Ann. Inst. Statist. Math.
Vol. 55, No. 2, 423–436 (2003)

NONPARAMETRIC ADAPTIVE DETECTION IN FADING CHANNELS BASED ON SEQUENTIAL MONTE CARLO AND BAYESIAN MODEL AVERAGING*

DONG GUO[1], XIAODONG WANG[1] AND RONG CHEN[2]

[1]*Department of Electrical Engineering, Columbia University, New York, NY 10027, U.S.A.*
[2]*Department of Information and Decision Science, University of Illinois at Chicago, Chicago, IL 60607, U.S.A. and Department of Business Statistics and Econometrics, Peking University, Beijing, China*

(Received May 29, 2002; revised February 12, 2003)

Abstract. Recently, a Bayesian receiver for blind detection in fading channels has been proposed by Chen, Wang and Liu (2000, *IEEE Trans. Inform. Theory*, **46**, 2079–2094), based on the sequential Monte Carlo methodology. That work is built on a parametric modelling of the fading process in the form of a state-space model, and assumes the knowledge of the second-order statistics of the fading channel. In this paper, we develop a nonparametric approach to the problem of blind detection in fading channels, without assuming any knowledge of the channel statistics. The basic idea is to decompose the fading process using a wavelet basis, and to use the sequential Monte Carlo technique to track both the wavelet coefficients and the transmitted symbols. Moreover, the algorithm is adaptive to time varying speed/smoothness in the fading process and the uncertainty on the number of wavelet coefficients (shrinkage order) needed. Simulation results are provided to demonstrate the excellent performance of the proposed blind adaptive receivers.

Key words and phrases: Fading channel, wavelet, adaptive shrinkage, Bayesian model averaging, sequential Monte Carlo, resampling.

1. Introduction

Signal detection in fading channels has been a key problem in communications and an array of methodologies have been developed to tackle this problem. Specifically, the optimal detector for flat-fading channels with known channel statistics are studied in Haeb and Meyr (1989) and Lodge and Moher (1990), which has a prohibitively high complexity. Suboptimal receivers in fading channels often employ a two-stage structure, with a channel estimation stage followed by a sequence detection stage. Other approaches include the method based on a combination of hidden Markov model and Kalman filtering in Collings and Moore (1994), and the method based on the expectation-maximization (EM) algorithm (Georghiades and Han (1997)).

Recently, Chen *et al.* (2000) develop a blind Bayesian receiver for flat-fading channels. It is based on the powerful sequential Monte Carlo (SMC) technique for numerical Bayesian computation. It achieves near-optimum performance without the use of any

*This work was supported in part by the U.S. National Science Foundation (NSF) under grants CCR-9875314, CCR-9980599, DMS-9982846, DMS-0073651 and DMS-0073601.

training/pilot symbols or decision feedback. They assume that the fading channel process follows a linear dynamic model (i.e., ARMA model), and the model parameters are known to the receiver. However, some practical fading processes exhibit spectral characteristics that require a very-high order ARMA model to fit. Moreover, in some applications, the channel fading statistics may not be known to the receiver at all. Hence in this paper, we address the problem of blind adaptive detection in fading channels with unknown channel statistics.

Our approach is to decompose the fading process using a wavelet basis, and then to use the SMC technique to estimate both the wavelet coefficients and the data symbols. Wavelet-based signal processing enjoys a very strong optimality property for general inverse problems in that their use can achieve accurate and parsimonious representation of the signal of interest. Some recent works have addressed the use of wavelet to model fading channels (Martone (2000)). In these methods, the shrinkage order is fixed *a priori*, and the wavelet coefficients are obtained by using training symbols and standard adaptive algorithms (e.g., LMS, RLS). Here our wavelet-based SMC receiver is blind in nature, i.e., without using any training symbols.

It is noted that although wavelet decomposition can perfectly reconstruct any sequence of finite length, truncation is often needed to filter out the noise, to achieve parsimony and to obtain better prediction. The shrinkage order used also reflects the trade-off between bias and smoothness. In fact, the shrinkage order directly controls the smoothness of the wavelet approximation. For a smooth underlying process, fewer number of wavelet coefficients are needed. When the underlying process is relatively volatile (e.g., large second derivative), a large number of wavelet coefficients are needed.

It is noted that fading process often exhibit time varying characteristics, particularly its speed or smoothness. For example, it may be very smooth for a period of time then change to a period of fast fading. To more accurately track and predict the fading process to ensure high reliability in signal extraction, we propose an adaptive algorithm that allows the change of shrinkage order in the wavelet approximation. Simulation has shown it provides significant improvement over non-adaptive algorithms.

The rest of this paper is organized as follows. The time varying wavelet representation of fading channels is discussed in Section 2. The nonparametric adaptive blind SMC receiver for flat-fading channels is presented in Section 3. Simulation results are provided in Section 4. Section 5 contains the conclusions.

2. Problem formulation

2.1 *Flat-fading channel model*

Consider a discrete-time baseband communication system signaling through a flat-fading channel with additive white Gaussian noise. The transmitted data symbols $\{s_t\}$ take values from a finite alphabet set $\mathcal{A} = \{a_1, a_2, \ldots, a_{|\mathcal{A}|}\}$. The input-output relationship is given by

$$y_t = s_t \alpha_t + e_t, \qquad t = 1, 2, \ldots \tag{2.1}$$

where y_t, α_t and e_t are respectively the received signal, the fading coefficient and the noise sample at time t. It is assumed that the processes $\{\alpha_t\}$, $\{s_t\}$ and $\{e_t\}$ are mutually independent and e_t assumes a complex Gaussian distribution,

$$e_t \sim \mathcal{N}_c(0, \sigma^2). \tag{2.2}$$

The fading process is assumed to be Rayleigh, that is, $\{\alpha_t\}$ is a zero-mean complex Gaussian process with a Jakes' autocorrelation function given by (Proakis (1995))

$$E\{\alpha_t \alpha_{t+j}^*\} = J_0(2\pi f_d T j), \tag{2.3}$$

where $J_0(\cdot)$ is the Bessel function of the first kind and zeroth order, f_d is the maximum Doppler shift, and T is the symbol interval. Note that in Chen *et al.* (2000), it is assumed that the fading process $\{\alpha_t\}$ follows an ARMA model. However, for practical fading processes, e.g., Jakes' fading processes, very-high order models are needed to fit the fading spectrum given by (2.3). Hence, in this paper, we drop such a model assumption and treat the general fading processes via a non-parametric approach using the wavelet decomposition.

