A Bayesian periodogram finds evidence for three planets in HD 11964

P. C. Gregory*†

Physics and Astronomy Department, University of British Columbia, 6224 Agricultural Rd., Vancouver, British Columbia V6T 1Z1, Canada

Accepted 2007 August 7. Received 2007 August 4; in original form 2007 June 19

ABSTRACT
A Bayesian multiplanet Kepler periodogram has been developed for the analysis of precision radial velocity data. The periodogram employs a parallel tempering Markov chain Monte Carlo algorithm. The HD 11964 data have been re-analysed using 1, 2, 3 and 4 planet models. Assuming that all the models are equally probable a priori, the three planet model is found to be ≥600 times more probable than the next most probable model which is a two planet model. The most probable model exhibits three periods of 38.02±0.06, 360±4 and 1924±44 d, and eccentricities of 0.22±0.11, 0.63±0.34 and 0.05±0.03, respectively. Assuming the three signals (each one consistent with a Keplerian orbit) are caused by planets, the corresponding limits on planetary mass (M sin i) and semimajor axis are (0.090±0.015 MJ, 0.253±0.009 au), (0.21±0.07 MJ, 1.13±0.04 au) and (0.77±0.08 MJ, 3.46±0.13 au), respectively. The small difference (1.3σ) between the 360-d period and one year suggests that it might be worth investigating the barycentric correction for the HD 11964 data.

Key words: methods: numerical – methods: statistical – techniques: radial velocities – stars: individual: HD 11964 – planetary systems.

1 INTRODUCTION
Improvements in precision radial velocity (RV) measurements and continued monitoring are permitting the detection of lower amplitude planetary signatures. One example of the fruits of this work is the detection of a super earth in the habitable zone surrounding Gliese 581 (Udry et al. 2007). This and other remarkable successes on the part of the observers are motivating a significant effort to improve the statistical tools for analysing radial velocity data (e.g. Loredo & Chernoff 2003; Cumming 2004; Loredo 2004; Ford 2005; Gregory 2005b; Ford 2006; Ford & Gregory 2006). Much of the recent work has highlighted a Bayesian Markov chain Monte Carlo (MCMC) approach as a way to better understand parameter uncertainties and degeneracies, and to compute model probabilities.

Gregory (2005a,b,c, 2007) presented a Bayesian MCMC algorithm that makes use of parallel tempering (PT) to efficiently explore a large model parameter space starting from a random location. It is able to identify any significant periodic signal component in the data that satisfies Kepler’s laws, and thus functions as a Kepler periodogram. This eliminates the need for a separate periodogram search for trial orbital periods which typically assume a sinusoidal model for the signal that is only correct for a circular orbit. In addition, the Bayesian MCMC algorithm provides full marginal parameters distributions for all the orbital elements that can be determined from radial velocity data. The samples from the parallel chains can also be used to compute the marginal likelihood for a given model (Gregory 2005a) for use in computing the Bayes factor that is needed to compare models with different numbers of planets. The parallel tempering MCMC algorithm employed in this work includes an innovative two stage adaptive control system that automates the selection of efficient Gaussian parameter proposal distributions. The annealing of the proposal distributions carried out by the control system combined with parallel tempering makes it practical to attempt a blind search for multiple planets simultaneously. This was done for the analysis of the current data set and for the analysis of the HD 208487 reported earlier (Gregory 2007).

This paper presents a Bayesian re-analysis of the existing 87 precision radial velocity measurements for HD 11964 published by Butler et al. (2006), who reported the detection of a single planet with a period of 2110 ± 270 d after removing a trend in the data. They remark that the 5.3 m s⁻¹ residuals are comparable to the 9 m s⁻¹ amplitude, placing the exoplanetary interpretation of the velocity variations somewhat in doubt.

2 ANALYSIS
The analysis of the HD 11964 data employed exactly the same Bayesian multiplanet Kepler periodogram that was previously described for the analysis of HD 208487 (Gregory 2007). The periodogram utilizes a parallel tempering Markov chain Monte Carlo algorithm which yields the probability density distribution for each model parameter and permits a direct comparison of the probabilities of models with differing numbers of planets. In parallel tempering, each chain corresponds to a different temperature. We parametrize the temperature by its reciprocal, β = 1/T which varies from 0
to 1. For parameter estimation purposes 12 chains \((\beta = \{0.05, 0.1, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.70, 0.80, 0.90, 1.0\})\) were employed and the final samples drawn from the \(\beta = 1\) chain, which corresponds to the desired target probability distribution. For \(\beta < 1\), the distribution is much flatter.

At intervals, a pair of adjacent chains on the tempering ladder are chosen at random and a proposal made to swap their parameter states. The mean number of iterations between swap proposals was set \(= 8\). A Monte Carlo acceptance rule determines the probability for the proposed swap to occur. This swap allows for an exchange of information across the population of parallel simulations. In the higher temperature simulations radically different configurations can arise, whereas in higher \(\beta\) (lower temperature) states, a configuration is given the chance to refine itself.

The samples from hotter simulations were also used to evaluate the marginal (global) likelihood needed for model selection, following section 12.7 of Gregory (2005a) and Gregory (2007). This is discussed more in Section 4. Marginal likelihoods estimated in this way require many more parallel simulations. For HD 11964, 40-\(\beta\) levels were used spanning the range \(\beta = 10^{-8} - 1.0\) with a mean interval between swaps \(= 3\).

