

# Comparing Bayesian Methods for Estimating Parameters of Noisy Sinusoids

*Dursun Üstündağ*

Department of Mathematics,  
Faculty of Science and Arts,  
Marmara University, Istanbul, Turkey  
[dustundag@marmara.edu.tr](mailto:dustundag@marmara.edu.tr)

*Mehmet Cevri*

Department of Mathematics  
Faculty of Science,  
İstanbul University, Istanbul, Turkey  
[mcevri@gmail.com.tr](mailto:mcevri@gmail.com.tr)

**Abstract**— This paper deals with parameter estimation of sinusoids within a Bayesian framework, where inferences about parameters require an evaluation of complicated high dimensional integrals or a solution of multi-dimensional optimization. Unfortunately, it is not possible in general to derive analytical Bayesian inferences. Therefore, the purpose of this paper is to study some of existing stochastic procedures, based on different sampling schemes and to compare their performances with respect to Cramér-Rao lower bound (CRLB), defined to be a limit on the best possible performance achievable for a method given a dataset. Furthermore, all simulations support their effectiveness and demonstrate their performances in terms of CRLB for different lengths of data sampling and signal-to-noise ratio (SNR) conditions.

**Keywords**— Bayesian inference; parameter estimation; Gibbs sampling; simulated annealing; parallel tempering; Cramér-Rao lower bound; power spectral density.

## I. INTRODUCTION

In experimental science, it is hard to find any experiment where we can measure directly desire quantities that characterize physical systems. Therefore, those quantities that we would like to determine are different from the data we measured. However, it contains at least some information about them so that extracting this information is subject to this paper.

In a wide range of applications, a discrete data set  $\mathbf{D} = \{d_i\}_{i=1}^N$  denoted as an output of a physical system that we want to be modeled is sampled from an un known function  $y(t)$  at discrete times  $\{t_i\}_{i=1}^N$ :

$$\begin{aligned} d_i &= y(t_i) \\ &= f(t_i; \boldsymbol{\theta}) + e(t_i), \quad (i = 1, \dots, N), \end{aligned} \quad (1)$$

where  $\boldsymbol{\theta}$  is a set of parameters that characterize behavior of physical systems  $f(t_i; \boldsymbol{\theta})$  which is time dependent systematic component, called model function whose choices depend on applications. We initially restrict our attention to the real

static<sup>1</sup> sinusoidal model which is a superposition of  $m$  sinusoids:

$$d_i = \sum_{j=1}^m a_{c_j} \cos(t\omega_j) + a_{s_j} \sin(t\omega_j) + e(t_i), \quad (2)$$

Due to its applicability, it has received a great interest in many fields of science [1, 36, 37]. Especially, the frequency parameter has been subject to extensive research since it enters the signal model in non-linear fashion. The term  $e(t)$  represents a random process at time  $t$ , due to measurement error. Then Equation (2) can commonly be written in the matrix-vector form:

$$\mathbf{D} = \mathbf{G}\mathbf{a} + \mathbf{e}, \quad (3)$$

where  $\mathbf{D}$  is  $(N \times 1)$  matrix of data points;  $\mathbf{e}$  is  $(N \times 1)$  matrix of independent identically distributed Gaussian noise samples with variance  $\sigma^2$ ;  $\mathbf{G}$  is  $(N \times 2m)$  matrix whose each column is a basis function evaluated at each point of time series and  $\mathbf{a}$  is  $(2m \times 1)$  matrix whose components are arranged in order of coefficients of cosine and sine terms.

The vector  $\boldsymbol{\theta} = \left\{ \left\{ a_{c_j}, a_{s_j}, \omega_j \right\}_{j=1}^m, \sigma^2 \right\}$  consists of parameters, which belong to signals and noise. Then, the goal of data analysis is usually to infer  $\boldsymbol{\theta}$  from  $\mathbf{D}$ . Besides estimating them, there are two additional important problems left. The first one we will not consider here is to assess whether or not the model function  $f(t_i; \boldsymbol{\theta})$  is appropriate for explaining the data, i.e., a number of sinusoids,  $m$ , may not be known a priori so that it is called model selection[33, 36]. The second one we studied here is to obtain an indication of uncertainties in parameter values by using different methods, i.e. some measures of how far they are away from their true values.

In this paper, we consider to introduce some improvements of Bayesian methods that use three different stochastic sampling procedures for estimating parameters of

<sup>1</sup> Static refers to that the amplitudes of the sinusoids do not change with time.

noisy sinusoids. We first combine BRETTHORST method with a simulated annealing (SA); secondly, we derive an extension of Gibbs (GIBBS) sampling to multi-dimensional cases and finally, we use parallel tempering (PT) algorithm to implement a flexible choice of priors. Although different numerical procedures for the Bayesian parameter estimation problem have been proposed by different researchers in statistical signal processing literature, there has been a little work about a comparison of their performances. Therefore, series of computer simulation studies with a variation in the signal to noise ratio (SNR) and the length of the data sample  $N$  are set up for a frequency estimation of a single sinusoid in order to assess the best achievable accuracy of parameter estimates with respect to CRLB.

## II. BAYESIAN DATA ANALYSIS

To estimate the characteristics (e.g., amplitudes, frequencies) of signals from  $\mathbf{D}$  the Bayesian framework is interested in the current context because it provides a mathematical foundation for making inferences about them and, as a consequent, provides a rigorous basis for quantifying uncertainties in their estimates. Therefore, the basic relationship quantifying those parameter inferences is given by Bayes' rule [3, 12, 13, 39]:

$$p(\boldsymbol{\theta}|\mathbf{D}, I) = \frac{p(\boldsymbol{\theta}|I)p(\mathbf{D}|\boldsymbol{\theta}, I)}{p(\mathbf{D}|I)}, \quad (4)$$

where  $I$  represents the prior information;  $p(\boldsymbol{\theta}|I)$  is the prior PDF of the parameter vector  $\boldsymbol{\theta}$  that encapsulates our state of knowledge of the parameters before the receipt of the measurements  $\mathbf{D}$ ;  $p(\mathbf{D}|\boldsymbol{\theta}, I)$  is termed the likelihood function when considered as a function of  $\boldsymbol{\theta}$ , but is known as the sampling distribution when considered as a function of  $\mathbf{D}$ ;  $p(\boldsymbol{\theta}|\mathbf{D}, I)$  is the posterior PDF of  $\boldsymbol{\theta}$ , that corresponds to the update of  $p(\boldsymbol{\theta}|I)$  incorporating the knowledge gained about  $\boldsymbol{\theta}$  after the receipt of the observations  $\mathbf{D}$ . By use of Bayesian inference, All information in  $\mathbf{D}$  relevant to the problem of estimating  $\boldsymbol{\theta}$  is summarized in the posterior PDF of  $\boldsymbol{\theta}$ . Furthermore,  $p(\mathbf{D}|I)$  is termed as evidence which is a measure of the probability but, it is constant in parameter estimation so that Equation (4) becomes

$$p(\boldsymbol{\theta}|\mathbf{D}, I) \propto p(\boldsymbol{\theta}|I)p(\mathbf{D}|\boldsymbol{\theta}, I). \quad (5)$$

To proceed further in the specification of  $p(\boldsymbol{\theta}|\mathbf{D}, I)$ , we now need to assign a functional form for the terms  $p(\boldsymbol{\theta}|I)$  and  $p(\mathbf{D}|\boldsymbol{\theta}, I)$ . Because different methods use different prior PDFs  $p(\boldsymbol{\theta}|I)$ , we postpone discussing their assignments to the next sections but, in the case of

independent measurements the assignment of a functional form for  $p(\mathbf{D}|\boldsymbol{\theta}, I)$  becomes

$$p(\mathbf{D}|\boldsymbol{\theta}, I) = \frac{1}{\prod_{i=1}^N \sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2}\chi^2(\boldsymbol{\theta})\right), \quad (6)$$

where

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^N \left(\frac{d(t_i) - f(t_i, \boldsymbol{\theta})}{\sigma_i}\right)^2. \quad (7)$$

After making assignments to the prior and posterior PDF, the problem turns out to search  $\boldsymbol{\theta}$  in a parameter space  $\mathfrak{S}$ :

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \mathfrak{S}} \{p(\boldsymbol{\theta}|\mathbf{D}, I)\}. \quad (8)$$

## III. BAYESIAN METHODS

### A. Bretthorst's Integral Method with SA

Let us rewrite the joint posterior PDF of all parameters in Equation (5):

$$p(\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s, \sigma^2 | \mathbf{D}, I) \propto p(\mathbf{D} | \boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s, \sigma^2, I) \times p(\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s, \sigma^2 | I). \quad (9)$$

