CHAPTER 6

Nonlinear Filtering Techniques

Patrick E. McSharry and Gari D. Clifford

6.1 Introduction

The ECG is routinely used to provide important clinical information. In practice, the utility of any diagnosis based on the ECG relies on the quality of the available signal. A typical ECG recorded in a clinical environment may be corrupted by one or more of the following: (1) electrical interference from surrounding equipment such as the effect of the electrical mains supply; (2) analog-to-digital conversion; and (3) movement and muscle artifacts. In order to employ the ECG signal for facilitating medical diagnosis, filtering techniques may be employed to *clean* the signal, thereby attempting to remove the distortions caused by these various sources of noise.

Many techniques for filtering are based on a spectral decomposition of the signal (see [1]). Such techniques include notch filters for removing the effect of the electrical mains supply, and both low and high bandpass filters for removing noise that is localized in particular regions of the frequency spectrum. These techniques all rely on the principle of linear superposition and there is a fundamental assumption that the underlying signal and the noise are active in different parts of the frequency spectrum. Section 3.1 provides a description of these techniques.

Linear filtering techniques are of limited use in cases where both the noise and signal occupy similar regions of the frequency domain. This restriction motivates the use of nonlinear filtering methods that do not rely on the linear assumptions underlying spectral analysis. In this chapter, three nonlinear techniques are described. These are (1) nonlinear noise reduction (NNR) [2], (2) independent component analysis (ICA) [3], and (3) model-based filtering [4]. For simplicity, these methods are demonstrated using univariate signals. Each of these techniques can easily be extended to multivariate signals arising from multiple leads and this will generally provide better performance.

Accurate metrics for evaluating the effectiveness of filtering techniques applied to the ECG are difficult to define due to the inherently complicated structure of the noise and the absence of knowledge about the underlying dynamical processes. Without having access to a noise-free ECG signal, the fact that the true underlying dynamics of a real ECG can never be known implies that one cannot distinguish between the clean ECG signal and the many sources of noise that can occur during recording in a clinical environment. While the availability of biomedical databases [5] provides a useful benchmark for comparing different techniques, this approach can never truly distinguish between noise and signal. ECGSYN, a dynamical model for generating ECG signals with known temporal and spectral

characteristics and prespecified average morphology [6] is used to compare and evaluate these techniques under a range of different conditions (see Section 4.3.2 for further details of ECGSYN).

The layout of the chapter is as follows. Section 6.2 provides an overview of nonlinear dynamics and the application of this theory to signals known as nonlinear time series analysis or nonlinear signal processing. This section describes how to reconstruct a multidimensional state space using a univariate signal and calculate nonlinear descriptive statistics such as Lyapunov exponents, correlation dimension, and entropy. The associated discussion also describes how to test the significance of these statistics and how failure to do so can lead to erroneous conclusions. Section 6.3 discusses the different forms of noise that affect the ECG and suggests different metrics for evaluating filtering techniques. Section 6.4 describes and compares two empirical filtering techniques, NNR and a state space implementation of ICA. Section 6.5 gives details of two model-based filtering approaches using ECGSYN to provide constraints on the underlying ECG signal; the first uses a nonlinear least squares parameter estimation technique and the second uses an extended Kalman filter.

These approaches illustrate different paradigms for utilizing nonlinear methods for filtering the ECG. Statistical techniques such as PCA and ICA make statistical assumptions about the relationship between the signal and the noise in a reconstructed state space. In contrast, the model-based approach makes an explicit assumption concerning the underlying structure of the ECG signal and this is encoded in a dynamical model such as ECGSYN. Nonlinear techniques are then used to find an optimal fit of this model to the data. For this reason, the model-based approach is tailor-made for ECG signals.

6.2 Nonlinear Signal Processing

Chaotic dynamics provide one possible explanation for the different complex and erratic patterns that appear in a large number of observed signals. Chaos refers to the existence of behavior so unpredictable as to appear random because of the inherent sensitivity to small perturbations in the initial conditions. This suggests that many complex systems could possibly be described by low-dimensional deterministic mathematical models. Although many real-world systems are undoubtedly nonlinear (a necessary but not sufficient condition for chaos), in practice, the quality of recorded signals is usually better suited to traditional linear analyses. Despite the fact that there is very little evidence of chaos in real systems, the field of nonlinear dynamics can help improve our understanding of many complicated systems [7]. The recent proliferation of high-speed computers and cheap storage for large databases suggests that the construction of data-driven nonlinear models is now feasible and that such developments have the potential to make a substantial contribution to science. Nonlinear signal processing relates to the data analysis that is required to construct, estimate, and evaluate these nonlinear models. The decision to pursue nonlinear models brings many challenges. As traditional techniques based on normal distributions and least squares are no longer valid, new techniques for parameter estimation [8] and model evaluation [9] are required.

6.2.1 State Space Reconstruction

Reconstructing a state space using an observed time series is usually the first step in building a model for describing nonlinear dynamics. Suppose that the underlying system dynamics of the ECG evolve on an attractor $\mathbf{A}$ according to $\mathbf{G} : \mathbf{A} \rightarrow \mathbf{A}$. Let τ_s be the sampling time of the recorded signal $s(t)$ that provides discrete observations, $s_n = s(n\tau_s)$. It is possible to construct a replica state space using a delay vector reconstruction [10–12] of the observations s_n defined by

$$\mathbf{s}_n = \left[s_n, s_{n+d} \ldots, s_{n+(m-1)d}\right] \in \Re^m \tag{6.1}$$

where m is the reconstruction dimension and $\tau_d = d\tau_s$ is the time delay. In order to reconstruct the dynamics, $\mathbf{G}$, of the system state space using a data-driven model, $\mathbf{F} : \mathbf{\Phi}(\mathbf{A}) \rightarrow \mathbf{\Phi}(\mathbf{A})$, it is necessary to ensure that the mapping $\mathbf{\Phi} : \mathbf{A} \rightarrow \mathbf{\Phi}(\mathbf{A})$ provides a faithful representation of the system's attractor.

Mathematical theory can be used to describe the conditions under which it should be possible to obtain a faithful representation of the underlying dynamics. This states that the reconstruction dimension, m, should satisfy $m > 2D_0$ where D_0 is the box-counting dimension [11, 13]. Unfortunately this theory is of little help when faced with noisy data of finite duration—the challenge of most interesting signal processing problems. For example, while the choice of τ_d is irrelevant in the theory, it is extremely important in practice. One approach for selecting τ_d is to identify the first minimum in the mutual information of the signal [14]. Another approach is based on a geometric interpretation of the reconstruction [15].

The fact the value of D_0 is unknown a priori implies that m must also be estimated. One technique for estimating m is known as the method of *false nearest neighbors* [16]. This method determines a sufficient value for m by varying the size of m and monitoring the number of false nearest neighbors associated with areas of the reconstructed state space that have self-intersections. Unfortunately, the detection of false nearest neighbors is subject to the choice of an arbitrary constant that varies with location in state space. By testing for consistency between the model dynamics and the observational uncertainty while incorporating the variation of the local instabilities of the nonlinear dynamics throughout state space, it is possible to calculate a robust estimate for the minimum value of m [9]. In practice, if the state space reconstruction is one component of a technique with an obvious application, such as the case of filtering, then it is advisable to determine values for τ_d and m by optimizing the accuracy of the filtering technique. This is possible when using the synthetic ECG signals generated by ECGSYN.

6.2.2 Lyapunov Exponents

The best known hallmark of chaotic dynamics is perhaps the inherent unpredictability of the future despite the fact that the underlying system has to obey deterministic equations of motion. This unpredictability may be quantified through the increasing average forecast error that results from larger prediction lead times. Such *sensitive dependence on initial condition* describes the inherent instability of the solutions that generate this unpredictability; two nearby initial conditions will, on average, diverge over time. Although many linear systems give rise to a slow rate of

divergence, it is the exponential divergence in nonlinear systems that is characteristic of chaotic systems.

A deterministic dynamical system may be described by a discrete map, $\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n)$ where $\mathbf{x}_n \in \Re^m$. The evolution of an infinitesimal uncertainty, ε_0, in the initial condition, $\mathbf{x}_0$, over a finite number of time steps, k, is given by $\varepsilon_k = \mathbf{M}(\mathbf{x}_0, k)\varepsilon_0$, where $\mathbf{M}(\mathbf{x}_0, k)$ is the linear tangent propagator formed by the product of the Jacobians along the k steps of the trajectory $\mathbf{M}(\mathbf{x}_0, k) = \mathbf{J}(\mathbf{x}_{k-1})\mathbf{J}(\mathbf{x}_{k-2}) \ldots \mathbf{J}(\mathbf{x}_0)$.

The linear dynamics that describe the evolution of the uncertainty quantified by ε_k may be analyzed using the singular value decomposition (SVD), $\mathbf{M} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$, where the columns of the orthogonal matrix $\mathbf{U}(\mathbf{V})$ are the left (right) singular vectors, respectively, and the entries of the diagonal matrix $\boldsymbol{\Sigma}$ are the singular values, $\sigma_i(\mathbf{x}_0, k)$, which are usually ranked in decreasing order [17]. The *finite time* Lyapunov exponents [18, 19] are defined as

$$\lambda_i^{(k)}(\mathbf{x}_0) = \frac{1}{k}\log_2 \sigma_i(\mathbf{x}_0, k), \quad i = 1, 2, \ldots, m \tag{6.2}$$

and depend on both the initial condition, $\mathbf{x}_0$, and the number of steps, k. The first finite time Lyapunov exponent $\lambda_1^{(k)}(\mathbf{x}_0)$ describes the maximum possible linear growth over the time k for which the linear propagator was defined. The Lyapunov exponents, Λ_i, are defined by taking the limit as k goes to infinity,

$$\Lambda_i = \lim_{k\to\infty} \lambda_i^{(k)}(\mathbf{x}_0), \quad i = 1, 2, \ldots, m \tag{6.3}$$

A system is said to be chaotic if the leading Lyapunov exponent, Λ_1, is positive, whereas a negative value indicates the existence of a stable fixed point. Indeed, the Lyapunov spectrum provides a means of classifying the dynamics of a system which could be useful for diagnosing *dynamical diseases* where transitions between stable and oscillatory behavior can indicate changes from health to illness or vice versa [20].

A number of approaches are available for estimating Lyapunov exponents [21–24]. By calculating the finite time Lyapunov exponents, one can monitor the convergence as a function of the amount of data available [19]. Confidence intervals should be calculated if one wishes to establish whether or not a system is chaotic [25]. In the case of an observed signal, it is advisable to first check for the existence of exponential growth. The combined effects of small data sets and noisy observations mean that it is difficult to calculate reliable estimates from signals. Techniques which specifically aim to calculate the maximal Lyapunov exponent for real signals [26, 27] are available from the TISEAN package [28].