For a one planet model the predicted radial velocity is given by
\[
v(t) = V + K \left[ \cos(\theta(t_i + \chi P) + \omega) + e \cos \omega \right],
\]
and involves the six unknown parameters as follows.

- \(V = \) constant velocity.
- \(K = \) velocity semi-amplitude.
- \(P = \) the orbital period.
- \(e = \) the orbital eccentricity.
- \(\omega = \) the longitude of periastron.
- \(\chi = \) the fraction of an orbit, prior to the start of data taking, that periastron occurred at. Thus, \(\chi P = \) the number of days prior to \(t_i = 0\) that the star was at periastron, for an orbital period of \(P\) days.
- \(\theta(t_i + \chi P)\) is the angle of the star in its orbit relative to periastron at time \(t_i\), also called the true anomaly.

We utilize this form of the equation because we obtain the dependence of \(\theta\) on \(t\) by solving the conservation of angular momentum equation:
\[
\frac{d\theta}{dt} = \frac{2\pi(1 + e \cos \theta(t_i + \chi P))}{P(1 - e^2)^{3/2}} = 0.
\]

Our algorithm is implemented in Mathematica and it proves faster for Mathematica to solve this differential equation than solve the equations relating the true anomaly to the mean anomaly via the eccentric anomaly. Mathematica generates an accurate interpolating function between \(t\) and \(\theta\), so the differential equation does not need to be solved separately for each \(t\). Evaluating the interpolating function for each \(t\) is very fast compared to solving the differential equation, so the algorithm should be able to handle much larger samples of radial velocity data than those currently available without a significant increase in computational time.

As described in more detail in Gregory (2007), we employed a re-parametrization of \(\chi\) and \(\omega\) to improve the MCMC convergence speed motivated by the work of Ford (2006). The two new parameters are \(\psi = 2\pi \chi + \omega\) and \(\phi = 2\pi \chi - \omega\). Parameters \(\psi\) is well determined for all eccentricities. Although \(\phi\) is not well determined for low eccentricities, it is at least orthogonal to the \(\psi\) parameter. In Gregory (2007), we recommended a uniform prior for \(\psi\) in the interval \(0 \leq \psi 

3 RESULTS

Fig. 1 shows the precision radial velocity data for HD 11964 from Butler et al. (2006) who reported a single planet with \(M \sin i = 0.61 \pm 0.10\) in a \(2110 \pm 270\) day orbit with an eccentricity of \(0.06 \pm 0.17\). Panels (b) and (c) show our best-fitting three planet light curve and residuals.

If we assume that all the models considered are equally probable a priori, then as shown in Section 4, the three planet model is

\footnote{In the absence of detailed knowledge of the sampling distribution for the extra noise, we pick a Gaussian because, for any given finite noise variance, it is the distribution with the largest uncertainty as measured by the entropy, that is, the maximum entropy distribution (Jaynes 1957; Gregory 2005a, section 8.7.4.).}
Table 1. Prior parameter probability distributions.

| Parameter | Prior | Lower bound | Upper bound |
|-----------|-------|-------------|-------------|
| Orbital frequency | $p(\ln f_1, \ln f_2, \ldots, \ln f_n | M_n, I) = \frac{n!}{(n/t+1)!}$ | 1/1.01 d | 1/1000 yr |
| Velocity $K_i$ (m s$^{-1}$) | Modified Jeffreys $^a$ | $0 (K_i = 1)$ | $K_{\text{max}} \left( \frac{3}{2\pi} \right)^{1/3} \frac{1}{\sqrt{1-\varepsilon^2}}$ |
| $V$ (m s$^{-1}$) | $-K_{\text{max}}$ | $K_{\text{max}}$ |
| $e_i$ eccentricity | Uniform | 0 | 1 |
| $\omega_i$ longitude of periastron | Uniform | 0 | $2\pi$ |
| $s$ extra noise (m s$^{-1}$) | $\frac{(s+s_0)}{s_0}$ | 0 ($s_0 = 1$) | $K_{\text{max}}$ |

$^a$Since the prior lower limits for $K$ and $s$ include zero, we used a modified Jeffreys prior of the form

$$p(X | M, I) = \frac{1}{X + X_0} \frac{1}{\ln(1 + \frac{X_{\text{max}}}{X_0})}$$

For $X < X_0, p(X | M, I)$ behaves like a uniform prior and for $X \gg X_0$ it behaves like a Jeffreys prior. The $\ln(1 + \frac{X_{\text{max}}}{X_0})$ term in the denominator ensures that the prior is normalized in the interval 0 to $X_{\text{max}}$.

$\geq 600$ times more probable than the next most probable model which is a two planet model. In this section, we mainly focus on the MCMC results for the three planet model.

Fig. 2 shows post-burn-in MCMC iterations for the parameters of a three planet model returned by the Kepler periodogram, starting from an initial location ($P_1 = 10, P_2 = 500$ and $P_3 = 2300$ d) in period parameter space. Similar results were obtained with other different starting positions. A total $10^6$ iterations were used with every tenth iteration stored. For display purposes only every hundredth stored point is plotted in the figure. The upper left-hand panel is a plot of log$_{10}$ (prior \times likelihood). The next two panels of the top row show the extra noise parameter $s$ and the constant velocity parameter $V$, respectively. The remaining panels show the orbital parameters for each of the three periods. The equilibrium solution corresponds to $P_1 = 38$ d, $P_2 = 360$ d and $P_3 = 1924$ d (for comparison, the two planet model MCMC yielded two solutions of $P_1 = 360$ and $P_2 = 1990$ d, and $P_2 = 38$ and $P_3 = 1932$ d, respectively). All the traces appear to have achieved an equilibrium distribution. The $X_i$ and $\omega_i$ traces were derived from the corresponding $\psi_i, \phi_i$ traces. The $\psi_i, \phi_i$ traces are not shown.

