International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:4, No:1, 2010

Computing Fractal Dimension of Signals using Multiresolution Box-counting Method

B. S. Raghavendra, and D. Narayana Dutt

Abstract**—**In this paper, we have developed a method to compute fractal dimension (FD) of discrete time signals, in the time domain, by modifying the box-counting method. The size of the box is dependent on the sampling frequency of the signal. The number of boxes required to completely cover the signal are obtained at multiple time resolutions. The time resolutions are made coarse by decimating the signal. The log-log plot of total number of boxes required to cover the curve versus size of the box used appears to be a straight line, whose slope is taken as an estimate of FD of the signal. The results are provided to demonstrate the performance of the proposed method using parametric fractal signals. The estimation accuracy of the method is compared with that of Katz, Sevcik, and Higuchi methods. In addition, some properties of the FD are discussed.

Keywords**—**Box-counting, Fractal dimension, Higuchi method, Katz method, Parametric fractal signals, Sevcik method.

I. Introduction

FRACTAL dimension (FD) is a useful concept in describing natural objects, which gives their degree of complexity [1], [2]. There are various closely related notions of fractional dimension. From the theoretical point of view, the most important are the Hausdorff dimension, the packing dimension and, more generally, the Rényi dimensions. On the other hand, the box counting dimension and correlation dimension are widely used in practice, may be due to their ease of implementation. The term FD generally refers to any of the dimension used for fractal characterization. This includes capacity dimension, correlation dimension, information dimension, Lyapunov dimension and Minkowski-Bouligand dimension [3]. However, in fractal geometry, the FD is a statistical quantity that gives an indication of how completely a fractal appears to fill the space, as one zooms down to finer and finer scales, accordingly there are many specific definitions of fractal dimension.

The FD is a measure of how complicated a self-similar figure is. Hence the FD can be considered as a relative measure of number of basic building blocks that form a pattern [4]. According to Mandelbrot [1], a fractal is a set for which the Hausdorff-Besicovitch dimension (D_h) strictly exceeds the topological dimension. Hence, every set with a non integer dimension D is a fractal. The Hausdorff dimension (also known as the Hausdorff-Besicovitch dimension) is a non-negative real number associated to any metric space. To define the Hausdorff dimension for a set X as non-negative real number (that is a number in the half-closed infinite interval $[0,\infty)$), we first consider the number $N(r)$ of balls of radius at most r required to cover X completely. Clearly, as r gets smaller $N(r)$ gets larger. Very roughly, if $N(r)$ grows in the same way as $1/r^D$ as r is squeezed down towards zero, then we say X has dimension D. In fact the rigorous definition of Hausdorff dimension is somewhat roundabout, since it first defines an entire family of covering measures for X. It turns out that Hausdorff dimension refines the concept of topological dimension and also relates it to other properties of the space such as area or volume.

The fractal dimension measures, described above, are derived from fractals which are formally (mathematically) defined. However, many real-world phenomena exhibit fractal properties. So it can often be useful to characterize the fractal dimension of a set of sampled data. The fractal dimension measures of time series cannot be derived exactly but must be estimated. Practical dimension estimates are very sensitive to numerical or experimental noise, and particularly sensitive to limitations on the amount of data.

The FD estimation algorithms give a number regardless of whether or not the object is fractal. It is also possible to have two different fractal sets having the same dimension. In addition, a fractal property can be spatial, it can be temporal, as in a series of data taken from a system over an interval of time, and it can be exact or statistical. Hence, the FD is applicable to sets that may not be self similar over all ranges of space or time. Furthermore, it is still possible and useful to apply the general idea to a natural system and define its FD. However, no physical object is truly a fractal because it does not have self-similar properties at all scales. This leads to the fact that fractal dimension analysis does not differentiate between fractal and non-fractal objects, but rather gives a measure of the appropriateness of describing the object using fractal models.

B. S. Raghavendra and D. Narayana Dutt are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore – 560 012, India (corresponding author phone: +91 80 2293 2742; fax: +91 80 2360 0563; e-mail: r.bobbi@ece.iisc.ernet.in).

Thus, any planar curve (waveform) with $1 < D_h < 2$ is a fractal. The FD is an important characteristic of signals and contains information about their structural complexity. In the

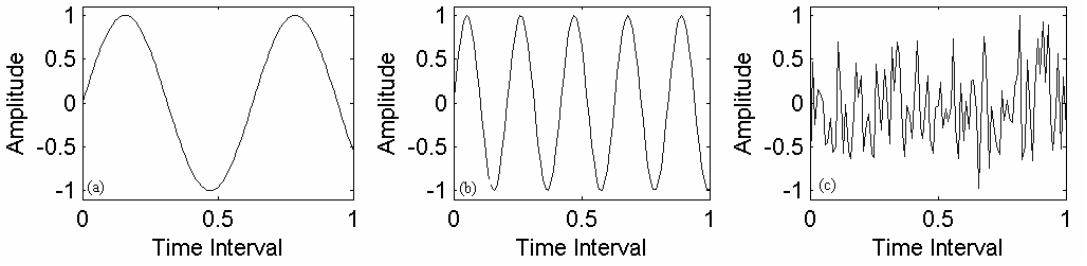

Fig. 1. Comparison of smooth and irregular waveforms, (a) sinusoidal signal of 10 Hz, (b) sinusoidal signal of 30 Hz, (c) random signal.

field of signal processing, the fractal models have proven useful for many applications. There are numerous signals such as speech [5], fractional Brownian motion (fBm), physiological signals [6]-[8], etc., with fractal properties such that their graph is a fractal set. Consequently the FD could reflect the signal complexity in the time domain. This complexity could vary with sudden occurrence of transient in signals. We would like to measure the complexity of signal waveforms by estimating their FD.

Consider the graph of functions as shown in the Figure 1, smooth sinusoidal curve of frequency 10 Hz, 20 Hz and a highly irregular random curve. Upon embedding these curves into a plane, it is evident that the irregular curve fills a larger region of the plane than the smooth one. The three curves have signal lengths of 6.6679, 18.9890 and 50.6929 respectively, and zero crossing rates of 4, 10 and 50 respectively. One definition of FD used in practice is a measure of this space filling property. Note that nothing has been stated explicitly regarding the self similarity of the irregular curve. Fractal dimension analysis is done because it gives a measure of the appropriateness of describing the structural complexity of objects.

FD estimators will give a number regardless of whether or not the object is fractal. The conceptual description of fractal dimension as a measure of an object's space filling property, establishes a basis for developing algorithms to estimate FD from experimental data. The FDs range from 1 to 2 for planar curves. To investigate the fractal structure experimentally, it is necessary to be able to relate the results of observation to fractal measures, such as dimension.

A very popular approach to obtain FD of signals is the box-counting method [9]. However, for signals the FD obtained by using box counting method is highly sensitive to the sampling frequency, and some times lead to over or under determination of the FD. In addition, most of the time series waveforms exist in the affine space where the axes have incompatible units, and there is no natural scaling between them. This means that distance along the time axis cannot be compared with distance along the amplitude axis.

The box counting method appears more suitable for determining the FD of self similar mass fractals [9], and less suited for measuring FD of self affine boundary fractals such as time series waveforms. For self affine boundary fractals, the measuring unit in determining their FD ought to be a straight line. However, measuring a profile by using a line with a varying resolution is computationally inefficient. Some times the FD of waveforms computed using box counting method is more than 2, which is in conflict with the definition of fractals in two dimensional spaces. These limitations necessitated definition of a new algorithm which is not only conceptually valid but also has a lower time complexity than the box counting method.

Many studies have been carried out to investigate the reliability of FD estimation with different algorithms applied to different FDs [10]-[13]. Numerous issues like quantization, number of data points, sampling methods, and role of noise have been addressed to help explain the existence of errors. Fractal complexity of signals in time domain is calculated using Katz's and Sevcik's methods. In time domain the method seems to be simple and may be used in many applications. The computation is quicker and simple to be done in real time.

The FD calculated this way is a measure of complexity of the curve representing the signal in a plane. Here the complexity refers to the degree of space filling of the signal in the 2D plane. The complexity of a signal may be characterized by its FD directly in time domain. Generally, signal complexity can be analyzed in time domain, frequency domain, or in the phase space of the system which generated the signal. Analysis in the frequency domain requires Fourier or wavelet transform of the signal, while analysis in the phase space requires embedding of the data in a higher dimensional space. However, the FD is a descriptive quantitative measure, a single number that quantifies complexity of a signal. The estimation of FD adopted here is derived from an operation directly on the signal and not on any phase space. This means that the data series does not have to be embedded into higher dimensional space for the FD estimation.

For signals, FD range between one and two. True waveforms can never become sufficiently convoluted to fill a plane. Thus the waveforms will never have FDs approximating the dimensionality of a plane ($D = 2.0$). The fractal dimensions of waveforms are a powerful tool for detection of transients in signals. FD analysis is frequently

used in biomedical signal processing applications including EEG data analysis. In particular, in the analysis of EEG, this feature has been used to identify and distinguish specific states of physiological function.

The FD and its variants are popular measures for characterizing complexity of signals in various fields [5], [14], [6]. In biomedical signal analysis, the FD is used as a quantitative measure to estimate complexity of discrete time physiological signals [7], [8], [12]. Such analysis of complexity of biomedical signals helps us to study physiological processes underlying the systems. The FD can be used to study dynamics of transitions between different states of systems like brain, as also in various physiological and pathological conditions [10], [14], [7]. More details on the general notion of FD and various ways to estimate FD of signals are discussed elsewhere [9], [16], [17].

