shows dynamical behavior which is neither periodic nor deterministically chaotic: The interbeat intervals seem to contain a random component. Although long term predictions are thus impossible, good predictions can be made for times smaller than one heart cycle. This fact is used in order to suppress measurement errors by a local geometric projection method which was originally developed for chaotic signals. The result constitutes evidence that techniques of time series analysis based on chaos theory can be useful despite the fact that very few natural phenomena have been actually established to be deterministically chaotic.

1 Introduction

Many natural phenomena show irregular, non periodic, behavior in time. The classical explanation is that random inputs are driving the system. In the last

decade or so, chaos theory has provided an alternative approach: nonlinear deterministic dynamical systems can evolve irregularly without the need of random inputs. This new concept is particularly attractive in biology and physiology where the systems are *expected* to be nonlinear and where often a simple, linear stochastic description fails to account for the rich structure of the signals. Consequently, considerable effort has been made to apply ideas from the theory of deterministic chaos to biological and physiological data.[1]

Although the introduction of the paradigm of deterministic chaos has lead to new insights in several respects, it must be said that the number of successful practical applications of chaotic methods to biological data has remained rather small. The reason is twofold. First, many of the methods of nonlinear time series analysis have been developed using long, clean computer generated sequences and they remain useful only for exceptionally clean experimental signals. A typical biological data set will be too short, too noisy and not stationary enough to yield meaningful results. Second, most of the theory deals with the case that the system is purely deterministic and generally low dimensional, which is not expected to be true for living systems. It is an open question, which of the theoretical results remain at least approximately valid when strict determinism is lacking and which of the data analysis techniques are at all applicable to high-dimensional data.

In this paper we will study human electrocardiograms (ECGs) as an exemplary physiological system where relatively clean, long and stationary measurements are possible. See Fig. 1 for a sample time series. The underlying physiological process, the electrochemical excitation of cardiac tissue, is nonlinear and the signals show both fluctuations and remarkable structures which are not explained by linear correlations; a random signal with the same linear correlations is unrecognizable as a clinical ECG. These structures make the ECG useful to cardiologists as a diagnostic tool [2]. Despite these characteristics, ECGs are probably not deterministic chaos: [3, 4] The length of the cardiac cycles, measured by the R–R intervals between two successive ventricular beats, fluctuates with an unpredictable component (see Fig. 2). These fluctuations make long term predictions impossible, although the dynamical evolution during one cycle is more or less confined to a typical shape with some variation related to the respiratory cycle. This shape also varies when the heart rate changes.

We will exploit the short term predictability in order to suppress measurement errors in an ECG using a nonlinear projection method developed

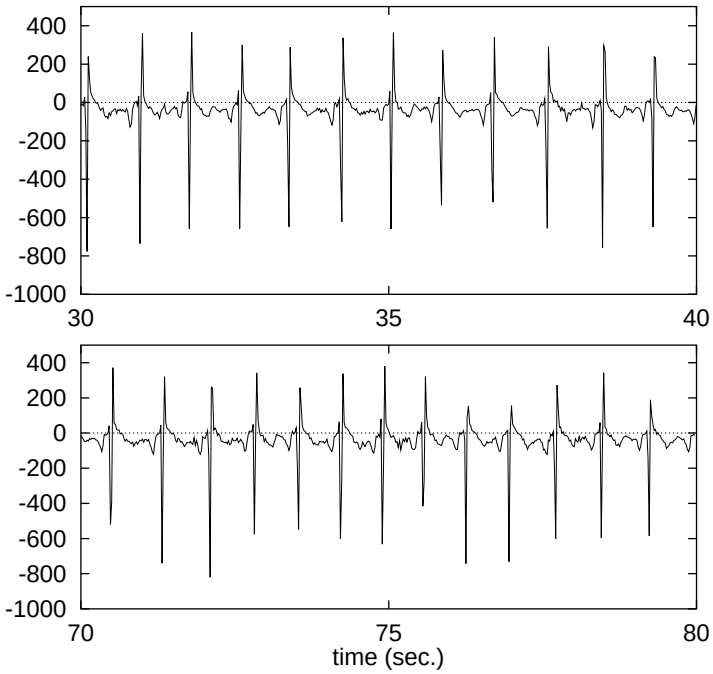

Figure 1: Two segments of a human ECG recording [5] Each panel shows ten seconds out of a total recording of 80 seconds which was used in this paper. The two segments differ mainly in the duration of the cycles. Other features, like the u–shaped P–wave prior to the big ventricular complex, remained qualitatively constant during the recording period.