The Gelman & Rubin (1992) statistic is typically used to test for convergence of the parameter distributions. In parallel tempering MCMC, new widely separated parameter values are passed up the line to the $\beta = 1$ simulation, and are occasionally accepted. Roughly every 100 iterations, the $\beta = 1$ simulation accepts a swap proposal from its neighbouring simulation. The final $\beta = 1$ simulation is thus an average of a very large number of independent $\beta = 1$ simulations. What we have done is divide the $\beta = 1$ iterations into 10 equal time intervals and intercompared the 10 differently independent average distributions for each parameter using a Gelman–Rubin test. For all of the three planet model parameters, the Gelman–Rubin statistic was $\leq 1.03$.

Fig. 3 shows the individual parameter marginal distributions for the three planet model. A correlation is clearly evident between $P_2$ and $e_2$, which is best seen in the joint marginals plotted in Fig. 4. Each dot is the result from one iteration. Table 2 gives our Bayesian three planet orbital parameter values and their errors. The parameter value listed is the median of the marginal probability distribution for the parameter in question and the error bars identify the boundaries of the 68.3 per cent credible region. The value immediately below in parentheses is the maximum a posteriori (MAP) value determined using the Nelder–Mead (1965) downhill simplex method. The values derived for the semimajor axis and $M \sin i$, and their errors, are based on the assumed mass of the star = 1.49 $\pm$ 0.15 $M_\odot$ (Valenti & Fischer 2005). Butler et al. (2006) assumed a mass of $= 1.12$ $M_\odot$ but also quote Valenti & Fischer (2005) as the reference. The last row gives the Bayesian estimate of the extra noise parameter (stellar jitter) for each model.

In Fig. 5, panel (a) shows the data, with the best-fitting $P_2$ and $P_3$ orbits subtracted, for two cycles of $P_1$ phase with the best-fitting $P_1$ orbit overlaid. Panel (b) shows the data plotted versus $P_2$ phase with the best-fitting $P_1$ and $P_3$ orbits removed. Panel (c) shows the data plotted versus $P_3$ phase with the best-fitting $P_1$ and $P_2$ orbits removed.

4 MODEL SELECTION

To compare the posterior probabilities of the $ith$ planet model to the one planet models, we need to evaluate the odds ratio, $O_{ij} = p(M_j | D, I) / p(M_i | D, I)$, the ratio of the posterior probability of model $M_j$ to model $M_i$. Application of Bayes’s theorem leads to

$$O_{ij} = \frac{p(M_j | I) p(D | M_j, I)}{p(M_i | I) p(D | M_i, I)} = \frac{p(M_j | I)}{p(M_i | I)} B_{ij},$$

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 marginal (global) likelihoods of the models. The MCMC algorithm produces
samples which are in proportion to the posterior probability distribution which is fine for parameter estimation but one needs the proportionality constant for estimating the model marginal likelihood. Clyde (2006) recently reviewed the state of techniques for model selection from a statistics perspective, and Ford & Gregory (2006) have evaluated the performance of a variety of marginal likelihood estimators in the extrasolar planet context.

In this work, we will compare the results from three marginal likelihood estimators: (i) parallel tempering, (ii) ratio estimator and (iii) restricted Monte Carlo. A brief outline of each method is presented in Sections 4.1, 4.2 and 4.3. The results are summarized in Section 4.4.

4.1 Parallel tempering estimator

The MCMC samples from all \((n_p)\) simulations can be used to calculate the marginal likelihood of a model according to equation (5) Gregory (2005a)

\[
\ln[p(D|M_i, I)] = \int d\beta \ln[p(D|M_i, X, I)]_\beta, \tag{5}
\]

where \(i = 0, 1, 2, 3, 4\) corresponds to the number of planets and \(X\) represent a vector of the model parameters which includes the extra Gaussian noise parameter \(s\). In words, for each of the \(n_p\) parallel simulations, compute the expectation value (average) of the natural logarithm of the likelihood for post-burn-in MCMC samples. It is necessary to use a sufficient number of tempering levels that we can estimate the above integral by interpolating values of

\[
\ln[p(D|M_i, X, I)]_\beta = \frac{1}{n} \sum \ln[p(D|M_i, X, I)]_\beta, \tag{6}
\]

in the interval from \(\beta = 0\) to 1, from the finite set. For this problem, we used 40 tempering levels in the range \(\beta = 10^{-3} \rightarrow 1.0\). Fig. 6 shows a plot of \(\ln[p(D|M_i, X, I)]_\beta\) versus \(\beta\). The inset shows a blow-up of the range \(\beta = 0.1 \rightarrow 1.0\).