There are various closely related notions of fractal dimension, and many algorithms have been proposed in the literature to estimate the FD of signals or time series data [17], [18]. It is proposed that the Higuchi's method of computation of FD is the robust and gives accurate estimation results [10], [11]. This method is also suitable for estimating FD of short segment of a time series, and hence it can be used for computing moving window estimates of FD for nonstationary signals, by segmenting them into short stationary frames.

Despite its popularity, issues of interpretation of the FD measure computed from signals and its relationship to their parameters have not been thoroughly addressed. The effect of various signal parameters such as amplitude, frequency, number of harmonics, noise power, signal bandwidth, etc., on its FD has not been addressed so far. For a particular class of signals, called $1/f$ process, where the power spectrum of the process follows a power law, that is $S(f) \approx c/|f|^{\gamma}$, where $S(f)$ is the power spectrum, c is a constant, f is frequency and γ is the power spectrum exponent, there exists a linear relationship between the power spectrum exponent and FD of the process, given by $FD = (5-\gamma)/2$ as described in [19]. However, the real world processes do not strictly follow the power law behavior and thus distribution of power over the frequencies may not follow the strict $1/f$ rule. The power may be concentrated over some specific frequencies. In such cases one has to find the relationship between the power spectrum of the signal and its FD numerically.

In this chapter, we deal with the problem of estimating FDs of topographically one dimensional signal waveforms, and we propose a new method, refer it as multiresolution box-counting method (MRBC), to estimate fractal dimension of signal waveforms. A little modification of this method results in another method; we refer it as multiresolution length method (MRL), which is also used to estimate FDs of signals. We test estimation accuracy of the proposed methods using parametric fractal signals such as, Weierstrass cosine function (WCF), Weierstrass-Mandelbrot cosine function (WMCF), Knopp function (KF), and fractional Brownian motion (fBm) signals, and also compare the estimation performance with that of Katz, Sevcik, and Higuchi methods. We show that our method performs comparable to Higuchi method but computationally less time consuming than the Higuchi method. In addition, we also study the issue of interpretation of the FD measure computed from signals and its relationship to the parameters such as amplitude, frequency, and noise power.

II. Methods

A. Box-counting method

There are many notions of FD and many algorithms are available to calculate them for topologically one dimensional curves [9], [16], the box-counting dimension is one among them. The box-counting dimension is motivated by the notion of determining space filling properties of a curve. In this approach, the curve is covered with a collection of area elements (square boxes), and the number of elements of a given size is counted to see how many of them are necessary to cover the curve completely. As the size of the area element approaches zero, the total area covered by the area elements will converge to the measure of the curve. This can be expressed mathematically as

$$D_B = \lim_{r \to 0} (\log N(r) / \log(1/r)),$$

where $N(r)$ is the total number of boxes of size r required to cover the curve entirely. However in practice, the box-counting algorithm estimates FD of the curve by counting the number of boxes required to cover the curve for several box sizes, and fitting a straight line to the log-log plot of $N(r)$ versus r. That is

$$\log N(r) = D_B \log(1/r) + C,$$

where C is a constant. The slope of the least square best fit straight line is taken as an estimate of the box-counting dimension D_B of the curve. This procedure is also called grid method and involves two dimensional processing of the curve at multiple grid sizes, which is computationally highly time consuming. In order to avid this drawback, we propose a new method of computation of signal waveforms by computing box areas at multiple time resolutions.

B. Multiresolution Box-counting Method

In this section, we propose a method to compute fractal dimension of waveforms. The proposed approach is described as follows. Consider a discrete time signal $S = \{s(1), s(2), ..., s(N)\}$ of sampling frequency f_s and having N number of sample points. Each of the sample points $s(i)$ in the sequence is represented as $(x(i), y(i))$, $i = 1, 2, ..., N$. The $x(i)$ are the abscissa, representing the monotonically increasing time at which the signal is sampled, and $y(i)$ are the ordinate values. Here, we have assumed that the discrete

time signal is sufficiently highly sampled with a rate of $1/f_s$, at least two times the Nyquist rate. At this sampling rate, the sample values represent the signal at the finest time resolution $r = 1/f_s$. Then the following calculations are made.

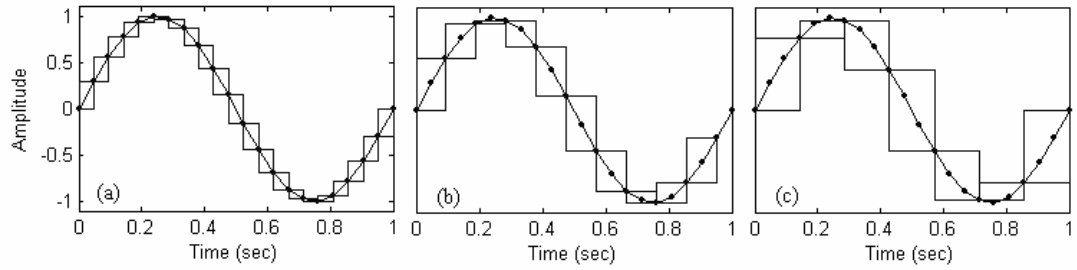

Fig. 2. Multiresolution box-counting approach for sinusoidal signal, (a) at the finest time resolution, and (b) and (c) at the next two coarse time resolution.

Step (1):

Consider the two points $s(i)$ and $s(i+1)$ on the curve representing the signal. The time interval between the points is $dt = x(i+1) - x(i) = 1/f_s$. The height between the points is $h = y(i+1) - y(i)$. The size of the box considered to cover the two points is dt, and the number of boxes of that size required to cover the points is $b(i) = \lceil |h|/dt \rceil$, where $\lceil a \rceil$ represents $ceil(a)$, the highest integer near to a. Then the value of $y(i+1)$ is updated as follows. If $h > 0$, then $y(i+1) = y(i) + |h| - dt$, and if $h < 0$, then $y(i+1) = y(i) - |h| + dt$. The procedure is repeated for all the points on the curve until the end point is reached. The total number of boxes required to cover the curve at the resolution r is calculated as $B(r) = sum(b(i))$, $i = 1, 2, ..., N-1$. This procedure is depicted in Figure 2(a), for a sinusoidal curve.

Step (2):

Now, consider the curve at the next coarse time resolution, by decimating the signal by a factor of two. That means, we leave every alternate points on the curve to get a time resolution $r = 2/f_s$. Now, the size of the box considered to cover the curve is $dt = 2/f_s$. The same procedure described in the step (1) is repeated at this time resolution and the total number of boxes required to cover the entire curve is calculated. The Figure 2(b) explains this step.

Step (3):

By repeating the above steps for many time resolutions, we get the number of boxes $B(r)$ to cover the curve, for $r = 1/f_s, 2/f_s, ..., R/f_s$, where R/f_s is the maximum coarse time resolution at which the curve is looked at.

Step (4):

The least-square linear best fitting procedure is applied to the graph $(r, B(r))$. The coefficient of linear regression of the plot of $\log(B(r))$ versus $\log(1/r)$ is taken as an estimate of the fractal dimension FD of the discrete time signal, and denoted as De.

In Figure 3, the plot of $\log(B(r))$ versus $\log(1/r)$, and least square best fit straight line to the graph $(1/r, B(r))$ is shown for three fractal signals of different FDs. The fractal signals used here are explained in detail in the next section.

Since the total number of boxes to cover the curve is calculated at multiple time resolutions, we refer this approach to as multiresolution box counting (MRBC) method. The sizes of the box considered are the time resolutions at which the curve/signal is looked at. A variant of this method is also proposed which is discussed below.

C. Multiresolution Length Method

The approach that we propose is described as follows. Consider a time series $S = \{s(1), s(2), ..., s(N)\}$ of length N. Each point $s(i)$ in the sequence S is represented as (x_i, y_i), $i = 1, 2, ..., N$. The x_i are abscissa and y_i are ordinate values. $x_i \in [0, t]$. If the points $s(1)$ and $s(2)$ are represented as (x_1, y_1) and (x_2, y_2) respectively, the Euclidean distance between them can be calculated as

$$dist(s1, s2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

We have assumed that the observed time series is sufficiently sampled with a high sampling rate. This time series is considered as a geometric object (curve) and further calculations are made on the object.