From the product rule of the probability using together with independent property the joint prior PDF of all parameters can then be written in the form:

$$p(\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s, \sigma^2 | I) \propto p(\mathbf{a}_c, \mathbf{a}_s | I) p(\boldsymbol{\omega} | I) p(\sigma^2 | I). \quad (10)$$

If  $p(\mathbf{a}_c, \mathbf{a}_s | I)$  is assigned to a uniform prior PDF and  $p(\sigma^2 | I)$  is a constant because  $\sigma$  is assumed to be known, then the joint prior PDF of all the parameters may be reduced to uninformative prior PDF for  $\boldsymbol{\omega}$  which is bounded in the interval  $(0, \pi)$ :

$$p(\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s, \sigma^2 | I) \propto \pi^{-m}. \quad (11)$$

This is because there is no variation at negative frequencies and the highest possible frequency corresponds to wave that under goes a complete cycle in two unit intervals, so that lower limit on the range is 0 and all the variation is accounted for by frequencies less than  $\pi$ . By using Equations (6) and (11) and dropping constant terms the posterior PDF of the parameters given in Equations (5) becomes

$$p(\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s, \sigma^2 | I) \propto \exp\left(-\frac{1}{2}\chi^2(\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s, \sigma^2)\right). \quad (12)$$

In order to obtain the marginal PDF of  $\boldsymbol{\omega}$ , we need to take the integration of Eq. (12) with respect to the amplitudes and the noise variance  $\sigma^2$ . If  $\sigma^2$  is known, then the posterior PDF of  $\boldsymbol{\omega}$  is given by

$$p(\boldsymbol{\omega} | \mathbf{D}, I) \propto \exp\left(\frac{m\mathbf{h}^2}{2\sigma^2}\right), \quad (13)$$

where  $h_l = \sum_{i=1}^N d_i H_l(t_i)$ , ( $l=1, \dots, 2m$ ) is a projection of data onto new orthogonal model functions  $H_l(t)$ :

$$H_j(t) = (\lambda_j)^{-1/2} \sum_{l=1}^{2m} \varphi_{jl} G_l(t, \omega). \quad (14)$$

Here  $\varphi_{lj}$  represents the  $j$ th component of the  $l$ th normalized eigenvector of the matrix  $\Omega = \mathbf{G}^T \mathbf{G}$ , with  $\lambda_l$  as the corresponding eigenvalue. If the variance of the noise  $\sigma^2$  is unknown, then it is known as a scale parameter so that the completely uninformative prior PDF for a scale parameter is the Jeffreys' prior [14], defined as  $p(\sigma^2 | I) \propto \frac{1}{\sigma^2}$   $0 < \sigma < \infty$ . By using this prior PDF in Equation (12) and integrating it with respect to  $\sigma^2$  we obtain

$$p(\omega | \mathbf{D}, I) \propto \left( 1 - \frac{m \overline{\mathbf{h}^2}}{N \mathbf{D}^2} \right)^{\frac{m-N}{2}}, \quad (15)$$

where  $\overline{\mathbf{h}^2}$  represents the mean squared observed projections. This is the form of the "Student's t distribution" with  $(N - m)$  degree of freedom.

The approach summarized above requires analytical or numerical approximation of integrals which is not given here but, we refer to Bretthorst's work [6]. Consequently, the Bayesian parameter estimation problem turns into maximization of the posterior PDF of  $\omega$  given in Equations (13) and (15) in the parameter space  $\mathfrak{Z} = \{0, \pi\}$ . Unfortunately, conventional algorithms [15, 16] based on the gradient direction fail to converge. Even when they converge, there is no assurance that they have found a global, rather than a local maximum. This is because the logarithm of the posterior PDF is so sharply peaked and highly nonlinear function of  $\omega$ . To overcome this problem, Bretthorst used a pattern search algorithm described by Hook-Jeeves [17] but, we found out that this approach does not converge unless the starting point is much closer to the optimum  $\omega$ . Therefore, we combined it with a simulated annealing (SA) algorithm [18, 19] to obtain a global maximum of the posterior PDF of the frequencies  $\omega$ . For detail information, we refer to our papers [26, 27, 28] and book's chapter [32].

### B. Gibbs Sampling

In order to avoid solving the difficult multivariate maximization problem in Section III.A, an alternative way proposed by Dou and Hodgson [7] combines Gibbs (GIBBS) sampling with Bayesian inference. Basically, it draws samples from the desired marginal distributions condition on the remaining unknown parameters, the data and the prior information. We extend this derivation for multiple frequency signals and summarize it below, but refer to their papers [7, 8, 39] and recently [34] for detail information.

Assume that  $\sigma^2$  is known and there is no any specific information about the parameters  $\{\omega, \mathbf{a}_c, \mathbf{a}_s\}$ . Then Equation (5) turns out to be the following form:

$$p(\omega, \mathbf{a}_c, \mathbf{a}_s | \mathbf{D}, \sigma^2, I) \propto p(\mathbf{D} | \omega, \mathbf{a}_c, \mathbf{a}_s, \sigma^2, I), \quad (16)$$

Because of  $p(\omega, \mathbf{a}_c, \mathbf{a}_s | I) \propto$  constant or flat prior PDF, called an uninformative prior then the mode of the posterior is the same as the maximum of the likelihood. Suppose also that  $a_{c_j}$  is the only unknown parameter among  $\{\mathbf{a}_{c_j}, \mathbf{a}_s, \omega\}$ , where  $\mathbf{a}_{c_j} = \{a_{c_1}, \dots, a_{c_{j-1}}, a_{c_{j+1}}, \dots, a_{c_m}\}$ . If the distribution of the noise is known as a-priori, the conditional PDF of  $a_{c_j}$  is considered to be as a univariate normal distribution:

$$p(a_{c_j} | \mathbf{a}_{c_j}, \mathbf{a}_s, \omega, \mathbf{D}, \sigma^2) \propto N(\hat{a}_{c_j}, \sigma^2 (\mathbf{X}_{c_j}^T \mathbf{X}_{c_j})^{-1}), \quad (17)$$

where

$$\hat{a}_{c_j} = \frac{\hat{\mathbf{D}}^{(1)} \mathbf{X}_{c_j}}{\mathbf{X}_{c_j}^T \mathbf{X}_{c_j}}, \quad \mathbf{X}_{c_j} = \begin{bmatrix} \cos(\omega_j t_1) \\ \cos(\omega_j t_2) \\ \vdots \\ \cos(\omega_j t_N) \end{bmatrix} \quad (18)$$

and  $\hat{\mathbf{D}}^{(1)} = \{d_i^{(1)}\}_{i=1}^N$  whose components are defined by

$$\hat{d}_i^{(1)} = d_i - \sum_{l=1}^m a_{c_l} \cos(\omega_l t_i) \delta_{lj} + a_{s_l} \sin(\omega_l t_i), \quad (19)$$

where  $\delta_{lj} = \begin{cases} 1 & l \neq j \\ 0 & l = j \end{cases}$  helps to eliminate the contribution of the cosine term of the  $j$ th sinusoid. However, in some cases where  $\sigma^2$  is unknown the joint posterior PDF of the parameters  $\{\omega, \mathbf{a}_c, \mathbf{a}_s, \sigma^2\}$  can be implemented in the form:

$$p(\omega, \mathbf{a}_c, \mathbf{a}_s, \sigma^2 | \mathbf{D}, I) \propto p(\mathbf{D} | \omega, \mathbf{a}_c, \mathbf{a}_s, \sigma^2, I) p(\sigma^2 | I) \quad (20)$$

In order to eliminate it, we assign Jeffreys' prior to  $p(\sigma^2 | I)$  and integrate Equation (20) with respect to  $\sigma^2$  so that we obtain a univariate Student's T- distribution:

$$p(a_{c_j} | \mathbf{a}_{c_j}, \mathbf{a}_s, \omega, \mathbf{D}, \sigma^2) \propto T(\hat{a}_{c_j}, s_{a_{c_j}}^2 (\mathbf{X}_{a_{c_j}}^T \mathbf{X}_{a_{c_j}})^{-1}, N - 1), \quad (21)$$

with

$$s_{a_{c_j}}^2 = \frac{1}{N - 1} (\hat{\mathbf{D}}^{(1)} - \hat{a}_{c_j} \mathbf{X}_{a_{c_j}})^T (\hat{\mathbf{D}}^{(1)} - \hat{a}_{c_j} \mathbf{X}_{a_{c_j}}). \quad (22)$$

In a similar way, the conditional PDF of  $a_{s_j}$  is given in the form of Equations (17) and (21) but,  $\hat{a}_{s_j}$  and  $\mathbf{X}_{a_{s_j}}$  are obtained by replacing  $a_{c_j}$  and sine terms in Equations (18) and (21) with  $a_{s_j}$  and cosine terms, respectively. In Equations (17) and (22),  $\hat{\mathbf{D}}^{(1)}$  term is calculated by eliminating the contribution of the sine term of the  $j$ th sinusoid in Equation (19) instead of cosine term.