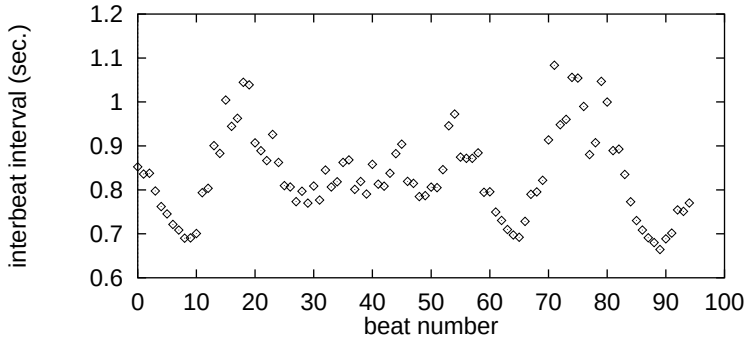

Figure 2: Interbeat intervals obtained from the ECG recording shown in Fig. 1. Analysis of much larger data bases of R–R–intervals show that such signals are not usually deterministic chaotic, except maybe for pathological cases.[3, 6]

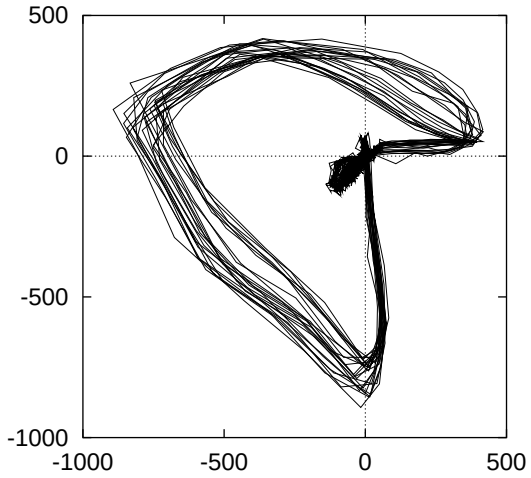

Figure 3: Delay representation with delay time 0.02 sec. of the ECG used in Fig. 4.

