INSTITUTE OF PHYSICS PUBLISHING CLASSICAL AND QUANTUM GRAVITY

Class. Quantum Grav. **22** (2005) S901–S911 doi:10.1088/0264-9381/22/18/S04

LISA source confusion: identification and characterization of signals

Richard Umstätter[1], Nelson Christensen[2], Martin Hendry[3], Renate Meyer[1], Vimal Simha[3], John Veitch[3], Sarah Vigeland[2] and Graham Woan[3]

[1] Department of Statistics, University of Auckland, Auckland, New Zealand
[2] Physics and Astronomy, Carleton College, Northfield, MN 55057, USA
[3] Department of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK

E-mail: richard@stat.auckland.ac.nz, nchriste@carleton.edu, martin@astro.gla.ac.uk, meyer@stat.auckland.ac.nz, vimal_simha@hotmail.com, jveitch@astro.gla.ac.uk, vigelans@carleton.edu and graham@astro.gla.ac.uk

Received 30 March 2005, in final form 15 July 2005
Published 23 August 2005
Online at stacks.iop.org/CQG/22/S901

Abstract
The Laser Interferometer Space Antenna (LISA) is expected to detect gravitational radiation from a large number of compact binary systems. We present a method by which these signals can be identified and have their parameters estimated. Our approach uses Bayesian inference, specifically the application of a Markov chain Monte Carlo method. The simulation study that we present here considers a large number of sinusoidal signals in noise, and our method estimates the number of periodic signals present in the data, the parameters for these signals and the noise level. The method is significantly better than classical spectral techniques at performing these tasks and does not use stopping criteria for estimating the number of signals present.

PACS numbers: 04.80.Nn, 02.70.Lq, 06.20.Dq

1. Introduction

LISA [1] is expected to detect a very large number of signals from compact binaries in the 10^{-2} mHz to 100 mHz band, making signal identification very difficult. Tens of thousands of signals could be present in the data with significant signal-to-noise ratios. In the 0.1 mHz to 3 mHz band there will be numerous signals from white dwarf binaries. Sources above 5 mHz should be resolvable, however, below 1 mHz there will be *source confusion*. In the 1 mHz to 5 mHz band we expect as many as 10^5 potential sources [2–4] resulting in an astoundingly difficult data analysis problem. We direct the reader to Barack and Cutler [2] and Nelemans

et al [5] for an in-depth description of the population of binary systems in the LISA operating band, and how LISA's performance is influenced by them.

The goal of this paper is to introduce the LISA data analysis community to a new approach for identifying and characterizing these numerous signals. This method is designed to work both in and out of the source confusion regime. We apply Bayesian Markov chain Monte Carlo (MCMC) methods to a simplified problem that will serve as an example of the technique. MCMC methods are a numerical means of parameter estimation, and are especially useful when there are a large number of parameters [6]. We have already applied MCMC methods to other gravitational radiation parameter estimation problems; for example, we have used a Metropolis–Hastings (MH) algorithm [7, 8] for estimating astrophysical parameters for gravitational wave signals from coalescing compact binary systems [9], and pulsars [10, 11]. We believe that MCMC methods could provide an effective means for identifying sources in LISA data. We summarize our reversible jump MCMC technique in this paper. A more detailed and comprehensive description can be found in our subsequent paper [12], where we extend the work we have initiated here.

Here we present a summary of our study of some simple simulated data, comprising a number of sinusoidal signals embedded in noise. By noise we mean the instrumental noise, plus weak unresolved sources. Our reversible jump MCMC algorithm infers the parameters for each (sufficiently large) sinusoidal signal, the magnitude of the noise and the number of signals present. In our approach we solve both the detection and parameter estimation problems without the need for evaluating formal model selection criteria. The method does not require a stopping criterion for determining the number of signals and produces results which compare very favourably with classical spectral techniques. The number of resolvable sources and the noise level are parameters in our model; unresolved sources, along with instrumental noise, contribute to the overall estimated noise level. A Bayesian analysis naturally encompasses Occam's razor and a preference for a simpler model [13]. In addition, our MCMC method is better than a classical periodogram at resolving signals that are very close in frequency, and we provide an explanation of how to identify these signals.

The MCMC method that we present here performs source modelling rather than source subtraction [14], and avoids the artefacts that can be generated by the sequential removal of signals (though clearly the modelled sources can still be subtracted from the data if desired). Signals for which there is sufficient evidence will be identified with a quoted confidence, and sources that are too weak to be detectable will simply contribute to the noise, the level of which we also estimate. We show that the noise level estimate from our method depends (as it should) on both the inherent detector noise level and the presence of many low-level signals too weak to be detected. A further benefit of using MCMC methods is that computation time does not show an exponential increase with the number of parameters [6].

The problem of identifying an unknown number of sinusoids is neither new nor simple [15, 16]. Previous studies have looked for a handful of unknown signals, here we show results for 100 signals. MCMC methods are robust and dynamic, and we believe that ultimately it will be possible to use them with LISA data to estimate the parameters of all modelled sources types. In the future we will make the model more complex, taking into account the orbit of the LISA spacecraft and binary source evolution.