In order to enable sampling for the frequency  $\omega$  we need Taylor series expansion of  $\chi^2(\omega)$  at  $\hat{\omega}$ :

$$\chi^2(\boldsymbol{\omega}) \approx \chi^2(\hat{\boldsymbol{\omega}}) + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 \chi^2(\boldsymbol{\omega})}{\partial \omega_i \partial \omega_j} \Big|_{\hat{\boldsymbol{\omega}}} (\omega_i - \hat{\omega}_i)(\omega_j - \hat{\omega}_j) + \dots, \quad (23)$$

where

$$\hat{\boldsymbol{\omega}} = \arg \min_{\boldsymbol{\omega} \in [0, \pi]^m} \chi^2(\boldsymbol{\omega}). \quad (24)$$

Then the conditional PDF of  $\omega_j$  turns out to be in the form:

$$p(\omega_j | \boldsymbol{\omega}_{-j}, \mathbf{a}_c, \mathbf{a}_s, \mathbf{D}, \sigma^2) \propto \exp \left( -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 \chi^2(\boldsymbol{\omega})}{\partial \omega_i \partial \omega_j} \Big|_{\hat{\boldsymbol{\omega}}} (\omega_i - \hat{\omega}_i)(\omega_j - \hat{\omega}_j) \right) \propto N(\hat{\omega}_j, \sigma^2 (\mathbf{X}_{\omega_j}^T \mathbf{X}_{\omega_j})^{-1}) \quad (25)$$

where

$$\mathbf{X}_{\omega_j} = \begin{bmatrix} -a_{c_j} t_1 \sin(\hat{\omega}_j t_1) + a_{s_j} t_1 \sin(\hat{\omega}_j t_1) \\ -a_{c_j} t_2 \sin(\hat{\omega}_j t_2) + a_{s_j} t_2 \sin(\hat{\omega}_j t_2) \\ \vdots \\ -a_{c_j} t_N \sin(\hat{\omega}_j t_N) + a_{s_j} t_N \sin(\hat{\omega}_j t_N) \end{bmatrix}. \quad (26)$$

If  $\sigma^2$  is unknown, Equation (25) becomes

$$p(\omega_j | \boldsymbol{\omega}_{-j}, \mathbf{a}_c, \mathbf{a}_s, \mathbf{D}, \sigma^2) \propto T(\hat{\omega}_j, s_{\omega_j}^2 (\mathbf{X}_{\omega_j}^T \mathbf{X}_{\omega_j})^{-1}, N-1). \quad (27)$$

with

$$s_{\omega_j}^2 = \frac{1}{N-1} (\mathbf{D} - \hat{\mathbf{D}})^T (\mathbf{D} - \hat{\mathbf{D}}), \quad (28)$$

where the components of the simulated data  $\hat{\mathbf{D}}$  is defined by the use of the current estimated values of the parameters. In numerical calculations, a systematic form of GIBBS algorithm proceeds in the following manner. Firstly, arbitrary starting values  $\{\mathbf{a}_{c,0}, \mathbf{a}_{s,0}, \boldsymbol{\omega}_0\}$  are chosen. At the each iteration of the GIBBS sampler, we cycle through the set of conditional distributions and draw one sample from each. When a sample is drawn from one conditional distribution, the succeeding distributions are updated with the new value of that sample. Then successively random drawings from the full conditional distributions described above are as follow:

$$\begin{aligned} a_{c_j,1} &\sim p(a_{c_j} | \{a_{c_{1,1}}, \dots, a_{c_{j-1,1}}, a_{c_{j+1,0}}, \dots, a_{c_{m,0}}\}, \mathbf{a}_{s_j,0}, \boldsymbol{\omega}_0, \mathbf{D}) \\ a_{s_j,1} &\sim p(a_{s_j} | \mathbf{a}_{c_j,1}, \{a_{s_{1,1}}, \dots, a_{s_{j-1,1}}, a_{s_{j+1,0}}, \dots, a_{s_{m,0}}\}, \boldsymbol{\omega}_0, \mathbf{D}) \\ \omega_{j,1} &\sim p(\omega_j | \mathbf{a}_{c_j,1}, \mathbf{a}_{s_j,1}, \{\omega_{1,1}, \dots, \omega_{j-1,1}, \omega_{j+1,0}, \dots, \omega_{m,0}\}, \mathbf{D}) \end{aligned} \quad (29)$$

After the first iteration, we get  $\{\mathbf{a}_{c,1}, \mathbf{a}_{s,1}, \boldsymbol{\omega}_1\}$ ; secondly  $\{\mathbf{a}_{c,2}, \mathbf{a}_{s,2}, \boldsymbol{\omega}_2\}$  and so on. Repeating this procedure  $K$  times, we obtain  $\{\mathbf{a}_{c,K}, \mathbf{a}_{s,K}, \boldsymbol{\omega}_K\}$ . In Bayesian context, for a large enough  $K$  the joint PDF can be replaced by the conditional PDF so that the parameters  $a_{c_j,K}$ ,  $a_{s_j,K}$  and  $\omega_{j,K}$  become random variables. Then we draw  $M$  random samples

of  $\{\mathbf{a}_{c,K}^l\}_{l=1}^M$ ,  $\{\mathbf{a}_{s,K}^l\}_{l=1}^M$  and  $\{\boldsymbol{\omega}_K^l\}_{l=1}^M$  from their marginal PDFs, respectively and using these samples, we obtain all of the estimates about the corresponding parameter, such as its most probable value, its mean and its marginal variances with respect to the most probable value etc. When  $\sigma^2$  is unknown, we do the same thing as above except that the random numbers are generated from the Student's T-distributions given in Equations (21) and (26).

### C. Parallel Tempering

The algorithm of GIBBS sampling overcomes some problems associated with BRETTHORST but, it faces with two serious drawbacks. First of all, it is only an approximate Bayesian inference scheme since Laplace approximation in Equation (25) is used in order to enable sampling for  $\boldsymbol{\omega}$ . Secondly, the optimum point  $\hat{\boldsymbol{\omega}}$  of  $\chi^2(\boldsymbol{\omega})$  given in Equation (24), is highly intractable since it involves minimization of a sharply peaked multimodal cost function which cannot be computed in closed form. To overcome these problems and provide a flexibility of choice of priors, we implement parallel tempering (PT) method [9, 10] that is originated with Swendsen [20], extended by Geyer [21] and later developed and successfully used in a number of general optimization problems[40].

In Bayesian analysis we need to specify a suitable prior PDF that should represent the best knowledge of the parameters. If  $\omega$  is considered as a location parameter and we know that there value is upper,  $\omega_{max}$  and lower bound  $\omega_{min}$ . If that is all we know about this parameter then the principle of maximum entropy [3] will lead us to assign uniform prior PDF:

$$p(\omega|I) = \begin{cases} \frac{1}{\omega_{max} - \omega_{min}} & \omega_{min} \leq \omega \leq \omega_{max} \\ 0 & \text{otherwise} \end{cases}. \quad (30)$$

Otherwise, it is a positive quantity so that it can be considered as a scale parameter. Therefore, one may then choose Jeffreys prior [31] as mentioned before but, this is not strictly speaking a probability at all because it cannot be normalized. To make this a proper, probability one must introduce an upper and lower bound for  $\omega$  and compute the normalization constant so that we get its modified version:

$$p(\omega|c) = \begin{cases} \frac{1}{(\omega+c) \ln \left( \frac{\omega_{max}+c}{\omega_{min}+c} \right)} & \omega_{min} \leq \omega \leq \omega_{max} \\ 0 & \text{otherwise} \end{cases}, \quad (31)$$

where  $c$  removes singularity at  $\omega = \omega_{min}$  and it is assigned to the mean of the standard deviation of the noise vector. In Bayesian calculations, the prior PDF provides an order of magnitude estimate for the parameters. These estimates are ascertained from known factors, such as the sampling time and the magnitude of data. On the other hand, we assign the

prior PDFs for the angular frequencies and amplitudes using broad uninformative Gaussians. This is because a bounded Gaussian correctly describes an order of magnitude of the estimate. However, these priors can be assigned by specifying a lower and upper parameter values so that an uninformative Gaussian prior PDF for  $\omega$  is defined in the form:

$$p(\omega|\mu_\omega, \sigma_\omega) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma_\omega^2}} \exp\left(-\frac{(\omega-\mu_\omega)^2}{2\sigma_\omega^2}\right) & \omega_{\min} \leq \omega \leq \omega_{\max} \\ 0 & \text{otherwise} \end{cases} \quad (32)$$

where

$$\mu_\omega = \frac{\omega_{\max} + \omega_{\min}}{2}, \quad \sigma_\omega = \frac{\omega_{\max} - \omega_{\min}}{3} \quad (33)$$

