

Bayesian re-analysis of the Gliese 581 exoplanet system

Philip C. Gregory^{*}

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

Accepted 2011 April 8. Received 2011 April 4; in original form 2010 December 28

ABSTRACT

A re-analysis of Gliese 581 HARPS and HIRES precision radial velocity data was carried out with a Bayesian multiplanet Kepler periodogram (from one to six planets) based on a fusion Markov chain Monte Carlo algorithm. In all cases, the analysis included an unknown parametrized stellar jitter noise term. For the HARPS data set, the most probable number of planetary signals detected is five with a Bayesian false alarm probability of 0.01. These include the 3.1498 ± 0.0005 , 5.3687 ± 0.0002 , $12.927^{+0.006}_{-0.004}$ and 66.9 ± 0.2 d periods reported previously plus a 399^{+14}_{-16} d period. Their orbital eccentricities are $0.0^{+0.2}_{-0.0}$, $0.00^{+0.02}_{-0.00}$, $0.10^{+0.06}_{-0.10}$, $0.33^{+0.09}_{-0.10}$ and $0.02^{+0.30}_{-0.01}$, respectively. The semimajor-axis and $M \sin i$ of the five planets are $(0.0285 \pm 0.0006 \text{ au}, 1.9 \pm 0.3 M_{\oplus})$, $(0.0406 \pm 0.0009 \text{ au}, 15.7 \pm 0.7 M_{\oplus})$, $(0.073 \pm 0.002 \text{ au}, 5.3 \pm 0.4 M_{\oplus})$, $(0.218 \pm 0.005 \text{ au}, 6.7 \pm 0.8 M_{\oplus})$ and $(0.7 \pm 0.2 \text{ au}, 6.6^{+2.0}_{-2.7} M_{\oplus})$, respectively.

The analysis of the HIRES data set yielded a reliable detection of only the strongest 5.37 and 12.9 d periods. The analysis of the combined HIRES/HARPS data again only reliably detected the 5.37 and 12.9 d periods. The detection of four planetary signals with the periods of 3.15, 5.37, 12.9 and 66.9 d was only achieved by including an additional unknown but parametrized Gaussian error term added in quadrature to the HIRES quoted errors. The marginal distribution for the σ of this additional error term has a well-defined peak at $1.8 \pm 0.4 \text{ m s}^{-1}$. It is possible that this additional error arises from unidentified systematic effects. We did not find clear evidence for a fifth planetary signal in the combined HIRES/HARPS data set. Based on the available data, our analysis does not support the claimed detection of a sixth planet Gliese 581g.

Key words: methods: statistical – methods: data analysis – techniques: radial velocities – planetary systems.

1 INTRODUCTION

A remarkable array of new ground-based and space-based astronomical tools have finally provided astronomers access to other solar systems with over 500 planets discovered to date, starting from the pioneering work of Campbell, Walker & Yang (1988), Wolszczan & Frail (1992), Mayor & Queloz (1995) and Marcy & Butler (1996). Recent interest has focused on the Gliese 581 (Gl 581) planetary system (also designated GJ 581 in the literature). It was already known to harbour three planets, including two super-Earth planets that straddle its habitable zone: Gl 581b with a period of 5.37 d (Bonfils et al. 2005b), Gl 581c (period of 12.9 d) and Gl 581d (period of 82 d) (Udry et al. 2007). Armed with additional HARPS data, Mayor et al. (2009) reported the detection of an additional planet Gl 581e with a minimum mass of $1.9 M_{\oplus}$ and a period of 3.15 d. They also corrected previous confusion about the orbital period of Gl 581d (the outermost planet) with a 1-yr alias at 82 d. The revised period is 66.8 d and positions the semimajor-axis

inside the habitable zone of the low-mass star. Vogt et al. (2010) reported the analysis of the combined HIRES and HARPS data set spanning 11 yr, claiming the detection of two additional planets, Gl 581f and Gl 581g. Gl 581f has a period of 433 d, a minimum mass of $7.0 M_{\oplus}$ and a semimajor-axis of 0.758 au. Gl 581g has a period of 36.6 d, a minimum mass of $3.1 M_{\oplus}$ and a semimajor-axis of 0.146 au. The estimated equilibrium temperature of Gl 581g is 228 K, placing it squarely in the middle of the habitable zone of the star. The Vogt et al. (2010) analysis assumed circular orbits for all six planets.

The excitement generated by the exoplanetary discoveries has spurred a significant effort to improve the statistical tools for analyzing data in this field (e.g., Lored & Chernoff 2003; Lored 2004; Cumming 2004; Gregory 2005a,b; Ford 2005, 2006; Ford & Gregory 2007; Cumming & Dragomir 2010). Much of this work has highlighted a Bayesian MCMC approach as a way to better understand parameter uncertainties and degeneracies and to compute model probabilities. Gregory (2009) and Gregory & Fischer (2010) presented a Bayesian hybrid or fusion Markov chain Monte Carlo (MCMC) algorithm that incorporates parallel tempering (PT), simulated annealing and a genetic crossover operation to facilitate the

^{*}E-mail: gregory@phas.ubc.ca

detection of a global minimum in χ^2 . This enables efficient exploration of a large model parameter space starting from a random location. When implemented with a multiplanet Kepler model,¹ it is able to identify any significant periodic signal component in the data that satisfies Kepler's laws and is able to function as a multiplanet Kepler periodogram.² In addition, the Bayesian MCMC algorithm provides full marginal parameter distributions. The algorithm includes an innovative adaptive control system (CS) that automates the selection of efficient parameter proposal distributions even if the parameters are highly correlated (Gregory 2011). A recent application of the algorithm (Gregory & Fischer 2010) confirmed the existence of a disputed second planet (Fischer et al. 2002) in 47 Ursae Majoris (47 UMa) and provided orbital constraints on a possible additional long-period planet with a period of $\sim 10\,000$ d.

This paper reports the results of a re-analysis of the HARPS (Mayor et al. 2009) and HIRES data (Vogt et al. 2010) for Gl 581 using the above-mentioned Bayesian multiplanet Kepler periodogram. Section 2 provides an introduction to our Bayesian approach and describes the adaptive fusion MCMC (FMCMC) algorithm. Section 3 gives the model equations and priors. Sections 4 and 5 present the parameter estimation and model selection results for the analysis of the HARPS data alone. Section 6 is devoted to the analysis of the HIRES data followed by the analysis of the combination of HIRES and HARPS data. The final two sections are devoted to discussion and conclusions.

2 THE ADAPTIVE FMCMC

The adaptive fusion³ MCMC (FMCMC) is a very general Bayesian non-linear model-fitting program. After specifying the model, M_i , data, D , and priors, I , Bayes' theorem dictates the target joint probability distribution for the model parameters which is given by

$$p(X|D, M_i, I) = Cp(X|M_i, I) \times p(D|M_i, X, I), \quad (1)$$

where C is the normalization constant and X represents the vector of model parameters. The first term on the right-hand side of the equation, $p(X|M_i, I)$, is the prior probability distribution of X , prior to the consideration of the current data D . The second term, $p(D|X, M_i, I)$, is called the likelihood and there is the probability that we would have obtained the measured data D for this particular choice of the parameter vector X , model M_i and prior information I . At the very least, the prior information, I , must specify the class of alternative models (hypotheses) being considered (hypothesis space of interest), and the relationship between the models and the data (how to compute the likelihood). In some simple cases, the log of the likelihood is simply proportional to the familiar χ^2 statistic. For further details of the likelihood function for this type of problems, see Gregory (2005b).

To compute the marginals for any subset of the parameters, it is necessary to integrate the joint probability distribution over the remaining parameters. For example, the marginal probability density

¹ For multiple planet models, there is no analytic expression for the exact radial velocity (RV) perturbation. In many cases, the RV perturbation can be well modelled as the sum of multiple independent Keplerian orbits which is what has been assumed in this paper.

² Following on from the pioneering work on Bayesian periodograms by Jaynes (1987) and Bretthorst (1988).

³ In earlier papers, the algorithm was referred to as a hybrid MCMC. We subsequently learned that this term already exists in the literature in connection with a Hamiltonian version of a MCMC. In this paper, we replace the term hybrid by fusion.

function (PDF) of the orbital period in a one-planet RV model fit is given by

$$\begin{aligned} p(P|D, M_1, I) &= \int dK \int dV \int de \int d\chi \int d\omega \int ds \\ &\times p(P, K, V, e, \chi, \omega, s|D, M_1, I) \\ &\propto p(P|M_1, I) \int dK \cdots \int ds \\ &\times p(K, V, e, \chi, \omega, s|M_1, I) \\ &\times p(D|M_1, P, K, V, e, \chi, \omega, s, I), \end{aligned} \quad (2)$$

where $p(P, K, V, e, \chi, \omega, s|D, M_1, I)$ is the target joint probability distribution of the RV model parameters (P, K, V, e, χ, ω) and s is an extra noise parameter which is discussed in Section 3. $p(P|M_1, I)$ is the prior for the orbital period parameter, $p(K, V, e, \chi, \omega, s|M_1, I)$ is the joint prior for the other parameters and $p(D|M_1, P, K, V, e, \chi, \omega, s, I)$ is the likelihood. For a five-planet model fit, we need to integrate over 26 parameters to obtain $p(P|D, M_1, I)$. Integration is more difficult than maximization; however, the Bayesian solution provides the most accurate information about the parameter errors and correlations without the need for any additional calculations, that is, Monte Carlo simulations. Bayesian model selection requires integrating over all the model parameters.

In high dimensions, the principal tool for carrying out the integrals is an MCMC based on the Metropolis algorithm. The greater efficiency of an MCMC stems from its ability, after an initial burn-in period, to generate samples in parameter space in direct proportion to the joint target probability distribution. In contrast, straight Monte Carlo integration randomly samples the parameter space and wastes most of its time-sampling regions of very low probability.

MCMC algorithms avoid the requirement for completely independent samples, by constructing a kind of random walk in the model parameter space such that the number of samples in a particular region of this space is proportional to the target posterior density for that region. The random walk is accomplished using a Markov chain, whereby the new sample, X_{t+1} , depends on the previous sample, X_t , according to a time-independent entity called the *transition kernel*, $p(X_{t+1}|X_t)$. The remarkable property of $p(X_{t+1}|X_t)$ is that after an initial burn-in period (which is discarded) it generates samples of X with a probability density proportional to the desired posterior $p(X|D, M_1, I)$ (e.g. see chapter 12 of Gregory 2005a for details).

The transition kernel, $p(X_{t+1}|X_t)$, is given by

$$p(X_{t+1}|X_t) = q(X_{t+1}|X_t)\alpha(X_t, X_{t+1}), \quad (3)$$

where $\alpha(X_t, X_{t+1})$ is called the acceptance probability and is given by equation (4). This is achieved by proposing a new sample X_{t+1} from a *proposal distribution*, $q(X_{t+1}|X_t)$, which is easy to evaluate and is centred on the current sample X_t . The proposal distribution can have almost any form. A common choice for $q(X_{t+1}|X_t)$ is a multivariate normal (Gaussian) distribution. With such a proposal distribution, the probability density decreases with the distance away from the current sample. The new sample X_{t+1} is accepted with a probability $\alpha(X_t, X_{t+1})$ given by

$$\alpha(X_t, X_{t+1}) = \min \left[1, \frac{p(X_{t+1}|D, I)}{p(X_t|D, I)} \frac{q(X_t|X_{t+1})}{q(X_{t+1}|X_t)} \right], \quad (4)$$

where $q(X_t|X_{t+1}) = q(X_{t+1}|X_t)$ for a symmetrical proposal distribution. If the proposal is not accepted, then the current sample X_t is repeated.

An important feature that prevents the FMCMC from becoming stuck in a local probability maximum is PT (Geyer 1991, and

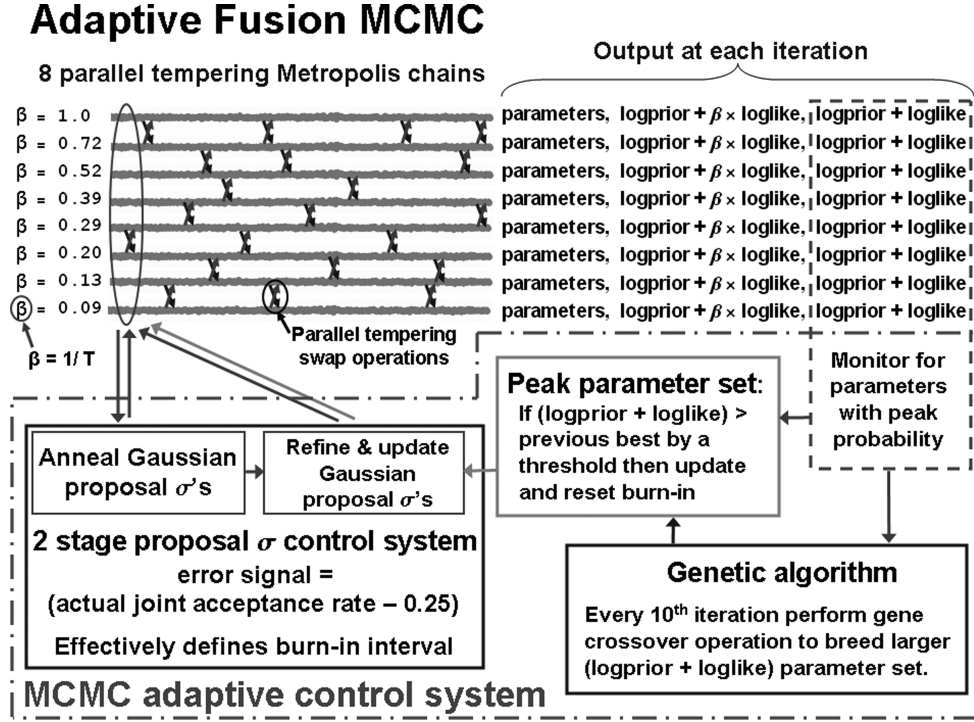


Figure 1. The first of the two schematics on the operation of the adaptive FCMC algorithm.

re-invented by Hukushima & Nemoto 1996 under the name *exchange Monte Carlo*). Multiple MCMC chains are run in parallel. The joint distribution for the parameters (X) of model M_i , for a particular chain, is given by

$$\pi(X|D, M_i, I, \beta) \propto p(X|M_i, I) \times p(D|X, M_i, I)^\beta. \quad (5)$$

Each MCMC chain corresponds to a different β , with the value of β ranging from zero to 1. When the exponent $\beta = 1$, the term on the left-hand side of the equation is the target joint probability distribution for the model parameters, $p(X|D, M_i, I)$. For $\beta \ll 1$, the distribution is much flatter.

In equation (5), an exponent $\beta = 0$ yields a joint density distribution equal to the prior. The reciprocal of β is analogous to a temperature; the higher the temperature, the broader the distribution. For parameter estimation purposes, eight chains were employed. A representative set of β values is shown in Fig. 1. At an interval of 10 iterations, a pair of adjacent chains on the tempering ladder are chosen at random and a proposal made to swap their parameter states. A Monte Carlo acceptance rule determines the probability for the proposed swap to occur (e.g. Gregory 2005a, equation 12.12). This swap allows for an exchange of information across the population of parallel simulations. In low- β (higher temperature) simulations, radically different configurations can arise, whereas in higher β (lower temperature) states, a configuration is given the chance to refine itself. The lower β chains can be likened to a series of scouts that explore the parameter terrain on different scales. The final samples are drawn from the $\beta = 1$ chain, which corresponds to the desired target probability distribution. The choice of β values can be checked by computing the swap acceptance rate. When they are too far apart, the swap rate drops to very low values. In this work, a typical swap acceptance rate of ≈ 30 per cent was employed but rates in a broad range from 0.15 to 0.5 were deemed acceptable as they did not exhibit any clear differences in performance. For a swap acceptance rate of 30 per cent, jumps to adjacent chains will occur at

an interval of ~ 230 iterations, while information from more distant chains will diffuse much more slowly. Recently, Atchadé, Roberts & Rosenthal (2010) have shown that under certain conditions, the optimal swap acceptance rate is 0.234. A future goal for a FCMC is to extend the CS to automate the selection of an optimal set of β values as well.

At each iteration, a single joint proposal to jump to a new location in the parameter space is generated from independent Gaussian proposal distributions (centred on the current parameter location), one for each parameter. In general, the values of σ for these Gaussian proposal distributions are different because the parameters can be very different entities. If the values of σ are chosen too small, successive samples will be highly correlated and will require many iterations to obtain an equilibrium set of samples. If the values of σ are too large, then proposed samples will very rarely be accepted. The process of choosing a set of useful proposal values of σ when dealing with a large number of different parameters can be very time-consuming. In a PT MCMC, this problem is compounded because of the need for a separate set of Gaussian proposal values of σ for each tempering chain. This process is automated by an innovative three-stage statistical CS (Gregory 2007b; Gregory 2009) in which the error signal is proportional to the difference between the current joint parameter acceptance rate and a target acceptance rate, α (typically $\alpha \sim 0.25$). Roberts, Gelman & Gilks (1997) showed that α exhibits a weak convergence to an asymptotic value of 0.234 as the number of parameters converges to ∞ under quite general conditions. A schematic of the first two stages of the adaptive CS is shown⁴ in Fig. 1. The third stage that handles highly correlated parameters is described in Section 2.1.

The first stage of the CS, which involves annealing the set of Gaussian proposal distribution values of σ , was first presented in

⁴ The interval between tempering swap operations is typically much smaller than is suggested by this schematic.

Gregory (2005a). An initial set of proposal values of σ (≈ 10 per cent of the prior range for each parameter) are used for each chain. During the major cycles of length n_1 iterations (typically $n_1 = 1000$), the joint acceptance rate is measured based on the current proposal values of σ and compared to a target acceptance rate. During the minor cycles of length n_2 (typically $n_2 = 200$), each proposal value of σ is separately perturbed to determine an approximate gradient in the acceptance rate [i.e. determine which parameter(s) are limiting the acceptance rate]. The values of σ are then jointly scaled by factors (maximum change is a factor of 10) that depend on the gradient, and on both the magnitude and sign of the error signal. This is done for each of the parallel chains. The first stage of the CS is turned off after the first instance that the error signal for the $\beta = 1$ chain changes sign and is less than a threshold $\text{tol} = 1.5 \times \sqrt{\alpha \times n_1}$, provided the total number of iterations is $\geq \text{minCS}$. Typically, minCS is in the range 10 000 (for a one-planet fit) to 30 000 (for a fit to three or more planets). While the values of σ are being annealed, the FCMCMC is homing in on a good set of parameter values and minCS helps to keep these two processes more in step.