2. Occam factors

One can approach the problem of identifying and enumerating sinusoidal signals in noise from a number of different directions, but one thing is clear. Discrete noisy data can be fitted *exactly* if one uses a sufficient number of components—the result is simply the discrete Fourier

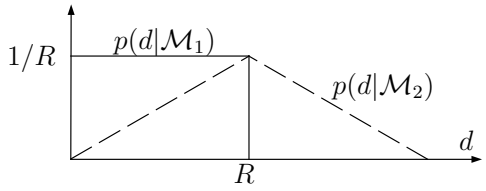

Figure 1. The evidences for models $\mathcal{M}_1$ (solid line, one parameter fit) and $\mathcal{M}_2$ (dashed line, two parameter fit) as a function of the datum d. Note that the evidence ratio always favours the simpler model $\mathcal{M}_1$ when the datum is equally consistent with either.

transform of the data. Classically, we proceed by estimating the noise floor of the spectrum and identify a threshold spectral power that divides the components between signals and noise. In this way we prevent the model from over-fitting the noise. In an iterative fitting procedure, this is achieved by halting when the statistics of the residuals fit the noise model well. One very attractive, and well-known, feature of Bayesian inference is that these ideas are within the fabric of the method. Indeed they are such a basic property of logical inference that there is no need to refer to ideas such as 'over-fitting' at all. More generally, the method discourages us from using models that have more degrees of freedom than are necessary for the problem in hand. Readers desiring a comprehensive and informative description of Bayesian methods are encouraged to consult the following sources [13, 17–22].

One can see how this works in a simple example. Take two data models, $\mathcal{M}_1$ and $\mathcal{M}_2$ constrained by a single datum, d. $\mathcal{M}_1$ has one parameter, a_1, to describe the datum, whereas $\mathcal{M}_2$ uses the sum of two parameters, $s = a_1 + a_2$ to describe the same datum. Which model is better? Here we have no noise and no random variables, so this is not a problem for orthodox statistics. However, if the datum is equally consistent with both models, we would clearly prefer the simpler model $\mathcal{M}_1$ in favour of $\mathcal{M}_2$.

Within the Bayesian framework we consider the odds ratio of the two models:

$$\mathcal{O}_{12} = \frac{p(\mathcal{M}_1|d)}{p(\mathcal{M}_2|d)} = \frac{p(\mathcal{M}_1)}{p(\mathcal{M}_2)}\frac{p(d|\mathcal{M}_1)}{p(d|\mathcal{M}_2)}. \tag{1}$$

We will set $p(\mathcal{M}_1)/p(\mathcal{M}_2)$ to unity, as we have no prior preference for either model, and take the priors for a_1 and a_2 to be each uniform in the range $0 \to R$, making the prior for s the convolution of two of these. The functional priors for a_1 under $\mathcal{M}_1$ and s under $\mathcal{M}_2$ are therefore

$$p(a_1|\mathcal{M}_1) = \frac{1}{R}; \qquad p(s|\mathcal{M}_2) = \begin{cases} s/R^2 & \text{for} \quad 0 < s < R \\ 2/R - s/R^2 & \text{for} \quad R < s < 2R. \end{cases} \tag{2}$$

The probability of the data, given either model, is simply a Dirac delta function centred on the value of the datum, so we can calculate the *evidences* $p(d|\mathcal{M}_1)$ and $p(d|\mathcal{M}_2)$ by marginalizing over the allowed parameter values,

$$p(d|\mathcal{M}_1) = \int_0^R \frac{1}{R}\delta(d - a_1)\,\mathrm{d}a_1 = \begin{cases} 1/R & \text{for} \quad 0 < d < R \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

$$p(d|\mathcal{M}_2) = \int_0^{2R} p(s|\mathcal{M}_2)\delta(d - s)\,\mathrm{d}s = \begin{cases} d/R^2 & \text{for} \quad 0 < d < R \\ 2/R - d/R^2 & \text{for} \quad R < d < 2R \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

as shown in figure 1. If the datum lies in the range $0 < d < R$, the odds ratio is $(1/R)/(d/R^2) = R/d$, so that $\mathcal{M}_1$ will always favoured over $\mathcal{M}_2$ in that range. If $d = R$ the odds ratio is unity, and if $R < d < 2R$ the only possible model is $\mathcal{M}_2$.

This demonstrates why the simpler model is favoured even when the datum is equally consistent with both: $\mathcal{M}_2$ has more flexibility than is necessary to explain the datum and so penalizes itself by spreading its evidence more thinly. Although in this example we consider a probability ratio to determine our favoured model, when more than two models are available, we can consider the model choice to be a parameter itself, and determine its marginal posterior probability in the usual way.

A slightly more pertinent, though still highly restricted, example would be a data set $\{d_k\}$ that consists of observations of the sum of m sinusoids of the form $A_i \sin(2\pi f_i t)$ at times $t = t_k$ and Gaussian noise with variance σ^2. We also know that $0 \leqslant m \leqslant 5$, and A_i and f_i can only take the discrete values $A_i \in \{1, \ldots, 5\}$ and $f_i \in \{0.01, 0.02, \ldots, 0.05\}$. For this discrete problem, assuming uniform priors on m, A_i and f_i for each sinusoid, we can identify the probability of any particular m, given the data, irrespective of the other signal parameters:

$$p(m|\{d_k\}) \propto \frac{1}{25^m} \sum_{A_1, f_1} \cdots \sum_{A_M, f_M} \exp(-\chi^2/2), \tag{5}$$

where

$$\chi^2 = \frac{1}{\sigma^2} \sum_k \left[d_k - \sum_{i=1}^{m} A_i \sin(2\pi f_i t_k) \right]^2. \tag{6}$$

Here it is the factor of 25^m, originating from the m normalized priors for the model parameters, that offsets the increasingly good χ^2 fit that might come from large values of m and provides our Occam factor.