2.2 *Wavelet representation of fading processes*

Consider the following wavelet regression representation of a segment of the fading process $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_K]^T$,

$$\boldsymbol{\alpha} = \tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{x}},$$

where $\tilde{\boldsymbol{x}}$ is the discrete wavelet transform (DWT) of $\boldsymbol{\alpha}$, and $\tilde{\boldsymbol{\Phi}}^T$ is the orthogonal matrix corresponding to the discrete wavelet transformation (Daubechies (1988, 1992)). The detailed construction of the matrix $\tilde{\boldsymbol{\Phi}}$ taking into account the edge effects can be found in (Guo *et al.* (2002)). Since the wavelet representation exhibits several useful properties, such as orthogonality, compact support, varying degrees of smoothness, and localization in time and scale (frequency), only a few large coefficients explain most of the functional form in the fading process, while the remaining majority are comparatively small and therefore can be discarded (Donoho and Johnstone (1994)). In addition, for any possibility of estimating s_t given y_t only, it is required the fading process to be smooth with certain predictability. When too many wavelet terms are used in the approximation, the process may become less smooth and one may loss the ability to estimate the symbol s_t. Note that in the extreme case when the full order is used, there always exists a $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{S}}$ such that

$$\boldsymbol{\alpha}^*(\boldsymbol{S}) = \tilde{\boldsymbol{\Phi}}_{\boldsymbol{S}}\tilde{\boldsymbol{x}},$$

holds perfectly for any sequence of $\boldsymbol{S} = [s_1, \ldots, s_K]^T$, where $\boldsymbol{\alpha}^*(\boldsymbol{S}) = [\alpha_1 s_1, \ldots, \alpha_K s_K]^T$. Hence $p(y_t \mid \tilde{\boldsymbol{x}}, \boldsymbol{S})$ does not depend on $\boldsymbol{S}$ and we loss all the ability to make inference on s_t.

The relationship between fading speed and the number of wavelet terms needed for proper approximation is demonstrated by the following numerical example. In this example, we choose the length of the fading process segment $\boldsymbol{\alpha}$ as 128, the Daubechies filter with order 2, and the decomposition level 7. Then the size of the wavelet coefficients $\tilde{\boldsymbol{x}}$ is 143. For a given fading process realization, we compute the wavelet coefficients $\tilde{\boldsymbol{x}}$, and then truncate it by keeping only the first κ elements. Hence, the fading process $\boldsymbol{\alpha}$ is approximated by

$$\hat{\boldsymbol{\alpha}}^{\kappa} = \tilde{\boldsymbol{\Phi}}[:, 1:\kappa]\tilde{\boldsymbol{x}}[1:\kappa].$$

The approximation error is then

$$\varepsilon^{\kappa} = \frac{1}{K} E\{\|\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}^{\kappa}\|^2\}.$$

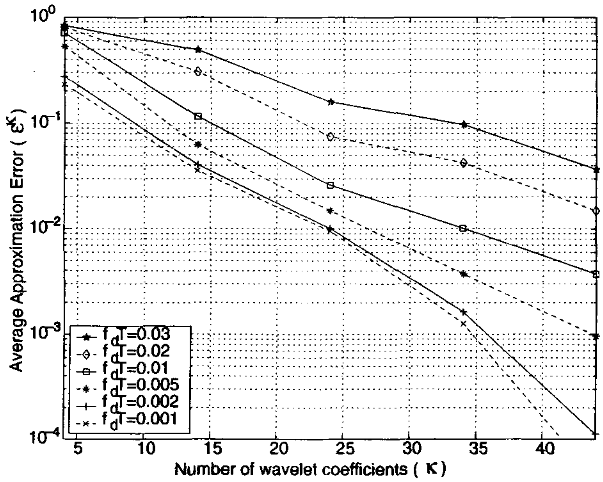

Fig. 1. The average approximation error versus the number of wavelet coefficients in the wavelet representation of the fading processes. The Daubechies filter with order 2 is used.

In Fig. 1, we plot ε^{κ} as a function of the number of wavelet coefficients for different values of normalized Doppler shift f_dT. It is seen that in general, for a fixed approximation error, the slower the fading process is, the less wavelet coefficients are needed to approximate the fading process. For fading processes with fading rate $f_dT \leq 0.01$, with 32 wavelet coefficients, the approximation error is below -20 dB. For a very fast fading process, e.g., $f_dT > 0.01$, more wavelet coefficients are needed to well approximate it. Similar observations are made for Daubechies filters with order higher than two. Various wavelet shrinkage methods exist for choosing the shrinkage order, such as the hard shrinkage method, the visual shrinkage method (Donoho and Johnstone (1994)), and the adaptive Bayesian shrinkage method (Chipman *et al.* (1997); Clyde and George (2000)).

Due to the time varying nature of the fading process, here we use an adaptive wavelet representation. Specifically, we assume the following model in state space form.

$$\text{(observation equation):} \qquad y_t = s_t \boldsymbol{\phi}_t \boldsymbol{x}_t + e_t$$

where $\boldsymbol{\phi}_t = \tilde{\boldsymbol{\phi}}_t[1 : \kappa_t]$, the t-th row of the fixed wavelet transform matrix $\tilde{\boldsymbol{\Phi}}$, and $\boldsymbol{x}_t = (x_{t,1}, \ldots, x_{t,\kappa_t})$. The state variables include κ_t, the shrinkage order at time t and $\boldsymbol{x}_t$, the wavelet coefficient at time t and s_t, the transmitted symbol at time t. Assume that the time-varying behavior is smooth, we adopt the following models for the state variables.