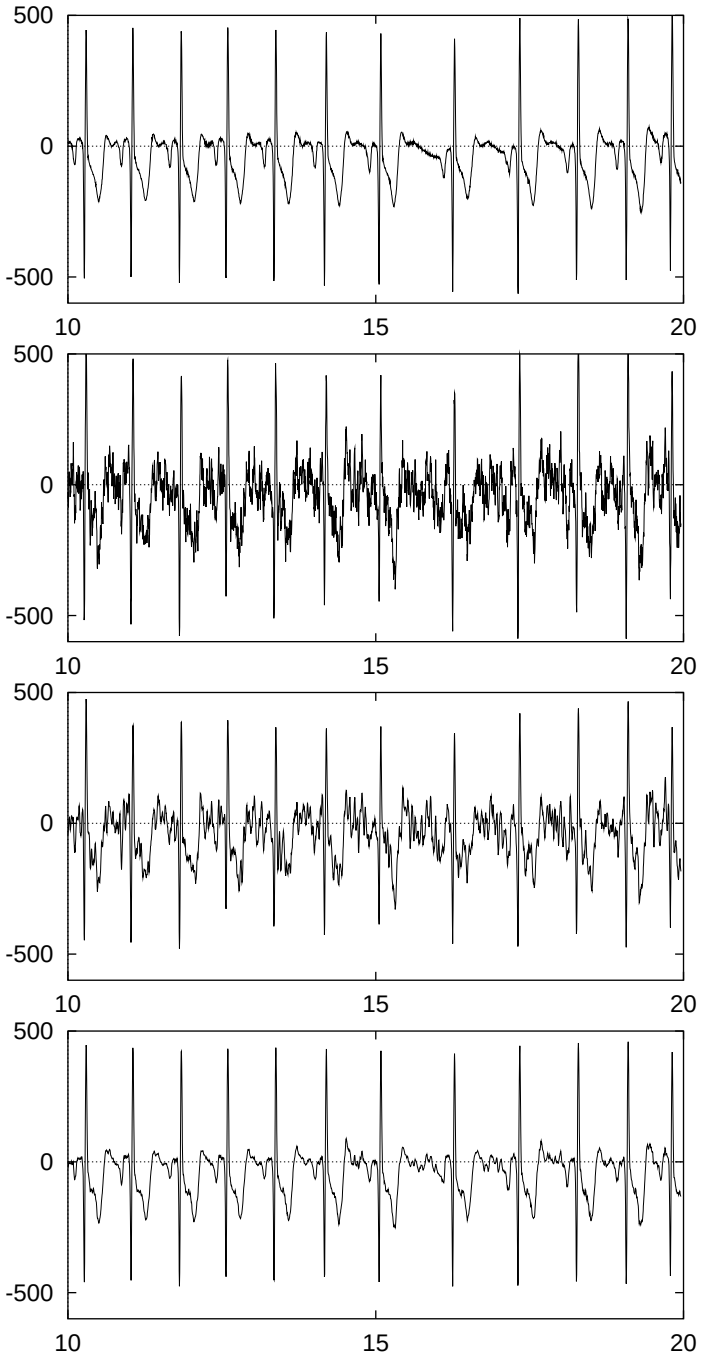

Figure 4: Noise reduction on an electrocardiogram. Upper: original data. Below: data artificially contaminated with 50% typical baseline noise. Third: data after cleaning with the optimal Wiener filter. Lower: data after nonlinear noise reduction. The error relative to the original signal is reduced by a factor of $r = 3.16$. Note the striking irregularity of the cycle lengths which seems not to affect the noise reduction procedure.

for chaotic signals. While the predominant QRS complexes which reflect the electrical depolarization of the ventricle are usually visible even in the presence of rather strong noise, more subtle features like the atrial P–wave (which normally occurs 120–200 ms before the peak of the QRS complex) may be concealed by errors which are due the imperfect transmission of the signal from the heart through different kinds of tissue, the electrode and the electronic equipment. These errors are particularly serious for ECGs taken during exercise (sweaty skin, muscle activity) and on long-term ambulatory (Holter) recordings where the experimental conditions can not be controlled that well. The peculiar twofold nature of ECGs (a pronounced pattern repeated with irregular intervals) makes it difficult to filter these signals with Fourier methods. The continuous part of the spectrum is due to both the measurement noise (we want to cancel) and to the irregular interbeat intervals (we want to preserve).

2 Nonlinear Processing

In many nonlinear methods, the first step in the analysis of deterministic time series is to reconstruct the phase space of the dynamical system, e.g. using delay coordinates $\mathbf{x}_n = (x_{n-m+1}, \ldots, x_n)$.

Due to the random fluctuations of the cycle lengths, we cannot expect the embedding theorems [8] to be completely valid here, especially near the beginning of each beat. Nevertheless, the time-delay representation is very useful to exploit the structure hidden in the scalar recordings, and the assumptions of the embedding theorems may be well enough satisfied to allow practical use to be made of the embedding technique. In this paper we will use approximate projections locally in the reconstructed phase space to reduce measurement noise in the time series.

Let us now outline the nonlinear projection technique used to reduce noise in the ECG signal. The technique has been discussed in detail for the case of a deterministically chaotic signal plus measurement noise in [9], examples with experimental data are given in [10]. ECGs are usually densely sampled in which case any of the local schemes reviewed in [11] should work about equally well. Consider for a moment a deterministic dynamical system written in m dimensional delay coordinates: $x_n = f(x_{n-m}, \ldots, x_{n-1})$. We perform a measurement which is subject to random fluctuations $y_n = x_n + \epsilon_n$.

Rewriting the dynamics in an implicit form $\tilde{f}(x_{n-m}, \ldots, x_n) = 0$ shows that in an $m+1$ dimensional delay coordinate space the noise free dynamics is constrained to an m-dimensional hypersurface. For the measured values y_n this is not exactly true, but the extension perpendicular to this hypersurface of the cloud of data points is only of the size of the noise level. Therefore one can hope to identify this direction and to correct the y_n by simply projecting them onto the subspace spanned by the clean data. In order to do this, one has to reconstruct this surface from the noisy data. Since ECGs cannot be assumed to be deterministic we cannot expect that even the noise free signal lies exactly on a low dimensional manifold. We will rather accept as an empirical fact that the observed data remain close to such a manifold.

