Reversible Jump MCMC for Joint Detection and Estimation of Sources in Coloured Noise

Jean-René Larocque and James P. Reilly
Department of Electrical and Computer Engineering,
McMaster University, 1280 Main St. W.,
Hamilton, Ontario,
Canada L8S 4K1

Abstract—
This paper presents a novel Bayesian solution to the difficult problem of joint detection and estimation of sources impinging on a single array of sensors in spatially coloured noise with arbitrary covariance structure. Robustness to the noise covariance structure is achieved by integrating out the unknown covariance matrix in an appropriate posterior distribution. The proposed procedure uses the Reversible Jump Markov Chain Monte Carlo method to extract the desired model order and direction of arrival parameters.

We show that the determination of model order is consistent provided a particular hyperparameter is within a specified range. Simulation results support the effectiveness of the method.

I. Introduction

Array signal processing has found important applications in diverse areas such as radar, sonar, communications, and seismic exploration. The array processing problem has to do with 1) the detection of the number of discrete signal components, and 2) estimation of the angles of arrival of the signals incident onto arrays of sensors. This detection and estimation problem is by now mature [1]-[12] etc. However, there is still considerable room for improvement beyond the current state of the art in this area. For example, current algorithms commonly make assumptions regarding the characteristics of the noise field and the incident signals; in particular, that the background noise is spatially white [1][2][3] or of known covariance. Further, in the typical scenario, the determination of model order (detection) and the estimation of desired signal parameters are executed independently rather than jointly. Array signal processing methods which can accommodate the unknown coloured noise case are desirable, since they offer improved performance in coloured noise over methods developed assuming white noise. Further, the coloured background noise case is often encountered in the practical scenario.

The direction of arrival (DOA) problem in the case where the noise has unknown covariance has been addressed in [4][5][6]. In these papers, the model order is assumed known, and the associated detection problem is not addressed. Joint detection and estimation has been addressed in [2] and [3] under the white noise assumption, and later generalized in [7] to the autoregressive noise case. Determination of model order for the unknown arbitrarily spatially coloured noise case has long been regarded as a difficult problem. Established methods, e.g., using information theoretic criteria [8][9], have been developed under the assumption of white noise, but exhibit considerable sensitivity to the whiteness assumption and suffer significant performance loss in the presence of coloured noise [11]. Methods for detection in unknown coloured noise have been proposed [10][11][12], but these techniques require in effect two separate sensor arrays where the signal is assumed correlated between the arrays and the noise uncorrelated. This assumption places restrictions on the geometry of the configuration and therefore these methods may have limited use in practice. In [13], this problem was addressed from the information theoretic stand point, and a solution to the joint detection and estimation problem in arbitrary unknown noise is proposed. The method however involves the minimization of a highly nonlinear objective function and is therefore prone to being trapped in local minima.

In this paper, we propose a new array processing technique [14][15] based on MCMC methods, for joint detection and estimation of sources in arbitrarily coloured noise, using a single sensor array. The proposed method can therefore adopt unrestricted array geometries. With this, and the fact that no assumptions are made on the noise covariance structure, the proposed method overcomes many of the difficulties associated with previous methods. The technique proposed in this paper is an extension of the method of [2] from the white noise case to the

coloured noise case.

The use of MCMC methods in array processing is advantageous for several reasons. First, by virtue of the *reversible jump Metropolis Hastings algorithm* [16], the MCMC techniques afford a joint detection and estimation procedure, which offers improved performance over each process being implemented independently. Second, the performance of the detection and estimation process for a given set of observed data can be readily estimated. Finally, unlike competing methods, they offer convergence to the global optimum with arbitrarily high probability, provided the number of samples is large enough.

Notation: In this paper, bold lower case symbols denote a vector, while bold upper-case quantities represent matrices. Superscript ^{T} and ^{H} denote the transpose and Hermitian transpose operations respectively. A subscript $_o$, e.g., $\boldsymbol{\phi}_o$, denotes the true value of the associated variable. The quantity $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ represents a normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, while $\mathrm{U}_(a, b]$ represents a uniform distribution over the open interval $(a, b]$. The symbol $\sim$ denotes "distributed as"; e.g., $x \sim U_{(\cdot,\cdot]}$ means the random variable x is distributed uniformly over some interval.

The problem formulation is presented in Section II, and in Section III, the desired marginal posterior distribution is developed. We perform the integration with respect to the unknown spatial covariance matrix by projecting the observed data onto the noise subspace $\mathcal{N}$. After assigning a Jeffreys' prior, the resulting posterior may then be integrated by comparison to a complex Wishart distribution. In Section IV, we apply the reversible jump MCMC algorithm to sample the posterior distribution and so obtain joint estimates of the desired parameters. In section V we discuss conditions for consistency of the model order estimate. Simulation results demonstrating the performance of the proposed method are given in Section VI, and the conclusions are presented in Section VII.

II. Problem Formulation

The signal model we consider consists of a complex data vector $\boldsymbol{y}(n) \in \mathcal{C}^M$ which represents the data received by a linear array of M sensors at the nth snapshot. The data vector is composed of incident narrow-band plane wave signals each at centre frequency ω_0 from k_o distinct sources embedded in Gaussian noise. The received vector at the nth snapshot can be written as

$$\begin{aligned} \boldsymbol{y}(n) = \boldsymbol{S}(\boldsymbol{\phi})\Big|_{\boldsymbol{\phi}=\boldsymbol{\phi}_o} \boldsymbol{a}(n) + \boldsymbol{\nu}(n), \\ n = 1, \ldots, N, \end{aligned} \tag{1}$$

where N is the number of observed snapshots. In particular, when the array elements are uniformly spaced, we assume $d_0 \leq \frac{\pi v}{\omega_0}$, where d_0 is the distance between the sensors, and v is the velocity of propagation. Each of these incident plane-wave signals impinges on the array of sensors at an angle $\theta_k, k = 1, \ldots, k_o$, to the normal of the array. Then $\boldsymbol{S}(\boldsymbol{\phi})$ is the $M \times k_o$ matrix, the kth column of which is

$$\begin{aligned} \boldsymbol{s}(\phi_k) = [1, e^{j\phi_k}, e^{j2\phi_k}, \ldots, e^{j(M-1)\phi_k}]^T, \\ k = 1, \ldots, k_o, \end{aligned} \tag{2}$$

with

$$\phi_k = \omega_0 d_0 \sin(\theta_k)/v. \tag{3}$$

The quantity $\boldsymbol{a}(n) \in \mathcal{C}^{k_o}$ represents the complex amplitudes of the incident signals at the nth snapshot. We assume the amplitudes are *iid* normally distributed between snapshots, with unknown and arbitrary means and covariance. The spatially coloured noise vector $\boldsymbol{\nu}(n)$ is an *iid* normally distributed noise vector distributed according to $\mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} \in \mathcal{C}^{M \times M}$ is an unknown and arbitrary covariance matrix. The signal and noise are assumed uncorrelated such that

$$E(\boldsymbol{a}^H \boldsymbol{S}^H \boldsymbol{\nu}) = \boldsymbol{0}. \tag{4}$$

For a hypothesized number of signals k, we designate by $\mathcal{S}(k)$ the signal subspace spanned by the vectors $[\boldsymbol{s}(\phi_1), \ldots, \boldsymbol{s}(\phi_k)]$. The $(M-k)$-dimensional orthogonal complement noise subspace of $\mathcal{S}(k)$ is denoted by $\mathcal{N}(k)$. In the following discussion, we simplify notation by writing $\mathcal{S}$ and $\mathcal{N}$ instead of $\mathcal{S}(k)$ and $\mathcal{N}(k)$ respectively.