The curve S is a time series looked at the finest time resolution say r_1. The total length of the curve at this resolution is calculated as

$$L = \sum_{i=1}^{N-1} dist(s_i, s_{i+1}).$$

This is the length L_{r1} at resolution $r_1 = 1/f_s$, where f_s is the sampling rate of the time series. Consider the time series at next coarse resolution by eliminating every alternate point (decimation by a factor 2). Now the resolution becomes $r_2 = 2/f_s$. Calculate the length L_{r_2} of the curve at this new time resolution. It is to be noted that as the resolution becomes coarser the estimate of length of the time series becomes less

accurate. Repeat the above procedure for different resolutions $r = r_1, r_2, ..., r_p$, where r_p is the maximum coarsest resolution at which the length of the curve is calculated. Let the lengths be denoted as L_r, $r = r_1, r_2, ..., r_p$. Draw a log-log graph of $(1/r_k)$ versus (L_r / r) and compute the slope α of the best fit straight

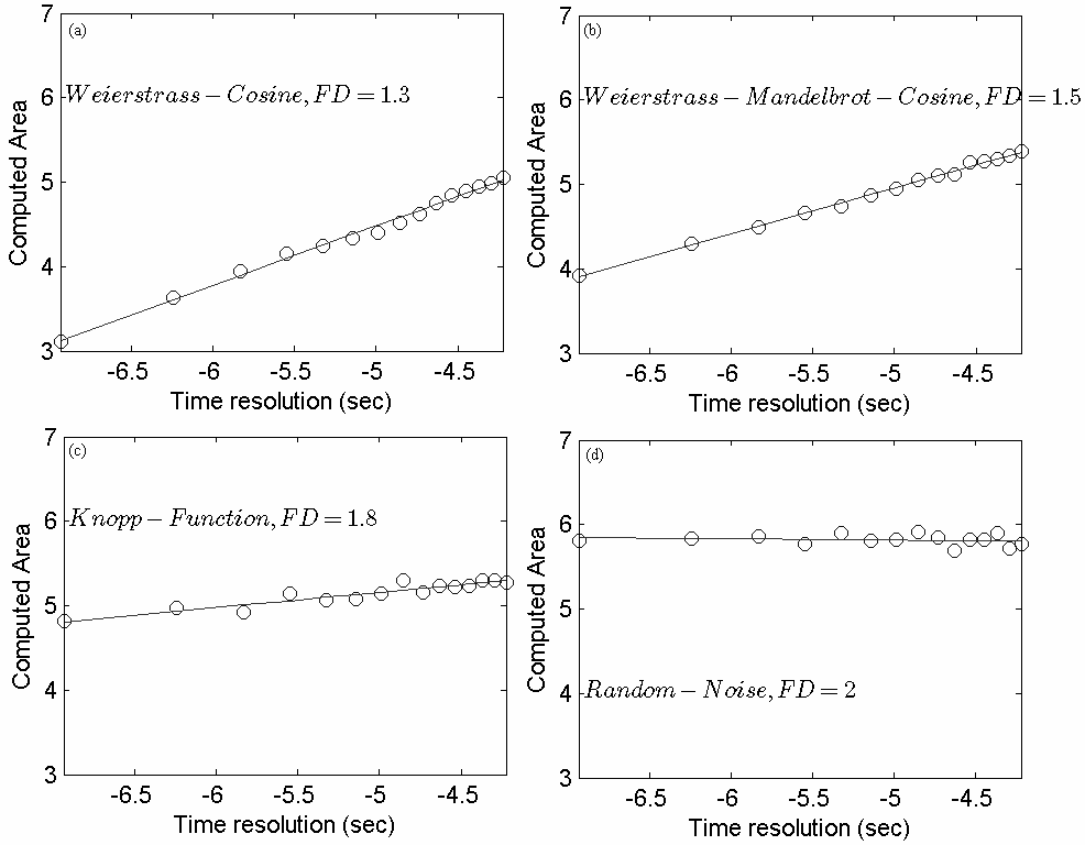

Fig. 3. Least square straight line fitting to log-log plot of total number of boxes required to cover the curve versus the size of the box (time resolution), for (a) Weierstrass cosine function, (b) Weierstrass-Mandelbrot cosine function (c) Knopp function, and (d) Fractional Brownian motion.

line to graph of points by linear regression method. Finally the fractal dimension of the time series is calculated as $D = \alpha$. We refer to this method as multiresolution length-based (MRL) method.

There are many methods available in the literature that deal with estimation of FD of time series waveforms. The methods such as Katz, Sevcik and Higuchi are considered here to compare the estimation accuracy results with the above proposed methods. They are explained briefly now.

D.Katz Method

This method is explained as follows [20]. Consider a waveform with sequence of points $[s_1, s_2, ..., s_N]^T$, where T represents transposition and N is the total number of samples in the sequence. The graph of the sequence is represented as $s_i = (x_i, y_i)$, $i = 1, 2, ..., N$, x_i are values of abscissa and y_i, are values of ordinate. In time series waveforms $x_i = t_i$, where t_i, $i = 1, 2, ..., N$ are monotonically increasing time instants at which the waveform is sampled. If the points s_1 and s_2 are represented as (x_1, y_1) and (x_2, y_2) respectively, the Euclidean distance between the points is computed as

$$dist(s_1, s_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}\ .$$

The fractal dimension of the waveform representing the time series is estimated using Katz method as follows.

The FD of the curve can be defined as

$$D = \frac{\log(L)}{\log(d)},$$

where L is the total length of the curve calculated as the sum of the distance between the successive data points as

$$L = \sum_{i=1}^{N-1} dist(i, i+1),$$

where $dist(i, j)$ is the distance between the points i and j on the curve, d is the diameter or planar extent of the curve, estimated as the distance between the first point and the point in the sequence that gives the farthest distance. For the

waveforms (signals) that do not cross themselves it can be expressed as $d = \max(dist(1,i))$, $i = 2,3,...,N$.

The FD computed in this manner depends on the measurement units used. If the units are different then so are the FDs. Katz's approach solved this problem by dividing the length by average step or average distance between the successive points, $\bar{a}$. This normalization results in

$$D_k = \frac{\log(L/\bar{a})}{\log(d/\bar{a})}.$$

Defining $n = L/\bar{a}$ the expression becomes,

$$D_k = \frac{\log(n)}{\log(n) + \log(d/L)}.$$

E. Sevcik Method

Sevcik [21] showed that approximate FD may be estimated from a set of N values sampled from a waveform. In this method, the FD estimate is derived from the definition of Hausdorff dimension (D_h). The D_h of a set in a metric space may be expressed as $D_h = \lim_{\varepsilon \to 0} \frac{-\log(N(\varepsilon))}{\log(\varepsilon)}$,

where $N(\varepsilon)$ is the number of open balls of radius ε needed to cover the set. In a metric space, given any point p, an open ball of center c and radius ε is a set of all points q for which $dist(p,q) < \varepsilon$. A curve of length L may be divided into $N(\varepsilon) = L/2\varepsilon$ segments of length 2ε and may be covered by $N(\varepsilon)$ balls of radius ε. Then the expression becomes

$$D_h = \lim_{\varepsilon \to 0}\left[-\frac{\log(L) - \log(2\varepsilon)}{\log(\varepsilon)}\right],$$

$$D_h = \lim_{\varepsilon \to 0}\left[1 - \frac{\log(L) - \log(2)}{\log(\varepsilon)}\right].$$

Sevcik proposes a double linear transformation of the curve into another normalized metric space, making all axes equal since the topology of a metric space does not change under linear transformation. For this, normalization of abscissa and ordinates are done as follows. $x_i^* = x_i / x_{\max}$, $i = 1,2,...,N$ and $y_i^* = (y_i - y_{\min})/(y_{\max} - y_{\min})$, $i = 1,2,...,N$ where $x_{\max} = \max(x_i)$ and $y_{\max} = \max(y_i)$, $y_{\min} = \min(y_i)$. These two linear transformations map the N points of the curve into another that belong to a unit square. The unit square can be visualized as covered by a grid of $N \times N$ cells. Calculating length L of the of the transformed waveform in the unit square and taking $\varepsilon = 1/2N'$, where $N' = N - 1$, the above equation becomes

$$D_s = \lim_{N' \to \infty}\left[1 + \frac{\log(L) - \log(2)}{\log(2(N-1))}\right].$$

The D_s is approximately equal to the fractal dimension D and the approximation improves as $N' \to \infty$.

F. Higuchi Method

Higuchi's method of computation of fractal dimension of the waveform is explained as follows [18]. An epoch of the waveform is represented by $y(1), y(2), ..., y(N)$, where N is the total number of samples in the epoch. From the given epoch, k new sub-epochs are constructed and represented by y_m^k, each of them is defined as

$$y_m^k = \{y(m), y(m+k), y(m+2k), ..., x(m+Mk)\},\ m = 1, 2, ..., k,$$

where m and k are integers, indicating initial time and interval time respectively, $M = \lfloor (N-m)/k \rfloor$, where $\lfloor a \rfloor$ denotes integer part of a. For each of the sub-epochs y_m^k constructed, the average length $L_m(k)$ is computed as

$$L_m(k) = \frac{1}{k}\left\{\frac{N-1}{Mk}\sum_{i=1}^{M}\left(\left|y(m+ik) - y(m+(i-1)k)\right|\right)\right\},$$

where $(N-1)/Mk$ is a normalization factor. The length of the epoch $L(k)$ for the time interval k is computed as the mean of the k values, for $m = 1, 2, ..., k$. That is $L(k) = \sum_{m=1}^{k} L_m(k)$.

If $L(k)$ is proportional to k^{-D}, the curve describing the shape of the epoch is fractal-like with the dimension D. Thus, if $L(k)$ is plotted against k, $k = 1, ..., k_{\max}$, on a double logarithmic scale, the points should fall on a straight line with a slope equal to $-D$. The least-square linear best fitting procedure is applied to the graph $(\ln(1/k), \ln(L(k)))$. The coefficient of linear regression of the plot of $\ln(L(k))$ versus $\ln(1/k)$ is taken as an estimate of the fractal dimension of the epoch. The value of interval time used is taken as $k = 1, 2, 3, 4$, and $k = [2^{(j-1)/4}]$ for k larger than 4, where $j = 11, 12, 13, ...$ and [.] denotes Gauss notation. We have used ten interval time values to compute Higuchi's fractal dimension.

III. Parametric Fractal Signals

To compute the accuracy of the proposed FD estimation methods and to compare the performance with the methods discussed, we have used parametric fractal waveforms which are briefly explained below.

