

Frequency Estimation, Multiple Stationary Nonsinusoidal Resonances With Trend¹

G. Larry Bretthorst

Department of Chemistry, Washington University, St. Louis MO 63130

Abstract. In this paper, we address the problem of frequency estimation when multiple stationary nonsinusoidal resonances oscillate about a trend in nonuniformly sampled data when the number and shape of the resonances are unknown. To solve this problem we postulate a model that relates the resonances to the data and then apply Bayesian probability theory to derive the posterior probability for the number of resonances. The calculation is implemented using simulated annealing in a Markov chain Monte Carlo simulation to draw samples from this posterior distribution. From these samples, using Monte Carlo integration, we compute the posterior probability for the resonance frequencies given the model indicators as well as a number of other posterior distributions of interest. For a single sinusoidal resonance, the Bayesian sufficient statistic is numerically equal to the Lomb-Scargle periodogram. For a nonsinusoidal resonance this statistic is a straightforward generalization of both the discrete Fourier transform and the Lomb-Scargle periodogram. Finally, we illustrate the calculations using data taken from two different astrophysical sources.

INTRODUCTION

Frequency estimation using data obtained from astrophysical sources presents some unique challenges because of the nature of the resonances and the data. The data are almost always nonuniformly sampled. The number of resonances is usually unknown. The shape of each resonance may differ. Each resonance may have both amplitude and phase modulation. Finally, the resonances may oscillate about an unknown trend. These difficulties make the use of the discrete Fourier transform and the Lomb-Scargle periodogram, [1, 2, 3], problematic at best and misleading at worst.

To solve any problem using Bayesian probability theory, one must have a model that relates the hypotheses of interest to the available data. In this first preliminary analysis, the model is of multiple nonsinusoidal stationary resonances oscillating about a trend, so, for now, we are going to ignore amplitude and phase modulation. Note that the presence of trend may make the signal nonstationary, but the resonances we are considering are not changing shape as a function of time, i.e., the resonances are stationary. One way nonsinusoidal resonances manifest themselves in the Fourier transform power spectrum and the Lomb-Scargle periodogram is as peaks at integer multiples of a fundamental resonance frequency. These peaks, called harmonics, expand the shape of the resonance, called a light curve in astrophysics, in a Fourier series. In this paper we will compute the

¹ in Bayesian Inference and Maximum Entropy Methods in Science and Engineering, C. W. Williams ed., pp 3-22, American Institute of Physics, 2003.

posterior probability for the number of stationary nonsinusoidal resonances independent of the light curve of the various resonances given the data and the prior information. This calculation yields a sufficient statistic that reduces to a discrete Fourier transform power spectrum and a Lomb-Scargle periodogram in the appropriate limits. Consequently, the calculation generalizes the discrete Fourier transform to account for the nonsinusoidal nature of the resonances.

THE MODEL

As noted in the introduction, to solve any problem using Bayesian probability theory one must relate the hypotheses of interest to the available data. The multiple resonance model which expands the unknown shape of the resonances in a harmonically related series may be written as

$$d_i = \sum_{\ell=0}^{n_T} T_\ell \mathcal{L}_\ell(t_i) + \sum_{j=0}^m \sum_{k=1}^{n_j} A_{jk} \cos(2\pi k f_j t_i) + B_{jk} \sin(2\pi k f_j t_i) + \text{error} \quad (1)$$

where d_i is the data value acquired at time t_i . The first sum represents the trend, where n_T is the number of expansion functions, $\mathcal{L}_\ell(t_i)$, needed to represent the trend down to the noise. The amplitude of the ℓ th expansion functions is T_ℓ . These expansion functions could be polynomials, a Fourier expansion or any other convenient set of functions. In the program that implements this calculation we used orthogonal polynomials because it made the program computationally more stable, but the use of orthogonal polynomials is not required. In the double sum, the number of resonances has been designated as m and it should be understood that when $m = 0$, the sum over the sinusoids does not appear. The light curve of the j th resonances has been harmonically expanded in a Fourier series having n_j terms. The cosine amplitude of the k th harmonic of the j th resonance has been labeled A_{jk} , and similarly the corresponding sine amplitude is B_{jk} . Finally, the fundamental frequency of each resonance is f_j . Consequently, $k f_j$ is the frequency of the k th harmonic of the j th resonance.

This model equation has a number of interesting and important special cases. For example, when none of the expansions are present, this model reduces to the single stationary frequency model and will reproduce all of the results given in [4, 5]. As another example, when only the trend expansion is present, this model reduces to a trend plus a sinusoid and using it will effectively “detrend” a data set. In a similar vein, when a single frequency is harmonically expanded this model is looking for nonsinusoidal resonances and will result in a sufficient statistic that is a generalization of the Lomb-Scargle periodogram to this important class of problems. So, while this model does not treat the nonstationary resonance case, it nonetheless contains a wealth of important special cases and we will have more to say about them when we apply the calculations to astrophysical data.

Writing the model in the form of Eq. (1) is convenient for explaining where each of the terms comes from, but it is a inconvenient computationally and for applying Bayesian probability theory. In Eq. (1) there are a number of different expansion functions and

sinusoids with amplitudes multiplying them. We are going to designate these amplitudes as $A_\ell \equiv \{T_0, \dots, T_{n_T}, A_{11}, \dots, A_{mn_m}, B_{11}, \dots, B_{mn_m}\}$ and their corresponding sinusoid or expansion functions as $\mathcal{G}_\ell(f_1, \dots, f_m, t_i)$. The expansion functions are not frequency dependent, while all of the sinusoids are; so the notation, $\mathcal{G}_\ell(f_1, \dots, f_m, t_i)$, is more general than needed the sense that some of these functions are not frequency dependent and some are. Nonetheless, we are designating these functions as if they were frequency dependent just as a reminder that some of them are. With this notation, the model can be written as

$$d_i = \sum_{\ell=0}^u A_\ell \mathcal{G}_\ell(f_1, \dots, f_m, t_i) + \text{error} \quad (2)$$

where the total number of functions, u , is given by

$$u \equiv 1 + n_T + \sum_{j=1}^m n_j. \quad (3)$$

The “1” does not appear if $n_T = 0$. The order of the one-to-one mapping of the sinusoids and expansion functions from Eq. (1) to Eq. (2) does not matter in the sense that the sufficient statistics we derive will be order independent. The important point is that Eq. (2) is easier to work with and performing the appropriate Bayesian calculations will be relatively straightforward.

We wish to solve two closely related problems, first estimate the number of resonances, and second, knowing the number of resonances estimate the frequencies. Estimating the number of resonances is the more fundamental problem and, as it will turn out, if we can solve this problem, estimating the frequencies may be done as a derivative calculation. Consequently, in what follows we will derive the posterior probability for the number of resonances given the data and the prior information. This posterior probability is denoted as $P(m|DI)$, where m is the number of resonances, D stands for all of the data, and I represents all of the prior information, for example the functional form of the model, the maximum number of resonances, etc.

The posterior probability for the number of resonances is a marginal posterior probability where all of the effects we do not care about have been removed using the sum and product rules of probability theory. The effects we do not care about include the amplitudes, the number of expansion functions and the number of harmonics needed to represent the light curve of each resonance down to the noise. To compute this posterior probability, one applies Bayes’ theorem to obtain

$$P(m|DI) = \frac{P(m|I)P(D|mI)}{P(D|I)}. \quad (4)$$

The numerator consists of two terms, the prior probability for the number of resonances given only the prior information, $P(m|I)$, and the direct probability for the data given the number of resonances and the prior information, $P(D|mI)$. The denominator, $P(D|I)$, is a normalization constant given by

$$P(D|I) = \sum_{m=0}^{\max} P(mD|I) = \sum_{m=0}^{\max} P(m|I)P(D|mI) \quad (5)$$

where max is an upper bound on the number of resonances.