The relative importance of different decades of \(\beta\) can be judged from Table 3. The second column gives the fractional error that would result if this decade of \(\beta\) was not included, and thus indicates the sensitivity of the result to that decade. The fractional error falls rapidly with each decade and for the lowest decade explored in this run, \(\beta = 10^{-6} \rightarrow 10^{-7}\), reaches 0.21. From Fig. 6, it is apparent that the steep drop in the curve that occurs below \(\beta = 10^{-6}\) shows a significant change in curvature in the direction of levelling off similar to that experienced in the case of HD 208784 (Gregory 2007) and HD 188133 (Ford & Gregory 2006). In the case of HD 208487 (two planet model), the fractional error reached 0.16 in the range \(\beta = 10^{-6} \rightarrow 10^{-7}\) and the error fell to 0.02 in the next decade. For HD 188133 (one planet model), the fractional error reached 0.26 in the range \(\beta = 10^{-5} \rightarrow 10^{-6}\) and the contribution to the fractional error for the next four decades was 0.14. Based on these comparisons, we estimate that ignoring lower decades of \(\beta\) would result in a systematic underestimate of \(p(D|M_i, I)\) of \(\sim 15\) per cent. A similar table for the two planet PT results gave the fractional error for the lowest decade at 0.11.

4.2 Marginal likelihood ratio estimator

Our second method\(^3\) was introduced by Ford & Gregory (2006). It makes use of an additional sampling distribution \(h(X)\). Our starting point is Bayes’ theorem

\[
p(X \mid M_i, I) = \frac{p(X \mid M_i, I)p(D \mid M_i, X, I)}{p(D \mid M_i, I)}. \tag{7}
\]

Re-arranging the terms and multiplying both sides by \(h(X)\), we obtain

\[
p(D \mid M_i, I)p(X \mid M_i, I)h(X) = p(X \mid M_i, I)p(D \mid M_i, X, I)h(X). \tag{8}
\]

Integrate both sides over the prior range for \(X\):

\[
p(D|M_i, I)_{RE} \int p(X \mid M_i, I)h(X) dX = \int p(X \mid M_i, I)p(D \mid M_i, X, I)h(X) dX. \tag{9}
\]

The ratio estimator of the marginal likelihood, which we designate by \(p(D|M_i, I)_{RE}\), is given by

\[
p(D \mid M_i, I)_{RE} = \frac{\int p(X \mid M_i, I)p(D \mid M_i, X, I)h(X) dX}{\int p(X \mid M_i, I)h(X) dX}. \tag{10}
\]

To obtain the marginal likelihood ratio estimator, \(p(D \mid M_i, I)_{RE}\), we approximate the numerator by drawing samples \(X^1, X^2, \ldots, X^{n_p}\).

\(^3\) Initially proposed by J. Berger, at an Exoplanet Workshop sponsored by the Statistical and Applied Mathematical Sciences Institute in 2006 January.

© 2007 The Author. Journal compilation © 2007 RAS, MNRAS 381, 1607–1616
According to Ford & Gregory (2006), we improve over the single multinormal by using a mixture of multivariate normals by setting 

\[ h(X) = \frac{1}{n_c} \sum_{j=1}^{n_c} h_j(X), \]

where we must determine a covariance matrix for each \( h_j(X) \) using the posterior sample. We choose each mixture component to be a

\footnote{According to Ford & Gregory (2006), the numerator converges more rapidly than the denominator.}

valid samples. Of course, we need to use the same truncated multinormal in the denominator of equation (10) so the normalization factor cancels. \( p(D | M_2, I)_{PT} \) converges much more rapidly than the parallel tempering estimator Gregory (2007) and the parallel tempering estimator, \( p(D | M_2, I)_{PT} \), required 40 \( \beta \) simulations instead of one.

\subsection*{4.2.1 Mixture model}

It is clear that a single multinormal distribution cannot be expected to do a very good job of representing the correlation between the parameters that is evident between \( P_2 \) and \( \epsilon_2 \) in Fig. 4. Following Ford & Gregory (2006), we improve over the single multinormal by using a mixture of multivariate normals by setting

\[ h(X) = \frac{1}{n_c} \sum_{j=1}^{n_c} h_j(X), \]

and approximate the denominator by drawing samples from a truncated multinormal. The factor required to normalize the truncated multinormal is just the ratio of the total number of samples from the full multinormal to the number of physical samples. Of course, we need to use the same truncated multinormal in the denominator of equation (10) so the normalization factor cancels. \( p(D | M_2, I)_{PT} \) converges much more rapidly than the parallel tempering estimator Gregory (2007) and the parallel tempering estimator, \( p(D | M_2, I)_{PT} \), required 40 \( \beta \) simulations instead of one.
multivariate normal distribution, \( h_j(X) = N(X | \mu_j, \Sigma_j) \), where we must determine a covariance matrix for each \( h_j \) using the posterior sample. First, we compute \( \rho \), defined to be a vector of the sample standard deviations for each of the components of \( X \), using the posterior sample. Next, define the distance between the posterior sample \( X_i \) and the centre of \( h_j(X) \), \( d_{ij}^2 = \sum_k (X_{ki} - X_{kj})^2 / \rho_k^2 \), where \( k \) indicates the element of \( X \) and \( \rho \). Now, draw another random subset of 100 \( n_\ell \) samples from the original posterior sample (without replacement), select the 100 posterior samples closest to each mixture component and use them to calculate the covariance matrix, \( \Sigma_j \), for each mixture component. Since the posterior sample is assumed to have fully explored the posterior, \( h(X) \) should be quite similar to the posterior in all regions of significant probability, provided that we use enough mixture components.