6.2.3 Correlation Dimension

A number of different approaches, both dynamical and geometrical, exist for estimating the number of *active* degrees of freedom in a given system. Techniques for estimating fractal dimensions have received a lot of attention. The motivation for this has been the realization that if a system is governed by low-dimensional deterministic dynamics, then a model with only a few degrees of freedom exists.

The ability to obtain accurate estimates has proven difficult. It is advisable that such dimension estimates should be supported by a quantification of the confidence in the estimate. In practice, it may be easier to construct a low-dimensional model from the data than to obtain a direct estimate of the underlying system's dimension.

The *box-counting dimension* of a self-similar point set is calculated by counting the number of hyper-cubes $M(\epsilon)$ of side ϵ required to cover the set, then for a self-similar set $M(\epsilon) \propto \epsilon^{-D_0}$ as $\epsilon \to 0$ and

$$D_0 = \lim_{\epsilon \to 0} \frac{\ln M(\epsilon)}{\ln(1/\epsilon)} \tag{6.4}$$

Renyi [29] defined a family of generalized dimensions which differ in the way regions with different densities are weighted, thereby giving more weight to regions which are visited more frequently and thus contain larger fractions of the measure. These *Renyi dimensions*, D_q, are a decreasing function of q: $D_{q_1} \leq D_{q_2}$ if $q_1 > q_2$. A measure for which D_q varies with q is called a *multifractal* measure [30].

The *correlation dimension*, D_2, may be estimated from experimental data and used to suggest a suitable reconstruction dimension, m, since $D_2 \leq D_0$. D_2 reflects how the probability that the distance between two randomly chosen points will be less than ε, scales as a function of ε. For a finite sequence of points $\{\mathbf{x}_i\}_{i=1}^{N}$, a quantify known as the correlation integral measures the fraction of pairs of points $(\mathbf{x}_i, \mathbf{x}_j)$ whose distance is smaller than ε [31],

$$C(\varepsilon, N) = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \left[\sum_{j=i+1}^{N} \Theta\left(\epsilon - ||\mathbf{x}_i - \mathbf{x}_j||\right) \right] \tag{6.5}$$

where Θ is the Heaviside step function: $\Theta(x) = 0$ if $x \leq 0$ and $\Theta(x) = 1$ if $x > 0$. If this correlation integral scales as a power law, $C \sim \varepsilon^{D_2}$, then

$$D_2 = \lim_{\varepsilon \to 0} \lim_{N \to \infty} \frac{d \ln C(\varepsilon, N)}{d \ln \varepsilon} \tag{6.6}$$

For recorded signals, it is generally difficult to identify a scaling region since the finite sample size, N, places a lower bound on ε. Furthermore the finite accuracy of the measurements and the sparseness of near neighbors limits the calculation of C when ε is small.

Interpreting estimates of D_2 is also complicated by the fact that infinite dimensional stochastic signals can lead to finite and low-dimensional dimension estimates [32]. In addition, estimates for C can be biased by temporal correlations as small spatial distances between pairs of points may arise simply because the difference between their times of observation was small. Restricting attention to pairs of points separated by a specific temporal window may help to reduce this problem [32]. In practice, this problem of spatio-temporal correlations will always be present and the best approach is to use a space-time-separation diagram to test when sufficient data are not available [33].

An alternative approach employs a maximum likelihood estimate of D_2 without estimating the slope of the correlation integral directly [34]. This approach has been

extended to provide a coherent estimate of D_2, which is consistent with measures at all smaller length scales [35]. When estimating D_2 from recorded signals, it is useful to know the rate of convergence of the estimate with the quantity of data, N, so as to determine confidence intervals. There have been numerous attempts at deriving a formula for determining the minimum length of the time series, N_{min}, required to obtain accurate dimension estimates, ranging from $N_{min} = 10^{D_2/2}$ [36] to $N_{min} = D_2^{42}$ [37]; see [25] for a discussion. For $D_2 = 10$, approximations of the amount of data required vary from 10^5 to 10^{42}.

In practice, estimation of D_2 is complicated by the fact that it is extremely difficult to maintain suitable experimental conditions in order to collect a sufficient quantity of data during a period when the underlying data generating process is stationary. In addition, measurement errors also restrict the estimation of D_2.

6.2.4 Entropy

Information theory provides a probabilistic approach to measuring the statistical dependence between random variables. Let X be a random variable with probability density function $p(x)$. The information corresponding to a particular probability $p(x)$ is defined as $-\log p(x)$. This essentially implies that the more probable an event, the less information it contains. The entropy of the distribution of X, a measure of the uncertainty or disorder, is given as the average information in X:

$$H = -\int p(x) \log p(x) dx \tag{6.7}$$

The more erratic the observations of the variable X, the higher the entropy.

For example, consider a distribution such that X is uniformly distributed between 0 and 1. The integral in (6.7) may be evaluated by dividing the interval [0, 1] into n segments of length $\varepsilon = 1/n$ giving $H(\varepsilon) = -\sum_{i=1}^{n} \frac{1}{n} \log \varepsilon = -\log \varepsilon = \log n$. This demonstrates that the larger n is, the more information is gained by knowing that the variable X takes a value in some interval of length $1/n$. Alternatively, if the variable X always falls into one of the n intervals, all the probabilities will be zero except for one which will be unity. In this case the entropy is given by $H(\varepsilon) = 0$. The small value for the entropy indicates the certainty that the variable X will always land in the same interval.

Similarly, for a sequence of n random variables, $X_1, \ldots, X_n$, the joint entropy is given by

$$H_n = \int \cdots \int p(x_1, \ldots, x_n) \log p(x_1, \ldots, x_n) dx_1 \ldots dx_n \tag{6.8}$$

where $p(x_1, \ldots, x_n)$ is the joint probability for the n variables $X_1, \ldots, X_n$. The Kolmogorov-Sinai (KS) entropy measures the mean rate of creation of information by measuring how much information each new observation brings. In practice, given a signal recorded with a sampling time τ_s and a state space with dimension n partitioned using a grid size ε, the KS entropy can be expressed as

$$H_{KS} = \lim_{\tau_s \to 0} \lim_{\varepsilon \to 0} \lim_{n \to \infty} (H_{n+1} - H_n) \tag{6.9}$$

In the case of a deterministic periodic systems, the KS entropy is zero because the initial condition specifies all future states. In contrast, the KS entropy is maximum for uncorrelated random processes because each new state is totally independent of the previous states [38].

The estimation of entropy for recorded signals is complicated by the usual problems of small data and noisy signals. Approximate entropy (ApEn) was proposed as a measure of system complexity and quantifies the unpredictability of fluctuations in a time series [39]. The difficulty in estimating the entropy of short noisy time series has motivated an alternative approach, known as sample entropy [40], with the advantage of being less dependent on the length of the time series.

6.2.5 Nonlinear Diagnostics

Although the concept of using nonlinear statistics, such as D_2 or Λ_1, to categorize the state of an observed system is appealing, one should be aware that this is complicated by the previously discussed problems of obtaining accurate estimates. The ability to measure the complexity of an observed system may be useful for classifying states of health and disease and could form the basis of a diagnostic tool.

For clinical applications it is important to rule out simple linear statistics before inventing complicated nonlinear measures. The possible existence of strong correlations between simple linear statistics and their nonlinear counterparts suggests that newly proposed statistics based on nonlinear dynamics should be tested against simple traditional benchmarks. Indeed, published results using the correlation integral to forecast epileptic seizures from the electroencephalogram EEG were reproduced using the variance of the EEG signal [41]. The fact that an increase in variance produces a decrease in the value of correlation integral implies that carefully designed statistical tests with clinical relevance are required [42].

In practice, nonlinear statistics should only be evaluated on segments of data that arise when the underlying dynamics are stationary. Application of a test for stationarity to RR intervals [43] using 48-hour Holter recordings from 23 healthy subjects during sinus rhythm demonstrated that while it was relatively easy to find stationary periods containing 1,024 RR intervals, only a few stationary segments of between 8,192 and 32,768 RR intervals were found [44]. Using statistical tests based on the correlation dimension, the authors were able to reject the hypothesis that the RR intervals represented a static transformation of a linear process and found evidence for time irreversibility. These results suggest that heart rate variability is driven by nonlinear processes and that the RR intervals may contain more information than can be extracted by linear analyses in the time and frequency domains.

Despite the drawbacks caused by the difficulty in estimating nonlinear statistics, a number of techniques have been proposed in the field of biomedical research. Both an effective correlation dimension [45] and a method based on the convergence and divergence of short-term maximum Lyapunov exponents from adaptively selected electrodes [46] have been used to provide predictions of epileptic seizures. Approximate entropy has been used to discern changing complexity from relatively small amounts of data arising from RR intervals [47]. Sample entropy [40] has also been employed to explore the multiple time scales involved in physiological dynamics.

A technique, known as multiscale entropy [48], can distinguish between RR intervals from subjects within healthy and pathologic groups such as atrial fibrillation and congestive heart failure.

Nonlinear dynamics has proven useful for constructing improved methods for predictive recognition of patients threatened by sudden cardiac death [49]. Such nonlinear methods have shown promise in classifying fatal cardiac arrhythmias such as ventricular tachycardia (VT) and ventricular fibrillation (VF). In a study of 17 chronic heart failure patients before the onset of a life-threatening arrhythmia and at a control time (without either VT or VF events), neither time nor frequency domain statistics showed significant differences between the VT-VF and the control time series, whereas methods based on symbolic dynamics and finite-time growth rates were able to discriminate significantly between both groups [50].

6.3 Evaluation Metrics

Noise usually describes the uncertainty in the data resulting from measurement errors or specifically the part of the data that does not directly reflect the underlying system of interest. Sources of noise commonly encountered in the ECG include (1) electrical interference, (2) analog-to-digital conversion, and (3) movement or muscle artifacts. There is an important difference between *observational uncertainty* and *dynamical uncertainty*. Observational uncertainty refers to measurement errors which are independent of the dynamics. Sources include finite precision measurements, truncation errors, and missing data (both temporal and spatial). In contrast, dynamical uncertainty refers to external fluctuations interacting with and changing internal variables in the underlying system. Dynamical uncertainty includes parametrical and structural uncertainty, both of which lead to model error. An example of dynamical uncertainty is where a parameter value, assumed constant in the equations describing the underlying dynamics, was actually varying during the time when the data were being recorded. In short, observational uncertainty obscures the state vectors whereas dynamical uncertainty changes the actual dynamics.