In this paper, we jointly estimate k_o and $\boldsymbol{\phi}_o$ using MCMC techniques. Determination of k_o using MCMC methods involves sampling a posterior distribution $p(k, \boldsymbol{\phi}|\boldsymbol{y})$ of varying model order. This consideration requires the use of the *reversible jump* MCMC method [16], which is treated in further detail in Sect. 4. We treat the unknown quantities $\boldsymbol{a}(n), n = 1, \ldots, N$ and $\boldsymbol{\Sigma}$ as nuisance parameters, which are integrated out. It is the elimination of $\boldsymbol{\Sigma}$ in this way which enables us to handle correlated noise with unknown covariance. Even though these parameters can be eliminated by the sampling process inherent in the MCMC technique, we choose

to integrate them analytically since the analysis is tractable, to produce a more efficient sampler for the parameters of interest.

After integration of the nuisance parameters, the parameter space $\mathbf{\Phi}$ of interest for joint detection of the model order and estimation of the incident signal angles, includes k and the corresponding angles $\phi_1 \dots \phi_k$. We denote the permissible set of k to be $\mathcal{K} = \{0, \dots, M-1\}$. To reflect the fact that the number of parameters changes with model order, we note that $\mathbf{\Phi}$ can be written as a finite union of subspaces as $\mathbf{\Phi} = \cup_{k=0}^{M-1}\{k\} \times \mathbf{\Phi}_k$, where

$$\mathbf{\Phi}_k = \{(0, 2\pi)^k : \ |\phi_i - \phi_j| > \epsilon\}, \qquad i \neq j, \tag{5}$$

where ϵ is a small number > 0, and $(i, j) = \{0, \dots, k\}$. The set $\mathbf{\Phi}_k$ is defined in this way so that $\boldsymbol{S}$ in (1) is always full rank. The set $\mathbf{\Phi_0} = \emptyset$.

III. Development of the Marginal Posterior Distribution

Since the N snapshots are *iid*, the total likelihood function of all the data can be expressed in the following form:

$$p(\boldsymbol{Y}|\boldsymbol{\phi}_k, \boldsymbol{A}, \boldsymbol{\Sigma}, k) = \frac{1}{\pi^{MN}\left|\boldsymbol{\Sigma}\right|^N} \times e^{-\sum_n (\boldsymbol{y}(n) - \boldsymbol{S}(\boldsymbol{\phi}_k)\boldsymbol{a}(n))^H \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}(n) - \boldsymbol{S}(\boldsymbol{\phi}_k)\boldsymbol{a}(n))}, \tag{6}$$

where $\boldsymbol{Y}$, $\boldsymbol{A}$ are all the data and amplitudes, respectively and $|\cdot|$ denotes determinant.

To proceed with the integration of the nuisance parameters $\boldsymbol{\Sigma}$ and $\boldsymbol{A}$, we first define an orthonormal matrix $\boldsymbol{U}(\boldsymbol{\phi}, k) \in \mathcal{C}^{M \times M}$ as in [5][7]

$$\boldsymbol{U}(\boldsymbol{\phi}, k) = \underset{M \times k}{[\boldsymbol{U}_s(\boldsymbol{\phi}, k)} \quad \underset{M \times (M-k)}{\boldsymbol{U}_\nu(\boldsymbol{\phi}, k)}], \tag{7}$$

where $\boldsymbol{U}_s(\boldsymbol{\phi}, k) \in \mathcal{S}$ and $\boldsymbol{U}_\nu(\boldsymbol{\phi}, k) \in \mathcal{N}$. We now transform the received data $\boldsymbol{y}(n)$ into $\boldsymbol{z}(n) \triangleq \boldsymbol{U}^H \boldsymbol{y}(n)$, with a signal component $\boldsymbol{z}_s(n) \in \mathcal{S}$ and a noise component $\boldsymbol{z}_\nu(n) \in \mathcal{N}$ defined respectively as

$$\boldsymbol{z}_s(n) = \boldsymbol{U}_s^H(\boldsymbol{\phi}, k)\boldsymbol{y}(n), \tag{8}$$

and

$$\boldsymbol{z}_\nu(n) = \boldsymbol{U}_\nu^H(\boldsymbol{\phi}, k)\boldsymbol{y}(n). \tag{9}$$

In the following analysis we consider only the neighbourhood around the true value where $\boldsymbol{\phi} \approx \boldsymbol{\phi}_o$. It is straightforward to show that for reasonably high values of N and SNR, the maximum of the final posterior distribution (22) of interest is close to $\boldsymbol{\phi} = \boldsymbol{\phi}_o$. Thus, for *maximum a posteriori* (MAP) estimation as we propose, this neighbourhood is the only region of interest. In this region, we have $\boldsymbol{S}^H \boldsymbol{U}_\nu \approx \boldsymbol{0}$. Then using (4), we can write

$$\boldsymbol{R}_{zz} = \mathrm{E}\left[\boldsymbol{z}\boldsymbol{z}^H\right] \approx \begin{bmatrix} \boldsymbol{R}_{\tilde{a}\tilde{a}} + \boldsymbol{U}_s^H \boldsymbol{\Sigma} \boldsymbol{U}_s & \boldsymbol{U}_s^H \boldsymbol{\Sigma} \boldsymbol{U}_{\boldsymbol{\nu}} \\ \boldsymbol{U}_\nu^H \boldsymbol{\Sigma} \boldsymbol{U}_s & \boldsymbol{U}_\nu^H \boldsymbol{\Sigma} \boldsymbol{U}_\nu \end{bmatrix}, \tag{10}$$

where $\boldsymbol{R}_{\tilde{\boldsymbol{a}}\tilde{\boldsymbol{a}}} \triangleq \mathrm{E}\{\tilde{\boldsymbol{a}}\tilde{\boldsymbol{a}}^H\}$ and $\tilde{\boldsymbol{a}}(n) = \boldsymbol{U}_s^H \boldsymbol{S}\boldsymbol{a}(n)$. In the following, we neglect the off-diagonal blocks of $\boldsymbol{R}_{zz}$ in (10). This assumption makes the analysis tractable, at the expense of a somewhat suboptimal solution[1]. Then locally around $\boldsymbol{\phi} \approx \boldsymbol{\phi}_o$ we can assume that $\boldsymbol{z}_s$ is independent of $\boldsymbol{z}_\nu$. Using (10) it follows that $\boldsymbol{z}_s|\tilde{\boldsymbol{a}}(n) \sim \mathrm{N}(\tilde{\boldsymbol{a}}(n), \boldsymbol{C})$, and $\boldsymbol{z}_\nu \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{W})$, where $\boldsymbol{C} \triangleq \boldsymbol{U}_s^H \boldsymbol{\Sigma} \boldsymbol{U}_s$, and $\boldsymbol{W} \triangleq \boldsymbol{U}_\nu^H \boldsymbol{\Sigma} \boldsymbol{U}_\nu$. Under these conditions the joint likelihood function of $\boldsymbol{z}_s$ and $\boldsymbol{z}_\nu$ is given as:

$$\begin{aligned} p(\boldsymbol{Z}_s, \boldsymbol{Z}_\nu|\tilde{\boldsymbol{A}}, \boldsymbol{\phi}, k, \boldsymbol{W}^{-1}) \approx \\ \pi^{-Nk}|\boldsymbol{C}^{-1}|^N \\ \times \exp\left\{-\sum_{n=1}^{N}(\boldsymbol{z}_s(n) - \tilde{\boldsymbol{a}}(n))^H \boldsymbol{C}^{-1}(\boldsymbol{z}_s(n) - \tilde{\boldsymbol{a}}(n))\right\} \\ \times \pi^{-N(M-k)}|\boldsymbol{W}^{-1}|^N \exp\left\{-\sum_{n=1}^{N} \boldsymbol{z}_\nu^H(n)\boldsymbol{W}^{-1}\boldsymbol{z}_\nu(n)\right\}. \end{aligned} \tag{11}$$