1. We assume κ_t follows a random walk in the interval $[\kappa_{\min}, \kappa_{\max}]$, with transition probability

$$P(\kappa_t = i + \nu \mid \kappa_{t-1} = i) = p_{i\nu}, \qquad \text{for} \quad \nu \in \{-1, 0, 1\}.$$

2. There is a random perturbation in the wavelet coefficients as the system evolves,

$$\begin{aligned} x_{t,i} &= x_{t-1,i} + \epsilon_{t,i}, \qquad \text{for} \quad i = 1, \ldots, \min\{\kappa_t, \kappa_{t-1}\} \\ \epsilon_{t,i} &\sim N(0, \sigma_x^2) \quad \text{independent for all } i. \end{aligned}$$

When $\kappa_t = \kappa_{t-1} + 1$, assign x_{t,κ_t} a prior distribution of $N(0, \sigma_x^2)$. That is

$$x_{t,\kappa_t} = \epsilon_{t,\kappa_t}$$

where $\epsilon_{t,\kappa_t} \sim N(0, \sigma_x^2)$, independent of $\epsilon_{t,k}$, $k = 1, \ldots, \kappa_{t-1}$.

In summary, we have

$$\boldsymbol{x}_t = \boldsymbol{x}_{t-1}^* + \epsilon_t \tag{2.4}$$

where

$$\boldsymbol{x}_{t-1}^* = \begin{cases} \boldsymbol{x}_{t-1} & \text{if} \quad \kappa_t = \kappa_{t-1} \\ \boldsymbol{x}_{t-1}[1 : \kappa_t] & \text{if} \quad \kappa_t = \kappa_{t-1} - 1 \\ [\boldsymbol{x}_{t-1} \ 0] & \text{if} \quad \kappa_t = \kappa_{t-1} + 1. \end{cases}$$

3. The transmitted symbol s_t is *i.i.d*, taking values in $\mathcal{A}$.

Remark 1. The wavelet transform matrix is ordered in terms of the smoothness. Hence, when fading becomes faster, κ_t tends to increase. Since we assume the change is smooth, we only allow κ_t to move one step at a time. Also, to ensure stability, the prior probability of 'stay', $P(\kappa_t = \kappa_{t-1})$, should be given a relative large value such as 0.8.

Remark 2. Because the wavelet transform matrix is orthogonal and the change in the fading process is smooth, it is reasonable to assume that, when κ_t increases, the wavelet coefficients (except the new one) will remain relatively stable. Also, to avoid large changes in the fading process, the new coefficient should be relatively small, achieved by imposing small σ_x^2.

Remark 3. The transmitted symbol s_t can adopt certain Markov property without complicating the extraction algorithm.

Remark 4. The above state space model is in fact a conditional linear dynamic model (Chen and Liu (2000)). Given a trajectory of the sequence $\{s_t, \kappa_t\}_{t=1}^n$, the system is linear and Gaussian.

Remark 5. The time-varying setting is useful even when the system is not really time-varying. It is known that the SMC tends to provide inaccurate results when some of the state variables (in our case, the shrinkage order κ and the wavelet coefficients $\boldsymbol{x}_t$) are fixed (Andrieu *et al.* (1999); Liu and West (2001)). Allowing small disturbances in the system, SMC is able to make movements in process. It is not necessary to use the random walk model for the random disturbances, except for its convenience and simplicity. Under this setting the variance of the prior distribution of x_t increases with t, though the posterior distribution gets compensated with more observations y_t. An Oerstein-Uhlenbeck process in the form of

$$x_{t,i} = a x_{t-1,i} + \sqrt{1 - a^2} e_{t,i}$$

provides a stationary prior for x_t and may result in better solutions.

Remark 6. Note that it is possible to specify a suitable shrinkage prior to avoid the crisp variable selection procedure. That is, one can use all the variables in the model, but imposes strong priors on the higher order wavelet coefficients to be close to zero (e.g., normal priors with mean zero and very small variances). It is a much simpler procedure and provides 'soft selection'. However, this procedure shifts the burden of model selection to prior specification, which may not be an easy task. Comparison between the two approaches is an interesting question, but out of the scope of this paper.

3. The nonparametric adaptive blind SMC receiver

Sequential Monte Carlo (SMC) method can be used to perform on-line filtering of nonlinear and non-Gaussian dynamic system. It utilizes the important concept of importance sampling, combined with sequential updating mechanism, to perform Monte Carlo estimation of the unobserved underlying state variables in a stochastic dynamic system. For detailed information and a wide range of applications, see Liu (2001) and Doucet *et al.* (2001).

For a special class of the state space models, the conditional linear dynamic models (CDLM), Chen and Liu (2000) proposed an efficient SMC algorithm, the mixture Kalman filter (MKF). The CDLM becomes a linear and Gaussian system, given a trajectory of a subset of the state variables (indicators). Using this feature, MKF generates Monte Carlo samples only in the indicator space, and marginalizes out the rest of the state variables using Kalman filter. It has been shown to be very efficient in dealing with fading channels (Chen *et al.* (2000)) and other applications (Andrieu *et al.* (2000), Liu and Chen (1995)).

Denote $\boldsymbol{Y}_t = (y_1, \ldots, y_t)$, $\boldsymbol{S}_t = (s_1, \ldots, s_t)$ and $\boldsymbol{K}_t = (\kappa_1, \ldots, \kappa_t)$. We apply the SMC method to the problem of on-line estimation of the *a posteriori* probability of the symbol s_t based on the received signals up to time t, without knowing the fading process α_t. That is, at time t, we need to estimate

$$p(s_t = a_i \mid \boldsymbol{Y}_t), \qquad a_i \in \mathcal{A}. \tag{3.1}$$

Then a hard MAP (maximum *a posteriori*) decision on symbol s_t is given by

$$\hat{s}_t = \arg\max_{a_i \in \mathcal{A}} p(s_t = a_i \mid \boldsymbol{Y}_t).$$

In order to implement the SMC, we need to obtain a set of Monte Carlo samples of the transmitted symbols, $\{(\boldsymbol{S}_t^{(j)}, w_t^{(j)})\}_{j=1}^m$, properly weighted with respect to $p(\boldsymbol{S}_t \mid \boldsymbol{Y}_t)$. Then the *a posteriori* symbol probability in (3.1) is approximated by

$$p(s_t = a_i \mid \boldsymbol{Y}_t) \cong \frac{1}{W_t} \sum_{j=1}^{m} 1(s_t^{(j)} = a_i) w_t^{(j)}, \qquad a_i \in \mathcal{A},$$

with $W_t \triangleq \sum_{j=1}^m w_t^{(j)}$.

The model in Section 2 is a CDLM where the indicator variables are (s_t, κ_t). It is easily seen that given the entire trajectory $(\boldsymbol{S}_t, \boldsymbol{K}_t)$, the system is linear and Gaussian with y_t as the observation and $\boldsymbol{x}_t$ as the state variable.