Figure 3. Marginal parameter probability distributions for the three planet model.
Table 2. Three planet model parameter estimates.

| Parameter | Planet 1 | Planet 2 | Planet 3 |
|-----------|---------|---------|---------|
| \( P \) (d) | 38.02^{+0.6}_{-0.5} | 360^{+5}_{-4} | 1925^{+44}_{-44} |
|             | (38.07) | (357)   | (1928)  |
| \( K \) (m s\(^{-1}\)) | 4.3^{+0.7}_{-0.3} | 6.1^{+3.0}_{-3.3} | 9.7^{+0.8}_{-0.8} |
|             | (4.8)   | (5.4)   | (10.0)  |
| \( e \)   | 0.23^{+0.10}_{-0.07} | 0.63^{+0.35}_{-0.33} | 0.05^{+0.03}_{-0.03} |
|             | (0.31)  | (0.63)  | (0.09)  |
| \( \omega \) (deg) | 123^{+4.1}_{-2.2} | 103^{+4.8}_{-4.4} | 195^{+8.8}_{-7.4} |
|             | (111)   | (107)   | (205)   |
| \( a \) (au) | 0.252^{+0.0085}_{-0.0085} | 1.132^{+0.039}_{-0.039} | 3.46^{+0.13}_{-0.13} |
|             | (0.253) | (1.124) | (3.46)  |
| \( \sin i \) (\( M_1 \)) | 0.090^{+0.014}_{-0.015} | 0.213^{+0.058}_{-0.067} | 0.77^{+0.08}_{-0.08} |
|             | (0.098) | (0.191) | (0.795) |
| Periastron passage (JD - 244 0000) | 12 737^{+6}_{-3} | 12 397^{+35}_{-32} | 10 535^{+401}_{-341} |
|             | (12 736) | (12 421) | (10 564) |
| \( s \) (m s\(^{-1}\)) | 4.9^{+0.5}_{-0.2} | 3.7^{+0.4}_{-0.4} | 2.4^{+0.4}_{-0.8} |
|             | (4.7)   | (3.3)   | (1.9)   |

4.3 Restricted Monte Carlo marginal likelihood estimate

We can also make use of Monte Carlo integration to evaluate the marginal likelihood as given by equation (13).

\[
p(D \mid M_1, I) = \int p(X \mid M_1, I)p(D \mid M_1, X, I)\,dX. \quad (13)
\]

Monte Carlo (MC) integration can be very inefficient in exploring the whole prior parameter range, but once we have established the significant regions of parameter space with the MCMC results, this is no longer the case. The outer borders of the MCMC marginal parameter distributions were used to delineate the boundaries of the volume of parameter space to be used in the Monte Carlo integration. RMC integration was carried out for models \( M_1, M_2 \) and \( M_3 \) based on \( 4 \times 10^6 \) samples and repeated three times.

4.4 Summary of model selection results

Table 4 summarizes the marginal likelihoods and Bayes factors comparing models \( M_0, M_2, M_3, M_4 \) to \( M_1 \). For model \( M_0 \), the marginal
likelihood was obtained by numerical integration. For $M_1$, the value and error estimate are based on the RMC method discussed in Section 4.3, the ratio estimator (RE) method (one mixture component), discussed in Section 4.2 and the RE method (100 mixture components). Each method was repeated three times on the same posterior sample to ascertain the variance of repeated trials. The quoted uncertainty is the standard deviation of the repeats. The sample error of the mean is a factor of $1/\sqrt{3}$ smaller. Since all three methods yield approximate marginal likelihoods, it is not clear which is the most accurate but we are inclined to favour the RE method.

| $\beta$ range | Fractional error |
|---------------|------------------|
| $10^{-1}$     | $9.59 \times 10^{10}$ |
| $10^{-1} - 10^{-2}$ | $4.55 \times 10^{13}$ |
| $10^{-2} - 10^{-3}$ | $225$ |
| $10^{-3} - 10^{-4}$ | $3.46$ |
| $10^{-4} - 10^{-5}$ | $1.32$ |
| $10^{-5} - 10^{-6}$ | $0.77$ |
| $10^{-6} - 10^{-7}$ | $0.51$ |
| $10^{-7} - 10^{-8}$ | $0.21$ |

The marginal likelihood for $M_2$ was estimated from the posterior samples from the $\beta = 1$ chain using the RE and the PT method which makes use of the samples from 40 tempering chains. The results for the 100 and 500 mixture components are in good agreement and are a factor of $\sim 2$ less than the one component RE results. It is to be expected that the multiple mixture component versions with 100 mixture components. All three estimates agree within 15 per cent.

For model $M_2$, two peaks in the joint posterior probability distribution were detected: (i) $P_1 = 362$ d, $P_2 = 1984$ d and (ii) $P_1 = 37.98$ d, $P_2 = 1897$ d. The contribution to the marginal likelihood from each peak was estimated from the posterior samples from the $\beta = 1$ chain after filtering the posterior samples in $P_1$ and $P_2$ to exclude samples from the other peak. The ratio estimator method was employed with three different mixture components 1, 100 and 500. For each peak, the results agreed well within a factor of better than 2. The 100 and 500 mixture components agreed more closely and appeared to be systematically lower than for the one component version. On the basis of these results, peak A is a factor of $\sim 10$ more probable. Combining the RE, 500 component results for both peaks yields a $p(D|M_s, I) = 2.3 \times 10^{-124}$ which is close to the $3.0 \times 10^{-124}$ value derived from the parallel tempering method which was discussed in Section 4.1. Our final estimate is $(2.5 \pm 0.5) \times 10^{-124}$.