The desired posterior distribution can then be written using Bayes' theorem as

$$\begin{aligned} p(\tilde{\boldsymbol{A}}, \boldsymbol{\phi}, k, \boldsymbol{W}^{-1}|\boldsymbol{Z}_s, \boldsymbol{Z}_\nu) \propto \\ p(\boldsymbol{Z}_s, \boldsymbol{Z}_\nu|\tilde{\boldsymbol{A}}, \boldsymbol{\phi}, k, \boldsymbol{W}^{-1}) \\ \times p(\tilde{\boldsymbol{A}}|\boldsymbol{\phi}, k, \boldsymbol{W}^{-1}) \cdot p(\boldsymbol{W}^{-1}|\boldsymbol{\phi}, k) \cdot p(\boldsymbol{\phi}|k) \cdot p(k). \end{aligned} \tag{12}$$

We now discuss the assignment of each of the prior distributions in (12). These priors are chosen to be non-informative where possible. When convenient, we also choose the structural form of these distributions for their desirable conjugate properties. The prior distributions are described as follows:

- $\tilde{\boldsymbol{A}}$ is assigned a non-informative prior distribution. This non-informative prior is implemented as a normal distribution with zero mean, whose covariance matrix $\boldsymbol{D}$ is such that the corresponding joint confidence region at a specified probability level is

[1]Simulation results have verified that the performance of the algorithm is within a few dB of the corresponding Cramer-Rao lower bound. Thus, the effect of neglecting the off-diagonal blocks is not substantial.

significantly larger than that of a normal distribution with mean zero and covariance matrix $\boldsymbol{C}$. Thus,

$$p(\tilde{\boldsymbol{A}}|\boldsymbol{\phi}_k, k, \boldsymbol{W}^{-1}) = \prod_{n=1}^{N} \mathrm{N}(\boldsymbol{0}, \boldsymbol{D}) \tag{13}$$

where

$$\boldsymbol{D} = d^2 \boldsymbol{I}_k, \tag{14}$$

and $\boldsymbol{I}_k$ is the identity matrix of size k. We discuss the choice of the hyper-parameter d^2 in Sect. 5.

• the prior distribution for $\boldsymbol{\phi}$ is chosen to be uniform[2]:

$$p(\boldsymbol{\phi}_k|k) = \mathrm{U}[0, 2\pi]^k. \tag{15}$$

• The prior on k is chosen to be Poisson with expectation Λ. Although this choice is not strictly non-informative, it results in a more efficient MCMC sampling procedure, and further, simulation results as shown in Section 6 have shown that performance is not sensitive to the choice of prior on k.

$$p(k) = \Lambda^k e^{-\Lambda}/k! \tag{16}$$

• $\boldsymbol{W}^{-1}$: We use a non-informative multi-dimensional Jeffreys' prior [17] for the unknown transformed noise covariance matrix. Jeffreys derived a general rule for obtaining a non-informative prior, which may be stated as: *The prior distribution for a set of parameters is taken to be proportional to the square root of the determinant of the Fisher information matrix.* A further extensive treatment on this topic is provided in [18] and in [5].

According to this principle, the non-informative prior distribution of the transformed noise covariance matrix can be written as [5]

$$p(\boldsymbol{W}^{-1} \mid \boldsymbol{\phi}, k) \propto \mid \boldsymbol{W}^{-1} \mid^{-(M-k)}. \tag{17}$$

The posterior distribution is then:

$$\begin{aligned} &p(k, \boldsymbol{\phi}, \boldsymbol{W}^{-1}, \tilde{\boldsymbol{A}}|\boldsymbol{Z}_s, \boldsymbol{Z}_\nu) \propto \\ &\quad |\boldsymbol{C}^{-1}|^N e^{-\sum_n (\tilde{\boldsymbol{a}}(n)-\boldsymbol{z}_s(n))^H \boldsymbol{C}^{-1} (\tilde{\boldsymbol{a}}(n)-\boldsymbol{z}_s(n))} \\ &\quad \times \pi^{-kN} |\boldsymbol{D}^{-1}|^N e^{-\sum_n \tilde{\boldsymbol{a}}(n)^H \boldsymbol{D}^{-1} \tilde{\boldsymbol{a}}(n)} \\ &\quad \times |\boldsymbol{W}^{-1}|^{N-(M-k)} e^{-\sum_n \boldsymbol{z}_\nu^H(n) \boldsymbol{W}^{-1} \boldsymbol{z}_\nu(n)} \\ &\quad \times \frac{\Lambda^k}{k!(2\pi)^k}, \end{aligned} \tag{18}$$

[2]Strictly speaking, this prior should be defined only over the region specified by (5). However, in practice, this constraint on the valid region for $\boldsymbol{\phi}$ is implemented in the algorithm which evaluates the desired posterior distribution (22). Thus we leave the prior for $\boldsymbol{\phi}$ in its more tractible form given above.

where superfluous constants independent of $\boldsymbol{\phi}$ or k have been absorbed into the constant of proportionality.

Remark 1: The hyper-parameter Λ could either be estimated ahead of time, or could be considered known as part of the design parameters. It could also be treated as a random variable, with its own prior distribution, to make the algorithm more robust. However, this complicates the problem and does not prove necessary. □

We now proceed to integrate out the nuisance parameters analytically. Since we have assumed the matrix $\boldsymbol{D}$ is "larger" than the matrix $\boldsymbol{C}$, then for sufficiently large N, the term $\tilde{\boldsymbol{a}}^H \boldsymbol{D}^{-1} \tilde{\boldsymbol{a}}$ in the second exponential on the right in (18) is small in comparison to the term $\tilde{\boldsymbol{a}}^H \boldsymbol{C}^{-1} \tilde{\boldsymbol{a}}$ in the first exponential. This simplifies the posterior to:

$$\begin{aligned} &p(k, \boldsymbol{\phi}, \boldsymbol{W}^{-1}, \tilde{\boldsymbol{A}}|\boldsymbol{Z}_s, \boldsymbol{Z}_\nu) \propto \\ &\quad \pi^{-kN} |\boldsymbol{C}^{-1}|^N e^{-\sum_n [\tilde{\boldsymbol{a}}(n)-\boldsymbol{z}_s(n)]^H \boldsymbol{C}^{-1} [\tilde{\boldsymbol{a}}(n)-\boldsymbol{z}_s(n)]} \\ &\times |\boldsymbol{W}^{-1}|^{N-(M-k)} e^{-\sum_n \boldsymbol{z}_\nu(n)^H \boldsymbol{W}^{-1} \boldsymbol{z}_\nu(n)} |\boldsymbol{D}^{-1}|^N \\ &\times \frac{\Lambda^k}{k!(2\pi)^k}. \end{aligned} \tag{19}$$

The first nuisance parameter (the amplitude of the sources $\tilde{\boldsymbol{a}}(n)$) can now be easily integrated out by comparison, since it only appears in an isolated Gaussian distribution. The posterior distribution can then be simplified to:

$$\begin{aligned} &p(k, \boldsymbol{\phi}, \boldsymbol{W}^{-1}|\boldsymbol{Z}_\nu) \propto \\ &\quad \left|\boldsymbol{W}^{-1}\right|^{N-(M-k)} \\ &\quad \times e^{\left\{-\sum_{n=1}^{N} \boldsymbol{z}_\nu(n)^H \boldsymbol{W}^{-1} \boldsymbol{z}_\nu(n)\right\}} \frac{\Lambda^k}{k!(2\pi)^k d^{2kN}} \\ &\quad = \left[|\boldsymbol{W}^{-1}|^{N-(M-k)} \exp\left\{-\mathrm{tr} N \hat{\boldsymbol{W}} \boldsymbol{W}^{-1}\right\}\right] \\ &\quad \times \frac{\Lambda^k}{k!(2\pi)^k (d^2)^{kN}}, \end{aligned} \tag{20}$$

where $\mathrm{tr}(\cdot)$ is the trace operator, $N\hat{\boldsymbol{W}}((\boldsymbol{\phi}, k)) \triangleq \sum_{n=1}^{N} \boldsymbol{z}_\nu(n) \boldsymbol{z}_\nu(n)^H$, and we have used (14) for $\boldsymbol{D}$. The noise covariance matrix can now be integrated out by comparing the term inside the square brackets above to a complex Wishart distribution on $\boldsymbol{W}^{-1}$, with order $p = M - k$ and parameter $N\hat{\boldsymbol{W}}$, i.e. the conventional roles of $\boldsymbol{W}$ and $\hat{\boldsymbol{W}}$ are reversed. This term integrates to $I(\hat{\boldsymbol{W}})$, given as [19]

$$I(\hat{\boldsymbol{W}}) = \pi^{\frac{1}{2}(M-k)(M-k-1)} \prod_{i=1}^{M-k} \Gamma(N-i+1) \left|N\hat{\boldsymbol{W}}\right|^{-N}, \tag{21}$$

where Γ is the Gamma function. The posterior distribution, after carrying out the integration and some manipulation is then:

$$p(k,\boldsymbol{\phi}|\boldsymbol{Z}_\nu) \propto \frac{\pi^{\frac{1}{2}(M-k)(M-k-1)}\prod_{i=1}^{M-k}\Gamma(N-i+1)}{(2\pi/\Lambda)^k k!(d^2)^{kN}} \times \left|N\hat{\boldsymbol{W}}\right|^{-N}. \tag{22}$$