The prior probability for the number of resonances has been sufficiently simplified that it could be assigned, however the direct probability for the data has not. The direct probability for the data is a marginal probability density function, where the marginalization is over the nuisance parameters. As noted, these include the amplitudes, the number of trend expansion functions, the number of harmonics and, when $P(m|DI)$ is being computed, the fundamental frequency of each resonance. If all of the amplitudes, the number of harmonics and the frequencies are designated by the unadorned symbols A , n and f respectively, then the direct probability for data may be computed as

$$P(D|mI) = \sum_{n_T=0}^{M_T} \sum_{n_1=1}^{M_H} \cdots \sum_{n_m=1}^{M_H} \int dA_0 \cdots dA_u \int df_1 \cdots df_m \int d\sigma P(DAnfn_T\sigma|mI) \quad (6)$$

where $P(DAnfn_T\sigma|mI)$ is the joint probability for the data and the nuisance parameters. We are using M_H to indicate the maximum number of harmonics. Also, we have added one parameter, σ , to represent what is known about the general scale of the errors.

Applying the product rule to the right-hand side of this equation one obtains

$$P(D|mI) = \sum_{n_T=0}^{M_T} \sum_{n_1=1}^{M_H} \cdots \sum_{n_m=1}^{M_H} \int dA_0 \cdots dA_u \int df_1 \cdots df_m \int d\sigma \quad (7)$$

$$\times P(Anfn_T\sigma|mI)P(D|Anfn_T\sigma mI)$$

where $P(Anfn_T\sigma|mI)$ is the joint prior probability for the parameters given the number of resonances, and $P(D|Anfn_T\sigma mI)$ is the direct probability for the data given the parameters.

As an aside, note that we can derive a number of other interesting probabilities using the samples from a Markov chain Monte Carlo simulation. To see how this is possible, note that the product, “prior \times direct probability,” is proportional to the joint posterior probability for all of the parameters given the model. So after obtaining samples from the posterior probability for the number of resonances, the samples with a given number of resonances can be used to generate estimates of the various parameters including the frequencies and number of harmonics, etc.

Continuing with the calculation, the direct probability for the data, the likelihood, cannot be further simplified and, shortly, we will have no choice but to assign a numerical value to represent it. However, the joint prior probability for the nuisance parameters has not been sufficiently simplified for assignment. We continue simplifying this probability by repeated application of the product rule:

$$P(Anfn_T\sigma|mI) \propto P(\sigma|I)P(n_T|I) \left[\prod_{\ell=0}^u P(A_\ell|I) \right] \left(\prod_{k=0}^m P(f_k|mI) \left[\prod_{n_k=1}^{M_H} P(n_k|mI) \right] \right) \quad (8)$$

where we have assumed logical independence of the various parameters, that is to say knowing the value of an amplitude tells us nothing about the number of harmonics or frequencies, etc. With this factorization the posterior probability for the number of

resonances may be written as

$$\begin{aligned}
P(m|DI) &\propto P(m|I) \sum_{j=0}^{M_T} \sum_{n_1=1}^{M_H} \cdots \sum_{n_m=1}^{M_H} \int dA_0 \cdots dA_u \int df_1 \cdots df_m \int d\sigma \\
&\times P(\sigma|I) P(n_T|I) \left[\prod_{\ell=0}^u P(A_\ell|I) \right] \left(\prod_{k=0}^m P(f_k|mI) \left[\prod_{n_k=1}^{M_H} P(n_k|mI) \right] \right) \\
&\times P(D|Anfn_TmI).
\end{aligned} \tag{9}$$

We have now reached the point in this calculation where we have no choice but to assign numerical values to represent these probabilities. For the standard deviation of the noise prior probability, we will assign a Jeffreys' prior probability. In model selection calculations assigning an improper prior probability can cause problems because it introduces a singularity into the calculations. However, here it does not cause a problem because the probability for every model contains exactly the same singularity and so it cancels out when these probabilities are normalized. For the discrete variables we are going to assign an exponential prior probability of the form:

$$P(n|LM) = C \begin{cases} \exp\{-n\} & \text{if } n \in \{0, 1, \dots, M\} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where n represents the hypotheses, for example the number of harmonics or the trend expansion order. The number C is a normalization constant and M is the maximum of the discrete variable. The normalization constant is easily computed for a discrete exponential because

$$\frac{1}{1-x} = x^0 + x^1 + x^2 + \cdots + x^M + \frac{x^{M+1}}{1-x} \tag{11}$$

where $x = 1/e$. Rearranging this somewhat, one obtains

$$\frac{1-x^{M+1}}{1-x} = x^0 + x^1 + x^2 + \cdots + x^M. \tag{12}$$

Identifying the right-hand side of this equation as $1/C$, the normalization constant, C , is given by

$$\frac{1}{C} = \frac{1-x^{M+1}}{1-x} \implies C = \frac{1-x}{1-x^{M+1}}. \tag{13}$$

We will assign a uniform prior probability for the frequencies. However, if you examine the model you will note an interesting thing: reordering the resonances still gives you an m resonance model. This seemingly trivial fact means that the posterior probability for the frequencies has $m!$ identical maxima. For example in a three frequency model, the posterior probability has the property that $P(f_1 f_2 f_3 | DI) = P(f_3 f_2 f_1 | DI) = P(f_2 f_3 f_1 | DI)$, etc. These identical maxima do not correspond to different physical phenomena, they simply correspond to different identifications of what is meant by frequency f_1 , f_2 , etc. In this calculation we will order the frequencies so that f_1 is the

lowest frequency, f_2 is the second lowest frequency, etc. With this convention we have restricted ourselves to a subspace whose volume is $m!$ smaller than the full space and

$$\prod_{j=1}^m P(f_j|mI) = \begin{cases} \frac{m!}{(H-L)^m} & \text{If } L \leq f_1 \leq f_2 \cdots f_m \leq H \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

where H and L are given high and low frequencies.

For the amplitudes, $P(A_j|I)$, we will assign a Gaussian prior probability of the form:

$$P(A_j|I) = \left(\frac{2\pi\sigma^2}{\beta^2 G_{jj}} \right)^{-\frac{1}{2}} \exp \left\{ -\frac{\beta^2 G_{jj} A_j^2}{2\sigma^2} \right\} \quad (15)$$

where this is an unbounded zero mean Gaussian. The squared length of the model function, G_{jj} defined in Eq. (21) below, is a number that ensures we are using the same prior information for all of the amplitudes. For sinusoids $G_{jj} \approx N/2$ and is almost frequency independent. Finally, the hyperparameter β is a guess at how strongly we believe the amplitude is zero relative to what the data are telling us. If $\beta \approx 0.001$, then this prior will change the maximum likelihood estimate of the amplitudes in about the third decimal place. In this calculation we will take β as a given and fix its value at $\beta = 0.001$. Although in a Markov chain Monte Carlo simulation there is no reason why β could not be treated like any other unknown nuisance hypotheses and removed using marginalization.

If we now substitute the priors into the posterior probability for the number of resonances, Eq. (9), and assign the likelihood using a Gaussian prior probability for the errors, one obtains

$$\begin{aligned} P(m|DI) &\propto \exp\{-m\} \frac{m!}{(H-L)^m} \\ &\times \sum_{n_T=0}^{M_T} \sum_{n_1=1}^{M_H} \cdots \sum_{n_m=1}^{M_H} \int df_1 \cdots df_m \int dA_1 \cdots dA_u \int \frac{d\sigma}{\sigma} \\ &\times C_{n_T} C_{n_1} \cdots C_{n_m} \exp\{-n_T - n_1 - \cdots - n_m\} \\ &\times \prod_{j=0}^u \left(\frac{2\pi\sigma^2}{\beta^2 G_{jj}} \right)^{-\frac{1}{2}} (2\pi\sigma^2)^{-\frac{N}{2}} \exp \left\{ -\frac{Q}{2\sigma^2} \right\} \end{aligned} \quad (16)$$

where we have added a subscript to the normalization constant, C_x , to indicate which prior probabilities these constants are associated with. Also constants that cancel on normalization have been dropped. The function Q is defined as

$$\begin{aligned} Q &\equiv \sum_{j=0}^u \beta^2 G_{jj} A_j^2 + \sum_{i=1}^N \left[d_i - \sum_{\ell=0}^u A_\ell \mathcal{G}_\ell(f_1, \dots, f_m, t_i) \right]^2 \\ &= N\bar{d}^2 - 2 \sum_{\ell=0}^u A_\ell T_\ell + \sum_{j=0}^u \sum_{k=0}^u A_j A_k g_{jk} \end{aligned} \quad (17)$$

where we again remind the reader, that when the number of model functions is zero, i.e., there is no signal, none of the sums in the above expression appear. The mean-square data value, $\overline{d^2}$, is defined as

$$\overline{d^2} \equiv \frac{1}{N} \sum_{i=1}^N d_i^2 \quad (18)$$

where N is the total data values. The projection of the data onto the j th model function, T_j , is defined by

$$T_j \equiv \sum_{i=1}^N d_i \mathcal{G}_j(f_1, \dots, f_m, t_i) \quad (19)$$

and g_{jk} is defined as

$$g_{jk} \equiv G_{jk}(1 + \beta^2 \delta_{jk}) \quad (20)$$

where δ_{jk} is the Kronecker delta function [9] and

$$G_{jk} \equiv \sum_{i=1}^N \mathcal{G}_j(f_1, \dots, f_m, t_i) \mathcal{G}_k(f_1, \dots, f_m, t_i). \quad (21)$$