This discrete problem can be solved by directly marginalizing over the nuisance amplitude and frequency parameters. However, problems with more and/or with continuous parameters are not approachable using such direct methods, and must be tackled in another way.

3. Parameter estimation

We consider the continuous case as a signal consisting of m superimposed sinusoids where m is unknown. Therefore, we confine our attention to a set of models $\{\mathcal{M}_m : m \in \{1, \ldots, M\}\}$ where M is the maximum allowable number of sinusoids. Let $\boldsymbol{d} = [d_1, \ldots, d_N]$ be a vector of N samples recorded at times $\boldsymbol{t} = [t_1, \ldots, t_N]$. Model $\mathcal{M}_m$ assumes that the observed data are composed of a signal plus noise: $d_j = \mathrm{f}_m(t_j, \boldsymbol{a}_m) + \epsilon_j$, for $j = 1, \ldots, N$, where the noise terms ϵ_j are assumed to be i.i.d. $N\left(0, \sigma_m^2\right)$ random variables. The signal of model $\mathcal{M}_m$ is assumed to be of the form

$$\mathrm{f}_m(t_j, \boldsymbol{a}_m) = \sum_{i=1}^{m} \left[A_i^{(m)} \cos\left(2\pi f_i^{(m)} t_j\right) + B_i^{(m)} \sin\left(2\pi f_i^{(m)} t_j\right) \right]. \tag{7}$$

Model $\mathcal{M}_m$ is therefore characterized by a vector

$$\boldsymbol{a}_m = \left[A_1^{(m)}, B_1^{(m)}, f_1^{(m)}, \ldots, A_m^{(m)}, B_m^{(m)}, f_m^{(m)}, \sigma_m^2 \right] \tag{8}$$

of $3m + 1$ unknown parameters. The objective is to find the model $\mathcal{M}_m$ that best fits the data. To this end, we use a Bayesian approach as in [21]. The joint probability of these data $\boldsymbol{d}$ given the parameter vector $\boldsymbol{a}_m$ and model $\mathcal{M}_m$ is

$$p(\boldsymbol{d}|m, \boldsymbol{a}_m) \propto \frac{1}{\sigma_m^N} \exp\left\{ -\frac{1}{2\sigma_m^2} \sum_{j=1}^{N} [d_j - \mathrm{f}_m(t_j, \boldsymbol{a}_m)]^2 \right\}. \tag{9}$$

We choose (now continuous) uniform priors for the amplitudes $A_i^{(m)}$, $B_i^{(m)}$ and frequency $f_i^{(m)}$ with ranges $A_i^{(m)} \in [-A_{\max}, A_{\max}]$, $B_i^{(m)} \in [-B_{\max}, B_{\max}]$ and $f_i^{(m)} \in [0, 0.5]$, respectively. Furthermore, we use a uniform prior for m over $\{1, \ldots, M\}$ and vague inverse gamma priors for σ_m^2. By applying Bayes' theorem, we obtain the posterior pdf

$$p(m, \boldsymbol{a}_m|\boldsymbol{d}) = \frac{p(m, \boldsymbol{a}_m)\, p(\boldsymbol{d}|m, \boldsymbol{a}_m)}{p(\boldsymbol{d})}, \tag{10}$$

where $p(\boldsymbol{d}) = \sum_{i=1}^{M} \int p(m, \boldsymbol{a}_m)\, p(\boldsymbol{d}|m, \boldsymbol{a}_m)\, \mathrm{d}\boldsymbol{a}_m$. We use a sampling-based technique for posterior inference via MCMC [6]. MCMC techniques only require the unnormalized posterior $p(m, \boldsymbol{a}_m|\boldsymbol{d}) \propto p(m, \boldsymbol{a}_m)\, p(\boldsymbol{d}|m, \boldsymbol{a}_m)$ to simulate from equation (10) in order to estimate the quantities of interest. However, as the dynamic variable of the simulation does not have fixed dimension, the classical MH techniques [7, 8] cannot be adopted when proposing trans-dimensional moves between models where the model indicator m determines the dimension $(3m+1)$ of the parameter vector $\boldsymbol{a}_m$. We therefore use the reversible jump Markov chain Monte Carlo (RJMCMC) algorithm [23, 24] for model determination, as in [16]. For transitions within the same model, we use the delayed rejection method [25, 26] which yields a better adaptation of the proposals in different parts of the state space.

