Preprint, Astrophysical Journal in press Jan. 17, 1999

BAYESIAN PERIODIC SIGNAL DETECTION. 1. ANALYSIS OF 20 YEARS OF RADIO FLUX MEASUREMENTS OF THE X-RAY BINARY LS I $+61°303$

P. C. Gregory
Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, V6T 1Z1

ABSTRACT

We extend the Bayesian method of Gregory and Loredo (1992) for the detection of a periodic signal of unknown shape and period, to the case where the noise sampling distribution is independent Gaussian. The analysis readily handles nonuniformily sampled data and allows for an unknown noise variance. The method is applied to the radio astronomy data for the interesting X-ray binary system LS I $+61°303$, which exhibits periodic radio outbursts with a period of 26.5 days. Several authors have suggested that the peak flux density of the outbursts exhibit a periodic or quasi-periodic modulation of approximately 1600 days. Our Bayesian analysis of the outburst peak flux densities provides strong support for such a modulation. We derive the posterior probability density function of the modulation period and the estimated mean shape of the modulation based on the available flux density data. The probability density for the modulation period exhibits a broad peak in the range 1599 to 1660 days (68 % credible region) with a mean value of 1632 days. The RMS flux density deviation from the mean shape, amounting to 45 mJy, is much larger than the measurement errors of $\approx$ 10 mJy which suggest additional complexity in the source which is yet to be understood. The next maximum in the long term flux modulation is predicted to occur near July 22, 1999 (Julian day 2,451,382).

Subject headings: Bayesian methods, period detection, Gregory-Loredo method, LS I $+61°303$, X-ray binaries, pulsars, time series analysis

1. INTRODUCTION

LS I $+61°303$ (V615 Cas, GT 0236+610, 2CG 135+01) is particularly interesting among high-mass X-ray binaries because of its strong variable emission from radio to X-ray and probably γ-ray (Gregory and Taylor 1978, Kniffen et al. 1997). At radio wavelengths it exhibits periodic radio outbursts with a period of 26.5 days (Taylor and Gregory 1982, 1984) and there is recent evidence for an approximately 5 fold 26.7 $\pm$ 0.2 day modulation of the 2-10 keV X-ray flux (Paredes et al. 1997). However, the radio outbursts are not stable in phase. Outburst maxima have been seen from phase 0.45 to 0.95, but bright maxima seem to occur near 0.6 (Paredes, Estella, & Rius 1990). Furthermore the peak flux of the outbursts are known to exhibit a long term ≈ 4 year modulation (Gregory et al. 1989, Marti 1993, Marti and Paredes 1995). The most recent estimate of the time scale of the long term modulation is 1598 ± 13 days (Peracaula 1997).

In the period 1977 August to 1992 August a total of 14 outbursts were recorded by a variety of groups. However, beginning in January 1994 (Ray et al. 1997) detailed monitoring was performed (several times a day) with the National Radio Astronomy Observatory Green Bank Interferometer (GBI). This has yielded high quality data for an additional 44 cycles to date. From the GBI data, Ray et al. (1997) reported a secular change in the outburst phase indicating either orbital period evolution, or a drift in orbital phase. Based on the first 2 years of the GBI data (28 cycles) they find only weak evidence for the proposed long term periodic outburst peak flux modulation.

In this paper we will focus on the variations in the outburst peak flux. In a companion paper (Gregory et al. 1998, Paper II) we address the nature of the outburst phase and its relationship to the outburst peak flux. We use Bayesian analysis to assess the relative probabilities of three hypotheses for explaining the outburst peak flux variations. Included is the hypothesis that the outburst peak flux is periodically modulated. For this purpose we exploit the Gregory-Loredo (GL) Bayesian method (Gregory and Loredo 1992a, b, 1993, 1996; henceforth referred to as GL1, GL2, GL3 and GL4). This method is well suited for the detection and measurement of a periodic signal when we have no specific prior knowledge of its shape. In GL1 and GL2 details of the theory were presented for the case of a time series where the appropriate noise sampling distribution is the Poisson distribution. For an example of its use in X-ray astronomy see Gregory and Loredo (1996); henceforth referred to as GL4. The focus of the current paper is to generalize the GL method to time series with independent Gaussian noise. In this respect the results are quite general and could be applied to a wide variety of time series. The analysis of LS I $+61^{\circ}303$ data in this and a companion paper (Gregory et al. 1998) serves as an interesting example of its capabilities.

In section 2 we consider the hypotheses to be tested and present the data. In section 3 we develop the appropriate equations for the GL method for the case where the noise sampling distribution is Gaussian. In section 4 we apply the method to an analysis of the LS I $+61^{\circ}303$ outburst flux data and show the case for a periodic modulation is very strong. In sections 5 and 6 we obtain the Bayesian estimate of the period and light curve for the flux modulation. In section 7 we discuss the results and examine the question of quasi-periodic flux modulation. Our conclusions are presented in section 8.

2. Hypothesis Space and Data

In a Bayesian analysis the first step is to define the hypothesis space of interest. In this problem we are interested in hypotheses concerning a time series with associated Gaussian errors. In the particular case of LS I $+61°303$ the time series consist of the times and peak flux densities of the radio outbursts.

We consider three hypotheses concerning the outburst peak flux behavior. They are:

ABBREVIATION	HYPOTHESIS
H_C	Deviations of the outburst peak flux values from a constant value are consistent with Gaussian random noise.
H_P	Outburst peak flux values have a modulation which is periodic but of a priori unknown shape and period.
H_{NP}	Outburst peak flux values have a non periodic modulation of unknown shape.

As we will see shortly hypotheses H_C and H_{NP} are limiting cases of a family of periodic models capable of describing a periodic structure of arbitrary shape for use with H_P. To this list we might like to add a hypothesis to represent a quasi-periodic modulation but this is beyond the scope of the present work.

The data consists of the estimated time and flux density corresponding to the peak of each outburst. These are listed in table 1 together with an error estimate for each quantity and a literature reference.

In addition to the long term variation in outburst flux, there was considerable variation in peak flux from outburst to outburst which was much larger than measurement errors of ≈ 10 mJy. The RMS variation in flux between adjacent outbursts was 33 mJy. Close inspection of the GBI data during periods when the 26.5 day outbursts are very weak or undetectable, indicates the existence of shorter duration flaring with amplitudes of up to 100 mJy. These flares also appear to be present at other times as well but are superposed on the larger 26.5 day outbursts. They are probably responsible for the 33 mJy outburst to outburst RMS flux variation noted above. Indeed the data on LS I $+61°303$ indicate a rich variety of structures much like the layers of an onion. We have adopted the approach of trying to understand one layer of complication at a time.

The objective in this paper is the study of the long term modulation of the 26.5 day outbursts leaving a detailed study of the flares for later. To this end we have assumed an error of 33 mJy for the well defined outburst and 50 mJy in three cases where the coverage was not as complete. Because of the complications to the outburst light curve arising from the flares we did not use any sophisticated peak finding routines, such as cross correlation analysis, but instead simply took the highest data value for each outbursts and estimated the time error from shape of the underlying outburst. In section 3 we discuss precisely how the flux error information is used in our Bayesian analysis.

As mentioned above there are times when the outbursts are very weak and not well enough defined to derive a meaningful error for the outburst time. However, there is still useful information concerning the flux modulation and since this occurs on a time scale of approximately 4 years the timing accuracy required is not high. So these outbursts were still included, if the coverage was sufficient, to provide information about the peak flux density. These are indicated by a blank in the timing error column. Similarly in two cases it was only possible to obtain a lower limit on the flux density but it was still possible to obtain useful information on the outburst times although with a larger estimated error.

Although the Julian date of the highest data points are given to two decimal places, to identify the peaks measurements used, we have assumed outburst timing error as quoted in the error column. Only the subset of Table 1 data values with a corresponding flux error listed were used in the flux modulation analysis of this paper. For this analysis we do not make any use of the outburst timing errors. In the companion paper (Gregory et al. 1998) we make use of the subset of Table 1 with a quoted timing error, in a study of the outburst phase variations. We caution against the use of Table 1 for studies of variability on still shorter time scales. Such analysis should be carried out with the original data set.

In this and the companion paper we do not address the question of the possible frequency dependence of the phenomena under study. The literature contains observations spanning a wide frequency range. The GBI observations at 2.8 and 8.3 GHz are dominant in the period after January 1994. Prior to this the majority of observations were at 5.0 and 10.5 GHz. To reduce any possible frequency dependence and access the longest time base we have restricted the current analysis to the range 5.0 to 10.5 GHz.

3. Gregory-Loredo Method, Gaussian Noise Case

The GL method addresses the problem of the detection and characterization of a periodic signal in a time series when we have no specific prior knowledge of the existence of such a signal or of its characteristics, including shape. The method was developed in GL1, GL2 and GL3 to deal with photon arrival time data in X-ray and γ-ray astronomy, where the appropriate sampling distribution is the Poisson distribution. A discussion of the usefulness of this approach relative to other common period finding algorithms is given in GL2 an GL4. In the current work we are dealing with radio flux density measurements, i.e. analyzing a sampled time series with Gaussian noise. Our analysis does not assume uniform sampling; the Bayesian approach allows us to draw optimal inferences about the nature of the signal for what ever data is available. We can represent any data value d_i by the following equation:

$$d_i = d_{pi} + e_i, \tag{1}$$

where d_{pi} = the value of d_i predicted by the model and e_i is a noise term representing the uncertainty in d_i. In general e_i consists of the random measurement errors plus any real signal in the data that cannot be explained by the model. For example suppose the data contain two periodic signals, P_a and P_b, and the model assumes only P_a is present. In this connection the P_b signal acts like an additional unknown noise term. In the absence of detailed knowledge of the noise distribution, other than that it has a finite variance, the maximum entropy principle tells us that a Gaussian distribution would be the most conservative choice (i.e. maximally non committal about the information we don't have). For a justification of this argument see the work of E. T. Jaynes (e.g. "Probability Theory: The Logic of Science", available at http://bayes.wustl.edu/, to be published by Cambridge University Press). In this paper we will assume the noise variance is finite and adopt a Gaussian distribution for e_i with a variance σ_i^2.

