

On hidden fractal model signal processing*

Vikram Krishnamurthy, John B. Moore

*Department of Systems Engineering, Research School of Physical Sciences and Engineering, Australian National University,
GPO Box 4, Canberra ACT 2601, Australia*

Shin-Ho Chung

Research School of Biological Sciences, Australian National University, GPO Box 4, Canberra ACT 2601, Australia

Received 7 June 1990

Revised 5 December 1990 and 18 March 1991

Abstract. Fractal stochastic processes are examples of semi-Markov processes where the signal behaviour is a function of the prefiltering bandwidth. In this paper we develop schemes for estimating such fractal models when they are hidden (imbedded) in noise. We reformulate this hidden fractal model (HFM) problem in the scalar case as a higher order scalar or first order 2-vector homogeneous hidden Markov model (HMM) problem in which the state consists of the signal augmented by the time to the last transition. With this reformulation, we can apply HMM signal processing techniques to obtain optimal estimates of the signals and signal model parameters, including transition probabilities and noise statistics. Also, the signal levels and fractal dimension can be estimated.

Zusammenfassung. Fraktale stochastische Prozesse sind Beispiele von Semimarkovprozessen, wobei das Signalverhalten eine Funktion der Vorfilterbandbreite ist. In diesem Beitrag entwickeln wir Methoden für die Schätzung solcher fraktaler Modelle, die in Rauschen 'versteckt' (eingebettet) sind. Wir formulieren dieses Problem der Bestimmung eines hidden fractal model (HFM) im skalareren Fall neu als ein Problem zur Bestimmung eines homogenen hidden Markov model (HMM), das entweder skalarwertig und von höherer Ordnung ist oder vektorwertig mit zwei Komponenten und von erster Ordnung. Dabei besteht der Zustand aus dem Signal, vermehrt um die Zeit bis zum letzten Übergang. Mit dieser Neuformulierung können wir HMM-Signalverarbeitungstechniken anwenden, um die optimalen Schätzwerte für die Signale und Signalmodellparameter einschließlich der Übergangswahrscheinlichkeiten und der Rauschstatistiken zu erhalten. Ebenso können die Signalpegel und die fraktale Dimension geschätzt werden.

Résumé. Les processus stochastiques fractals sont des exemples de processus semi markoviens pour lesquels le comportement du signal est fonction de la largeur de bande du préfiltrage. Nous développons dans cet article des méthodes permettant d'estimer de tels modèles fractals quand ils sont cachés (immergés) dans le bruit. Nous reformulons le problème de ce modèle fractal caché (en anglais hidden fractal model, HFM) dans le cas scalaire comme un problème de modèle de Markov caché (en anglais hidden Markov model, HMM) scalaire d'ordre supérieur ou vectoriel de dimension deux et d'ordre un dans lequel l'état est constitué par le signal augmenté du temps jusqu'à la dernière transition. A l'aide de cette reformulation, nous pouvons appliquer les techniques de traitement de signaux HMM afin d'obtenir les estimations optimales des paramètres du signal et du modèle de signal, ce qui inclue les probabilités de transition et la statistique du bruit. De plus, les niveaux de signal et la dimension fractale peuvent eux aussi être estimés.

Keywords. Semi-Markov processes, hidden Markov model processing, maximum likelihood estimation, hidden fractal models.

* Partially supported by D.S.T.O. Australia and Boeing (BCAC), USA.

1. Introduction

Discrete-state hidden Markov models (HMMs) with time-invariant transition probabilities have been widely used to model noisy physical systems in areas such as communication systems, speech processing and biological signal processing. However, in some physical systems, the rate of transition from one state to another could well be a function of the bandwidth of the preprocessing filters, referred to by some authors as the time scale in which it is observed and so the transition probabilities depend on the time to the last transition. Such 'fractal' stochastic processes have time-varying transition probabilities described in terms of a fractal dimension. The longer the time spent in a particular state, the greater the probability of remaining in the particular state. The notion of an HMM then generalizes naturally to that of a hidden fractal model (HFM), being a useful and broad class of HMMs that are non-homogeneous, i.e., with time varying transition probabilities. HFMs can be characterized in terms of state signal levels, time-varying state transition probabilities and noise characteristics. In this paper one scheme for estimating such HFMs is proposed.

This paper has been in part motivated by our curiosity to see if HMM signal processing techniques naturally generalize for HFMs, and if so, whether there is any incentive for working with such techniques instead of HMM techniques in some of the areas of application of HMM signal processing. One particular class of signals of interest to our work is that reported in the biophysical literature concerning ion channel currents in cell membranes [14]. Our previous work has been with first order HMM representations of ion channels [3]. However, other workers [8–10] suggest that fractal models [11] could be more relevant than first order Markov models. Their proposed schemes cannot cope with background noise. However, often the channel currents are very small and obscured by unavoidable background noise. In such cases when a fractal process is imbedded in noise, these techniques are inadequate (as we

shall illustrate in simulation studies) and so motivate a more sophisticated HFM processing approach.

The paper is organized as follows: In Section 2 we formulate the scalar HFM problem as a 2-vector HMM problem and review learning and estimation objectives of HMM schemes. In Section 3, estimation and re-estimation schemes are presented for vector HMMs. These schemes are also formulated for higher order hidden Markov models. In Section 4, we present simulation studies. Finally, some conclusions are presented in Section 5.

2. HFM signal processing

2.1. Finite state fractal models

Binary-state continuous-time fractal model

For simplicity we first describe the simplest binary-state fractal model, termed a channel with an open state and closed state. Later the more general case of having N_s states is considered. Define $k_1(t)\Delta t$ and $k_2(t)\Delta t$ as the time-varying transition probabilities of leaving the open and closed states, respectively, in the time interval Δt . These probabilities are in terms of the time to the last transition, denoted t . Let $P(t)$ be the probability that the channel remains in the closed state over the interval $[0, t]$. It can then be shown [9, 10] that

$$P(t)k_1(t) = -\frac{dP(t)}{dt}$$

$$\text{or } \int \frac{dP(t)}{P(t)} = - \int k_1(t) dt. \quad (2.1)$$

Of course, $f(t) = -P(t)k_1$ is the associated probability density function since $P(t) = \int_0^t f(\tau) d\tau$. When the kinetic rate constants $k_1(t)$, $k_2(t)$ are independent of the time to the last transition and are then denoted by k_1 , k_2 , the model is first order homogeneous Markov. For fractal models, the kinetic rate constants are assumed to be time-dependent as follows [9]:

$$k_1(t) = k_1 t^{1-D_1}, \quad k_2(t) = k_2 t^{1-D_2},$$

$$1 \leq D_1, D_2 < 2, \quad (2.2)$$

where D_1, D_2 are termed the fractal dimensions. Thus the transition rate matrix \mathbf{H} is

$$\mathbf{H} = \begin{bmatrix} -k_1 t^{1-D_1} & k_1 t^{1-D_1} \\ k_2 t^{1-D_2} & -k_2 t^{1-D_2} \end{bmatrix}. \quad (2.3)$$

The corresponding transition probabilities are the solutions of the matrix differential equation $\dot{\mathbf{A}}(t) = \mathbf{H}(t)\mathbf{A}(t)$, $\mathbf{A}(0) = \mathbf{I}$, where $\mathbf{A}(t) = (a_{ij}(t))$, $a_{ij}(t) = P(s_t = q_j | s_0 = q_i)$. It is not in general possible to obtain an expression for these transition probabilities in a closed form. However when $D_1 = D_2$ then $\mathbf{A}(t) = \exp(\int_0^t \mathbf{H}(\sigma) d\sigma)$ and the transition probabilities can be obtained in a closed form. Clearly, the smaller the time to the last transition, the greater the probability of a transition in a subsequent relatively small time interval Δt . There is an analogy with the more familiar example of fractal coastlines, where the 'ruggedness' is invariant of the spatial resolution. For a channel with fractal kinetics, the degree of 'flickering', or more precisely the fractal dimensions D_1, D_2 are invariant of the temporal resolution which is determined by the bandwidth of the measurement apparatus. This means that the effective rate constant of the channel is higher when measured at a finer temporal resolution.