3.1. The RJMCMC for model determination

To sample from the joint posterior $p(m, \boldsymbol{a}_m|\boldsymbol{d})$ via MCMC, we need to construct a Markov chain simulation with state space $\cup_{m=1}^{M}(m \times \mathbb{R}^{3m+1})$. When a new model is proposed we attempt a step between state spaces of different dimensionality. Suppose that at the nth iteration of the Markov chain we are in state $(k, \boldsymbol{a}_k)$. If model $\mathcal{M}_{k'}$ with parameter vector $\boldsymbol{a}'_{k'}$ is proposed, a reversible move has to be considered in order to preserve the detailed balance equations of the Markov chain. Therefore, the dimensions of the models have to be matched by involving a random vector $\boldsymbol{r}$ sampled from a proposal distribution with pdf q, say, for proposing the new parameters $\boldsymbol{a}'_{k'} = \mathrm{t}(\boldsymbol{a}_k, \boldsymbol{r})$ where t is a suitable deterministic function of the current state and the random numbers. Here we focus on transitions that either decrease or increase models by one signal, i.e. $k' \in \{k-1, k+1\}$. We use equal probabilities $p_{k\mapsto k'} = p_{k'\mapsto k}$ to either move up or down in dimensionality. Without loss of generality, we consider $k < k'$.

If the transformation $\mathrm{t}_{k\mapsto k'}$ from $(\boldsymbol{a}_k, \boldsymbol{r})$ to $\boldsymbol{a}'_{k'}$ and its inverse $\mathrm{t}^{-1}_{k\mapsto k'} = \mathrm{t}_{k'\mapsto k}$ are both differentiable, then reversibility is guaranteed by defining the acceptance probability for increasing a model by one signal according to [23] by

$$\alpha_{k\mapsto k'}(\boldsymbol{a}'_{k'}|\boldsymbol{a}_k) = \min\left\{1, \frac{p(\boldsymbol{a}'_{k'}, k')\, p(\boldsymbol{d}|\boldsymbol{a}'_{k'}, k')\, p_{k\mapsto k'}}{p(\boldsymbol{a}_k, k)\, p(\boldsymbol{d}|\boldsymbol{a}_k, k)\, q(\boldsymbol{r})\, p_{k'\mapsto k}} |J_{k\mapsto k'}|\right\}, \tag{11}$$

where $|J_{k\mapsto k'}| = \left|\frac{\partial \mathrm{t}(\boldsymbol{a}_k, \boldsymbol{r})}{\partial(\boldsymbol{a}_k, \boldsymbol{r})}\right|$ is the Jacobian determinant of this transformation and $q(\boldsymbol{r})$ is the proposal distribution.

In this context, two types of transformations, 'split-and-merge' and 'birth-and-death', are obvious choices. In a 'split-and-merge' transition, the proposed parameter vector $\boldsymbol{a}'_{k'}$ comprises all $(k-1)$ subvectors of $\boldsymbol{a}_k$ except a randomly chosen subvector $\boldsymbol{a}_{(i)} = \left(A_i^{(k)}, B_i^{(k)}, f_i^{(k)}\right)'$ which is replaced by two three-dimensional subvectors,

$$\boldsymbol{a}'_{(i_1)} = \left(A_{i_1}^{(k')}, B_{i_1}^{(k')}, f_{i_1}^{(k')}\right) \qquad \text{and} \qquad \boldsymbol{a}'_{(i_2)} = \left(A_{i_2}^{(k')}, B_{i_2}^{(k')}, f_{i_2}^{(k')}\right)$$

with roughly half the amplitudes but about the same frequency as $\boldsymbol{a}_{(i)}$.

A three-dimensional Gaussian random vector (with mean zero), $\boldsymbol{r} = (r_A, r_B, r_f)$, changes the current state $\boldsymbol{a}_{(i)}$ to the two resulting states $\boldsymbol{a}'_{(1i)}$, $\boldsymbol{a}'_{(2i)}$ through a linear transformation

$$\begin{aligned} \mathrm{t}_{k\mapsto k'}(\boldsymbol{a}_{(i)}, \boldsymbol{r}) &= \left(\tfrac{1}{2}A_i^{(k)} + r_A, \tfrac{1}{2}B_i^{(k)} + r_B, f_i^{(k)} + r_f, \tfrac{1}{2}A_i^{(k)} - r_A, \tfrac{1}{2}B_i^{(k)} - r_B, f_i^{(k)} - r_f\right) \\ &= \left(A_{i_1}^{(k')}, B_{i_1}^{(k')}, f_{i_1}^{(k')}, A_{i_2}^{(k')}, B_{i_2}^{(k')}, f_{i_2}^{(k')}\right). \end{aligned} \tag{12}$$

By analogy, the inverse transformation $\mathrm{t}^{-1}_{k\mapsto k'} := \mathrm{t}_{k'\mapsto k}$ accounts for the merger of two signals. Note that the determinant of the Jacobian of the transformation $\mathrm{t}_{k\mapsto k'}$ is $|J_{k\mapsto k'}| = 2$, and that of its inverse is $1/2$. We use a multivariate normal distribution, $N\left[\mathbf{0}, \mathrm{diag}\left(\sigma_A^2, \sigma_B^2, \sigma_f^2\right)\right]$, for the proposal distribution $q(\boldsymbol{r})$. Care has to be taken in choosing suitable values for the proposal variances to achieve reasonable acceptance probabilities in equation (11).