In a Bayesian analysis where the variance of e_i is unknown, but assumed to be the same for all data points, we can treat σ as an unknown parameter. The full Bayesian inference is a joint posterior distribution for all of the parameters; but its implication for any parameter of interest can be summarized by integrating out the uninteresting parameters. For example if we are interested in the period, we treat σ as a nuisance parameter and integrate out this parameter. This procedure of integrating out nuisance parameters is called marginalization. Marginalizing over σ has the desirable effect of treating anything in the data that can't be explained by the model as noise and this leads to the most conservative estimates of model parameters. We can also use Bayes's theorem to compute $p(\sigma \mid D, \mathrm{Model}, I)$ to learn about the effective noise in fitting the model to the data. For a more detailed discussion of marginalization see sections 2.2 and 2.3 of GL2.

In the current problem we do not assume all the σ_i are the same. We let s_i = the experimenter's estimate of σ_i, prior to fitting the model and examining the model residuals. Our estimates, s_i, are given in Table 1. The σ_i values are not known but the s_i values are our best estimates which also contain information on the relative weight we want to associate with each point. Since we do not know the absolute values of the σ_i we introduce a parameter called the noise scale parameter, b, to allow for this. It could also be called a noise weight parameter. Several different definitions of b are possible including $\sigma_i^2 = b\ s_i^2$ and $\sigma_i = b\ s_i$. The definition we will use in this paper is given by,

$$\frac{1}{\sigma_i^2} = \frac{b}{s_i^2}. \tag{2}$$

Again marginalizing over b has the desirable effect of treating anything in the data that can't be explained by the model as noise, leading to the most conservative estimates of model parameters. Since b is a scale parameter we will assume a Jeffreys prior (see section 3.2). A nice feature of this prior is that it corresponds to to assuming a Jeffreys prior for σ_i as well.

We can also use Bayes's theorem to compute $p(b \mid D, \text{Model}, I)$. If the most probable estimate of $b \approx 1$, then the model is doing a good job accounting for everything that is not noise based on our s_i estimates. If $b < 1$ then either the model is not accounting for significant features in the data or the initial noise estimates, s_i, were low. We hasten to point out that the Bayesian approach to model selection involves the direct comparison of the global probabilities of competing models. The global probability of any model requires marginalizing over all the model parameters which gives rise to an Occam penalty associated with each parameter (see GL2 section 2.3 for details).

3.1. Periodic Model

Hypothesis H_P represents a class of models with periodic structure. The periodic models describe the signal plus any arbitrary offset with a stepwise function, resembling a histogram, with m phase bins per period. Clearly, such a stepwise model is capable of approximating a light curve of arbitrary shape. Although such a model is well suited to describe "spiky" periodic signals, our simulations in GL2 demonstrated that this model can usefully detect smooth signals and accurately estimate their frequency and shape.

H_P represents the whole class of stepwise periodic model for which M_m is one specific member with m bins. The Bayesian posterior probability for a periodic model contains a term which quantifies Occam's razor, penalizing successively more complicated periodic models for their greater complexity even though they are assigned equal prior probabilities. We will be interested in calculating $p(H_P \mid D, I)$ by marginalizing over a range of M_m models corresponding to a prior range of m values, from $m = 2$ to 12.

$$p(H_P \mid D, I) = \sum_{m=2}^{m_{\max}} p(M_m \mid D, I), \tag{3}$$

where from Bayes's theorem,

$$p(M_m \mid D, I) = \frac{p(M_m \mid I) p(D \mid M_m, I)}{p(D \mid I)}. \tag{4}$$

The term in the denominator is a normalization constant, obtained by summing the products of the priors and the global likelihoods of all models being considered.

We consider each member of this class to be equally probable *a priori*, so that the probability assigned to the periodic class is spread equally among the $\nu = m_{\max} - 1$ members of this class. Thus,

$$p(M_m \mid I) = \frac{1}{\nu}\, p(H_P \mid I) \tag{5}$$

Alternatively, we could view our stepwise models as a single model, and m as a discrete parameter in this model; the $1/\nu$ factor in equation (5) then plays the role of a flat prior distribution for the parameter m.

For each choice of m the model contains $m+3$ parameters: an angular frequency ω (or alternatively a period $P = 2\pi/\omega$), a phase (position of first bin relative to the start of the data), an unknown noise scale factor, b, plus m parameters describing the shape of the periodic light curve. A remarkable feature of the stepwise model is that it enables marginalization of the m shape parameters to be performed analytically, leaving only the period and phase to be integrated numerically.

3.2. Likelihood Function for the Stepwise Model

The next step is to evaluate the global likelihood $p(D \mid M_m, I)$ for the stepwise model. We will let r_j represent the light curve value in bin j. The value of the subscript j corresponding to any particular sample time t is given by,

$$j(t) = \mathrm{int}[1 + m\{(\omega t + \phi) \bmod 2\pi\}/2\pi]. \tag{6}$$

We sometimes denote the full set of m values of r_j by the symbol $\vec{r}$.

We can expand the global likelihood according to equation (7) where we substitute for θ our specific model parameters.

$$\begin{aligned} p(D \mid M_m, I) &= \int d\omega \int d\phi \int db \int d\vec{r}\; p(\omega \mid M_m, I) p(\phi \mid M_m, I) p(b \mid M_m, I) \\ &\times\; p(\vec{r} \mid M_m, I) p(D \mid \omega, \phi, b, \vec{r}, M_m, I). \end{aligned} \tag{7}$$

Next we need to select priors for the parameters and compute the likelihood $p(D \mid \omega, \phi, b, \vec{r}, M_m, I)$. Following the arguments given in GL2, the prior density for ϕ we take to be uniform over the interval $[0, 2\pi]$,

$$p(\phi \mid M_m, I) = \frac{1}{2\pi}, \tag{8}$$

and b and ω to have a Jeffreys prior of the form,

$$p(b \mid M_m, I) = \frac{1}{b \ln \frac{b_{hi}}{b_{lo}}}, \tag{9}$$

In our calculations we set $b_{LO} = 0.05$ and $b_{HI} = 1.95$.

$$p(\omega \mid M_m, I) = \frac{1}{\omega \ln \frac{\omega_{\rm hi}}{\omega_{\rm lo}}}, \tag{10}$$

where $[\omega_{\rm lo}, \omega_{\rm hi}]$ is a prior range for ω and the $\ln \frac{\omega_{\rm hi}}{\omega_{\rm lo}}$ factor is a normalization constant ensuring that the integral of $p(\omega \mid M_m, I)$ over the prior range is equal to 1. This density is uniform in the logarithm of ω. The Jeffreys prior arises from an invariance property: invariance of conclusions with respect to scale changes in time (E. T. Jaynes 1968 and his unpublished "Probability Theory: The Logic of Science"). In addition, this prior is form-invariant with respect to reparameterization in terms of P. That is, an investigator working in terms of P, and using a $1/P$ prior, will reach the same conclusions as an investigator working in terms of ω, and using a $1/\omega$ prior. This would not be true for a prior of another form, for example a uniform prior. To the extent that parameterization in terms of P and ω are both equally "natural", this form of invariance is desirable. Of course if the prior range of a scale parameter is small when expressed as a fraction of the center value, then the conclusions will not differ significantly from those arrived at by assuming a uniform prior.

The likelihood function requires there to be at least several periods in the data set so we set $P_{hi} = 2\pi/\omega_{\rm lo} = T/3$, where $T = 7523$ d, the data duration. We set $P_{lo} = 2\pi/\omega_{\rm hi} = 800$ d, or $1/2$ the previous estimate of the period. Changing the value of $\omega_{\rm hi}$ by a factor of a few does not greatly affect our results, because $\omega_{\rm hi}$ enters calculations only through the logarithmic factor.

3.2.1. *Evaluating $p(D \mid \omega, b = 1, \phi, \vec{r}, M_m, I)$*

We first write the likelihood for the m bin model, M_m, as a function of all the parameters assuming $b = 1$, which corresponds to $\sigma_i = s_i$. Equation (6) tells us which stepwise bin any particular sample d_i falls in, for a given ω and ϕ. The predicted value of d_i in bin j is r_j. The difference $(d_i - r_j)$ is assumed to have a Gaussian distribution with variance σ_i^2. Assuming each d_i is independent we can now write the desired likelihood as a product of Gaussians.