Technically, we proceed as follows. In an embedding space of dimension $m+1$ we compute the covariance matrix of all delay vectors in a small neighborhood of a given point which we want to correct. The eigenvectors of this matrix are the semi–axes of the best approximating ellipsoid of this cloud of points. Now the important assumption is that the clean signal lives near a smooth manifold with dimension $d < m+1$, and that the variance of the noise is smaller than the signal. Then for the noisy data the covariance matrix has large eigenvalues corresponding to the directions occupied by the signal and small eigenvalues in all other directions. Therefore we move the vector under consideration towards the subspace of large eigenvectors to get rid of the noisy components. See [9] for details.

If we want to compute the correction for the n-th embedding vector $\mathbf{y}_n$, we first form a small neighborhood $\mathcal{U}$ around this point. The indices of the set of points that fall into this neighborhood is denoted $\mathcal{U}_n$, so that the neighboring points are $y_j, j \in \mathcal{U}_n$. $|\mathcal{U}_n|$ is the number of points in the neighborhood. In the following, the neighborhood size was set to the smallest value that gave $k_{\min}$ neighbors, but no smaller than 50 units (cf. Fig. 1). From the points $\mathbf{y}_k, k \in \mathcal{U}_n$ we construct the mean

$$\eta_i = \frac{1}{|\mathcal{U}_n|} \sum_{k \in \mathcal{U}_n} y_{k-m+i}, \quad i = 0, \ldots, m \tag{1}$$

and the $(m+1) \times (m+1)$ covariance matrix

$$C_{ij} = \frac{1}{|\mathcal{U}_n|} \sum_{k \in \mathcal{U}_n} y_{k-m+i} y_{k-m+j} - \eta_i \eta_j \,. \tag{2}$$

We then introduce a diagonal weight matrix R and define a transformed version of the covariance matrix $\Gamma_{ij} = R_{ii}C_{ij}R_{jj}$. In order to penalize corrections based on the first and last coordinates in the delay window we put $R_{00} = R_{mm} = r$ where r is large. The other values on the diagonal of R are 1. The Q orthonormal eigenvectors of the matrix $\mathbf{\Gamma}$ with the smallest eigenvalues are called $\mathbf{e}_q$, $q = 1, \ldots, Q$. The projector onto the subspace spanned by these vectors is then

$$\mathcal{Q}_{ij} = \sum_{q=1}^{Q} e_{q,i} e_{q,j}. \tag{3}$$

Finally the i-th component of the correction $\theta_{\mathbf{n}}$ is given by

$$\theta_{n,i} = \frac{1}{R_{ii}} \sum_{j=0}^{m} \mathcal{Q}_{ij} R_{jj} (\eta_j - y_{n-m+j}). \tag{4}$$

This gives us the correction b_- which can be added to each embedded vector to bring the point towards the manifold spanned by the $m + 1 - Q$ largest eigenvectors. Note that the R penalty matrix effectively makes the 2 largest eigenvalues lie in the subspace spanned by the first and last coordinates of the embedding space, and prevents the correction vector from having any components in these directions.

This correction is done for each embedding vector, such that we end up with a set of corrected vectors in embedding space. Since each element of the scalar time series occurs in $m + 1$ different embedding vectors, we finally have as many different suggested corrections, of which we simply take the average. Therefore in embedding space the corrected vectors do not precisely lie on the local subspaces but are only moved towards them. In contrast to strictly deterministic data we do not iterate the procedure here: the signal is not expected to be strictly confined to the manifold, only to be close to it.