The objective is to estimate the parameters of this highly non-linear function, as the following MAP (*maximum a posteriori*) estimate:

$$\{\hat{k},\hat{\boldsymbol{\phi}}\} = \arg\max_{k,\boldsymbol{\phi}} p(k,\boldsymbol{\phi}|\mathbf{Z_v}). \tag{23}$$

IV. The Reversible Jump MCMC Algorithm

We now propose the Reversible Jump Metropolis-Hastings (MH) algorithm [16][20] to perform the Bayesian computation in extracting the parameters of interest from the posterior distribution (22). The proposed method is similar to that presented in [2] which describes the white noise case. The following presentation is therefore brief, with the intention of presenting sufficient detail only to distinguish the coloured noise case from the white noise case. A further treatment on model order detection using MCMC methods is given in [21].

This procedure samples directly from different model orders from the joint distribution on $\boldsymbol{\Phi} = \{\boldsymbol{\phi}_k, k\}$. In effect, the process jumps between subspaces of different dimensions, thus visiting all model orders for $k \in \mathcal{K}$.

In the reversible jump case, candidate samples $\{\boldsymbol{\phi}^\star, k^\star\}$ are chosen from a *set* of proposal distributions, which are randomly accepted according to an acceptance ratio that ensures reversibility, and therefore the invariance of the Markov chain with respect to the desired posterior distribution. Here, we choose our set of proposal distributions to correspond to the following set of moves:

1. the *birth* move, valid for $k < M$. Here, a new incident plane wave is proposed at random on $(0, 2\pi]$.
2. the *death* move, valid for $k > 0$. Here, a randomly chosen incident plane wave is removed.
3. the *update* move. Here, the parameters describing the incident plane wave are updated for a fixed value of k.

The probabilities for choosing each move are denoted u_k, b_k and d_k, respectively, such that $u_k + b_k + d_k = 1$ for all k. In accordance with [16], we choose:

$$\begin{aligned} b_k &= c\ \min\{\frac{p(k+1)}{p(k)}, 1\}, \\ d_{k+1} &= c\ \min\{\frac{p(k)}{p(k+1)}, 1\}, \end{aligned} \tag{24}$$

where $p(\cdot)$ is the prior probability of the kth model according to (16), and c is the tuning parameter for the ratio of update moves to jump moves. We choose $c = 0.5$ so that the probability of a jump is between 0.5 and 1 at each iteration [16].

An acceptance ratio r is then generated according to:

$$r = \frac{p(\boldsymbol{\phi}^\star, k^\star|\boldsymbol{Z}_\nu)q(\boldsymbol{\phi}^{(i)}, k^{(i)}|\boldsymbol{\phi}^\star, k^\star)}{p(\boldsymbol{\phi}^{(i)}, k^{(i)}|\boldsymbol{Z}_\nu)q(\boldsymbol{\phi}^\star, k^\star|\boldsymbol{\phi}^{(i)}, k^{(i)})}. \tag{25}$$

where $q(\cdot|\cdot)$ is the proposal distribution corresponding to the respective move type. An acceptance parameter α is then defined as

$$\alpha = \min\{r, 1\}. \tag{26}$$

Then, the proposed candidate $(\boldsymbol{\phi}^\star, k^\star)$ is accepted as the current state at iteration i with probability α.

The set of accepted candidates represents a set of samples drawn form the posterior distribution of interest. These samples can then be used to construct a histogram from which the desired statistical inferences can be made.

This description is summarized as follows:

Reversible Jump MCMC

1. Initialization: set $\boldsymbol{\Phi}^{(0)} = (\boldsymbol{\phi}^{(0)}, k^{(0)})$
2. Iteration i,
   - Sample $u \sim [0, 1]$
   - if $(u < b_{k^{(i)}})$ then execute a "birth move" (see part B).
   - else if $(u < b_{k^{(i)}} + d_{k^{(i)}})$ then execute a "death move" (see part B).
   - else, execute an update move (see part A).
3. $i \leftarrow i + 1$, goto step 2

A. Update move

Here, we assume that the current state of the algorithm is $(\boldsymbol{\Phi}_k, \{k\})$. When the update move is selected, the algorithm samples only on the space of $\boldsymbol{\Phi}_k$, for k fixed, using a proposal distribution $q(\cdot|\cdot) = U_{(0,2\pi]}$. The acceptance ratio $r = r_{update}$

from (25) and (22) for the update move, in the case of coloured noise is therefore:

$$r_{update}(\boldsymbol{\phi}_k^\star, k, \boldsymbol{\phi}_k, k) = \frac{\left|N\hat{\boldsymbol{W}}(\boldsymbol{\phi}^\star, k)\right|^{-N}}{\left|N\hat{\boldsymbol{W}}(\boldsymbol{\phi}, k)\right|^{-N}}, \quad (27)$$

$$\alpha_{update} = \min[r_{update}, 1]. \quad (28)$$

The candidate $\boldsymbol{\phi}^\star$ is then accepted as the current state ($\boldsymbol{\phi}_k^{(i+1)} = \boldsymbol{\phi}_k^\star$), with probability α_{update}. The mixing performance of the proposed method is enhanced by selecting randomly between two types of proposal distribution for the update case: one involves a global exploration of the parameter space, while the other involves a local exploration. Further detail is presented in [2].

B. Birth and Death moves

In the death move case, we assume the current state is in $(\boldsymbol{\Phi}_{k+1}, \{k+1\})$ and we wish to determine whether the state is in $(\boldsymbol{\Phi}_k, \{k\})$ at the next iteration. This involves the removal of an incident signal, which is chosen randomly amongst the $(k+1)$ existing incident signals. The proposal distribution $q(\boldsymbol{\phi}_k^\star, k|\boldsymbol{\phi}_{k+1}, k+1)$ for the death move is therefore chosen as

$$\begin{aligned} q(\boldsymbol{\phi}_k^\star, k|\boldsymbol{\phi}_{k+1}, k+1) &= p(k) \div \binom{k+1}{1} \\ &\propto \frac{\Lambda^k}{k!}\frac{1}{(k+1)}. \end{aligned} \quad (29)$$

Similarly, in the birth move case, we assume the current state is $(\boldsymbol{\Phi}_k, \{k\})$ and we wish to determine whether the next state is in $(\boldsymbol{\Phi}_{k+1}, \{k+1\})$. This involves the addition of a new incident signal, which is proposed uniformly over $(0, 2\pi]$. The proposal distribution $q(\boldsymbol{\phi}_{k+1}^\star, (k+1)|\boldsymbol{\phi}_k, k)$ for the birth move is therefore

$$\begin{aligned} q(\boldsymbol{\phi}_{k+1}^\star, (k+1)|\boldsymbol{\phi}_k, k) &= p(k+1) \times \frac{1}{2\pi} \\ &\propto \frac{\Lambda^{k+1}}{(k+1)!}\frac{1}{2\pi}. \end{aligned} \quad (30)$$

For the death move, a candidate state $(\boldsymbol{\phi}_k^\star, k)$ is then sampled from (29). The acceptance ratio $r = r_{death}$ from (25) and (22) is then given as

$$\begin{aligned} r_{death}(\boldsymbol{\phi}_k^\star, k, \boldsymbol{\phi}_{k+1}, k+1) &= \frac{\left|N\hat{\boldsymbol{W}}(\boldsymbol{\phi}^\star, k)\right|^{-N}}{\left|N\hat{\boldsymbol{W}}(\boldsymbol{\phi}, k+1)\right|^{-N}} \\ &\times \pi^{M-k-1}\Gamma(N-M+k+1)(k+1)d^{2^N}. \end{aligned} \quad (31)$$

The quantity α_{death} is then defined according to

$$\alpha_{death} = \min[r_{death}, 1]. \quad (32)$$

In [16], it is shown that a sufficient condition for reversibility is that the acceptance ratio $r = r_{birth}$ for the birth move be given as

$$\alpha_{birth} = \min[1, \frac{1}{r_{death}}]. \quad (33)$$

The following block describes the algorithm for the birth move.