$$p(D \mid \omega, b = 1, \phi, \vec{r}, M_m, I) = \prod_{j=1}^{m} \left[(2\pi)^{-\frac{n_j}{2}} \{ \prod_{i=1}^{n_j} (s_i)^{-1} \} \exp\left(-\frac{\alpha}{2}\right) \right], \tag{11}$$

where n_j = number of samples in bin j and,

$$\alpha \;=\; \sum_{i=1}^{n_j} \frac{(d_i - r_j)^2}{s_i^2} = \sum_{i=1}^{n_j} \frac{d_i^2}{s_i^2} - 2r_j \sum_{i=1}^{n_j} \frac{d_i}{s_i^2} + r_j^2 \sum_{i=1}^{n_j} \frac{1}{s_i^2}$$

$$
\begin{aligned}
&= W_j\{\overline{d^2_{Wj}} - 2r_j\overline{d_{Wj}} + r_j^2\} \\
&= W_j\left[\{r_j^2 - 2r_j\overline{d_{Wj}} + (\overline{d_{WJ}})^2\} + \overline{d^2_{Wj}} - (\overline{d_{Wj}})^2\right] \\
&= W_j(r_j - \overline{d_{Wj}})^2 + \chi^2_{Wj}, \qquad (12)
\end{aligned}
$$

where,

$$W_j = \sum_{i=1}^{n_j} \frac{1}{s_i^2}, \qquad (13)$$

and,

$$\overline{d_{Wj}} = \frac{\sum_{i=1}^{n_j} \frac{d_i}{s_i^2}}{\sum_{i=1}^{n_j} \frac{1}{s_i^2}} = \frac{\sum_{i=1}^{n_j} \frac{d_i}{s_i^2}}{W_j}, \qquad (14)$$

and,

$$\overline{d^2_{Wj}} = \frac{\sum_{i=1}^{n_j} \frac{d_i^2}{s_i^2}}{W_j}, \qquad (15)$$

and,

$$
\begin{aligned}
\chi^2_{Wj} &= \sum_{i=1}^{n_j} \frac{(d_i - \overline{d_{Wj}})^2}{s_i^2} \\
&= W_j\{\overline{d^2_{Wj}} - (\overline{d_{Wj}})^2\}. \qquad (16)
\end{aligned}
$$

3.2.2. Marginalizing over $\vec{r}$

To marginalize over the shape parameters, $\vec{r}$, we need to introduce a prior for $\vec{r}$ and evaluate,

$$p(D \mid \omega, b = 1, \phi, M_m, I) = \prod_{j=1}^{m}\left[(2\pi)^{-\frac{n_j}{2}}\{\prod_{i=1}^{n_j}(s_i)^{-1}\}\exp\left(-\frac{\chi^2_{Wj}}{2}\right)R\right], \qquad (17)$$

where,

$$R = \int_{r_{min}}^{r_{max}} dr_j\ p(r_j \mid M_m, I)\exp\left(-\frac{W_j(r_j - \overline{d_{Wj}})^2}{2}\right). \qquad (18)$$

In general r_j is not necessarily a positive quantity so we treat it as a location parameter and assume a flat prior distribution for each r_j in the range r_{min} to r_{max}, thus

$$p(r_j \mid M_m, I) = \frac{1}{\Delta r} = \frac{1}{r_{max} - r_{min}}. \qquad (19)$$

In the present problem r_j stands for a flux density which is always positive. The prior range used in the analysis of the LS I +61°303 data was $r_{min} = 0.0$ mJy to $r_{max} = 400$ mJy In terms of this prior equation (17) becomes,

$$
\begin{aligned}
p(D \mid \omega, b=1, \phi, M_m, I) &= (2\pi)^{-\frac{N}{2}}(\Delta r)^{-m}\{\prod_{i=1}^{N}(s_i)^{-1}\}\exp\left(-\sum_{j=1}^{m}\frac{\chi^2_{Wj}}{2}\right) \\
&\times \prod_{j=1}^{m}\left[\int_{r_{min}}^{r_{max}} dr_j \exp\left(-\frac{W_j(r_j-\overline{d_{Wj}})^2}{2}\right)\right].
\end{aligned} \tag{20}
$$

We can evaluate the integral over r_j in terms of the complimentary error function,

$$
\mathrm{erfc}(y) = \frac{2}{\sqrt{\pi}}\int_y^{\infty}\exp(u^2)du \tag{21}
$$

With $u^2 = W_j(r_j - \overline{d_{Wj}})^2/2$, we can write,

$$
\int_{r_{min}}^{r_{max}} dr_j \exp(-\frac{W_j(r_j-\overline{d_{Wj}})^2}{2}) = \sqrt{\frac{\pi}{2}}(W_j)^{-1/2}\left[\mathrm{erfc}(y_{jmin}) - \mathrm{erfc}(y_{jmax})\right], \tag{22}
$$

where,

$$
y_{jmin} = \sqrt{\frac{W_j}{2}}(r_{min} - \overline{d_{W_j}}) \quad ; \quad y_{jmax} = \sqrt{\frac{W_j}{2}}(r_{max} - \overline{d_{W_j}}) \tag{23}
$$

Substituting into equation (20) we obtain finally,

$$
\begin{aligned}
p(D \mid \omega, b=1, \phi, M_m, I) &= (2\pi)^{-\frac{N}{2}}(\Delta r)^{-m}\{\prod_{i=1}^{N}(s_i)^{-1}\}(\frac{\pi}{2})^{m/2}\exp(-\sum_{j=1}^{m}\frac{\chi^2_{Wj}}{2}) \\
&\times \prod_{j=1}^{m}\{W_j^{-1/2}\left[\mathrm{erfc}(y_{jmin}) - \mathrm{erfc}(y_{jmax})\ \right]\}.
\end{aligned} \tag{24}
$$

3.2.3. Evaluating $p(D \mid \omega, \phi, b, M_m, I)$ for $b \neq 1$

In the general case $b \neq 1$, which results in a few minor changes to our earlier equations. We need to replace all references to W_j by bW_j and s_i by $s_i b^{-\frac{1}{2}}$ with the exception that both $\overline{d_{Wj}}$ and $\overline{d^2_{Wj}}$ remain the same. Equation (24) becomes,

$$
\begin{aligned}
p(D \mid \omega, b, \phi, M_m, I) &= (2\pi)^{-\frac{N}{2}}(\Delta r)^{-m}\{\prod_{i=1}^{N}(s_i)^{-1}\}(\frac{\pi}{2})^{m/2} b^{\frac{N-m}{2}}\exp(-\frac{b}{2}\sum_{j=1}^{m}\chi^2_{Wj}) \\
&\times \prod_{j=1}^{m}\{W_j^{-1/2}\left[\mathrm{erfc}(y_{jmin}) - \mathrm{erfc}(y_{jmax})\ \right]\},
\end{aligned} \tag{25}
$$

where,

$$y_{jmin} = \sqrt{\frac{bW_j}{2}}(r_{min} - \overline{d_{W_j}}) \quad ; \quad y_{jmax} = \sqrt{\frac{bW_j}{2}}(r_{max} - \overline{d_{W_j}}) \tag{26}$$

We can now rewrite equation (7) as,

$$\begin{aligned} p(D \mid M_m, I) &= \frac{(2\pi)^{-\frac{N}{2}}(\Delta r)^{-m}\{\prod_{i=1}^{N}(s_i)^{-1}\}(\frac{\pi}{2})^{m/2}}{2\pi \ln\frac{\omega_{hi}}{\omega_{lo}} \ln\frac{b_{hi}}{b_{lo}}} \int \frac{d\omega}{\omega} \int d\phi \int \frac{db}{b} b^{\frac{N-m}{2}} \exp(-\frac{b}{2}\sum_{j=1}^{m}\chi_{Wj}^2) \\ &\times \prod_{j=1}^{m}\{W_j{}^{-1/2}\,[\mathrm{erfc}(y_{jmin}) - \mathrm{erfc}(y_{jmax})\ ]\}. \end{aligned} \tag{27}$$

For very small data sets it is possible that there will be less than two samples in one or more bins. See appendix A for a discussion on a modification to handle small data sets.

We can extract useful information about what the data and our prior information have to say about the noise scale parameter, b, by computing the marginal posterior probability density of b. In Figure 1(a) we have plotted what we call the projected probability of b for the prior range of $b = 0.05 \ - \ 1.95$. It is equal to the product of the projection of the multidimensional likelihood, $p(D \mid \omega, \phi\, b, m, I)$ on to the b axis times our prior for b. In this work we use the projected probability because in practice is often a reasonably good approximation to the marginal and is much easier to compute. The most probable value of $b = 0.68$ means that the effective noise for model H_P, namely everything that can't be fit by this model, is $\approx 1/\sqrt{0.68} = 1.21$ times the estimated noise sigma.

3.3. Likelihood Function for the Constant Model, H_C

The constant model is just a special case of a periodic model when there is only one bin (i.e. $m = 1$).

$$p(D \mid H_C) = p(D \mid M_1) \tag{28}$$

Consequently the model has only two parameters: the unknown value of the constant signal level, A, and the noise scale parameter, b. The prior ranges used were $\Delta A = 400$ mJy and $b = 0.05$ to 1.95. The global likelihood for this model is given by,

$$\begin{aligned} p(D \mid M_1) &= \frac{(2\pi)^{-\frac{N}{2}}(\Delta A)^{-1}\{\prod_{i=1}^{N}(s_i)^{-1}\}(\frac{\pi}{2})^{1/2}}{\ln\frac{b_{hi}}{b_{lo}}} \int_{b_{lo}}^{b_{hi}} \frac{db}{b} b^{\frac{N}{2}-1} W^{-1/2} \exp(-\frac{b}{2}\chi_W{}^2) \\ &\times \ [\mathrm{erfc}(y_{A_{min}}) - \mathrm{erfc}(y_{A_{max}})\ ], \end{aligned} \tag{29}$$