Substituting $k_1(t)$ from (2.2) into (2.1) yields expressions for the distribution $P(t)$ and density function $f(t)$:

$$P(t) = \exp\left(-\frac{k_1}{2-D_1} t^{2-D_1}\right), \quad (2.4)$$

$$f(t) = k_1 t^{1-D_1} \exp\left(-\frac{k_1}{2-D_1} t^{2-D_1}\right).$$

Thus $P(t)$ satisfies the Weibull distribution [4]. Also $f(t)$ is the frequency histogram of closed times. Notice that for a first order Markov process, $D = 1$ and $\log(f(t))$ versus t is a straight line.

The rate of channel openings and closing should depend inversely on the time scale, so that $D_i \geq 1$. To normalize the probability distribution requires that $\lim_{t \rightarrow 0} P(t)$ exists, which is true only if $D_i < 2$. Thus the fractal dimension D_i is restricted to the range $1 \leq D_i < 2$.

Computing the fractal dimensions

One commonly used technique for the noise-free observations is described in [8, 9]. It is shown in [8] that D_i is overestimated on the average by 10 percent. However, because of the simplicity of this scheme we shall use a similar analysis with signal estimates derived from HFM processing in Section 3.3.

Discrete-time fractal model

We now consider the discrete-time version of the continuous time fractal model obtained by appropriate lowpass (anti-aliasing) filtering and sampling at greater than or equal to the Nyquist sampling rate. The filter bandwidth then determines the temporal resolution.

Let $f_s = 1/T_s$ be the sampling frequency. Define the discrete-time transition probability matrix $\mathbf{A}(t_k) = (a_{ij}(t_k))$, where $a_{ij}(t_k) \triangleq P(s_{k+1} = q_j | s_k = q_i)$, and t_k is the time to the last transition at time k , i.e., the time the process s_k has spent in the current state q_i . For convenience we work with the number of discrete-time samples after the last transition rather than the actual time. Thus for a sampling time interval of T_s , t_k is quantized to a finite number N_i of possible integer values $\tau_1, \dots, \tau_{N_i}$. Without any computational effort and memory constraints, it would be reasonable to take τ_i as the integer i , for $i = 1, \dots, N_i$. An upper bound on N_i is the total number of observations in the data set. However, as described later, state aggregation allows N_i to be bounded by more realistic values, usually 5 to 10.

The discrete transition probabilities of the sampled fractal process cannot be obtained in an analytical form in general. However, it is easily shown that

$$\begin{aligned} a_{11}(t_k) - a_{21}(t_k) &= a_{22}(t_k) - a_{12}(t_k) \\ &= \exp\left(k_1 \frac{t_k^{2-D_1} - (t_k-1)^{2-D_1}}{2-D_1} \right. \\ &\quad \left. + k_2 \frac{t_k^{2-D_2} - (t_k-1)^{2-D_2}}{2-D_2} \right). \end{aligned} \quad (2.5)$$

This illustrates the behaviour of the transition probabilities: with increasing t_k , $a_{11}(t_k)$ and $a_{22}(t_k)$ increase and asymptotically approach 1; $a_{12}(t_k)$, $a_{21}(t_k)$ asymptotically approach 0. We shall also use (2.5) subsequently when dealing with state aggregation.

If $D_1 = D_2$ then it is easily shown that $\mathbf{A}(t_k)$ can be expressed in a closed form as

$$\mathbf{A}(t_k) = \frac{1}{k_1 + k_2} \begin{bmatrix} k_2 + k_1 c(t_k) & k_1 - k_1 c(t_k) \\ k_2 - k_2 c(t_k) & k_1 + k_2 c(t_k) \end{bmatrix},$$

$$c(t_k) = \exp\left(-\frac{k_1 + k_2}{2-D} (t_k^{2-D} - (t_k - 1)^{2-D})\right). \quad (2.6)$$

Notice that when $D_i = 1$, the transition probabilities are independent of time and s_k reduces to a homogeneous first order Markov process.

Multi-state discrete-time semi-Markov model

We now 'generalize' the two-state discrete-time fractal model to a discrete semi-Markov model with a finite arbitrary number of states. Consider a discrete-time, finite-state stochastic process s_k , $k \geq 0$, where for each k , s_k is a random variable taking on a finite number N_s of possible states q_1, \dots, q_{N_s} . The fractal model is a particular case of semi-Markov models whose transition probabilities are functions of the time to the last transition:

$$a_{ij}(t_k) : t_k \mapsto [0, 1] \quad \text{for } i, j \in [1, N_s]. \quad (2.7)$$

2.2. Reformulation of fractal models as Markov models

Define the 2-vector process \mathbf{S}_k as $\mathbf{S}_k = (t_k, s_k)$ for each $k \geq 0$. Clearly \mathbf{S}_k is a finite-state process with $N = N_s N_t$ states. Here N_t is taken as the maximum duration time in any state considering the observation sequence of length T . It is easily shown that the 2-vector stochastic process \mathbf{S}_k , as defined above with $\tau_i = i$ for $i = 1, \dots, N_t$ is a homogeneous, first order Markov process. Notice

that

$$t_{k+1} = \begin{cases} t_k + 1 & \text{if } s_{k+1} = s_k \text{ and } t_k < N_t, \\ 1 & \text{otherwise.} \end{cases} \quad (2.8)$$

So t_{k+1} depends only on t_k , s_k and s_{k+1} . Also from (2.8) we have

$$P(\mathbf{S}_{k+1} | \mathbf{S}_k) = \begin{cases} a_{ii}(\tau_h) & \text{if } \mathbf{S}_k = (\tau_h, q_i) \\ & \text{and } \mathbf{S}_{k+1} = (\tau_h + 1, q_i), \\ & 1 \leq \tau_h < N_t, \\ a_{ij}(\tau_h) & \text{if } \mathbf{S}_k = (\tau_h, q_i) \\ & \text{and } \mathbf{S}_{k+1} = (1, q_j), \\ & i \neq j, 1 \leq \tau_h \leq N_t, \\ 0 & \text{otherwise,} \end{cases} \quad (2.9)$$

which is independent of k .

REMARK. If for some integer $N'_t < N_t$, τ_i is defined more generally as $\tau_i = i$, $i = 1, \dots, N'_t - 1$ and $\tau_{N'_t} = i : N'_t \leq i \leq N_t$ in Lemma 2.1, the first order Markov property still holds. However, it is readily shown that \mathbf{S}_k is then not necessarily homogeneous. For certain functions $a_{ij}(t_k)$ that 'saturate' beyond some t_k , \mathbf{S}_k is homogeneous and aggregation of states in the saturation region is possible as discussed below.

NOTATION. We denote the set of $N = N_s N_t$ states $\{(\tau_1, q_1), \dots, (\tau_{N_t}, q_{N_s})\}$ as $\{\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_N\}$, although not necessarily in the same order. We will denote elements of this set by integer subscripts, usually m or n . So a fractal model with the set of states $\{\tau_h, q_i\}$ is viewed as a vector Markov model with states $\{\mathbf{Q}_m\}$. Also for $\mathbf{Q}_m = (\tau_h, q_i)$ and $\mathbf{Q}_n = (\tau_l, q_j)$, where $m, n \in [1, N]$, $\tau_h, \tau_l \in [1, N_t]$, $q_i, q_j \in [1, N_s]$, denote the transition probabilities of the homogeneous \mathbf{S}_k process by $\mathbf{A} = (a_{mn})$:

$$a_{mn} = a_{(h,i),(l,j)} \triangleq P(\mathbf{S}_{k+1} = \mathbf{Q}_n | \mathbf{S}_k = \mathbf{Q}_m). \quad (2.10)$$

2.3. Aggregation

For computational purposes it is important to pose the question if it is possible to quantize the

time to the last transition to $N'_i < T$ states, with arbitrarily small error for N'_i suitably large? We propose to do so by 'aggregating' [5] the states $(\tau_{N'_i}, q_i), (\tau_{N'_i+1}, q_i), \dots, (\tau_{N_i}, q_i)$ into an 'aggregated' state

$$\mathbf{Q}_{N'_i,i} = \{(\tau_{N'_i}, q_i), (\tau_{N'_i+1}, q_i), \dots, (\tau_{N_i}, q_i)\}. \quad (2.11)$$

Consider a norm bound $\varepsilon \geq 0$ such that

$$|a_{ij}(t_k) - a_{ij}(t_{k-1})| < \varepsilon \quad \text{for } i, j \in [1, N_s] \\ \text{and } N'_i < t_k < N_i = T. \quad (2.12)$$

For the special case when $\varepsilon = 0$, our objective is accomplished by virtue of the following definition and theorem from [5].