If the desired threshold, tol , is not reached within maxCS iterations, then the first stage of the CS is turned off and the second stage initiated (see below). Typically, maxCS is in the range 15 000 (for a one-planet fit) to 60 000 (for a fit to three or more planets). After maxCS iterations, it is more efficient to employ the mathematics of the second stage of the CS. Proposals to swap parameter values between chains are allowed during major cycles but not within minor cycles. The code used for the first two stages of the CS is given in the MCMC-powered Kepler periodogram example of the MATHEMATICA support material accompanying the textbook ‘Bayesian Logical Data Analysis for the Physical Sciences’ (Gregory 2005a). The software is freely available on the Cambridge University Press website for this book.

The annealing of the proposal values of σ occurs while the FCMCMC is homing in on any significant peaks in the target probability distribution. Concurrent with this, another aspect of the annealing operation takes place whenever the Markov chain is started from a location in parameter space that is far from the best-fitting values. This automatically arises because all the models considered incorporate an extra additive noise term (Gregory 2005b) whose probability distribution is Gaussian with zero mean and with an unknown standard deviation s . When the χ^2 of the fit is very large, the Bayesian Markov chain automatically inflates s to include anything in the data that cannot be accounted for by the model with the current set of parameters and the known measurement errors. This results in a smoothing out of the detailed structure in the χ^2 surface and, as pointed out by Ford (2006), allows the Markov chain to explore the large-scale structure in parameter space more quickly. The chain begins to decrease the value of the extra noise as it settles in near the best-fitting parameters. An example of this is shown in Fig. 2 for a two-planet fit to the HARPS data as discussed in Section 4.1. The three panels show the evolution of the $\log_{10}(\text{prior} \times \text{likelihood})$, the s parameter and the two period parameters. In the early stages, the extra noise is inflated to around 16 m s^{-1} and then rapidly decays to much lower values as it homes in on a good solution. This is similar to simulated annealing, but does not require choosing a cooling scheme. In this example, the starting parameter values were far from the best and the MCMC algorithm finds several less probable solutions en route to a final best choice. Initially, it homes in on the two periods of 5.37 and 66.9 d and the CS switches off around iteration 220 000, but around iteration 500 000 it detects a much more probable solution for the period combinations of 5.37

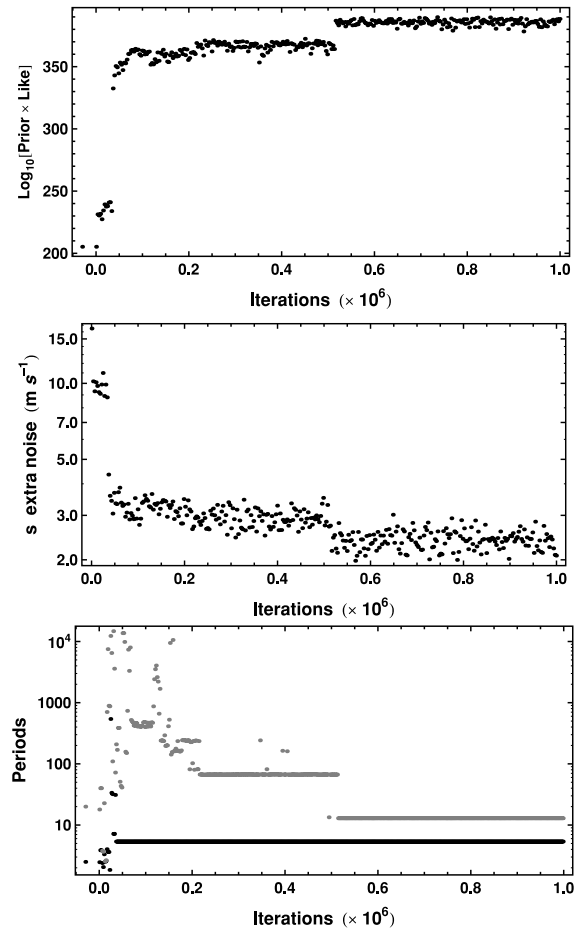


Figure 2. The upper panel is a plot of the $\log_{10}(\text{prior} \times \text{likelihood})$ versus MCMC iteration for a blind two-planet fit to the HARPS data. The middle panel is a similar plot for the extra noise term s . Initially, s is inflated and then rapidly decays to a much lower level as the better-fitting parameter values are approached. The lower panel shows the values of the two unknown period parameters versus iteration number. The two starting periods of 2.5 and 20 d are shown on the left-hand side of the plot at a negative iteration number.

and 12.9 d. In this example, the adaptive CS switched on again briefly following the detection of the much improved solution.

Although the first stage of the CS achieves the desired joint acceptance rate, it often happens that a subset of the proposal values of σ are too small leading to an excessive autocorrelation in the FCMCMC iterations for these parameters. Part of the second stage of the CS corrects for this. The goal of the second stage is to achieve a set of proposal values of σ that equalize the FCMCMC acceptance rates when new parameter values are proposed separately and achieve the desired acceptance rate when they are proposed jointly. Details of the second stage of the CS are given in Gregory (2007b).

The first stage is run only once at the beginning, but the second stage can be executed repeatedly, whenever a significantly improved parameter solution emerges. Frequently, the algorithm homes in on the most significant peak within the span of the first stage of the CS and the second stage improves the choice of proposal values of σ based on the highest probability parameter set. Occasionally, a new higher (by a user-specified threshold) target probability parameter set emerges after the first two stages of the CS have completed. The CS has the ability to detect this and automatically reactivate the second stage. In this sense, the CS is adaptive. If this happens, the iteration corresponding to the end of the CS is reset. The

requirement that the transition kernel be time-independent means that $q(X_{t+1}|X_t)$ be time-independent, so useful FCMCMC simulation data are obtained only after the CS is switched off.

The adaptive capability of the CS can be appreciated from an examination of Fig. 1. The upper left-hand portion of the figure depicts the FCMCMC iterations from the eight parallel chains, each corresponding to a different tempering level β as indicated on the extreme left-hand side. One of the outputs obtained from each chain at every iteration (shown at the upper extreme right-hand side) is the log prior + log likelihood. This information is continuously fed into the CS which constantly updates the most probable parameter combination, regardless of which chain the parameter set occurred in. This is passed to the ‘peak parameter set’ block of the CS. Its job is to decide if a significantly more probable parameter set has emerged since the last execution of the second stage of the CS. If so, then the second stage of the CS is re-run using the new more probable parameter set which is the basic adaptive feature of the existing CS.

The CS also includes a genetic algorithm block which is shown in the bottom right-hand side of Fig. 1. The current parameter set can be treated as a set of genes. In the present version, one gene consists of the parameter set that specifies one orbit. On this basis, a three-planet model has three genes. At any iteration, there exist within the CS the most probable parameter set to date, X_{\max} , and the current most probable parameter set of the eight chains, X_{cur} . At regular intervals (user specified), each gene from X_{cur} is swapped for the corresponding gene in X_{\max} . If either substitution leads to a higher probability, then it is retained and X_{\max} updated. The effectiveness of this operation can be tested by comparing the number of times the gene crossover operation gives rise to a new value of X_{\max} with the number of new X_{\max} arising from the normal PT FCMCMC iterations. The gene crossover operations prove to be very effective and give rise to new X_{\max} values approximately three times more often. Of course, most of these swaps lead to

very minor changes in probability but occasionally big jumps are created.

Gene swaps from $X_{\text{cur}2}$, the parameters of the second most probable current chain, to X_{\max} can also be utilized. This gives rise to new values of X_{\max} at a rate approximately half that of swaps from X_{cur} to X_{\max} . Crossover operations at a random point in the entire parameter set did not prove as effective except in the single-planet case where there is only one gene. Further experimentation with this concept is ongoing.

2.1 Highly correlated parameters

The part of the algorithm described above (Fig. 1) is most efficient when working with model parameters that are independent of or transformed to new independent parameters. New parameter values are jointly proposed based on independent Gaussian proposal distributions (‘I’ scheme), one for each parameter. Initially, only this ‘I’ proposal system is used and it is clear that if there are strong correlations between any parameters, then the values of σ of the independent Gaussian proposals will need to be very small for any proposal to be accepted and consequently convergence will be very slow. However, the accepted ‘I’ proposals will generally cluster along the correlation path. In the optional third stage of the CS shown in Fig. 3, every second accepted ‘I’ proposal is appended to a correlated sample buffer. Only the 300 most recent additions to the buffer are retained. A ‘C’ proposal is generated from the difference between a pair of randomly selected samples drawn from the correlated sample buffer, after multiplication by a constant. The value of this constant is computed automatically by another CS module which ensures that the ‘C’ proposal acceptance rate is close to 25 per cent. With very little computational overhead, the ‘C’ proposals provide the scale and direction for efficient jumps in a correlated parameter space.

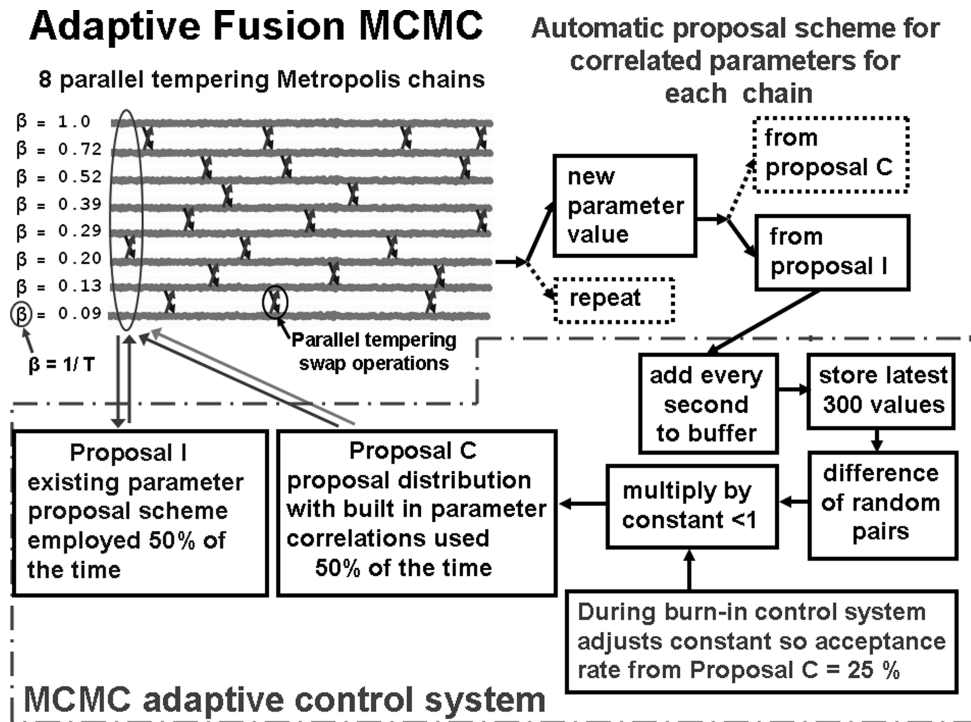


Figure 3. Schematic outlining the operation of the third stage of the adaptive FCMCMC algorithm that handles correlated parameters.

The final proposal distribution is a random selection of ‘I’ and ‘C’ proposals such that each is employed 50 per cent of the time. The overhead to generate the ‘C’ proposals is minimal. The combination ensures that the whole parameter space can be reached and that the FCMC chain is aperiodic. The PT feature operates as before to avoid becoming trapped in a local probability maximum.

Because the ‘C’ proposals reflect the parameter correlations, large jumps are possible allowing for much more efficient movement in parameter space than can be achieved by the ‘I’ proposals alone. Once the first two stages of the CS have been turned off, the third stage continues until a minimum of an additional 300 accepted ‘I’ proposals have been added to the buffer and the ‘C’ proposal acceptance rate is within the range ≥ 0.22 and ≤ 0.28 . At this point, further additions to the buffer are terminated and this sets a lower bound on the burn-in period.

Full details on the operation and testing of the combined ‘I’ and ‘C’ proposal scheme are given in Gregory (2011).

3 MODELS AND PRIORS

In this section, we describe the model-fitting equations and the selection of priors for the model parameters. We have investigated the Gl 581 data using models ranging from one to six planets. For a one-planet model, the predicted RV is given by

$$v(t_i) = V + K[\cos\{\theta(t_i + \chi P) + \omega\} + e \cos \omega], \quad (6)$$

which involves the following six unknown parameters:

- V = a constant velocity;
- K = velocity semiamplitude;
- P = the orbital period;
- e = the orbital eccentricity;
- ω = the longitude of periastron;
- χ = the fraction of an orbit, prior to the start of data taking, that periastron occurred at. Thus, χP = the number of days prior to $t_i = 0$ that the star was at periastron, for an orbital period of P days;
- $\theta(t_i + \chi P)$ = the true anomaly, the angle of the star in its orbit relative to periastron at time t_i .

We utilize this form of the equation because we obtain the dependence of θ on t_i by solving the conservation of angular momentum equation

$$\frac{d\theta}{dt} - \frac{2\pi[1 + e \cos \theta(t_i + \chi P)]^2}{P(1 - e^2)^{3/2}} = 0. \quad (7)$$

Our algorithm is implemented in MATHEMATICA and it proves faster for MATHEMATICA to solve this differential equation than to solve the equations relating the true anomaly to the mean anomaly via the eccentric anomaly. MATHEMATICA generates an accurate interpolating function between t and θ so the differential equation does not need to be solved separately for each t_i . Evaluating the interpolating function for each t_i is very fast compared to solving the differential equation. Details on how equation (7) is implemented are given in the Appendix A.

The model RV accuracy is limited by the accuracy of this interpolating function which varies with the eccentricity, e . As shown in Appendix A, for $e = 0$ to 0.8 the maximum model RV error, expressed as a fraction of the K parameter, is $\leq 2.2 \times 10^{-5}$. The situation degrades progressively for larger values of e but is still $\leq 2.8 \times 10^{-3}$ for $e = 0.98$, rising to $\leq 1.2 \times 10^{-2}$ for $e = 0.99$. Even for $e = 0.99$ the fractional error is $< 10^5$ over most of the interpolation range, only rising above this towards the very end of the range.

As described in more detail in Gregory (2007a), we employed a re-parametrization of χ and ω to improve the MCMC convergence speed motivated by the work of Ford (2006). The two new parameters are $\psi = 2\pi\chi + \omega$ and $\phi = 2\pi\chi - \omega$. The parameter ψ is well determined for all eccentricities. Although ϕ is not well determined for low eccentricities, it is at least orthogonal to the ψ parameter. We use a uniform prior for ψ in the interval 0 to 4π and a uniform prior for ϕ in the interval -2π to $+2\pi$. This insures that a prior that is wraparound continuous in (χ, ω) maps into a wraparound continuous distribution in (ψ, ϕ) . To account for the Jacobian of this re-parametrization, it is necessary to multiply the Bayesian integrals by a factor of $(4\pi)^{-n_{\text{plan}}}$, where n_{plan} = the number of planets in the model. Also, by utilizing the orthogonal combination (ψ, ϕ) , it was not necessary to make use of the ‘C’ proposal scheme outlined in Section 2.1 which typically saves about 25 per cent of the execution time.

In a Bayesian analysis, we need to specify a suitable prior for each parameter. These are tabulated in Table 1. For the current problem, the prior given in equation (5) is the product of the individual parameter priors. Detailed arguments for the choice of each prior are given in Gregory (2007a).

Gregory (2007a) discussed two different strategies to search for the orbital frequency parameter space for a multiplanet model: (i) an upper bound on $f_1 \leq f_2 \leq \dots \leq f_n$ is utilized to maintain the identity of the frequencies; and (ii) all f_i are allowed to roam over the entire frequency range and the parameters re-labelled afterwards. Case (ii) was found to be significantly more successful at converging on the highest posterior probability peak in fewer iterations during repeated blind frequency searches. In addition, case (ii) more easily permits the identification of two planets in 1:1 resonant orbits. We adopted approach (ii) in the current analysis.

All the models considered in this paper incorporate an extra noise parameter, s , that can allow for any additional noise beyond the known measurement uncertainties.⁵ We assume the noise variance is finite and adopt a Gaussian distribution with a variance s^2 . Thus, the combination of the known errors and extra noise has a Gaussian distribution with variance $= \sigma_i^2 + s^2$, where σ_i is the standard deviation of the known noise for the i th data point. For example, suppose the star actually has two planets and the model assumes only one is present. As regards the single-planet model, the velocity variations induced by the unknown second planet act like an additional unknown noise term. Other factors like star-spots and chromospheric activity can also contribute to this extra velocity noise term which is often referred to as stellar jitter. Several researchers have attempted to estimate the stellar jitter for individual stars based on statistical correlations with observables (e.g. Saar & Donahue 1997; Saar, Butler & Marcy 1998; Wright 2005). In general, nature is more complicated than our model and known noise terms. Marginalizing s has the desirable effect of treating anything in the data that cannot be explained by the model and known measurement errors as noise, leading to conservative estimates of orbital parameters (see sections 9.2.3 and 9.2.4 of Gregory 2005a for a tutorial demonstration of this point). If there is no extra noise, then the posterior probability distribution for s will peak at $s = 0$. The upper limit on s was set equal

⁵ 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	$p(\ln f_1, \ln f_2, \dots, \ln f_n M_n, I) = \frac{n!}{[\ln(f_H/f_L)]^n}$ (n = number of planets)	1/1.1 d	1/1000 yr
Velocity K_i (m s^{-1})	Modified Jeffreys prior ^a $\frac{(K+K_0)^{-1}}{\ln \left[1 + \frac{K_{\max}}{K_0} \left(\frac{P_{\min}}{P_i} \right)^{1/3} \frac{1}{\sqrt{1-e_i^2}} \right]}$	0 ($K_0 = 1$)	$K_{\max} \left(\frac{P_{\min}}{P_i} \right)^{1/3} \frac{1}{\sqrt{1-e_i^2}}$ $K_{\max} = 2129$
V (m s^{-1})	Uniform	$-K_{\max}$	K_{\max}
Eccentricity e_i	(a) Uniform (b) Eccentricity noise bias correction filter	0 0	1 0.99
Orbit fraction χ	Uniform	0	1
Longitude of periastron ω_i	Uniform	0	2π
Extra noise s (m s^{-1})	$\frac{(s+s_0)^{-1}}{\ln \left(1 + \frac{s_{\max}}{s_0} \right)}$	0 ($s_0 = 1$)	K_{\max}

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

$$p(X|M, I) = \frac{1}{X + X_0} \frac{1}{\ln \left(1 + \frac{X_{\max}}{X_0} \right)} \quad (8)$$

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

to K_{\max} . We employed a modified Jeffreys prior for s with a knee, $s_0 = 1 \text{ m s}^{-1}$.