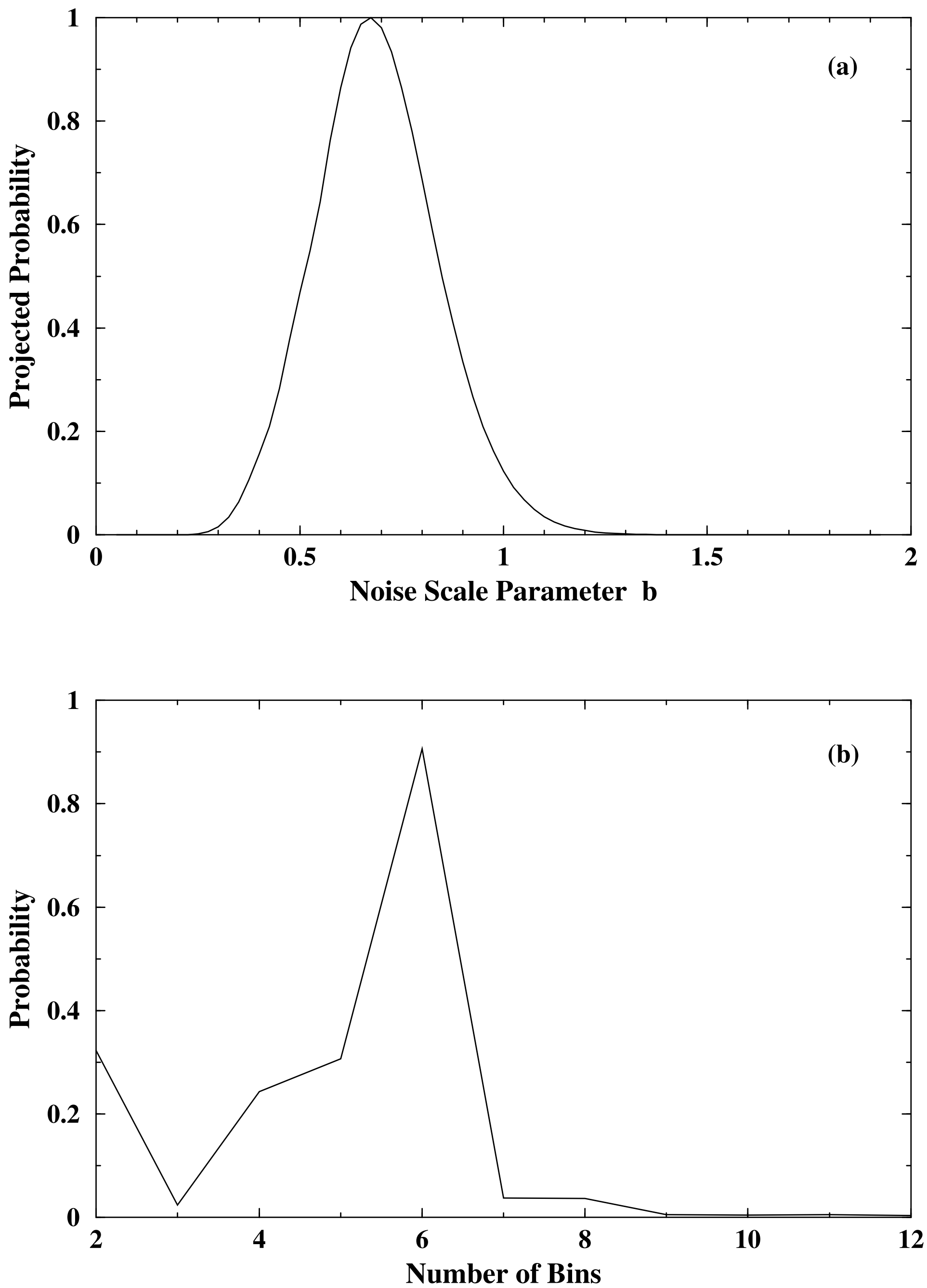


Fig. 1.— Panel (a) shows the projected probability of the noise scale parameter b and panel (b) the marginal probability of the number of bins, m, in the stepwise model.

where,

$$W = \sum_{i=1}^{N} \frac{1}{s_i^2}, \tag{30}$$

and,

$$y_{A_{min}} = \sqrt{\frac{bW}{2}}(A_{min} - \overline{d_W}) \;\; ; \;\; y_{A_{max}} = \sqrt{\frac{bW}{2}}(A_{max} - \overline{d_W}) \tag{31}$$

3.4. Likelihood Function for the Non Periodic Modulation Model, H_{NP}

The non periodic modulation model is just another special case of a periodic model, when the period is equal to the duration of the data. In general the number of bins required to explain a non periodic modulation is greater than for the periodic modulation so we need to consider a wider range of m. Suppose the data does contain a real periodic modulation whose shape can be adequately described with 5 bins. To attempt to describe this modulation with a non periodic model to the same accuracy would require $5 \times T/P$ bins, where T = data duration and P = the period. In this case we would expect the probability of the periodic model to be much greater due to the heavier Occam penalty that would be levied against the non periodic model because of the much larger number of bins that would be required. Of course the periodic model would suffer an additional Occam factor due to the unknown period search range but here the penalty increases only logarithmically with range. A prior range of $m = 2$ to 20 was used for the non periodic model.

The non periodic modulation model has $m + 2$ parameters: m shape parameters, bin phase and the noise scale parameter, b. H_{NP} represents the whole class of stepwise non periodic models for which N_m is one specific member with m bins. The global likelihood for this model is given by,

$$p(H_{NP} \mid D, I) = \sum_{m=2}^{m_{\max}} p(N_m \mid D, I), \tag{32}$$

The global likelihood for N_m, the same as that for M_m without the ω terms, is given by,

$$\begin{aligned} p(D \mid N_m) \;\; &= \;\; \frac{(2\pi)^{-\frac{N}{2}}(\Delta r)^{-m}\{\prod_{i=1}^{N}(s_i)^{-1}\}(\frac{\pi}{2})^{m/2}}{2\pi \ln \frac{b_{hi}}{b_{lo}}} \int d\phi \int \frac{db}{b} b^{\frac{N-m}{2}} \exp(-\frac{b}{2}\sum_{j=1}^{m}\chi_{Wj}^2) \\ &\times \;\; \prod_{j=1}^{m}\{{W_j}^{-1/2}\,[\mathrm{erfc}(y_{jmin}) - \mathrm{erfc}(y_{jmax})\;]\}. \end{aligned} \tag{33}$$

4. Model Selection

The first question is a detection problem, "Is there evidence for a periodic signal regardless of the exact shape, period and phase of the signal?" We address the detection question by using Bayesian probability theory to compare the relative probabilities of the three competing hypotheses presented in section 2

It is useful to consider the ratios of the probabilities of the H_P to each of the other two models. The ratio, $O_{PC} = p(H_P \mid D, I)/p(H_C \mid D, I)$, is called the *odds ratio* in favor of model H_P over the constant model H_C. Application of Bayes's theorem leads to,

$$O_{PC} = \frac{p(H_P \mid I)}{p(H_C \mid I)} \frac{p(D \mid H_P)}{p(D \mid H_C)} \equiv \frac{p(H_P \mid I)}{p(H_C \mid I)} B_{PC} \tag{34}$$

where the first factor is the prior odds ratio, and the second factor is called the *Bayes factor*. The Bayes factor is the ratio of the global likelihoods of the models. In a similar fashion we can write

$$O_{PNP} = \frac{p(H_P \mid I)}{p(H_{NP} \mid I)} \frac{p(D \mid H_P)}{p(D \mid H_{NP})} \equiv \frac{p(H_P \mid I)}{p(H_{NP} \mid I)} B_{PNP} \tag{35}$$

B_{PC} is computed from equations (3), (27), (28) and (29) and B_{PNP} from the first two equations plus (32) and (33).

Bayesian model comparison relies on the ratio of global likelihoods, not maximum likelihoods. A crucial consequence of the marginalization procedure used to calculate global likelihoods is that the Bayes factor automatically favors simpler models unless the data justify the complexity of more complicated alternatives. Each parameter of a model that is marginalized over, in the process of computing the global likelihood, introduces an Occam's factor penalizing the model for wasted prior parameter space that gets ruled out by the data (see GL2, section 2.3). Thus the greater the model complexity the more Occam's penalties it must overcome in comparison with a simpler model. Thus even if we assign equal prior probabilities to competing models, simpler models can have larger posterior probabilities than their competitors. The calculation thus balances model simplicity with goodness-of-fit, allowing us to determine both whether there is evidence for a periodic signal and the optimum number of bins for describing the structure in the data.

The global likelihoods and the most probable value of the noise parameter, b, were computed for each model. Recall that by marginalizing over b, anything in the data that cannot be described by the model is assumed to be noise which leads to the most conservative estimates of model parameters. If $b < 1$ then either the model is not accounting for significant features in the data or the initial noise estimates, s_i, were low. Equation (2) can be used to compute the derived values of the outburst flux density noise sigma given the most probable b and the initial estimate of the typical noise of $s_i = 33$ mJy. The most probable values of b and derived noise sigma are listed in Table 2.

It is clear that the periodic model is capable of explaining more of the structure in the data than the other two models because its noise parameter is significantly closer to one. The computed Bayes factors are $B_{PC} = 1.2 \times 10^8$ and $B_{PNP} = 1.4 \times 10^5$. Not surprisingly the Bayes's factors also strongly support the periodic model.

If we assume that a priori the three models are equally probable then the two odds ratios are equal to their corresponding Bayes's factors. Keeping in mind that in this paper we are restricting our hypothesis space to the three models, then the probability of H_P can be related to the odds ratios in the following way.

$$\begin{aligned} p(H_P \mid D, I) &= \frac{p(H_P \mid D, I)}{p(H_P \mid D, I) + p(H_C \mid D, I) + p(H_{NP} \mid D, I)} \\ &= \frac{1}{1 + \frac{1}{O_{PC}} + \frac{1}{O_{PNP}}} \end{aligned} \tag{36}$$

This leads to a value of $p(H_P \mid D, I) = 0.999989$. We note that the estimated noise sigma for even the periodic model of 45 mJy is considerably higher than the measurement uncertainty (5 to 10 mJy). This implies that there is scope for still more complex models to describe deeper levels of the sources behavior. Eventually we hope to expand our hypothesis space to include additional more complex models.

We also derive the most probable number of bins required to describe the periodic modulation for our existing state of knowledge. The probability versus bin number m is shown in Figure 1(b) and has a maximum for $m = 6$.