DEFINITION. A Markov chain can be strictly aggregated with respect to aggregated states, if for every a priori (non-aggregated) state probability vector π , the aggregated process is a (first order) Markov chain and the transition probabilities do not depend on the choice of π .

THEOREM 2.1 (Strict aggregation property) [5]. *A necessary and sufficient condition for a Markov chain to be aggregated with respect to a set of aggregated states is that the transition probabilities have the same value for all possible transitions between any pair of aggregated states. These common transition probabilities form the transition matrix for the aggregated chain.*

ε aggregation property

The aggregation (2.11) under (2.12) with ε non-zero leads to errors that can be bound in terms of ε as follows:

$$|P(\mathbf{S}_k \in \mathbf{Q}_n | \mathbf{S}_{k-1} \in \mathbf{Q}_m, \dots, \mathbf{S}_1 \in \mathbf{Q}_l) \\ - P(\mathbf{S}_k \in \mathbf{Q}_n | \mathbf{S}_{k-1} \in \mathbf{Q}_m)| \leq \varepsilon. \quad (2.13)$$

We see that the aggregated process at time k depends on the states at time $k-2, \dots, 1$ only by an amount ε . Moreover, with $\varepsilon \rightarrow 0$, \mathbf{S}_k tends to a first order Markov chain. So according to the above definition and theorem, the Markov chain \mathbf{S}_k

can be aggregated with respect to the partition $\{(\tau_1, q_i), \dots, (\tau_{N_i-1}, q_i), \mathbf{Q}_{N'_i,i}\}$, where $\mathbf{Q}_{N'_i,i}$ is defined in (2.11). The fractal model with transition probabilities that asymptotically converge (see (2.5)) satisfies (2.12) and so can be aggregated. For the rest of this paper we set $N_i = N'_i \ll T$ where N'_i is suitably large to result in negligible error. Of course, now (2.8) is modified as

$$t_{k+1} = \begin{cases} t_k + 1 & \text{if } s_{k+1} = s_k \text{ and } t_k < N_i, \\ t_k & \text{if } s_{k+1} = s_k \text{ and } t_k = N_i, \\ 1 & \text{otherwise.} \end{cases} \quad (2.14)$$

Transition probability matrix sparseness

Notice that with $\tau_h = h$ for $h < N_i = N'_i$, the relationship between (2.7) and (2.10) for $1 \leq \tau_h < N_i$ is

$$a_{ii}(\tau_h) = a_{(h,i),(h+1,i)} \\ \text{and} \quad (2.15)$$

$$a_{ij}(\tau_h) = a_{(h,i),(1,j)}, \quad i \neq j.$$

Clearly \mathbf{A} has $(N_s N_i)^2 = N^2$ elements. However, since t_{k+1} is restricted as in (2.14) to only three possible values, simple calculations show that $N^2 - N_s^2 N_i$ elements of \mathbf{A} are zero. For $i, j \in [1, \dots, N_s]$, only the following elements of \mathbf{A} are not necessarily zero:

$$a_{(h,i),(h+1,i)}, \quad \tau_h \neq N_i; \\ a_{(h,i),(1,j)}, \quad i \neq j; \quad \text{and } a_{(N_i,i),(N_i,i)}. \quad (2.16)$$

Consequently, in any scheme to estimate \mathbf{A} , only the $N_s^2 N_i$ elements of \mathbf{A} in (2.16) need be estimated.

2.4. Hidden models

Fractal and Markov processes imbedded in noise are called hidden fractal models (HFMs) and hidden Markov models (HMMs), respectively. Consider that the fractal process s_k and hence the associated vector Markov process $\mathbf{S}_k = (t_k, s_k)$ are hidden, that is indirectly observed by measurements y_k . We denote the sequence y_1, y_2, \dots, y_k

by \mathbf{Y}_k . The vector of probability functions $\mathbf{b}(\cdot) = (b_m(\cdot)) = P(y_k | \mathbf{S}_k = \mathbf{Q}_m)$ where $\mathbf{Q}_m = (\tau_h, q_i)$ is assumed invariant of the times k and τ_h . So $b_m(y_k) = b_i(y_k) = P(y_k | s_k = q_i)$. Also assume the independence property $P(y_k | s_k = q_i, s_{k-1} = q_j, \mathbf{Y}_{k-1}) = P(y_k | s_k = q_i)$. Further assume that the initial state probability vector $\boldsymbol{\pi} = (\pi_m)$ is defined from $\pi_m = P(\mathbf{S}_1 = \mathbf{Q}_m)$. The transition probabilities \mathbf{A} are defined as in (2.10). The vector HMM for the \mathbf{S}_k process is denoted $\lambda = (\mathbf{A}, \mathbf{b}(\cdot), \boldsymbol{\pi})$. Of course λ also denotes the HFM for the \mathbf{S}_k process.

A special case of interest is when the measurements y_k consist of a signal corrupted by zero mean, normally distributed white noise as follows: $y_k = s_k + w_k$, $w_k \sim N[0, \sigma_w^2]$. Then $b_i(y_k) = (\sqrt{2\pi}\sigma_w)^{-1} \exp(-(y_k - q_i)^2 / (2\sigma_w^2))$.

2.5. Learning and estimation objectives

Given the fractal signal model reformulated as a Markov model as described above, and given observations y_1, y_2, \dots, y_k denoted \mathbf{Y}_k , there are three-related HMM problems which can be solved [7, 13] and interpreted to yield results for HFMs.

(1) Evaluate the likelihood of a given model λ_i generating the given sequence of data, denoted $P(\mathbf{Y}_k | \lambda_i)$. This allows comparison of a set of models $\{\lambda_i\}$ to select the most likely, given the observations. In simulation studies presented subsequently we use this likelihood function to compare different order Markov models and fractal models.

(2) Estimate signal statistics such as a posteriori probabilities for an assumed model λ and data sequence. For a fixed interval T , estimate

$$\begin{aligned} \gamma_{k|T} &= (\gamma_{k|T}(m)); \\ \gamma_{k|T}(m) &\triangleq P(\mathbf{S}_k = \mathbf{Q}_m | \mathbf{Y}_T, \lambda). \end{aligned} \quad (2.17)$$

From $\gamma_{k|T}(m)$, maximum a posteriori (MAP) state estimates conditioned on λ , denoted $\{\hat{\mathbf{S}}_k | \lambda\}$, are

$$\hat{\mathbf{S}}_k^{\text{MAP}} = \mathbf{Q}_n, \quad (2.18)$$

where

$$n = \arg \max_{1 \leq m \leq N} \gamma_{k|T}(m).$$

Of course if $\mathbf{Q}_n = (\tau_h, q_j)$ then $\hat{s}_k^{\text{MAP}} = q_j$ in (2.18).

(3) Processing of the observations based on a model assumption λ , and adjustment (re-estimation) of the model parameters (functions) $(\mathbf{A}, \mathbf{b}(\cdot), \boldsymbol{\pi})$, such that the likelihood of the updated model $\bar{\lambda} = (\bar{\mathbf{A}}, \bar{\mathbf{b}}(\cdot), \bar{\boldsymbol{\pi}})$ given the observation sequences, $P(\mathbf{Y}_T | \bar{\lambda}) > P(\mathbf{Y}_T | \lambda)$, and repeating until convergence. The objective is to achieve the most likely model λ^{ML} amongst the set $\lambda = (\mathbf{A}, \mathbf{b}(\cdot), \boldsymbol{\pi})$, given the observations.