The second 'birth-and-death' transformation consists of the creation of a new signal with parameter triple $\boldsymbol{a}'_{(i)}$ independent of other existing signals in the current model $\mathcal{M}_k$. The one-to-one transformation in this case is very simply given by $\mathrm{t}_{k\mapsto k'}(\boldsymbol{r}) = \boldsymbol{r} = \boldsymbol{a}'_{(k+1)}$ with Jacobian equal to 1.

Here, q is the three-dimensional pdf from which we draw proposals for the additional signal. We use independent uniform distributions with frequency range $0 \leqslant f \leqslant 0.5$ and amplitude range $\left(A_i^2 + B_i^2\right)^{1/2} < l_A$, where l_A is the radius for the two amplitudes of the signal in polar coordinates. Again, the radii of the uniform proposal densities have to be tuned to achieve an optimal acceptance rate.

3.2. The delayed rejection method for parameter estimation

For transitions within a model $\mathcal{M}_m$, classical MCMC methods can be applied. Here, however, we use an adaptive MCMC technique, the delayed rejection method [25–27] that we have successfully applied to estimate parameters of pulsars [11]. The idea behind the delayed rejection method is that persistent rejection indicates that locally, the proposal distribution is badly calibrated to the target. Therefore, the MH algorithm is modified so that on rejection, a second attempt to move is made with a proposal distribution that depends on the previously rejected state. In this context, when a proposed MH move is rejected from a bold normal distribution with large variance a second candidate can be proposed with a timid proposal distribution for sampling the parameters for the individual sinusoids. Hence, the main objective of the first stage is a coarse scan of the parameter space and therefore we choose the variances of the parameters about one order of magnitude smaller than the prior ranges of the corresponding parameters. Once a mode is found, we aim to draw representative samples in the second stage.

The precision of the frequency in a single-frequency model depends on the amplitude, the variance σ^2 of the noise and the number of samples N of the data set [21, 28]. The precision of the frequency has been derived in [21] by a Gaussian approximation to the posterior pdf of the frequency and calculation of its standard deviation, given by $\sigma''_f = (2\pi)^{-1}[48\sigma^2/N^3(A^2 + B^2)]^{1/2}$. We therefore choose proposals with this standard deviation.

3.3. Starting values

The starting values of a Markov chain are crucial for the length of the burn-in period, i.e. the time needed for the chain to achieve convergence to the real posterior distribution. We perform a fast Fourier transformation (FFT) prior to the simulation and use corresponding estimates as

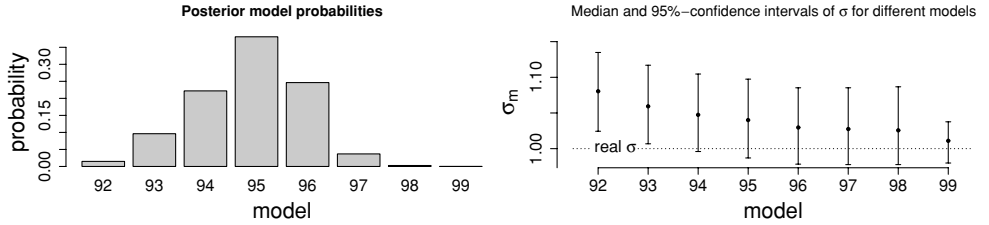

Figure 2. Posterior distribution of model number m with 5%, 50% and 95% posterior confidence intervals of the corresponding noise levels.

starting values. Schuster introduced the periodogram [29]

$$C(f) = \frac{1}{N}[R(f)^2 + I(f)^2], \tag{13}$$

where $R(f) = \sum_{j=1}^{N} d_j \cos(2\pi f t_j)$ and $I(f) = \sum_{j=1}^{N} d_j \sin(2\pi f t_j)$ are the real and imaginary parts from the sums of the discrete Fourier transformation. As a starting value for m, we use the number m_0 of local maxima in the periodogram that exceed a certain noise level (lower than the expected one). We use the frequencies corresponding to the local maxima in the periodogram, $f_{0,i}$, as starting values for $f_{0,i}^{(m_0)}$ and $A_{0,i} = 2R(f_{0,i})/N$, $B_{0,i} = 2I(f_{0,i})/N$ as starting values for $A_i^{(m_0)}$ and $B_i^{(m_0)}$, respectively.

3.4. Identifying the sinusoids

Although the RJMCMC offers great ease in model selection, we still encounter the *label-switching* problem, a general problem due to invariance of the likelihood under relabelling that has been extensively discussed in the context of mixture models [30]. The sinusoids that are contained in the model $\hat{m}$ with the highest posterior probability of m are permutations of $\hat{m}$ coexistent sinusoids out of a number of sinusoids that we do not know but can estimate by the upper limit $m_{\max}$ of the marginal posterior of m. Therefore, the parameter vector sampled in each iteration of the Markov chain (corresponding to model $\hat{m}$) is a permutation of $\hat{m}$ parameter triples determining $\hat{m}$ out of $m_{\max}$ sinusoids. The problem is to determine which parameter triple belongs to which sinusoid.

The parameter that contributes significantly to identifying a sinusoid is its frequency. We thus calculate the marginal posterior of the frequency and obtain the $m_{\max}$ strongest peaks together with their frequency ranges by finding the threshold that separates those peaks. It is still possible that individual peaks contain more than one sinusoid or even none. This can be assessed by a histogram similar to that in figure 2 but restricted to the frequency range under consideration. To separate more than one present sinusoids, we then consider the two amplitudes and apply an agglomerative hierarchical cluster analysis that involves all three parameters. We use a modified Ward technique [31] that minimizes the within group variance using a normalized Euclidean distance between the parameters by adjusting the frequency range to the much larger range of the amplitudes.