In the following, only the effects of noise due to observational uncertainty are considered. Let τ_s be the sampling time of the recorded signal so that the observed time series is $y_n = y(n\tau_s)$. A simple description of observational uncertainty is provided by additive measurement error where the observed time series is

$$y_n = x_n + \epsilon_n \tag{6.10}$$

x_n is the true state vector and ϵ_n represents the unobserved measurement error. This measurement error term is usually described by a random variable, for example an identically and normally distributed (IND) process, $\epsilon \sim N(0, \sigma^2_{\text{noise}})$, where σ^2_{noise} is the variance of the noise. If the variance of the signal is σ^2_{signal}, then the SNR is defined as

$$\gamma = \frac{\sigma_{\text{signal}}}{\sigma_{\text{noise}}} \tag{6.11}$$

After employing a particular technique for *cleaning* a noisy dataset, a performance metric is required to assess its failure or success. Let the cleaned signal be z_n. Following Schreiber and Kaplan [2], a *noise reduction factor* is defined as

$$\chi = \sqrt{\frac{\langle y_n - x_n \rangle^2}{\langle z_n - x_n \rangle^2}} \tag{6.12}$$

where $\langle \cdot \rangle$ denotes the average calculated by summing over the observed time series, indexed by n. The value of χ provides a measure of the degree by which the RMS error is reduced. Unlike the investigation of Schreiber and Kaplan [2], the ECG model may be used to obtain a truly noise-free signal x_n so that the value of χ may be viewed as the actual noise reduction factor and not merely a lower bound. The higher the value of χ, the better the noise reduction procedure, whereas $\chi = 1$ indicates no improvement since similar accuracy could have been achieved by using the noisy signal, y_n, instead of z_n.

An alternative measure of noise reduction performance is given by a measure of the linear correlation between the cleaned signal, z_n, and the original noise-free signal, x_n. The cross-correlation coefficient ρ between x_n and z_n is given by [51]

$$\rho = \frac{\langle [x_n - \mu_x][z_n - \mu_z] \rangle}{\sigma_x \sigma_z} \tag{6.13}$$

where μ_x and σ_x are the mean and standard deviation of x_n, and μ_z and σ_z are the mean and standard deviation of z_n. A value of $\rho \sim 1$ reflects a strong correlation, $\rho \sim -1$ implies a strong anticorrelation, and $\rho \sim 0$ indicates that x_n and z_n are uncorrelated. This means that a value of $\rho = 1$ suggests that the noise reduction technique has removed all the noise from the observed signal.

6.4 Empirical Nonlinear Filtering

The ability to reconstruct a multidimensional state spaces from a univariate signal means that a number of multivariate techniques can now be applied. In the following, nonlinear noise reduction and independent component analysis are described. The two techniques are then employed to filter a synthetic ECG signal with known characteristics produced by ECGSYN. The performance is measured as a function of the SNR using both the noise reduction factor and the correlation between the cleaned signal and the original noise-free signal.

6.4.1 Nonlinear Noise Reduction

The ECG signal cannot be classified as either periodic or deterministically chaotic. Although the dynamics of the cardiac interbeat interval time series is similar to a $1/f$-like noise process [52, 53], midway between the complete randomness of white noise and the much smoother Brownian motion, the ECG signal displays limited predictability over times less than one heartbeat since each beat contains the familiar

P wave, QRS complex, and T wave. Schreiber and Kaplan [2] successfully applied a technique, originally constructed for removing noise from chaotic signals [54], to ECG signals. This short-term predictability may be used to reduce measurement errors by a local geometric method. The basic idea behind nonlinear noise reduction (NNR) is to use the manifold of the underlying dynamical system to project out the noise. This may be achieved by using a local linear model to predict a particular point in the state space while using its neighbors to construct a local linear map.

Following [2], consider a state space reconstruction with dimension m and time delay d, such that the underlying noise-free time series x_n is described by the evolution of a deterministic dynamical system,

$$x_n = f(x_{n-md}, \ldots, x_{n-2d}, x_{n-d}) \tag{6.14}$$

By rewriting the dynamics represented by (6.14) in the implicit form of

$$g(x_{n-md}, \ldots, x_{n-d}, x_n) = 0 \tag{6.15}$$

it is apparent that the noise-free dynamics are constrained to an m-dimensional hypersurface. While this is only approximately true for the observed noisy time series, $y_n = x_n + \epsilon_n$, one can still attempt to estimate the noise-free value x_n from the noisy y_n by projecting them onto the subspace spanned by the filtered data expressed through (6.15). The objective is to estimate this subspace from the noisy data while assuming that the observed time series lies close to a low-dimensional manifold. NNR relies on the assumption that the signal lies on a manifold with dimension less than $m+1$, and that the variance of the noise is smaller than that of the signal.

In order to calculate a noise-free estimate for a particular reconstructed state vector, $\mathbf{y}_n$, its local neighborhood is defined; let $\mathcal{B}_n$ denote the indices of the set of points that form this neighborhood, y_j, $j \in \mathcal{B}_n$, and $|\mathcal{B}_n|$ be the number of neighbors. The mean value, μ_{y_i}, of each coordinate, $i = 0, \ldots, m$, is given by

$$\mu_{y_i} = \frac{1}{|\mathcal{B}_n|} \sum_{k \in |\mathcal{B}_n|} y_{k-(m+i)d} \tag{6.16}$$

The covariance matrix, C_{ij}, of the points in the neighborhood, $\mathcal{B}_n \in \Re^{m+1}$, is

$$C_{ij} = \frac{1}{|\mathcal{B}_n|} \sum_{k \in |\mathcal{B}_n|} y_{k-(m+i)d}\ y_{k-(m+j)d} - \mu_{y_i}\mu_{y_j} \tag{6.17}$$

and its eigenvectors give the principal axes of an ellipsoid that approximates this cloud of neighbors. Corrections based on the first and last coordinates in the delay vector may be penalized by using a diagonal weight matrix, $\mathbf{R}$, to transform the covariance matrix, $\Gamma_{ij} = R_{ii} C_{ij} R_{jj}$, where $R_{00} = R_{mm} \gg 1$ and all other diagonal values are equal to one [7]. The Q orthonormal eigenvectors of the matrix, Γ_{ij}, with the smallest eigenvectors are called $\mathbf{e}_q$ for $q = 1, \ldots, Q$. A projection matrix, P_{ij}, may be defined as

$$P_{ij} = \sum_{q=1}^{Q} e_{q,i} e_{q,j} \tag{6.18}$$

so that the ith component of the correction vector, $\mathbf{c}_n$, is given by

$$c_{n,i} = \frac{1}{R_{ii}} \sum_{j=0}^{m} P_{ij} R_{jj} (\mu_{y_j} - y_{n-(m+j)d}) \tag{6.19}$$

This correction vector can then be added to each reconstructed state vector in order to move it towards the manifold spanned by the $m + 1 - Q$ largest eigenvectors. The role of the penalty matrix, $\mathbf{R}$, is to force the two largest eigenvalues to lie in the subspace spanned by the first and second coordinates of the reconstructed state space and ensures that the correction vector will not have any components in these directions. Following this procedure, given that each scalar observation occurs in $m + 1$ different reconstructed state vectors, this will provide as many suggested corrections. Taking the average of all these corrections gives the overall correction used for filtering the time series.

In the case of data from a deterministic system, it may be beneficial to iterate this process a number of times to clean the time series. For the ECG, where the signal is only expected to lie close to a manifold, the best results are obtained using only one iteration [2]. This type of filtering is nonlinear in the sense that the effective filter given by the local linear map varies throughout state space depending on the local dynamics. In particular it has the ability to remove noise from the recorded signal, even in cases when the underlying signal and the noise overlap in the frequency domain.

This NNR technique is available as *nrlazy* from the TISEAN software package [28, 55], and it requires the choice of various parameters such as the reconstruction dimension, m, the time delay, d, and the neighborhood radius, r. Using ECGSYN, it is possible to generate noise-free artificial ECG signals such that the correct answer is known a priori. Within this setting, it is possible to optimize the filtering procedure by performing a thorough search of the parameter space required to implement NNR.

6.4.2 State Space Independent Component Analysis

ICA is a statistical technique for decomposing a dataset into independent subparts [3, 56]. Using the delay reconstruction described in Section 6.2.1, the observed univariate ECG signal, $y_i = y(i\tau_s)$, is transformed into an $m \times n$ matrix,

$$\mathbf{Y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \\ y_{1+d} & y_{2+d} & \cdots & y_{n+d} \\ \vdots & \vdots & & \vdots \\ y_{1+(m-1)d} & y_{2+(m-1)d} & \cdots & y_{n+(m-1)d} \end{bmatrix} \tag{6.20}$$

where each column of $\mathbf{Y}$ contains one reconstructed state vector as defined by (6.1) with reconstruction dimension m and delay d. Note that the observed ECG signal, y_i, is assumed to have a mean of 0 and a standard deviation of 1, achieved by removing the mean, μ_y, of y and dividing by its standard deviation, σ_y. After application of

the ICA algorithm, the resulting cleaned signal is rescaled by multiplying by σ_y and adding μ_y so that it is compatible with y_i.

In mathematical terms, the problem may be expressed as

$$\mathbf{Y} = \mathbf{BX} \tag{6.21}$$

where $\mathbf{X}$ is an $m \times n$ matrix containing the independent source signals, $\mathbf{B}$ is the $m \times m$ mixing matrix, and $\mathbf{Y}$ is an $m \times n$ matrix containing the observed (mixed) signals. ICA algorithms attempt to find a separating or demixing matrix $\mathbf{W}$ such that

$$\mathbf{X} = \mathbf{WY} \tag{6.22}$$

In practice, iterative methods are used to maximize or minimize a given cost function such as mutual information, entropy, or the kurtosis (fourth-order cumulant), which is given by

$$kurt(\mathbf{Y}) = E\{\mathbf{Y}^4\} - 3(E\{\mathbf{Y}^2\})^2 \tag{6.23}$$

where $E\{\mathbf{Y}\}$ is the expectation of $\mathbf{Y}$. The following analysis uses Cardoso's multidimensional ICA algorithm *jadeR* [3], which is based upon the joint diagonalization of cumulant matrices, because it combines the benefits of both PCA and ICA to provide a stable deterministic solution. (ICA suffers from a scaling and column ordering problem due to the indeterminacy of solution to scalar multipliers to and column permutations of the mixing matrix.)

Most ICA methods assume there are at least as many independent measurement sensors as the number of sources that one wishes to separate. Following James et al. [56], ICA was applied to the embedding matrix $\mathbf{Y}$, with the assumption of one signal and one noise source. Cardoso's *jadeR* algorithm was used for the ICA. An estimate $\hat{\mathbf{X}}$ of the sources $\mathbf{X}$ was obtained from

$$\hat{\mathbf{X}} = \mathbf{WY} \tag{6.24}$$

where $\hat{\mathbf{X}}$ is the ICA estimate of $\mathbf{X}$. Note that due to the scaling and inversion indeterminacy problem of ICA, both $\pm$ each row of $\mathbf{X}$ must be considered. The scaling problem is addressed by multiplying by σ_y and adding μ_y. The row with the highest correlation with the original noise-free signal is chosen as the best estimate, $z(t)$, of noise-free signal $x(t)$.