A. Weierstrass cosine Function

The WCF [11] is defined as

$$W_H(t) = \sum_{k=0}^{\infty} \gamma^{-kH} \cos(2\pi\gamma^k t),\ 0 < H < 1,$$

where $\gamma > 1$. The function is continuous but nowhere differentiable, and its fractal dimension is $D = 2 - H$. If γ is integer, then the function is periodic with period one. We synthesized discrete time WCFs of various fractal dimensions by controlling the parameter H, and by sampling $t \in [0, 1]$ at $N+1$ equidistant points, using a fixed $\gamma = 5$ and truncating the infinite series so that the summation is done only for

$0 \le k \le k_{\max}$ and choosing $k_{\max} = 100$. Figure 4 shows waveforms of sampled WCF for fractal dimensions 1.2, 1.5 and 1.8.

B. Weierstrass-Mandelbrot cosine Function

This function is derived from Weierstrass-Mandelbrot function (WMF) $W(t)$ which is a scaling fractal curve [9]. The WMF of fractal dimension D is defined as

$$W(t) = \sum_{k=-\infty}^{\infty} \frac{(1 - e^{ib^k t})e^{i\phi_k}}{b^{(2-D)k}}, \quad 1 < D < 2,$$

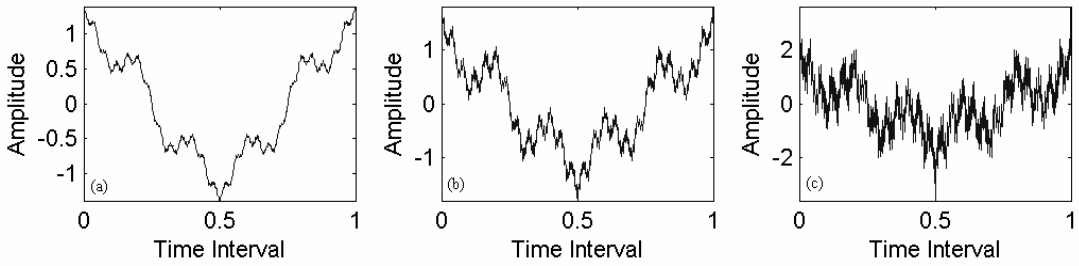

Fig. 4. Weierstrass cosine function, N=1025, g = 5, M = 100, (a) D = 1.2, (b) D = 1.5, (c) D = 1.8.

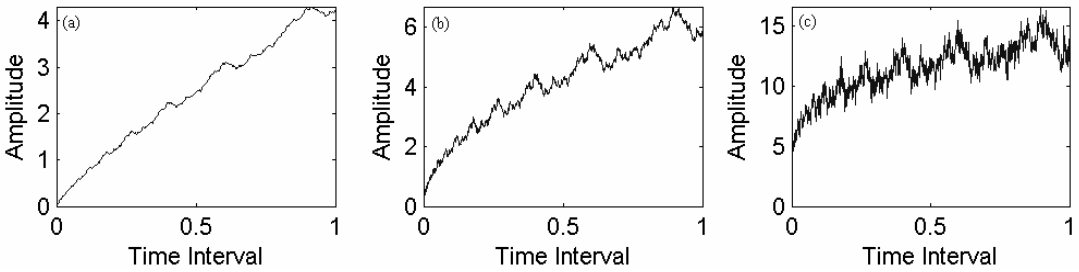

Fig. 5. Weierstrass-Mandelbrot cosine function, N = 1025, b = 1.5, M = 100, (a) D = 1.2, (b) D = 1.5, (c) D = 1.8.

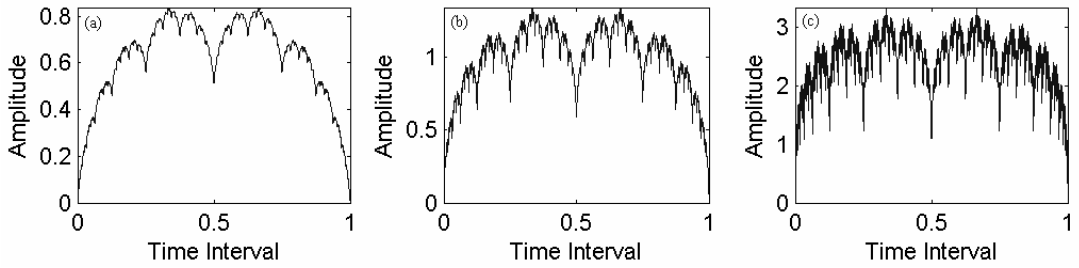

Fig. 6. Knopp function (Takagi function) for different parameter values of a, N = 1025, (a) a = 0.60, FD = 1.263, (b) a = 0.75, FD = 1.585, (c) a = 0.90, FD = 1.848.

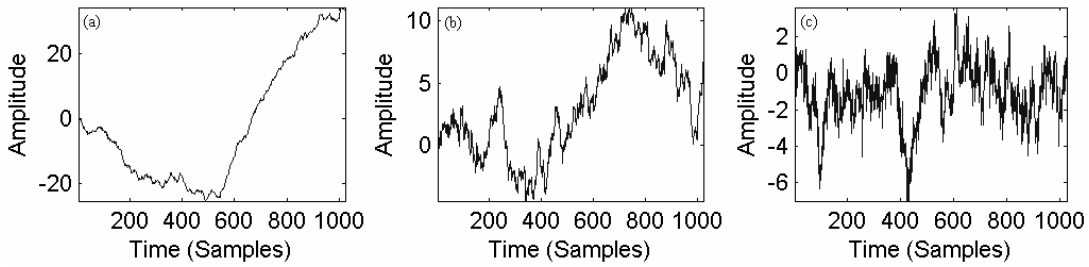

Fig. 7. Fractional Brownian motion, N = 1024, (a) FD = 1.2, (b) FD = 1.5, (c) FD = 1.8.

where ϕ_n is an arbitrary phase, and each choice of ϕ_n defines a specific function $W(t)$. This function is continuous but has no derivatives at any point. If we set $\phi_n = 0$ and taking real

part of $W(t)$ to obtain Weierstrass-Mandelbrot cosine function (WMCF) as $C(t) = \sum_{k=-\infty}^{\infty} \frac{(1-\cos b^k t)}{b^{(2-D)k}}$.

Figure 5 shows waveforms of discrete time WMCF for the value of fractal dimension equal to 1.2, 1.5, and 1.8, for $b = 1.5$.

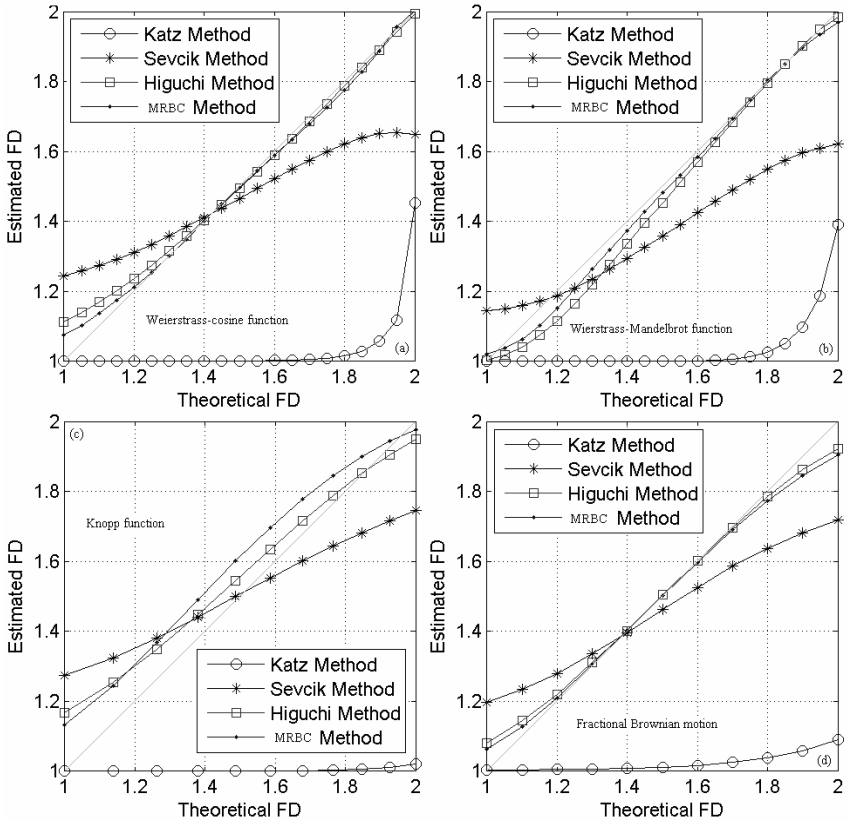

Fig. 8. Plot of estimated FD versus theoretical FD of synthetic fractal signals using Katz, Sevcik, Higuchi, and MRA method, for (a) Weierstrass-cosine function, (b) Weierstrass-Mandelbrot function, (c) Knopp (Takagi) function, (d) fractional Brownian motion. The number of samples in each of the signals is 1024.

where ϕ is the distance close to integer, that is $\phi(t) = |bt - round(bt)|$, b is an integer greater than one and a is a real number $a \in [0, 1]$. This function is everywhere continuous but nowhere differentiable if $ab \geq 1$. We set $b = 2$ and $a \in [1/2, 1]$. If $K(t)$ is defined with $1/2 < a < 1$ and $t \in [0, 1]$ then $K(t)$ has a Bouligand dimension of

$D = \log(4a)/\log(b)$

We synthesize discrete time TF by sampling $t \in [0,1]$ at $N+1$ equidistant points using a fixed value of parameter $b = 2$, with a maximum limiting value $k_{\max} = 100$ and for different values of parameter a, to get waveforms with different fractal dimension. Three sample waveforms are shown in Figure 6 for fractal dimensions of 1.263, 1.585, and 1.848.