The marginal likelihood estimates, Bayes factors and probabilities for the five models. The last two columns list the MAP value of extra noise parameter, $s$, and the rms residual.

| Model | Method | Mixture components | Marginal Likelihood | Bayes factor nominal | Probability nominal | $s$ (m s$^{-1}$) | rms residual (m s$^{-1}$) |
|-------|--------|-------------------|---------------------|---------------------|-------------------|----------------|-------------------------|
| $M_0$ | Exact  |                    | $6.86 \times 10^{-138}$ | $2.7 \times 10^{-10}$ | $2.5 \times 10^{-18}$ | 7.6            | 8.0                     |
| $M_1$ | RMC    |                    | $(2.31 \pm 0.01) \times 10^{-128}$ |                     |                   |                |                         |
|       | RE     | 1                  | $(2.86 \pm 0.07) \times 10^{-128}$ |                     |                   |                |                         |
|       | RE     | 100               | $(2.50 \pm 0.06) \times 10^{-128}$ |                     |                   |                |                         |
|       | Summary|                    | $(2.50 \pm 0.4) \times 10^{-128}$ | 1.0                 | $9 \times 10^{-9}$ | 4.7            | 5.3                     |
| $M_{2A}$ | RMC   |                    | $(1.5 \pm 0.2) \times 10^{-124}$ |                     |                   |                |                         |
| $M_{2A}$ | RE    | 1                  | $(3.0 \pm 0.16) \times 10^{-124}$ |                     |                   |                |                         |
| $M_{2A}$ | RE    | 100               | $(2.08 \pm 0.06) \times 10^{-124}$ |                     |                   |                |                         |
| $M_{2A}$ | RE    | 500               | $(1.98 \pm 0.08) \times 10^{-124}$ |                     |                   |                |                         |
| $M_{2B}$ | RMC   |                    | $(3.3 \pm 0.2) \times 10^{-125}$ |                     |                   |                |                         |
| $M_{2B}$ | RE    | 1                  | $(3.86 \pm 0.26) \times 10^{-125}$ |                     |                   |                |                         |
| $M_{2B}$ | RE    | 100               | $(3.28 \pm 0.15) \times 10^{-125}$ |                     |                   |                |                         |
| $M_{2B}$ | RE    | 500               | $(2.95 \pm 0.18) \times 10^{-125}$ |                     |                   |                |                         |
| $M_2$ | RE (A+B) | 500 | $(2.3 \pm 0.08) \times 10^{-124}$ |                     |                   |                |                         |
| $M_2$ | RMC (A+B) |          | $(1.8 \pm 0.3) \times 10^{-124}$ |                     |                   |                |                         |
| $M_2$ | PT | $(2^{0.5}) \times 10^{-124}$ |                          |                     |                   |                |                         |
| $M_2$ | Summary|                    | $(2.5 \pm 0.5) \times 10^{-124}$ | $1.0 \times 10^4$ | $9 \times 10^{-5}$ | 3.3            | 4.1                     |
| $M_3$ | RMC    |                    | $(1.8 \pm 0.3) \times 10^{-121}$ |                     |                   |                |                         |
| $M_3$ | RE     | 1                  | $(14 \pm 3) \times 10^{-119}$ |                     |                   |                |                         |
| $M_3$ | RE     | 100               | $(4.95 \pm 0.44) \times 10^{-119}$ |                     |                   |                |                         |
| $M_3$ | RE     | 500               | $(5.00 \pm 0.44) \times 10^{-119}$ |                     |                   |                |                         |
| $M_3$ | PT     |                    | $2.8 \times 10^{-120}$ |                     |                   |                |                         |
| $M_3$ | Summary|                    | $(2.8^{18}_{1/10}) \times 10^{-120}$ | $1.1 \times 10^8$ | 0.99991         | 1.9            | 3.0                     |
| $M_4$ | RE     | 100               | $3.2 \times 10^{-126}$ |                     |                   |                |                         |
| $M_4$ | RE     | 100               | $1.5 \times 10^{-125}$ |                     |                   |                |                         |
| $M_4$ | RE     | 100               | $3.7 \times 10^{-125}$ |                     |                   |                |                         |
| $M_4$ | RE     | 100               | $1.0 \times 10^{-124}$ |                     |                   |                |                         |
| $M_4$ | Summary|                    | $\leq 1.9 \times 10^{-125}$ | $\leq 760$ | $M_4$ excluded | 1.5            | 2.5                     |
The thin solid black curves show three repeats of the parallel tempering marginal likelihood method versus iteration number for the three planet model. The dashed curves show six repeats of the ratio estimator method using only one mixture component. The thick black curves show the result for four repeats of the ratio estimator method using 100 mixture components and the grey curves correspond to trials using 500 mixture components.

**Figure 7.** The thin solid black curves show three repeats of the parallel tempering marginal likelihood method versus iteration number for the three planet model. The dashed curves show six repeats of the ratio estimator method using only one mixture component. The thick black curves show the result for four repeats of the ratio estimator method using 100 mixture components and the grey curves correspond to trials using 500 mixture components.

Nominal model probabilities excluding model $M_s$ are given in Column 6. The results overwhelmingly favour the three planet model.

Columns 7 and 8 list the most probable values of the extra noise parameter, $s$, and the rms residuals in m s$^{-1}$, respectively.