3. Estimation and re-estimation

We now proceed to solve Problems 1 to 3 above for the HFM reformulated as an HMM. The formulae of this section turn out to be identical to more familiar ones for scalar HMMs, as in [3], but with scalars s_k replaced by 2-vectors \mathbf{S}_k and the states q_i, q_j replaced by $\mathbf{Q}_m, \mathbf{Q}_n$ which can be associated with the pair of scalar states (τ_h, q_i) and (τ_l, q_j) .

Forward-backward procedure

Recursive formulae for updating the forward variable α_k and the backward variable β_k are readily calculated [3, 13] for $\mathbf{Q}_n = (\tau_h, q_j)$ for any $1 \leq \tau_h < N_t$ as

$$\begin{aligned} \alpha_k(n) &= \left(\sum_{m=1}^N \alpha_{k-1}(m) a_{mn} \right) b_j(y_k), \\ \alpha_1(n) &= \pi_n b_j(y_1), \\ \beta_k(m) &= \sum_{n=1}^N a_{mn} b_j(y_{k+1}) \beta_{k+1}(n), \quad \beta_T(m) = 1. \end{aligned} \quad (3.1)$$

Updating α_k and β_k requires the order of $N^2 M$ multiplications and additions at each time instant k because only $N^2 M$ components of \mathbf{A} are not necessarily zero, see (2.16). Also to avoid numerical problems, appropriate scaling must be used as in [13].

The optimal a posteriori probabilities associated with Problem 2 are given from [3, 13] as

$$\gamma_k = (\gamma_k(m)); \quad \gamma_k(m) = \frac{\alpha_k(m)\beta_k(m)}{\sum_{m=1}^N \alpha_k(m), \beta_k(m)}. \quad (3.2)$$

Then using (2.18) we obtain MAP fixed-interval smoothed estimates. The likelihood function, which is the end result of the Problem 1, is calculated as

$$L_T \triangleq P(Y_T | \lambda) = \sum_{m=1}^N \alpha_T(m). \quad (3.3)$$

Baum-Welch re-estimation formulae

Re-estimation of \mathbf{A}, π are given from [13]

$$\begin{aligned} \bar{a}_{mn} &= \frac{\sum_{k=1}^{T-1} \zeta_k(m, n)}{\sum_{k=1}^{T-1} \sum_{n=1}^N \zeta_k(m, n)}; \\ \bar{\pi}_m &= \frac{\sum_{n=1}^N \zeta_1(m, n)}{\sum_{m=1}^N \sum_{n=1}^N \zeta_1(m, n)}, \end{aligned} \quad (3.4)$$

where $\zeta_k(m, n) = \alpha_k(m)a_{mn}b_n(y_{k+1})\beta_{k+1}(n)$. For updating $\mathbf{b}(\cdot)$ see [13]. It is proved in [7] that with the re-estimation formulae the model $\bar{\lambda} = (\bar{\mathbf{A}}, \bar{\mathbf{b}}(\cdot), \bar{\pi})$ is more likely than λ , given the observations, in that $P(Y_T | \bar{\lambda}) \geq P(Y_T | \lambda)$.

Fractal analysis

We propose two ways of computing the fractal dimensions and initial setpoints.

- (1) When an analytical expression for the transition probabilities of the semi-Markov process exists (e.g. the fractal dimensions of a fractal process are equal for the various states), the parameters of the function a_{ij} (see (2.7)) are obtained by fitting functions of the form a_{ij} to the Baum-Welch estimates $a_{(h,i),(h+1,j)}, 1 \leq \tau_h < N_i$ (see (2.15)) in a least squares sense. However, because of the finite length of the

observation sequence, for large values of τ_h it is often the case that there are not sufficient such transitions to get a reliable estimate of $a_{(h,i),(h+1,j)}$. Also, we ignore $a_{(1,i),(1,j)}$ since it includes all events outside the bandwidth of interest. Thus typically for a few thousand observations, it makes sense to use $a_{(h,i),(h+1,j)}$ only for $2 \leq \tau_h \leq 5$ (say) in the least squares fit.

- (2) When the fractal dimensions for the various states are unequal then using the same analysis techniques mentioned in Section 2.1 with the MAP state estimates (2.18) of the most likely model, the fractal dimensions and initial setpoints can be calculated.

4. Computer studies

4.1. Simulations and discussion

HFM data

Let us first work with simple scalar binary level processes ($N_s = 2$). We study various 10 000 point sample paths of fractal processes with fractal dimensions $D_i = 1.0, 1.2, 1.5, 1.8, i \in \{1, 2\}$. In generating the various sample paths, the NAG library random generator G05CAF was used. The states q_i are separated by intervals $\Delta q = 1$.

Transition probability estimates. The solid lines in Figs. 1a and 1b show the transition probability estimates $a_{(h,i),(h+1,i)}$ for $\tau_h = h, h = 1$ to 10 of the noise free sample paths. In Fig. 1a the initial setpoints were taken as $k_i = 0.255$. In Fig. 1b different initial setpoints were taken for different fractal dimensions: for $D_i = 1.0, 1.2, 1.5$ and 1.8, we selected $k_i = 1.95, 1.9, 0.7, 0.55$, respectively. (For comparison we also show the true transition probabilities.) Not surprisingly, with finite length data segments, for $\tau_h > 6$, the estimates of the transition probabilities associated with the generated noise free sample paths differ significantly from those of the fractal signal generating system. The differences increase with increasing τ_h . Also, increasing the data length to 100 000 for example, is not enough to eliminate such differences for large τ_h .