6.4.3 Comparison of NNR and ICA

The performance of NNR and ICA was compared using a synthetic ECG signal generated by ECGSYN. Both techniques were used to remove the distortions arising from a stochastic noise source. The noise was assumed to result from additive measurement errors represented by a normal distribution with zero mean. Signal to noise ratios of $\gamma = 10, 5, 2.5$ were considered. The effect of IND additive measurement errors with $\gamma = 10$ on the ECG signal is shown in Figure 6.1.

The NNR technique produced optimal results for a delay of $d = 1$. For a signal to noise ratio of $\gamma = 10$, NNR was first applied to a coarse range of values of

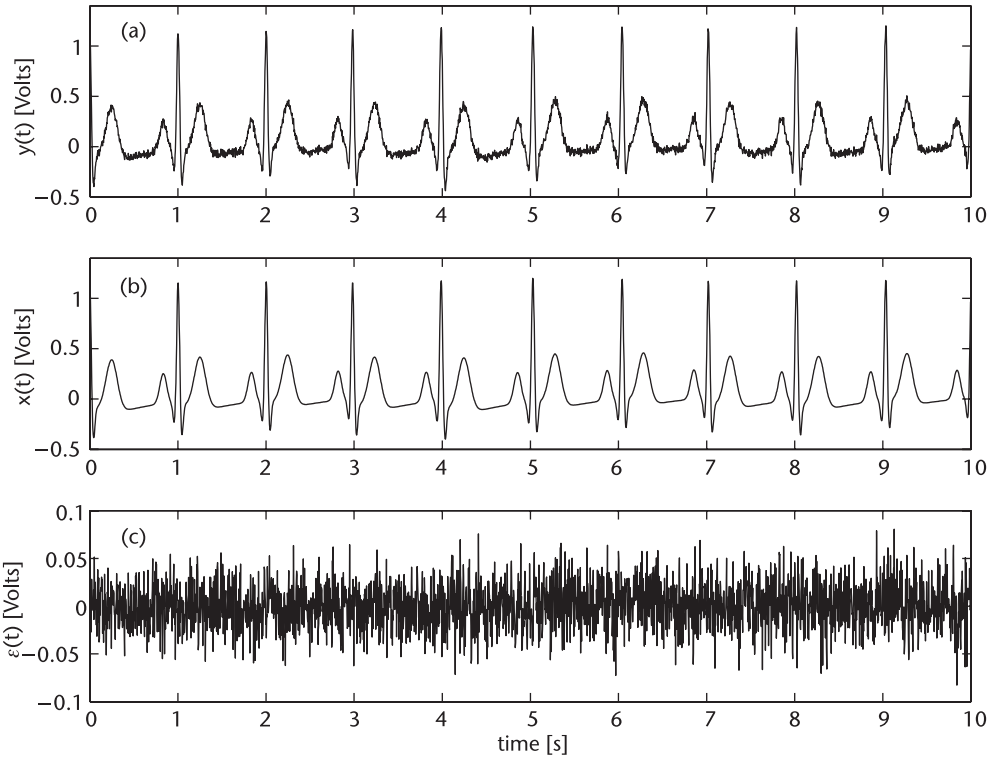

Figure 6.1 ECG signal generated by ECGSYN using additive IND measurement errors with a signal-to-noise ratio of $\gamma = 10$: (a) synthetic ECG signal with additive measurement noise, $y(t)$; (b) noise-free synthetic ECG signal, $x(t)$; and (c) measurement errors $\epsilon(t)$.

m and r in order to illustrate how the performance of the technique varied with these two parameters. This gave an optimal noise reduction factor of $\chi = 2.09$ for $m = 16$ and $r = 0.08$ (Figure 6.2). A closer examination of the dependence of χ on m was obtained by taking a cross-section of the surface shown in Figure 6.2 at $r = 0.08$ where m was sampled at all integer values between 1 and 120. As shown in Figure 6.3(a) there is a maximum noise reduction factor of $\chi = 2.2171$ at $m = 20$. The various time series and the error involved in the noise reduction process are illustrated in Figure 6.4. This shows that while the cleaned signal, $z(t)$, [Figure 6.4(c)] closely resembles the original noise-free signal, $x(t)$, [Figure 6.4(b)], there still remains considerable structure in the error, $z(t)-x(t)$, [Figure 6.4(d)]. This structure is particularly evident and appears larger around the QRS complex. As pointed out by Schreiber and Kaplan [2], the NNR technique attempts to minimize the resulting RMS error and does not directly aim to recover other key characteristics of the ECG that may be of more clinical relevance to the physician. Despite this, NNR does recover the peaks and troughs that define the morphology of the ECG. Both the P waves and T waves are clearly visible in Figure 6.4(c) and their positions and magnitudes remain faithful to that of the original noise free ECG shown in Figure 6.4(b).

The NNR technique gave optimal results for neighborhoods of different sizes depending on the signal to noise ratio: (1) $r = 0.08$ for $\gamma = 10$, (2) $r = 0.175$ for $\gamma = 5$, and (3) $r = 0.4$ for $\gamma = 2.5$. Figure 6.3 shows both the noise reduction

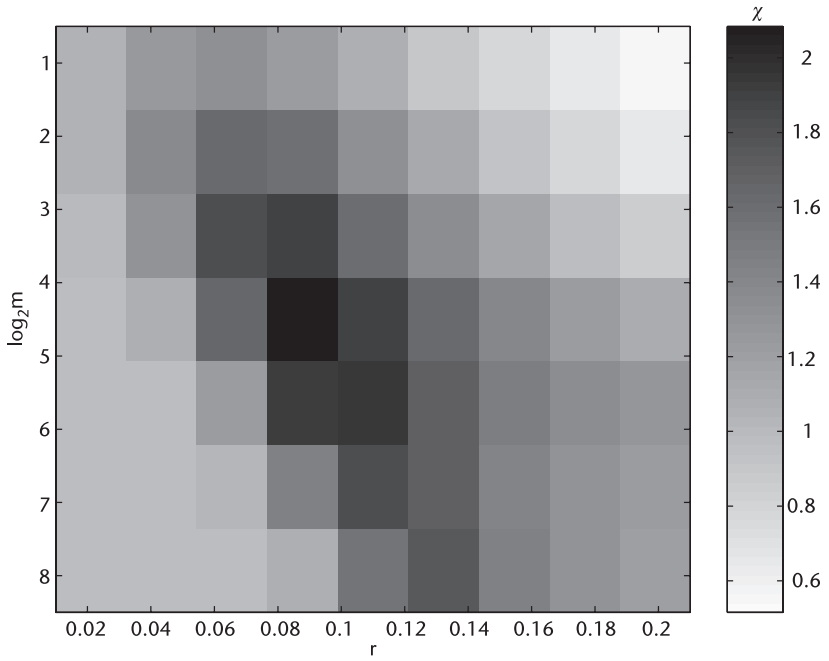

Figure 6.2 Noise reduction factor, χ, for different values of the reconstruction dimension, m, and neighborhood size, r, using NNR applied to data with a signal-to-noise ratio of $\gamma = 10$.

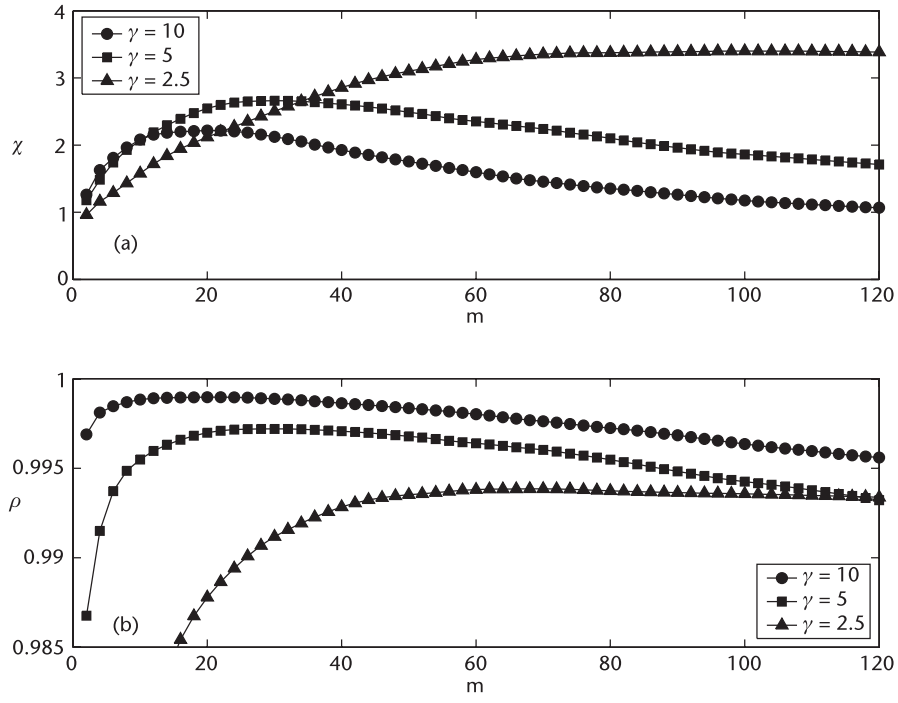

Figure 6.3 Variation in (a) noise reduction factor, χ, and (b) correlation, ρ, with reconstruction dimension, m, for NNR applied to data with signal to noise ratios of $\gamma = 10$ (•), $\gamma = 5$ (■), and $\gamma = 2.5$ (▲), having neighborhood sizes of $r = 0.08, 0.175, 0.4$, respectively.