Birth Move

- Propose a new direction of arrival $\phi_c \sim U_{(0,2\pi]}$

$\boldsymbol{\phi}_{k+1}^{(i+1)} = [\boldsymbol{\phi}_k^{(i)} \vdots \phi_c]$,
- Evaluate α_{birth} with (33).
- Sample $u \sim U_{[0,1]}$.
- if $(u \leq \alpha_{birth})$ then the state of the Markov Chain becomes $(\boldsymbol{\phi}_{k+1}^{(i+1)}, k+1)$, else it remains at $(\boldsymbol{\phi}_k^{(i)}, k)$.

■

The description for the death move is similar, with appropriate modifications.

Remark 2: At this point, it is enlighting to take the logarithm of the previously obtained posterior distribution (22):

$$\begin{aligned} \log p(k, \boldsymbol{\phi}_k|\boldsymbol{Z}_\nu) = \\ \gamma - N\log(\left|N\hat{\boldsymbol{W}}\right|) - kN\log(d^2) \\ + k\log(\Lambda/2\pi) + \log(k!) \\ + \frac{1}{2}(M-k)(M-k-1)\log(\pi) \\ + \log(\prod_{i=1}^{M-k}\Gamma(N-i+1)) \end{aligned} \quad (34)$$

where γ is a constant. In this form, the similarities with previous model selection criteria such as AIC, MDL, D-MAP of Djuric[22] or W-MDL of Wax [13] are made apparent. The first term represents the likelihood term, while the remaining ones jointly constitute a "penalty term", which is dependent on the prior for $\tilde{\boldsymbol{a}}$ and k. □

V. Model Order Determination

In this section, we discuss conditions which must apply on the hyper-parameter d^2 in (13) for consistent determination of model order. That is, we show that if d^2 falls within a specified range, and if certain assumptions on the posterior $p(\boldsymbol{\phi}, k|\boldsymbol{Z}_\nu)$ are satisfied, then the estimated $\hat{k} \to k_o$ as $N \to \infty$.

The marginal posterior distribution for k is given as

$$p(k|\boldsymbol{Z}_\nu) \propto \int_{\boldsymbol{\Phi}_k} p(k, \boldsymbol{\phi}|\boldsymbol{Z}_\nu) d\boldsymbol{\phi}. \tag{35}$$

Let the eigenvalues of $N\hat{\boldsymbol{W}}(\boldsymbol{\phi}, k)$ in (22) at $\boldsymbol{\phi} = \hat{\boldsymbol{\phi}}$, where $\hat{\boldsymbol{\phi}}$ is the MAP estimate, be given as $\lambda_1, \lambda_2, \ldots, \lambda_{M-k}$ arranged in *ascending*, rather than the usual *descending* order. For the asymptotic case we are considering, for moderate values of SNR, the joint posterior distribution $p(\boldsymbol{\phi}, k|\boldsymbol{Z}_\nu)$ concentrates at $\boldsymbol{\phi} = \hat{\boldsymbol{\phi}}$, which will be near the true value $\boldsymbol{\phi}_o$ if $k = k_o$[3]. Thus (35) can be written as

$$\begin{aligned} p(k|\boldsymbol{Z}_\nu) \sim & \left[\frac{\pi^{\frac{1}{2}(M-k)(M-k-1)}}{(2\pi/\Lambda)^k k!} \right] \\ & \times \left[\frac{\prod_{i=1}^{M-k} \Gamma(N-i+1)}{d^{2kN}} \left(\prod_{i=1}^{M-k} \lambda_i \right)^{-N} \right] \end{aligned} \tag{36}$$

where $\sim$ indicates "approximately proportional to".

Let us define the event E_i as the declaration of a model order in error by i signals; i.e., E_i occurs when we declare $\hat{k} = k_o + i$ or $\hat{k} = k_o - i$. In the following analysis, we assume $P(E_1) > P(E_2) > \ldots > P(E_{M-1})$, as in e.g., [23]. This implies $p(k|\boldsymbol{Z}_\nu)$ is unimodal in k. From this assumption, a necessary and sufficient condition for consistent detection of model order is therefore

$$\lim_{N\to\infty} \frac{p(k_o+1|\boldsymbol{Z}_\nu)}{p(k_o|\boldsymbol{Z}_\nu)} \to 0, \tag{37}$$

and

$$\lim_{N\to\infty} \frac{p(k_o-1|\boldsymbol{Z}_\nu)}{p(k_o|\boldsymbol{Z}_\nu)} \to 0. \tag{38}$$

[3]Here we have a situation which is typical of model order detection problems. When $k \geq k_o$, then under good conditions, k_o elements of $\hat{\boldsymbol{\phi}}$ will be close to the corresponding elements of $\boldsymbol{\phi}_o$, and $k - k_o$ elements of $\hat{\boldsymbol{\phi}}$ will be extraneous. These extraneous values can easily be determined because their associated amplitudes will not be statistically significant. In the case where $k < k_o$, then generally no elements of $\hat{\boldsymbol{\phi}}$ will be close to any in $\boldsymbol{\phi}_o$.

Using (36) we can write

$$\begin{aligned} \frac{p(k_o+1|\boldsymbol{Z}_\nu)}{p(k_o|\boldsymbol{Z}_\nu)} \approx & \\ & \frac{\left[\frac{\pi^{\frac{1}{2}(M-k_o-1)(M-k_o-2)}}{(2\pi/\Lambda)^{k_o+1}(k_o+1)!} \right]}{\left[\frac{\pi^{\frac{1}{2}(M-k_o)(M-k_o-1)}}{(2\pi/\Lambda)^{k_o} k_o!} \right]} \\ & \times \frac{\left[\frac{\prod_{i=1}^{M-k_o-1} \Gamma(N-i+1)}{(d)^{2N(k_o+1)}} \left(\prod_{i=1}^{M-k_o-1} \lambda_i \right)^{-N} \right]}{\left[\frac{\prod_{i=1}^{M-k_o} \Gamma(N-i+1)}{(d)^{2k_oN}} \left(\prod_{i=1}^{M-k_o} \lambda_i \right)^{-N} \right]} \\ & = \frac{\pi^{-(M-k_o-1)} (\lambda_{M-k_o})^N (\frac{\Lambda}{2\pi})}{\Gamma(N-M+k_o+1) d^{2N}(k_o+1)}, \end{aligned} \tag{39}$$

where we have used the property $\Gamma(n+1) = (n+1)\Gamma(n)$ and that the first $(M-k_o-1)$ smallest eigenvalues are common to both the numerator and to the denominator. Since for $k = k_o$ and $\boldsymbol{\phi} = \boldsymbol{\phi}_o$, and large N, we have $\hat{\boldsymbol{W}} \to \boldsymbol{U}_\nu^H \boldsymbol{\Sigma} \boldsymbol{U}_\nu \approx \boldsymbol{U}_\nu^H \boldsymbol{R}_{yy} \boldsymbol{U}_\nu$, where $\boldsymbol{R}_{yy} \triangleq \mathrm{E}\{\boldsymbol{y}\boldsymbol{y}^H\}$. Since the eigenvalues of $\boldsymbol{U}_\nu^H \boldsymbol{R}_{yy} \boldsymbol{U}_\nu$ are the same as the nonzero eigenvalues of $\boldsymbol{P}_{\mathcal{N}} \boldsymbol{R}_{yy} \boldsymbol{P}_{\mathcal{N}}^H$, where $\boldsymbol{P}_{\mathcal{N}}$ is the projector onto $\mathcal{N}$, the eigenvalue λ_{M-k_o} above is the largest eigenvalue of the covariance matrix formed from the data projected onto $\mathcal{N}$.