Finally, a suitable prior PDF for  $\omega_j$  can be taken as a combination of these three prior PDFs described above:

$$p(\omega_j|I) = \begin{cases} \frac{1}{3(\pi)} + \frac{1}{3(\omega_j+c)\ln(\frac{\omega_{\max}}{\omega_{\min}+c})} + \frac{1}{3\sqrt{2\pi\sigma_\omega^2}} \exp\left(-\frac{(\omega_j-\mu_\omega)^2}{2\sigma_\omega^2}\right) & \omega_{\min} \leq \omega_j \leq \omega_{\max} \\ 0 & \text{otherwise} \end{cases} \quad (34)$$

By using similar arguments, the prior PDF for the amplitude  $a_{c_j}$  can also be taken as

$$p(a_{c_j}|I) = \begin{cases} \frac{1}{2(D_{\max}-D_{\min})} + \frac{1}{2\sqrt{2\pi\sigma_{a_s}^2}} \exp\left(-\frac{(a_{c_j}-\mu_{a_s})^2}{2\sigma_{a_s}^2}\right) & D_{\min} \leq a_{c_j} \leq D_{\max} \\ 0 & \text{otherwise} \end{cases} \quad (35)$$

where

$$\mu_{a_s} = \frac{D_{\max} + D_{\min}}{2}, \quad \sigma_{a_s} = \frac{D_{\max} - D_{\min}}{3} \quad (36)$$

A similar prior PDF is also assigned to the coefficient  $a_{s_j}$  by replacing  $a_{c_j}$  with  $a_{s_j}$  in Equation (35). Putting Equations (32) and (34) into Equation (10), we obtain the posterior PDF  $p(\boldsymbol{\omega}, \mathbf{a} | \mathbf{D}, I)$ , denoted here as a tempered PDF  $\pi(\boldsymbol{\omega}, \mathbf{a} | \mathbf{D}, \beta, I)$ :

$$\pi(\boldsymbol{\omega}, \mathbf{a} | \mathbf{D}, \beta, I) = p(\boldsymbol{\omega}, \mathbf{a} | I) \exp((\beta \ln(p(\mathbf{D} | \boldsymbol{\omega}, \mathbf{a}, I))) \quad (37)$$

Now, the problem turns out finding the parameter values that maximize Equation (37) using the PT algorithm. Starting from a given initial sample  $X_t = \{\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s\}$  over the state space, it basically uses a stochastic transition function to produce a new sample using proposal PDF  $q(X_{t+1} | X_t)$ , which is considered here to be a multivariate Normal distribution with a mean equaled to current sample  $X_t$  and a deviation  $\boldsymbol{\sigma}_X$  named as a step size which is taken to be a square root of CRLB for estimated parameters [21]. It consists of two main updating steps. The first one is the state update of each chain in which there exists  $n\beta$  multiple copies of Markov chain Monte Carlo (MCMC) simulations[20], which are run simultaneously in parallel

each at different values of tempering parameter  $\beta$ . Actually, MCMC algorithm generates desired samples  $X_t$  by constructing a kind of a random walk in a model parameter space so that it is accepted as a new sample called  $X_{t+1}$  by satisfying

$$u_1 \leq \alpha(X_t, X_{t+1}), \quad (38)$$

where  $u_1$  is a random variable drawn from uniform distribution  $U_{(0,1)}$  and an acceptance probability  $\alpha(X_t, X_{t+1})$  is defined by

$$\alpha(X_t, X_{t+1}) = \min\left\{1, \frac{p(X_{t+1} | \mathbf{D}, I)}{p(X_t | \mathbf{D}, I)}\right\} \quad (39)$$

After a number of iterations on each replica of the MCMC simulations, the current samples are considered probabilistically for exchanges between different tempering levels. This is the second update step called the swapping between two neighboring chains at each  $n_s$  step if a random

number  $u_2$  drawn from  $U_{(0,1)}$  satisfies  $u_2 \leq \frac{1}{n_s}$ . Then at

time  $t$  the simulation  $\beta_i$  in the state  $X_{t,i}$  and the simulation  $\beta_{i+1}$  in the state  $X_{t,i+1}$  can be interchanged if a random number  $u_3$  drawn from  $U_{(0,1)}$  satisfies

$$u_3 \leq \alpha(X_{t,i}, X_{t,i+1}), \quad (1 \leq i \leq n\beta - 1), \quad (40)$$

where

$$\alpha(X_{t,i}, X_{t,i+1}) = \min\left\{1, \frac{\pi(X_{t,i+1} | \mathbf{D}, \beta_i, I) \cdot \pi(X_{t,i} | \mathbf{D}, \beta_{i+1}, I)}{\pi(X_{t,i} | \mathbf{D}, \beta_i, I) \cdot \pi(X_{t,i+1} | \mathbf{D}, \beta_{i+1}, I)}\right\} \quad (41)$$

This is called a probability of the swap acceptance. As expected, after an initial burn-in period this proposed method generates samples  $X_t$  with a PDF equal to the desired posterior PDF  $p(X_t | \mathbf{D}, I)$ . Finally, inferences about parameters are based on these samples drawn from the output corresponding to the lowest temperature chain ( $\beta = 1$ ). For detail information about the algorithm, we refer to our recent papers in [30, 31, 38].

#### IV. COMPUTER SIMULATIONS

In this section, we demonstrate the performance of our algorithms by some simulation results. We first generated a simulated data vector according to a signal model with two closed harmonic frequencies:

$$d_i = 0.5403 \cos(0.3t_i) - 0.8415 \sin(0.3t_i) - 0.4161 \cos(0.31t_i) - 0.9093 \sin(0.31t_i) + e_i \quad (42)$$

Here  $t_i$  runs over the symmetric time interval  $-T$  to  $T$  in  $(2T + 1) = N = 512$  integer steps and the components of  $e_i$  are generated from the zero mean Gaussian distribution with a deviation  $\sigma = 1$ . Then they were added to the simulated

data samples and obtained the noisy data  $\{d_i\}_{i=1}^N$ , shown in Figure 1(a). Thus, we carried out the Bayesian analysis of the noisy data, assuming that we know the mathematical form of the signal model but, not the values of its parameters.

All three Bayesian methods were coded in Mathematica programming language because it provides much flexible and efficient computer programming environment. Furthermore, it also contains a large collection of built-in functions and results much shorter computer codes than those written in C or FORTRAN programming languages. They were run on a workstation with four processors which of each has got Intel Core 2 Quad Central Processing Unit (CPU). As an initial estimate of the frequencies  $\omega_0$  for each stochastic maximization procedures, it is possible to take random choices from the interval  $(0, \pi)$ . However, it is better to start with the locations of the peaks with the greatest magnitudes as an initial estimate of  $\omega$  automatically from the Fourier power spectral density (FPSD) graph by using a computer code written in Mathematica. Then we carried on calculating the coefficients  $\mathbf{a}_c$  and  $\mathbf{a}_s$  as initial values for the amplitudes, respectively.

In the case of  $\sigma=1$ , the output of the computer simulations of all methods is illustrated in Table 1. It can be seen that the estimated two frequencies and their corresponding amplitudes from the noisy signal are quoted as  $(value) \pm (standard\ deviation)$ . If one uses each of those estimated values of parameters in the signal model to restore it, we almost obtained similar results shown in Figure 1(b).

In general, we secondly consider a multiple harmonic frequency signal model:

$$d_i = 0.540302 \cos(0.1 t_i) - 0.841471 \sin(0.1 t_i) - 0.832294 \cos(0.15 t_i) - 1.81859 \sin(0.15 t_i) - 4.94996 \cos(0.3 t_i) - 0.70560 \sin(0.3 t_i) - 1.30729 \cos(0.31 t_i) + 1.51360 \sin(0.31 t_i) + 0.850087 \cos(t_i) + 2.87677 \sin(t_i) + e_i, \quad (43)$$

$(i = 1, 2, \dots, N)$

In a similar way, we obtained the noisy data shown in Figure 3(a) and ran Mathematica codes of the proposed algorithms again. We found the estimated values of parameters using each method and tabulated in Table 2. It is observed that each of them provides estimated parameter values with almost similar estimation accuracies, especially for the angular frequencies.