5. ESTIMATING THE PERIOD

Now that we have established the presence of a periodic flux modulation we apply Bayes's theorem to determine the probability distribution of the modulation period or frequency, ω. We do this first for the m bin periodic model and later marginalize over m.

$$p(\omega \mid D, M_m, I) = p(\omega \mid M_m, I)\,\frac{p(D \mid \omega, M_m, I)}{p(D \mid M_m, I)}. \tag{37}$$

In parameter estimation questions like this the global likelihood, $p(D \mid M_m, I)$, merely plays the role of a normalization constant. The result is,

$$p(\omega \mid D, M_m, I) \;=\; \frac{C}{\omega} \int_{b_l o}^{b_h i} \frac{db}{b}\, b^{\frac{N-m}{2}} \int_{\phi=0}^{\phi=2\pi} d\phi\; \exp(-\frac{b}{2} \sum_{j=1}^{m} \chi^2_{Wj})$$

$$\times \prod_{j=1}^{m} \{ W_j{}^{-1/2} \left[\mathrm{erfc}(y_{jmin}) - \mathrm{erfc}(y_{jmax}) \ \right] \}, \tag{38}$$

where $C =$ is a normalization constant, equal to the integral of the right hand side of above equation over ω. The marginal contains all the information provided by the data about the frequency when we are not interested in the exact shape of the light curve or its phase. This distribution summarizes the information provided by the data about the frequency. In particular, a "best" frequency can be found by locating the maximum of equation (38) (giving the mode) or by calculating the posterior mean. The probability that the true frequency is in any specified range can be found simply by integrating the equation over that range.

Note that the normalization constants for the priors, Δr, 2π, and $\ln(\omega_{\mathrm{hi}}/\omega_{\mathrm{lo}})$, do not appear, making the posterior very insensitive to the prior ranges for the parameters. This is a general characteristic of Bayesian calculations: the results of parameter estimation tend to be very insensitive to the prior ranges of parameters, being essentially equal to limiting results obtained with infinite range. This is in contrast to model comparison calculations, which can depend more sensitively on the prior parameter ranges.

These calculations are all conditional on the choice of a particular model, M_m. But the Bayesian model comparison calculations of the previous section do not isolate a single model; rather, they assign a probability to each possible model (just as Bayesian parameter estimation does not produce a single point in parameter space, but rather a posterior "bubble"). Formally, our estimate of the frequency should include the information provided by all of the models, essentially marginalizing over m, which we can consider a discrete nuisance parameter. We can perform this calculation as follows.

We let $(m > 1)$ stand for the proposition, "the signal is periodic, not constant." Then a complete description of our knowledge of the frequency of the periodic signal is given by the marginal distribution,

$$\begin{aligned} p(\omega \mid m > 1, D, I) &= \sum_{m=2}^{m_{\max}} p(M_m, \omega \mid m > 1, D, I) \\ &= \sum_{m=2}^{m_{\max}} p(M_m \mid D, I)\, p(\omega \mid D, M_m, I), \end{aligned} \tag{39}$$

where

$$\begin{aligned} p(M_m \mid D, I) &= \frac{p(M_m \mid I)\, p(D \mid M_m, I)}{\sum_{m=2}^{mmax} p(M_m \mid I)\, p(D \mid M_m, I)} \\ &= \frac{p(D \mid M_m, I)}{\sum_{m=2}^{mmax} p(D \mid M_m, I)} \end{aligned} \tag{40}$$

and $p(\omega \mid D, M_m, I)$ is from equation (38). Equation (39) is a weighted sum of the marginal posteriors for all the periodic models being considered, from those with $m = 2$ to $m = m_{\max} = 12$.

Figure 2(a) shows the posterior probability of the flux modulation period, obtained from equation (39).

Figure 2(b) shows an exploded view of the peak flux modulation period probability distribution The solid curve is the result after marginalizing the noise scale parameter over the prior range $b = 0.05$ to 1.95 . For comparison we show the results assuming two fixed values of b. The dashed curve corresponds to $b = 1$ and the dot dashed curve to $b = 1.8$. Recall that the typical flux value in Table 1 of ≈ 33 mJy, based on the scatter of adjacent outburst fluxes is already much larger than the individual flux measurement errors of ≈ 10 mJy. The value of $b = 1.8$ comes from assuming σ_i equal to the typical flux measurement error of 10 mJy, or $b = \sqrt{33/10} = 1.8$. In section 4 we found that the most probable value of $b = 0.68$. Recall that if $b < 1$ then either the model is not accounting for significant features in the data or the initial noise estimates, s_i, were low. The three choices for b: (a) marginalizing over a prior range, (b) $b = 1.0$, and (c) $b = 1.8$, constitute three different models. We can readily evaluate the global likelihood for each. The relative global likelihood of the three is $2.1 \times 10^6 : 1.1 \times 10^6 : 1.0$. Thus, in spite of its greater complexity, the model which treats b as a free parameter and marginalizes over it has the highest probability. Figure 2(b) demonstrates that marginalizing over b leads to the most conservative estimate of the period, in that it has a broader probability distribution. This example illustrates the potential danger of a simple maximum likelihood analysis in which the measurement errors are assumed to be the sole source of uncertainty. In this way one might be lead to conclude falsely that the dot dashed curve in Figure 2(b) was our best estimate of the probability distribution for the modulation period.

The mean of the probability distribution is at 1632 days with a 68.3% credible region (posterior bubble) extending from 1599 to 1660 days. The 68.3% credible region includes other recent estimates of 1598 ± 13 days obtained by a very different analyses (Marta 1997) and 1605 ± 24 Marti 1993).

6. ESTIMATING THE PERIODIC LIGHT-CURVE SHAPE

In this section we estimate the shape of the periodic modulation together with the uncertainty of the shape. The light curve is entirely specified by the stepwise function $r(t) = r_{j(t)}$ where j is the bin number corresponding to a time t. The estimate of $r(t)$ at one particular time is not simply the estimate of a particular r_j parameter, because the value of j corresponding to t depends on the frequency and phase, which are usually not known. Additionally, the estimate will depend on which model (number of bins) is chosen. Formally, since the shape of the signal is determined by the choice of the model and by the values of *all* of the model parameters, the full Bayesian shape estimate for any given value of m is the joint posterior distribution, $p(M_m, \omega, b, \phi, \vec{r} \mid D, I)$. We

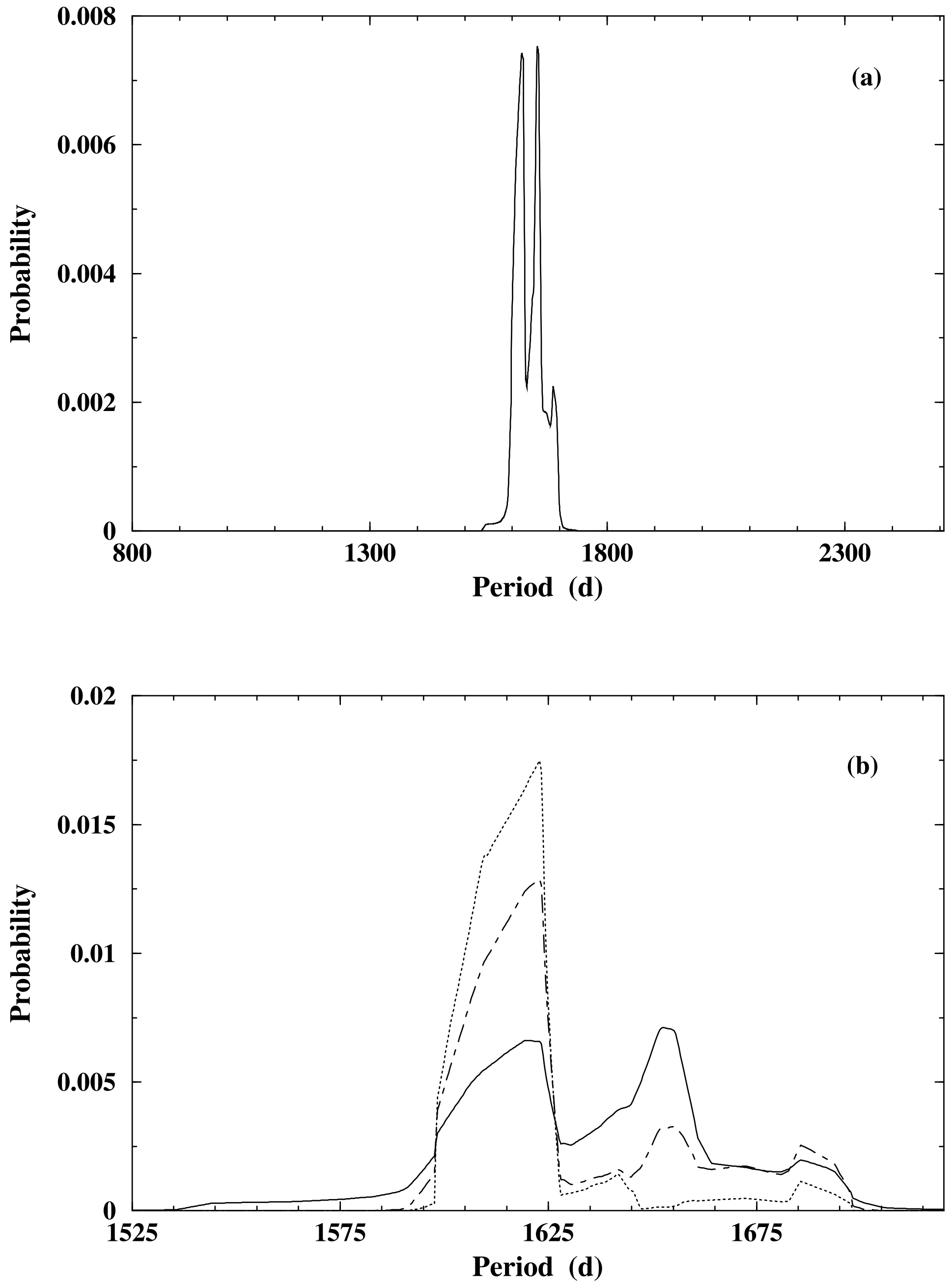


Fig. 2.— Panel (a) shows the marginal probability of the peak flux modulation period of the LS I +61°303 outbursts. Panel (b) gives an exploded view of the peak flux modulation period probability distribution. The solid curve is the result after marginalizing over the noise scale parameter b, the dashed curve is for $b = 1$ and the dot dashed curve for $b = 1.8$.

will call this distribution the *shape distribution.* The shape distribution is so complicated as to be almost useless by itself. In this section we seek to summarize its implications in an intuitively accessible manner.