3 Results

To demonstrate the effectiveness of the nonlinear noise reduction technique, we apply it to two signals (upper panels in Figs. 4 and 5) from a publically available database [5], which we artificially contaminate with noise. Let x_n be the original ECG signal (which itself contains some measurement noise),

y_n the artificially contaminated signal, and c_n the result of cleaning y_n. We can define the *noise reduction factor* to be

$$r = \sqrt{\frac{\langle (y_n - x_n)^2 \rangle}{\langle (c_n - x_n)^2 \rangle}}, \tag{5}$$

the factor by which the rms error is reduced. Since x_n was not really noise free we could succeed in cleaning y_n beyond the accuracy of x_n, whence r is a lower bound on the true noise reduction factor. Let us stress here that the rms error is not fully appropriate to represent the quality of noise reduction on ECGs. Many features of clinical relevance, like the P–waves, contribute little to the variance of the signal, which is dominated by the QRS complex. This means that a procedure which minimizes the rms error does not necessarily minimize the distortion of clinically important features. This issue is also discussed in the next section.

We applied the procedure to two fairly clean ECG signals (upper panels in Figs. 4 and 5) which we artificially contaminated with measurement noise (second panels in Figs. 4 and 5). The signals we used are contained in a publically available database [5] of the Massachusetts Institute of Technology and Beth Israel Hospital (Boston). The first example is an ECG sampled at the clinically typical rate of 250 samples per second. The data was contaminated with colored noise which has the same power spectrum as that part of the signal itself which is close to the baseline (all points not belonging to one of the QRS complexes). The noise amplitude is 50% (the signal–to–noise ratio is -10dB).

In the second example (Fig. 5) we used Gaussian white noise of an rms amplitude of 25% of that of the signal. With 250 samples per second, white noise would be easily cleaned by a low pass filter. This becomes impossible for more coarsely sampled data, like they appear e.g. in Holter recordings. To demonstrate that our method is substantially different from a low pass filter, the data was resampled to 50 points per second.

In both cases shown in Figs. 4 and 5 we embedded the time series using a delay window of 500ms (corresponding to $m = 125$ and $m = 25$ dimensions respectively) and projected onto a two-dimensional manifold locally. The radius of the neighborhoods was chosen to be 50 units.

Table 1 summarizes the resulting rms noise reduction factors for both examples. The first column shows the time span covered by the delay coor-