The usual way the result from a spectral analysis is displayed is in the form of a power spectral density (PSD) [3, 6] that shows the strength of the variation (energy) as a function of frequency. In Fourier transform spectroscopy this is typically taken as the squared magnitude of the discrete Fourier transform of the data. In order to display our results in the form of a power spectral density, it is necessary to give an attention to its definition that shows how much power is contained in a unit frequency.

According to Bretthorst, the Bayesian power spectral density (BPSD) is defined as the expected value of the power of the signals over the joint posterior PDF:

$$BPSD(\omega) = \frac{N}{2} \int \sum_{j=1}^m (\mathbf{a}_{c_j}^2 + \mathbf{a}_{s_j}^2) p(\omega, \mathbf{a}_c, \mathbf{a}_s | D, \sigma, I) d\mathbf{a}_c d\mathbf{a}_s \quad (44)$$

Performing integrals analytically over  $\mathbf{a}_c$  and  $\mathbf{a}_s$  by using orthogonal model functions  $H$  defined in Equation (14), the PSD can therefore be approximated as

$$BPSD(\omega) = 2 \left( \sigma^2 + \sum_{j=1}^{2m} h_j^2 \right) \sum_{k=1}^m \sqrt{\frac{b_{kk}}{2\pi\sigma^2}} \exp\left(\frac{-b_{kk}(\hat{\omega}_k - \omega)}{2\sigma^2}\right), \quad (45)$$

where

$$b_{jk} = -m \frac{\partial^2 h^2}{\partial \omega_j \partial \omega_k} \Big|_{\substack{\omega_j = \hat{\omega}_j \\ \omega_k = \hat{\omega}_k}}. \quad (46)$$

This function stresses information about the total energy carried by the signal and about the accuracy of each line. Fourier and Bayesian PSDs for two signal models are shown in Figures 2 and 3. Fourier PSDs in Figures 2 (a) and 3 (b) indicate only one of two and four of five well separated frequencies, respectively but, Bayesian PSDs in Figure 2 (b)-(d) and 3(c)-(e) show that all two and the five frequencies are well separated, respectively while their heights are indicative of the resolution. A comparison of them implies that frequencies obtained by using the proposed methods are separated very well although the separation of the sinusoids is less than the Nyquist step [3,

6, 24], defined as  $|\omega_1 - \omega_2| < \frac{\pi}{N}$ . These results demonstrate

ability of resolving closely spaced frequencies by use of the methods based on Bayesian inference.

Moreover, we initially assumed that the values of the random noise in data were drawn from the Gaussian distribution. Figure 4 shows the exact and estimated PDF of the noise using each of the proposed methods. It is seen that the estimated (dotted) PDF is closer to its true (solid) PDF and the histogram of the errors, which is known as nonparametric estimator of the PDF of the noise is also much closer to its true PDF.

Performance of estimators is often compared against the CRLB, which is an inversion of the Fisher Information Matrix  $J(\omega)$  whose elements are the expectation of the second derivative of the Log likelihood function with respect to the parameter vector  $\omega$  :

$$J(\omega) = E \left[ -\frac{\partial^2 \ln p(D|\omega, I)}{\partial \omega^2} \right] \quad (47)$$

Assuming that the matrix  $J(\omega)$  is diagonal for a large  $N$  so that its inversion is straightforward. In this case, the diagonal elements yield the lower bound for the variance of the estimates  $\hat{\omega}$  asymptotically and we can write

$$\sigma_{CRLB}^2(\omega_j) \geq \frac{24\hat{\sigma}^2}{N^3(\hat{a}_{c_j}^2 + \hat{a}_{s_j}^2)}, \quad (j = 1, 2, \dots, m) \quad (48)$$

where  $\hat{\sigma}^2$  represents the estimated variance of the noise. A set of computer simulation was therefore conducted to evaluate the sinusoidal frequency estimation performances of the proposed algorithms. We first generated 64 data samples with different noise levels from a single real tone frequency signal model and calculated the mean square error (MSE) of the estimated frequencies after 50 independent trials under the same SNR, defined as

$$SNR = 10 \log \left( \frac{a_c^2 + a_s^2}{\sigma^2} \right), \quad (49)$$

which varies from 0 to 20 decibel (dB) so that we plotted the logarithmic values of MSE as a function of the SNR and showed it in Figure 5(a). It can be seen that BRETTHORST, GIBBS and PT estimators have thresholds about 3 dB, 4 dB and 5 dB of the SNRs, respectively and the curves for the frequency estimates obtained by using these methods also follow nicely the CRLB all the way down approximately to 5 dB. As getting lower SNRs all estimates starts to deteriorate significantly. To see changes in performances we calculated an efficiency parameter  $\eta$  [34], defined as

$$\eta = \left( \frac{CRLB}{MSE} \right) \times 100. \quad (50)$$

This indicates closeness to the CRLB. Table 3 contains the MSEs and the values of efficiency parameter  $\eta$  for the frequency estimation and indicates that BRETTHORST is the most efficient among the others so that BRETTHORST is called better estimator at SNR = 20dB and  $N = 64$ . The above argument treats with only the case in which a data set of  $N = 64$  is used for estimation. However, one may ask how to estimation accuracy varies with the length of data sampling  $N$ . To answer it, we secondly generated data sets with different lengths  $N$  from a single real tone frequency signal model under SNR = 10 dB. After independent 20 trials we calculated the values of MSEs of the estimated frequency obtained by each of methods so that we plotted their logarithmic values as a function of the length of data sampling  $N$ , which varies from 64 to 676, shown in Figure 5(b). From these results, larger  $N$  makes higher accuracy but, requires larger consumption of CPU time. In addition, Table 3 also indicates that PT is slightly more efficient method because it give a value of efficiency parameter  $\eta$  closed to CRLB for the frequency estimation at  $N = 676$  under SNR=10 dB. As expected, lower SNR and less data sampling deteriorate the performances of the methods. As a result, computer simulations show that the performances of the proposed methods are close to optimal with a minimum variance, which is close to the predictions made by the CRLB.

The computational complexity of the methods depends on the length of data samples  $N$ , number of parameters  $m$  and some control parameters that need to be tuned before performance becomes optimal and vary with the methods such as annealing schedule in BRETTHOST, tempering levels  $n\beta$  and steps for exchanging  $n_s$ , maximum number

of iterations in PT, number of Gibbs simulations  $K$ , and number of sampling  $M$  in GIBBS. Discussions about the control parameters are given in papers [31, 32] where the proposed methods were introduced separately. By fixing those parameters, Figure 6 shows only CPU time of different simulations taken by BRETTHOST in a variety of number of data samples  $N$  and parameters  $m$ . GIBBS and PT give also similar results given in papers [31] and [32], respectively. Consequently, it indicates that as the length of the data samples or the number of parameters increases all the methods requires larger consumption of CPU time.

## CONCLUSIONS

In this paper, three Bayesian approaches with different sampling procedures to estimate parameters of sinusoids embedded in Gaussian noise are partly improved and outputs of their computer simulations are compared and discussed.

Overall results show that Bayesian approach can not only give us the best estimates for the parameters but, also tell us uncertainties associated with their estimated values. Although they are more computationally intensive than the usual power spectrum methods, they are the best suited to those datasets where the values are noisy and aliasing that causes difficulties in the interpretation of the power spectrum. Computer simulation results allow us to show outstanding performances for separating two closely spaced sinusoids at very low SNRs and also demonstrate how the state-of-the art Bayesian inference schemes for the frequency in the static sinusoidal model work. Moreover, their performances for single frequency estimation are demonstrated via computer simulations at different data lengths  $N$  and various SNRs and indicate that errors variance obtained by each of them tracks that of CRLB closely at low SNR values, showing no signs of thresh holding down to SNR=5 dB.

In analyzing experimental data, one has enough prior information in a given experiment to select the best model among a finite set of model functions so that Bayesian inference helps us to accomplish it. This is, in general, called model selection but, it is known here as a detection of number of sinusoids that is a big part of spectral analysis. A joint detection and estimation problem for complex signal models is our current interest and therefore will deserve further investigations. In addition, Mathematica has grown in breadth and depth to become today an unparalleled platform for all forms of computations.