6.1. Conditional Mean of r_j

As a first step toward a more informative summary of the shape distribution, we now seek the posterior mean of r_j when m, ω, b, and ϕ are specified. The mean (expected value) of r_j is given by,

$$\langle r_j \mid \omega, b, \phi, m\rangle = \int_{r_{min}}^{r_{max}} dr_j \; r_j \; p(r_j \mid \omega, b, \phi, D, M_m, I), \tag{41}$$

where,

$$p(r_j \mid \omega, b, \phi, D, M_m, I) = p(r_j \mid M_m, I) p(D \mid \omega, b, \phi, r_j, M_m, I)/p(D \mid \omega, b, \phi, M_m, I). \tag{42}$$

Combining results from section 3.2.1 and 3.2.3 we obtain,

$$\begin{aligned} p(D \mid \omega, b, \phi, r_j, M_m, I) &= \int dr_1 \cdots \int dr_{j-1} \int dr_{j+1} \cdots \int dr_m \; p(r_1, \cdots, r_{j-1}, r_{j+1}, \cdots, r_m \mid M_m, I) \\ &\times \; p(D \mid \omega, b, \phi, \vec{r}, M_m, I), \end{aligned} \tag{43}$$

and,

$$p(D \mid \omega, b, \phi, M_m, I) = \int dr_j \; p(r_j \mid M_m, I) p(D \mid \omega, b, \phi, r_j, M_m, I), \tag{44}$$

where,

$$p(D \mid \omega, b, \phi, \vec{r}, M_m, I) = \prod_{j=1}^{m} \left[(2\pi)^{-\frac{n_j}{2}} \{ \prod_{i=1}^{n_j} (s_i)^{-1} \} b^{\frac{n_j}{2}} \exp(-\frac{b\chi^2_{Wj}}{2}) \Lambda \right], \tag{45}$$

where,

$$\Lambda = \exp(-\frac{bW_j(r_j - \overline{d_{Wj}})^2}{2}). \tag{46}$$

Substituting equations (44), (43) and (45) into equation (42) and simplifying yields,

$$p(r_j \mid \omega, b, \phi, D, M_m, I) = \frac{\Delta r \exp(-\frac{bW_j(r_j - \overline{d_{Wj}})^2}{2})}{\int_{r_{min}}^{r_{max}} dr_j \exp(-\frac{bW_j(r_j - \overline{d_{Wj}})^2}{2})}. \tag{47}$$

We now have all the components to evaluate equation (41),

$$\langle r_j \mid \omega, b, \phi, m\rangle = \frac{\int_{r_{min}}^{r_{max}} dr_j \; r_j \; \exp(-\frac{bW_j(r_j - \overline{d_{Wj}})^2}{2})}{\int_{r_{min}}^{r_{max}} dr_j \exp(-\frac{bW_j(r_j - \overline{d_{Wj}})^2}{2})}. \tag{48}$$

Equation (48) can be evaluated in terms of the complimentary error function,

$$\langle r_j \mid \omega, b, \phi, m\rangle = \overline{d_{Wj}} + \frac{\sqrt{2}(\exp(-bW_j(r_{max} - \overline{d_{Wj}})^2/2) - \exp(-bW_j(r_{min} - \overline{d_{Wj}})^2/2))}{\sqrt{\pi b W_j}\left[\mathrm{erfc}(\frac{bW_j}{2}(r_{min} - \overline{d_{Wj}})) - \mathrm{erfc}(\frac{bW_j}{2}(r_{max} - \overline{d_{Wj}}))\right]}. \qquad (49)$$

Provided the boundaries of the integral are far from the peak (in units of the width of the integrand) as they are in this problem, to a very good approximation one is just integrating over full Gaussians yielding only the first term. In this paper we will use the equation,

$$\langle r_j \mid \omega, b, \phi, m\rangle = \overline{d_{Wj}} \qquad (50)$$

The standard deviation of the marginal posterior for r_j, defined by

$$\begin{aligned} \langle \sigma_j \mid \omega, b, \phi, m\rangle &= \langle (r_j - \langle r_j\rangle)^2\rangle^{1/2} \\ &= \left[\langle r_j^2\rangle - \langle r_j\rangle^2\right]^{1/2} \\ &\approx \left[\langle r_j^2\rangle - \overline{d_{Wj}}^2\right]^{1/2}, \end{aligned} \qquad (51)$$

provides a measure of the uncertainty of the value of r_j. Following the analysis of $\langle r_j \mid \omega, b, \phi, m\rangle$ we can express equation (51) as:

$$\langle \sigma_j \mid \omega, b, \phi, m\rangle^2 = (bW_j)^{-1} +$$

$$\frac{\sqrt{2}((r_{max} - \overline{d_{Wj}})\exp(-bW_j(r_{max} - \overline{d_{Wj}})^2/2) - (r_{min} - \overline{s_{Wj}})\exp(-bW_j(r_{min} - \overline{d_{Wj}})^2/2))}{\sqrt{\pi b W_j}\left[\mathrm{erfc}(\frac{bW_j}{2}(r_{min} - \overline{d_{Wj}})) - \mathrm{erfc}(\frac{bW_j}{2}(r_{max} - \overline{d_{Wj}}))\right]} \qquad (52)$$

Again to a very good approximation only the first term is important, thus

$$\langle \sigma_j \mid \omega, b, \phi, m\rangle = \sqrt{(bW_j)^{-1}}, \qquad (53)$$

where W_j is determined by the estimated data errors s_i according to equation (13). To get a better understanding for equation (53) assume all s_i = constant = s. Then we can rewrite equation (53) as,

$$\langle \sigma_j \mid \omega, b, \phi, m\rangle = \frac{s}{\sqrt{bn_j}}. \qquad (54)$$

Thus we see the "root-n" dependence we might have guessed. The noise scale parameter factor, $\sqrt{b}$, occurring in the denominator allows for our earlier parameterization of the noise σ (see equation (2)).

6.2. Averaging Over Frequency, Noise Scale Parameter and Phase

Recall that all of these results are conditional on ω, b, ϕ (and the choice of m). When ω, b and ϕ are unknown, we need to estimate $r(t)$ at a particular time, t, which may not correspond to estimating a particular r_j, since the uncertainty in ω, b and ϕ make $j(t)$ uncertain. We can take into account our uncertainty in ω, b and ϕ by marginalizing. This can all be done numerically using the results already found. For example, for $\langle r(t) \mid m \rangle$ we have,

$$\begin{aligned}
\langle r(t) \mid m \rangle &= \int d\omega \int db \int d\phi \int d\vec{r}\, r_{j(t)}\, p(\omega, b, \phi, \vec{r} \mid D, M_m, I) \\
&= \int d\omega \int db \int d\phi \left[\int d\vec{r}\, r_{j(t)}\, p(\vec{r} \mid \omega, b, \phi, D, M_m, I) \right] p(\omega, b, \phi \mid D, M_m, I) \\
&= \int d\omega \int db \int d\phi\, \langle r_{j(t)} \mid \omega, b, \phi, m \rangle\, p(\omega, b, \phi \mid D, M_m, I),
\end{aligned} \tag{55}$$

that is, we just average the conditional result we found above over the marginal posterior for ω, b and ϕ (conditional on m), choosing the appropriate value of j in the integral as ω and ϕ vary. $p(\omega, b, \phi \mid D, M_m, I)$ can be evaluated from equations (25), (26), (10), (9), and (8).

$$p(\omega, b, \phi \mid D, M_m, I) = \frac{(\omega b)^{-1} \prod_{j=1}^{m} \{(bW_j)^{-1} F \exp(-\frac{b}{2} \sum_{j=1}^{m} \chi_{Wj}^2)\}}{\int \frac{d\omega}{\omega} \int \frac{db}{b} \int d\phi \prod_{j=1}^{m} \{(bW_j)^{-1} F \exp(-\frac{b}{2} \sum_{j=1}^{m} \chi_{Wj}^2)\}}, \tag{56}$$

where,

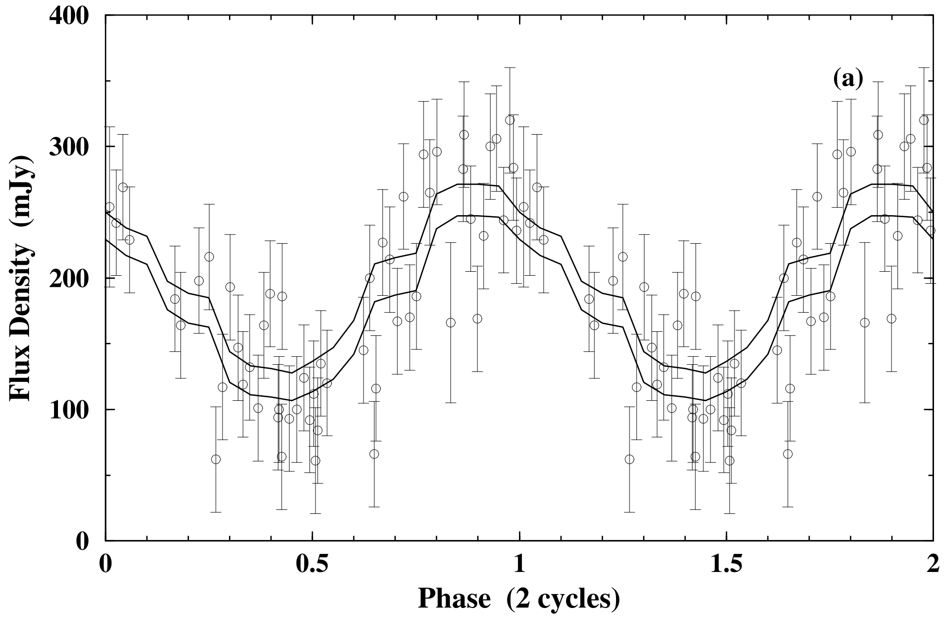
$$F = \left[\mathrm{erfc}(\frac{bW_j}{2}(r_{min} - \overline{d_{Wj}})) - \mathrm{erfc}(\frac{bW_j}{2}(r_{max} - \overline{d_{Wj}})) \right]. \tag{57}$$

Calculation of the second moment proceeds analogously, allowing us to calculate the standard deviation of the estimate of $r(t)$.