C. Takagi Function

The TF also called Knopp function (KF) [21] is defined as $K(t) = \sum_{k=0}^{\infty} a^k \phi(b^k t)$,

D. Fractional Brownian Motion (fBm)

Fractional Brownian motions are non stationary and self similar stochastic processes, which are of great importance for modeling processes which exhibit long-term dependencies, such as $1/f$ type processes. We have used wavelet-based synthesis approach to generate fBms, and the approach is explained in [22]. For generating fBm waveforms, we have made use of the Matlab command $wfbm(H, N)$, which generates N sample fBm of Hurst exponent H. The fractal dimension of the waveform is computed using the relation $FD = 2 - H$. Three samples of fBm for H equal to 0.8, 0.5, and 0.2 (corresponding FDs 1.2, 1.5 and 1.8, respectively) are shown in Figure 7.

IV. Results

A. Results on Papmetric Fractal Signals

The proposed FD estimation methods are applied to synthesized mathematical fractal signals for studying their performance. The parametric fractal signals considered here are WCF, WMCF, TF and fBm. Since the fBm is a stochastic fractal, the estimation accuracy is averaged for 100 realizations of the time series. The fractal dimensions are increased from 1 to 2 in steps, and corresponding estimate values are computed using the two methods for all the four synthetic parametric fractal waveforms. Figure 8 (a)-(d) show the plot of estimated values versus theoretical values for the waveforms synthesized using WCF, WMCF, TF and fBm respectively. In Figure 9, the performance is compared for Higuchi, MRBC and MRL methods, in order to have high clarity plot. Note that, perfect reproduction of the true fractal dimension should yield a straight line of slope equal to one.

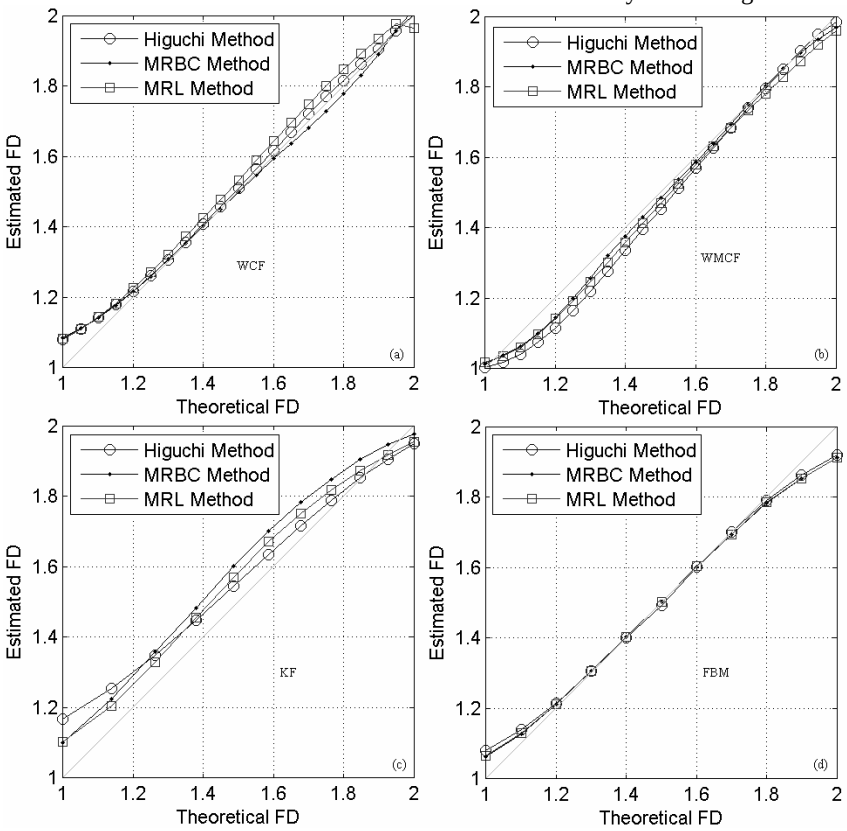

Fig. 9. Plot of estimated FD versus theoretical FD of synthetic fractal signals using Higuchi, and MRA method, for (a) Weierstrass-cosine function, (b) Weierstrass-Mandelbrot function, (c) Knopp (Takagi) function, (d) fractional Brownian motion. The number of samples in each of the signals is 1024.

From the results, it is observed that the proposed MRBC and MRL methods have provided the more accurate estimates of the fractal dimension for all the four parametric fractal waveforms than the Katz and Sevcik methods. In addition, the proposed methods have shown comparable estimation performance as that of Higuchi method. And also, a little bias is observed in the value of MRBC, MRL, and Higuchi FDs for the TF (KF) and fBm cases, and the values have become saturated towards fractal dimension of 1 and 2 for fBm waveforms. The Katz method is less accurate and has not provided linear variation but shown an exponential variation with increase of theoretical fractal dimension. Furthermore, the Sevcik method has shown a saturation at the beginning (near FD from 1 to 1.2) and towards end (near FD from 1.8 to 2), for all the four signals that we have used.

There is considerable amount of error which is expected due to sampling of continuous functions for which the true FD is defined. Since mathematically defined parametric fractal signals are sampled versions of non-band-limited fractal functions, some degree of fragmentation is lost during sampling. However, the true FD refers to the continuous time signal. Hence, the fractal dimensions estimation algorithms can offer only an approximate of true FD. In addition, the specific approach used to synthesize fractal signals, such as wavelets for fBm synthesis, affects the relationship between the degree of their fragmentation and the true FD. Thus, it may also affect the performance of the FD estimation algorithms.

B. Computation Time Taken

The time taken to compute FDs for various lengths of the data are compared for all the considered methods, and is plotted in Figure 10. The results show that for sequences of large length, the Higuchi's method has taken longer time to compute their FDs. It is also observed that the MRBC and MRL methods are comparatively faster than Higuchi and Sevcik method. The time taken to compute FD is increased with the length of the time series for Higuchi method, and the variation is almost quadratic.

C. Stochastic and Chaotic Signal

Consider a random process with linear temporal correlations, such as the autoregressive process of order one AR(1), $x_{i+1} = \alpha x_i + \varepsilon_i$, where $-1 < \alpha < 1$, and ε_i is a normally distributed random variable with mean zero and standard deviation one [23]. The standard deviation of the process is $\sigma_x = 1/\sqrt{1-\alpha^2}$, and its autocorrelation $\rho_k = \alpha^{|k|}$, where k is the time delay.

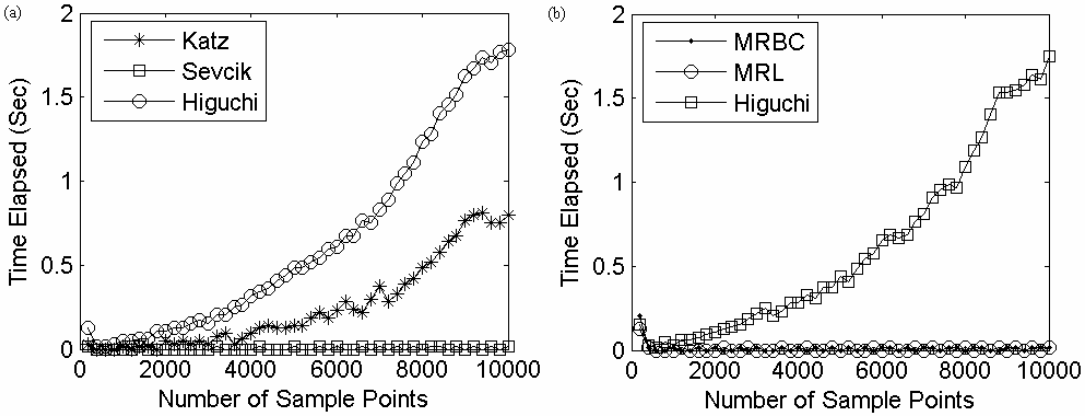

Fig. 10. Plot of time taken to estimate FD versus number of sample points in the signal, (a) comparison of Katz, Sevcik and Higuchi methods, (b) comparison of Higuchi, MRBC and MRL methods.

The nonlinear deterministic system known as the skew-tent map, given by $y_{i+1} = \begin{cases} 1/a & 0 \le y_i \le a \\ (1-y_i)/(1-a) & a \le y_i \le 1 \end{cases}$ is a non invertible transformation of unit interval into itself, with the parameter a chosen to satisfy $0 < a < 1$, and its invariant measure is uniform on the unit interval. This dynamic system is a chaotic system.