The amplitude integrals are Gaussian quadrature integrals and evaluating such integrals is straightforward, one obtains

$$\begin{aligned} P(m|DI) &\propto \exp\{-m\} \frac{m!}{(H-L)^m} \sum_{n_T=0}^{M_T} \sum_{n_1=1}^{M_H} \cdots \sum_{n_m=1}^{M_H} \int df_1 \cdots df_m \int \frac{d\sigma}{\sigma} \\ &\times C_{n_T} C_{n_1} \cdots C_{n_m} |g_{jk}|^{-\frac{1}{2}} \exp\{-n_T - n_1 - \cdots - n_m\} \prod_{j=0}^u \left(\beta^2 G_{jj}\right)^{\frac{1}{2}} \quad (22) \\ &\times (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{N\overline{d^2} - \overline{h^2}}{2\sigma^2}\right\} \end{aligned}$$

where $|g_{jk}|$ is the determinant of the g_{jk} matrix. The sufficient statistic, $\overline{h^2}$, is defined as

$$\overline{h^2} = \sum_{j=0}^u T_j \mathcal{A}_j \quad (23)$$

and is the total squared projection of the data onto the model (see [4] for more on this type of sufficient statistic). Note that in the above the sufficient statistic is a function of the number of resonances, the frequencies, the trend expansion order and the number harmonics for each resonance. We could have indicated this in the notation, but it would have cluttered things up more than it already is. Consequently, when dealing with this quantity keep in mind that it is a function of both the continuous and discrete parameters. Finally, the expected amplitudes, \mathcal{A}_j , are the amplitudes that maximize the joint posterior probability for all the parameters and are given by

$$\sum_{\ell=0}^u \mathcal{A}_\ell g_{j\ell} = T_j. \quad (24)$$

The integral over the standard deviation of the noise prior probability can be transformed into a gamma function integral and we omit the details of evaluating this integral, one obtains

$$\begin{aligned}
P(m|DI) \propto & \exp\{-m\} \frac{m!}{(H-L)^m} \sum_{n_T=0}^{M_T} \sum_{n_1=1}^{M_H} \cdots \sum_{n_m=1}^{M_H} \int df_1 \cdots df_m C_{n_T} C_{n_1} \cdots C_{n_m} \\
& \times \exp\{-n_T - n_1 - \cdots - n_m\} |g_{jk}|^{-\frac{1}{2}} \prod_{j=0}^u (\beta^2 G_{jj})^{\frac{1}{2}} \left[\frac{Nd^2 - h^2}{2} \right]^{-\frac{N}{2}}
\end{aligned} \tag{25}$$

where we have dropped some additional constants that cancel on normalization. The posterior probability for the frequency is of the form of Student's t -distribution [10] and it is in this form that we will apply it in the next section.

The calculation has proceeded as a model selection calculation and like most model selection calculations it has buried within it a parameter estimation calculation. In particular the quantity on the last line of Eq. (25) is proportional to the posterior probability for the resonance frequencies given the model indicators,

$$P(f_1 \cdots f_m | m, n_T, n_1, \dots, n_m, D, I) \propto |g_{jk}|^{-\frac{1}{2}} \prod_{j=0}^u (\beta^2 G_{jj})^{\frac{1}{2}} \left[\frac{Nd^2 - h^2}{2} \right]^{-\frac{N}{2}} \tag{26}$$

where some additional quantities that are constants for a given model have been dropped. In the following section we will apply both Eq. (25) and Eq. (26) to two examples using astrophysical data sets.

EXAMPLES USING ASTROPHYSICAL DATA

In this section we are going to apply the calculations presented in the previous section to two astrophysical data sets. Initially we will apply the calculations to a data set where the calculation will reduce to single stationary frequency estimation, then we will apply it to a more complicated case to illustrate the behavior of the calculation when a nonsinusoidal resonance is present. The first data set we will analyze is data from the X-Ray Binary LS I +61⁰ 303. This data is shown in Fig. 1(C) along with the error bars given by Gregory [11]. Gregory analyzed this data using a Bayesian calculation for the frequency independent of the shape of the resonance. He finds a single resonance with a 68% credible region from 1599 to 1660 days.

To analyze this data using the calculations given in the previous section we are going to initially assume that Gregory is correct and analyze this data using a single resonance model, then we will use the full calculation to determine the number of resonances and whether the resonance is sinusoidal. Assuming there is only a single stationary sinusoidal resonance, we will apply Eq. (26) with a constant offset, one harmonic, and a single frequency. Consequently, the model reduces to a single stationary sinusoidal resonance oscillating about a constant offset. Because Eq. (26) has only a

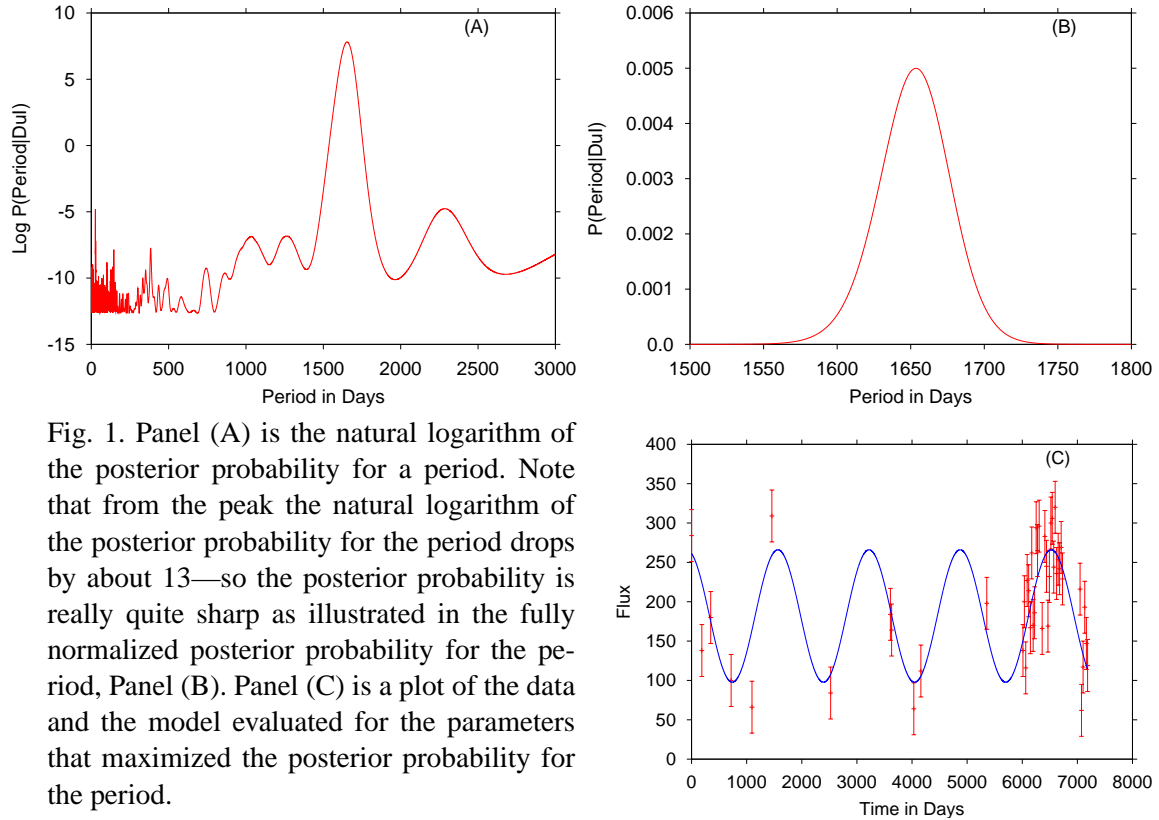


Fig. 1. Panel (A) is the natural logarithm of the posterior probability for a period. Note that from the peak the natural logarithm of the posterior probability for the period drops by about 13—so the posterior probability is really quite sharp as illustrated in the fully normalized posterior probability for the period, Panel (B). Panel (C) is a plot of the data and the model evaluated for the parameters that maximized the posterior probability for the period.