We used two different choices of priors for eccentricity, a uniform prior and eccentricity noise bias correction filter that is described in the next section.

3.1 Eccentricity bias

In Gregory & Fischer (2010), the velocities of model fit residuals were randomized in multiple trials and processed using a one-planet version of the FCMCMC Kepler periodogram. In this situation, periodogram probability peaks are purely the result of the effective noise. The orbits corresponding to these noise-induced periodogram peaks exhibited a well-defined statistical bias⁶ towards high eccentricity. They offered the following explanation for this effect. To mimic a circular velocity orbit, the noise points need to be correlated over a larger fraction of the orbit than they do to mimic a highly eccentric orbit. For this reason, it is more likely that noise will give rise to spurious highly eccentric orbits than low-eccentricity orbits.

Gregory & Fischer (2010) characterized this eccentricity bias and designed a correction filter that can be used as an alternate prior for eccentricity to enhance the detection of planetary orbits of low or moderate eccentricity. On the basis of our understanding of the mechanism underlying the eccentricity bias, we expect the eccentricity prior filter to be generally applicable to searches for low-amplitude orbital signals in precision RV data sets. The PDF for this filter is shown by the solid black curve in Fig. 4 and is given by

$$\text{PDF}(e) = 1.3889 - 1.5212e^2 + 0.53944e^3 - 1.6605(e - 0.24821)^8. \quad (9)$$

⁶ The bias found using multiple sets of randomized residuals from a five-planet fit to 55 Cancri combined Lick and Keck data agreed closely with the bias found for multiple sets of randomized residuals from both two- and three-planet fits to 47 UMa Lick data.

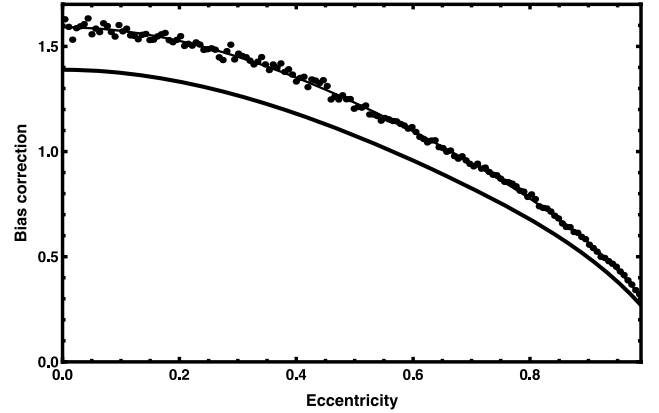


Figure 4. The best-fitting polynomial (thin black curve) to the reciprocal of the mean of the eccentricity bias determined by Gregory & Fischer (2010). After normalization, this yields the eccentricity noise bias correction filter (lower solid black curve).

In a related study, Shen & Turner (2009) explored least- χ^2 Keplerian fits to synthetic RV data sets. They found that the best-fitting eccentricities for a low signal-to-noise ratio, $K/\sigma \leq 3$, and a moderate number of observations, $N_{\text{obs}} \leq 60$, were systematically biased to higher values, leading to the suppression of the number of nearly circular orbits. More recently, Zakamska, Pan & Ford (2011) found that eccentricities of planets on nearly-circular orbits are preferentially overestimated, with a typical bias of one to two times the median eccentricity uncertainty in a survey, for example, 0.04 in the Butler et al. catalogue (Butler et al. 2006). When performing a population analysis, they recommend using the mode of the marginalized posterior eccentricity distribution to minimize potential biases.

In the analysis of the Gl 581 data, we used the eccentricity noise bias correction filter as the eccentricity prior on fits of all the models, with occasional runs using a uniform eccentricity prior to test the robustness of our conclusions.

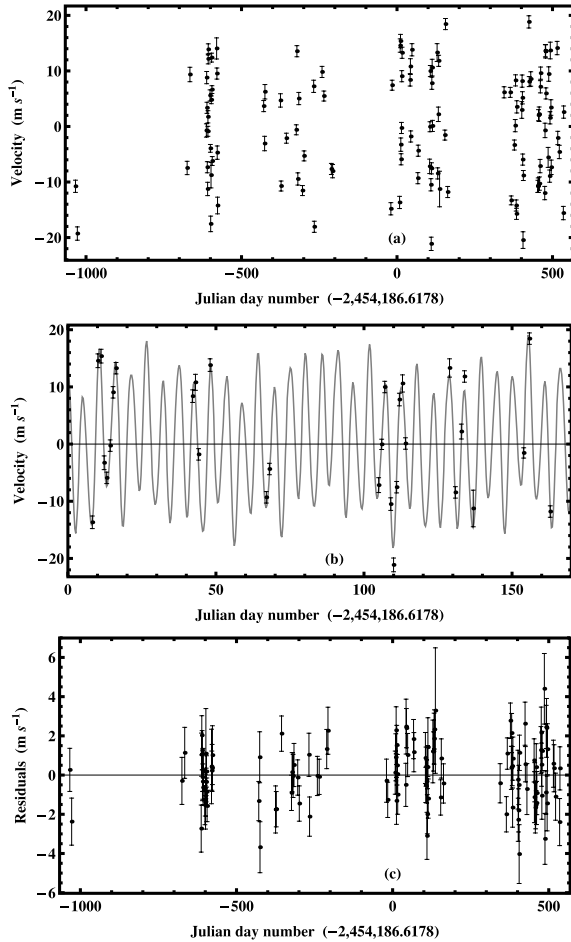


Figure 5. Panel (a) shows the HARPS observations of Gl 581. Panel (b) shows a blow-up of the mean five-planet model fit compared to the data and panel (c) shows the residuals.

4 ANALYSIS OF THE HARPS DATA

The HARPS data (Mayor et al. 2009) were retrieved electronically.⁷ A mean velocity of $-9.208\,0205\text{ km s}^{-1}$ was subtracted and the remainder converted to units of m s^{-1} . Panel (a) of Fig. 5 shows the HARPS observations of Gl 581. Panel (b) shows a blow-up of a portion of the mean five-planet model fit compared to the data, and panel (c) shows the residuals. The zero reference time is the mean time of the HARPS observations which corresponds to a Julian Day number = 245 4186.6178.

4.1 Two-planet model

The results of our two-planet Kepler periodogram analysis of these data are shown in Figs 2 and 6. The upper panel of Fig. 2 shows a plot of the $\log_{10}(\text{prior} \times \text{likelihood})$ versus FCMC iteration for a two-planet fit of the HARPS data. The lower panel shows the values of the two unknown period parameters versus iteration number. The two starting periods of 2.5 and 20 d are shown on the left-hand side of the plot at a negative iteration number. The larger of the two period parameters finds both the 67 and 12.9 d periods, but the latter has a much larger $\log_{10}(\text{prior} \times \text{likelihood})$ value. The median value of the extra noise parameter $s = 2.37\text{ m s}^{-1}$.

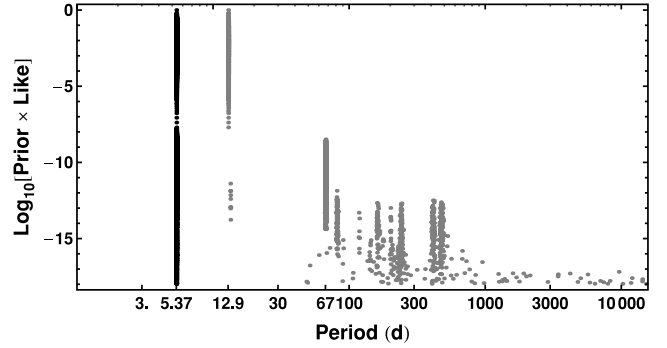


Figure 6. A plot of the two unknown period parameters versus a normalized value of $\log_{10}(\text{prior} \times \text{likelihood})$ for the two-planet FCMC Kepler periodogram of the HARPS data.

Fig. 6 shows a plot of a sample of the FCMC two period parameters versus a normalized value of $\log_{10}(\text{prior} \times \text{likelihood})$, that is, a two-planet periodogram. Only values within 18 decades of the maximum $\log_{10}(\text{prior} \times \text{likelihood})$ are plotted but without regards to whether the values occurred before or after burn-in. The two prominent periods are 5.37 and 12.9 d. The second period parameter exhibited many other peaks, but these were all at least eight decades less probable.

4.2 Three-planet model

The results of our three-planet Kepler periodogram analysis are shown in Figs 7–10. The upper panel of Fig. 7 shows a plot of the $\log_{10}(\text{prior} \times \text{likelihood})$ versus FCMC iteration for a three-planet fit of the HARPS data. The lower panel shows the FCMC values of the three unknown period parameters versus iteration number. The three starting periods of 2.5, 20 and 100 d are shown on the

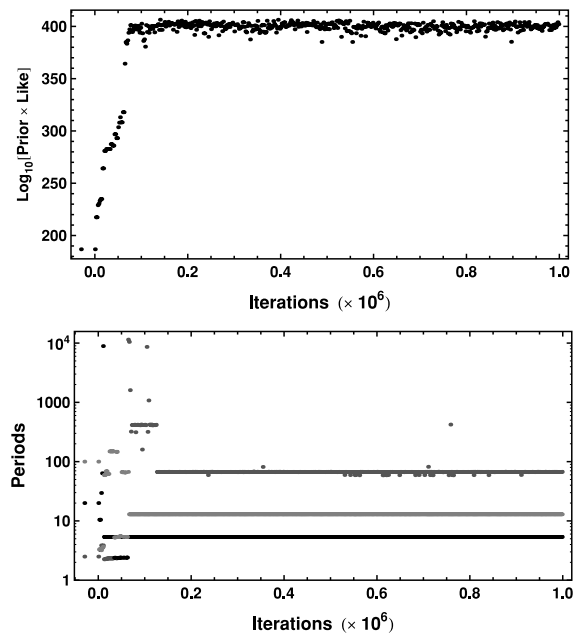


Figure 7. The upper panel is a plot of $\log_{10}(\text{prior} \times \text{likelihood})$ versus iteration for the three-planet FCMC Kepler periodogram of the HARPS data. The lower panel shows the values of the three unknown period parameters versus iteration number. The three starting periods of 2.5, 20 and 100 d are shown on the left-hand side of the plot at a negative iteration number.

⁷ <http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/A+A/507/487>

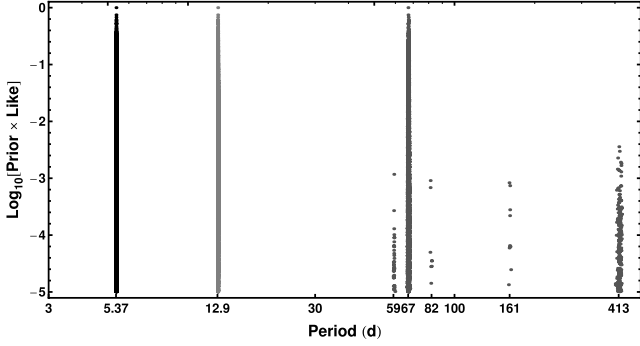


Figure 8. A plot of the three period parameter values versus a normalized value of $\log_{10}(\text{prior} \times \text{likelihood})$ for the three-planet FCMC Kepler periodogram of the HARPS data.

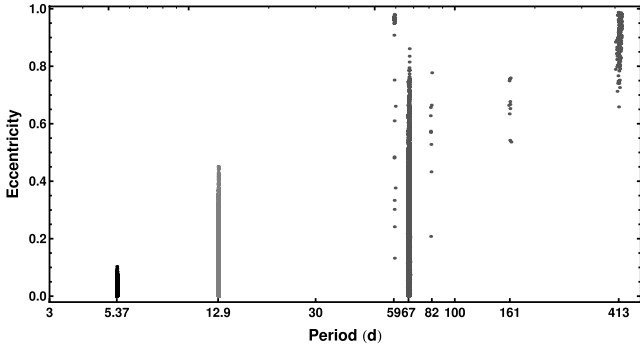


Figure 9. A plot of eccentricity versus period for the three-planet FCMC Kepler periodogram of the HARPS data.

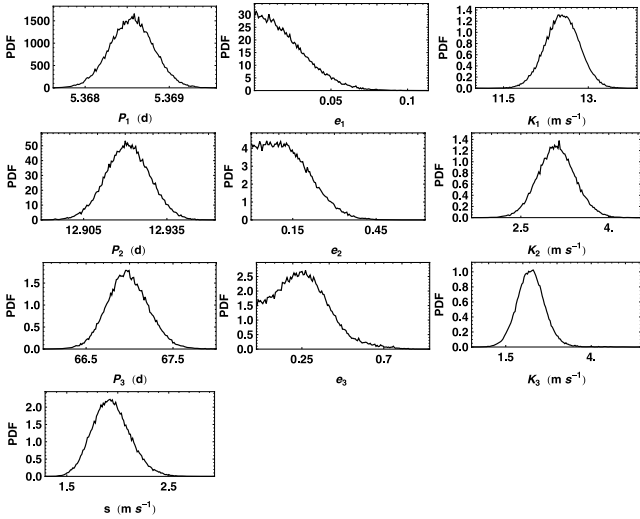


Figure 10. A plot of a subset of the FCMC parameter marginal distributions for a three-planet fit of the HARPS data.

left-hand side of the plot at a negative iteration number. The burn-in period for this run was 0.13×10^6 iterations.

Fig. 8 shows a plot of a sample of the FCMC three period parameters versus a normalized version of $\log_{10}(\text{prior} \times \text{likelihood})$, that is, a three-planet periodogram. Only values within five decades of the maximum $\log_{10}(\text{prior} \times \text{likelihood})$ are plotted but without regards to whether the values occurred before or after burn-in. Three prominent periods were clearly detected: 5.37, 12.9 and 66.9 d. The third period parameter exhibited four other peaks, but these were all

more than two decades less probable. The most probable of these has a period of ~ 413 d. The spectral peak at 82 d coincides with a 1-yr alias ($1/67 - 1/365 \sim 1/82$) of the dominant 67 d period. For more on RV aliases, see Dawson & Fabrycky (2010). The 59-d peak is close but not coincident with the other 1-yr alias ($1/67 + 1/365 \sim 1/57$).

Fig. 9 shows a plot of eccentricity versus period for a sample of the FCMC parameter samples for the three-planet model. There is clearly a large uncertainty in the eccentricity of the 67 d period which extends down to low eccentricities. The 413 d period peak exhibits very large eccentricity values. Gregory & Fischer (2010) showed that it is more likely that noise will give rise to spurious highly eccentric orbits than low-eccentricity orbits. To mimic a circular velocity orbit, the noise points need to be correlated over a larger fraction of the orbit than they do to mimic a highly eccentric orbit. Even though we are using the noise-induced eccentricity prior proposed in Gregory & Fischer (2010), we still observe a preponderance of high-eccentricity orbital solutions in the low- K -value regime.

Fig. 10 shows a plot of a subset of the FCMC parameter marginal distributions for the three-planet fit of the HARPS data after filtering out the post-burn-in FCMC iterations that correspond to the three dominant period peaks at 5.37, 12.9 and 66.9 d. The bottom panel shows the marginal for the unknown standard deviation, s , of the additive Gaussian extra noise term which has a median value of 1.93 m s^{-1} . The Bayesian analysis automatically inflates s to account for anything in the data that the model and quoted measurement errors cannot account for including stellar jitter.

The three-planet model was also run using a flat uniform eccentricity to compare with the results obtained with the noise-induced eccentricity prior. Fig. 11 shows a comparison of the eccentricity marginals for the noise-induced prior (solid black curve) and the uniform prior (grey dashed curve).

4.3 Four-planet model

A one-planet fit to the residuals of the three-planet fit above yielded a dominant Keplerian orbit with a period of 3.15 d. The results of our four-planet Kepler periodogram analysis are shown in Figs 12 and 13. All four period parameters were free to roam within a search range extending from 1.1 d to 10 times the data duration. Another run that extended the period search range down to 0.5 d yielded the same four periods. The median value of the extra noise parameter $s = 1.36 \text{ m s}^{-1}$.

4.4 Five-planet model

The results of a five-planet Kepler periodogram analysis of the HARPS data are shown in Figs 14–16. The starting period for the fifth period was set to 300 d and the most probable period was found to be ~ 400 d. The best set of parameters from the four-planet fit were used as start parameters. The fifth period parameter shows three peaks but the 400 d period is almost 1000 times stronger than the other two. We filtered the five-planet MCMC results to only include the fifth period values from the 400-d peak region and used the Gelman & Rubin (1992) statistic to test for convergence. In a PT 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. We divided the $\beta = 1$ iterations into 10 equal time-intervals and intercompared the 10 different essentially independent

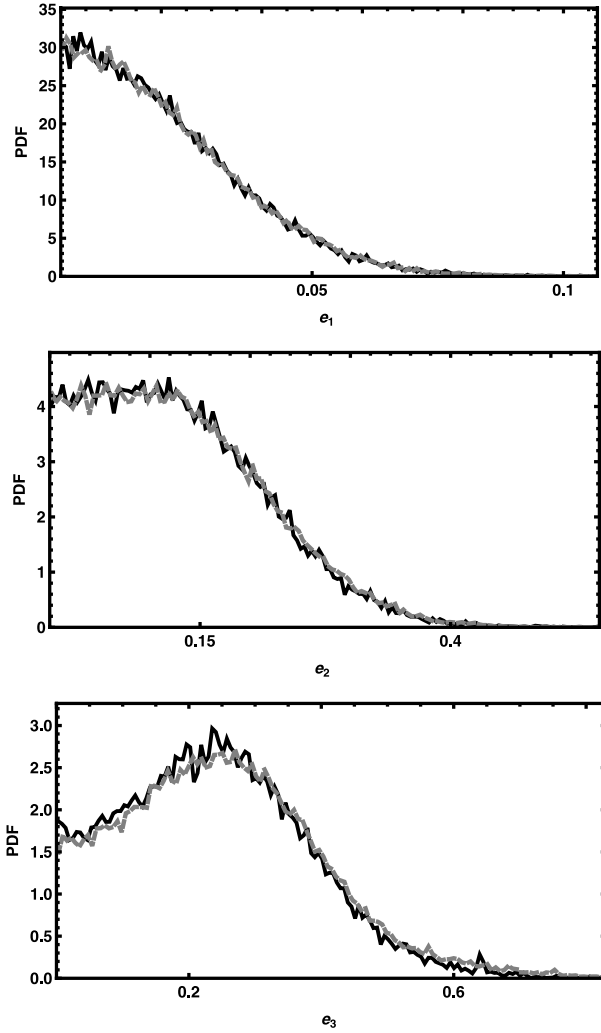


Figure 11. The eccentricity marginals for (a) the noise-induced eccentricity prior (solid black curve) and (b) the uniform prior (grey dashed curve).

average distributions for each parameter using a Gelman–Rubin test. For the five-planet model results, the Gelman–Rubin statistic was ≤ 1.01 .