If the parameter values of these two systems are chosen such that $\alpha = 2a - 1$, then both systems will have identical power spectra [24]. To enhance the similarity between these two systems, a measurement function $z_i = h(y_i)$ can be used to transform the output of the skew-tent map y_i, so that the probability density function (pdf) of z_i is also normally distributed with mean zero and standard deviation σ_x, as in the case of the AR(1) process. The measurement function used is

$$z_i = \frac{\sqrt{2}}{\sigma_x} \Phi\left[2\left(y_i - \frac{1}{2}\right)\right],$$

where Φ is the inverse error function [25]. The autocorrelation functions of x_i and z_i will differ somewhat, although those of x_i and y_i are identical. Figure 12 shows the statistical properties of x_i and z_i for $a = 0.95$. Although the two time series shown in Figure 11 look different to the eye, their pdfs are identical, and their autocorrelation functions are similar, even though some disparity is introduced as the measurement function Φ is nonlinear. The difference in the underlying dynamical equations is best seen by investigating their return maps (delay reconstructions), as shown in Figure 12(c) and 12(f). Note that linear statistics, such as variance, autocorrelation function, and indeed any quantity defined with respect to power spectrum, fail to distinguish between the two systems illustrated in the Figure 12. The time series is segmented into epochs of length 1000 samples to get 901 epochs. The FD is found for each of the epochs and mean and standard deviation values are calculated and presented in the Table 1. The Katz FD values are similar for both stochastic and deterministic signals. On the other hand, the other methods (Sevcik, Higuchi, MRBC and MRL) have shown a clear distinction in FDs between the two time series.

TABLE I
FD OF STOCHASTIC AND DETERMINISTIC SIGNALS

	Stochastic		NL Deterministic	
	Mean	Std	Mean	Std
Katz	1.0120	0.0004	1.0113	0.0004
Sevcik	1.6239	0.0134	1.5608	0.0219
Higuchi	1.5933	0.0267	1.3833	0.0260
MRBC	1.5973	0.0282	1.3815	0.0278
MRL	1.5964	0.0282	1.3800	0.0278

In addition, we have also performed a study to interpret FD measure of signals in terms of their parameters such as amplitude, frequency, etc. The variation of FD with these parameters are computed and plotted, and the results are discussed as follows.

D.Effect of Waveform Amplitude

To test the effect of waveform amplitude on its fractal dimension, the amplitude of WCF is varied from 1 to 200 in steps of 10, and corresponding fractal dimension of the waveforms are computed using the five methods. The results are plotted as fractal dimension against the amplitude as shown in Figure 13. It is observed from the plots that the Katz method is sensitive to the waveform amplitude, and as the amplitude is increased the fractal dimension is also increased. However, the rest of the methods are insensitive to change in amplitude of the waveform. Hence, one should be careful in applying Katz method to find fractal dimensions, particularly for biomedical recordings, where the variance of the epochs is changing. And also note that because of this nature the Katz method finds applications in detection of transients in signals such as epileptic seizures.

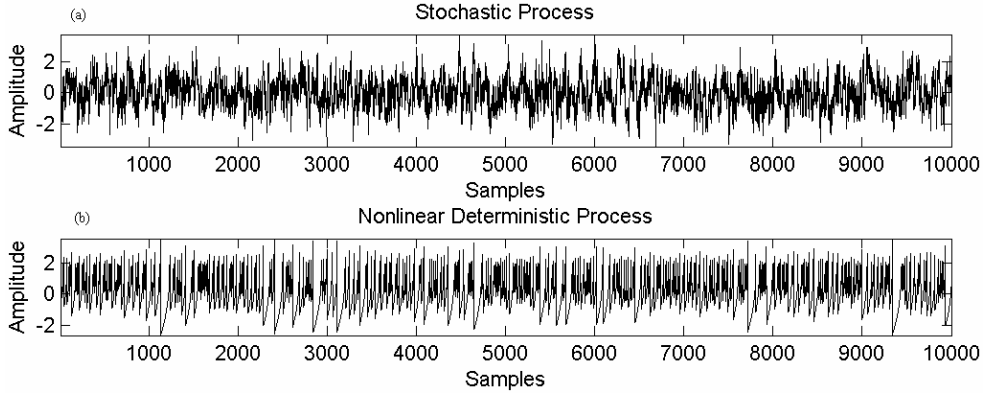

Fig. 11. (a) Time series of stochastic AR(1) process, (b) Time series of nonlinear deterministic process.

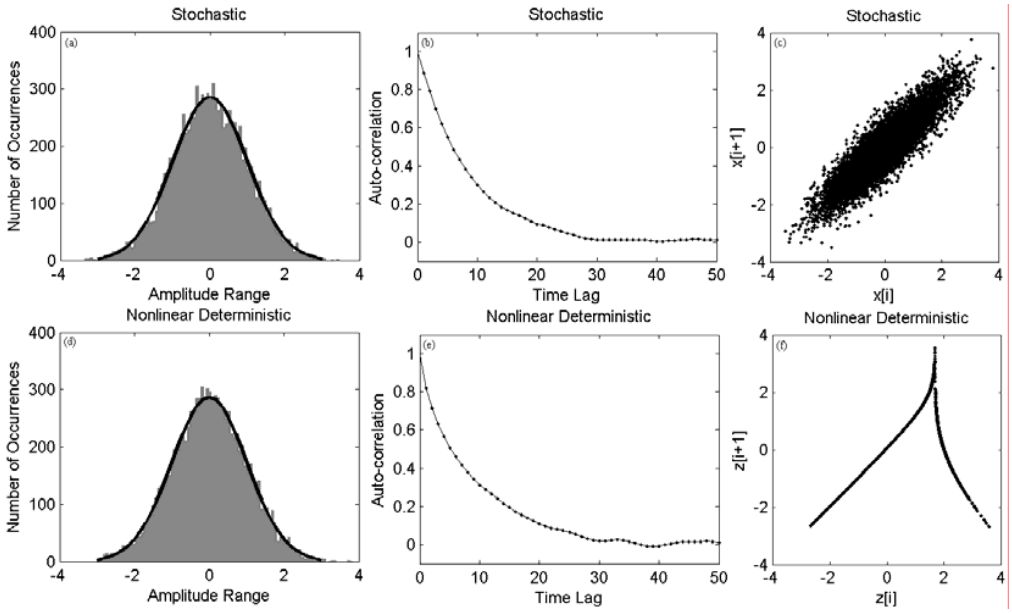

Fig. 12. (a) histogram, (b) auto-correlation, and (c) return map of stochastic process, (d) histogram, (e) auto-correlation, and (f) return map of non-linear deterministic process.

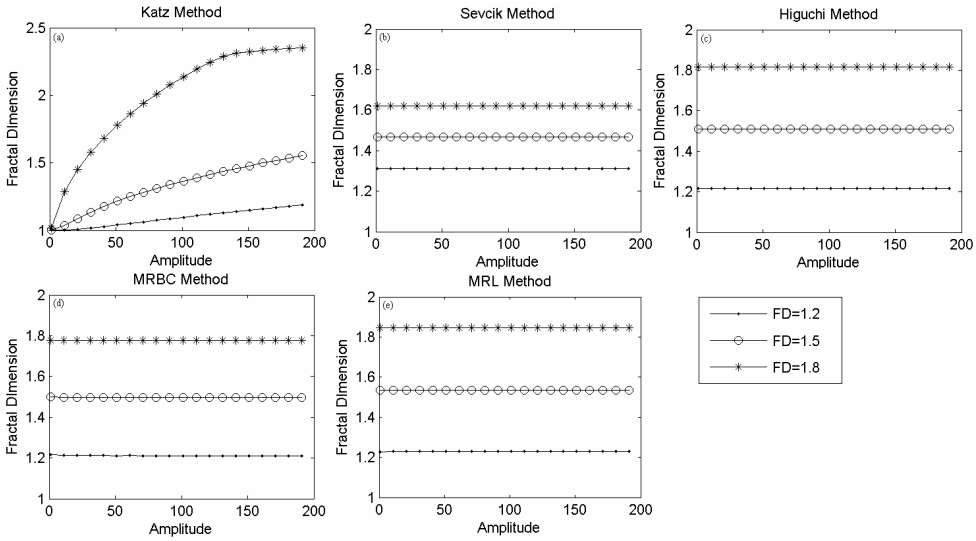

Fig. 13. Effect of waveform amplitude on FD, (a) Katz method, (b) Sevcik method, (c) Higuchi method, (d) MRBC method, (e) MRL method.