Following the MKF algorithm (Chen and Liu (2000)), we construct an adaptive blind receiver as follows. Assume at time $t-1$ we have a properly weighted sample

$(\boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{K}_{t-1}^{(j)}, w_{t-1}^{(j)})$ with respect to the target distribution $p(\boldsymbol{S}_{t-1}, \boldsymbol{K}_{t-1} \mid \boldsymbol{Y}_{t-1})$. Since given the trajectory $(\boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{K}_{t-1}^{(j)})$, the system is linear and Gaussian, we have

$$p(\boldsymbol{x}_{t-1} \mid \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{K}_{t-1}^{(j)}, \boldsymbol{Y}_{t-1}) = N(\boldsymbol{\mu}_{t-1}^{(j)}, \boldsymbol{\Sigma}_{t-1}^{(j)}).$$

Note that, with given new (s_t, κ_t) and observation y_t at time t, the mean and variance matrix can be easily updated to $(\boldsymbol{\mu}_t^{(j)}, \boldsymbol{\Sigma}_t^{(j)})$ with the Kalman filter.

Then at time t, we generate the samples of (s_t, κ_t) using an efficient trial sampling distribution

$$q(s_t, \kappa_t \mid \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{K}_{t-1}^{(j)}, \boldsymbol{Y}_t) = p(s_t, \kappa_t \mid \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{K}_{t-1}^{(j)}, \boldsymbol{Y}_t). \tag{3.2}$$

For this trial distribution, the importance weight updating can be seen as (Liu and Chen (1998)),

$$w_t^{(j)} = w_{t-1}^{(j)} \cdot p(y_t \mid \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{K}_{t-1}^{(j)}, \boldsymbol{Y}_{t-1}).$$

Note that $p(y_t \mid \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{K}_{t-1}^{(j)}, \boldsymbol{Y}_{t-1})$ can be computed by

$$\begin{aligned} (3.3)\quad & p(y_t \mid \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{K}_{t-1}^{(j)}, \boldsymbol{Y}_{t-1}) \\ & \propto \sum_{a_i \in \mathcal{A}} \sum_{\nu=-1}^{1} \underbrace{p(y_t \mid \boldsymbol{S}_{t-1}^{(j)}, s_t = a_i, \kappa_t = \kappa_{t-1}^{(j)} + \nu, \boldsymbol{Y}_{t-1}) p(s_t = a_i) p(\kappa_t = \kappa_{t-1}^{(j)} + \nu)}_{\triangleq \gamma_{t,i,\nu}^{(j)}}. \end{aligned}$$

In order to calculate the terms in (3.2) and (3.3), we use the following one-step Kalman filter. First we need to adjust the mean and variance matrix for $\boldsymbol{x}_{t-1}$ to accommodation the change of model order. It is easily seen that the $\boldsymbol{x}_{t-1}^*$ defined in (2.4) follows $N(\boldsymbol{\mu}_{t-1}^*, \boldsymbol{\Sigma}_{t-1}^*)$, where (i) for $\kappa_t = \kappa_{t-1}$, $\boldsymbol{\mu}_{t-1}^* = \boldsymbol{\mu}_{t-1}$, $\boldsymbol{\Sigma}_{t-1}^* = \boldsymbol{\Sigma}_{t-1}$; (ii) for $\kappa_t = \kappa_{t-1} - 1$, $\boldsymbol{\mu}_{t-1}^* = \boldsymbol{\mu}_{t-1}[1 : \kappa_t]$, $\boldsymbol{\Sigma}_{t-1}^* = \boldsymbol{\Sigma}_{t-1}[1 : \kappa_t, 1 : \kappa_t]$; and (iii) for $\kappa_t = \kappa_{t-1} + 1$, $\boldsymbol{\mu}_{t-1}^* = [\boldsymbol{\mu}_{t-1}\ \ 0]$ and

$$\boldsymbol{\Sigma}_{t-1}^* = \begin{bmatrix} \boldsymbol{\Sigma}_{t-1} & \boldsymbol{0} \\ 0 & 0 \end{bmatrix}.$$

With the state space model (2.4) and (2.1) and given $\boldsymbol{S}_t$ and $\boldsymbol{K}_t$, one step Kalman filter can be used for updating $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$. Specifically, we have

$$\begin{aligned} P_t &= \boldsymbol{\Sigma}_{t-1}^* + \boldsymbol{I}_{\kappa_t} \sigma_x^2 \\ S_t &= s_t^2 \boldsymbol{\phi}_t P_t \boldsymbol{\phi}_t^T + \sigma^2 \\ \boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-1}^* + s_t P_t \boldsymbol{\phi}_t S_t^{-1} (y_t - s_t \boldsymbol{\phi}_t \boldsymbol{\mu}_t^*) \\ \boldsymbol{\Sigma}_t &= P_t - s_t^2 P_t \boldsymbol{\phi}_t S_t^{-1} \boldsymbol{\phi}_t^T P_t. \end{aligned}$$

Furthermore, for $\nu \in \{-1, 0, 1\}$,

$$p(y_t \mid \boldsymbol{S}_{t-1}^{(j)}, s_t = a_i, \boldsymbol{K}_{t-1}^{(j)}, \kappa_t = \kappa_{t-1}^{(j)} + \nu, \boldsymbol{Y}_{t-1}) \sim N(\mu_{t,i,\nu}^{(j)}, \sigma_{t,i,\nu}^{2(j)}), \tag{3.4}$$

where

$$\mu_{t,i,\nu}^{(j)} = s_t^{(j)} \boldsymbol{\phi}_t \boldsymbol{\mu}_{t-1}^{(j)} \quad \text{and} \quad \sigma_{t,i,\nu}^{2(j)} = S_t. \tag{3.5}$$

Therefore, $\gamma_{t,i,\nu}^{(j)}$ in (3.4) can be computed by

$$\gamma_{t,i,\nu}^{(j)} = \frac{1}{\sigma_{t,i,\nu}^{(j)}} \exp\left(-\frac{\|y_t - \mu_{t,i,\nu}^{(j)}\|^2}{\sigma_{t,i,\nu}^{2(j)}}\right) p(s_t = a_i) p(\kappa_t = \kappa_{t-1}^{(j)} + \nu). \tag{3.6}$$