Fig. 16 shows a plot of a subset of the FCMC parameter marginal distributions for the five-planet fit of the HARPS data after filtering out the post-burn-in FCMC iterations that correspond to the five dominant period peaks at 3.15, 5.37, 12.9, 66.9 and 400 d. The median value of the extra noise parameter $s = 1.16 \text{ m s}^{-1}$.

4.5 Six-planet model

We also carried out a six-planet Kepler periodogram analysis of the HARPS data and the results are shown in Figs 17–20. The best set of parameters from the five-planet fit were used as start parameters and the starting period for the sixth period was set to 36 d. The most probable sixth period found was 34.4 d. A 34.4 d period also appeared as a secondary peak in the five-planet fit and is evident in Fig. 14. Fig. 19 shows a plot of eccentricity versus period for the five-planet FCMC Kepler periodogram.

Fig. 20 shows a plot of a subset of the FCMC parameter marginal distributions for the six-planet fit of the HARPS data after filtering out the post-burn-in FCMC iterations that corre-

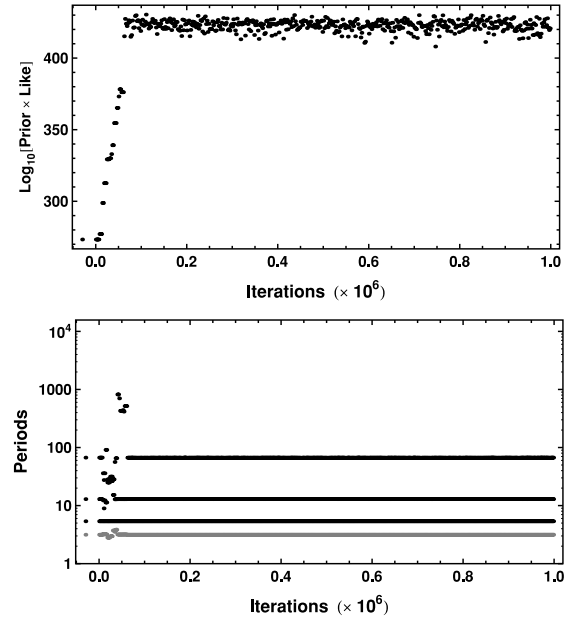


Figure 12. The upper panel is a plot of $\log_{10}(\text{prior} \times \text{likelihood})$ versus iteration for the four-planet FCMC Kepler periodogram of the HARPS data. The lower panel shows the values of the four unknown period parameters versus iteration number.

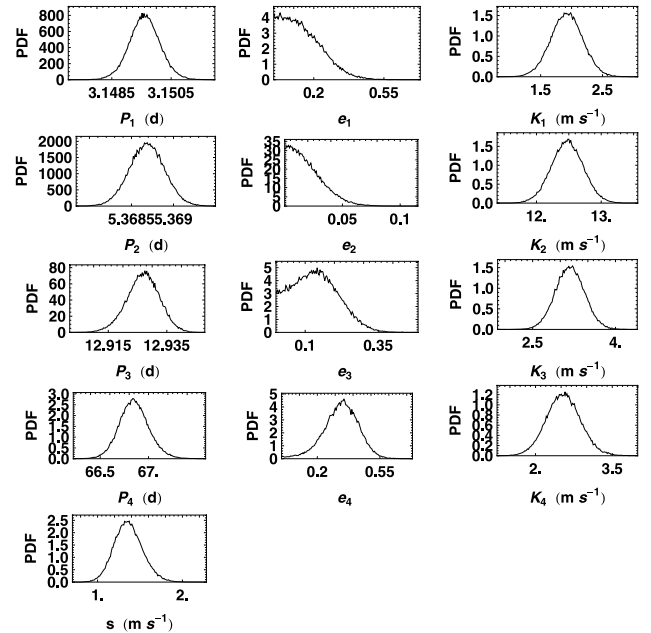


Figure 13. A plot of a subset of the FCMC parameter marginal distributions for a four-planet fit of the HARPS data.

spond to the six dominant period peaks at 3.15, 5.37, 12.9, 34.4, 66.9 and 400 d. The median value of the extra noise parameter $s = 1.00 \text{ m s}^{-1}$. There is considerable agreement with the four- and five-planet marginals shown earlier which leads us to conclude that the 66.9-d orbit is significantly eccentric with an $e \approx 0.34$. There is an indication that $e \approx 0.12$ for the 12.9-d orbit. If the 34.4-d orbit is real, then it would also appear to have a significant eccentricity of $0.49^{+0.22}_{-0.17}$.

Phase plots for the six-planet model are shown in Fig. 21. The top left-hand panel shows the data and model fit versus 3.15 d orbital

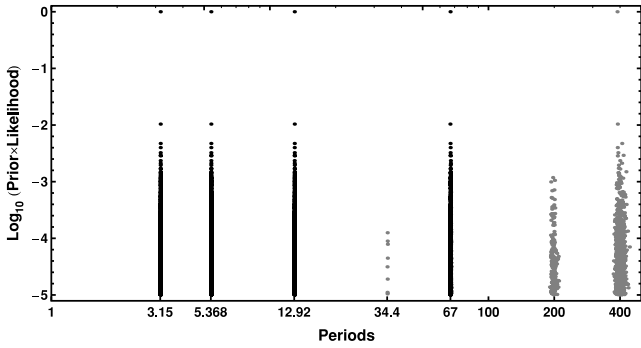


Figure 14. A plot of the five period parameter values versus a normalized value of $\log_{10}[\text{prior} \times \text{likelihood}]$ for the five-planet FCMCMC Kepler periodogram of the HARPS data. The fifth period parameter points are shown in grey.

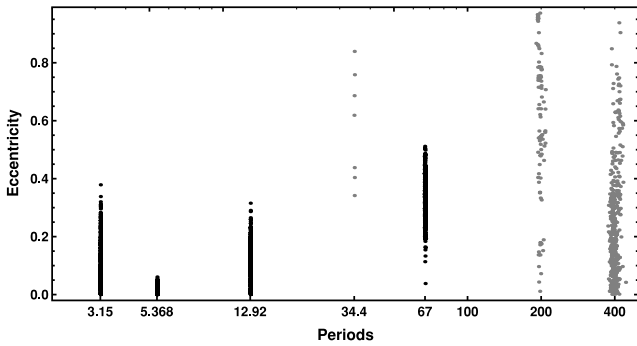


Figure 15. A plot of eccentricity versus period for the five-planet FCMCMC Kepler periodogram of the HARPS data. The fifth period parameter points are shown in grey.

phase after removing the effects of the five other orbital periods. To construct this phase plot, we first filter out the post-burn-in FCMCMC iterations that correspond to the six dominant period peaks at 3.15, 5.37, 12.9, 34.4, 66.9 and 400d. The FCMCMC output for each of these iterations is a vector of the six-planet orbital parameter set plus V . To compute the 3.15-d phase plot data, we subtract the mean velocity curve for the other five planets plus V from the measured set of velocities. This is done by taking a sample of typically 200 FCMCMC iterations and for each iteration we compute the predicted velocity points for that realization of the five-planet plus V parameter set. We then construct the average of these model prediction data sets and subtract that from the data points. These residuals for the set of observation times are converted to residuals versus phase using the mode of the marginal distribution for the 3.15 d period parameter. An orbital phase model velocity fit is then computed at 100 phase points for each realization of the 3.15-d planet parameter set obtained in the same sample of 200 iterations as above. At each of these 100 phase points, we construct the mean model velocity fit and mean ± 1 standard deviation. The upper and lower solid curves in Fig. 21 are the mean FCMCMC model fit ± 1 standard deviations. Thus, 68.3 per cent of the FCMCMC model fits fall between these two curves.

The other panels correspond to a phase plot for the other five periods. In each panel, the quoted period is the mode of the marginal distribution. It is clear that for the 3.15, 5.37, 12.9 and 66.9 d periods the separation of the fit curves is small compared to the amplitude. For the 395 d period phase plot, the wide range of possible orbits that can fit between the upper and lower curves is reflected by the broad extent of the marginal distributions of the parameters.

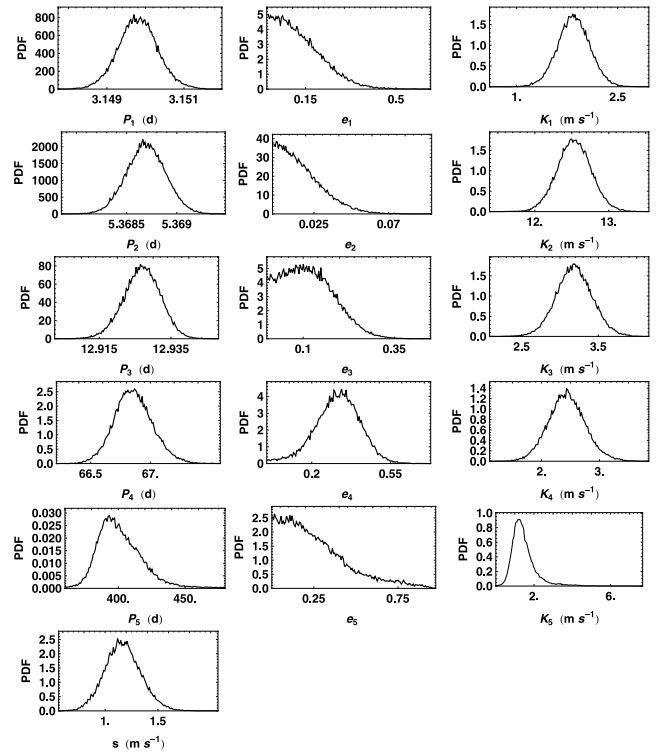


Figure 16. A plot of a subset of the FCMCMC parameter marginal distributions for a five-planet fit of the HARPS data.

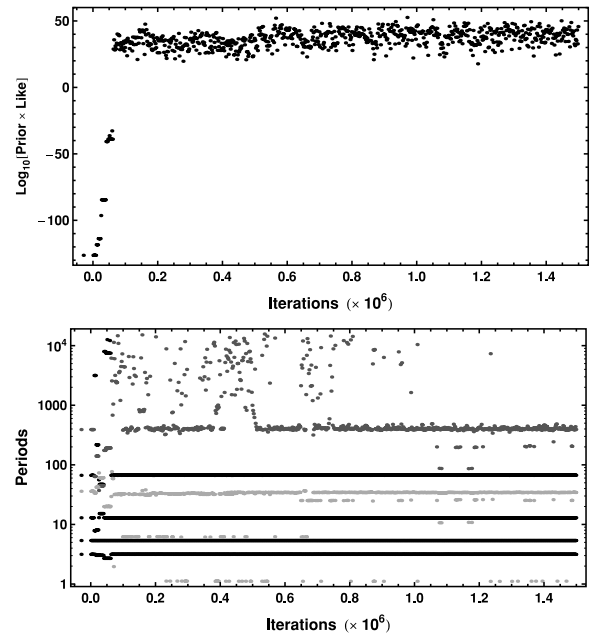


Figure 17. The upper panel is a plot of $\log_{10}(\text{prior} \times \text{likelihood})$ versus iteration for the FCMCMC six-planet fit of the HARPS data. The lower panel shows the values of the six unknown period parameters versus iteration number. The six starting periods are shown on the left-hand side of the plot at a negative iteration number.

5 MODEL SELECTION FOR HARPS ANALYSIS

One of the great strengths of the Bayesian analysis is the built-in Occam's razor. More complicated models contain larger numbers

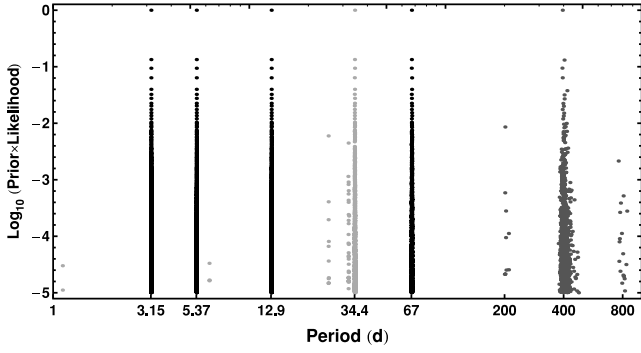


Figure 18. A plot of the six period parameter values versus a normalized value of $\log_{10}(\text{prior} \times \text{likelihood})$ for the six-planet FCMCMC Kepler periodogram of the HARPS data. The fifth period parameter points are shown in dark grey and the sixth in light grey.

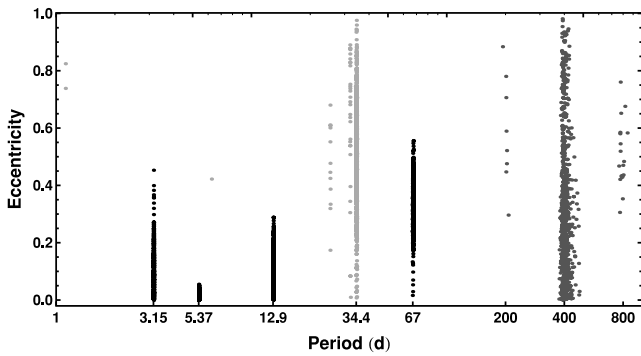


Figure 19. A plot of eccentricity versus period for the six-planet FCMCMC Kepler periodogram of the HARPS data.

of parameters and thus incur a larger Occam penalty, which is automatically incorporated in a Bayesian model selection analysis in a quantitative fashion (see e.g. Gregory 2005a). The analysis yields the relative probability of each of the models explored.

To compare the posterior probability of the i th planet model to that of the four-planet model, we need to evaluate the odds ratio, $O_{i4} = p(M_i|D, I)/p(M_4|D, I)$, the ratio of the posterior probability of model M_i to model M_4 . Application of Bayes' theorem leads to

$$O_{i4} = \frac{p(M_i|I)}{p(M_4|I)} \frac{p(D|M_i, I)}{p(D|M_4, I)} \equiv \frac{p(M_i|I)}{p(M_4|I)} B_{i4}, \quad (10)$$

where the first factor is the prior odds ratio and the second factor is called *Bayes' factor*, B_{i4} . Bayes' factor is the ratio of the marginal (global) likelihoods of the models. The marginal likelihood for model M_i is given by

$$p(D|M_i, I) = \int dX p(X|M_i, I) \times p(D|X, M_i, I). \quad (11)$$

Thus, Bayesian model selection relies on the ratio of marginal likelihoods, not maximum likelihoods. The marginal likelihood is the weighted average of the conditional likelihood, weighted by the prior probability distribution of the model parameters and s . This procedure is referred to as marginalization.

The marginal likelihood can be expressed as the product of the maximum likelihood and the Occam penalty (see Gregory 2005a). Bayes' factor will favour the more complicated model only if the maximum-likelihood ratio is large enough to overcome this penalty. In the simple case of a single parameter with a uniform prior of width

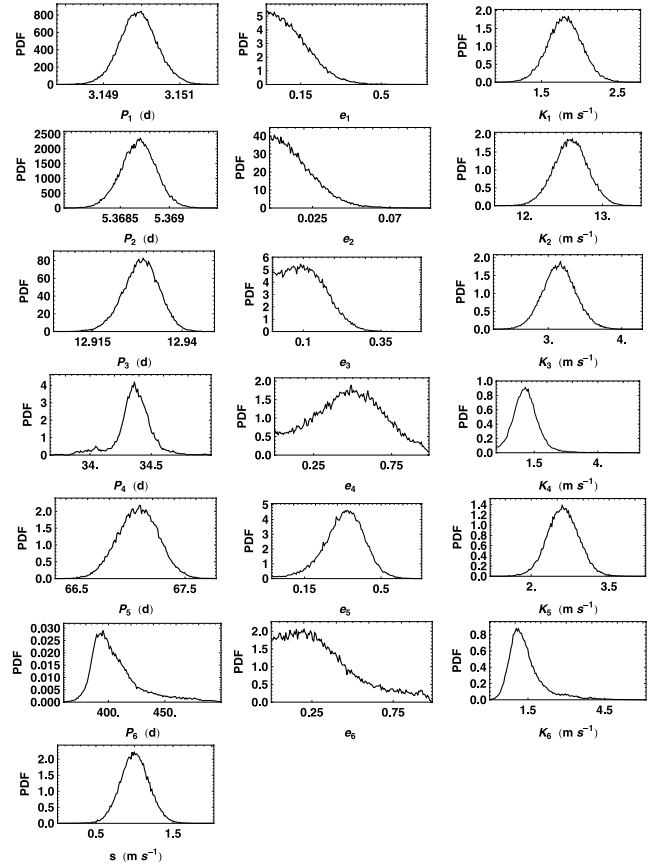


Figure 20. A plot of a subset of the FCMCMC parameter marginal distributions for a six-planet fit to the HARPS data.