### 5 Discussion

In this paper, we have demonstrated that a sophisticated Bayesian analysis of the published data for HD 11964 finds strong evidence for two additional planets. Is it likely that there are many other cases among the 200 published RV data sets that this type of analysis would yield evidence for additional planets, or is the HD 11964 system likely to be unique or rare? To date, the Bayesian MCMC Kepler periodogram has been run on only a small number of data sets including HD 73526 Gregory (2005b) and HD 208487 Gregory (2007), which both yielded evidence for an additional planet. It thus appears likely that the algorithm is capable of detecting many additional exoplanet candidates in the published RV data. Although the current implementation of the algorithm is not particular fast (19 h for a typical three planet model run of 10$^6$ iterations with 12 tempering chains), it has many advantages that were outlined in Section 1.

One source of error in the measured velocities is ‘jitter’, which is due in part to flows and inhomogeneities on the stellar surface. Wright (2005) gives a model that estimates, to within a factor of roughly 2 (Butler et al. 2006), the jitter for a star based upon a star’s activity, colour, Teff and height above the main sequence. For HD 11964, Butler et al. (2006) quote a jitter estimate of 5.7 m s$^{-1}$, based on Wright’s model. Our models $M_0$ to $M_4$ employ instead an extra Gaussian noise nuisance parameter, $s$, with a prior upper bound of equal to $K_{max} = 2129$ m s$^{-1}$. Anything that cannot be explained by the model and published measurement uncertainties (which do not include jitter) contributes to the extra noise term. Of course, if we are interested in what the data have to say about the size of the extra noise term, then we can readily compute the marginal posterior for $s$. The marginal for $s$ for $M_1$ is shown in the middle panel of the sixth row in Fig. 3. The marginal for $s$ shows a pronounced peak with a median of 2.4 m s$^{-1}$ and a MAP value of 1.9 m s$^{-1}$. The MAP value will do a better job of modelling correlated parameters than a single component model.

Fig. 7 shows the behaviour of the PT marginal likelihood estimator when it is computed using different numbers of iterations taken from a particular run. The results for the three planet model, using three such MCMC runs, are shown by the thin black curves. The iteration number indicated on the abscissa is the total number of iterations executed but only a fraction (typically every tenth or less) were saved and used for this analysis. For comparison, the results from repeated trials of the RE marginal likelihood estimates versus iteration are shown. The dashed curves are RE using one component. The thick black curves are RE with 100 mixture components and the grey curves correspond to 500 components. It is apparent that two of the PT runs have not yet converged. The third appears to have levelled off at a value of $p(D \mid M_1, I) = 2.8 \times 10^{-12}$ but, of course, more iterations would be desirable. All of the PT results argue for a value significantly less than for the RE method. Finally, the result of two RMC trials is even lower at 1.8 $\times$ $10^{-12}$. It is a difficult question to decide which estimate is best. As a summary, we have quoted the PT result with errors that span the other two methods. In spite of the large uncertainty, the evidence favouring the three planet model is very strong. Assuming equal model priors, then from the lower limit for $p(D \mid M_1, I)$ of $1.8 \times 10^{-12}$ and the upper limit for $p(D \mid M_2, I)$ of $3.0 \times 10^{-12}$, we conclude that for a fair bet the odds in favour of $M_3$ over $M_2$ are $\geq 600$. Similarly, the odds in favour of $M_3$ over $M_1$ are $\geq 6 \times 10^6$.

Table 4 also gives an estimate for $p(D \mid M_4, I)$ based on four repeats of the ratio estimator with 100 components. Because of the large spread in results, our summary is the geometric mean of the individual values. In view of the results for the three planet model, we consider the RE estimates for $p(D \mid M_4, I)$ as an upper limit.

Column 5 of Table 4 gives the nominal Bayes factor comparing each model to the one planet model. Assuming equal model priors, the probability of model $M_i$ is given by

$$p(M_i \mid D, I) = \frac{p(D \mid M_i, I)}{\sum_{j=0}^4 p(D \mid M_j, I)}$$

(14)
of \( s \) for all our models is tabulated in Table 4. For \( M_1 \), the MAP value is \( 4.9 \text{ m s}^{-1} \) which is well within the factor of 2 uncertainty of the jitter estimate given in Butler et al. (2006) based on Wright’s model. The results of our Bayesian model selection analysis indicate that a three planet model is \( \geq 6 \times 10^8 \) times more probable than a one planet model with the previously estimated jitter.

It is interesting to compare the performance of the three marginal likelihood estimators employed in this work to their performance in the two planet fit for HD 208487 (Gregory 2007). For HD 208487, the parallel tempering estimator, based on 34 chains, required \( \sim 1.5 \times 10^6 \) iterations for convergence. The two separate runs agreed within a factor of 2.2. The average of the two HD 208487 PT results agreed with the RMC and RE (one component) within 20 per cent. For the two planet fit of HD 11964 with 40 chains, convergence required \( 5 \times 10^6 \) iterations. Since there were two peaks in the posterior, the RE and RMC had to be run on each peak separately and the two peak contributions added before comparing to the PT result which integrates over the entire posterior. The single PT run agreed with the RMC and RE estimates within a factor of \( \sim 2 \). For both HD208487 and HD 11964, the RMC and the RE results for a two planet model agreed within \( \sim 25 \) per cent.

For HD 11964, the model \( M_3 \) results from the three methods spanned a much larger range. Further, it is clear that two of the three PT runs have not converged, in one case after \( 6 \times 10^6 \) iterations with 40 tempering chains which took 16 d on a fast single core PC. This experience suggests that it may not be feasible to compute parallel tempering marginal likelihoods for models involving three or more planets, which will typically require \( \geq 40 \) chains. Parallel computing could help but there is still a need for more efficient, accurate and well-calibrated methods for computing the marginal likelihoods. MCMC is great for parameter estimation, so perhaps more effort is required to include the number of planets as an additional parameter as has been done in other areas.