The trial distribution in (3.2) can be computed as follows,

$$\begin{aligned} & p(s_t = a_i, \kappa_t = \kappa_{t-1}^{(j)} + \nu, \mid \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{K}_{t-1}^{(j)}, \boldsymbol{Y}_t) \\ & \propto p(y_t \mid \boldsymbol{S}_{t-1}^{(j)}, s_t = a_i, \boldsymbol{Y}_{t-1}) p(s_t = a_i) p(\kappa_t = \kappa_{t-1}^{(j)} + \nu) \\ & = \gamma_{t,i,\nu}^{(j)}. \end{aligned} \tag{3.7}$$

Finally, we summarize the nonparametric adaptive blind SMC receiver algorithm as follows:

0. Initialization: For each $j = 1, 2, \ldots, m$:
- Sample $\kappa_0^{(j)}$ uniformly from $[\kappa_{\min}, \kappa_{\max}]$.
- Set $\boldsymbol{\Sigma}_0^{(j)} = 1000 \boldsymbol{I}_{\kappa_0^{(j)}}$. Draw $\boldsymbol{\mu}_0^{(j)} = 0$.
- Set $w_0^{(j)} = 1$.

The following steps are implemented at time t ($t = 1, \ldots, K_0$) to update each weighted sample. For $j = 1, \ldots, m$:

1. For each $a_i \in \mathcal{A}$ and $\nu \in \{-1, 0, 1\}$, compute $(\mu_{t,i,\nu}^{(j)}, \sigma_{t,i,\nu}^{2(j)})$ and $\gamma_{t,i,\nu}^{(j)}$ given by (3.5) and (3.6), respectively.

2. Draw $(s_t^{(j)}, \kappa_t^{(j)})$ from $\mathcal{A}$ and $\kappa_{t-1}^{(j)} + \nu$, $\nu \in \{-1, 0, 1\}$ with probability

$$p(s_t = a_i, \kappa_t = \kappa_{t-1}^{(j)} + \nu \mid \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{K}_{t-1}^{(j)}, \boldsymbol{Y}_t) \propto \gamma_{t,i,\nu}^{(j)}.$$

Append $s_t^{(j)}$ to $\boldsymbol{S}_{t-1}^{(j)}$ to obtain $\boldsymbol{S}_t^{(j)}$ and $\kappa_t^{(j)}$ to $\boldsymbol{K}_{t-1}^{(j)}$ to obtain $\boldsymbol{K}_t^{(j)}$.

3. Compute the importance weight

$$w_t^{(j)} \propto w_{t-1}^{(j)} \sum_{a_i \in \mathcal{A}} \sum_{\nu \in \{-1,0,1\}} \gamma_{t,i,\nu}^{(j)}.$$

4. Suppose the imputed sample $s_t^{(j)} = a_i$ and $\kappa_t^{(j)} = \kappa_{t-1}^{(j)} + \nu$, then let $\boldsymbol{\mu}_t^{(j)}$ and $\boldsymbol{\Sigma}_t^{(j)}$ be the corresponding mean and variance matrix of $\boldsymbol{x}_t$ at time t.

5. Resampling.

Remark 7. The resampling procedure is an important step in SMC, as shown in Doucet *et al.* (2000) and Liu and Chen (1998). Roughly speaking, resampling is to duplicate the streams with large importance weights, while eliminate the ones with small importance weights. Heuristically, resampling can provide chances for good sample streams to amplify themselves and hence "rejuvenate" the sampler to produce a better result for future states as system evolves.

Remark 8. If the speed of fading does not change over time, then one can assume there is a 'best' (but unknown) wavelet order κ to approximate the fading process. Then the blind receiver can be modified as follows: At time $t = 0$, $\kappa_0^{(j)}$ is sampled from a prior distribution on κ (say, uniform distribution on $(\kappa_{\min}, \kappa_{\max})$), and they are fixed subsequently, i.e. $\kappa_t^{(j)} = \kappa_{t-1}^{(j)}$. In this case, how 'good' each order is is represented by the weight of the streams, and the estimation of the symbols is essentially done through averaging results from different model size κ, i.e. the Bayesian model averaging approach (Madigan and Raftery (1994), Volinsky *et al.* (1997), Madigan and York (1995), Hoeting *et al.* (1999), Raftery *et al.* (1997), George and McCulloch (1993)). Furthermore, if the wavelet coefficients does not change over time, one can further set $\sigma_x^2 = 0$ in the algorithm.

Remark 9. The algorithm requires $3 \times |\mathcal{A}|$ one-step Kalman filter operations to generate one sample in each iteration. When $|\mathcal{A}|$ is large, this may create significant computational burden. In such cases, one can simplify the algorithm by generating s_t and κ_t separately. By assuming that, if there is a change ($\nu = -1$ or 1) in κ_t at time t, its effect on the current y_t is relatively small. Hence it is relatively safe to sample s_t assuming $\kappa_t = \kappa_{t-1}$. Specifically, steps 1–3 are changed to

1*. For each $a_i \in \mathcal{A}$, compute $(\mu_{t,i,0}^{(j)}, \sigma_{t,i,0}^{2(j)})$ and $\gamma_{t,i,0}^{(j)}$ given by (3.5) and (3.6), respectively. Draw $s_t^{(j)}$ from $\mathcal{A}$ with probability $\gamma_{t,i,0}$. Append $s_t^{(j)}$ to $\boldsymbol{S}_{t-1}^{(j)}$ to obtain $\boldsymbol{S}_t^{(j)}$.

2*. Assume the sampled $s_t^{(j)} = a_{i^*}$. For $\nu = -1, 0, 1$, compute $(\mu_{t,i^*,\nu}^{(j)}, \sigma_{t,i^*,\nu}^{2(j)})$ and $\gamma_{t,i^*,\nu}^{(j)}$ given by (3.5) and (3.6), respectively. Draw $\nu = \nu^*$ with probability $\gamma_{t,i^*,\nu}$. Append $\kappa_t^{(j)}$ to $\boldsymbol{K}_{t-1}^{(j)}$ to obtain $\boldsymbol{K}_t^{(j)}$.