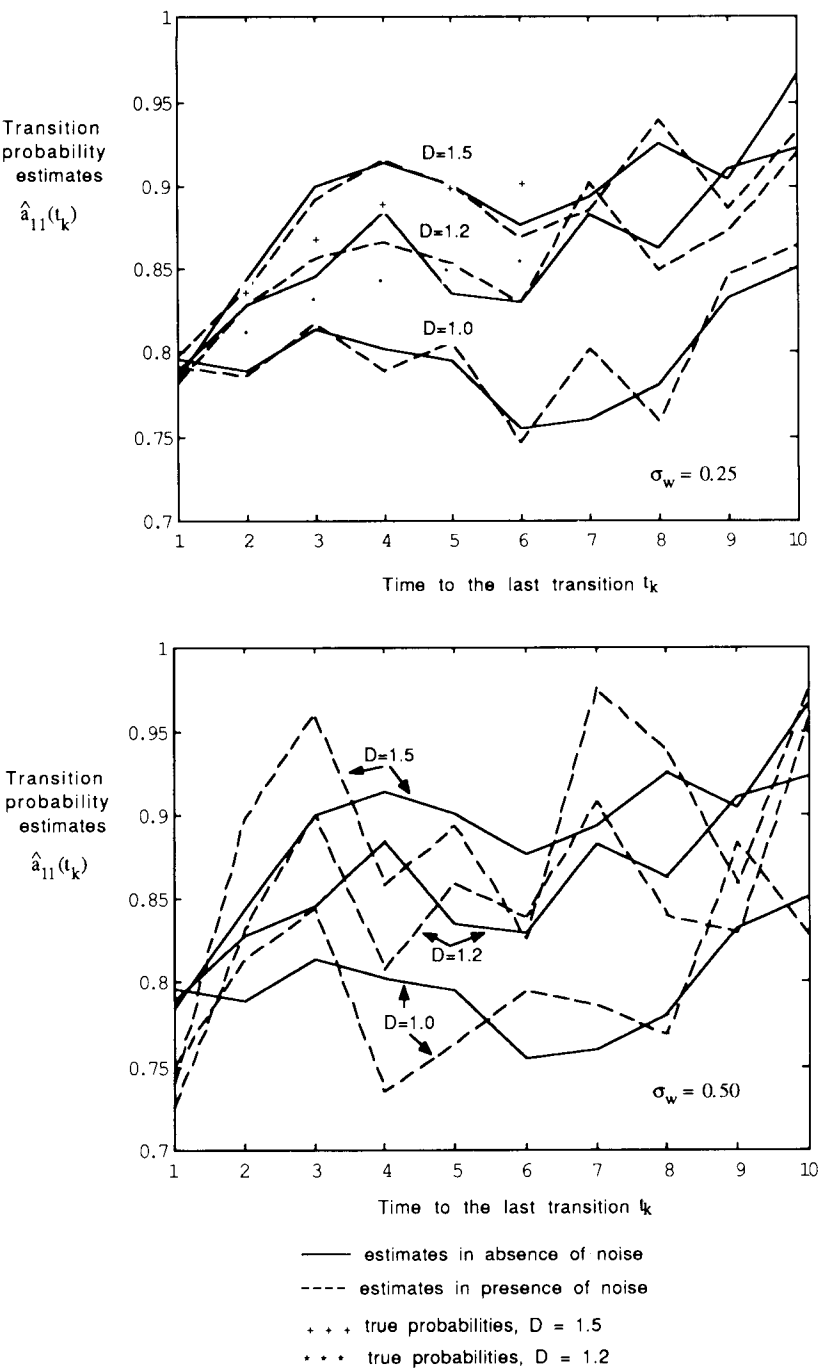


Fig. 1a. Estimates for computer generated HFM data with $k_i = 0.255$.

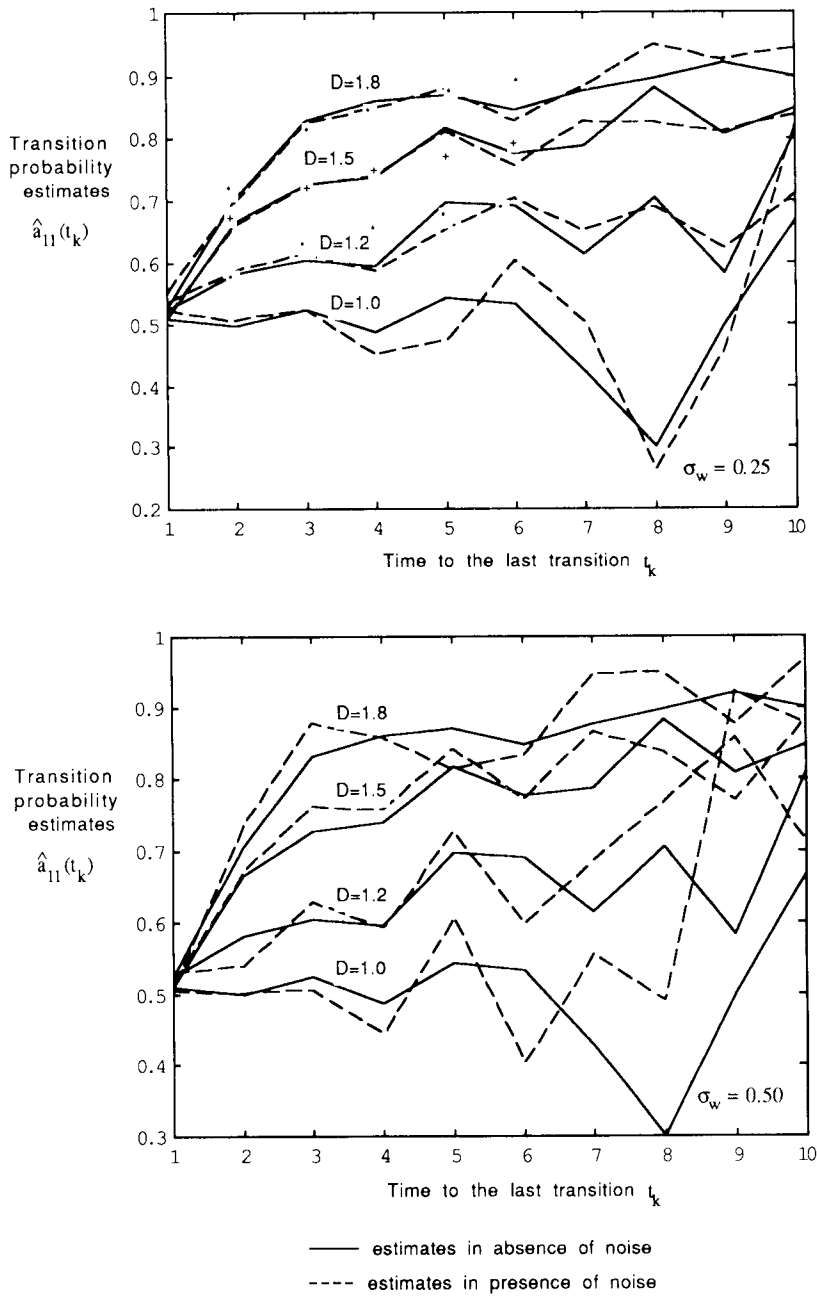


Fig. 1b. Estimates for computer generated HFM data with $k_i = 1.95, 1.9, 0.7, 0.55$ for $D = 1.0, 1.2, 1.5, 1.8$, respectively.

Further increases only detect inadequacies in the random number generator.

Actually, aggregation so that τ_h/h decreases with increasing h , would effectively smooth the curves so that they approach the exponential shape of the idealized fractal model. Case studies on this are omitted since they add no further insights. In the absence of this aggregation, the departure of the noise-free transition probability estimates of finite data length sequences from the idealized fractal model behaviour reveals the inadequacy of the random number generator. It points out the advantage of our approach of first allowing more general model estimates than fractal models, and then approximating these by fractal models.

To the computer generated 10 000 point fractal sample paths is added zero mean Gaussian noise with standard deviations σ_w of 0.25 and 0.5, respectively. The dotted lines in Figs. 1a and 1b show the transition probability estimates of these hidden fractal processes using 10 repeated applications of the Baum–Welch re-estimation formulae. Notice that the Baum–Welch estimates after 10 passes follow the noise free estimates reasonably closely. Of course as σ_w increases, the estimates deteriorate. The results show the robustness of our scheme in

the presence of white measurement noise with known mean and variance.

Model likelihood. We study in more detail the sample path generated from a fractal model with fractal dimensions $D_1=1.5$, and initial setpoints $k_i=0.7$ as described above. Based on the statistics of the generated noise free sample path over 10 000 points, the estimated fractal dimensions are $D_1=1.554$ and $D_2=1.508$. These are calculated by least squares fitting (2.6) to $a_{(h,i),(h+1,j)}$ for $1 \leq \tau_h < 6$. To the computer generated fractal model data was added zero mean Gaussian noise with standard deviations σ_w of 0.1, 0.25, 0.5 and 1. The states q_i are separated by intervals $\Delta q = 1$. (For completeness, after 10 successive applications of the Baum–Welch re-estimation formulae, \hat{D}_1 was calculated using a least squares fit to be 1.533 when $\sigma_w = 0.25$, and 1.450 when $\sigma_w = 0.5$).

To quantitatively compare how well various scalar HMM and HFM models fit the data, we calculate the likelihood function L (see (3.3)). Let us denote the most likely n th-order HMM estimate by $\lambda^{(n)}$ and the most likely HFM estimate with the time to the last transition taking N_t values as $\lambda_f^{(N_t)}$. Also define $L^{(n)} \triangleq P(Y_T | \lambda^{(n)})/P(Y_T | \lambda^{(1)})$ and $L_f^{(N_t)} \triangleq P(Y_T | \lambda_f^{(N_t)})/P(Y_T | \lambda^{(1)})$. Figure 2a shows

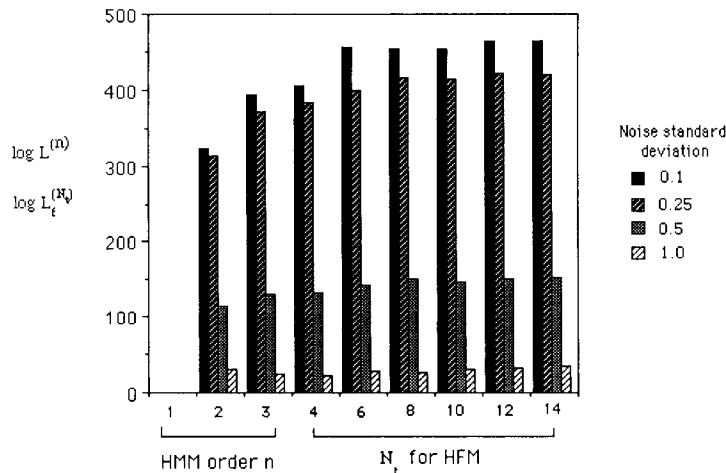


Fig. 2a. Comparison of HMMs and HFMs for HFM data in terms of log likelihood functions $\log L^{(n)}$ and $\log L_f^{(N_t)}$ normalized to the first order HMM case.