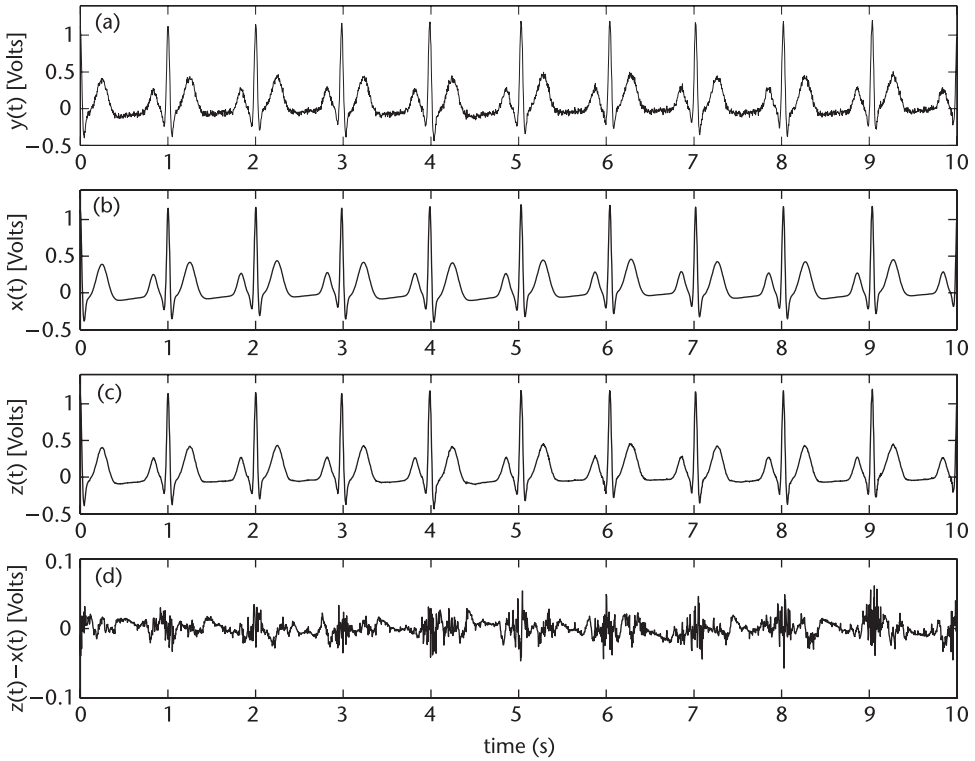

Figure 6.4 Illustration of NNR: (a) original noisy ECG signal, $y(t)$; (b) underlying noise-free ECG, $x(t)$; (c) noise-reduced ECG signal, $z(t)$; and (d) remaining error, $z(t)-x(t)$. The signal-to-noise ratio was $\gamma = 10$ and the NNR used parameters $m = 20$, $d = 1$, and $r = 0.08$.

factor, χ, and the correlation, ρ, as a function of the reconstruction dimension, m, for signals having $\gamma = 10, 5, 2.5$. For $\gamma = 10$, maxima occur at $\chi = 2.2171$ and $\rho = 0.9990$, both with $m = 20$. For the intermediate signal to noise ratio, $\gamma = 5$, maxima occur at $\chi = 2.6605$ and $\rho = 0.9972$ with $m = 20$. In the case of $\gamma = 2.5$, the noise reduction factor has a maximum, $\chi = 3.3996$ at $m = 100$, whereas the correlation, $\rho = 0.9939$, has a maximum at $m = 68$.

ICA gave best results for all signal to noise ratios for a delay of $d = 1$. As may be seen from Figure 6.5, optimizing the noise reduction factor, χ, or the correlation, ρ, gave maxima at different values of m. For $\gamma = 10$, the maxima were $\chi = 26.7265$ at $m = 7$ and $\rho = 0.9980$ at $m = 9$. For the intermediate signal to noise ratio, $\gamma = 5$, the maxima are $\chi = 18.9325$ at $m = 7$ and $\rho = 0.9942$ at $m = 9$. Finally for $\gamma = 2.5$, the maxima are $\chi = 10.8842$ at $m = 8$ and $\rho = 0.9845$ at $m = 11$. A demonstration of the effect of optimizing the ICA algorithm over either χ or ρ is illustrated in Figure 6.6. While both the χ-optimized cleaned signal [Figure 6.6(b)] and the ρ-optimized cleaned signal [Figure 6.6(d)] are similar to the original noise-free signal [Figure 6.6(a)], an inspection of their respective errors, [Figure 6.6(c)] and [Figure 6.6(e)], emphasizes their differences. The χ-optimized outperforms the ρ-optimized in recovering the R peaks.

A summary of the results obtained using both the NNR and ICA techniques are presented in Table 6.1. These results demonstrate that NNR performs better in terms of providing a cleaned signal which is maximally correlated with the original

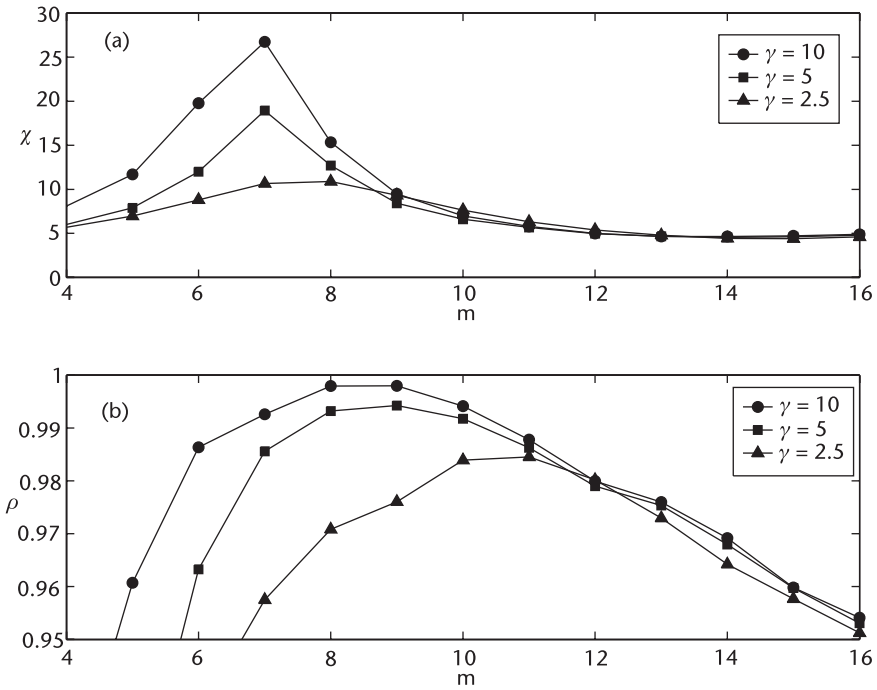

Figure 6.5 Variation in (a) noise reduction factor, χ, and (b) correlation, ρ, for ICA with reconstruction dimension, m, and delay, $d = 1$. The signal-to-noise ratios are $\gamma = 10$ (•), $\gamma = 5$ (■), and $\gamma = 2.5$ (▲).

Table 6.1 Noise Reduction Performance in Terms of Noise Reduction Factor, χ, and Correlation, ρ, for Both NNR and ICA for Three Signal-to-Noise Ratios, $\gamma = 10, 5, 2.5$

Method	*Measure*	$\gamma = 10$	$\gamma = 5$	$\gamma = 2.5$
NNR	χ	2.2171	2.6605	3.3996
ICA	χ	26.7265	18.9325	10.8842
NNR	ρ	0.9990	0.9972	0.9939
ICA	ρ	0.9980	0.9942	0.9845

noise-free signal, whereas ICA performs better in terms of yielding a cleaned signal which is closer to the original noise-free signal, as measured by an RMS metric.

The decision between seeking an optimal χ or ρ depends on the actual application of the ECG signal. If the morphology of the ECG is of importance and the various waves (P, QRS, T) are to be detected, then perhaps a large value of ρ is of greater relevance. In contrast, if the ECG is to be used to derive RR intervals for generating an RR tachogram, then the location in time of the R peaks are required. In this latter case, the noise reduction factor, χ, is preferable since it penalizes heavily for large squared deviations and therefore will favor more accurate recovery of extrema such as the R peak.

6.5 Model-Based Filtering

The majority of the filtering techniques presented so far involve little or no assumptions about the nature of either the underlying dynamics that generated the signal or the noise that masks it. These techniques generally proceed by attempting to

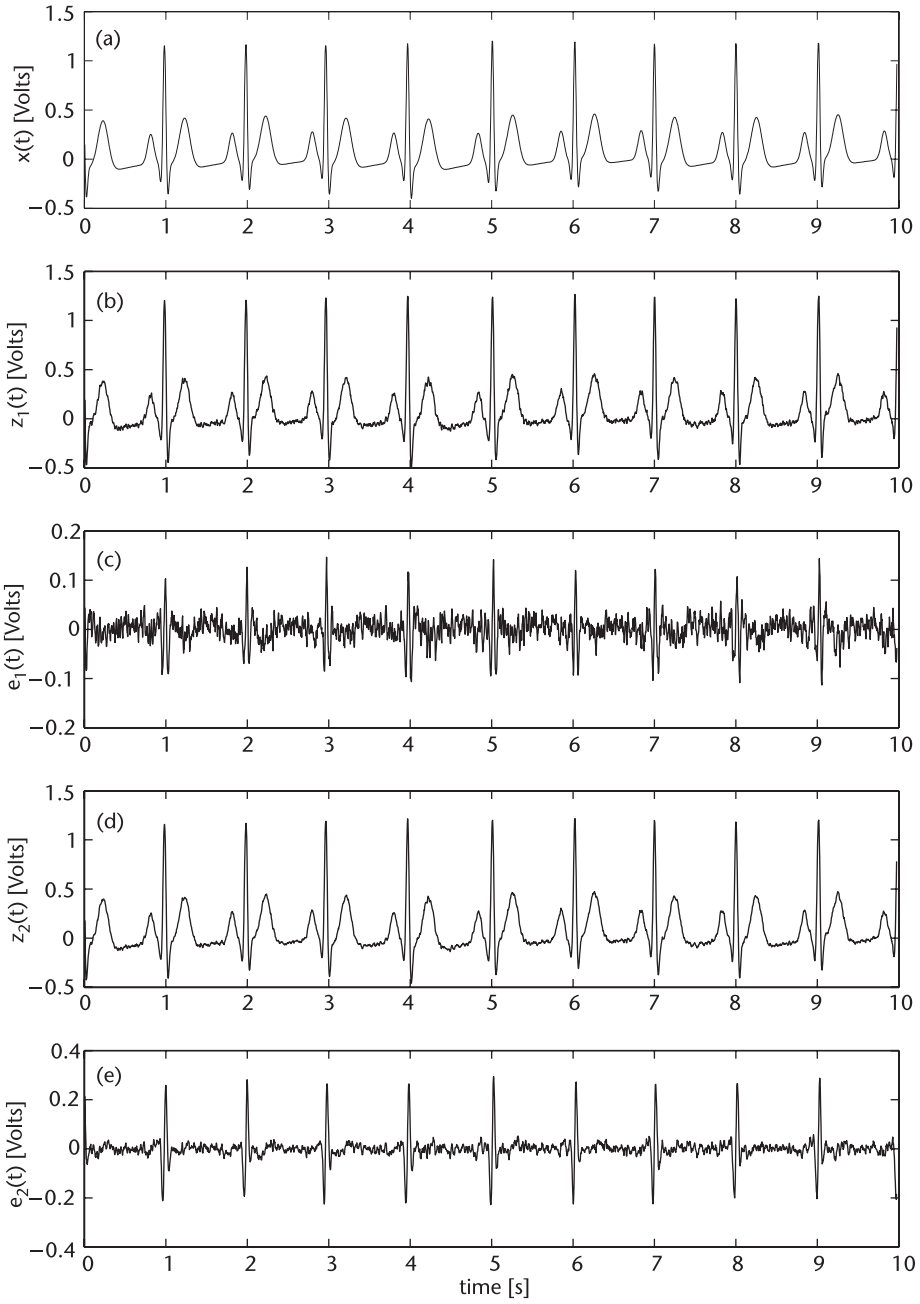

Figure 6.6 Demonstration of ICA noise reduction for $\gamma = 10$: (a) original noise-free ECG signal, $x(t)$; (b) χ-optimized noise-reduced signal, $z_1(t)$, with $m = 7$; (c) error, $e_1(t) = z_1(t) - x(t)$; (d) ρ-optimized noise-reduced signal, $z_2(t)$, with $m = 8$; and (e) error, $e_2(t) = z_2(t) - x(t)$.

separate the signal and noise using the statistics of the data and often rely on a set of assumed heuristics; there is no explicit modeling of any of the underlying sources. If, however, a known model of the signal (or noise) can be built into the filtering scheme, then it is likely that a more effective filter can be constructed.