### 6 CONCLUSIONS

In this paper, we further demonstrated the capabilities of an automated Bayesian parallel tempering MCMC approach to the analysis of precision radial velocities. The method is called a Bayesian Kepler periodogram because it is ideally suited for detecting signals that are consistent with Kepler’s laws. However, it is more than a periodogram because it also provides full marginal posterior distributions for all the orbital parameters that can be extracted from radial velocity data. Moreover, it is a very general algorithm that can be applied to many other non-linear model fitting problems.

The HD 11964 data (Butler et al. 2006) have been re-analysed using 1, 2, 3 and 4 planet models. The most probable model exhibits three periods of 38.02\(+0.06\)\,-0.05\) days and 192.4\(+4.4\),\,-4.3 days, and eccentricities of 0.29\,+0.11\),\,-0.22\), 0.63\,+0.34\),\,-0.17\), and 0.05\,+0.03\),\,-0.05\) respectively. Assuming the three signals (each one consistent with a Keplerian orbit) are caused by planets, the corresponding limits on planetary mass (\( M \sin i \)) and semimajor axis are

\[
(0.90\,+0.15\,-0.14\)\), \( 0.25\,+0.09\),\,-0.06\) au), \( (0.21\,+0.06\),\,-0.07\) \( M_1 \), \( 1.13\,+0.04\),\,-0.04\) au) \]

and \( (0.77\,+0.08\),\,-0.08\) \( M_1 \), \( 3.46\,+0.13\),\,-0.13\) au), respectively. Based on our three planet model results, the remaining unaccounted for stellar jitter is \( \sim 1.9 \text{ m s}^{-1} \). The small difference (1.3\( \sigma \)) between the 360-d period and one year raise some concern about a possible instrumental effect and we suggest that it might be worth investigating the barycentric correction for the HD 11964 data.

Considerable attention was paid to the topic of Bayesian model selection. For model fitting involving \( \leq 2 \) planets, all three marginal likelihood estimators were in good agreement. For a three planet fit, the RMC and RE results differed by a factor of \( \sim 300 \) and each differed from the PT result by a factor of \( \sim 17 \). Further improvements on the model selection side of this problem are clearly needed, requiring the development of more efficient, accurate and well-calibrated methods for computing the marginal likelihoods.

### ACKNOWLEDGMENTS

The author would like to thank John Skilling for suggesting an improvement to the sampling during the re-parametrization stage of the algorithm. This research was supported in part by grants from the Canadian Natural Sciences and Engineering Research Council at the University of British Columbia.

### REFERENCES

Brethorst G. L., 1988, Bayesian Spectrum Analysis and Parameter Estimation, Springer-Verlag, New York
Butler R. P. et al., 2006, ApJ, 646, 505
Cumming A., 2004, MNRAS, 354, 1165
Clyde M., 2006, in Babu G. J., Feigelson E. D., eds, Statistical Challenges in Modern Astronomy IV. Astron. Soc. Pac., San Francisco, in press
Ford E. B., 2005, AJ, 129, 1706
Ford E. B., 2006, ApJ, 620, 481
Ford E. B., Gregory P. C., 2006, in Babu G. J., Feigelson E. D., eds, Statistical Challenges in Modern Astronomy IV. Astron. Soc. Pac., San Francisco, in press (astro-ph/0608328)
Gelman A., Rubin D. B., 1992, Statist. Sci., 7, 457
Gregory P. C., 2005a, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica Support, Cambridge Univ. Press, Cambridge
Gregory P. C., 2005b, ApJ, 631, 1198
Gregory P. C., 2005c, in Abbas A. E., Morris R. D., Castle J. P., eds, AIP Conf. Proc. Vol. 803, Bayesian Inference and Maximum Entropy Methods in Science and Engineering. Am. Inst. Phys., New York, p. 139
Gregory P. C., 2007, MNRAS, 374, 1321
Jaynes E. T., 1957, Stanford University Microwave Laboratory Report 421 (Reprinted: 1988, in Erickson G. J., Smith C. R., eds, Maximum Entropy and Bayesian Methods in Science and Engineering. Kluwer Academic Press, Dordrecht, p. 1)
Jaynes E. T., 1987, in Smith C. R., Erickson G. L., Reidel D., eds, Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems. Reidel, Dordrecht, p. 1
Loredo T., 2004, in Erickson G. J., Zhai Y., eds, AIP Conf. Proc. Vol. 707, Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 23rd International Workshop. Am. Inst. Phys., New York, p. 330
Loredo T. L., Chernoff D., 2003, in Feigelson E. D., Babu G. J., eds, Statistical Challenges in Modern Astronomy III, Springer, New York, p. 57
Nelder J. A., Mead R., 1965, Comput. J., 7, 308
Saar S. H., Donahue R. A., 1997, ApJ, 485, 519
Saar S. H., Butler R. P., Marcy G. W., 1998, ApJ, 498, L153
Valenti J. A., Fischer D. A., 2005, ApJS, 159, 141
Wright J. T., 2005, PASP, 117, 657
Udry S. et al., 2007, A&A, 469, L43

This paper has been typeset from a\( \LaTeX\)/\( \LaTeX\) file prepared by the author.