## ACKNOWLEDGEMENTS

This work is part of the projects, whose names are "Positron lifetimes spectrum analysis with Bayesian statistical

inference" with a number FEN - A -060510-0135 and "Comparison of Bayesian methods for recovering sinusoids" with a number FEN-D-150513-0187 supported by Marmara University, Istanbul, Turkey.

## REFERENCES

1. Kay S.M.: "Fundamentals of Statistical Signal Processing: Estimation Theory", Prentice Hall, 1993.
2. E.T. Jaynes: "Bayesian Spectrum and Chirp Analysis", In *Proceedings of the Third Workshop on Maximum Entropy and Bayesian Methods* Ed. C. Ray Smith and D. Reidel, Boston 1987, 1-37.
3. E.T. Jaynes: "Probability Theory: The Logic of Science", Cambridge University Press, Cambridge, UK, 2003.
4. D.S. Sivia, J. Skilling: "Data Analysis: A Bayesian Tutorial", Oxford University Press Inc., New York, 2006.
5. G. D' Agostini: "Bayesian inference in processing experimental data: principles and basic applications", Report on Processing Physics 2003; **66**; 1383-1419.
6. G.L. Bretthorst: "Bayesian Spectrum Analysis and Parameter Estimation", Lecture Notes in Statistics, Springer-Verlag Berlin Heidelberg New York, 1997.
7. L. Dou and R. J. W. Hodgson: "Bayesian inference and Gibbs sampling in spectral analysis and parameter estimation I", *Inverse Problem* 1995; **11**; 1069-1085.
8. L. Dou and R. J. W. Hodgson: "Bayesian inference and Gibbs sampling in spectral analysis and parameter estimation II", *Inverse Problem* 1995; **11**; 121-137.
9. P.C. Gregory: "Bayesian Logical Data Analysis for the Physical Science", Cambridge University Press, United Kingdom, 2005.
10. P.C. Gregory: "A Bayesian Kepler Periodogram Detects a Second Planet in HD 208487", *Mon. Not. R. Astron. Soc.* 2006; 1-14.
11. S.M. Kay: "Accurate frequency estimation at low signal-to-noise ratio", *IEEE Transactions on Acoustic, Speech and Signal Processing* 1984; **ASSP-32**; 540-547.
12. H.L. Harney: "Bayesian Inference: Parameter Estimation and Decisions", Springer-Verlag, Berlin Heidelberg, 2003.
13. M. Bernardo, A.F.M. Smith: "Bayesian Theory", Willey series in Probability and Statistics, New York, 2000.
14. H. Jeffreys: "Theory of Probability", Oxford University press, 1988.
15. D. W. Marquardt: "An algorithm for least-squares estimation of nonlinear parameters"; 1963; **11**; 431-441.
16. W.H Press, B.P Flannery, S.A Teukolshy and W.T. Vetterling: "Numerical Recipes in C: The Art of Computing", Second Edition, Cambridge University Press, 1995.
17. T.R. Hooke and T.A. Jevees: "Direct search solution of numerical and statistical problems", *Journal of Association of Computer Machinery* 1962; **5**: 212-229.
18. W.L. Goffe, G.D. Ferier and J. Rogers: "Global optimization of statistical functions with simulated annealing", *Journal of Econometrics*; 1994; **60**; 65-100.
19. A. Corana, M. Marchesi, C. Martini, S. Ridella: "Minimizing multimodal functions of continuous variables with the simulated annealing algorithm", *ACM Transactions on Mathematical Software* 1987; **13**: 262-280.
20. R. H. Swendsen and J.S. Wang: "Replica Monte Carlo simulation of spin-glasses", *Physical Review of Letters*; 1986; **57**; 2607.
21. C. J. Geyer: "In computing science and statistics, *Proceedings of the 23rd Symposium on the Interface*", New York, 1991.
22. N. Metropolis, A. Rosenbluth, M. Rosenblatt, A. Teller and E. Teller: "Equation of states calculations by fast computing machines", *Journal of Chemical Physics* 1953; **21**; 1087-1092.
23. J. Ireland: "Simulated annealing and Bayesian posterior distribution analysis applied to spectral emission line fitting", *Journal of Solar Physics* 2007; **243**: 237-252.
24. A. Schuster: "The Periodogram and its Optical Analogy", *Proceedings of the Royal Society of London* 1905; **77**; 136.
25. J.W. Cooley, J.W. Tukey: "An algorithm for the machine calculation of complex Fourier series", *Mathematics of Computation* 1965; **19**; 297-301.
26. D. Üstündağ and M. Cevri: "Estimating parameters of sinusoids from noisy data using Bayesian inference with simulated annealing", *Wseas Transactions on Signal Processing* 2008; **7**; 432-441.
27. D. Üstündağ and M. Cevri: "Bayesian Parameter Estimation of Sinusoids with Simulated Annealing", *8Th Wseas International Conference On Signal Processing, Computational Geometry And Artificial Vision (Iscgav'08)*, Rhodes, Greece 2008; 106-112.
28. D. Üstündağ and M. Cevri: "Recovering sinusoids from noisy data using Bayesian inference with simulated annealing" *Mathematical and Computational Applications* 2011; **16**; 382-391.
29. L.Tierney: "Markov chains for exploring posterior distributions", *The Annals of Statistics* 1994; **22**; 1701-1728.
30. M. Cevri and D. Üstündağ: "Bayesian Estimation of Sinusoidal Signals via Parallel Tempering", *Proceedings of the 3rd WSEAS international symposium on Wavelets theory and applications in applied mathematics, signal processing & modern science, Istanbul, Turkey, 2009*, 67-72.
31. M. Cevri and D. Üstündağ: "Bayesian recovery of sinusoids from noisy data with parallel tempering", *IET Signal Processing* 2012; **6**; 673-683, 2012.
32. D. Üstündağ and Mehmet Cevri: "Bayesian Recovery of Sinusoids with Simulated Annealing", *Advances, Applications and Hybridizations*, Ed. Dr. Marcos Sales Guerra Tsuzuki, InTech, DOI: 10.5772/50449, 2012. Available from: <http://www.intechopen.com/books/simulated-annealing-advances-applications-and-hybridizations/bayesian-recovery-of-sinusoids-with-simulated-annealing>.
33. D. Üstündağ: "Recovering Sinusoids from Data Using Bayesian inference with RJMCMC", *Seventh International Conference on Natural Computation. Shanghai, China, 2011*; 1850-1854.
34. M. Cevri and D. Üstündağ: "Performance analysis of Gibbs sampling for Bayesian extracting sinusoids", *Proceedings of the*

2013 International Conference on Systems, Control, Signal Processing and Informatics (SCSI 2013), Rhodes Island, Greece, 2013; 128-134.

35. B. Ristic, S. Arulampalam and N. Gordon: "Beyond the Kalman Filter: Particle Filters for Tracking Applications", Artech House, London, 2004.
36. V. Trees and L. Harry: "Detection, Estimation, and Modulation Theory", John Wiley & Sons, New York, 1971.
37. H.T. Li, P.M. Djuric: "An iterative MMSE procedure for parameter estimation of damped sinusoidal signals", Signal Processing 1996; **51**: 105-120.
38. N. Rathore, M. Chopra and J.J. de Pablo: "Optimal allocation of replicas in parallel tempering simulations", The journal of Chemical Physics 2005; **122**: 024111-1-7.
39. J.K. Nielsen: "Sinusoidal parameter estimation-A Bayesian approach," Master Thesis, Aalborg University, 2009.
40. Y. Li, M. Mascagni, A. Gorin: "A decentralized parallel implementation for parallel tempering algorithm", Parallel Computing 2009;**35**; 269-283.