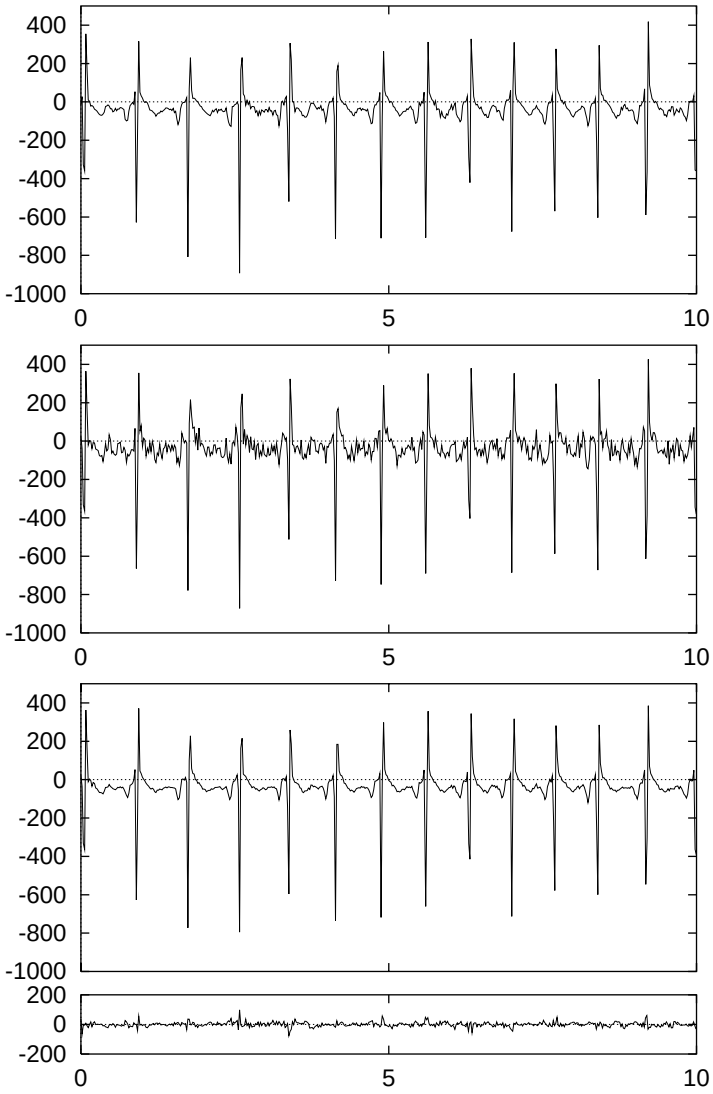

Figure 5: Noise reduction on an electrocardiogram (from the same sequence used for Figs. 1 and 2). Upper: original data. Middle: data artificially contaminated with 25% Gaussian white noise. Lower: data after nonlinear noise reduction. Many details (P and T waves) of the original series are recovered. The error relative to the original signal is reduced by a factor of $r = 1.91$. The lowest panel shows the remaining noise component, which is largest during the QRS complexes.

dinates used. A window of 200ms e.g. corresponds to an embedding dimension of 50 in the first and 10 in the second example. These high dimensional embeddings are necessary to cover a significant fraction of a typical cycle length. Trials using a delay time greater than one sample and smaller embedding dimensions were found to yield inferior results consistently. The noise reduction factor is smaller for the white noise example due to the lower noise level and the smaller sampling rate. Otherwise, the table shows that results are quite stable against changes to the parameters. We repeated the analysis with ECGs from different subjects and noise levels and always found very similar results. Let us note however that in order to further suppress already very small noise, like for example in the "clean" signals we used, it is necessary to have enough data in neighbourhoods as small as the noise amplitude. This requires many more ECG cycles and stationarity might become a serious issue.

For a comparison we also applied an optimal Wiener filter [12] in the spectral domain. In order to distinguish the signal and the noise in the spectrum we made use of the spectrum of the noise which was known in this particular case. Thus the performance of the filter can be regarded as a very optimistic example: in real examples we have to guess the spectrum of the noise from the data. Even with this aditional knowledge, the Wiener filter gave only weak noise reduction ($r = 1.33$) for the first example and almost no noise reduction ($r = 1.05$) for the second. Fig. 6 shows power spectral estimates for the first example. The original clean data, the signal after Wiener filtering, and the signal after our nonlinear noise reduction procedure have very similar spectra. The spectra of the noise added (which was also used for designing the Wiener filter. This produces a filter that is somewhat more effective than the one that would be designed in practice, where the exact spectrum of the noise is unknown.) and the noise subtacted by the nonlinear noise reduction procedure are consistent as well. Although both the linear and nonlinear cleaning procedures were able to recover the spectrum of the original signal, the nonlinear algorithm yields much better noise reduction, see Fig. 4 and Table I. Although the power spectrum of the Wiener filtered signal and the original are quite similar, the signals themselves are quite different. Clearly the phases of the Fourier components are altered by the addition of the noise in a manner that the Wiener filter cannot correct. For the second example, the Wiener filtered signal is almost indistinguishable from the noisy one and thus not shown.

t_m	k	d	$r^{1)}$	$r^{2)}$
200ms	20	1	2.88	1.55
		2	2.80	1.57
		3	2.65	1.59
	50	1	2.60	1.62
		2	2.50	1.70
		3	2.35	1.78
300ms	20	1	2.90	1.70
		2	2.77	1.73
		3	2.66	1.79
	50	1	3.02	1.75
		2	2.93	1.79
		3	2.81	1.83
500ms	20	1	3.17	1.82
		2	3.12	1.86
		3	3.04	1.92
	50	1	3.20	1.87
		2	3.16	1.91
		3	3.12	1.99
Wiener filter			1.33	1.05

t_m=embedding time window
k =minimal number of neighbors
d =dimension of manifold
r =noise reduction factor
[1)] first example
[2)] second example

Table 1: Performance of nonlinear noise reduction with different choices of the parameters of the algorithm.