Using a similar development we also have

$$\frac{p(k_o-1|\boldsymbol{Z}_\nu)}{p(k_o|\boldsymbol{Z}_\nu)} \approx \frac{\Gamma(N-M+k_o) d^{2N}(k_o)}{\pi^{-(M-k_o)} (\lambda_{M-k_o+1})^N (\frac{\Lambda}{2\pi})}. \tag{40}$$

In this case, $k = k_o - 1$, and since the estimated dimension of the noise subspace is too large, it incorporates part of $\mathcal{S}$. The quantity λ_{M-k_o+1} may thus be associated with the smallest signal eigenvalue of $\boldsymbol{R}_{yy}$ projected into $\mathcal{S}$, which is the same as the smallest signal eigenvalue of $\boldsymbol{R}_{yy}$.

We can now evaluate the limits of (39) and (40). To do so, we require the *Stirling approximation* [24] to the Γ-function, which is valid for large values of the argument:

$$\Gamma(x) \approx \sqrt{2\pi} e^{-x} x^{(x-\frac{1}{2})}. \tag{41}$$

We also note from the definition of $N\hat{\boldsymbol{W}}$ that the eigenvalues λ_i are directly proportional to N, and thus they can be written as

$$\lambda_i = N\tilde{\lambda}_i, \qquad i = 1, \ldots, M-k, \tag{42}$$

where $\tilde{\lambda}_i$ is the normalized version of λ_i. Substituting (41) and (42) into (39) we find that (37) is satisfied if

$$\tilde{\lambda}_{M-k_o} < \frac{d^2}{e}. \tag{43}$$

In a similar way, using (40), (38) is satisfied if

$$\tilde{\lambda}_{M-k_o+1} > \frac{d^2}{e}. \tag{44}$$

Therefore, by combining (43) and (44), we see that detection of model order is consistent if the hyperparameter d^2 is chosen so that

$$\tilde{\lambda}_{M-k_o} < \frac{d^2}{e} < \tilde{\lambda}_{M-k_o+1}, \tag{45}$$

i.e., the quantity $\frac{d^2}{e}$ must lie in the gap between the largest projected normalized noise eigenvalue and the smallest normalized signal eigenvalue.

Strictly speaking, this procedure for determining d^2 cannot be used for consistent detection because (45) depends on the unknown k_o. However, it should be possible in the practical scenario to propose an *ad hoc* scheme to approximate (45). For example, it is usually possible to form an estimate of $\mathbf{\Sigma}$ during periods where it is known with reasonable certainty there are no signals present. The largest eigenvalue of this matrix could then be used as an upper bound on $\tilde{\lambda}_{M-k_o}$. If some *a priori* knowledge on the DOAs were available, then an estimate of $\boldsymbol{U}_\nu$ can be evaluated and a better estimate of $\tilde{\lambda}_{M-k_o}$ could be determined. In either case, d^2 could be given as an empirically determined constant times this eigenvalue estimate.

VI. Simulation Results

In this Section, we present simulation results of the sampling scheme developed in Sections 3 and 4. The estimates are obtained as the MAP estimator of the histogram of the samples, as defined in (23).

The spatially coloured noise is generated with an AR process of order 2 with roots $0.95e^{-j1.07\pi}$ and $0.95e^{-j0.88\pi}$ as in [7], with excitation from complex white noise samples of equal variance $\sigma_w^2/2$ for both the real and imaginary parts. Figure 1 shows the directional spectrum of the spatially coloured noise.

The hyperparameter d^2 in (14) was assigned the value of 1000, in accordance with the criterion (45). Figures 2 and 3 show typical results for 10000 iterations (after a sufficient burn-in period of 5000 iterations, based on the observation of the chain behaviour) of the Reversible Jump Sampler with $N = 30$ observations of a circular array made of 5 equi-spaced sensors (with a radius of $0.102m$) at 1.86GHz, when the SNR is 2 dB, where the hyperparameter Λ in (16) was assigned the true value of two. The SNR is defined as:

$$SNR = \frac{a_1^2}{2\sigma_w^2}. \tag{46}$$

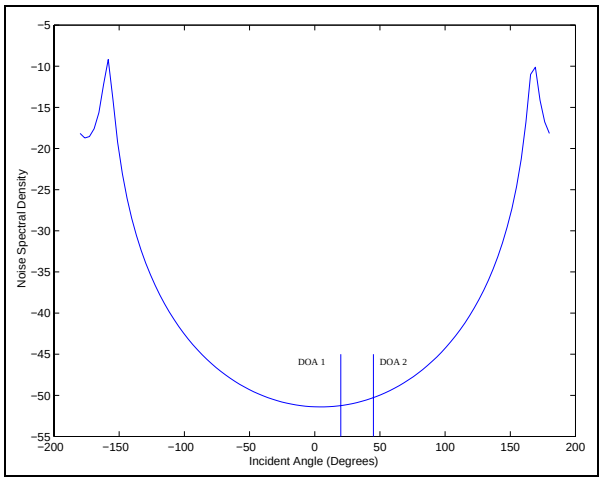

Fig. 1. Spectrum of the spatially coloured noise used in simulations.

The characteristics of the signals and parameters used to obtain these simulation results are summarized in Table 6.1. This scenario is very difficult, using only a small number of snapshots at low SNR with two sources well within a beamwidth of the receiver array.

Amplitude	DOA	Sensors	Snapshots	SNR	MCMC iterat
10	20°	5	30	2 dB	10000
10	45°				

TABLE 6.1
Signal parameter values for the simulation results.

It is interesting to observe the evolution of the instantaneous model probabilities (top portion of Figure 2) and how they reach an equilibrium value. The bottom portion of Figure 2 shows the posterior histogram of the estimated number of sources after burn-in. Figure 3 shows the histograms $\hat{P}(\boldsymbol{\phi}|k = 2)$. Clearly, from visual inspection of the figures, we see the algorithm detected the right number of sources and has estimated their respective DOAs as 19.5° and 46.0°, which are close to the true values.

The next figure, Figure 4, shows the behavior of the algorithm when the hyperparameter Λ in (16) is initialized to $\Lambda = 5$. This represents the scenario where the prior distribution favours the wrong choice of model order. As expected, the algorithm takes somewhat longer to converge to the proper model order, but the correct a posteriori estimate is ultimately obtained. This supports the claim that the algorithm is not sensitive to this hyperparame-

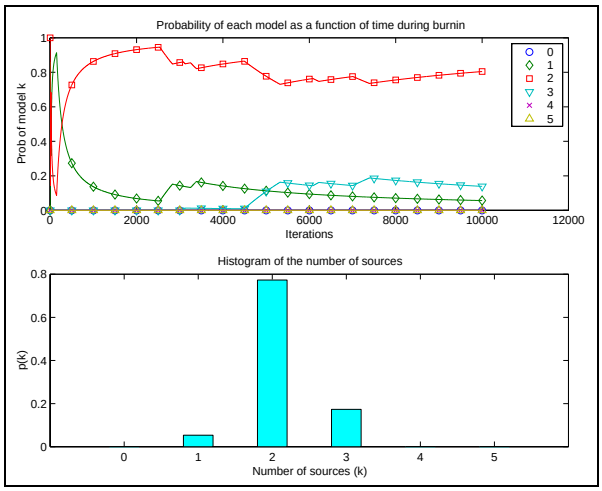

Fig. 2. Instantaneous estimate of the model probability vs. iteration index (top half); Histogram of the number of sources after burn-in (bottom half), for the hyperparameter $\Lambda = 2$ (from (16)).