FIGURE 1. Bayesian Analysis of X-Ray Binary LS I +61° 303.

single unknown parameter, the period (the reciprocal of the frequency), we can scan the period over any interval we choose. In this first demonstration we will scan this period from 1 up to 3000 days and plot the natural logarithm of Eq. (26). This plot is displayed in Fig. 1(A). The spectrum of white noise is uniform when plotted as a function of frequency. When the natural logarithm of the posterior probability is computed for white noise, it is also uniform. However, when plotted as a function of period, the magnitude of the oscillations remains uniform but the width of the oscillations stretch for long periods. This stretching is an artifact due to plotting as a function of period. For long periods we are seeing the region around zero frequency. So the rapid oscillations at short period and broad oscillations at high periods are the noise. But what about the region around the maximum, near 1650 days?

The maximum is about 8 and from this peak the logarithm of the posterior probability drops to roughly -5; a drop of about 13, a seemingly small number. But when this posterior probability is normalized, Fig. 1(B), 13 e-foldings (about 442,413) is more than adequate to suppress all of the spurious oscillations and yields a very symmetric

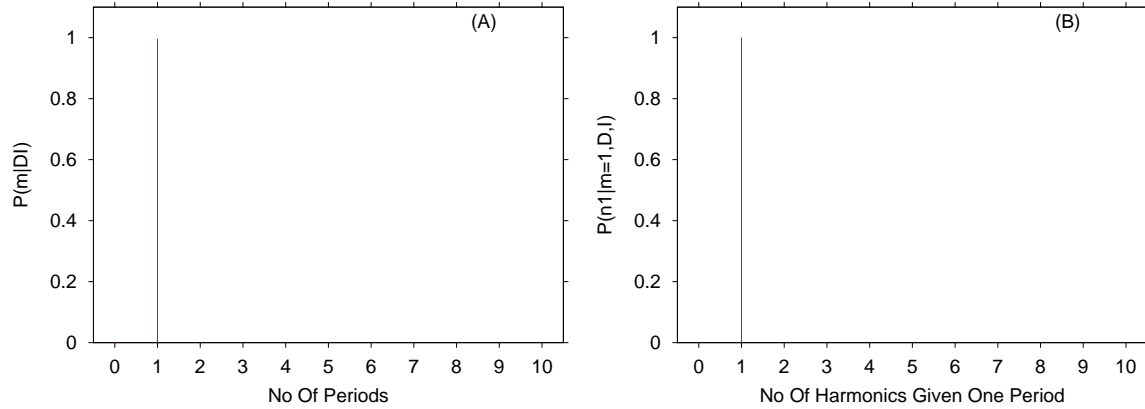


FIGURE 2. Panel (A) is the posterior probability for the number of periods, Eq. (25). In the numerical simulation, all of the simulations settled into a model with one period. Panel (B) is the posterior probability for the number of harmonics given the number of periods is one. This posterior probability was derived from the output from the Markov chain Monte Carlo simulations. After the annealing phase of the Monte Carlo simulation, all of the models contained exactly one period and this period had exactly one harmonic.

posterior probability. If one uses the mean and standard deviation to estimate the period, then the period is estimated to be 1650 ± 50 . In good agreement with the estimate given by Gregory.

In Fig. 1(C) we have overlaid the data with the model generated from the parameters that maximized the posterior probability for the period, where the amplitudes used in this plot were computed from Eq. (24). When this model is overlaid onto the data the period virtually jumps out at one giving good visual confirmation that the period is really there. However, because we have used Eq. (26) we have essentially assumed the signal is present. The only way to be certain that the period is present is to apply Eq. (25) to this data set where we allow the possibility that the number of resonances, m , may be zero. This calculation is illustrated in Fig. 2, where panel (A) is the posterior probability for the number of periods and panel (B) is the posterior probability for the number of harmonics given a one period model.

The numerical calculations are implemented using Markov chain Monte Carlo with simulated annealing. In the annealing phase of the calculation, an annealing parameter is introduced into Eq. (25). If the annealing parameter is designated as α , then the annealing parameter raises everything in Eq. (25), except the prior probability for the number of resonances, $\exp\{-m\}$, to the α power. As the simulations begin, $\alpha = 0$, and everything in Eq. (25) is wiped out except the prior probability for the number of resonances. So the simulations start by exploring the prior probability for the number of resonances. The program that implements the Markov chain Monte Carlo analysis runs multiple independent simulations. In this case 100 simulations were processed. When the simulations are initialized, the parameters (the number of periods, number of harmonics, the periods, etc.) are all set by drawing samples from their respective prior probabilities. When the annealing parameter is zero, the Markov chain is exploring the prior probability for the number of resonances. As the annealing parameter increases slowly from zero to one, the other terms in Eq. (25) become more and more important.

The priors move the simulations towards simpler models, while the data tends to move the simulations towards more complex models. The Markov chain balances these two forces to select the number of resonances and harmonics. Moves between models tend to occur when the annealing parameter is small. As this parameter is increased, moves across models tend to stop because such moves decrease the posterior probability: either the fit to the data is much worse, or the fit to the data is essentially unchanged but prior probabilities decrease, so the model is rejected. The simulations usually settle into a distribution populating only a few of the terms in the sums in Eq. (25) because only high probability models contribute significantly to these sums. Similarly, as a function of the periods, the simulations cluster about the largest peak in the integrand as a function of the periods; again because it is only this peak location that contributes significantly to the integrals. In the example we are discussing, all of the simulations finished the analysis having one period, one harmonic, and the periods were all clustered near 1650. So this analysis just confirms what Gregory has already indicated privately; the resonance is a single stationary sinusoid.

Next we want to illustrate the behavior of this calculation as the resonance becomes nonsinusoidal. To do this we will use a data set from the MACHO project. This ongoing project has been building a catalogue of millions of stars with high signal-to-noise observations for over 10 years. The people working on this project, in particular Doug Welch, were kind enough to point me at a number of rather unusual variable stars. These stars are “Blazhko-type” RR Lyrae variables. These RR Lyrae’s have natural variations on time scales of about half a day and also weeks or months. The cause of this behavior is still not understood. The data shown in Fig. 3 as the asterisks are magnitudes in the blue region of the spectrum. They are nonuniformly sampled covering about 3000 days. If the period is about half a day, then the run of the data cover some 6000 or so oscillations; a tremendous run of data. Because the precision of a stationary frequency estimate varies inversely with the total sampling time, and thus inversely with the total number of periods covered by the run of the data, we can expect resonances in this data to be very precisely determined. The apparent structure in Fig. 3 reflects the sampling scheme, not the resonances in the data. There are gaps in the data that span hundreds of days and it is these gaps that the eye is naturally drawn to. On the scale of this plot, a period of half a day that has been nonuniformly sampled would look like noise. Finally, the plus signs on this plot should be ignored for now. They are a functions of the residuals computed from a model we will discuss shortly.