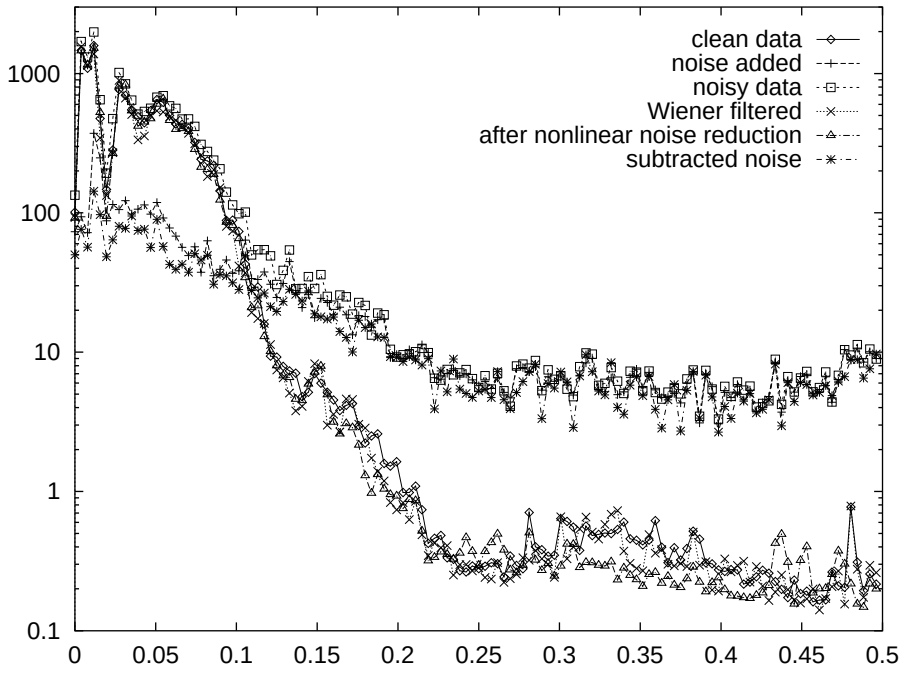

Figure 6: Power spectral estimates for the ECGs shown in Fig. 4. We show the spectra of the original, the noisy, the Wiener filtered and the nonlinearly cleaned data as well as the spectrum of the added noise which we also used for the Wiener filter. The window length for the estimates in this figure is 128.

4 Discussion

The aspects of the ECG that are utilized by clinicians are not in general well represented in the spectral domain, and so cannot be expected to be well preserved by linear filtering. This also means that the rms error is not a fully satisfactory measure for the clinical usefulness of noise reduction on an ECG. The clinically relevant features of the ECG are described in many textbooks on the art and science of ECG interpretation (e.g., see [2]). Here, we mention just a few: the interval between the onset of the P wave and the onset of the QRS complex; the existence of a negative going Q wave at the onset of the QRS complex and whether it is big or small compared to the R wave; the existence of notches in the QRS complex; the level of the ECG between the S and T waves; the width of the QRS complex; the interval between the onset of the QRS complex and the end of the T wave. Evaluating the effect of filtering on the clinical readability of the ECG ultimately requires assessing the accuracy of clinician's diagnoses, which can be quite a difficult undertaking. An example of such an assessment in the context of ECG compression algorithms is given in [13] where the authors remark that "large amplitude errors during rapidly changing portions of the QRS complex are more tolerable than much smaller errors in the baseline which may conceal or mimic P–waves." The performance of our method on such features is found to be satisfactory by inspection of Figs. 4,6, and 7; for instance, the clinical used interval between the P–wave and QRS complex is easily discerned in the cleaned data.

Since we know little about the dynamical nature of the ECG, we hesitate to use projections down to a one dimensional manifold in order not to destroy possible degrees of freedom which do not contribute significantly to the rms error but may be clinically relevant. An example would be the variation in ECG morphology during the respiration cycle or due to changes in the heart rate. As Figs. 4 and 5 show, the method is able to handle quite large fluctuations in the heart rate (see also Fig. 2). However, for nonstationary ECGs which change qualitatively in time we would suggest to modify the present algorithm by weighting neighbours according to their separation in time from the current point. The implementation of such a modified algorithm is subject to ongoing research.