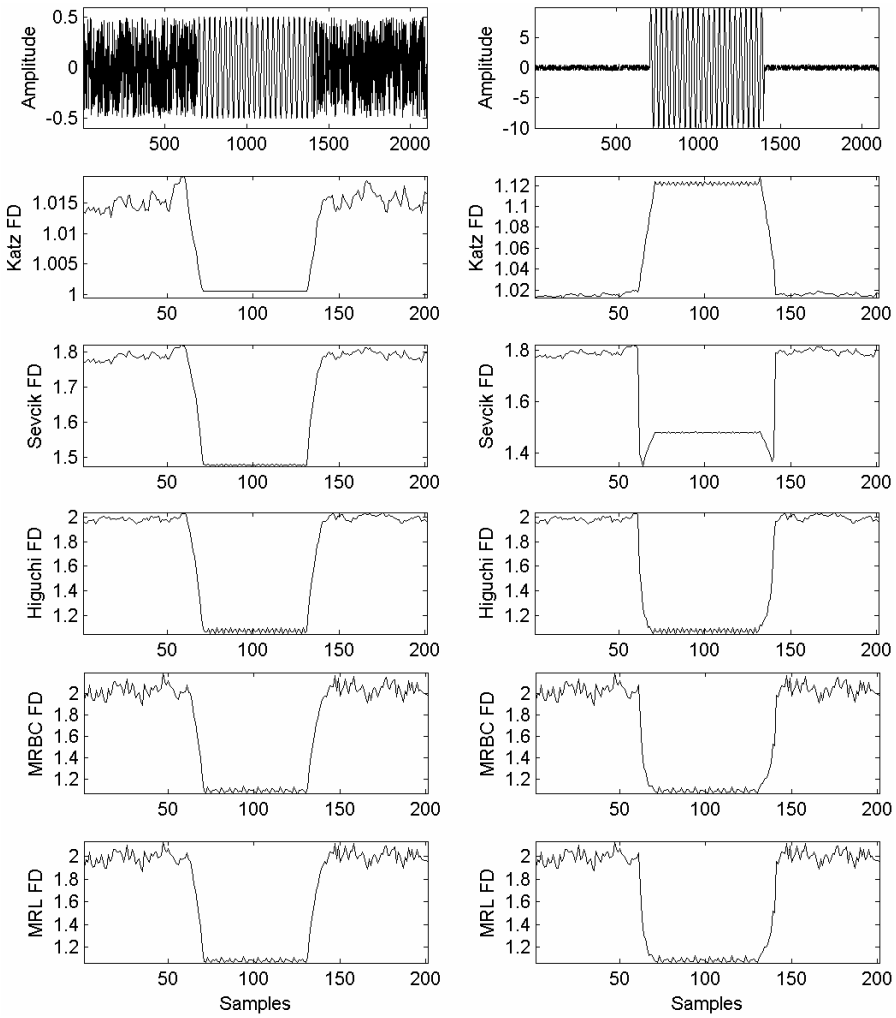

Fig. 14. Cascade of random and sinusoidal waves with constant variance (first row, left) and step increase in variance (first row, right), fractogram (STFD) using Katz method (second row), Sevcik method (third row), Higuchi's method (fourth row), MRBC method (fifth row), MRL method (sixth row).

E. Effect of Variance

We have also simulated two waveforms by cascading random and sinusoid waves as follows. The sinusoid wave of frequency 17 Hz is used here and its sampling frequency is 1000 Hz. The first waveform is

$$s_1 = [rand(1, N) - 0.5, 0.5\sin(2\pi fn), rand(1, N) - 0.5]^T$$

where the MATLAB command $rand(1, N)$ generates uniformly distributed random numbers, and T represents transpose. The second waveform is same as the first waveform but amplitude of sinusoid is changed to 10. That is

$$s_2 = [rand(1, N) - 0.5, 10\sin(2\pi fn), rand(1, N) - 0.5]^T .$$

The values of N chosen here is 600. The short-time fractal dimension (fractogram) of the cascade waveforms is plotted by computing fractal dimensions of moving window of 100 samples with an overlap of 50 samples. This test is carried out to check the effect of changing variance of a waveform on its fractal dimension. The time series and STFD computed are depicted in the Figure 14. The sinusoidal waveforms have the property of less space filling than the random waveforms. This is also shown in the FDs in the left column of the figure. However, due to its sensitivity to variance, the Katz FD value has increased for the sinusoidal waveforms as shown in the right column of the Figure 3.14.

F. Effect of Sampling Frequency

The effect of sampling frequency on the estimated fractal dimensions is tested by simulating waveforms of WCF of amplitude 100. The sampling frequency of the waveform is varied from 501 Hz to 4300 Hz in steps of 300 Hz, keeping the time duration constant, and corresponding fractal dimensions are computed using all the five methods. This test is carried out for waveforms of three different fractal dimensions such as 1.2, 1.5 and 1.8 as shown in Figure 15. The estimation accuracy is improved for all the methods except for Katz method as the sampling frequency is increased. However, the Katz fractal dimension is decreased as the sampling frequency

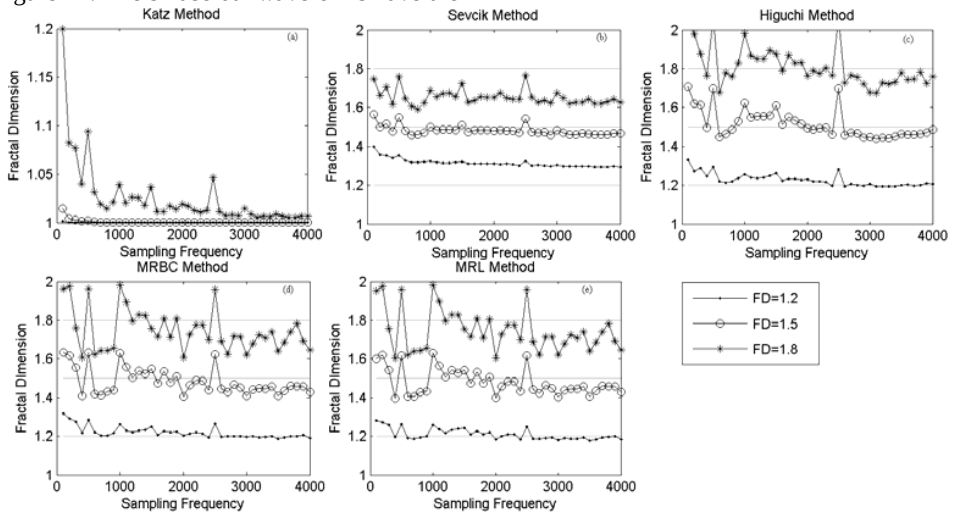

Fig. 15. Effect of sampling frequency of waveforms on FD estimate, (a) Katz method, (b) Sevcik method, (c) Higuchi method, (d) MRBC method, (e) MRL method.

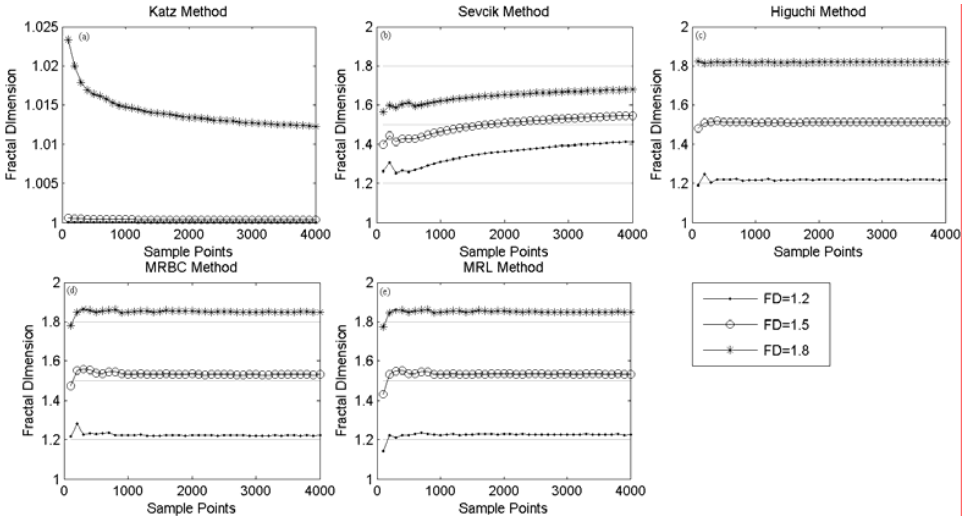

Fig. 16. Effect of waveform length (number of sample points) on FD estimate, (a) Katz method, (b) Sevcik method, (c) Higuchi method, (d) MRBC method, (e) MRL method.

of the waveform is increased. This is because, as the number of samples N increase, the ratio d/L in the Katz equation approaches a constant. Hence the fractal dimension rapidly decreases towards one.

G.Effect of Signal Length

The FDs of signals of various sample lengths are computed and plotted in the Figure 16, to study the effect of number of sample points of a signal on its FD value. The test is performed for signals of three different fractal dimensions (1.2, 1.5 and 1.8). The FD estimate of Katz method is not accurate and has shown a decrease of FD as the samples in the signal is increased. Even though the Sevcik method is less accurate, FD computed using this method increased and reached saturation. The Higuchi, MRBC and MRL methods showed constant values of FDs irrespective of their sample points except for the signals with very less number of samples. For higher FDs, a little bias in the FD estimate is observed in all the three cases.

H.Effect of Noise

To study the effect of noise, we have added noise to the WCF in steps and corresponding FD is computed. The variation of FD versus noise amplitude is shown in the Figure 17. Since the Katz method is sensitive to amplitude of waveforms, the result has shown a linear variation. In the case of other methods, the FD variation is increased for less value of noise amplitude and reached saturation. The sevcik FD reached to a value of 1.7 and Higuchi, MRBC and MRL FD reached towards 2 which is the FD of random noise waveforms. It is also observed that the waveform with higher

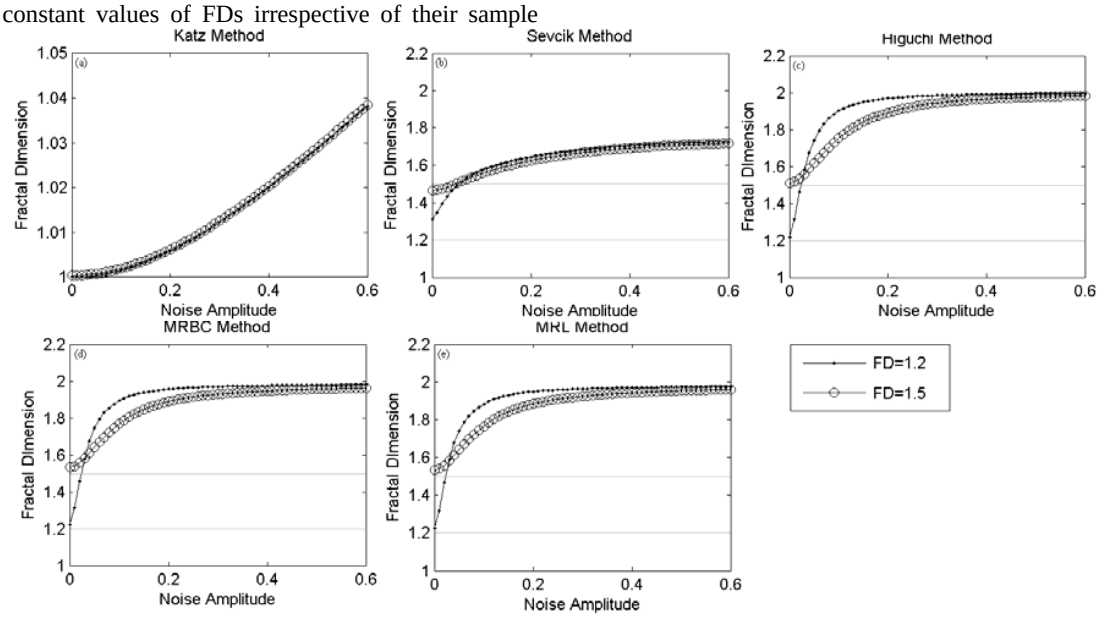

Fig. 17. Effect of noise power on FD estimate, (a) Katz method, (b) Sevcik method, (c) Higuchi method, (d) MRBC method, (e) MRL method.