4. Results

We created an artificial data set of 1000 samples from $m = 100$ sinusoids. The sinusoids were randomly chosen with maximum amplitudes 1 and the noise standard deviation was

$\sigma = 1$. We will present our results with dimensionless units. We chose a uniform prior for m on $\{0, 1, 2, \ldots, M = 60000\}$, and set $A_{\max} = B_{\max} = 5$. The Markov chain ran for 10^8 iterations and was thinned by storing every 1000th iteration. The first 5×10^6 iterations where considered as burn-in and discarded. The MCMC simulation was implemented in C on a 2.8 GHz Intel P4 PC and took about 43 h to run. Figure 2 gives the histogram of the posterior model probabilities obtained by the reversible jump algorithm. As each model $\mathcal{M}_m$ is characterized by a different noise level σ_m, we have also plotted the posterior distributions of the noise standard deviations for increasing model order. Note that σ_m decreases with higher model order m since a model comprising more sinusoids accounts for more noise. This illustrates our previous statement that the total noise is the sum of instrumental noise and any unresolved sources. Here, we subsequently choose and analyse model $m = 95$, corresponding to the posterior mode of m as the best fitting model. The five unresolved signals were either too small in amplitude to be detected or could not be resolved from neighbouring sinusoids. In the first case the sinusoids will contribute to the overall noise level determination for model $m = 95$. In the second case we estimate different parameters for the involved sinusoids.

We used all MCMC samples corresponding to model $\mathcal{M}_{95}$. For ease of notation, we denote the parameter vector of model $\mathcal{M}_{95}$ by $(A_1, B_1, f_1, \ldots, A_{95}, B_{95}, f_{95})$. The complete line power spectrum density can be estimated by the product of the conditional expectation of the energy $E\left(A_i^2 + B_i^2 \middle| \boldsymbol{d}, m, f_i\right)$ of each sinusoid i given its frequency f_i, and the posterior pdf of f_i given the data, $p(f_i|\boldsymbol{d})$. One of the advantages of the Bayesian spectrum analysis is the possibility to calculate confidence areas for the spectrum. Therefore, we group our MCMC samples and calculate posterior confidence intervals for each frequency bin. A sufficient width for the bins can be assessed by the frequency accuracy $\sigma_f = (2\pi \cdot \mathrm{snr})^{-1}(48/N^3)^{1/2}$ given by [21], where 'snr' is the signal-to-noise ratio. The ability to resolve periodic sources is set by this expression, and we explore this topic in greater detail in our subsequent publication [12]. In our example, the choice of 30 000 bins is sufficient to resolve sinusoids with an snr of about 2. Figure 3 shows the real signals, the Bayes spectrum and the classical Schuster periodogram mentioned in equation (13).

The plot for the real signals displays an individual energy contribution for each sinusoid i of $\left(A_i^2 + B_i^2\right)N/2$. Normally a theoretical spectrum would consist of delta functions with infinitely large energy peaks since the energy contribution is concentrated on an interval of infinitely small width. Therefore, we just plotted the energy contribution on the energy scale that yields a similar scaling as obtained by the periodogram.

In order to be able to display 95% confidence areas of the spectrum, we present three magnified areas in figure 4. In plot (*a*) we see a sinusoid with rather small energy. The accuracy of the frequency estimation is worse compared with the sinusoid of graph (*b*) which has a significantly larger energy. The third graph (*c*) shows two very close sinusoids. The frequency estimation is very inaccurate due to the interference of the two signals. This is consistent with theoretical results by [21]. The interference of the two close signals is due to a phase shift of 175°. The interference and hence the frequency estimation depends upon whether the sinusoids are orthogonal [21] or not. Nevertheless, we are able to identify the existence of two signals while the periodogram only reveals the existence of one. This displays the utility of our method in a source confusion regime. Our method is able to separate and distinguish nearby peaks better than classical techniques. Classically confused signals are distinguishable in they are sufficiently strong; a comprehensive discussion of this point is presented in our subsequent study [12].

The estimates of the amplitudes, however, always show huge values and confidence intervals for sinusoids close in frequency. The huge energies are merely restricted by the choice of priors for the amplitudes. The reason for this is due to the possible combinations to