In conclusion, we demonstrated that the scope of nonlinear techniques of time series analysis developed for deterministically chaotic signals is not

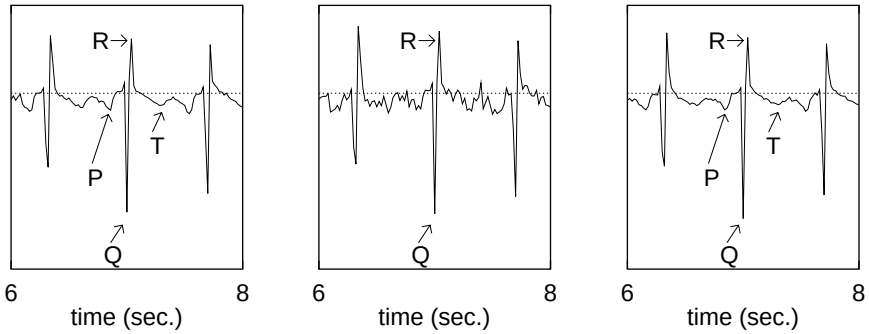

Figure 7: Identification of clinically relevant features (same data as in Fig. 5). Left: clean data, middle: noisy data, right: noisy data after nonlinear noise reduction. Note: The QRS complex is the large pair of deflections shown in the middle of each frame. In cardiologists' notation, the QRS complex can be composed of one or more "waves" Here, a Q and an R wave are present. Were the recording made with opposite polarity, this same QRS complex would be denoted as having an R and S wave."

necessarily confined to such signals. We feel that applications which do not strictly require determinism, like the one we presented above, are potentially more fruitful than the repeated but still dubious claims that deterministic chaos has been found in one or the other biological time series.

This paper does not attempt to present a filtering technique which can be relied on in clinical situations. This would require much broader studies on healthy and diseased subjects, in particular with respect to possible artifacts which are not accounted for by the rms error measure. This scepticism is not limited to our particular approach to noise reduction in ECGs: Any method that can remove noise from a fluctuating signal without creating artifacts needs to be based on some knowledge about the nature of the fluctuations. Noise reduction in ECG signals remains hazardous as long as the fluctuation of the heart rate are not better understood. Nonetheless, the ECG provides an illustration of a situation in which signal processing techniques originally developed for chaotic data can be naturally and effectively applied to practical data and can provide a superior alternative to traditional linear methods (such as the Wiener filter) in a stochastic environment.

We are indebted to Leon Glass, Peter Grassberger and Holger Kantz for stimulating discussions. This work was supported by the SFB 237 of

the Deutsche Forschungsgemeinschaft, the Natural Sciences and Engineering Research Council of Canada, and the Fonds de la recherche en santé du Québec.

References

[1] J. Bélair, L. Glass, U. an der Heiden, and J. Milton, CHAOS **5**, 1 (1995) and several other papers in the same volume.

[2] A. L. Goldberger, *Clinical Electrocardiography*, Mosby, St. Louis (1977).

[3] J. H. Lefebvre, D. A. Goodings, M. V. Kamath, and E. L. Fallen, CHAOS **3**, 267 (1993).

[4] J. K. Kanters, N.-H. Holstein-Rathlou, and E. Agner, J. Cardiovasc. Electrophysiol. **5**, 591-601 (1994)

[5] MIT-BIH Polysomnographic Database, Biomedical Engineering Centre, MIT, Cambridge (1991).

[6] T. Schreiber, unpublished (1994).

[7] L. Glass and D. Kaplan, *Time series analysis of complex dynamics in physiology and medicine* in A. S. Weigend and N. A. Gershenfeld, eds., "Time Series Prediction: Forecasting the Future and Understanding the Past", proceedings vol. XV, Santa Fe Institute Studies in the Science of Complexity, Addison-Wesley. (1993).

[8] F. Takens, "Detecting strange attractors in turbulence", in Lecture Notes in Mathematics vol. 898, Springer Verlag, Berlin (1981), T. Sauer, J. Yorke, and M. Casdagli, J. Stat. Phys. **65** 579 (1991).

[9] P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath, and T. Schreiber, Chaos **3**, 127 (1993).

[10] H. Kantz, T. Schreiber, I. Hoffmann, T. Buzug, G. Pfister, L.G. Flepp, J. Simonet, R. Badii, and E. Brun, Phys. Rev. E **48**, 1529 (1993).

[11] E. J. Kostelich and T. Schreiber, Phys. Rev. E **48** 1752 (1993).

[12] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*, Cambridge University Press (1988).

[13] G. B. Moody, R. G. Mark, and A. L. Goldberger, *Evaluation of the 'TRIM' ECG Data Compressor,* IEEE Computers in Cardiology, 167–170 (1988).