The simplest model-based filtering is based upon the concept of Wiener filtering, presented in Section 3.1. An extension of this approach is to use a more realistic

model for the dynamics of the ECG signal that can track changes over time. The advantage of such an approach is that once a model has been fitted to a segment of ECG, not only can it produce a filtered version of the waveform, but the parameters can also be used to derive wave onsets and offsets, compress the ECG, or classify beats. Furthermore, the quality of the fit can be used to obtain a confidence measure with respect to the filtering methods.

Existing techniques for filtering and segmenting ECGs are limited by the lack of an explicit patient-specific model to help isolate the required signal from contaminants. Only a vague knowledge of the frequency band of interest and almost no information concerning the morphology of an ECG are generally used. Previously proposed adaptive filters [57, 58] require another reference signal or some ad hoc generic model of the signal as an input.

6.5.1 Nonlinear Model Parameter Estimation

By employing a dynamical model of a realistic ECG, known as ECGSYN (described in detail in Section 4.3.2), a tailor-made approach for filtering ECG signals is now described.

The model parameters that are fit basically constitute a nondynamic version of the model described in [6, 59] and add an extra parameter for each asymmetric wave (only the T wave in the example given here). Each symmetrical feature of the ECG (P,Q,R, and S) is described by three parameters incorporating a Gaussian (amplitude a_i, width b_i) and the phase $\theta_i = 2\pi/t_i$ (or relative position with respect to the R peak). Since the T wave is often asymmetric, it is described by the sum of two Gaussians (and hence requires six parameters) and is denoted by a superscripted $-$ or $+$ to indicate that they are located at values of θ (or t) slightly to either side of the peak of the T wave (the original θ_T that would be used for a symmetric model). The vertical displacement of the ECG, z, from the isoelectric line (at an assumed value of $z = 0$) is then described by an ordinary differential equation,

$$\dot{z}(a_i,b_i,\theta_i) = -\sum_{i\in\{P,Q,R,S,T^-,T^+\}} a_i \Delta\theta_i \exp\left(\frac{-\Delta\theta_i^2}{2b_i^2}\right) \tag{6.25}$$

where $\Delta\theta_i = (\theta - \theta_i) mod(2\pi)$ is the relative phase. Numerical integration of (6.25) using an appropriate set of parameter values, a_i, b_i, and θ_i, leads to the familiar ECG waveform.

One efficient method of fitting the ECG model described above to an observed segment of the signal $s(t)$ is to minimize the squared error between $s(t)$ and $z(t)$. This can be achieved using an 18-dimensional nonlinear gradient descent in the parameter space [60]. Such a procedure has been implemented using two different libraries, the Gnu Scientific Libraries (GSL) in C, and in Matlab using the function *lsqnonlin.m*.

To minimize the search space for fitting the parameters, (a_i, b_i, and θ_i), a simple peak-detection and time-aligned averaging technique is performed to form an average beat morphology using at least the first 60 beats centred on their R peaks. The template window length is unimportant, as long as it contains all the $PQRST$ features and does not extend into the next beat. This method, including outlier

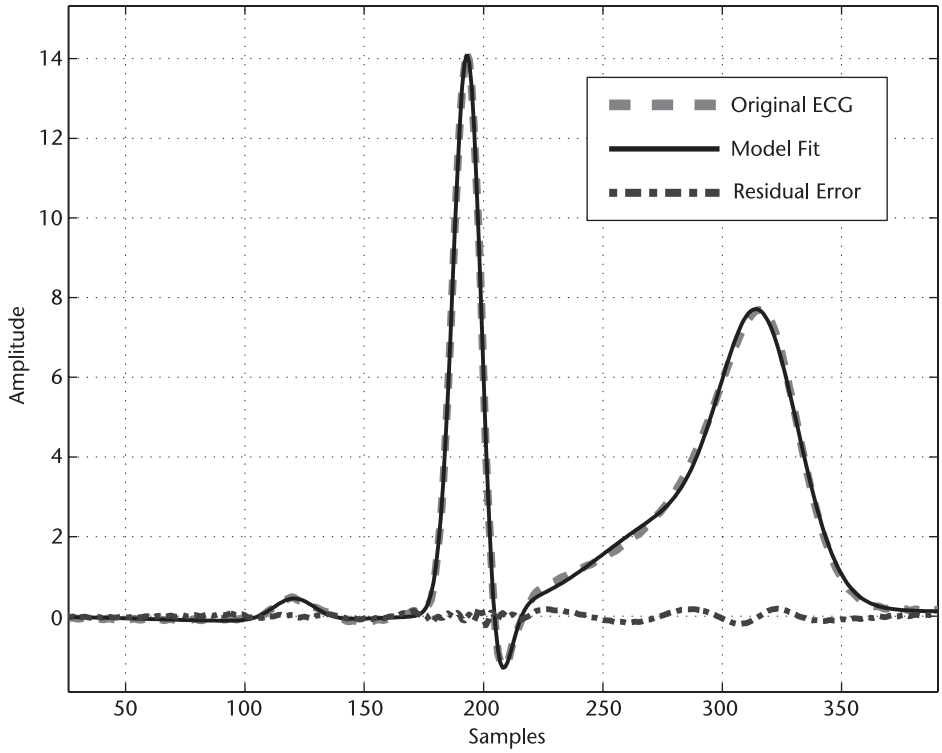

Figure 6.7 Original ECG, nonlinear model fit, and residual error.

rejection, is detailed in [61]. T^- and T^+ are initialized ±40 ms either side of θ_T. By measuring the heights, widths, and positions of each peak (or trough), good initial estimates of the model parameters can be made. Figure 6.7 illustrates an example of a template ECG, the resulting model fit, and the residual error.

Note that it is important that the salient features that one might wish to fit (the P wave and QRS segment in the case of the ECG) are sampled at a high enough frequency to allow them to contribute sufficiently to the optimization. In empirical tests it was found that when $F_s < 450$ Hz, upsampling is required (using an appropriate antialiasing filter). With $F_s < 450$ Hz there are often fewer than 30 sample points in the QRS complex and this can lead to some unrealistic fits that still fulfill the optimization criteria.

One obvious application of this model-fitting procedure is the segmentation of ECG signals and feature location. The model parameters explicitly describe the location, height, and width of each point (θ_i, a_i, and b_i) in the ECG waveform, in terms of a well-known mathematical object, a Gaussian. Therefore, the feature locations and parameters derived from these (such as the P, Q, and T onset and hence the PR and QT interval) are easily extracted. Onsets and offsets are conventionally difficult to locate in the ECG, but using a Gaussian descriptor, it is trivial to locate these points as two or three standard deviations of b_i from the θ_i in question. Similarly, for ECG features that do not explicitly involve the P, Q, R, S, or T points (such as the ST segment), the filtering aspect of this method can be applied.

Furthermore, the error in the fitting procedure can be used to provide a confidence measure for the estimates of any parameters extracted from the ECG signal.

A related application domain for this model-based approach is (lossy) compression with a rate of $(F_s/3k : 1)$ per beat, where $k = n + 2m$ is the number of features or turning points used to fit the heartbeat morphology (with n symmetric and m asymmetric turning points). For a low F_s ($\approx$ 128 Hz), this translates into a compression ratio greater than 7:1 at a heart rate of 60 bpm. However, for high sampling rates ($F_s = 1,024$) this can lead to compression rates of almost 60:1.

Although classification of each beat in terms of the values of a_i, b_i, and θ_i is another obvious application for this model, it is still unclear if the clustering of the parameters is sufficiently tight, given the sympathovagal and heart-rate induced changes typically observed in an ECG. It may be necessary to normalize for heart-rate dependent morphology changes at least. This could be achieved by utilizing the heart rate modulated compression factor α, which was introduced in [59]. However, clustering for beat typing is dependent on population morphology averages for a *specific* lead configuration. Not only would different configurations lead to different clusters in the 18-dimensional parameter space, but small differences in the exact lead placement relative to the heart would cause an offset in the cluster. A method for determining just how far from the standard position the recording is, and a transformation to project back onto the correct position would be required. One possibility could be to use a procedure designed by Moody and Mark [62] for their ECG classifier *Aristotle*. In this approach, the beat clusters are defined in a space resulting from a Karhunen-Loève (KL) decomposition and therefore an estimate of the difference between the classified KL-space and the observed KL-space is made. Classification is then made after transforming from the observation to classification space in which the training was performed. By measuring the distance between the fitted parameters and pretrained clusters in the 18-dimensional parameter space, classification is possible. It should be noted that, as with all classifiers, if an artifact closely resembles a known beat, a good fit to the known beat will obviously arise. For this reason, setting tolerances on the acceptable error magnitude may be crucial and testing on a set of labeled databases is required.

By fitting (6.25) to small segments of the ECG around each QRS-detection fiducial point, an idealistic (zero-noise) representation of each beat's morphology may be derived. This leads to a method for filtering and segmenting the ECG and therefore accurately extracting clinical parameters even with a relatively high degree of noise in the signal. It should be noted that since the model is a compact representation of oscillatory signals with few turning points compared to the sampling frequency and it therefore has a bandpass filtering effect leading to a lossy transformation of the data into a set of integrable Gaussians distributed over time. This approach could therefore be used on any band-limited waveform. Moreover, the error in each fit can provide beat-by-beat confidence levels for any parameters extracted from the ECG and each fit can run in real time (0.1 second per beat on a 3-GHz P4 processor).

The real test of the filtering properties is not the residual error, but how distorted the clinical parameters of the ECG are in each fit. In Section 3.1, an analysis of the sensitivity of clinical parameters to the color of additive noise and the SNR is given together with an independent method for calculating the noise color and SNR. An online estimate of the error in each derived fit can therefore be made. By titrating

colored noise into real ECGs, it has been shown that errors in clinical parameters derived from the model-fit method presented here are clinically insignificant in the presence of high amounts of colored noise. However, clinical features that include low-amplitude features such as the P wave and the ST level are more sensitive to noise power and color. Future research will concentrate on methods to constrain the fit for particular applications where performance is substandard.

An advantage of this method is that it leads to a high degree of compression and may allow classification in the same manner as in the use of KL basis functions (see Chapter 9). Although the KL basis functions offer a similar degree of compression to the Gaussian-based method, the latter approach has the distinct advantage of having a direct clinical interpretation of the basis functions in terms of feature location, width, and amplitude. Using a Gaussian representation, onsets and offsets of waves are easily located in terms of the number of standard deviations of the Gaussian away from the peak of the wave.