**Table 2.** The best estimates of parameters for a multiple frequency sinusoidal signal model

$\theta$	TRUE VALUES	BRETTTHO RSTS	GIBBS	PT
$\omega_1$	0.1000	0.1001 $\pm 0.0005$	0.0989 $\pm 0.0004$	0.1005 $\pm 0.0004$
$\omega_2$	0.1500	0.1502 $\pm 0.0003$	0.1499 $\pm 0.0002$	0.1502 $\pm 0.0002$
$\omega_3$	0.3000	0.2997 $\pm 0.0002$	0.3001 $\pm 0.0001$	0.2998 $\pm 0.0001$
$\omega_4$	0.3100	0.3097 $\pm 0.0004$	0.3102 $\pm 0.0002$	0.3094 $\pm 0.0003$
$\omega_5$	1.000	0.9999 $\pm 0.0002$	1.0000 $\pm 0.0001$	1.0001 $\pm 0.0001$
$a_{c_1}$	0.5403	0.5669 $\pm 0.0645$	0.5740 $\pm 0.0657$	0.5645 $\pm 0.0573$
$a_{s_1}$	-0.8414	-0.8922 $\pm 0.0646$	-0.7589 $\pm 0.0662$	-0.7664 $\pm 0.0586$
$a_{c_2}$	-0.8322	-0.8873 $\pm 0.0662$	-0.8922 $\pm 0.0654$	-0.7985 $\pm 0.0604$
$a_{s_2}$	-1.8185	-1.8215 $\pm 0.0661$	-1.8140 $\pm 0.0662$	-1.8304 $\pm 0.066$
$a_{c_3}$	-4.9499	-4.7931 $\pm 0.0649$	-5.0060 $\pm 0.0658$	-4.9055 $\pm 0.0934$
$a_{c_3}$	-0.7056	-0.7841 $\pm 0.0655$	-0.7615 $\pm 0.0655$	-0.7161 $\pm 0.0915$
$a_{c_4}$	-1.3072	-1.3671 $\pm 0.0653$	-1.1790 $\pm 0.0651$	-1.3965 $\pm 0.1178$
$a_{s_4}$	1.5136	1.4932 $\pm 0.0628$	1.4840 $\pm 0.0656$	1.4921 $\pm 0.0681$
$a_{c_5}$	0.8500	0.8879 $\pm 0.0667$	0.9373 $\pm 0.0655$	0.8414 $\pm 0.0651$
$a_{s_5}$	2.8767	2.9218 $\pm 0.0649$	2.9260 $\pm 0.0657$	2.8529 $\pm 0.0612$

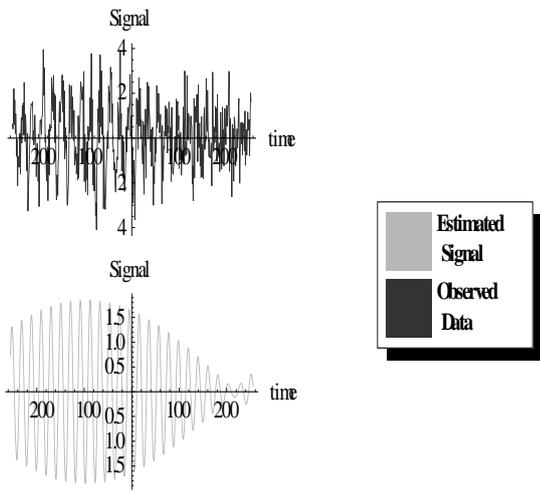
TABLES AND FIGURES

**Table 1.** The best estimates of parameters for two frequencies sinusoidal signal model

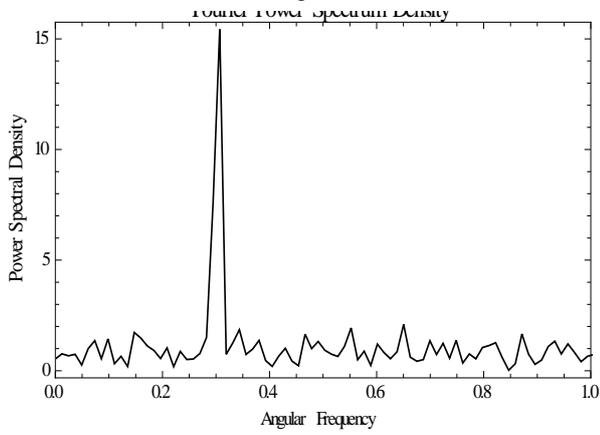
$\theta$	True Values	BRETTTHORSTS	GIBBS	PT
$\omega_1$	0.3	0.3001 $\pm 0.0001$	0.3001 $\pm 0.0004$	0.3002 $\pm 0.0003$
$\omega_2$	0.31	0.3102 $\pm 0.0001$	0.3108 $\pm 0.0004$	0.3103 $\pm 0.0005$
$a_{c_1}$	0.5403	0.4721 $\pm 0.06$	0.4821 $\pm 0.06$	0.4286 $\pm 0.05$
$a_{s_1}$	-0.8415	0.8653 $\pm 0.06$	-0.8300 $\pm 0.06$	1.0960 $\pm 0.04$
$a_{c_2}$	-0.4161	-0.3952 $\pm 0.06$	-0.3852 $\pm 0.06$	-0.3111 $\pm 0.04$
$a_{s_2}$	-0.9093	-0.7816 $\pm 0.06$	-0.9005 $\pm 0.06$	-0.9168 $\pm 0.04$

**Table 3.** Performance comparison of Bayesian methods for single frequency estimation of noisy sinusoid

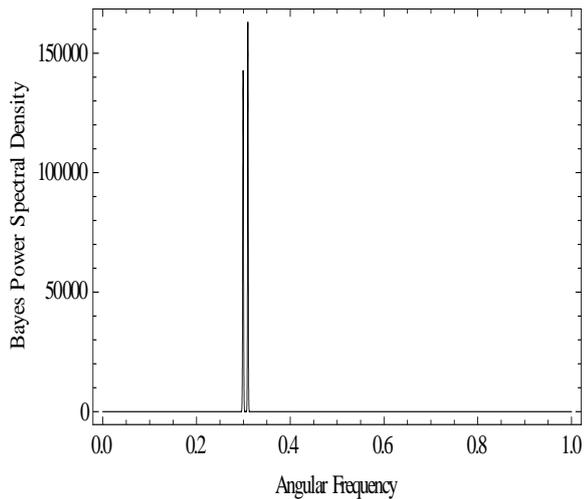
Methods	SNR = 20dB		N = 672	
	MSE(dB)	$\eta$	MSE(dB)	$\eta$
<b>GIBBS</b>	-72.0742	110	-81.371	122
<b>BRETTTHORST</b>	-74.7158	106	-83.007	120
<b>PT</b>	-68.266	116	-85.010	117
<b>CRLB</b>	-79.3572	100	-99.357	100



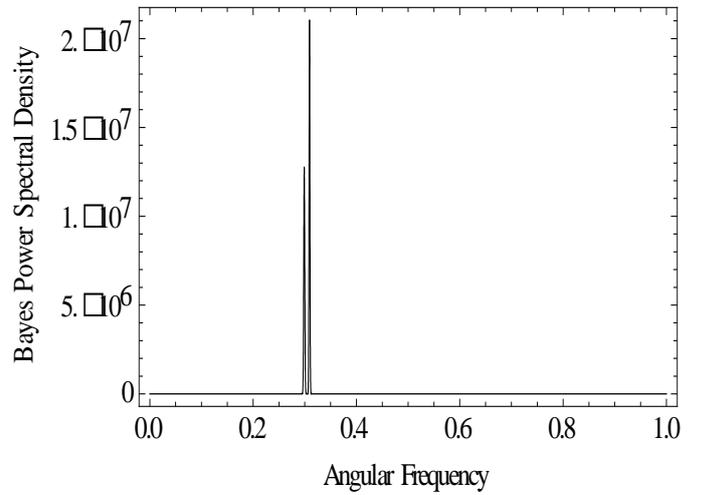
**Figure 1.** Signal models function with two frequencies corrupted by random noise and its restoration. (a) Observed data (b) Estimated signal.



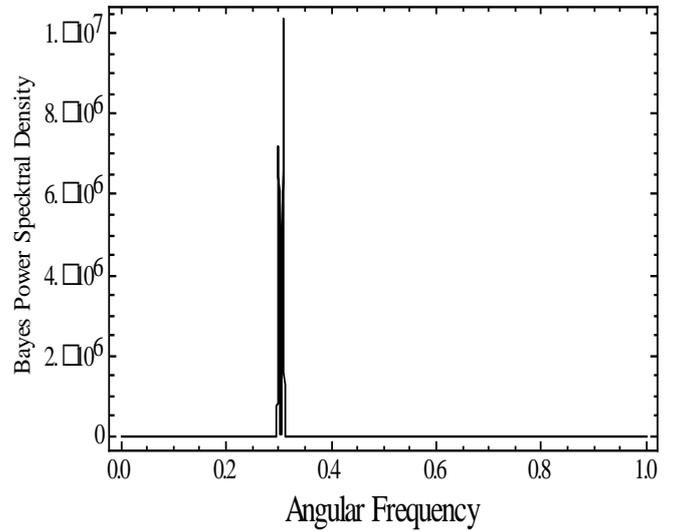
(b)