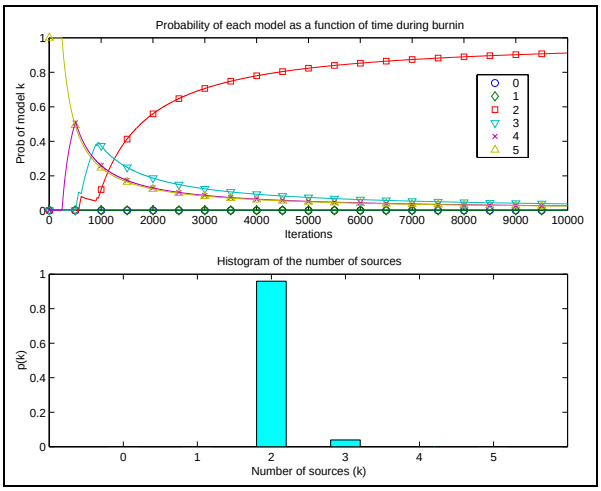

Fig. 4. Similar to Figure 2, but for $\Lambda = 5$. Instantaneous estimate of the model probability (top half); Histogram of the number of sources after burn-in (bottom half)

.

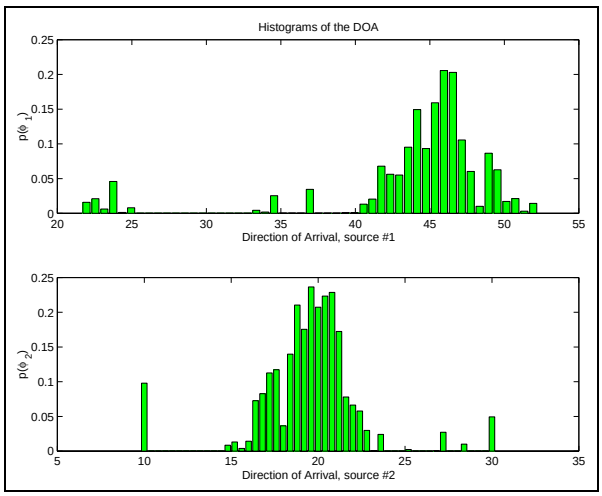

Fig. 3. Histogram of the DOA's after burn-in: Source 1 (top) and Source 2 (bottom).

ter.

The previous results represent only one realization of the noise. In order to assess the performance of the algorithm for DOA estimation only, in terms of mean- squared error of the estimates as a function of the SNR for multiple noise realizations, the algorithm was applied to 50 Monte Carlo noise realizations, for a range of values of SNR, from -8dB to 16dB, with the other parameter values given as before in Table 6.1. Specifically, the value of d^2 was again held at the value 1000, which was verified to satisfy (45) over the entire range of SNR values considered. The results are shown in Figure 5. Even though two sources are simulated, only the results for the source corresponding to DOA_1 for each respective method are shown; otherwise, the figure becomes too dense. The performance for each method is shown for the case where k is assigned its true value. The curves for DOA_2 behave similarly, but are degraded somewhat, due to the fact that, as shown in Figure 1, the source at DOA_2 receives more noise than that at DOA_1.

Figure 5 shows the Cramér-Rao lower bound for this simulation scenario, evaluated from the results in [25][26]. In this figure, it may be observed that the performance of the MCMC method almost achieves the CRLB. The slight degradation may be caused by neglecting the off-diagonal terms of the covariance matrix in the development of the posterior distribution (11).

In Figure 5 the performance of the MCMC method is compared with the algorithm developed by Wax (W-MDL) [13] and implemented by the alternating projection [27] algorithm (as suggested in [13]), where a steepest descent gradient method was used for the one dimensional optimization. There were two methods used to initialize the Wax method. When the initialization was done *randomly*, performance was severely degraded due to convergence of the algorithm to local minima, as shown in the figure. However, when the W-MDL method was initialized to the *true* value of the DOA parameters, then the performance is similar to that achieved by the MCMC method with random initialization. With the MCMC method, 20,000 samples were used, of which the first 15,000 were dis-

carded as burn-in. The MCMC reliably determines the global optimum (e.g., at 2 dB SNR, 49/50 of the Monte Carlo trials were successful in this respect). These results demonstrate that, even though this implementation of the Wax method is significantly faster, the proposed MCMC method has much better global convergence performance than W-MDL. This behaviour suggests the need for a global search procedure for the Wax method.

A close relationship between MCMC methods and the simulated annealing technique [28] for global optimization has been cited [29]. We may therefore regard the optimization inherent in the DOA estimation portion of the MCMC procedure as being related to a simulated annealing algorithm. Since a global search procedure for the DOA estimation portion of the Wax method could also be implemented using a simulated annealing method or equivalent, the proposed method requires about the same order of computation as the W-MDL method when estimating DOAs alone with global convergence. However, when model order as well as the DOAs are to be determined, by virtue of the reversible jump MCMC algorithm, the proposed MCMC procedure concentrates its samples on the most probable values of k, and hence does not waste excess effort on less likely values. In contrast, since the Wax method must expend equal effort over all feasible values of k, it is less efficient.

We also include performance results of the so-called maximum likelihood (ML) method for DOA estimation [27], which was developed explicitly assuming spatially white noise. From this figure, it is apparent that the proposed DOA estimation method which assumes arbitrary noise characteristics has significant performance benefits over one which assumes white noise. It is interesting to note that the ML DOA estimate, assuming k is known, is given as

$$\hat{\boldsymbol{\phi}} = \arg\min_{\boldsymbol{\phi}} \text{trace}\left[\boldsymbol{U}_\nu^H \boldsymbol{R}_{YY} \boldsymbol{U}_\nu\right] \qquad (47)$$

where $\boldsymbol{R}_{YY}$ is the covariance matrix of the observed data. In comparison, the DOA MAP estimate for the proposed method, for the case where k is known, may be written from (22) as

$$\hat{\boldsymbol{\phi}} = \arg\min_{\boldsymbol{\phi}} \left|\boldsymbol{U}_\nu^H \boldsymbol{R}_{YY} \boldsymbol{U}_\nu\right|. \qquad (48)$$

That is, the proposed method minimizes the determinant of a matrix, whereas the ML method minimizes the trace of the same matrix. Further comparisons of the two methods are given in [5].

The probability of detection for the MCMC and W-MDL methods is shown in Table 6.2. Both detection procedures provide similar performance, as indicated in Table 6.2. It is worthy of note that classical methods for detection of model order, such as MDL and AIC [8] (which were developed under the white noise assumption) tend to consistently estimate the number of signals as M in the presence of coloured noise, regardless of the true value, even at high values of SNR [11].

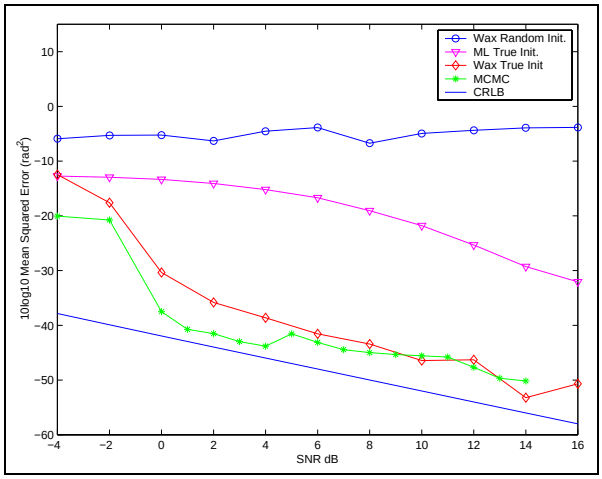

Fig. 5. Mean Squared Error of the DOA estimates vs. SNR, over 50 Monte Carlo runs for the MCMC, W-MDL and ML methods, and the corresponding Cramér Rao bound. The figure shows curves for two different initializations of the Wax method: up-triangles and down-triangles which correspond to initialization to the true and random values, respectively. The ML method was initialized at the true values, whereas the MCMC method was initialized arbitrarily. Curves are shown only for DOA_1 for reasons of clarity.