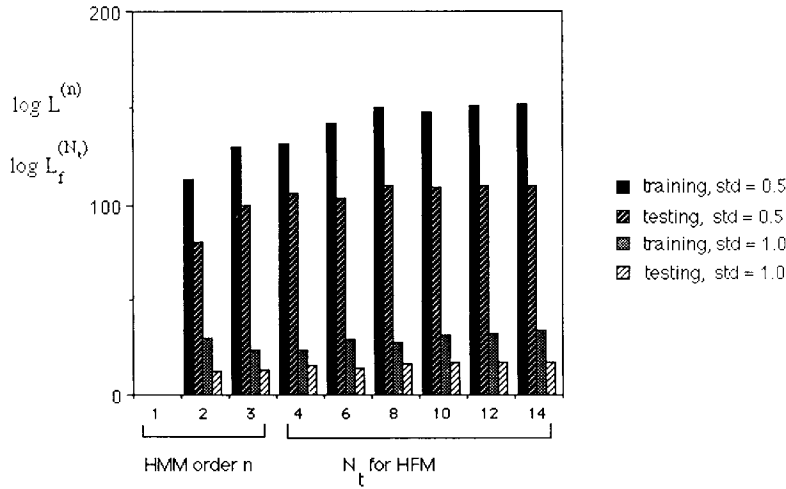


Fig. 2b. Comparison of improvement in log likelihood function for training and testing data sets.

$\log(L^{(n)})$ for $n = 1$ to 3 and $\log(L_f^{(N_t)})$ for N_t taking on values between 4 and 14. We can draw the following conclusions:

- (1) When the noise has a small variance, there is a dramatic increase in $L^{(n)}$ with increasing n . This shows that higher order HMMs model the HFM data much better than lower order HMMs. The increase in $L^{(n)}$ with increasing n , however, is less significant when the variance of the noise increases. Notice also that $\log(L^{(2)} - L^{(1)})$ is much greater than $\log(L^{(3)} - L^{(2)})$. Simulations suggest a decaying 'exponential' improvement in $\log(L^{(n)})$ with increasing n . The noise variance determines the rate of 'exponential' decay.
- (2) Simulations show that when using an HFM model, $L_f^{(N_t)}$ increases with increasing N_t . Again, the relative increase decreases with increasing noise variance. Also in our example there is no significant improvement for $N_t > 6$. Since computational cost is proportional to N_t , there is incentive to use a minimum N_t without sacrificing significantly on the likelihood function.
- (3) In all cases the convergence rate of the likelihood function to its final value decreases with increasing noise variance.
- (4) Simulations not reported here also show that the larger the fractal dimension and the initial

setpoints, the greater the improvement in the likelihood function when using higher order HMMS and HFMs compared to first order HMM processing.

The estimated models from the above sample paths (training data) were used to see how well they fit other sample paths with the same statistics (test data). For $\sigma_w = 0.1, 0.25$ the increase in $L(n)$ and $L_f^{N_t}$ with increasing n and N_t for the test data were very close to that for the training data. Figure 2b compares the relative improvements for $\sigma_w = 0.5, 1$. Notice that there is still an increase in $L(n)$ and $L_f^{N_t}$ with increasing n and N_t ; however it is less than that for the training data. Thus with increasing σ_w , the increase in $L(n)$ and $L_f^{N_t}$ with n, N_t for the test data gets smaller compared to the training data.

Signal estimates. Figure 3a shows a typical segment (200 points) of the above computer generated fractal signal s_k . Zero mean, white Gaussian noise with standard deviation $\sigma_w = 0.5$ is added to s_k yielding the observations y_k . Using these observations, HFM processing is used to obtain the MAP signal estimates \hat{s}_k , see (2.18).

Figure 3b compares the signal estimation capabilities of HFM processing with first order scalar HMM processing for fractal processes corrupted by zero mean white Gaussian noise of different variances. The number of errors per 1000

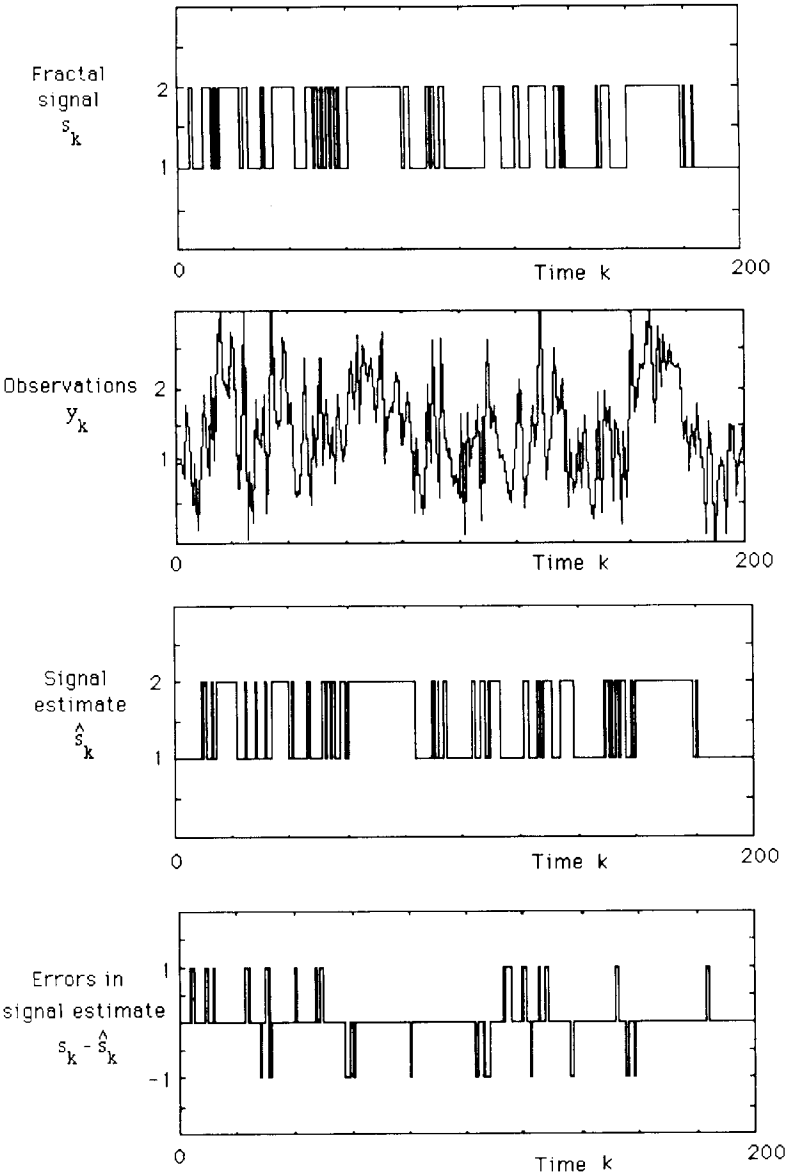


Fig. 3a. HFM signal estimation.

points is plotted versus the noise standard deviation normalized to the state spacing level Δq . The signal estimates were obtained as the MAP signal estimates after 10 successive applications of the Baum–Welch re-estimation formulae.