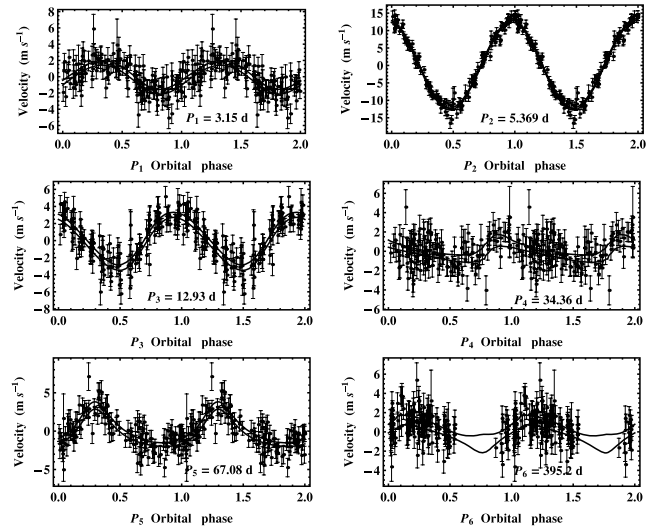


Figure 21. Phase plots for the six-planet model fit to the HARPS data. The top left-hand panel shows the data and model fit versus 3.15 d orbital phase after removing the effects of the five other orbital periods. The upper and lower curves are the mean FCMCMC model fit ± 1 standard deviation. The other five panels correspond to the phase plot for the other five periods.

ΔX , and a centrally peaked likelihood function with characteristic width δX , the Occam factor is $\approx \delta X / \Delta X$. If the data are useful, then generally $\delta X \ll \Delta X$. For a model with m parameters, each parameter will contribute a term to the overall Occam penalty. The

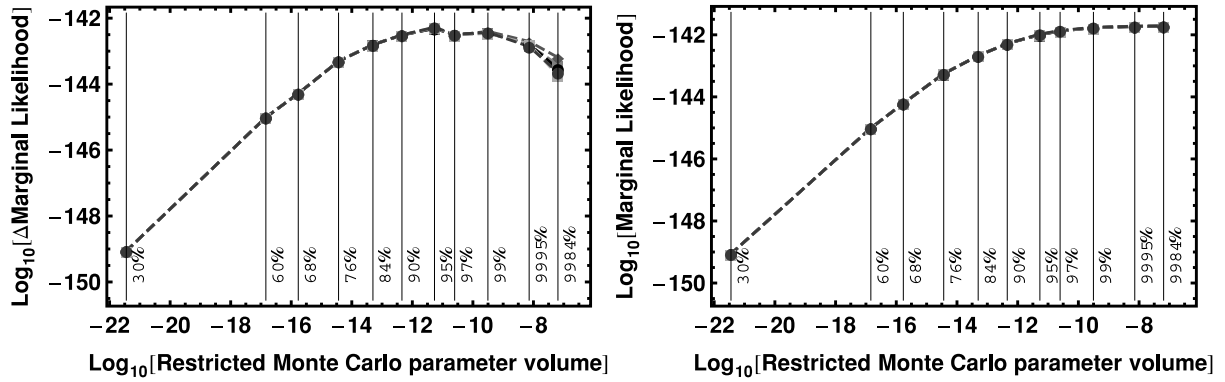


Figure 22. The left-hand panel shows the contribution of the individual nested intervals to the NRMC marginal likelihood for the three-planet model. The right-hand panel shows the integral of these contributions versus the parameter volume of the credible region.

Occam penalty depends not only on the number of parameters, but also on the prior range of each parameter (prior to the current data set, D), as symbolized in this simplified discussion by ΔX . If two models have some parameters in common, then the prior ranges for these parameters will cancel in the calculation of Bayes' factor. To make good use of Bayesian model selection, we need to fully specify priors that are independent of the current data D . The sensitivity of the marginal likelihood to the prior range depends on the shape of the prior and is much greater for a uniform prior than a Jeffreys prior (see e.g. Gregory 2005a). In most instances, we are not particularly interested in the Occam factor itself, but only in the relative probabilities of the competing models as expressed by Bayes' factors. Because the Occam factor arises automatically in the marginalization procedure, its effect will be present in any model selection calculation. Note that no Occam factors arise in parameter estimation problems. Parameter estimation can be viewed as model selection where the competing models have the same complexity so the Occam penalties are identical and cancel out.

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 et al. (2007) reviewed the state of techniques for model selection from a statistical perspective, and Ford & Gregory (2006) have evaluated the performance of a variety of marginal likelihood estimators in the exoplanet context.

Estimating the marginal likelihood is a very big challenge for models with large numbers of parameters, for example, our six-planet model has 32 parameters. In this work, we employ the nested restricted MC (NRMC) method described in Gregory & Fischer (2010) to estimate the marginal likelihoods. MC integration can be very inefficient in exploring the whole prior parameter range because it randomly samples the whole volume. The fraction of the prior volume of parameter space containing significant probability rapidly declines as the number of dimensions increases. For example, if the fractional volume with significant probability is 0.1 in one dimension, then in 32 dimensions the fraction might be of the order of 10^{-32} . In RMC integration, this is much less of a problem because the volume of the parameter space sampled is greatly restricted to a region delineated by the outer borders of the marginal distributions of the parameters for the particular model.

In NRMC integration, multiple boundaries are constructed based on credible regions ranging from 30 to ≥ 99 per cent, as needed. We are then able to compute the contribution to the total integral from each nested interval and sum these contributions. For example, for the interval between the 30 and 60 per cent credible regions, we

generate random parameter samples within the 60 per cent region and reject any sample that falls within the 30 per cent region. Using the remaining samples, we can compute the contribution to the NRMC integral from that interval.

The left-hand panel of Fig. 22 shows the contributions from the individual intervals for five repeats of the NRMC evaluation for the three-planet model. The right-hand panel shows the summation of the individual contributions versus the volume of the credible region. The credible region listed as 99.95 per cent is defined as follows. Let X_{U99} and X_{L99} correspond to the upper and lower boundaries of the 99 per cent credible region, respectively, for any of the parameters. Similarly, let X_{U95} and X_{L95} be the upper and lower boundaries of the 95 per cent credible region for the parameter. Then, $X_{U99.95} = X_{U99} + (X_{U99} - X_{U95})$ and $X_{L99.95} = X_{L99} + (X_{L99} - X_{L95})$. Similarly, $X_{U99.84} = X_{U99} + (X_{U99} - X_{U84})$.

The NRMC method is expected to underestimate the marginal likelihood in higher dimensions and this underestimate is expected to become worse the larger the number of model parameters becomes, that is, increasing number of planets (Gregory 2007c). When we conclude, as we do, that the NRMC-computed odds in favour of the five-planet model compared with the four-planet model is $\sim 10^2$ we mean that the true odds is $\geq 10^2$. Thus, the NRMC method is conservative. One indication of the break-down of the NRMC method is the increased spread in the results for repeated evaluations.

We can readily convert Bayes' factors to a Bayesian false alarm probability (FAP) which we define in equation (12). For example, in the context of claiming the detection of m planets, FAP_m is the probability that there are actually fewer than m planets, that is, $m - 1$ or less.

$$FAP_m = \sum_{i=0}^{m-1} (\text{probability of } i \text{ planets}). \quad (12)$$

If we assume a priori (absence of the data) that all models under consideration are equally likely, then the probability each model is related to Bayes' factors by

$$p(M_i | D, I) = \frac{B_{i4}}{\sum_{j=0}^N B_{j4}}, \quad (13)$$

where N is the maximum number of planets in the hypothesis space under consideration, and of course $B_{44} = 1$. For the purpose of computing FAP_m , we set $N = m$. Substituting Bayes' factors, given in Table 2, into equation (12) gives

$$FAP_5 = \frac{(B_{04} + B_{14} + B_{24} + B_{34} + B_{44})}{\sum_{j=0}^5 B_{j4}} \approx 10^{-2}. \quad (14)$$

Table 2. Marginal likelihood estimates, Bayes' factors relative to model 4 and FAPs. The last two columns list the MAP estimate of the extra noise parameter, s , and the rms residual.

Model	Periods (d)	Marginal likelihood	Bayes' factor nominal	FAP	s (m s ⁻¹)	rms residual (m s ⁻¹)
M_0		6.10×10^{-197}	2.0×10^{-59}			9.8
M_1	(5.37)	$(4.221 \pm 0.003) \times 10^{-155}$	1.4×10^{-17}	1.4×10^{-42}	3.5	3.6
M_2	(5.37, 12.9)	$(1.94 \pm 0.01) \times 10^{-145}$	6.5×10^{-8}	2.2×10^{-10}	2.4	2.6
M_3	(5.37, 12.9, 66.9)	$(3.0^{+0.7}_{-0.5}) \times 10^{-142}$	10^{-4}	6.5×10^{-4}	1.9	2.2
M_4	(3.15, 5.37, 12.9, 66.9)	$(3.0^{+1.1}_{-0.6}) \times 10^{-138}$	1.0	10^{-4}	1.4	1.7
M_5	(3.15, 5.37, 12.9, 66.9, 399)	$(3.0^{+2.1}_{-0.65}) \times 10^{-136}$	10^2	0.01	1.2	1.5
M_6	(3.15, 5.37, 12.9, 34.4, 66.9, 399)	$(6.7^{+2.4}_{-1/3}) \times 10^{-141}$	2.2×10^{-3}	0.999 978	1.0	1.4

Table 3. Five-planet model parameter estimates from HARPS analysis. The value immediately below in parentheses is the MAP estimate.

Parameter	planet 1	planet 2	planet 3	planet 4	planet 5
P (d)	$3.1498^{+0.0005}_{-0.0005}$ (3.149 77)	$5.3687^{+0.0002}_{-0.0002}$ (5.368 66)	$12.927^{+0.006}_{-0.004}$ (12.9316)	$66.85^{+0.15}_{-0.16}$ (66.747)	399^{+14}_{-16} (387.6)
K (m s ⁻¹)	$1.85^{+0.24}_{-0.22}$ (1.93)	$12.53^{+0.23}_{-0.22}$ (12.39)	$3.18^{+0.22}_{-0.24}$ (3.40)	$2.43^{+0.31}_{-0.31}$ (2.75)	$1.3^{+0.4}_{-0.5}$ (1.62)
e	$0.0^{+0.2}_{-0.0}$ (0.197)	$0.00^{+0.02}_{-0.00}$ (0.022)	$0.10^{+0.06}_{-0.10}$ (0.155)	$0.33^{+0.09}_{-0.10}$ (0.38)	$0.02^{+0.30}_{-0.01}$ (0.79)
ω (°)	133^{+81}_{-75} (140)	40^{+98}_{-82} (-1)	234^{+43}_{-43} (234)	334^{+25}_{-23} (326)	281^{+77}_{-100} (310)
a (au)	$0.0285^{+0.0006}_{-0.0006}$ (0.0285)	$0.0406^{+0.0009}_{-0.0009}$ (0.406)	$0.0730^{+0.0016}_{-0.0016}$ (0.730)	$0.218^{+0.005}_{-0.005}$ (0.218)	$0.72^{+0.24}_{-0.24}$ (0.71)
$M \sin i$ (M_E)	$1.91^{+0.26}_{-0.25}$ (1.984)	$15.7^{+0.7}_{-0.7}$ (15.50)	$5.29^{+0.43}_{-0.43}$ (5.63)	$6.7^{+0.8}_{-0.8}$ (7.38)	$6.6^{+2.0}_{-2.7}$ (5.14)
Periastron passage (JD - 245 0000)	$4182.6^{+0.6}_{-0.7}$ (4184)	$4182^{+1.4}_{-1.2}$ (4186)	$4168.9^{+1.6}_{-1.4}$ (4184)	$4137^{+3.5}_{-3.8}$ (4134)	3803^{+82}_{-114} (3828)

For the five-planet model, we obtain a low FAP $\approx 10^{-2}$. The Bayesian FAPs for one-, two-, three-, four-, five- and six-planet models are given in the fourth column of Table 2.

Table 2 gives the NMRC marginal likelihood estimates, Bayes' factors and FAPs for zero-, one-, two-, three-, four-, five- and six-planet models which are designated M_0, \dots, M_6 . The last two columns list the maximum a posteriori (MAP) estimate of the extra noise parameter, s , and the rms residual. For each model, the NRM calculation was repeated five times and the quoted errors give the spread in the results, not the standard deviation. Bayes' factors that appear in the third column are all calculated relative to model 4.

A summary of the five-planet model parameters and their uncertainties are given in Table 3. The quoted value is the median of the marginal probability distribution for the parameter in question (except eccentricity which uses the mode), and the error bars identify the boundaries of the 68.3 per cent credible region.⁸ The value immediately below in parentheses is the MAP estimate, the value at the maximum of the joint posterior probability distribution. It is not

uncommon for the MAP estimate to fall close to the borders of the credible region. In one case, the eccentricity of the fifth planet, the MAP estimate falls well outside the 68.3 per cent credible region which is one reason why we prefer to quote median or mode values as well. The semimajor-axis and $M \sin i$ values are derived from the model parameters, assuming a stellar mass of $0.31 \pm 0.02 M_\odot$ (Delfosse et al. 2000). The quoted errors on the semimajor-axis and $M \sin i$ include the uncertainty in the stellar mass.

Although the NRM estimate of Bayes' factor for the six-planet model is much lower than for either the four- or the five-planet models, we can still infer the orbital parameters of the most probable additional planetary signal in the six-planet fit. The period = 34.4 ± 0.1 d, the eccentricity = $0.49^{+0.22}_{-0.17}$, and the semimajor-axis and $M \sin i$ are 0.140 ± 0.003 au and $2.3^{+0.8}_{-0.7} M_\oplus$, respectively.

6 ANALYSIS OF THE HIRES DATA AND COMBINATION OF HIRES AND HARPS DATA

In this section, we present results on fits to the HIRES data alone and the combination of HIRES and HARPS data. Panel (a) of Fig. 23 shows the combined HIRES (grey points) and HARPS (black points) data for Gl 581. Panel (b) shows a blow-up of a portion of the mean four-planet model fit compared to the data and panel (c) shows the residuals. The same reference time was used for the combined data set as for the HARPS only data. The

⁸ In practice, the probability density for any parameter is represented by a finite list of values, p_i , representing the probability in discrete intervals δX . A simple way to compute the 68.3 per cent credible region, in the case of a marginal with a single peak, is to sort the p_i values in descending order and then sum the values until they approximate 68.3 per cent, keeping track of the upper and lower boundaries of this region as the summation proceeds.

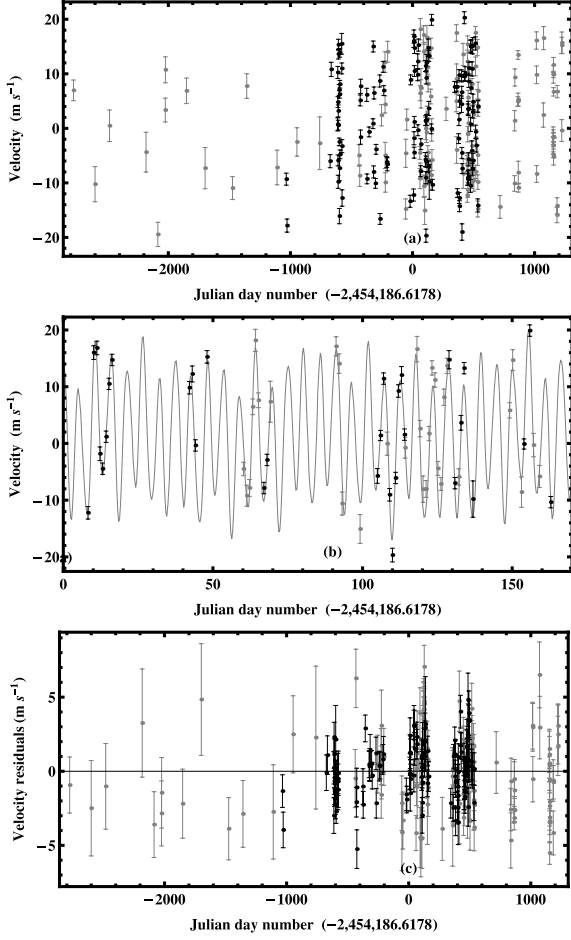


Figure 23. Panel (a) shows the combined HIRES (grey points) and HARPS (black points) data set for Gl 581. Panel (b) shows a blow-up of the mean four-planet model fit compared to the data and panel (c) shows the residuals.

HIRES data (Vogt et al. 2010) consisted of 122 velocity measurements spanning a range of 11 yr and with quoted errors ranging from 0.53 to 4.82 m s⁻¹. Fig. 24 shows a comparison of the velocity differences (HIRES–HARPS) for the nearest pairs of samples versus the sample time-difference. A cluster of 10 points with sample time-differences between 0.2 and 0.3 s is indicated by the ellipse. Since these time-differences are small compared with the shortest known orbital period of 3.15 d, they provide an indication of the

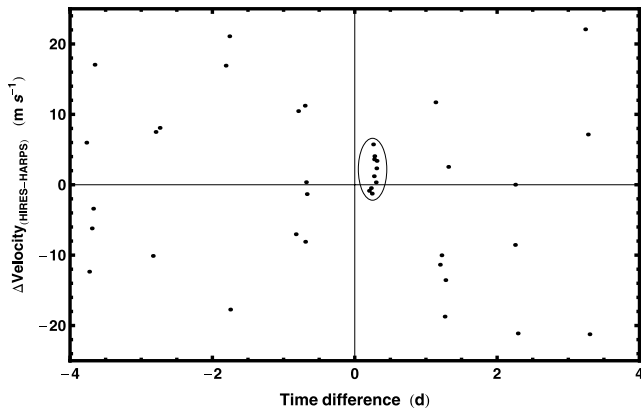


Figure 24. A comparison of the velocity differences (HIRES–HARPS) for the nearest pairs of samples with the sample time-difference.

velocity offset between the two data sets after the removal of the HARPS mean velocity of $-9.208\,0205\text{ km s}^{-1}$. The mean velocity difference for these 10 samples is 1.8 m s⁻¹.

The standard deviation of the velocity differences for these 10 pairs is 2.37 m s⁻¹. For comparison, the mean value of the quoted errors for each pair added in quadrature was 1.94 m s⁻¹.