Note that the estimate provided by equation (55) is essentially a weighted superposition of stepwise light curves, each with different phase, noise scale factor and frequency , with weights given by the probability densities for various choices of frequency, b and phase. It is thus not a stepwise function, but rather a somewhat smoothed version of a stepwise function.

6.3. Superposing Stepwise Models

All of our light curve estimates so far have been conditional on the choice of a single best (most probable) number of bins. But, as we noted in our derivation of equation (39), Bayesian model comparison leads to a probability distribution for the models, $p(M_m \mid D, I)$; it does not isolate a single model. Thus a more complete estimate of the rate should marginalize over m as well as the various model parameters. We can calculate the marginal posterior mean for $r(t)$ as follows;



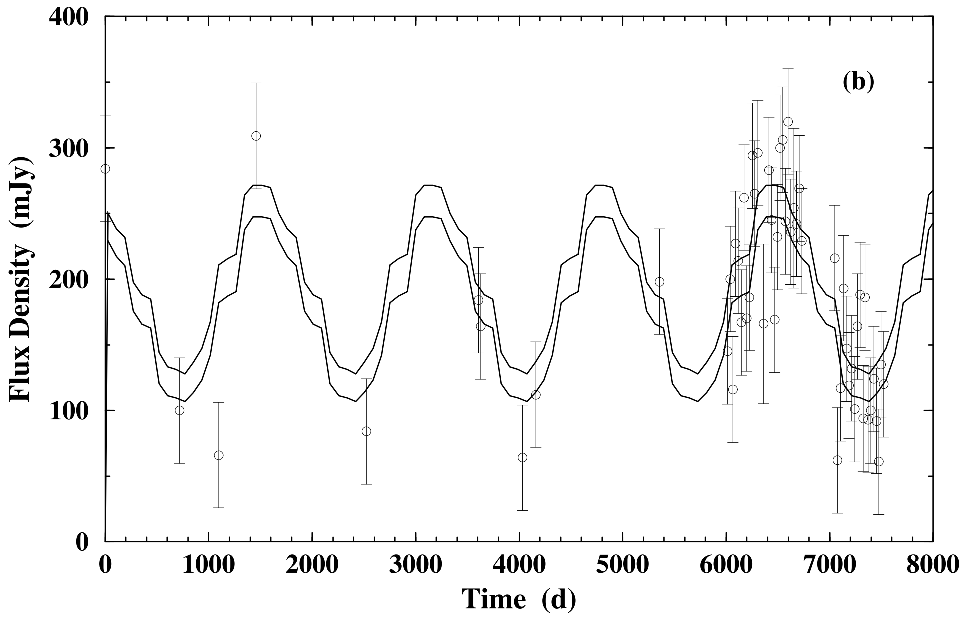


Fig. 3.— Panel (a) shows the shape estimate of the outburst peak flux modulation for LS I +61°303 and measured values plotted for two cycles of phase. The solid curves shown are the mean flux ± 1 one standard deviation. Panel (b) compares the Bayesian flux modulation light curve with the data.

$$\langle r(t) \mid m > 1 \rangle = \sum_{m=2}^{m_{\max}} p(M_m \mid D, I)\ \langle r(t) \mid m \rangle, \tag{58}$$

where $\langle r(t) \mid m \rangle$ is given by equation (55), and $p(M_m \mid D, I)$ can be calculated from equation 4. An analogous equation holds for the second moment, allowing calculation of the standard deviation for $r(t)$. The resulting estimate is even more "smoothed" than that produced by equation (55) because it contains contributions with different numbers of bins. It may still have a significantly "boxy" shape, however.

Figure 3(a) shows the shape estimate for the outburst peak flux modulation for LS I +61°303 plotted for two cycles of phase, derived from equation (58), with phase derived from time assuming the most probable modulation period of 1653 days. The solid curves are the mean flux ± 1 standard deviation. The raw data are also plotted. The error bars on the data points reflect the effective noise in the data after removal of the estimated shape of the flux modulation. They are much larger than the real flux measurement errors which are ≈ 10 mJy. The difference between the effective and real errors represents further structure in the data beyond what can be explained by the present model. In appearance the estimated light curve is a somewhat distorted sine wave.

Figure 3(b) shows a comparison of our computed Bayesian flux modulation light curve with the data versus time. In this plot the zero point of the time axis corresponds to Julian Date 2,443,366.775 by convention.

7. Discussion

It is clear that of the 3 models considered the periodic modulation model, H_P, is the clear winner. Here we consider whether we can distinguish between a periodic and quasi-periodic flux modulations. The plot of the probability of the flux modulation period (Fig. 2(b)) shows a broad maximum with two clear peaks separated by 34 days and a very weak 3rd peak again separated from the middle by 34 days. Recall that in a Bayesian analysis a probability distribution quantifies what we know about the period for our present state of knowledge or information. Even if the modulation were perfectly periodic our state of information (noisy nonuniform sampling with large gaps) might not result in a probability distribution with a single peak. To investigate this further we generated two simulated data sets by sampling the derived periodic flux light curve shown in Figure 3(a) with the same sampling in time as our measured outburst peaks and added Gaussian noise with $\sigma = s_i$. The simulated data sets were processed in exactly the same way as the real data set. The posterior probability of the flux modulation period for the two simulated data sets is shown in Figure 4. One (dashed curve) exhibits a single peak at 1646 days while the second (solid curve) has two peaks at 1587 and 1639 days. This comparison suggests that the multiple peak structure seen in the real data is not the result of the nonuniform data sampling but simply due to the effective noise. In light of this the mean of the probability density function in Figure 2(b) of 1632 days is probably a more robust estimator of the modulation period than the peak value.

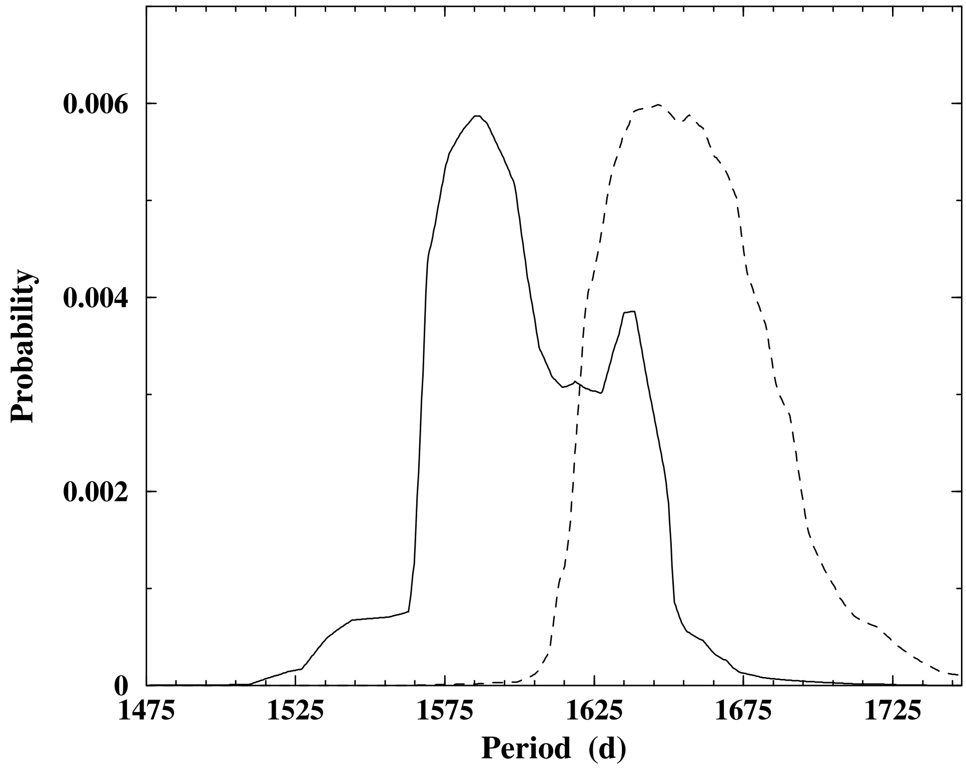


Fig. 4.— The marginal probability of the modulation period for the peak flux for two simulated data sets.

The results from the analysis of these simulated data sets do not argue for the need for variations in the modulation period, but we can't rule it out.

8. Conclusions

In this paper we have developed the equations for the GL method for the detection of a periodic signal of unknown shape and period, for the case where the noise sampling distribution is independent Gaussian. We show how to compute the global likelihood for our stepwise periodic models. The stepwise model, can just as easily represent a non periodic variable light curve or a constant light curve ($m = 1$). Therefore, our calculations can be used to compare both constant and variable non periodic alternatives to periodic models.