Notice in Fig. 3b that the HFM signal estimates are significantly better than the first order HMM estimates when the fractal dimension D_1 is relatively large (e.g., $D_1 = 1.95$). For relatively small

D_1 (e.g., $D_1 = 1.4$), however, despite large variations in the likelihood functions, simulations show that the actual number of errors in the signal estimates do not differ significantly with different HMM and HFM model classes. This is not surprising since for low values of the fractal dimension and initial setpoint, $a_{ij}(t_k)$ does not vary significantly with t_k and so the HFM approaches a first order HMM. This demonstrates the robustness

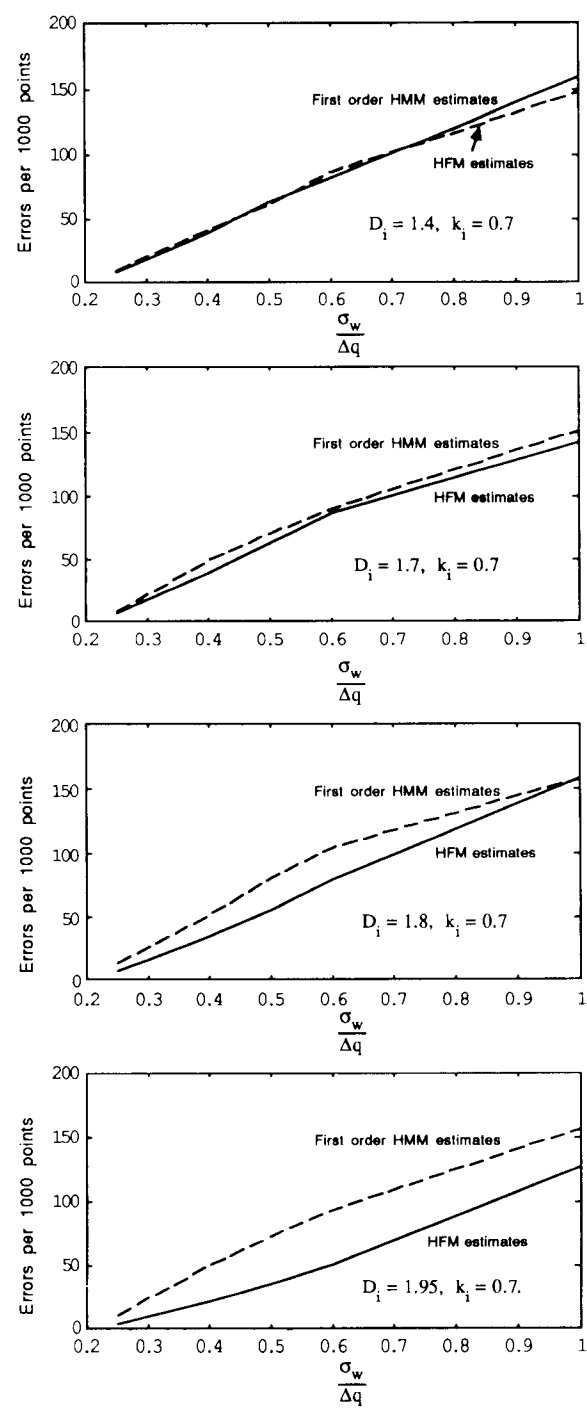


Fig. 3b. Signal estimate error performance.

of lower order HMM schemes in signal estimation. Thus for signal estimation purposes of HFMs with low initial setpoints and fractal dimensions, scalar first order HMM processing remains relatively attractive compared to HFM or higher order HMM processing due to its computational efficiency.

Comparison with histogram evaluation of HFM data

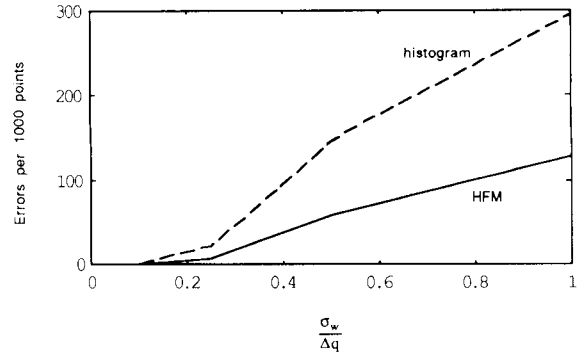
Here, we compare our HFM techniques with a conventional fractal analysis method in Section 2.1 which neglects noise.

We used the techniques described in Section 2.1 on the above hidden fractal data to estimate D_i in a least squares sense. Not surprisingly, with increasing noise variance, both the error probability (the number of errors in state estimates divided by the number of observations) as well as the error \hat{D}_i in the estimate of D_i increase dramatically when using this technique (see Fig. 4). This illustrates the usefulness of the HFM schemes in noisy environments. Notice that with increasing noise variance, the histogram techniques estimate D_i as being close to 1, irrespective of their true value (see Fig. 4(b)). This suggests that in noisy environments the conventional analysis technique wrongly models fractal processes as homogeneous scalar first order Markov processes.

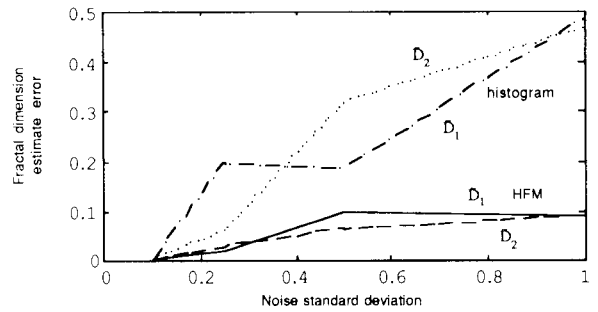
As discussed above, there is no significant difference in the number of errors in signal estimates with different classes of HMM and HFM models with fractal dimensions close to 1. Also, scalar first order HMM processing is much more computationally efficient than higher order HMM or HFM processing. Is it then possible to use first order HMM processing to obtain signal estimates and then use the histogram techniques described above? Simulations show that this does not yield results comparable to HFM processing unless the noise standard deviation is significantly smaller than the state level spacing (usually less than a quarter of the state level spacing).

First-order HMM data

From the conclusions in the previous subsection it might be inferred that using a larger n or N_t is



(a) : Error Performance



(b) : Error in fractal dimension estimates

Fig. 4. Comparison of HFM and existing histogram schemes.

better in modelling an observation sequence (in terms of the likelihood function) than a smaller n or N_t . However, this is not so as illustrated below.

A 10 000 two-state ($N_s = 2$) scalar first-order Markov chain with transition probabilities $a_{ii} = 0.8$, $a_{ij} = 0.2$, $i \neq j$ was generated with states separated by $\Delta q = 1$. To this was added zero mean Gaussian noise with standard deviation σ_w of 0.25. Simulations show that there is no increase in $L^{(n)}$ or $L_f^{(N_t)}$ with increasing n and N_t .

Thus fitting HMMs and HFMs to an observation sequence is not analogous to polynomial fitting where a higher order polynomial always improves the fit.

4.2. Cell membrane data

The data to model currents through ion channels in cell membranes was obtained as follows: An outside-out membrane patch was excised from a

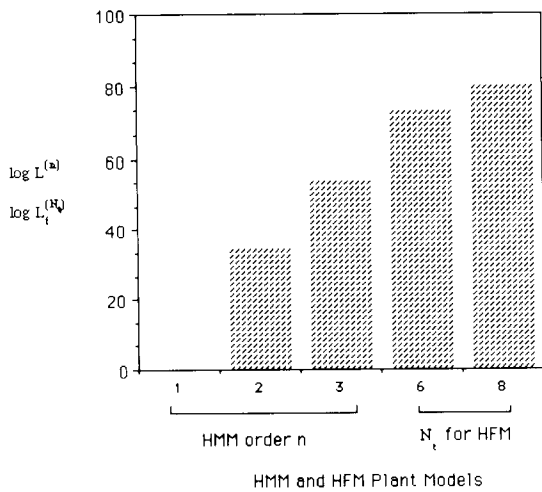


Fig. 5. Comparison of HMMs and HFMs for ion channel data.

cultured hippocampal neuron. In response to the application of $10\text{ }\mu\text{M}$ γ -aminobutyric acid, the channel was activated to allow chloride ions to cross the membrane [3]. The resulting analog signals were then digitized by lowpass filtering to 5 kHz using a 4-pole Bessel filter and sampling at 10 kHz.