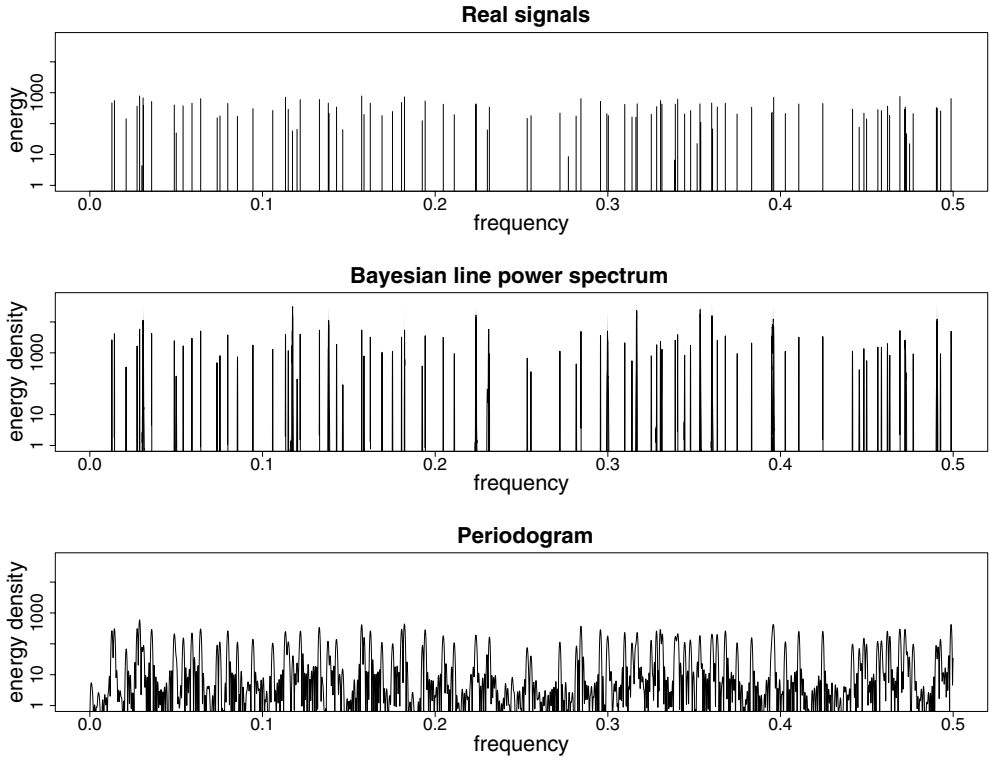

Figure 3. Comparison of real signals, classical Schuster periodogram and Bayesian spectrum. Note that the total energy of the three spectra is similar but the lines have different accuracies. Therefore, the energy of a single line is more concentrated due to the better accuracy of the Bayesian spectrum estimate resulting in higher peaks. The theoretical spectrum consists of delta functions for each sinusoid which would yield infinite peaks. Therefore, an energy contribution of $\frac{N}{2}(A_i^2 + B_i^2)$ for each sinusoid with frequency f_i is shown.

express a sum of sinusoids when the observation time is insufficiently long with respect to the distance in frequency. In this case we cannot make accurate statements about the amplitudes and hence energies of both sinusoids.

If we take a look at the single peak of the periodogram the energy that is considered is subject to the data from a discrete and finite observation time, given by $\sum_{j=1}^{N} \mathrm{f}_m(t_j, \boldsymbol{a}_m)$. This, however, does not reflect the energy contribution of the real signals. The Bayesian estimates of the amplitudes are honest by yielding large confidence areas for the energies of sinusoids close in frequency, but produce small confidence intervals for isolated sinusoids.

5. Discussion

We have presented a Bayesian approach to identifying a large number of unknown periodic signals embedded in noisy data. A reversible jump MCMC technique can be used to estimate the number of signals present in the data, their parameters, and the noise level. This approach allows for simultaneous detection and parameter estimation, and does not require a stopping criterion for determining the number of signals. The MCMC method compares favourably with classical spectral techniques.

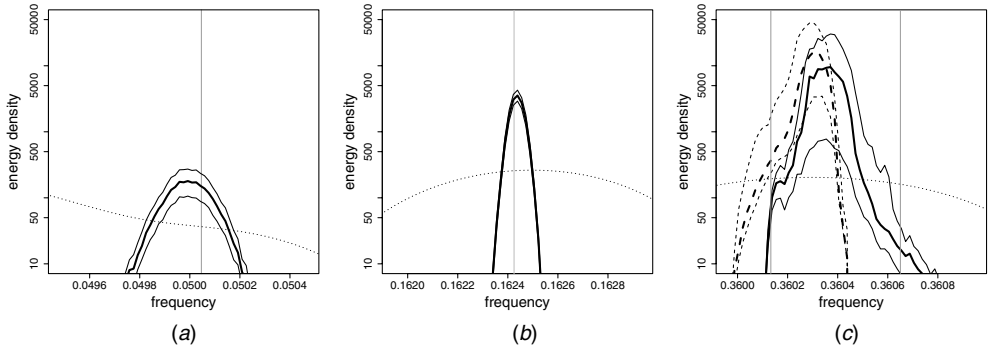

Figure 4. Three magnified areas for a comparison of the classical periodogram (dotted line) and the Bayesian spectrum (continuous lines or dashed if referring to more than one sinusoid). The areas within the thin lines (continuous or dashed) show the 95% confidence areas of the energy for a sinusoid whereas the bold lines within the 95% confidence areas indicate the median of the energy. The vertical lines show the delta functions of the theoretical spectrum. The energies are stated. (*a*) Median and 95% confidence area of energy for sinusoid: $E = 78.2$ [45.96,119.9]. Energies from corresponding real parameters: $E = 49.9$ at $f = 0.050\,0462$. (*b*) Median and 95% confidence area of energy for sinusoid: $E = 476.1$ [389.9, 570.7]. Energies from corresponding real parameters: $E = 459.5$ at $f = 0.162\,425$. (*c*) Median and 95% confidence area of energy for sinusoids: $E = 3497.2$ [819.7, 10 800.8], $E = 2731.6$ [261.8, 8809.1]. Energies from corresponding real parameters: $E = 462.7$ at $f = 0.360\,133$, $E = 66.9$ at $f = 0.360\,65$.