6.1 Two-planet fit to HIRES data

The results of our two-planet Kepler periodogram analysis are shown in Figs 25–27. The upper panel of Fig. 25 shows a plot of the $\log_{10}(\text{prior} \times \text{likelihood})$ versus FCMC iteration for a two-planet fit of the HIRES data. The lower panel shows the values of the two unknown period parameters versus iteration number. The two starting periods of 5.37 and 12.9 d are shown on the left-hand side of the plot at a negative iteration number. The median value of the extra noise parameter $s = 2.69\text{ m s}^{-1}$ compared to 2.37 m s⁻¹ for the HARPS two-planet fit. Fig. 26 shows the two-planet

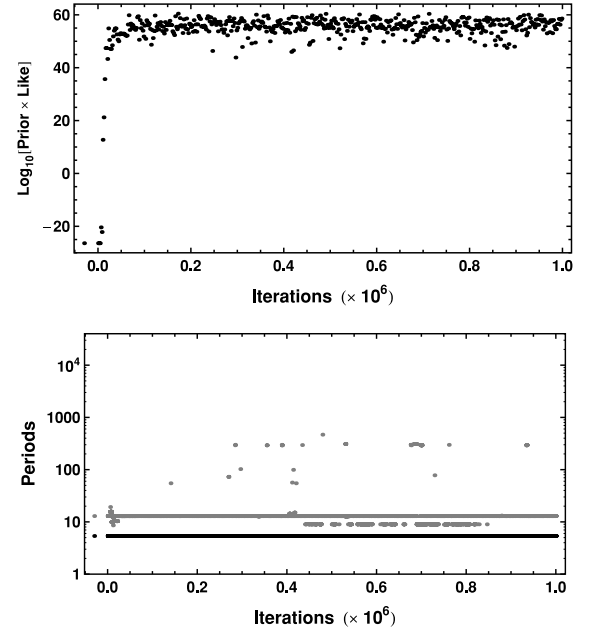


Figure 25. The upper panel is a plot of $\log_{10}(\text{prior} \times \text{likelihood})$ versus iteration for a sample of the FCMC two-planet fit of the HIRES data. The lower panel shows values of the two unknown period parameters versus iteration number. The two starting periods of 5.37 and 12.9 d are shown on the left-hand side of the plot at a negative iteration number.

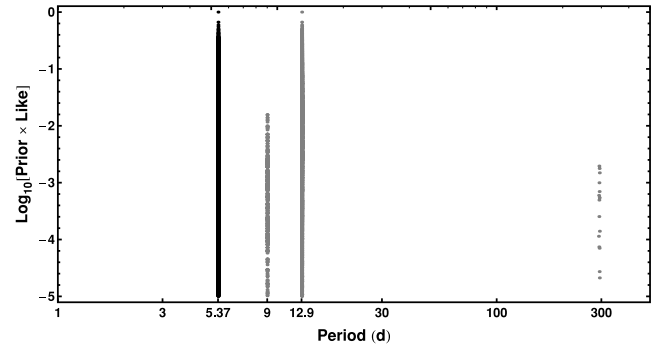


Figure 26. A plot of the two period parameter values versus a normalized value of $\log_{10}(\text{prior} \times \text{likelihood})$, the two-planet Kepler periodogram of the HIRES data.

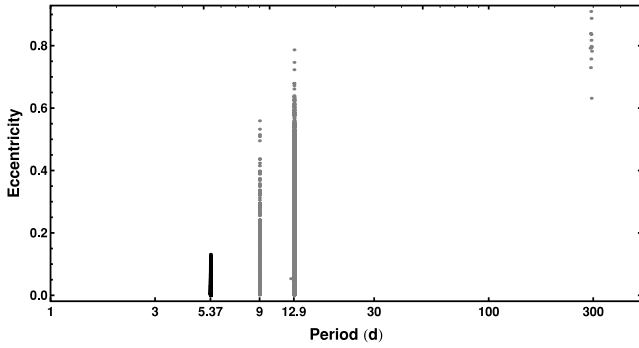


Figure 27. A plot of eccentricity versus period for the two-planet FCMC fit of the HIRES data.

Kepler periodogram. Only values within five decades of the maximum $\log_{10}(\text{prior} \times \text{likelihood})$ are plotted but without regards to whether the values occurred before or after burn-in. Two prominent periods were clearly detected: 5.37 and 12.9 d. The second period parameter exhibited two other peaks but these were significantly less probable. The most probable of these has a period of ~ 9 d.

Fig. 27 shows a plot of eccentricity versus period for a sample of the FCMC parameters for the two-planet model. The dominant 5.37- and 12.9-d peaks and the weaker 9-d peak allow for low-eccentricity orbits. The peak around 300 d has a high value of eccentricity typical of noise.

6.2 Three-planet fit to HIRES data

Two three-planet runs were carried out on the HIRES-only data starting with the best periods (5.37, 12.9 and 66.9 d) found from the HARPS analysis, but only the 5.37 d period (largest amplitude) was successfully detected. The best of these two runs (Fig. 28) detected

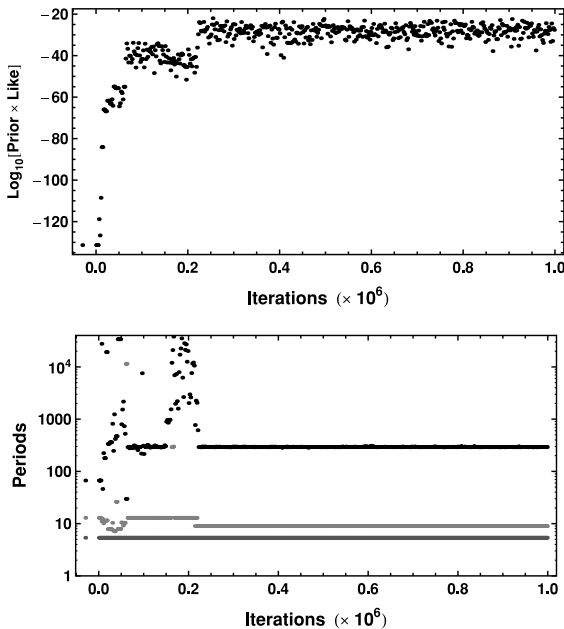


Figure 28. The upper panel is a plot of $\log_{10}(\text{prior} \times \text{likelihood})$ versus iteration for a sample of the FCMC three-planet fit of the HIRES data. The lower panel shows the values of the three unknown period parameters versus iteration number. The starting periods of 5.37, 12.9 and 66.9 d are shown on the left-hand side of the plot at a negative iteration number.

three dominant periods of 5.37, 8.99 and ~ 300 d. A much weaker peak was found at a period of 12.9 d. The HIRES fit extra noise parameter was $s = 2.2 \text{ m s}^{-1}$ compared to the HARPS three-planet fit where $s = 1.7 \text{ m s}^{-1}$. Fig. 29 shows a plot of the three-planet Kepler periodogram. Only values within five decades of the maximum $\log_{10}(\text{prior} \times \text{likelihood})$ are plotted but without regards to whether the values occurred before or after burn-in. Three prominent periods were clearly detected: 5.37, 9 and 300 d. The second period parameter (shown in grey) exhibited a second much less probable peak at 12.9 d and the third period parameter (black) exhibited many weak peaks.

Fig. 30 shows a plot of eccentricity versus period for a sample of the FCMC parameters for the three-planet model. The 5.37- and 9-d peaks exhibit low-eccentricity orbits. The peak around 300 d has a high value of eccentricity.

Fig. 31 shows a blow-up of the above eccentricity versus period plot in the vicinity of the 300-d peak which is dominated by two high-eccentricity features typical of noise. We conclude that there is no clear evidence for a third period in the HIRES data alone and suspect that the presence of the strong high-eccentricity 300-d complex may contribute to the dominance of the 9 d period over the 12.9 d period found in the two-planet fit.

For both the two- and three-planet fits, the extra noise parameter is larger for the HIRES data than for the HARPS data. One possibility is that the quoted HIRES errors have been systematically underestimated which we can model by an extra Gaussian noise term added in quadrature to the quoted HIRES errors with $\sigma = ds_{\text{HIRES}}$. We can obtain a crude estimate of ds_{HIRES} from the

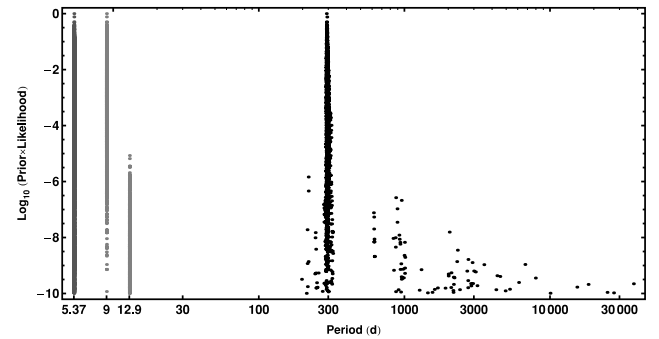


Figure 29. A plot of the three period parameter values versus a normalized value of $\log_{10}(\text{prior} \times \text{likelihood})$, the three-planet Kepler periodogram of the HIRES data.

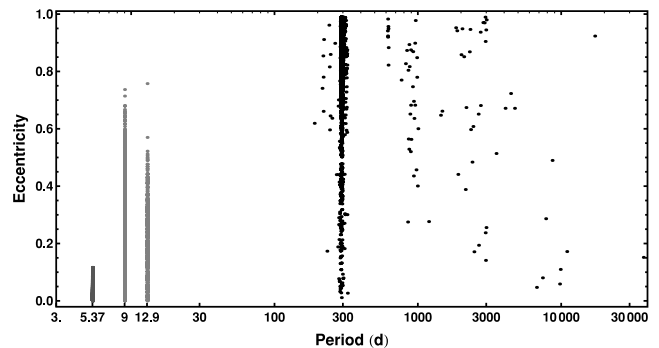


Figure 30. A plot of eccentricity versus period for the three-planet FCMC fit of the HIRES data.

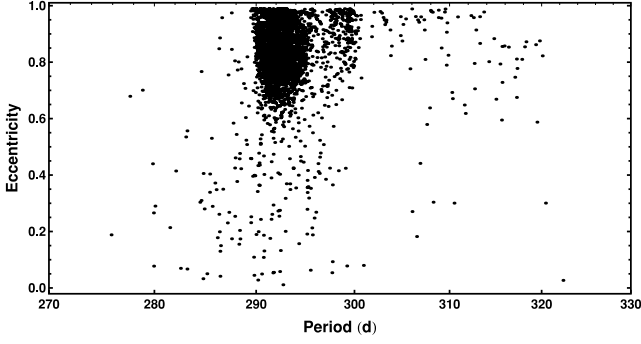


Figure 31. A blow-up of the above eccentricity versus period plot in the vicinity of the 300-d peak.

HIRES and HARPS s parameter values for the two-planet case where both analyses yielded the same two periods. The result is $ds_{\text{HIRES}} = \sqrt{2.69^2 - 2.37^2} = 1.3 \text{ m s}^{-1}$. This suggests that in the analysis of the combined data set, we should include an extra Gaussian noise term added in quadrature to the quoted HIRES errors with σ , labelled ds_{HIRES} , as an additional unknown parameter.

6.3 Two-planet fit to the combined HIRES/HARPS data

Based on the above results, we decided to use the following noise model for the j th data point for the combined HIRES/HARPS two-planet analysis:

$$\sigma_{j/\text{HARPS}} = \sqrt{\sigma_{j/\text{quoted}}^2 + s^2}, \quad (15)$$

$$\sigma_{j/\text{HIRES}} = \sqrt{\sigma_{j/\text{quoted}}^2 + ds_{\text{HIRES}}^2 + s^2}. \quad (16)$$

The best two-planet orbital parameters were employed as start coordinates for the combined HIRES/HARPS analysis. We also incorporated an additional unknown parameter dc to allow for a possible difference in the constant velocity offsets of the HIRES and HARPS data. Based on our analysis of Fig. 24, our best estimate of $dc \approx 1.8 \text{ m s}^{-1}$. We assumed a Gaussian prior⁹ with zero mean and $\sigma = 3 \text{ m s}^{-1}$.

The two-planet FCMC fit of the HIRES/HARPS data confirmed the 5.37 and 12.9 d periods. Fig. 32 shows a plot of a subset of the FCMC parameter marginal distributions for the two-planet fit of the data after filtering out the post-burn-in FCMC iterations that correspond to the two dominant period peaks at 5.37 and 12.9 d. The bottom row shows the marginals for dc , ds_{HIRES} and s . The maximum a posteriori and median values found for dc are 1.65 and 1.67 m s^{-1} , respectively, that is, very close to the crude estimate of 1.8 m s^{-1} made earlier.

The upper panel of Fig. 33 shows the marginal distribution for the unknown ds_{HIRES} parameter in the two-planet FCMC fit of the combined HIRES/HARPS data. The black curve in the lower panel shows the marginal distribution for the common extra noise parameter s in the HIRES/HARPS fit. The light grey curve is the

⁹ Strictly speaking, the Bayesian choice of the prior for the offset velocity should be independent of the data. As a check, we carried out a two-planet fit to the combined HIRES/HARPS data using a Gaussian prior with a much larger $\sigma = 1 \text{ km s}^{-1}$. No significant difference was detected.

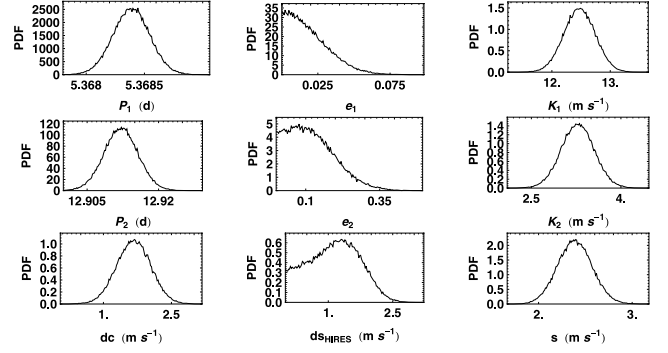


Figure 32. A plot of a subset of the FCMC parameter marginal distributions for a two-planet fit of the combined HIRES/HARPS data.

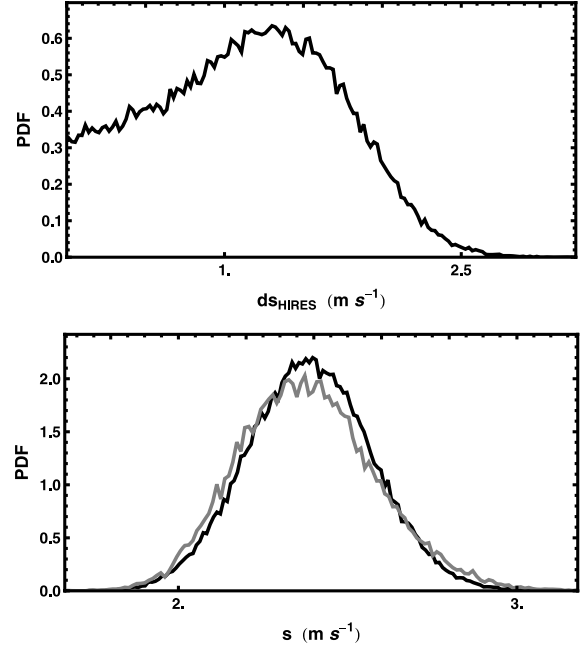


Figure 33. The upper panel shows the marginal distribution for the unknown ds_{HIRES} parameter in the two-planet FCMC fit of the combined HIRES/HARPS data. The black curve in the lower panel shows the marginal distribution for the common extra noise parameter s in the HIRES/HARPS fit. The light-grey curve is the same quantity for the two-planet HARPS-only fit.

same quantity for the two-planet HARPS-only fit. Clearly there is very good agreement between the two s parameter estimates.

6.4 Three-planet fit to the combined HIRES/HARPS data

For the three-planet combined analysis, we employed the same noise model as used in Section 6.3. Again the best three-planet orbital parameters from the HARPS analysis were employed as start coordinates for the combined HIRES/HARPS analysis. All three period parameters were free to roam within a search range extending from 1.1 d to 10 times the data duration.

Fig. 34 shows a plot of a sample of the FCMC three-planet periodogram. Only values within five decades of the maximum $\log_{10}(\text{prior} \times \text{likelihood})$ are plotted but without regards to whether the values occurred before or after burn-in. Three prominent periods were clearly detected: 5.37, 12.9 and 66.9 d. The third period parameter exhibited other peaks but these were significantly less

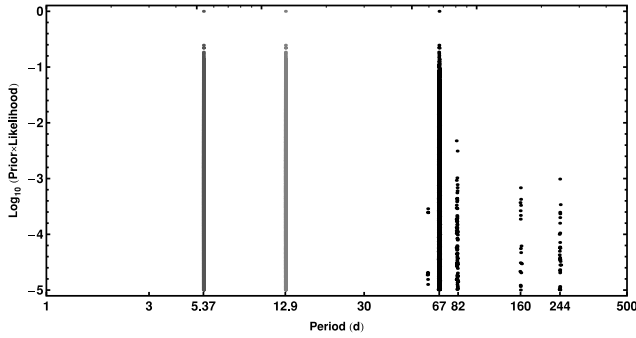


Figure 34. A plot of the three period parameter values versus a normalized value of $\log_{10}(\text{prior} \times \text{likelihood})$ for the three-planet FCMCMC Kepler fit of the combined HIRES/HARPS data.

probable, the one at 82 d is consistent with being a 1-yr alias of the 66.9 d period.

Fig. 35 shows a plot of eccentricity versus period for a sample of the FCMCMC parameters for the three-planet model. The dominant 5.37-, 12.9- and 66.9-d peaks allow for low-eccentricity orbits.

Fig. 36 shows a plot of a subset of the FCMCMC parameter marginal distributions for the three-planet fit of the data after filtering out the post-burn-in FCMCMC iterations that correspond to the three dominant period peaks at 5.37, 12.9 and 66.9 d. The bottom row shows the marginals for dc , ds_{HIRES} and s . ds_{HIRES} is more accurately defined than in the two-planet analysis. The maximum a pos-

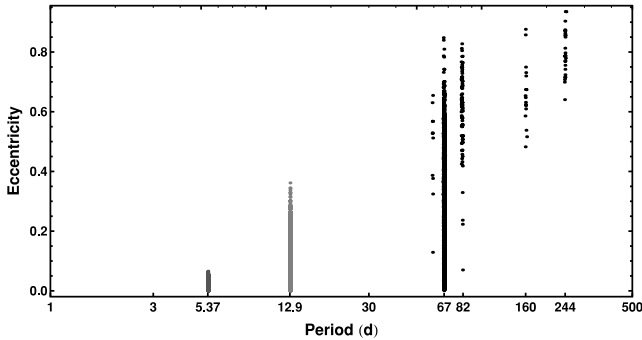


Figure 35. A plot of eccentricity versus period for the three-planet FCMCMC fit of the combined HIRES/HARPS data.