(1) *Comparison.* We compared the performance of first and higher order HMM and HFM schemes in modelling the above data. Figure 5 shows the performance in terms of the likelihood functions $L^{(n)}$ and $L_t^{(N_t)}$ for a 10 000 point data segment. From Fig. 5, we conclude that using a higher order HMM or an HFM with a larger N_t fits the observations better than a first order HMM.

(2) *Determining the fractal dimensions.* Here we compute the fractal dimensions using the second method in Section 3.3. The open and closed durations measured from the MAP signal estimates were used to construct the histograms of various bin sizes. Figure 6(a) shows the histograms for the open state for different bin sizes. The lines on each histogram are obtained from least squares fits using the second through fourth bins. The negative value of the slopes are the effective rate constants k_{eff} . Figures 7a and 7b are plots of $\log k_{\text{eff}}$ versus $\log t_{\text{eff}}$ for the open and closed times. From Figs. 7a and 7b the fractal dimensions for the open and

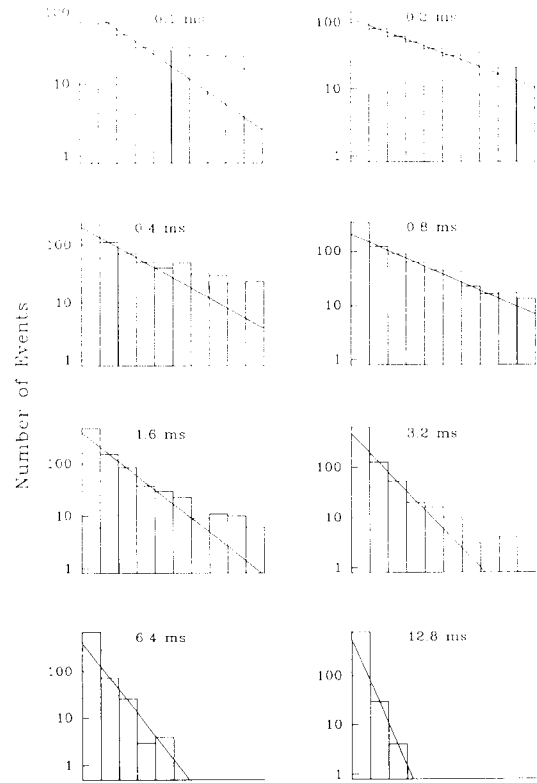


Fig. 6. Open-time histograms.

closed states are calculated to be 1.625 and 1.868. The approximately straight lines in Figs. 7a and 7b show that these channels can be reasonably represented by a model with fractal kinetics.

We do *not* claim here that HFMs for ion channels in cell membranes are better than HMMs in *all* cases. Whether the gating kinetics of single channel currents can best be described by Markov or fractal models is in dispute, see [6, 12]. We propose that the HFM techniques be employed in extensive studies to identify the approximate model class most consistent with the gating behaviour of single channel currents.

5. Conclusions

Existing techniques for estimating fractal stochastic processes assume that there is insignificant noise present in the observations. In this

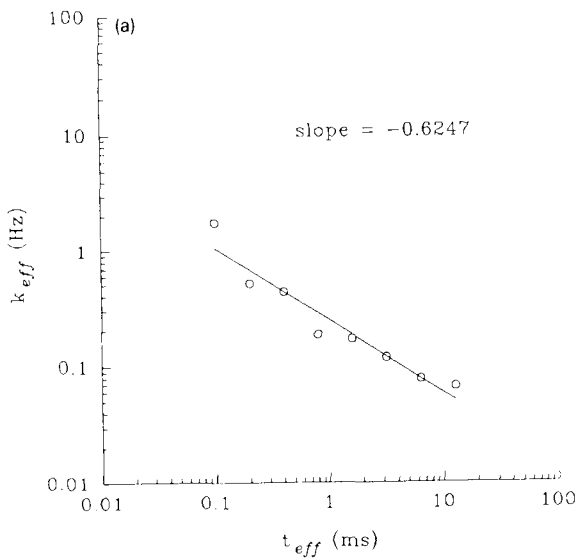


Fig. 7a. Determining the fractal dimension of open state.

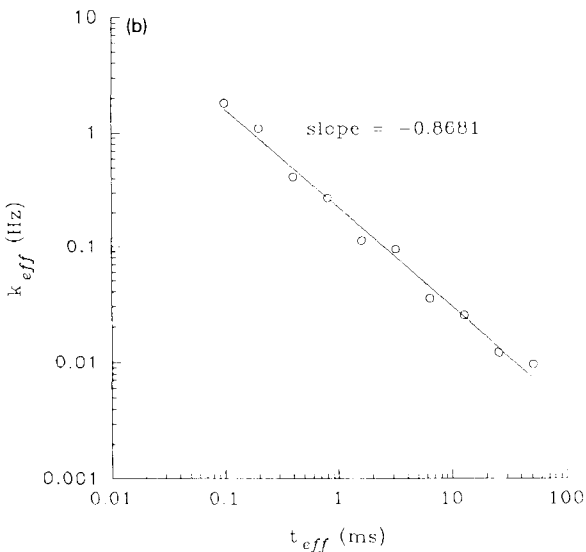


Fig. 7b. Determining the fractal dimension of closed state.

paper we have proposed schemes for estimating fractal stochastic processes with discrete levels which are imbedded in white (or near white) noise. The key is to reformulate the signal model as a hidden semi-Markov model and apply known estimation techniques. These yield optimal estimates of the signals and signal model parameters, includ-

ing level transition probabilities and noise statistics. Also the fractal dimensions can be estimated. We have demonstrated their effectiveness by means of simulation studies and also illustrated their application to a biological signal processing problem.

References

- [1] L.E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains", *Ann. Math. Statist.*, Vol. 37, 1966, pp. 1554-1563.
- [2] L.E. Baum, T. Petrie, G. Soules and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains", *Ann. Math. Statist.*, Vol. 41, No. 1, 1970, pp. 164-171.
- [3] S.H. Chung, J.B. Moore, L. Xia, L.S. Premkumar and P.W. Gage, "Hidden Markov model techniques for extracting small ionic currents from noise", *Phil. Trans. Roy. Soc. London Ser. B*, Vol. 329, 1990, pp. 265-285.
- [4] D.R. Cox, *Renewal Theory*, Science Paperbacks, London, 1962.
- [5] J.G. Kemeny and J.L. Snell, *Finite Markov Chains*, Van Nostrand, Reinhold, New York, 1960.
- [6] S.J. Korn and R. Horn, "Statistical discrimination of fractal and Markov models of single-channel gating", *Biophys. J. Biophys. Soc.*, Vol. 54, November 1988, pp. 871-877.
- [7] S.E. Levinson, L.R. Rabiner and M.M. Sondhi, "An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition", *The Bell System Techn. J.*, Vol. 62, No. 4, 1983, pp. 1035-1074.
- [8] L.S. Liebovitch and J.M. Sullivan, "Fractal analysis of a voltage-dependent potassium channel from cultured mouse hippocampal neurons", *Biophys. J. Biophys. Soc.*, Vol. 52, December 1987, pp. 979-988.
- [9] L.S. Liebovitch et al., "Fractal model of ion-channel kinetics", *Biochimica et Biophysica Acta, Elsevier*, Vol. 896, 1987, pp. 173-180.
- [10] L.S. Liebovitch et al., "Ion channel kinetics: A model based on fractal scaling rather than multi-state Markov processes", *Math. Biosci.*, Vol. 84, 1987, pp. 37-68.
- [11] B.B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, 1983.
- [12] O.B. McManus et al., "Fractal models are inadequate for the kinetics of four different channels", *Biophys. J. Biophys. Soc.*, Vol. 54, November 1988, pp. 859-870.
- [13] L.R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", *Proc. IEEE*, Vol. 77, No. 2, 1989, pp. 257-285.
- [14] B. Sakmann and E. Naher, eds., *Single Channel Recording*, Plenum, New York, 1983.