3*. Compute the importance weight

$$w_t^{(j)} \propto w_{t-1}^{(j)} \frac{\gamma_{t,i^*,\nu^*}^{(j)} \sum_i \gamma_{t,i,0}^{(j)} \sum_{\nu=-1}^{1} \gamma_{t,i^*,\nu}^{(j)}}{\gamma_{t,i^*,0}}.$$

The resulting $(\boldsymbol{S}_t^{(j)}, \boldsymbol{K}_t^{(j)}, w_t^{(j)})$ is properly weighted with respect to $P(\boldsymbol{S}_t, \boldsymbol{K}_t \mid \boldsymbol{Y}_t)$. This algorithm requires $(|\mathcal{A}| + 3)$ one-step Kalman filter operations for each sample.

Remark 10. From the recursive procedure described above, we get the samples $\{(\boldsymbol{S}_{t+\delta}^{(j)}, w_{t+\delta}^{(j)})\}_{j=1}^m$ at time $(t+\delta), \delta > 0$, which are properly weighted with respect to $p(\boldsymbol{S}_{t+\delta} \mid \boldsymbol{Y}_{t+\delta})$. Hence, focusing on $\boldsymbol{S}_t$ at time $(t+\delta)$, we obtain a delayed estimation of the symbol

$$p(s_t = a_i \mid \boldsymbol{Y}_{t+\delta}) \cong \frac{1}{W_{t+\delta}} \sum_{j=1}^{m} 1(s_t^{(j)} = a_i) w_{t+\delta}^{(j)}, \qquad a_i \in \mathcal{A}, \tag{3.8}$$

with $W_{t+\delta} \triangleq \sum_{j=1}^m w_{t+\delta}^{(j)}$. Since the weights $\{w_{t+\delta}^{(j)}\}_{j=1}^m$ contain information about the signals $(y_{t+1}, \ldots, y_{t+\delta})$, the estimation in (3.8) is usually more accurate. Note that such a delayed estimation method incurs no additional computational cost (i.e., cpu time), but it requires some extra memory for storing $\{s_{t+1}^{(j)}, \ldots, s_{t+\delta}^{(j)}\}_{j=1}^m$.

4. Simulation results

In this section, we present some simulation examples to illustrate the performance of the proposed nonparametric adaptive SMC receivers in flat-fading channels.

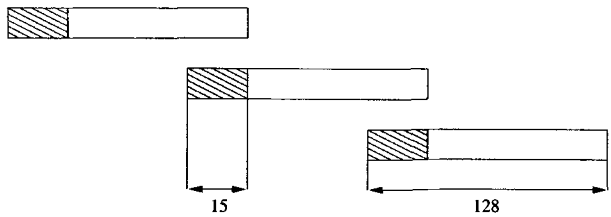

Fig. 2. The frame structure for the nonparametric SMC receivers. The block size is 128, and the adjacent blocks overlap by 15 symbols.

The BPSK modulation is employed in the simulation, i.e., the transmitted symbols $\{s_t\}$ take values from ± 1. The characteristics of the fading channels are described in Section 2. The Jakes' fading process is generated using the frequency spectrum method (Proakis (1995)). In the decomposition of the fading process, the Daubechies wavelet filter with order 2 is used to construct the reconstruction matrix. Our simulations show that little performance improvement can be achieved with Daubechies filters of order higher than 2; whereas with Daubechies filter of order 1 (i.e., Haar wavelet), the performance degradation is significant.

The signal frame structure is shown in Fig. 2. Each data block contains $K_0 = 128$ symbols. Adjacent blocks overlap by 15 symbols to allow the SMC filter to reach the steady state. To speed up the convergence, for each data block, the values of the mean and the covariance of wavelet coefficients $\{\boldsymbol{\mu}_0^{(j)}, \boldsymbol{\Sigma}_0^{(j)}\}_{j=1}^m$ are initialized as the corresponding values $\{\boldsymbol{\mu}_{K_0}^{(j)}, \boldsymbol{\Sigma}_{K_0}^{(j)}\}_{j=1}^m$ at the end of the previous block.

The performance of the proposed adaptive SMC receivers is compared with that of the receivers with perfect channel state information (CSI). In flat-fading channels, the receiver with CSI makes a decision on symbol s_t according to $\hat{s}_t = \text{sign}(\{\Re\{\alpha_t^* y_t\})$ (Proakis (1995)). We call the performance of the receiver with CSI the "known channel bound".

First we consider the simpler case in which the fading speed does not change within the data block. The performance comparison is shown in Figs. 3 and 4, for flat-fading channels with normalized Doppler $f_d T = 0.005$ and $f_d T = 0.01$, respectively.

The SMC receiver with fixed shrinkage order is first implemented, by fixing κ and setting $\sigma_x^2 = 0$ in the adaptive algorithm. For the adaptive approach, we set $\sigma_x = 0.001$, and $p(\kappa_t = \kappa_{t-1}) = 0.8$ and $p(\kappa_t = \kappa_{t-1} \pm 1) = 0.1$. In all settings, the number of the Monte Carlo samples drawn at each time is set at $m = 100$. The resampling procedure is employed in the SMC and the threshold of effective sample size is $\bar{m}_t = m/10$. The delayed-weight method is used with $\delta = 6$.

The bit error rate (BER) versus the signal-to-noise ratio is plotted in Fig. 3 for different settings. It is seen that in this case the best performance (close to the known channel bound) for fixed order model is achieved using eight wavelet coefficients. With four wavelet coefficients, the performance is a bit worse. And the performance is substantially degraded if the shrinkage order is very large (e.g., 15, 25 and 32).

Figure 4 shows the BER performance of the SMC blind receiver in a flat-fading channel with normalized Doppler $f_d T = 0.01$. Because the fading speed is faster, the best order is increased to 15 (when $E_b/N_0 < 20$ dB).

On the other hand, without assuming any knowledge of the fading speed, the adap-

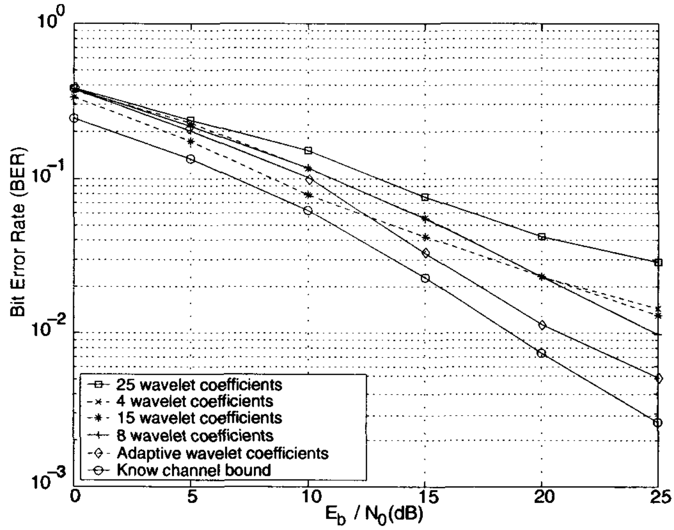

Fig. 3. The BER performance of the SMC receivers. The channel is flat-fading with $f_dT = 0.005$. The delayed-weight method is used with $\delta = 6$.