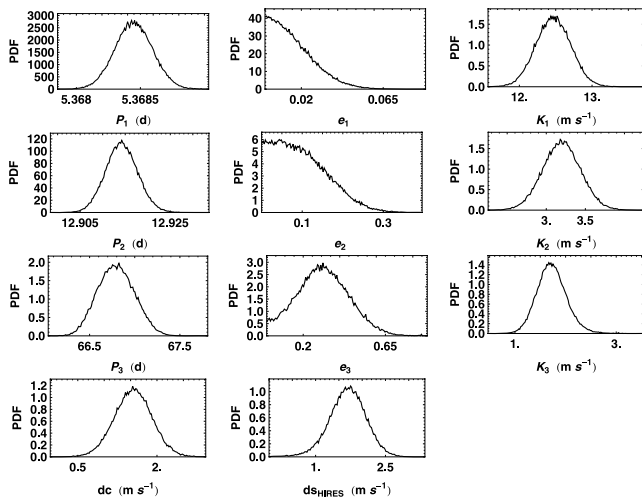


Figure 36. A plot of a subset of the FCMCMC parameter marginal distributions for a three-planet fit of the combined HIRES/HARPS data.

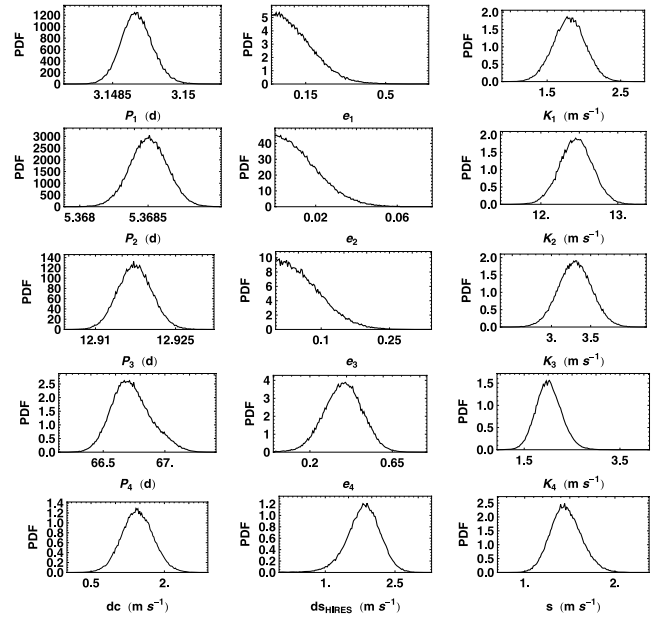


Figure 37. A plot of a subset of the FCMCMC parameter marginal distributions for a four-planet fit of the combined HIRES/HARPS data.

teriori and median values found for ds_{HIRES} are 1.67 and 1.55 m s^{-1} , respectively.

6.5 Four-planet fit to the combined HIRES/HARPS data

The best four-planet orbital parameters from the HARPS-only analysis were employed as start coordinates for the combined HIRES/HARPS analysis. As before we incorporated the same noise model as used in Section 6.3 and the unknown parameter dc to allow for a possible difference in the constant velocity offsets of the HIRES and HARPS data.

The four-planet Kepler periodogram found the four starting periods of 3.15, 5.37, 12.9 and 66.9 d and no other peaks. Fig. 37 shows the marginal a posteriori densities of a subset of the parameters.

The overall span of the HIRES data is 11 yr, while the HARPS data interval is only 4.3 yr. Perhaps the effective stellar jitter is significantly larger on the 11-yr span and this contributes to the ds_{HIRES} term. To investigate this, we constructed a new joint HIRES/HARPS data set which only included the 79 HIRES measurements that occurred within the span of the HARPS data.

Table 4 shows a comparison of the estimates of the dc , ds_{HIRES} and s parameters from the two-, three-, and four-planet fits to the full and partial joint HIRES/HARPS data sets. The parameter value listed is

Table 4. Bayesian estimates of the parameters dc , ds_{HIRES} and s from the two-, three- and four-planet fits to the full and partial joint HIRES/HARPS data sets. The value immediately below in parentheses is the MAP estimate.

Parameter	Two-planet full data	Three-planet full data	Four-planet full data	Four-planet partial data
dc (m s^{-1})	$1.67^{+0.40}_{-0.35}$ (1.64)	$1.54^{+0.35}_{-0.36}$ (1.63)	$1.45^{+0.32}_{-0.31}$ (1.53)	$1.99^{+0.36}_{-0.36}$ (1.5)
ds_{HIRES} (m s^{-1})	$1.15^{+1.05}_{-0.35}$ (0.75)	$1.55^{+0.58}_{-0.47}$ (1.67)	$1.84^{+0.35}_{-0.33}$ (1.85)	$1.76^{+0.45}_{-0.39}$ (1.70)
s (m s^{-1})	$2.39^{+0.25}_{-0.11}$ (2.32)	$2.0^{+0.2}_{-0.2}$ (1.83)	$1.45^{+0.17}_{-0.17}$ (1.34)	$1.39^{+0.15}_{-0.18}$ (1.33)

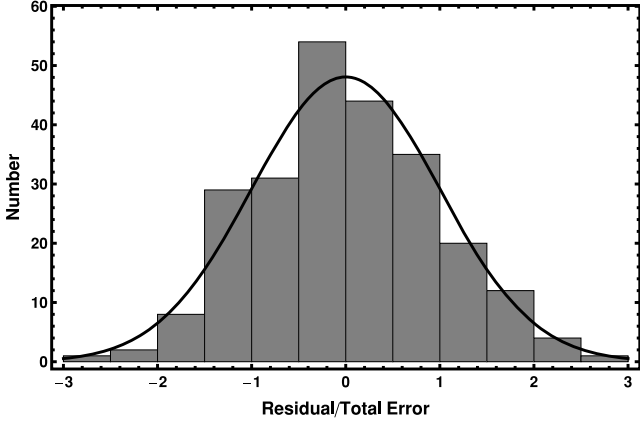


Figure 38. A histogram of the r_{1j} values compared to a suitably normalized Gaussian with zero mean and standard deviation = 1.

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 MAP estimate, the value at the maximum of the joint posterior probability distribution. It is clear from the table that the four-planet fit to the partial joint data set yielded distributions for ds_{HIRES} and s that are very similar to those obtained using the full joint data set. As expected, the extra noise parameter s decreases as the number of planets increase. The estimates of the offset parameter, dc , and the HIRES extra noise parameter, ds_{HIRES} , agree within the quoted uncertainties and these uncertainties are smallest for the four-planet model fit.

We carried out an analysis of the four-planet normalized fit residuals (r_{1j}) to the full joint data set. We define r_{1j} as

$$r_{1j} = \frac{1}{n_i} \sum_{i=1}^{n_i} \frac{\text{residual}_{ij}}{\sqrt{\text{error}_j^2 + ds_{\text{HIRES}_i}^2 + s_i^2}}, \quad (17)$$

where j is an index for the combined HIRES/HARPS data set and n_i is the number of post-burn-in FMCMC equilibrium samples used in computing the mean value of residuals/(effective noise σ) for each measurement (typically we use $n_i \sim 200$). The effective noise σ is given by

$$\text{effective noise } \sigma = \sqrt{\text{error}_j^2 + ds_{\text{HIRES}_i}^2 + s_i^2}, \quad (18)$$

which is the quadrature sum of the quoted error, ds_{HIRES_i} , and the extra noise term, s_i . Of course, ds_{HIRES_i} only contributes to the effective noise σ of the HIRES data values.

Fig. 38 shows a histogram of the r_{1j} values compared to a suitably normalized Gaussian with zero mean and standard deviation = 1. Within the uncertainties, r_{1j} is consistent with a Gaussian distribution. The reduced χ^2 of r_{1j} is given as

$$\chi_1^2 = \frac{1}{n_t - n_p} \sum_{j=1}^{n_t} r_{1j}^2 = 1.008, \quad (19)$$

where n_t is the total number of combined HIRES/HARPS data points and $n_p = 24$ is the number of fit parameters.

The four-period phase plots are shown in Fig. 39. The top left-hand panel shows the data and model fit versus 3.15 d orbital phase after removing the effects of the three other orbital periods. The upper and lower solid curves are the mean FMCMC model fit ± 1 standard deviation. The HIRES data points are shown in grey and the error bars are the quoted errors added in quadrature with our median

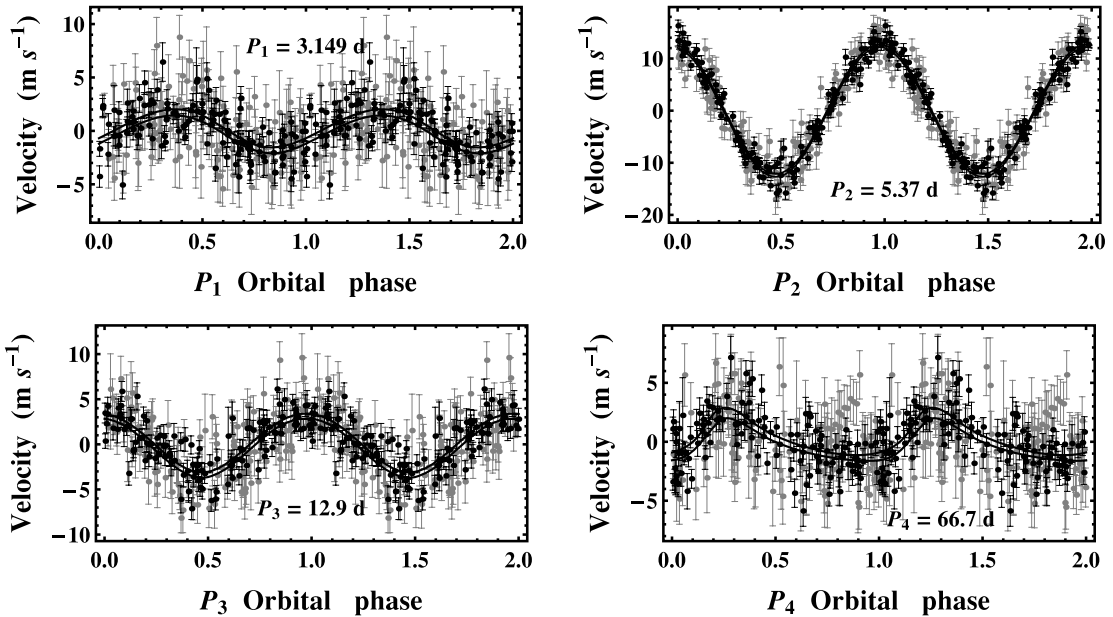


Figure 39. Phase plots for the four most probable periods derived from the combined HIRES/HARPS data. The top left-hand panel shows the data and model fit versus 3.15 d orbital phase after removing the effects of the five other orbital periods. The upper and lower curves are the mean FMCMC model fit ± 1 standard deviation. The other three panels correspond to a phase plot for the other three periods. The HARPS data points are shown in black and the HIRES data points are shown in grey.

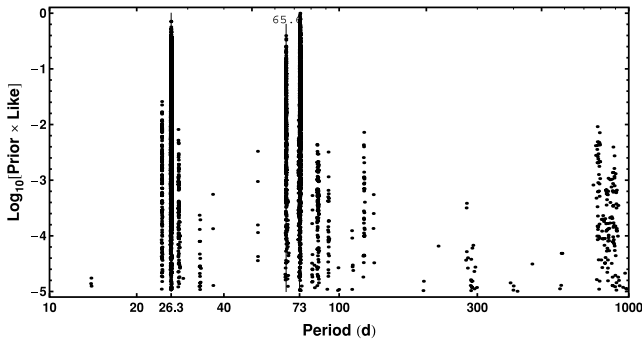


Figure 40. A plot of the period parameter versus a normalized value of $\log_{10}(\text{prior} \times \text{likelihood})$ for the one-planet FCMC Kepler fit to the HIRES residuals after removing the 3.15-, 5.37- and 12.9-d orbits.

estimate of $ds_{\text{HIRES}} = 1.84 \text{ m s}^{-1}$. The other panels correspond to the phase plot for the other three periods. Examination of the 66.9 d period phase plot indicates that the HIRES data do not lend much support to the 66.9 d period. In fact, around a phase of 0.8, there are a series of ~ 11 points that exhibit a systematic trend away from the mean light curve. To examine this further, we carried out a one-planet FCMC fit to the HIRES residuals after subtracting off the 3.15, 5.37 and 12.9 d mean orbits. Fig. 40 shows the one-planet Kepler periodogram of these residuals. Three prominent peaks are present at 26.3, 65.6 and 73 d along with many minor peaks. The thin solid vertical lines highlight these peaks. The strongest peak has a period of 73 d which explains the absence of good support for a 66.9 d period in Fig. 39.

6.6 Five-planet fit to the combined HIRES/HARPS data

Figs 41 and 42 show the results of a five-planet fit to the combined HIRES/HARPS data set. The evidence for a ~ 400 d period is more confused than the HARPS-only results shown in Figs 14 and 15. The combined HIRES/HARPS results show a dominant high-eccentricity peak with a period of 472 d. Again, weak high-eccentricity peaks are a common characteristic of noise.

7 DISCUSSION

Our analysis of the combined HIRES/HARPS data set argues strongly that the HIRES errors have been systematically underestimated. If we model this by the additive extra noise term ds_{HIRES} , as in equation (20), we conclude that $ds_{\text{HIRES}} = 1.84^{+0.35}_{-0.33}$. An alternative

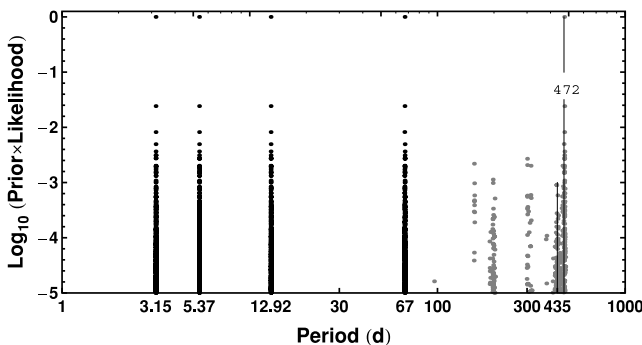


Figure 41. A plot of the five period parameters versus a normalized value of $\log_{10}(\text{prior} \times \text{likelihood})$ for the five-planet FCMC Kepler fit to the combined HIRE/SHARPS data.

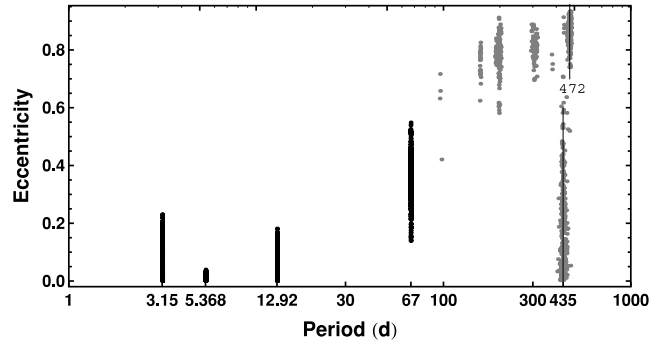


Figure 42. A plot of eccentricity versus period for the five-planet FCMC fit to the combined HIRE/SHARPS data.

possibility that we can test is that the HIRES data are systematically too low by a common factor. To check this out, we again carried out the four-planet analysis of the combined HIRES/HARPS data using the following noise model:

$$\sigma_{\text{HIRES}_j} = \sqrt{(b_{\text{HIRES}} \times \sigma_{\text{quoted}_j})^2 + s^2}. \quad (20)$$

We assumed a uniform prior for the unknown parameter b_{HIRES} in the range 0.5–4.0. The results were qualitatively very similar to the results obtained with the ds_{HIRES} noise term. Following similar calculations to those outlined in Section 6.5, we computed the reduced χ^2 of the residuals divided by the total effective noise σ and obtained a value of 1.009, indistinguishable from the reduced χ^2 obtained using the ds_{HIRES} parameter. Fig. 48 shows the marginal probability distributions for the b_{HIRES} and s parameters. The dashed curve in the lower panel is the marginal for s from the four-planet fit to the HARPS-only data. Employing the b_{HIRES} parametrization results in a significantly larger estimate in the s term than was required by the HARPS-only analysis or the ds_{HIRES} parametrization of the combined HIRES/HARPS data.

Although the two different parametrizations of the HIRES extra noise lead to similar values of the reduced χ^2 , we can gain additional insight into which parametrization is better from a more microscopic exploration by binning the four-planet fit residuals/(effective noise) and examining the χ^2 values of the individual bins. If we add two χ^2 random variables, one of which has ν_1 degrees of freedom and the other has ν_2 degrees of freedom, then their sum will be χ^2 with $\nu_1 + \nu_2$ degrees of freedom (see e.g. Gregory 2005a). Similarly, we can take the quantity r_{1j} of equation (19) which is χ^2 with $(n_t - n_p)$ degrees of freedom and divide the r_{1j} values into k bins. Suppose the k th bin has $n_k r_{1j}$ values, then χ^2 of these values will be given by

$$\chi_k^2 = \frac{1}{(n_k/n_t) \times (n_t - n_p)} \sum_{j=1}^{n_k} r_{1j}^2, \quad (21)$$

which is χ^2 with $(n_k/n_t) \times (n_t - n_p)$ degrees of freedom.

Fig. 44 shows a plot of the χ^2 statistic versus the original quoted errors that have been binned into 0.5 m s^{-1} bins for the four-planet fit to the combined HIRES/HARPS data using the ds_{HIRES} parametrization. Fig. 45 shows a similar plot for the b_{HIRES} parametrization. Ideally, the χ^2 distribution would be flat with a value of 1.0. It is clear that the ds_{HIRES} parametrization achieves a flatter distribution than the b_{HIRES} parametrization. Further the noise model involving ds_{HIRES} leads to a jitter noise estimate s which is much closer to the value expected from the HARPS-only four-planet analysis.