FD value has shown a slow variation with noise amplitude than the waveform with less FD.

V. Discussion and Conclusion

In this paper, we have proposed a method to compute FD of signal waveforms, based on counting the number of boxes to cover the waveform entirely at multiple resolutions. We referred this technique as MRBC method. A modification of this method resulted in another method; we referred it as MRL method. In MRBC method, total number of boxes which cover the entire waveform is calculated at multiple time resolutions, whereas signal length at multiple time resolutions are computed in MRL method. We have used these two methods to compute FD of various signal waveforms, parametric fractal signals, as also sinusoids, and random signals.

Since the parametric fractal signals are mathematically defined and their true (theoretical) FD can be calculated, we have used these waveforms to compare the performance of the proposed methods with other methods, by plotting graph of estimated FD versus true FD. The parametric fractal signals considered in this study are WCF, WMCF, KF (TF), and fBm signals. Since the fBm is a stochastic fractal signal, the FD values are computed for 100 different realizations of the time series at each of FD value, and the results are averaged to get the FD value. Other methods discussed in this chapter, such as Katz, Sevcik, and Higuchi, are used for comparing the results.

The proposed MRBC and MRL methods have shown superior performance in estimating FD of waveforms than Katz and Sevcik methods, while the accuracy results are comparable to that of Higuchi method. Furthermore, the MRBC and MRL methods have taken less time to compute FD compared to Higuchi method. The Katz method has shown poor performance in estimating FD whereas the FD values of Sevcik method have become saturated at low and high FD values. In addition, the MRBC and MRL methods do not require specifying the value of interval time as it is required in the Higuchi method. It is also observed from the results that the Higuchi method has taken more time to compute the FD of long-length time series data, other results being similar.

In the usual box-counting method of computing FD, covering the graph of one dimensional discrete time signal by grids involves two dimensional processing of the signal at multiple scales. However, the proposed MRBC and MRL methods involve one dimensional processing of the signal at multiple time resolutions. The methods can yield results that are invariant with respect to shifting of the domain of the signal and scaling of its dynamic range. In addition, the methods can be applied to arbitrary time signals and used to measure short-time FD (fractogram) of time varying signals.

Since the MRBC method has given more accurate FD results compared to other methods which are discussed here in this chapter, we have used this method to compute FD of various signal waveforms including chaotic, stochastic, and sinusoids. In an experimental study, the FD has clearly discriminated stochastic and chaotic signal waveforms for which mean, variance, and autocorrelation are similar except the return maps. Thus, the FD finds applications in distinguishing signals having similar second order statistics but of different nature.

In addition, we have made a study of effect of various parameters, such as signal amplitude, frequency, sampling frequency, signal length, noise power, noise band-width, autocorrelaton, on FD. It is found that the MRBC, MRL, Sevcik, and Higuchi methods are insensitive to wave form amplitudes (variance). However, the Katz method has shown sensitive dependence on signal amplitude (variance). As the amplitude is increased, the Katz FD of the waveform is increased.

In conclusion, the proposed MRBC and MRL method gives comparable performance in estimating FD of waveforms as Higuchi method, computation time taken being very less. This property may be used in real-world applications, as in clinical settings to compute structural changes in signal waveforms.

References

[1] Mandelbrot BB. The fractal geometry of nature. Freeman, New York, 1983.
[2] Barnsley M F. Fractals Everywhere, 2nd ed, New York: Academic Press Professional; 1993.
[3] Mandelbrot BB, Hudson RL. The (Mis)Behavior of Markets, A Fractal View of Risk, Ruin and Reward. Basic Books, New York, 2004.
[4] Falconer K. Fractal geometry: Mathematical foundations and applications. Wiley, New York, 1990.
[5] Maragos P, Potamianos A. Fractal dimensions of speech sounds: Computation and application to automatic speech recognition. J of the Acoustical Society of America 1999; 105 (3): 1925-1932.
[6] Hadjileontiadis LJ, Rekanos IT. Detection of explosive lung and bowel sounds by means of fractal dimension. IEEE Sig Proc Lett 2003; 10(10): 311-314.
[7] Pradhan N, Dutt DN. Use of running fractal dimension for the analysis of changing patterns in electroencephalograms. Comput Biol Med 1993; 23(5): 381-388.
[8] Guzman-Vargan L, Angulo-Brown F. Simple model of the aging effect in heart inter-beat time series. Phys Rev E 2003; 67: 05290.1-05290.4.
[9] Feder J. Fractals, New York: Plenum Press; 1988.
[10] Accardo A, Affinito M, Carrozzi M, Bouquet F. Use of the fractal dimension for the analysis of electroencephalographic time series. Biol Cyber 1997; 77: 339-350.
[11] Esteller R, Vachtsevanos G, Echauz J, Litt B. A comparison of waveform fractal dimension algorithms. IEEE Trans Biomed Eng 2001; 48(2): 177-183.
[12] Klonowski W. Chaotic dynamics applied to signal complexity in phase space and in time domain. Chaos Solitons and Fractals 2002; 14: 1379-1387.
[13] Maragos P, Sun F-K. Measuring the fractal dimension of signals: Morphological covers and iterative optimization. IEEE Trans Signal Proc 1993; 41(1): 108-121.
[14] Pitsikalis V, Maragos P. Filtered dynamics and fractal dimension for noisy speech recognition. IEEE Sig Proc Lett 2006; 13(11): 711-714.
[15] Spacic S, Kalauzi A, Grbic G, Martac L, Culic M. Fractal analysis of rat brain activity after injury. Med & Biol Eng & Comput 2005; 43: 345-348.
[16] Tricot C. Curves and fractal dimension, Springer-Verlag, New York, 1995.

[17] Schepers HE, van Beek JHGM, Bassingthwaighte JB. Four methods to estimate the fractal dimension from self-affine signals. IEEE Engg Med Bio 1992; 6: 57-64.
[18] Higuchi T. Approach to an irregular time series on the basis of the fractal theory. Physica D 1988; 31: 277-283.
[19] Wornell G, Oppenheim A. Estimation of spectral signals from noisy measurements using wavelets. IEEE Trans Sig Proc 1992; 40(3): 611-623.
[20] Katz M. Fractals and the analysis of waveforms. Comput Biol Med 1988; 18: 145-156.
[21] Dubuc B, Dubuc S. Error bounds on the estimation of fractal dimension. SIAM J Numer Ana 1996; 33(2): 602-626.
[21] Sevcik C. On fractal dimension of waveforms. Chaos Solitons & Fractals 2006; 28: 579-580.
[22] Flandrin P. Wavelet analysis and synthesis of fractional Brownian motion. IEEE Trans Info Theor 1992; 38(2): 910-917.
[23] Chatfield C. The analysis of time series, 4th ed, Chapman and Hall, London, 1989.
[24] Smith LA. The maintenance of uncertainty, in Nonlinearity in geophysics and astrophysics 1997, vol CXXXIII, Int School of Physics, 177-246.
[25] Press WH, Flannery BP, Teukolsky SA, Vetterling WT. Numerical recipes in C, 2nd ed, CUP Cambridge, 1992.

B. S. Raghavendra was born in Bobbi-Thirthahalli, India. He received B.E. degree from Bangalore University in 2001, and M. Tech. degree in Electronics and Communication Engineering from National Institute of Technology Karnataka, Surathkal, India, in 2004. Currently he is PhD student in the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India. His research interests include biomedical signal processing, nonlinear dynamics, and computational neuroscience.

D. Narayana Dutt obtained his Ph.D degree in Electrical Communication Engineering from the Indian Institute of Science, Bangalore, India. He has been a member of faculty in the same Institute for the past several decades. He worked earlier in the areas of acoustics and speech signal processing. He has been working for the past several decades in the area of applications of digital signal processing to the analysis of biomedical signals. He has been collaborating with researchers from NIMHANS, Bangalore, in applying signal processing techniques to various problems in the area of neuroscience. He worked in the Biomedical Engineering Research Centre at Nanyang Technological University in Singapore, as a visiting faculty and worked in the application of digital signal processing to cardiovascular signals. He has published a large number of papers in this area in leading international journals.