The sufficient statistic generated by the Bayesian calculation is very closely related to a discrete Fourier transform and when no trend is present and the data are uniformly sampled, it reduces approximately to the discrete Fourier transform power spectrum. Similarly, when no trend is present, but the data are nonuniformly sampled, the numerical value of this statistic is equal to the value of the Lomb-Scargle periodogram, see [5]. So in a sense this sufficient statistic is a generalization of the discrete Fourier transform to the case when trend and nonsinusoidal resonances are present. We would like to illustrate the behavior of this statistic in two regimes: first when it essentially reduces to a Lomb-Scargle periodogram, Fig. 4(A) and second when the resonances are nonsinusoidal, Fig. 4(B). Panel (A) is the natural logarithm of the posterior probability for the period, the reciprocal of the frequency, given a stationary sinusoid oscillating about a constant trend. The presence of multiple peaks in the sufficient statistic, and thus in the

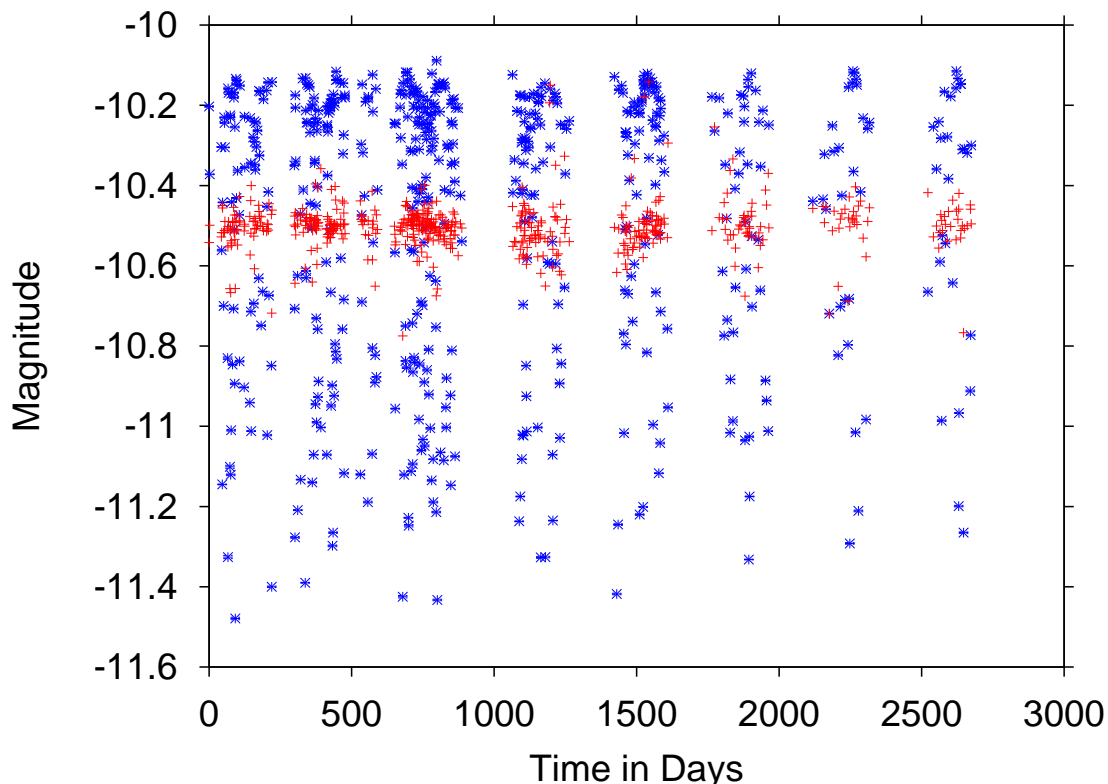


FIGURE 3. MACHO RR Lyrae variable 6.5722.3: This data, the asterisks, are stellar magnitudes in the blue region of the spectrum. The resonance frequency is on the order of half a day, so on the scale of this plot, individual oscillations are not visible. The structure apparent in this data just reflect the sampling schedule, i.e., how often the telescope actually stares at the star. For this data the larger gaps span hundreds of days.

logarithm of the posterior probability, reflects the presence of nonsinusoidal resonances, the sampling scheme, as well as the presence of multiple resonances. Panel (B) is the natural logarithm of the posterior probability for a period given a nonsinusoidal resonance oscillating about a constant trend. The nonsinusoidal resonance was expanded in a fifth order harmonic series, so there are five times as many peaks in this plot as in panel (A). If the expansion had been tenth order, there would have been ten times as many peaks. These peaks occur whenever the frequency of any sinusoid in the model matches either a fundamental or harmonic frequency in the data. When these frequencies match, the projection of the data onto that sinusoid increases and thus the sufficient statistic, and the logarithm of the posterior probability for a frequency has a maximum. Note that every frequency with a large probability density in panel (A) is also prominent in panel (B), but with an even larger density. This increase in probability is due to the harmonics. Harmonic models are n frequency models, when all n frequencies in the model match n frequencies in the data, the total squared projection of the model onto the data has a maximum and the posterior has a global maximum. It is only this global maximum that is relevant to nonsinusoidal frequency estimation.

Both panel (A) and (B) are the logarithm of the posterior probability for a single

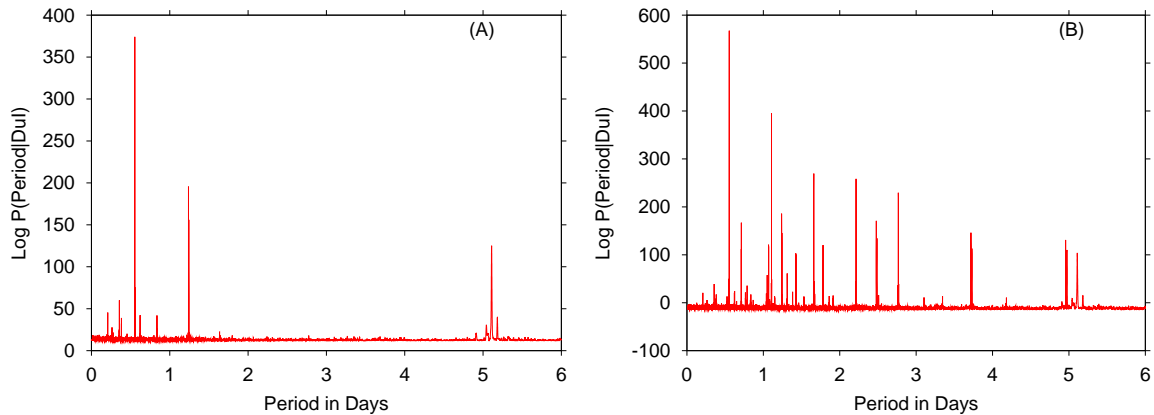


FIGURE 4. Panel (A) is the natural logarithm of the posterior probability for a single stationary sinusoid oscillating about a constant trend and is very closely related to both the discrete Fourier transform and the Lomb-Scargle periodogram. The peaks in this plot reflect both the nonsinusoidal nature of the resonances as well as the presence of multiple resonances. Panel (B) is the natural logarithm of the posterior probability for a single stationary nonsinusoidal oscillation about a constant trend and is a generalization of the discrete Fourier transform and the Lomb-Scargle periodogram. Note that in both cases these statistics have multiple peaks, however, the presence of those peaks reflects the model used, the sampling scheme, as well as evidence for multiple resonances.

period, so it is only the *single* largest peak in these two different log probabilities that are important in estimating the period. In the panel (A) the largest peak has a height of approximately 375 while the second largest peak has a height of about 190. If we were to normalize this posterior probability the second peak would be about 80 orders of magnitude smaller than the peak near 0.5. If you examine panel (B) you will find the difference between the height of the two largest peaks is essentially unchanged so again it is only the peak near 0.5 that is relevant. In this calculation the plotted function, though closely related to the periodogram, is a probability density rather than a power spectrum. Thus the amplitude of the largest peak is inversely proportional to the frequency uncertainty. In panel (A) the highest peak climbs from something slightly greater than zero to 375, while the peak in panel (B) climbs from something slightly less than zero to roughly 675, so one would expect that the model used to generate panel (B) will make a frequency estimate that is roughly two times more precise than the model used to produce the plot in panel (A). Normally to gain a factor of 2 in the precision of a parameter estimate, one would have to quadruple the number of data values or double the signal-to-noise ratio. In this example including the prior information about the harmonics is roughly equivalent to doubling the effective signal-to-noise because the harmonic expansion fits the data much better. Consequently, the mean-square residual has been reduced by roughly a factor of 4, and so the estimated signal-to-noise ratio is about a factor of two improved.