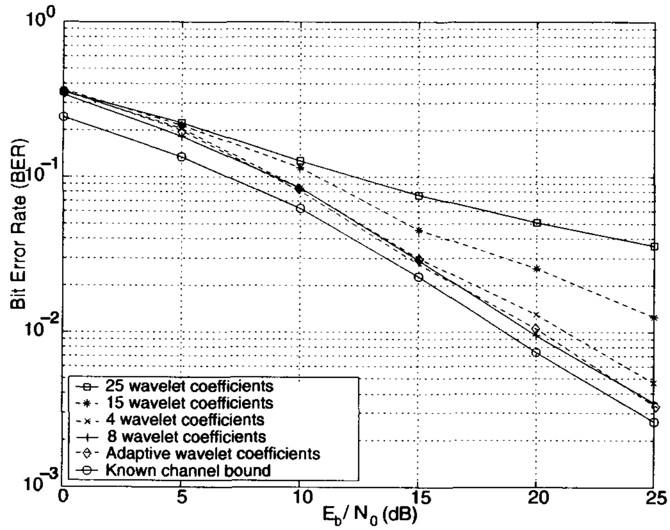

Fig. 4. The BER performance of the SMC receivers. The channel is flat-fading with $f_dT = 0.01$. The delayed-weight method is used with $\delta = 6$.

tive SMC receiver is able to perform as well as the best fixed order receiver in both cases, except in some low signal to noise ratio cases (less than 10 dB). The adaptive nature allows the receiver to 'choose' the correct order in process.

We next show the performance of the adaptive SMC receiver for fading channels with time varying fading speed. Fading processes with normalized Doppler shift at $f_dT = 0.005$ and $f_dT = 0.01$ are generated and appended (smoothly at $t = 75$) to form a block of length 128. The adaptive receiver is employed. Figure 5 shows the weighted

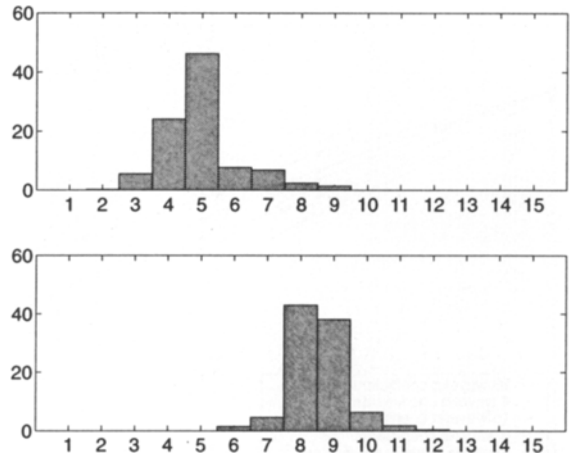

Fig. 5. The histogram for the shrinkage order associated with the sample streams at times $t = 50$ (top) and $t = 100$ (bottom) for $E_b/N_0 = 15$ dB. The fading process is composed of two equal-length parts: the normalized Doppler shift is $f_dT = 0.005$ in the first part and $f_dT = 0.01$ in the second part. The result is the average over 100 simulations.

histogram of the shrinkage order associated with each sample stream at the first region (upper panel) and the second region (lower panel) $E_b/N_0 = 15$ dB. The result is the average over 100 simulations. It is seen that the shrinkage order κ_t focuses on the range of $(3, 8)$ for the first region, which is close to the optimum fixed shrinkage order for the $f_dT = 0.005$ (see Fig. 3). Meanwhile, for the second region, the shrinkage order focuses on the range of $(7, 11)$, also close to the optimum fixed shrinkage order $(8, 15)$ for $f_dT = 0.01$ (see Fig. 4).

We also check the performance between the adaptive SMC receiver and the fixed order SMC receiver at the two different regions. Figure 6 shows the error rate at each time point (120 simulations) at $E_b/N_0 = 15$ dB, It is easily seen that the adaptive receiver performs the best in both regions while fixed order receivers do well in one region but fail in another.

Figure 7 illustrates the overall BER performance for different SNR values for flat-fading channels with time varying Doppler values. It is seen that the proposed adaptive SMC receiver performs the best.

5. Conclusions

In this paper, we have developed a new nonparametric adaptive Bayesian receiver technique for blind detection in fading channels with unknown and time varying channel statistics. It is based on wavelet modelling of the fading process and the sequential Monte Carlo method for online Bayesian inference. Moreover, a novel adaptive blind Bayesian receiver is developed. To cope with the uncertainty in fading speed and the time varying feature, a special structure of the wavelet approximation to the fading process is developed and utilized for SMC filtering. The performance of the blind adaptive receivers for flat-fading channels are demonstrated via computer simulations. It is seen that the adaptive receiver works as well as the 'optimal' fixed order receiver for time-invariant fading processes, without assuming the knowledge of the fading speed. When the fading

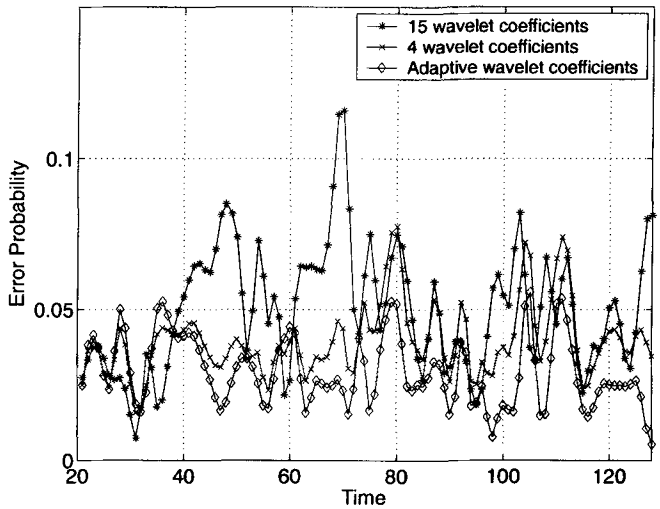

Fig. 6. The BER performance of the blind adaptive receiver at different time points for SNR = 15. The channel is flat-fading under time-varying Dopplers.