Our motivation for this research is to address the difficulty that LISA will ultimately encounter in having too many signals present. LISA may see 100 000 signals from binary systems in the 1 mHz to 5 mHz band. We see our work as a new method that could help LISA to identify and characterize these signals. The work here is a simplified problem, one that neglects the time evolution of the signal and modulation due to LISA's orbit. The next step is to deal with these more complicated signals, and to develop a realistic strategy for applying our MCMC methods to more realistic LISA data. The inclusion of Doppler shifts due to the LISA orbit could potentially improve the ability to resolve signals that are very close in frequency; this will be especially true when two sources are in different parts of the sky and experience non-equal Doppler shifts. What is considered to be the confusion frequency limit for LISA could decrease when the Doppler modulations of signal frequencies are folded into the problem. We believe that MCMC methods, such as those presented here, provide a practical and highly effective method of identifying and characterizing the large number of signals that will exist in the LISA data.

Acknowledgments

This work was supported by National Science Foundation grants PHY-0071327 and PHY-0244357, the Royal Society of New Zealand Marsden fund award UOA204, Universities UK and the University of Glasgow.

References

[1] Danzmann K and Rüdiger A 2003 *Class. Quantum Grav.* **20** S1
[2] Barack L and Cutler C 2004 *Phys. Rev.* D **70** 122002
[3] Benacquista M J, DeGoes J and Lunder D 2004 *Class. Quantum Grav.* **21** S509–14
[4] Crowder J and Cornish N J 2004 *Phys. Rev.* D **70** 082004

[5] Nelemans G, Yungleson L R and Protegies Zwart S F 2001 *Astron. Astrophys.* **375** 890
[6] Gilks W R, Richardson S and Spiegelhalter D J 1996 *Markov Chain Monte Carlo in Practice* (London: Chapman and Hall)
[7] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 *J. Chem. Phys.* **21** 1087
[8] Hastings W K 1970 *Biometrika* **57** 97
[9] Christensen N, Meyer R and Libson A 2004 *Class. Quantum Grav.* **21** 317
[10] Christensen N, Dupuis R J, Woan G and Meyer R 2004 *Phys. Rev.* D **70** 022001
[11] Umstätter R, Meyer R, Dupuis R J, Veitch J, Woan G and Christensen N 2004 *Class. Quantum Grav.* **21** S1655
[12] Umstätter R, Christensen N, Hendry M, Meyer R, Simha V, Veitch J, Vigeland S and Woan G 2005 *Phys. Rev.* D **72** 022001
[13] Jaynes E T and Bretthorst G L 2003 *Probability Theory: The Logic of Science* (Cambridge: Cambridge University Press)
[14] Cornish N J and Larson S L 2003 *Phys. Rev.* D **67** 103001
[15] Richardson S and Green P J 1997 *J. R. Stat. Soc.* B **59** 731
[16] Andrieu C and Doucet A 1999 *IEEE Trans. Signal Process.* **47** 2667
[17] Loredo T J 1990 From Laplace to supernova SN 1987A: Bayesian inference in astrophysics *Maximum-Entropy and Bayesian Methods* ed P Fougere (Dordrecht: Kluwer) pp 81–142
[18] Neal R M 1993 Probabilistic inference using Markov chain Monte Carlo methods *Technical Report* CRG-TR-93-1 (Department of Computer Science, University of Toronto)
[19] Kass R E, Carlin B P, Gelman A and Neal R M 1998 *Am. Stat.* **52** 93
[20] Gregory P C 2005 *Bayesian Logical Data Analysis for the Physical Sciences* (Cambridge: Cambridge University Press)
[21] Bretthorst G L 1988 *Bayesian Spectrum Analysis and Parameter Estimation (Springer Lecture Notes in Statistics* vol 48*)* (Berlin: Springer)
[22] Gelman A, Carlin J B, Stern H S and Rubin D B 2003 *Bayesian Data Analysis* 2nd edn (London: Chapman and Hall)
[23] Green P J 1995 *Biometrika* **82** 711
[24] Green P J 2003 *In Highly Structured Stochastic Systems* ed P J Green, N L Hjort and S Richardson (Oxford: Oxford University Press)
[25] Mira A 1998 Ordering, slicing and splitting Monte Carlo Markov chain *PhD Thesis* University of Minnesota
[26] Tierney L and Mira A 1999 *Stat. Med.* **18** 2507
[27] Green P J and Mira A 2001 *Biometrika* **88** 1035
[28] Jaynes E T 1987 Bayesian spectrum analysis and chirp analysis *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems* ed C Ray Smith and G J Erickson (Dordrecht: Reidel) pp 1–37
[29] Schuster A 1905 *Proc. R. Soc.* **77** 136
[30] Celeux G, Hurn M and Robert C P 2000 *J. Am. Stat. Assoc.* **95** 957–70
[31] Ward J H 1963 *J. Am. Stat. Assoc.* **58** 236–44