6.5.2 State Space Model-Based Filtering

The extended Kalman filter (EKF) is an extension of the traditional Kalman filter that can be applied to a nonlinear model [63, 64]. In the EKF, the full nonlinear model is employed to evolve the states over time while the Kalman filter gain and the covariance matrix are calculated from the linearized equations of motion. Recently, Sameni et al. [65] used an EKF to filter noisy ECG signals using the realistic artificial ECG model, ECGSYN described earlier in Section 2.2. The equations of motion were first transformed into polar coordinates:

$$\begin{aligned} \dot{r} &= r(1-r) \\ \dot{\theta} &= \omega \\ \dot{z} &= -\sum_i a_i \Delta\theta_i \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right) - (z - z_0) \end{aligned} \tag{6.26}$$

Using this representation, the phase, θ, is given as an explicit state variable and r is no longer a function of any of the other parameters and can be discarded. Using a time step of size δt, the two-dimensional equations of motion of the system, with discrete time evolution denoted by n, may be written as

$$\begin{aligned} \theta(n+1) &= \theta(n) + \omega\delta t \\ z(n+1) &= z(n) - \sum_i \delta t a_i \Delta\theta_i \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right) + \eta\delta t \end{aligned} \tag{6.27}$$

where $\Delta\theta_i = (\theta - \theta_i) mod(2\pi)$ and η is a random additive noise. Note that η replaces the previous baseline wander term and describes all the additive sources of process noise.

In order to employ the EKF, the nonlinear equations of motion must first be linearized. Following [65], one approach is to consider θ and z as the underlying state variables and the model parameters, $a_i, b_i, \theta_i, \omega, \eta$ as process noises. Putting all these together gives a process noise vector,

$$\mathbf{w}_n = [a_P, \ldots, a_T, b_P, \ldots, b_T, \theta_P, \ldots, \theta_T, \omega, \eta]^\dagger \tag{6.28}$$

with covariance matrix $\mathcal{Q}_n = E\{\mathbf{w}_n \mathbf{w}_n^\dagger\}$ where the subscript † denotes the transpose.

The phase of the observations ψ_n, and the noisy ECG measurements s_n are related to the state vector by

$$\begin{bmatrix} \psi_n \\ s_n \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_n \\ z_n \end{bmatrix} + \begin{bmatrix} \nu_n^{(1)} \\ \nu_n^{(2)} \end{bmatrix} \tag{6.29}$$

where $\boldsymbol{\nu}_n = [\nu_n^{(1)}, \nu_n^{(2)}]^\dagger$ is the vector of measurement noises with covariance matrix $\mathbf{R}_n = E\{\boldsymbol{\nu}_n \boldsymbol{\nu}_n^\dagger\}$.

The variance of the observation noise in (6.29) represents the degree of uncertainty associated with a single observation. When $\mathbf{R}_n$ is high, the EKF tends to ignore the observation and rely on the underlying model dynamics for its output. When $\mathbf{R}_n$ is low, the EKF's gain adapts to incorporate the current observations. Since the 17 noise parameters in (6.28) are assumed to be independent, $\mathcal{Q}_k$ and $\mathbf{R}_n$ are diagonal. The process noise η is a measure of the accuracy of the model, and is assumed to be a zero-mean Gaussian noise process.

Using this EKF formulation, Sameni et al. [65] successfully filtered a series of ECG signals with additive Gaussian noise. An example of this can be seen in Figure 6.8. Future developments of this model are therefore very promising, since the

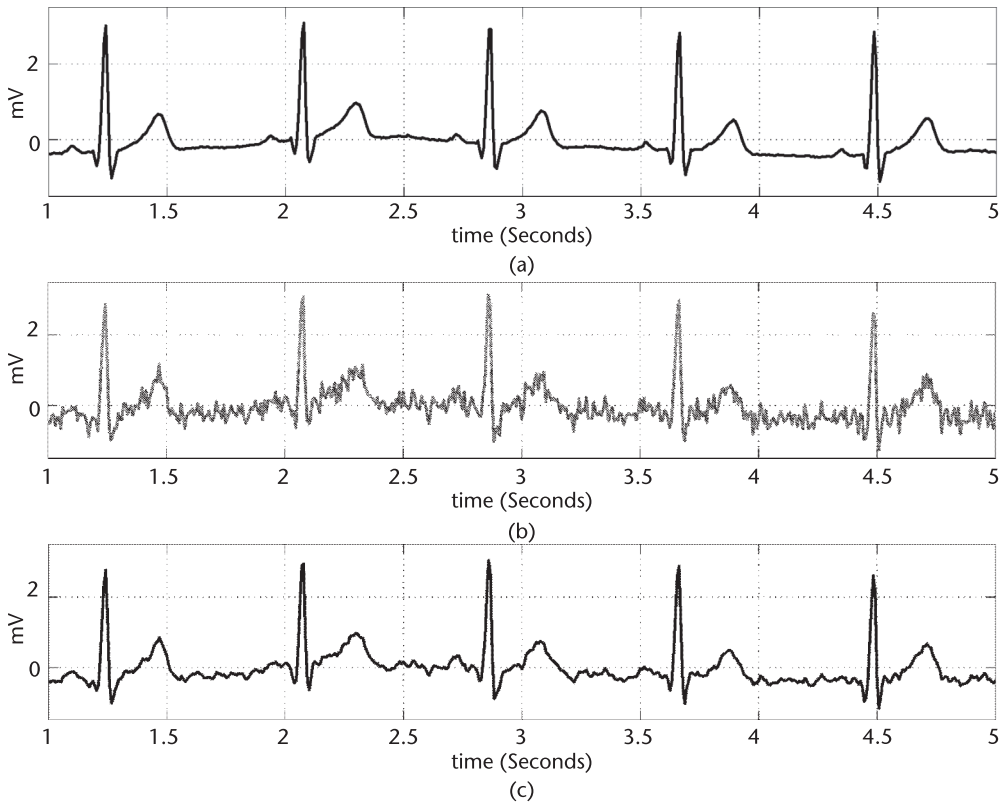

Figure 6.8 Filtering of noisy ECG using EKF: (a) original signal; (b) noisy signal; and (c) denoised signal.

combination of a realistic model and a robust tracking mechanism make the concept of online signal tracking in real time feasible. A combination of an initialization with the nonlinear gradient descent method from Section 6.5.1 to determine initial model parameters and noise estimates, together with subsequent online tracking, may lead to an optimal ECG filter (for normal morphologies). Furthermore, the ability to relate the parameters of the model to each PQRST morphology may lead to fast and accurate online segmentation procedures.

6.6 Conclusion

This chapter has provided a summary of the mathematics involved in reconstructing a state space using a recorded signal so as to apply techniques based on the theory of nonlinear dynamics. Within this framework, nonlinear statistics such as Lyapunov exponents, correlation dimension, and entropy were described. The importance of comparing results with simple benchmarks, carrying out statistical tests and using confidence intervals when conveying estimates was also discussed. This is important when employing nonlinear dynamics as the basis of any new biomedical diagnostic tool.

An artificial electrocardiogram signal, ECGSYN, with controlled temporal and spectral characteristics was employed to illustrate and compare the noise reduction performance of two techniques, nonlinear noise reduction and independent components analysis. Stochastic noise was used to create data sets with different signal-to-noise ratios. The accuracy of the two techniques for removing noise from the ECG signals was compared as a function of signal-to-noise ratio. The quality of the noise removal was evaluated by two techniques: (1) a noise reduction factor and (2) a measure of the correlation between the cleaned signal and the original noise-free signal. NNR was found to give better results when measured by correlation. In contrast, ICA outperformed NNR when compared using the noise reduction factor. These results suggest that NNR is superior at recovering the morphology of the ECG and is less likely to distort the shape of the P, QRS, and T waves, whereas ICA is better at recovering specific points on the ECG such as the R peak, which is necessary for obtaining RR intervals.

Two model-based filtering approaches were also introduced. These methods use the dynamical model underlying ECGSYN to provide constraints on the filtered signal. A nonlinear least squares parameter estimation procedure was used to estimate all 18 parameters required to specify the morphology of the ECG waveform. In addition, an approach using the extended Kalman filter applied to a discrete two-dimensional adaptation of ECGSYN in polar coordinates was also employed to filter an ECG signal.

The correct choice of filtering technique depends not only on the characteristics of the noise and signal in the time and frequency domains, but also on the application. It is important to test a candidate filtering technique over a range of possible signals (with a range of signal to noise ratios and different noise processes) to determine the filter's effect on the clinical parameter or attribute of interest.

References

[1] Papoulis, P., *Probability, Random Variables, and Stochastic Processes*, 3rd ed., New York: McGraw-Hill, 1991.

[2] Schreiber, T., and D. T. Kaplan, "Nonlinear Noise Reduction for Electrocardiograms," *Chaos*, Vol. 6, No. 1, 1996, pp. 87–92.

[3] Cardoso, J. F., "Multidimensional Independent Component Analysis," *Proc. ICASSP'98*, Seattle, WA, 1998.

[4] Clifford, G. D., and P. E. McSharry, "Method to Filter ECGs and Evaluate Clinical Parameter Distortion Using Realistic ECG Model Parameter Fitting," *Computers in Cardiology*, September 2005.

[5] Moody, G. B., and R. G. Mark, "Physiobank: Physiologic Signal Archives for Biomedical Research," MIT, Cambridge, MA, http://www.physionet.org/physiobank, updated June 2006.

[6] McSharry, P. E., et al., "A Dynamical Model for Generating Synthetic Electrocardiogram Signals," *IEEE Trans. Biomed. Eng.*, Vol. 50, No. 3, 2003, pp. 289–294.

[7] Kantz, H., and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge, U.K.: Cambridge University Press, 1997.

[8] McSharry, P. E., and L. A. Smith, "Better Nonlinear Models from Noisy Data: Attractors with Maximum Likelihood," *Phys. Rev. Lett.*, Vol. 83, No. 21, 1999, pp. 4285–4288.

[9] McSharry, P. E., and L. A. Smith, "Consistent Nonlinear Dynamics: Identifying Model Inadequacy," *Physica D*, Vol. 192, 2004, pp. 1–22.

[10] Packard, N., et al., "Geometry from a Time Series," *Phys. Rev. Lett.*, Vol. 45, 1980, pp. 712–716.

[11] Takens, F., "Detecting Strange Attractors in Fluid Turbulence," in D. Rand and L. S. Young, (eds.), *Dynamical Systems and Turbulence*, New York: Springer-Verlag, Vol. 898, 1981, p. 366.

[12] Broomhead, D. S., and G. P. King, "Extracting Qualitative Dynamics from Experimental Data," *Physica D*, Vol. 20, 1986, pp. 217–236.

[13] Sauer, T., J. A. Yorke, and M. Casdagli, "Embedology," *J. Stats. Phys.*, Vol. 65, 1991, pp. 579–616.