One might wonder if two models having such very different structure will even agree in their parameter estimates to an order of magnitude. To illustrate that the period estimate doesn't vary much as the number of harmonics increases, look at Fig. 5(A). The curve with the lowest peak in this plot is the posterior probability for a stationary sinusoid oscillating about a constant, while the highest peak is the posterior probability

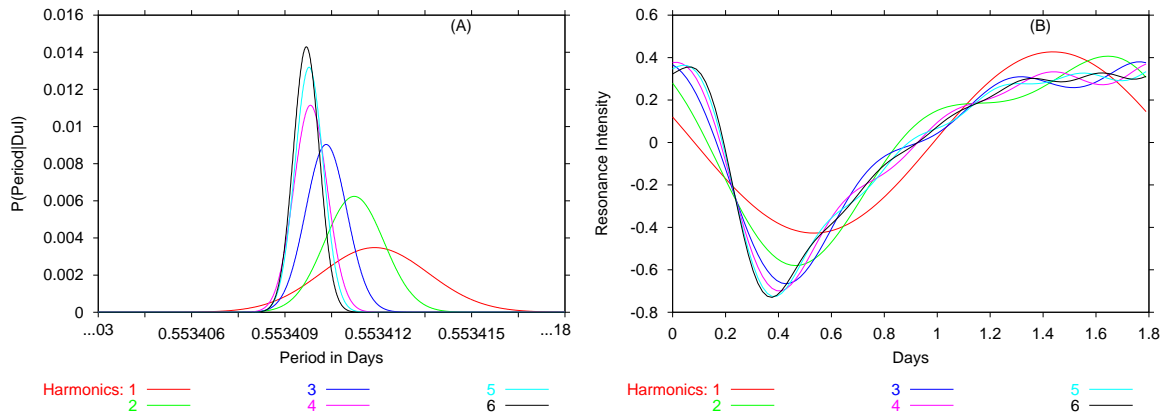


FIGURE 5. Each curves in Panel (A) is the posterior probability for the period, as the number of harmonics increases from the stationary sinusoidal case, the curve with the lowest peak, to the 6 harmonic case the precision of the frequency estimate becomes better and better. Panel (B) shows the light curve for increasing number of harmonics. Note that that as the number of harmonics increases, the light curve converges to a well defined shape that is fairly independent of the number of harmonics.

for the period using a 6 harmonic model. As just indicated, the estimated period is about a factor of two better when the 6 harmonic model is used. The remaining curves are the period estimates from the 2, 3, 4 and 5 harmonic models. Note the precision of the period estimates becomes better and better for increasing numbers of harmonics. One might wonder if this trend would continue indefinitely, the answer is no; the height of the posterior probability for the period decreases after the six harmonic model.

In addition to displaying the posterior probability as a function of the number of harmonics, we have also displayed the light curve for models containing one through six harmonics, panel (B). Here the light curve starts out as a sinusoid, as it must, and then rapidly converges to a nonsinusoidal shape that is fairly independent of the number of harmonics used in the model.

In the example using the data obtained by Gregory, we plotted both the data and the model signal generated from the parameters that maximized the posterior probability for a stationary resonance, Fig. 1(C). This plot helped to guide the eye in identifying the presence of the signal. We would like to do something like that here, but it really is not possible to overlay the model signal and the data because the run of the data is so long there would be more than 6000 oscillations to plot. Any such overlay would look like a solid mass unless we zoomed into a small section of the data and then we would not see how well the data as a whole are represented by the model.

The way this problem is typically solved in astrophysics is to generate a folded light curve. In a folded light curve one subtracts the estimated trend from the data, and then plots the data modulo the period of the primary resonance, in this case the period at 0.5534094 days. This has some significant advantages in that it compresses all of the data onto a single period. One can then overlay this folded light curve with the light curve from the model and everything can be compared nicely in a single plot. The folded light curve of the data along with the light curve generated from the 6 harmonic model are shown in Fig. 6. Note that the light curve of the data appears rather noisy, but, in spite of this, the light curve of the 6 harmonic model seems to fit the folded light curve

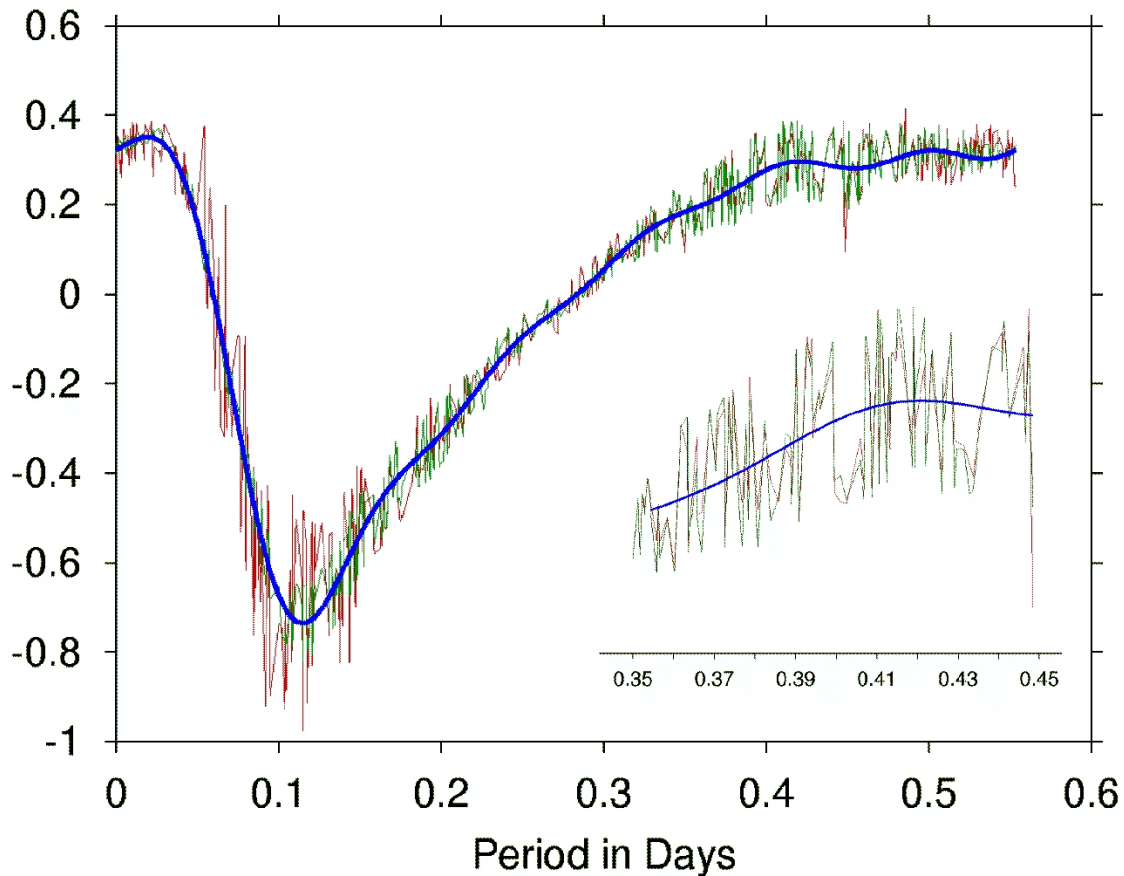


FIGURE 6. The folded light curve of the data has been overlaid with the light curve generated from the 6 harmonic model. Note how this model light curve follows the noisy light curve of the data. But the noise level of the data is much smaller than the deviations seen in this light curve. Indeed these deviations are not noise, but are actually evidence of additional resonances in the data, see text for details.

of the data rather well. However, we must add one cautionary note, the noise level of the data is much much smaller than the apparent noise in the folded light curve. Folding the data in this way presupposes a single resonance. If there are multiple resonances, then these resonances are also folded but with the wrong period. Consequently, these periods appear as if they were noise because the nonuniform sampling, samples them quasi randomly. We will have more to say about the apparent noise level in this plot shortly.