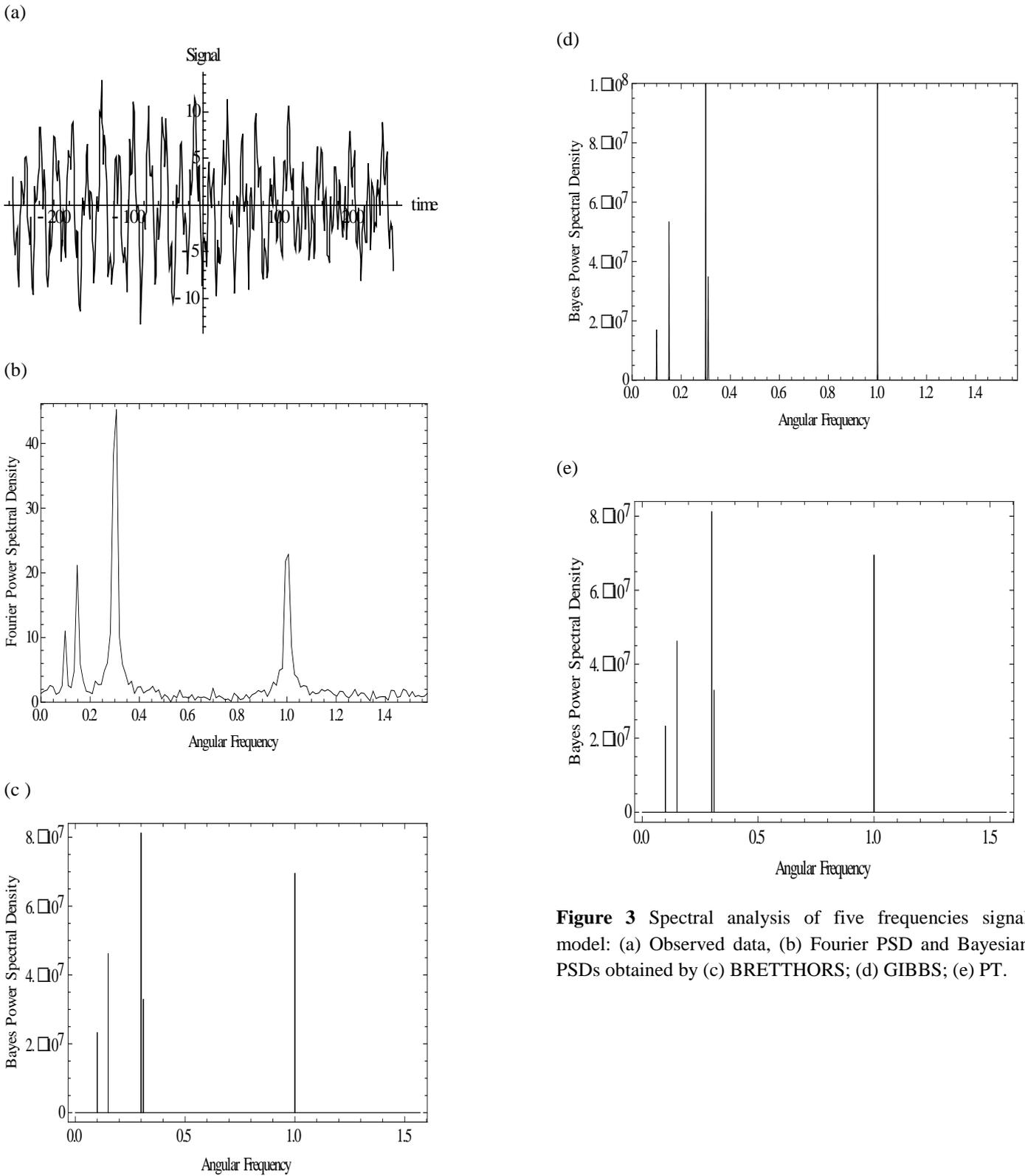
(c)



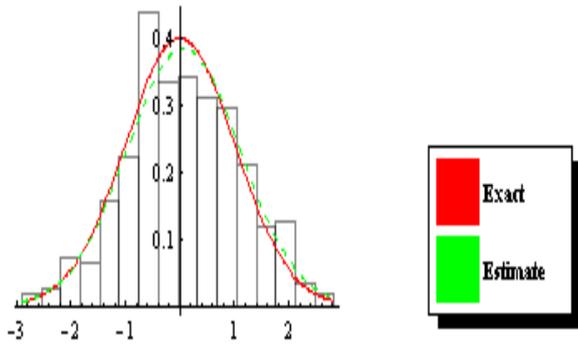
(d)



**Figure 2** Spectral analysis of two frequencies signal model: (a) Fourier PSD and Bayesian PSDs obtained (b) BRETTHORST; (c) GIBBS; (d) PT.

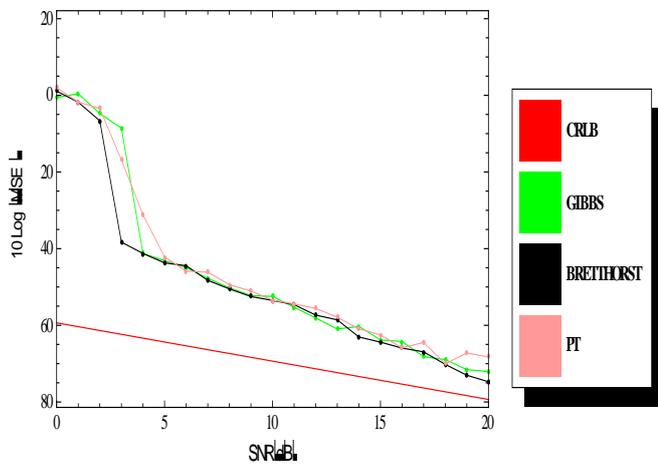


**Figure 3** Spectral analysis of five frequencies signal model: (a) Observed data, (b) Fourier PSD and Bayesian PSDs obtained by (c) BRETTHORS; (d) GIBBS; (e) PT.

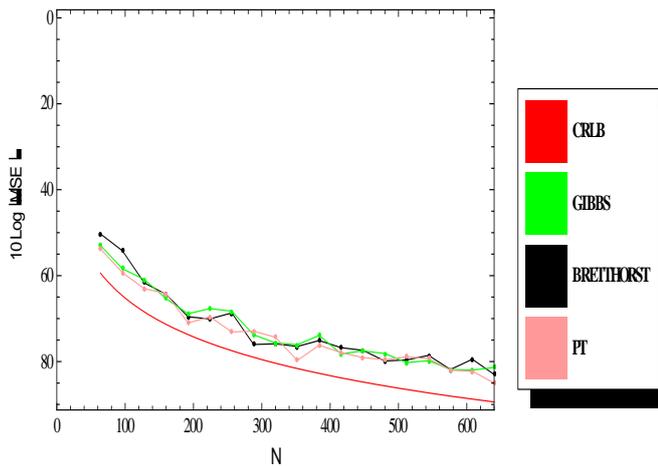


**Figure 4.** Comparison of exact and estimate PDFs of the noise in data

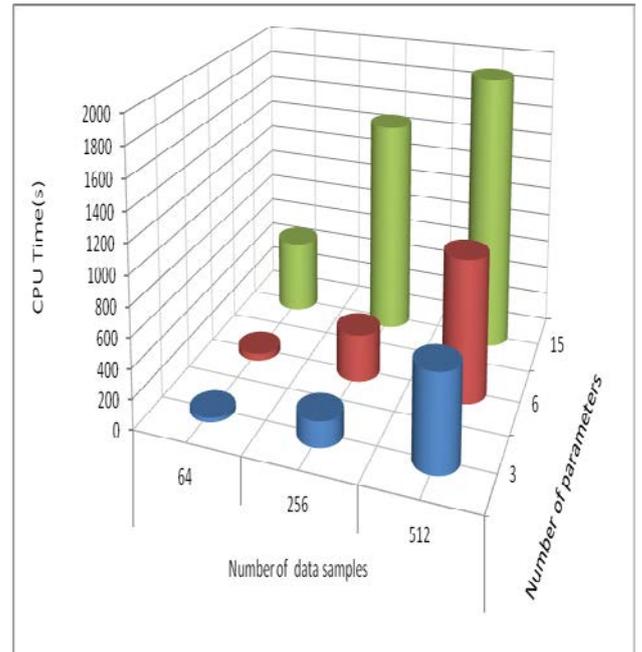
(a)



(b)



**Figure 5.** Performance comparison of Bayesian methods for a single frequency sinusoid: (a) Estimation error variance versus SNR with  $\omega = 0.3$  and  $N = 64$ ; (b) Estimation error variance versus number of data length  $N$  with  $\omega = 0.3$  and SNR = 10dB .



**Figure 6.** CPU times versus with Number of parameters and data samples for BRETHORST.