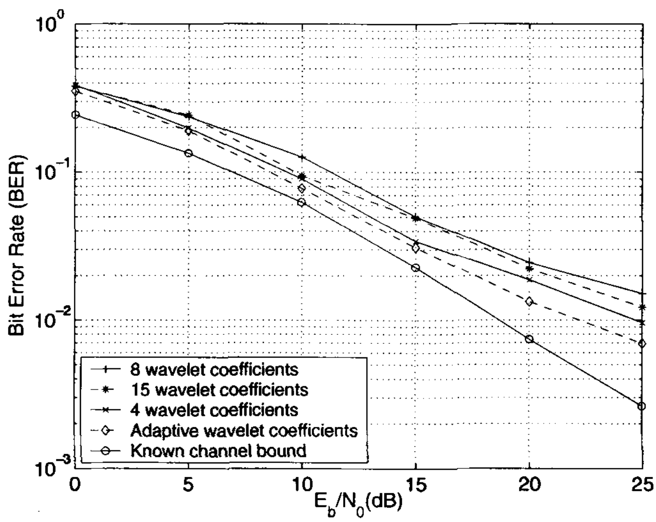

Fig. 7. The BER performance of the blind adaptive receiver. The channel is flat-fading under time varying Dopplers.

is indeed time varying, the adaptive receiver outperform all fixed order receivers.

References

Andrieu, C., de Freitas, J. F. G. and Doucet, A. (1999). Sequential MCMC for Bayesian model selection, *Proceedings of IEEE Higher Order Statistics Workshop, Israel*. (Available http://citeseer.nj.nec.com/andrieu99sequential.html)

Andrieu, C., de Freitas, J. F. G. and Doucet, A. (2000). Sequential Bayesian estimation and model selection for dynamic kernel machines, Tech. Report, TR 341, Cambridge University. (Available http://citeseer.nj.nec.com/andrieu00sequential.html)

Chen, R. and Liu, J. S. (2000). Mixture Kalman filters, *J. Roy. Statist. Soc. Ser. B*, **62**, 493–509.

Chen, R., Wang, X. and Liu, J. S. (2000). Adaptive joint detection and decoding in flat-fading channels via mixture Kalman filtering, *IEEE Trans. Inform. Theory*, **46**, 2079–2094.

Chipman, H., Kolaczyk, E. and McCulloch, R. (1997). Adaptive Bayesian wavelet shrinkage, *J. Amer. Statist. Assoc.*, **92**, 1413–1421.

Clyde, M. and George, E. (2000). Flexible empirical Bayes estimation for wavelets, *J. Roy. Statist. Soc. Ser. B*, **62**, 681–698.

Collings, I. B. and Moore, J. B. (1994). An HMM approach to adaptive demodulation of QAM signals in fading channels, *International Journal of Adaptive Control and Signal Processing*, **8**, 457–474.

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets, *Comm. Pure Appl. Math.*, **41**, 425–455.

Daubechies, I. (1992). *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia.

Donoho, D. and Johnstone, I. (1994). Multiple shrinkage and subset selection in wavelets, *Biometrica*, **81**, 425–455.

Doucet, A., Godsill, S. and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering, *Statist. Comput.*, **10**, 197–208.

Doucet, A., de Freitas, J. F. G. and Gordon, N. (2001). *Sequential Monte Carlo in Practice*, Cambridge University Press, Cambridge.

Georghiades, C. N. and Han, J. C. (1997). Sequence estimation in the presence of random parameters via the EM algorithm, *IEEE Trans. Commun.*, **COM-45**, 300–308.

George, E. and McCulloch, R. (1993). Variable selection via Gibbs sampling, *J. Amer. Statist. Assoc.*, **88**, 881–889.

Guo, D., Wang, X. and Chen, R. (2002). Wavelet-based sequential Monte Carlo blind receivers in fading channels with unknown channel statistics, *IEEE International Conference in Communications (ICC 2002)*, New York.

Haeb, R. and Meyr, H. (1989). A systematic approach to carrier recovery and detection of digitally phase modulated signals on fading channels, *IEEE Trans. Commun.*, **37**, 748–754.

Hoeting, A. J., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial, *Statist. Sci.*, **14**, 382–417.

Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*, Springer, New York.

Liu, J. S. and Chen, R. (1995). Blind deconvolution via sequential imputations, *J. Amer. Statist. Assoc.*, **90**, 567–576.

Liu, J. S. and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems, *J. Amer. Statist. Assoc.*, **93**, 1032–1044.

Liu, J. and West, M. (2001). Combined parameter and state estimation in simulation-based filtering, *Sequential Monte Carlo Methods in Practice* (eds. A. Doucet, N. de Freitas and N. Gordon), 197–217, Springer, New York.

Lodge, J. H. and Moher, M. L. (1990). Maximum likelihood sequence estimation of CPM signals transmitted over Rayleigh flat-fading channels, *IEEE Trans. Commun.*, **38**, 787–794.

Madigan, D. and Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window, *J. Amer. Statist. Assoc.*, **89**, 1535–1546.

Madigan, D. and York, J. (1995). Bayesian graphical models for discrete data, *Internat. Statist. Rev.*, **63**, 215–232.

Martone, M. (2000). Wavelet-based separating kernels for array processing of cellular DS/CDMA signals in fast fading, *IEEE Trans. Commun.*, **48**, 979–995.

Proakis, J. G. (1995). *Digital Communications*, 3rd ed., McGraw-Hill, New York.

Raftery, A. E., Madigan, D. and Hoeting, J. (1997). Bayesian model averaging for linear regression models, *J. Amer. Statist. Assoc.*, **92**, 179–191.

Volinsky, C. T., Madigan, D., Raftery, A. E. and Kronmal, R. A. (1997). Bayesian model averaging in proportional hazard models: Assessing the risk of a stroke, *J. Roy. Statist. Soc. Ser. C*, **46**, 433–448.