Further results shown in Figure 6 demonstrate that the probability of an error in detection diminishes towards zero with increasing N for this choice of d^2, thus verifying the development of Section 5.

VII. Conclusion

A new application of the Reversible Jump MCMC method was developed and presented for the problem of joint detection and estimation of sources impinging an array of sensors in spatially coloured noise. This work extends previous methods to the case of noise with arbitrary covariance structure using only a single array. This method is based on the formulation of a posterior *pdf* which has all the nuisance parameters integrated out. Consistent detection has been verified, for values of d^2 in the range given by (45). Simulation results support the effectiveness of the method, and demonstrate reliable

			SNR (dB)						
			-4	-3	-2	-1	0	1	2
M		1	6	14	28	26	40	18	30
C		2	64	82	72	74	60	82	70
M	k	3	28	4					
C		4	2						
		5							
W		1	2						2
M		2	74	82	66	78	72	78	78
D	k	3	18	10	20	12	20	20	10
L		4	4	8	12	2	8	2	6
		5	2		2	8			4

TABLE 6.2
PROBABILITY OF DETECTION (IN %) VS SNR

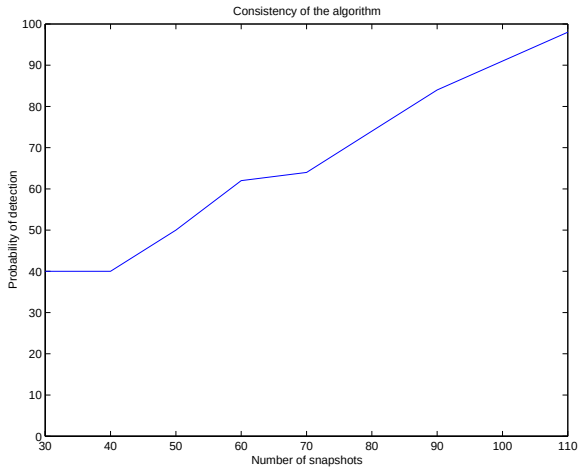

Fig. 6. Probability of detection vs. number of observations.

detection of the number of sources in coloured noise. Although MCMC approaches are computationally intensive, a significant advantage as demonstrated in this work is that they provide the global solution to the difficult problem of joint detection and estimation, with robustness to the initial guess of the parameter values. In comparison, classical approaches have been shown to suffer degradations in DOA estimation performance due to local solutions, unless *a priori* knowledge of the DOA estimates is available, or a global search procedure is used.

VIII. ACKNOWLEDGEMENTS

The authors are most grateful to Dr. Christophe Andrieu for helpful discussions, and to the anonymous reviewers whose comments have resulted in significant improvements to the manuscript. The authors also wish to acknowledge and thank the following institutions for their support of this project: Mitel Corporation, Kanata, Ontario, Canada; The Centre for Information Technology Ontario (CITO); and the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] R.O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, pp. 284–280, Mar. 1986.

[2] C. Andrieu and A. Doucet, "Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC," *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2667–2676, Oct. 1999.

[3] Q. Wu and D.R. Fuhrmann, "A parametric method for determining the number of signals in narrow-band direction finding," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 39, pp. 1848–1857, Aug. 1991.

[4] V. Nagesha and S. Kay, "Maximum likelihood estimation for array processing in coloured noise," *IEEE Transactions on Signal Processing*, vol. 44, no. 2, pp. 169–180, Feb. 1996.

[5] K.M. Wong, J.P. Reilly, Q. Wu, and S. Qiao, "Estimation of the directions of arrival of signals in unknown correlated noise, Part 1: The MAP approach and its implementation," *IEEE Transactions on Signal Processing*, vol. 40, no. 8, pp. 2007–2017, August 1992.

[6] M. Agrawal and S. Prasad, "A modified likelihood function approach to DOA estimation in the presence of unknown spatially correlated Gaussian noise using a uniform linear array," *IEEE Transactions on Signal Processing*, vol. 48, no. 10, pp. 2743–2749, Oct. 2000.

[7] C-M. Cho and P.M. Djuric, "Bayesian detection and estimation of cisoids in colored noise," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2943–2951, Dec. 1995.

[8] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, no. 2, pp. 387–392, Apr. 1985.

[9] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[10] J.P. LeCadre, "Parametric methods for spatial signal processing in the presence of unknown coloured noise fields," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 965–983, July 1989.

[11] W. Chen, J.P. Reilly, and K.M. Wong, "Detection of the number of signals in noise with banded covariance matrices," *IEE Proceedings on Radar, Sonar and Navigation*, vol. 143, pp. 289–294, 1996.

[12] J.J. Fuchs, "Estimation of the number of signals in the presence of unknown correlated sensor noise," *IEEE Transactions on Signal Processing*, vol. 40, pp. 1053–1061, May 1992.

[13] M. Wax, "Detection and localization of multiple sources in noise with unknown covariance," *IEEE Transactions on Signal Processing*, vol. 40, no. 1, pp. 245–249, Jan. 1992.

[14] J-R. Larocque and J.P. Reilly, "Reversible jump MCMC for joint detection and estimation of directions of arrival in coloured noise," in *Proceedings of the IMA International Conference on Mathematics in Signal Processing*, University of Warwick, England, Dec. 2000.

[15] J-R. Larocque and J.P. Reilly, "Application of the reversible jump MCMC method to real-life propagation measurements," in *Proceedings of the International*

Conference on Acoustics, Speech, and Signal Processing, Istanbul, Turkey, June 2000.

[16] P.J. Green, "Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.

[17] H. Jeffreys, *Theory of Probability*, Oxford University Press, London, 3rd edition, 1961.

[18] G.E.P. Box and G. Tiao, *Bayesian Inference in Statistical Analysis*, Addison Wesley, Reading, M.A., 1973.

[19] N.R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution," *Annals of Mathematical Statistics*, vol. 34, pp. 152–171, 1963.

[20] W.K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[21] C. Andrieu, P.M. Djuric, A. Doucet, "Model selection by MCMC computation", *Signal Processing*, Vol 81, January, 2001, pp. 19-37.

[22] P.M. Djuric, "A model selection rule for sinusoids in white Gaussian noise," *IEEE Transactions on Signal Processing*, vol. 44, no. 7, pp. 1744–1751, July 1996.

[23] Q.T. Zhang, K.M. Wong, P.C. Yip and J.P. Reilly, "Statistical analysis of the performance of information theoretic criteria in the detection of number of signals in array processing", IEEE Trans. ASSP, Vol. 37, No. 10, Oct. 1989, pp. 1557-1567.

[24] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, Dover Publications, New York, 1965.

[25] A.B. Gershman, P. Stoica, M. Pesavento, and E. Larsson, "Stochastic Cramér-Rao bound for direction estimation in unknown noise fields," *IEEE Transactions on Signal Processing*, 2001, To appear.

[26] H. Ye and R. DeGroat, "Maximum likelihood DOA estimation and asymptotic Cramér-Rao bounds for additive unknown colored noise," *IEEE Transactions on Signal Processing*, vol. 43, no. 4, pp. 938–949, Apr. 1995.

[27] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 10, pp. 1553–1560, Oct. 1988.

[28] P.J. Van Laarhoven and E.H.L. Arts, *Simulated annealing: Theory and Applications*, Reidel Publishers, Amsterdam, 1987.

[29] C. Andrieu, A. Doucet, P. Duvaut, W. Fitzgerald, S. Godsill, "Simulation-based computational methods for statistical signal processing", Tutorial notes, *ICASSP98*, Seattle, Washington.