As an alternative method of presentation, note that the plot of the logarithm of the posterior probability for a single stationary sinusoid oscillating about a constant trend, Fig. 4(A), is relatively simple and does contain information about the resonances and the harmonics. If we plot the logarithm of the posterior probability for this data and for a model signal generated from the parameters that maximized the posterior probability then we would be able to see how much of the structure in Fig. 4(A) is accounted for by the model signal. We noted before that this log posterior is essentially the Lomb-Scargle periodogram and in the discussion which follows we will refer to this log posterior

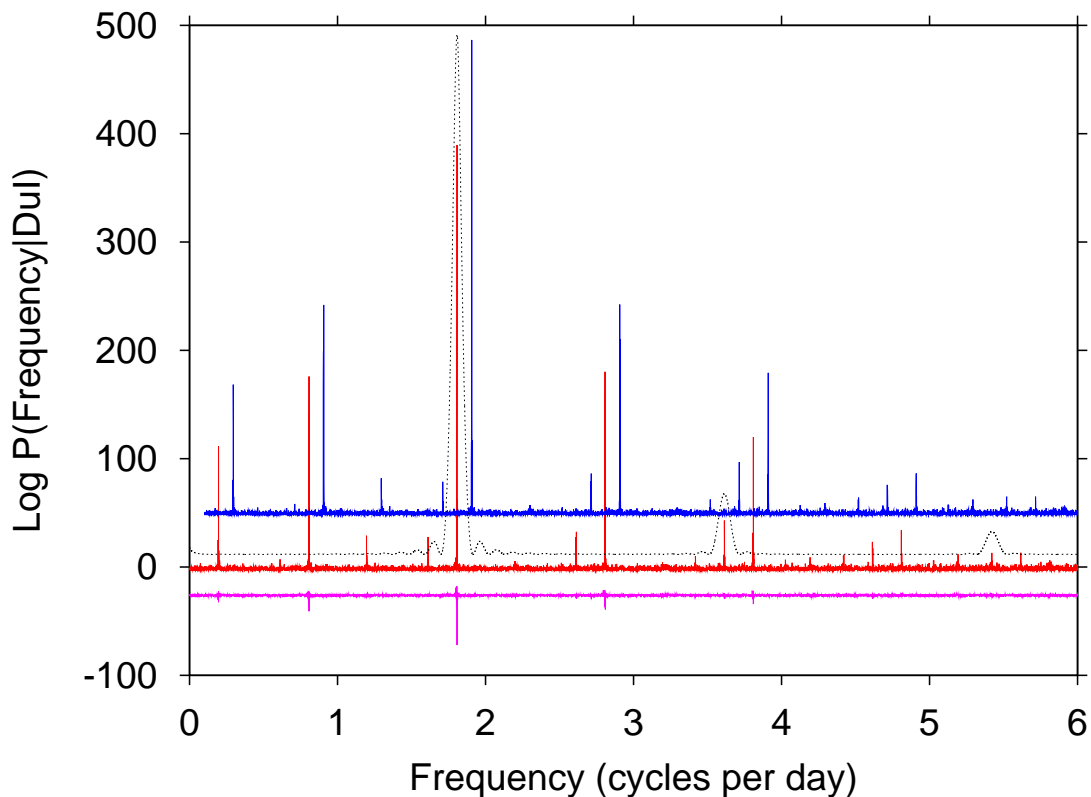


FIGURE 7. We have plotted the logarithm of the posterior probability for a single stationary resonance oscillating about a constant offset for the data and the model signal computed from the maximum of the posterior probability for a model containing six harmonics. The log posterior probability for the model has been shifted upward and to the right so that one can examine both plots carefully, and just below these two curves we have plotted their difference.

probability as the Lomb-Scargle periodogram.

We wish to compare the Lomb-Scargle periodogram of the data to that calculated from the time domain model signal generated from the maximum posterior probability estimate of the period. We wish to do this to see visually how much of the structure in the Lomb-Scargle periodogram is accounted for by the model. Although we have not yet applied the general formalism to this data, we will anticipate one result and use a six harmonic model. So we will generate a time domain model from the parameters that maximize the posterior probability for a single resonance given a six harmonic model. The Lomb-Scargle periodogram of the data and this time domain model signal are shown in Fig. 7. The periodograms in this plot have been plotted as a function of frequency, not period; the harmonic structure of the resonances are more apparent as a function of frequency. We have shifted the Lomb-Scargle periodogram of the time domain model signal upward and to the right because when it was plotted without shifting, these two periodograms overlapped exactly; one simply could not see the differences.

The shifted Lomb-Scargle periodogram was generated from data produced using a time domain model of a single resonance with six harmonics. A six harmonic model contains exactly six stationary sinusoids. Yet its Lomb-Scargle periodogram contains

at least 15 larger peaks, thousands of smaller peaks, and no noise; where did these peaks all come from? The answer is that these pseudo alias peaks are the result of the sampling schedule. If one generates a uniformly sampled time domain model signal from the same parameters used to generate the shifted curve shown in Fig. 7 and then computes the Lomb-Scargle periodogram, one obtains exactly six peaks as one should. We have plotted this Lomb-Scargle periodogram as the dotted line shown in Fig. 7. This Lomb-Scargle periodogram was shifted upward, and its height was scaled to make it fit conveniently on this plot. Consequently, one should ignore both the height and width of this periodogram, the location of the peaks is the only important item. Three of these six peaks are visible in Fig. 7, the other three are outside of the spectral window shown. The three visible peaks sit right on top of the three harmonics associated with the fundamental period near 1.8 cycles per day. Essentially *all* of the other structure in the Lomb-Scargle periodogram of this data is due to the sampling schedule.

In addition to the three periodograms discussed so far, Fig. 7 also contains a plot of the difference between the periodogram of the data and the model signal. This difference has been shifted downward and is shown at the bottom of the figure. These small differences are due to two additional resonances in this data. To detect these additional periods we ran the Markov chain Monte Carlo simulations that implement the full Bayesian calculations. When we ran the simulation, we specified 6 days as the maximum period. We did this for two reasons, first when the single resonance model is used and one examines periods longer than 6 days, nothing is found in the Lomb-Scargle periodogram; while not proof that no long periods exists, it is nonetheless indicative that none exist. Second, assuming there are three resonances, the single resonance model estimated the period to about one part in 10^{-6} , if we assume similar resolution for the other resonance, then the maximum of the posterior occupies a volume of about $10^{-18}/36 \approx 10^{-20}$ of the total volume; a tremendous volume for any algorithm to search. Consequently, when we ran the simulations we wished to keep the maximum period as small as possible.

When we ran this simulation, the posterior probability for the number of resonances had a maximum at three resonances. We have plotted some of the outputs from this analysis in Fig. 8. Panels (A), (C) and (E) are the posterior probability for the three frequencies. Each of these frequencies have been determined to about six significant figures. Panels (B), (D) and (F) are the posterior probabilities for the number of harmonics in these resonances.

Resonance 1 and 3 are stationary sinusoidal resonances. Only resonance 2 is non-sinusoidal, requiring 6 harmonics to expand the light curve. Resonance 2 is the single largest resonance in the data and it was this resonance that was picked out by the single frequency model. The shape of the light curve generated from this resonance is the same as that shown in Fig. 5 (for the 6 harmonic model) and in Fig 6. However, the shape of the light curve for all three resonances taken together is very different from the shape of the light curve for resonance 2 alone. To illustrate this we generated a model signal from the parameters that maximized the joint posterior probability for the parameters given a 3 resonance model. We then folded this model signal in exactly the same way the data were folded and displayed it in Fig. 6. If you examine Fig. 6, especially in the expansion on the lower right, you will see that the light curve of the model signal overlaps the light curve of the data almost exactly. The noise like behavior in the light curve of the data, is not noise at all, rather it is the other two resonances that cannot be accounted for in