Each parameter of a model that is marginalized over, in the process of computing the global likelihood, introduces an Occam's factor penalizing the model for wasted prior parameter space that gets ruled out by the data. Thus the greater the model complexity the more Occam's penalties it must overcome in comparison with a simpler model. The calculation thus balances model simplicity with goodness-of-fit, allowing us to determine both whether there is evidence for a periodic signal and the optimum number of bins for describing the structure in the data.

The method does not require uniform sampling and has the ability to determine the effective noise in the data (everything in the data that cannot be described by the model is treated as noise), and properly allows for this in the computation of the models global likelihood and parameter estimates. In section 5 we illustrated the dangers of simply assuming the measurement errors in a maximum likelihood analysis.

We have applied the method to an interesting time series consisting of over 20 years of radio measurements of LS I +61°303, and confirmed the existence of a periodic or quasi-periodic modulation of the radio outburst peak flux with a mean period of 1632 days with a 68.3% credible region extending from 1599 to 1660 days. We also derived the mean shape of the modulation which is consistent with an distorted sine wave. Currently the system is known to exhibit two periodicities: the 26.5 day periodic outbursts and the 1632 day outburst peak flux modulation. The estimated noise sigma from this analysis of 45 mJy is considerably higher than the measurement uncertainties which implies that there is scope for still more complex models to describe deeper levels of the source's radio properties which are now just beginning to be perceived and which are likely to contribute greatly to our understanding of accretion regimes in binary systems. In particular there is evidence for shorter duration flaring (of order 100 mJy) which is particularly noticeable at times when the peak outburst flux is a minimum, but is apparent at other times as well. It is hoped that GBI monitoring program, operated by the NRAO, will continue the high quality measurements of LS I +61°303 to permit these deeper source properties to be disentangled. In a companion paper (Gregory et al 1998) we address the nature of the outburst phase variations and their relationship to the outburst peak flux.

The author acknowledges helpful discussions with Marta Peracaula in compiling Table 1, the outburst data. During 1994-1996.3 (ref. 7 in Table 1) the Green Bank Interferometer was operated by the NRAO for the Naval Research Laboratory; and during the time period 1996.8-1998 (ref. 8 in Table 1) the instrument was operated by the NRAO in support of the NASA high energy program. The NRAO is operated by Associated Universities, Inc., under contract with the National Science Foundation. This research was supported in part by a grant from the Canadian Natural Sciences and Engineering Research Council at the University of British Columbia.

A. Small Data Sets

We have already indicated that the GL method is effective even for nonuniform sampling. Here we address the question of how small a data set the method can deal with. From the analysis presented in this paper it is clear that the statistic χ^2_{Wj} defined by equation (16) plays a central role. For example equation (25) for $p(D \mid \omega, b, \phi, M_m, I)$, depends very sensitively on the term $\exp(-\frac{b}{2}\sum_{j=1}^{m}\chi^2_{Wj})$. To permit useful inferences about light curve properties, there must be two or more samples in each bin. For uniform sampling this means the sample size must be $> 2\ m_{max}$. Nonuniform samples satisfying this condition may still have less than two samples in one or more bins for certain combinations of ω, ϕ, m. In this case $\chi^2_{Wj} = 0$, assuming we set $W_j = 1$ for the undefined $n_j = 0$ case. Bins with less than 2 samples would thus contribute to an artificial reduction in $\sum_{j=1}^{m}\chi^2_{Wj}$ and a corresponding enhancement in $p(D \mid \omega, b, \phi, M_m, I)$.

A simple solution to this problem is to introduce another weight, w_j, which is equals 0 if there are less than 2 samples in the bin and equals 1 otherwise and,

$$\text{replace} \quad \sum_{j=1}^{m}\chi^2_{Wj} \quad \text{by} \quad \frac{\sum_{j=1}^{m} w_j\ \chi^2_{Wj}}{\frac{1}{m}\ \sum_{j=1}^{m} w_j} \tag{A1}$$

This effectively assigns a value of χ^2_{Wj} to the bins with less than two samples equal to the average value of χ^2_{Wj} of bins with > 2 samples. This eliminates the spurious enhancements in $p(D \mid \omega, b, \phi, M_m, I)$ that would otherwise occur.

REFERENCES

Gregory, P. C. and Taylor, A. R. 1978, Nature, 272, 704

Gregory, P. C., Xu, Huang-Jian, Backhouse, C. J., Reid, A. 1989, ApJ, 339, 1054

Gregory, P. C., and Loredo, T. J. 1992a, in Maximum Entropy and Bayesian Methods, ed. C. R. Smith, G. J. Erickson, and P. O. Neudorfer (Kluwer, Dordrecht), 79

Gregory, P. C., and Loredo, T. J. 1992b, ApJ, 398, 146

Gregory, P. C., and Loredo, T. J. 1993, in Maximum Entropy and Bayesian Methods, Paris, 1992, ed. Mohammad-Djafari and G. Demoment (Kluwer, Dordrecht), 225

Gregory, P. C., and Loredo, T. J. 1996, ApJ, 473, 1059

Gregory, P. C., Peracaula, M., and Taylor, A. R., ApJ(in press Jan. 1999)

Jaynes, E. T. 1968, IEEE Transactions on System Science & Cybernetics, Vol.SSC-4, No. 3, p. 227

Kniffen, D. A. et al. 1997, ApJ, 486, 126

Marti, J. 1993, PhD Thesis, Universitat de Barcelona

Marti, J. and Paredes, J. M. 1995, A&A, 298, 151

Nelson, R. 1989, Ph.D. Thesis, Manchester University

Peracaula, M. 1997, PhD Thesis, Universitat de Barcelona

Paredes, J. M., Estelle, R., Ruis, A. 1990, A&A, 232, 337

Paredes, J.M., Marti, J., Peracaula, M. & Ribo, M. 1997, A&A, 320, L25

Ray, P. S., Foster, R.S., Waltman, E. B. et al. 1997, ApJ, 491, 381

Taylor, A. R. and Gregory, P. C. 1982, ApJ255, 210

Taylor, A. R. and Gregory, P. C. 1984, ApJ283, 273

This preprint was prepared with the AAS LaTeX macros v4.0.

Table 1. Estimated times and flux densities of outbursts.

Time days	Error days	Flux mJy	Error mJy	frequency GHz	Ref. #
43382.94	1.0	284.	33.	5.0	1
43570.00[a]	2.5	138.		10.5	1
43729.99[a]	2.5	180.		5.0	1
44100.97	2.0	100.	33.	5.0	1
44479.93		66.	33.	5.0	1
44840.03	1.0	309.	33.	5.0	2
45907.00	1.0	84.	33.	5.0	3
46990.63	1.0	184.	33.	5.0	4
47013.02	1.0	164.	33.	5.0	4
47416.00	2.0	64.	33.	8.4	5
47544.20	1.0	112.	33.	8.4	5
48738.80	1.0	198.	33.	8.4	6
49395.88	1.0	138.	33.	8.3	7
49421.77	1.0	200.	33.	8.3	7
49446.61	2.0	116.	33.	8.3	7
49472.19	1.0	227.	33.	8.3	7
49500.09	1.0	214.	33.	8.3	7
49530.19	1.0	167.	33.	8.3	7
49555.05	1.0	262.	33.	8.3	7
49580.98	1.0	170.	33.	8.3	7
49607.76	1.0	186.	33.	8.3	7
49634.92	1.0	294.	33.	8.3	7
49659.69	1.0	265.	33.	8.3	7
49689.71	1.0	296.	33.	8.3	7
49744.37	2.0	166.	50.	8.3	7
49794.32	1.0	283.	33.	8.3	7
49823.41	1.0	245.	33.	8.3	7
49850.01	1.0	169.	33.	8.3	7
49875.25	1.0	232.	33.	8.3	7
49902.04	1.0	300.	33.	8.3	7
49926.86	1.0	306.	33.	8.3	7
49954.94	1.0	244.	33.	8.3	7
49980.74	1.0	320.	33.	8.3	7
50006.58	1.0	236.	33.	8.3	7
50034.83	2.0	254.	50.	8.3	7
50060.72	1.0	242.	33.	8.3	7
50088.72	1.0	269.	33.	8.3	7
50114.35	1.0	229.	33.	8.3	7
50431.66	1.0	216.	33.	8.3	8
50458.60		62.	33.	8.3	8
50486.47		117.	33.	8.3	8
50516.55	1.0	193.	33.	8.3	8
50547.31		147.	33.	8.3	8
50568.25	2.0	119.	33.	8.3	8

Table 1—Continued

Time days	Error days	Flux mJy	Error mJy	frequency GHz	Ref. #
50594.12	1.0	132.	33.	8.3	8
50627.13		101.	33.	8.3	8
50650.07	1.0	164.	33.	8.3	8
50675.00	1.0	188.	33.	8.3	8
50707.88		94.	33.	8.3	8
50722.83	1.0	186.	33.	8.3	8
50752.82		93.	33.	8.3	8
50781.63		100.	33.	8.3	8
50810.71		124.	33	8.3	8
50834.64		92.	33.	8.3	8
50856.58		61.	33.	8.3	8
50877.55		135.	33.	8.3	8
50903.40		120.	33.	8.3	8

References. — (1) Taylor and Gregory 1982; (2) Taylor and Gregory 1984; (3) Gregory et al. 1989; (4) Nelson 1989; (5) Paredes et al. 1990; (6) Peracaula and Paredes (Private Communication); (7) Ray et al. 1997; (8) http://info.gb.nrao.edu/gbint/gbidata.html

[a]Estimated time with larger error because only the outburst rise or decay measured.

Table 2. Noise scale parameter, b, and RMS residuals values.

Model	b Value	RMS Residual (mJy)
H_C Constant	0.2	74.
H_P Periodic	0.68	45.
H_{NP} Nonperiodic	0.33	56.