[14] Fraser, A. M., and H. L. Swinney, "Independent Coordinates for Strange Attractors from Mutual Information," *Phys. Rev. A*, Vol. 33, 1986, pp. 1134–1140.

[15] Rosenstein, M. T., J. J. Collins, and C. J. De Luca, "Reconstruction Expansion as a Geometry-Based Framework for Choosing Proper Time Delays," *Physica D*, Vol. 73, 1994, pp. 82–98.

[16] Kennel, M. B., R. Brown, and H. D. I. Abarbanel, "Determining Embedding Dimension for the Phase-Space Reconstruction Using a Geometrical Construction," *Phys. Rev. A*, Vol. 45, No. 6, 1992, pp. 3403–3411.

[17] Strang, G., *Linear Algebra and Its Applications*, San Diego, CA: Harcourt College Publishers, 1988.

[18] Lorenz, E. N., "A Study of the Predictability of a 28-Variable Atmospheric Model," *Tellus*, Vol. 17, 1965, pp. 321–333.

[19] Abarbanel, H. D. I., R. Brown, and M. B. Kennel, "Local Lyapunov Exponents Computed from Observed Data," *Journal of Nonlinear Science*, Vol. 2, No. 3, 1992, pp. 343–365.

[20] Glass, L., and M. Mackey, *From Clocks to Chaos: The Rhythms of Life*, Princeton, NJ: Princeton University Press, 1988.

[21] Wolf, A., et al., "Determining Lyapunov Exponents from a Time Series," *Physica D*, Vol. 16, 1985, pp. 285–317.

[22] Sano, M., and Y. Sawada, "Measurement of the Lyapunov Spectrum from a Chaotic Time Series," *Phys. Rev. Lett.*, Vol. 55, 1985, pp. 1082–1085.

[23] Eckmann, J. P., et al., "Liapunov Exponents from Time Series," *Phys. Rev. A*, Vol. 34, No. 6, 1986, pp. 4971–4979.

[24] Brown, R., P. Bryant, and H. D. I. Abarbanel, "Computing the Lyapunov Spectrum of a Dynamical System from an Observed Time Series," *Phys. Rev. A*, Vol. 43, No. 6, 1991, pp. 2787–2806.

[25] McSharry, P. E., "The Danger of Wishing for Chaos," *Nonlinear Dynamics, Psychology, and Life Sciences*, Vol. 9, No. 4, October 2005, pp. 375–398.

[26] Rosenstein, M.T., J. J. Collins, and C.J. De Luca, "A Practical Method for Calculating Largest Lyapunov Exponents from Small Data Sets," *Physica D*, Vol. 65, 1994, pp. 117–134.

[27] Kantz, H., "Quantifying the Closeness of Fractal Measures," *Phys. Rev. E*, Vol. 49, No. 6, 1994, pp. 5091–5097.

[28] Hegger, R., H. Kantz, and T. Schreiber, "TISEAN: Nonlinear Time Analysis Software," Max-Planck-Institut für Physik Komplexer System, http://www.mpipks-dresden.mpg.de/~tisean/, December 2005.

[29] Renyi, A., *Probability Theory*, Amsterdam: North Holland, 1971.

[30] Ott, E., T. Sauer, and J. A. Yorke, *Coping with Chaos*, New York: John Wiley & Sons, 1994.

[31] Grassberger, P., and I. Procaccia, "Characterization of Strange Attractors," *Phys. Rev. Lett.*, Vol. 50, 1983, pp. 346–349

[32] Theiler, J. T., "Spurious Dimension from Correlation Algorithms Applied to Limited Time-Series Data," *Phys. Rev. A*, Vol. 34, 1986, pp. 2427–2432.

[33] Provenzale, A., et al., "Distinguishing Between Low-Dimensional Dynamics and Randomness in Measured Time Series," *Physica D*, Vol. 58, 1992, pp. 31–49.

[34] Takens, F., "On the Numerical Determination of the Dimension of an Attractor," in B. L. J. Braaksma, H. W. Broer, and F. Takens, (eds.), *Dynamical Systems and Bifurcations*, Volume 1125 of *Lecture Notes in Mathematics*, Berlin, Germany: Springer, 1985, pp. 99–106.

[35] Guerrero, A., and L. A. Smith, "Towards Coherent Estimation of Correlation Dimension," *Phys. Lett. A*, Vol. 318, 2003, pp. 373–379.

[36] Ruelle, D., "Deterministic Chaos: The Science and the Fiction," *Proc. R. Soc. Lond. A.*, Vol. 427, 1990, pp. 241–248.

[37] Smith, L. A., "Intrinsic Limits on Dimension Calculations," *Phys. Lett. A*, Vol. 133, 1988, pp. 283–288.

[38] Eckmann, J. P., and D. Ruelle, "Ergodic Theory of Chaos and Strange Attractors," *Rev. Mod. Phys.*, Vol. 57, 1985, pp. 617–656.

[39] Pincus, S. M., "Approximate Entropy as a Measure of System Complexity," *Proc. Natl. Acad. Sci.*, Vol. 88, 1991, pp. 2297–2301.

[40] Richman, J. S., and J. R. Moorman, "Physiological Time Series Analysis Using Approximate Entropy and Sample Entropy," *Am. J. Physiol.*, Vol. 278, No. 6, 2000, pp. H2039–H2049.

[41] McSharry, P. E., L. A. Smith, and L. Tarassenko, "Prediction of Epileptic Seizures: Are Nonlinear Methods Relevant?" *Nature Medicine*, Vol. 9, No. 3, 2003, pp. 241–242.

[42] McSharry, P. E., L. A Smith, and L. Tarassenko, "Comparison of Predictability of Epileptic Seizures by a Linear and a Nonlinear Method," *IEEE Trans. Biomed. Eng.*, Vol. 50, No. 5, 2003, pp. 628–633.

[43] Isliker, H., and J. Kurths, "A Test for Stationarity: Finding Parts in Time Series APT for Correlation Dimension Estimates," *Int. J. Bifurcation Chaos*, Vol. 3, 1993, pp. 1573–1579.

[44] Braun, C., et al., "Demonstration of Nonlinear Components in Heart Rate Variability of Healthy Persons," *Am. J. Physiol. Heart Circ. Physiol.*, Vol. 275, 1998, pp. H1577–H1584.

[45] Lehnertz, K., and C. E. Elger, "Can Epileptic Seizures Be Predicted? Evidence from Nonlinear Time Series Analysis of Brain Electrical Activity," *Phys. Rev. Lett.*, Vol. 80, No. 22, 1998, pp. 5019–5022.

[46] Iasemidis, L. D., et al., "Adaptive Epileptic Seizure Prediction System," *IEEE Trans. Biomed. Eng.*, Vol. 50, No. 5, 2003, pp. 616–627.

[47] Pincus, S. M., and A. L. Goldberger, "Physiological Time-Series Analysis: What Does Regularity Quantify?" *Am. J. Physio. Heart Circ. Physiol.*, Vol. 266, 1994, pp. H1643–H1656.

[48] Costa, M., A. L. Goldberger, and C. K. Peng, "Multiscale Entropy Analysis of Complex Physiologic Time Series," *Phys. Rev. Lett.*, Vol. 89, No. 6, 2002, p. 068102.

[49] Voss, A., et al., "The Application of Methods of Non-Linear Dynamics for the Improved and Predictive Recognition of Patients Threatened by Sudden Cardiac Death," *Cardiovascular Research*, Vol. 31, No. 3, 1996, pp. 419–433.

[50] Wessel, N., et al., "Short-Term Forecasting of Life-Threatening Cardiac Arrhythmias Based on Symolic Dynamics and Finite-Time Growth Rates," *Phys. Rev. E*, Vol. 61, 2000, pp. 733–739.

[51] Chatfield, C., *The Analysis of Time Series*, 4th ed., London, U.K.: Chapman and Hall, 1989.

[52] Kobayashi, M., and T. Musha, "1/f Fluctuation of Heartbeat Period," *IEEE Trans. Biomed. Eng.*, Vol. 29, 1982, p. 456.

[53] Goldberger, A. L., et al., "Fractal Dynamics in Physiology: Alterations with Disease and Ageing," *Proc. Natl. Acad. Sci.*, Vol. 99, 2002, pp. 2466–2472.

[54] Schreiber, T., and P. Grassberger, "A Simple Noise Reduction Method for Real Data," *Phys. Lett. A.*, Vol. 160, 1991, p. 411.

[55] Hegger, R., H. Kantz, and T. Schreiber, "Practical Implementation of Nonlinear Time Series Methods: The Tisean Package," *Chaos*, Vol. 9, No. 2, 1999, pp. 413–435.

[56] James, C. J., and D. Lowe, "Extracting Multisource Brain Activity from a Single Electromagnetic Channel," *Artificial Intelligence in Medicine*, Vol. 28, No. 1, 2003, pp. 89–104.

[57] "Filtering for Removal of Artifacts," Chapter 3 in R. M. Rangayyan, *Biomedical Signal Analysis: A Case-Study Approach*, New York: IEEE Press, 2002, pp. 137–176.

[58] Barros, A., A. Mansour, and N. Ohnishi, "Removing Artifacts from ECG Signals Using Independent Components Analysis," *Neurocomputing*, Vol. 22, No. 1–3, 1998, pp. 173–186.

[59] Clifford, G. D., and P. E. McSharry, "A Realistic Coupled Nonlinear Artificial ECG, BP and Respiratory Signal Generator for Assessing Noise Performance of Biomedical Signal Processing Algorithms," *Proc. of Intl. Symp. on Fluctuations and Noise 2004*, Vol. 5467-34, May 2004, pp. 290–301.

[60] More, J. J., "The Levenberg-Marquardt Algorithm: Implementation and Theory," *Lecture Notes in Mathematics*, Vol. 630, 1978, pp. 105–116.

[61] Clifford, G., L. Tarassenko, and N. Townsend, "Fusing Conventional ECG QRS Detection Algorithms with an Auto-Associative Neural Network for the Detection of Ectopic Beats," *Proc. of 5th International Conference on Signal Processing, 16th IFIP WorldComputer Congress*, Vol. III, 2000, pp. 1623–1628.

[62] Moody, G. B., and R. G. Mark, "QRS Morphology Representation and Noise Estimation Using the Karhunen-Loève Transform," *Computers in Cardiology*, Vol. 16, 1989, pp. 269–272.

[63] Kay, S. M., *Fundamentals of Statistical Signal Processing: Estimation Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[64] Grewal, M. S., and A. P. Andrews, *Kalman Filtering: Theory and Practice Using Matlab*, 2nd ed., New York: John Wiley & Sons, 2001.

[65] Sameni, R., et al., "Filtering Noisy ECG Signals Using the Extended Kalman Filter Based on a Modified Dynamic ECG Model," *Computers in Cardiology*, Vol. 32, 2005, pp. 1017–1020.