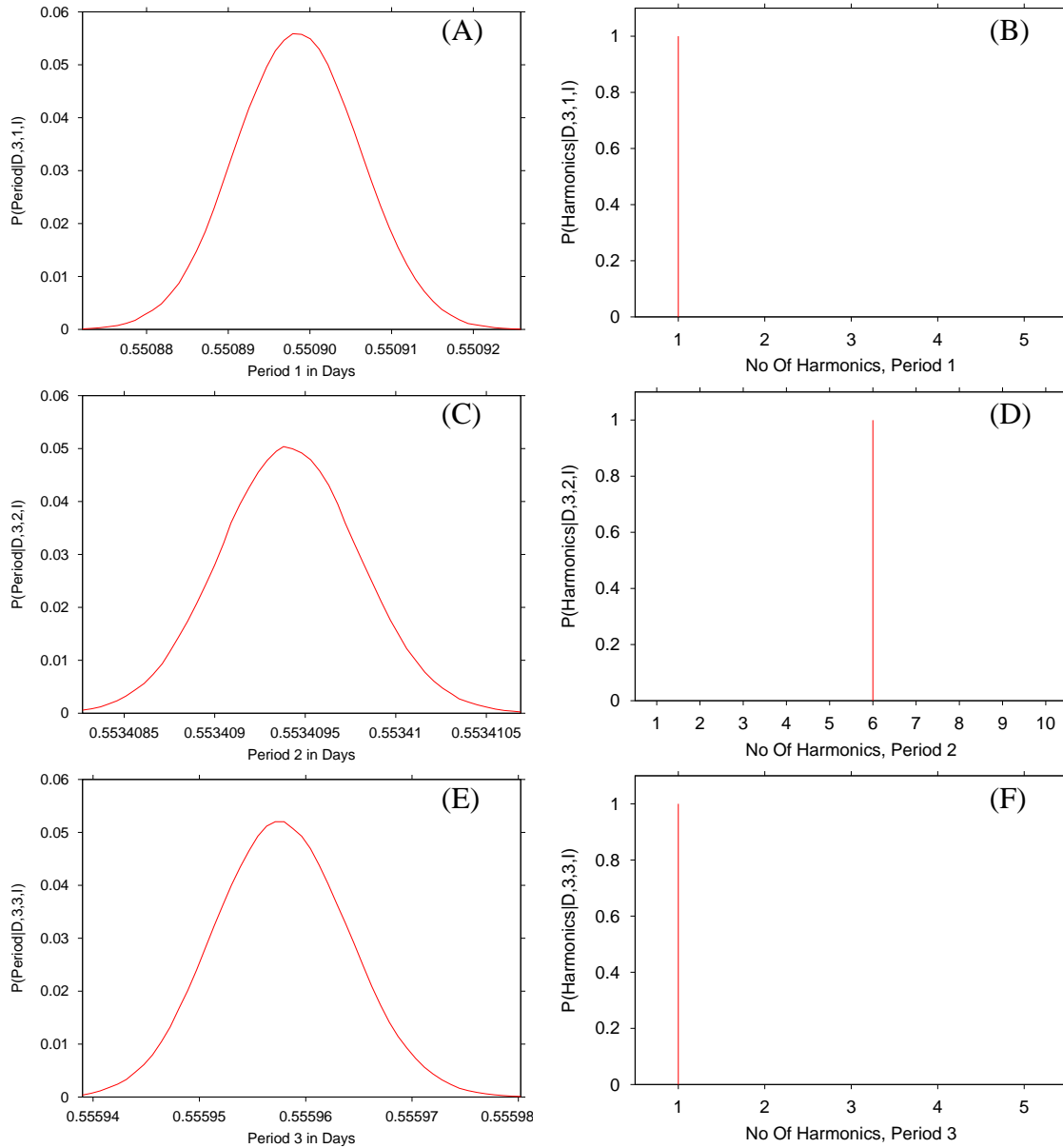


FIGURE 8. Panels (A), (C) and (E) are the posterior probabilities for the three frequencies. Panels (B), (D) and (F) are the posterior probabilities for the number of harmonics given the three frequency model.

a folded light curve. As noted, these other resonances when sampled quasi randomly appear as if they were noise to the eye.

In addition to the folded light curve of the model signal and other outputs discussed, we have also plotted the residuals, the difference between the data and the model (the same model used to produce the folded light curve). These residuals, plus signs, are plotted on top of the data in Fig. 3. We have displaced these residuals by the constant offset so that they may be plotted on the same scale as the data. There is no apparent pattern in these residuals. However, the error bars on the data are much smaller than the fluctuations visible in the residuals. Given the size of the error bars these residual

fluctuations are real in the sense that there are fluctuations in the data that are not accounted for by nonsinusoidal resonances. There are at least two possible sources for these fluctuations: the star could have fluctuations other than periodicities, and there could be systematic errors in the data, for example atmospheric fluctuations, that cannot be controlled by the experimenters. Almost certainly both contribute to these fluctuations.

Two of the three resonances in this data are very close together. It would be an interesting question to determine if, for some reason, the fundamental frequency of a single resonance shifted. To make this determination involves the calculations presented in this paper with two additional model selection calculations: first determine if a change occurred and then determine if the light curve is the same before and after the change. Such calculations would indeed be very interesting but they are beyond the scope of this paper.

SUMMARY AND CONCLUSIONS

Nonsinusoidal resonances manifest themselves in the discrete Fourier transform and the Lomb-Scargle periodogram as peaks at integer multiples of a fundamental frequency. These multiple peaks essentially expand the shape of the resonance in a Fourier series. This fact can be used to model nonsinusoidal resonances and relates the resonance frequencies to the data. This model contains both continuous parameters and discrete parameters and one must devise a formalism that allows one to estimate both types of parameter. For Bayesian probability theory, making such an estimate is a matter of straightforward calculation: no new principles are needed. However, from a computational standpoint the difficulties of exploring multiple discrete dimensions and multiple continuous parameters are formidable. As explained in the previous section, the maximum of the joint posterior probability for the parameters can occupy an incredibly small fraction of the total hyper-volume. Nonetheless this Bayesian calculation can be implemented using Markov chain Monte Carlo with simulated annealing. In the simulated annealing phase of the numerical calculation, the program computes the posterior probability for the model, thus selecting the number of resonances, the number of harmonics and fundamental frequencies. After the annealing phase completes, samples from the simulations are used to approximate various posterior distributions including, the posterior probability for the number of resonances, the resonance frequencies and the number of harmonics for each resonance. Thus the use of simulated annealing, Markov chain Monte Carlo and Bayesian probability theory can solve the problem of multiple nonsinusoidal frequency estimation, while simultaneously determining the number of resonances, the light curve of each resonance and the trend.

ACKNOWLEDGMENTS

The author wishes to acknowledge support from NIH grants CA-83060, NS-35912, and NS-41519. Also, this paper utilizes public domain data obtained by the MACHO

Project, jointly funded by the US Department of Energy through the University of California, Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48, by the National Science Foundation through the Center for Particle Astrophysics of the University of California under cooperative agreement AST-8809616, and by the Mount Stromlo and Siding Spring Observatory, part of the Australian National University.

REFERENCES

1. Lomb, N. R. (1976), *Astrophysics and Space Science*, **39**, pp. 447-462.
2. Scargle, J. D. (1982), *Astrophysical Journal*, **263**, pp. 835-853.
3. Scargle, J. D. (1989), *Astrophysical Journal*, **343**, pp. 874-887.
4. Bretthorst, G. Larry (1988), "Bayesian Spectrum Analysis and Parameter Estimation," in *Lecture Notes in Statistics*, **48**, J. Berger, S. Fienberg, J. Gani, K. Krickenberg, and B. Singer (eds), Springer-Verlag, New York, New York.
5. Bretthorst, G. Larry, "Nonuniform Sampling: Aliasing and Bandwidth," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, edited by Rychert, J. T., G. J. Erickson and C. R. Smith, American Institute of Physics, New York, 1999, AIP Conference Proceedings 567, pp. 1-28.
6. Bayes, Rev. T. (1763), *Philos. Trans. R. Soc. London* **53**, 370; reprinted in *Biometrika* **45**, 293 (1958), and "Facsimiles of Two Papers by Bayes," with commentary by W. Edwards Deming Hafner, New York, 1963. See also [7].
7. Bayes, Rev. T. (1763), *Philos. Trans. R. Soc. London* **53**, 269-271.
8. Jeffreys, H. (1939), "Theory of Probability," Oxford University Press, London.
9. Kronecker, L. (1901), *Vorlesungen über Zahlentheorie*, Teubner, Leipzig; republished by Springer-Verlag, 1978.
10. Student (1908), *Biometrika*, **6**, 1-24.
11. Gregory, P. C. (1999) *Astrophysical Journal*, 520, p. 361-375.
12. Gregory, P. C. and Thomas J. Loredo (1992), *Astrophysical Journal*, 398, p. 146.