Our HARPS-only analysis provides evidence for five planetary signals while incorporating the HIRES data seems to degrade the evidence for both the 66.9 and ~ 400 d periods even when we allow

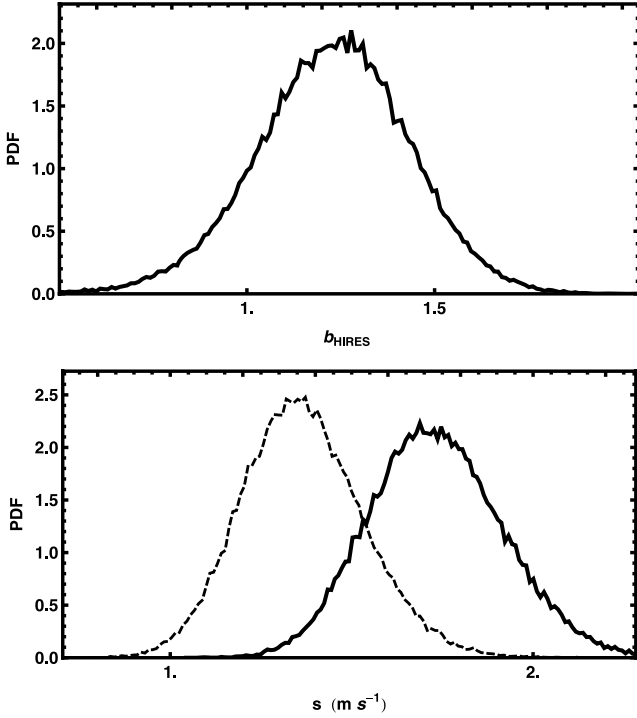


Figure 43. A plot of the marginal probability distributions for the b_{HIRES} and s parameters for a four-planet FCMC fit to the combined HIRES/HARPS data using the noise model of equation (20). The dashed curve in the lower panel is the marginal for s from the four-planet fit to the HARPS only data.

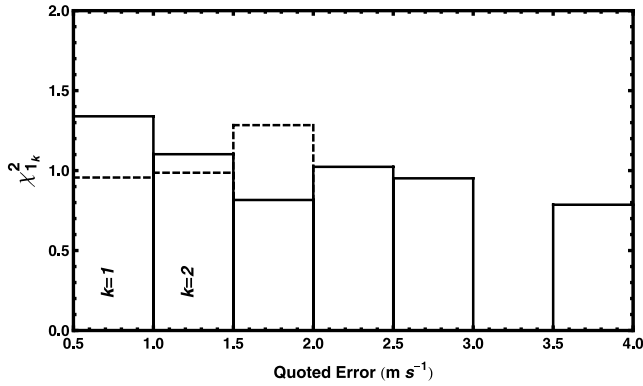


Figure 44. A plot of the χ^2_k statistic from equation (21) versus binned values of the quoted errors for the four-planet fit of the combined HIRES/HARPS data using the ds_{HIRES} parametrization. The solid and dashed histograms show the binned HIRES data and binned HARPS data, respectively.

for an extra HIRES noise term of the order of 1.8 m s^{-1} . We suspect this extra noise term may arise from as-yet-unidentified systematics which may be the reason for the degradation in the quality of fits beyond three planets. Alternatively, could unidentified HARPS systematics be responsible for the extra periods evident in their lower noise data?

8 CONCLUSIONS

A Bayesian re-analysis of published HARPS and HIRES precision RV data for Gl 581 was carried out with a multiplanet Kepler periodogram (from one to six planets) based on our FCMC algorithm. In all cases, the analysis included an unknown parametrized stel-

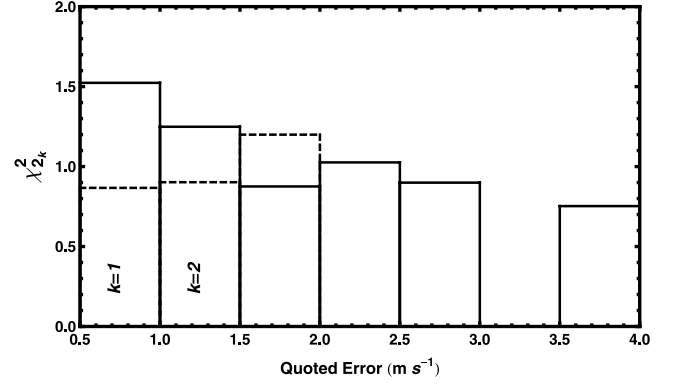


Figure 45. A plot of the χ^2_k statistic from equation (21) versus binned values of the quoted errors for the four-planet fit of the combined HIRES/HARPS data using the b_{HIRES} parametrization. The solid and dashed histograms show the binned HIRES data and binned HARPS data, respectively.

lar jitter noise term. For the HARPS data set, the most probable number of planetary signals detected is 5. The Bayesian FAP for the five-planet model is 0.01. The five periods include the 3.15, 5.37, 12.9 and 66.9 d periods reported previously plus a 399^{+14}_{-16} d period. The orbits of four out of the five planets are consistent with low-eccentricity orbits, the exception being the 66.9-d orbit where $e = 0.33^{+0.09}_{-0.10}$. The semimajor-axis and $M \sin i$ of the five planets are $(0.0285 \pm 0.0006 \text{ au}, 1.91^{+0.26}_{-0.25} M_{\oplus})$, $(0.0406 \pm 0.0009 \text{ au}, 15.7^{+0.7}_{-0.7} M_{\oplus})$, $(0.0730 \pm 0.0016 \text{ au}, 5.29 \pm 0.43 M_{\oplus})$, $(0.218 \pm 0.005 \text{ au}, 6.7 \pm 0.8 M_{\oplus})$ and $(0.72 \pm 0.24 \text{ au}, 6.6^{+2.0}_{-2.7} M_{\oplus})$, respectively.

In light of the Vogt et al. (2010) report of a sixth companion with a period of 36.6 d, we carried out a six-planet fit to the HARPS data which detected multiple period possibilities. The strongest of these, with a period = 34.4 ± 0.1 d and eccentricity of $0.49^{+0.22}_{-0.17}$, had a peak $\log_{10}(\text{prior} \times \text{likelihood})$ 100 times larger than the others. The inferred semimajor-axis and $M \sin i$ are $0.140 \pm 0.003 \text{ au}$ and $2.3^{+0.8}_{-0.7} M_{\oplus}$, respectively. The Bayesian FAP for the six-planet model is extremely large, 0.999 978, so we are unable to support any claim for a sixth companion on the basis of the current data.

The analysis of the HIRES data set yielded a reliable detection of only the strongest 5.37 and 12.9 d periods. The analysis of the combined HIRES/HARPS data again only reliably detected the 5.37 and 12.9 d periods. Detection of four planetary signals with the periods of 3.15, 5.37, 12.9, and 66.9 d was only achieved by including an additional unknown but parametrized Gaussian error term added in quadrature to the HIRES quoted errors. The marginal probability density of σ for this additional HIRES Gaussian noise term has a well-defined peak at $1.84^{+0.35}_{-0.33} \text{ m s}^{-1}$. Phase plots indicate that incorporating the HIRES data seems to degrade the evidence for the 66.9 and ~ 400 d periods even when we allow for the extra HIRES noise term. We suspect this extra noise term may arise from unidentified systematics which may be the reason for the degradation in the quality of fits beyond three planets. Alternatively, could unidentified HARPS systematics be responsible for the extra periods evident in their lower noise data? Independent experimental confirmation of the HARPS low-noise-level results would be very desirable.

ACKNOWLEDGMENT

The author would like to thank Wolfram Research for providing a complementary license to run gridMathematica.

REFERENCES

- Atchadé Y. F., Roberts G. O., Rosenthal J. S., 2010, *Stat. Comput.* (doi:10.1007/s11222-010-9192-1)
- Bonfils X., Forveille T., Delfosse X. et al., 2005, *A&A*, 443, L15
- Bretthorst G. L., 1988, *Bayesian Spectrum Analysis and Parameter Estimation*. Springer-Verlag, New York
- Butler R. P. et al., 2006, *ApJ*, 646, 505
- Campbell B., Walker G. A. H., Yang S., 1988, *ApJ*, 331, 902
- Clyde M. A., Berger J. O., Bullard F., Ford E. B., Jeffreys W. H., Luo R., Paulo R., Loredo T., 2007, in Babu G. J., Feigelson E. D., eds, *ASP Conf. Ser. Vol. 371, Statistical Challenges in Modern Astronomy IV*. Astron. Soc. Pac., San Francisco, p. 224
- Cumming A., 2004, *MNRAS*, 354, 1165
- Cumming A., Dragomir D., 2010, *MNRAS*, 401, 1029
- Dawson R. I., Fabrycky D. C., 2010, *ApJ*, 722, 937
- Delfosse X., Forveille T., Segransan D., Beuzit J.-L., Udry S., Perrier C., Mayor M., 2000, *A&A*, 364, 217
- Fischer, D. A., Marcy G. W., Butler R. P., Laughlin G. L., Vogt S. S., 2002, *ApJ*, 564, 1028
- 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, *ASP Conf. Ser. Vol. 371, Statistical Challenges in Modern Astronomy IV*. Astron. Soc. Pac., San Francisco, p. 189
- Gelman A., Rubin D. B., 1992, *Stat. Sci.*, 7, 457
- Geyer C. J., 1991, in Keramidas E. M., ed., *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*. Interface Foundation, Fairfax Station, p. 156
- 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., 2007a, *MNRAS*, 374, 1321
- Gregory P. C., 2007b, in Knuth K. H., Caticha A., Center J. L., Giffin A., Rodríguez C. C., eds, *AIP Conf. Ser. Vol. 954, Bayesian Inference and Maximum Entropy Methods in Science and Engineering*. Am. Inst. Phys., New York, p. 307
- Gregory P. C., 2007c, *MNRAS*, 381, 1607
- Gregory P. C., 2009, *JSM Proceedings*. American Statistical Association, Denver (arXiv:0902.2014v1 [astro-ph.EP])
- Gregory P. C., 2011, *MNRAS*, 410, 94
- Gregory P. C., Fischer D. A., 2010, *MNRAS*, 403, 731
- Hukushima K., Nemoto K., 1996, *J. Phys. Soc. Japan*, 65, 1604
- Jaynes E. T., 1957, in Erickson G. J., Smith C. R., eds, *Stanford University Microwave Laboratory Report 421*, Reprinted in ‘Maximum Entropy and Bayesian Methods in Science and Engineering’, (1988). Kluwer, Dordrecht, p. 1
- Jaynes E. T., 1987, in Smith C. R., Erickson G. L., eds, *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems*. D. Reidel, Dordrecht, p. 1
- Loredo T. L. and Chernoff D., 2003, in Feigelson E. D., Babu G. J., eds, *Statistical Challenges in Modern Astronomy III*. Springer-Verlag, New York, p. 57
- Loredo T., 2004, in Erickson G. J., Zhai Y., eds, *AIP Conf. Ser. Vol. 707, Bayesian Inference and Maximum Entropy Methods in Science and Engineering*. Am. Inst. Phys., New York, p. 330
- Marcy G. W., Butler R. P., 1996, *ApJ*, 464, L147
- Mayor M., Queloz D., 1995, *Nat*, 378, 355
- Mayor M. et al., 2009, *A&A*, 507, 487
- Roberts G. O., Gelman A., Gilks W. R., 1997, *Ann. Appl. Probab.*, 7, 110
- Saar S. H., Donahue R. A., 1997, *ApJ*, 485, 319
- Saar S. H., Butler R. P., Marcy G. W., 1998, *ApJ*, 498, L153
- Shen Y., Turner E. L., 2009, *ApJ*, 685, 553
- Udry S. et al., 2007, *A&A*, 469, L43
- Vogt S. S., Butler R. P., Rivera E. J., Haghighipour N., Henry G. W., Williamson M. H., 2010, *ApJ*, 723, 954
- Wolszczan A., Frail D., 1992, *Nat*, 355, 145
- Wright J. T., 2005, *PASP*, 117, 657
- Zakamska N. L., Pan M., Ford E. B., 2011, *MNRAS*, 410, 1895

APPENDIX A: ACCURACY OF MATHEMATICA MODEL RADIAL VELOCITIES

As explained in Section 3, we convert the observation times, t_i , to orbital angles, θ_i , by solving the conservation of angular momentum equation. As it stands, equation (7) runs into problems when the period P is small because MATHEMATICA’s NDSolve produces an interpolation function that spans the entire time-range. If there are many cycles, then a very large number of iteration steps are required and the procedure slows to a crawl. To avoid this, we first convert t_i to q_i , the corresponding fraction of one orbit, using the equation

$$q_i = \text{mod}[t_i/P + \chi, 1], \quad (\text{A1})$$

where again χ = the fraction of an orbit, prior to the start of data taking, that periastron occurred at. The relationship between q_i and θ_i is given by

$$\frac{d\theta}{dq} - \frac{2\pi[1 + e \cos \theta(q)]^2}{(1 - e^2)^{3/2}} = 0. \quad (\text{A2})$$

Note that the relationship between θ and q depends only on the eccentricity parameter, e .

MATHEMATICA generates an accurate interpolating function between q and θ so the differential equation does not need to be solved separately for each q_i . The solution of the differential equation is the largest component in the timing budget. Evaluating the interpolating function for each q_i is very fast compared to solving the differential equation, so the algorithm should be able to handle much larger samples of RV data than those currently available without a significant increase in computational time. For example, an increase in the sample size from 35 to 220 resulted in only an 18 per cent increase in execution time. Of course, for a large enough sample size the interpolation operation will begin to dominate and after that the execution time will scale with the number of data points.

We now address the question of the accuracy of the model RVs which are limited by the accuracy of the MATHEMATICA interpolating function. This was accomplished as follows:

- (1) Divide the interval $\theta = 0$ to 2π into $n = 10^7$ equal parts labelled θ_j . Let θ_{acc} represents this set of accurate θ_j values.
- (2) Evaluate the corresponding accurate RV_{acc}/K values, model RVs divided by K , where $\text{RV}_{\text{acc}}/K = [\cos(\theta_{\text{acc}} + \omega) + e \cos \omega]$.
- (3) Convert the θ_j values to q_j values by computing the orbital area swept out between each pair of θ_j values. According to Kepler’s Law of Areas, dq_j is proportional to that area increment. The computed total area swept out in the interval $\theta = 0$ to 2π was found to agree with theory to better than 1 part in 10^{11} .
- (4) Convert the computed q_j values into a set of interpolated θ values (θ_{int}) using MATHEMATICA’s solution of equation (A2).
- (5) Compute the corresponding set of interpolated $\text{RV}_{\text{int}}/K = [\cos(\theta_{\text{int}} + \omega) + e \cos \omega]$.

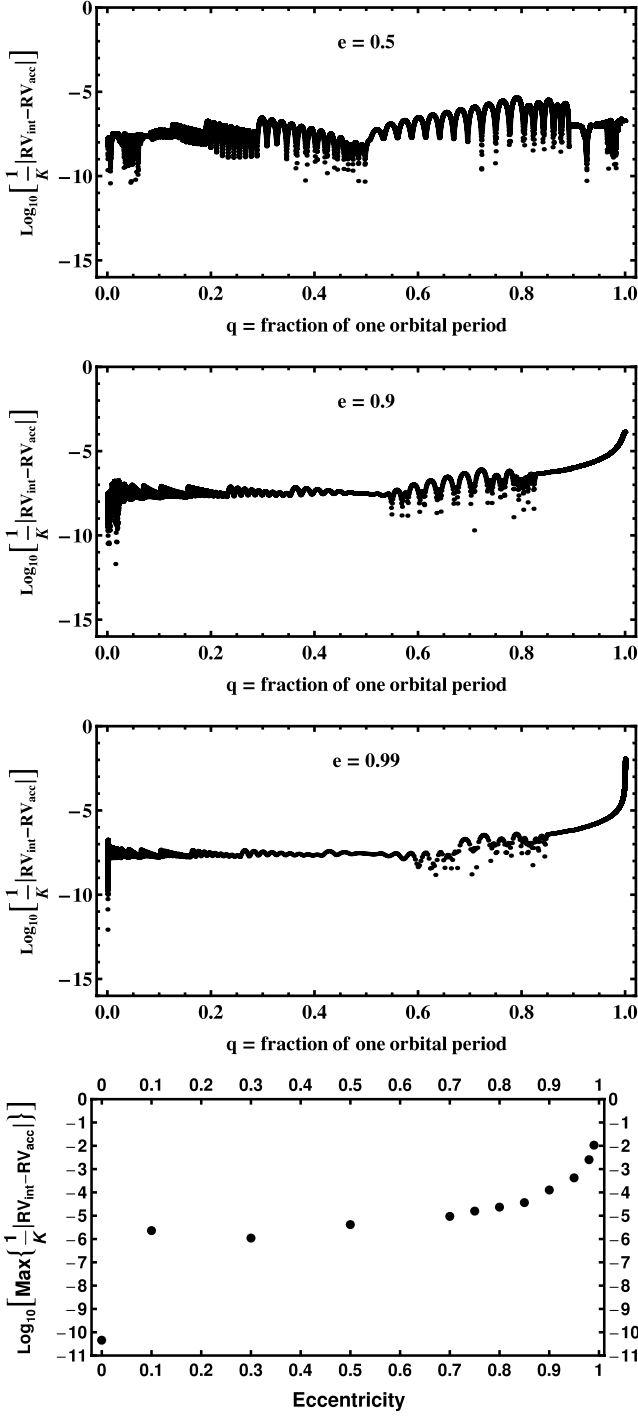


Figure A1. The top three panels show plots of the \log_{10} of the absolute magnitude of the model RV error as a fraction of the K value versus q , the fraction of the orbital period, for three different values of eccentricity. The bottom panel shows the \log_{10} of the maximum value of the above fractional RV error versus eccentricity.

(6) Compute $\frac{1}{K}(\text{RV}_{\text{int}} - \text{RV}_{\text{acc}})$, the RV error as a fraction of K :

$$\begin{aligned} \frac{1}{K}(\text{RV}_{\text{int}} - \text{RV}_{\text{acc}}) &= \cos(\theta_{\text{int}} + \omega) - \cos(\theta_{\text{acc}} + \omega) \\ &= \cos \omega (\cos \theta_{\text{int}} - \cos \theta_{\text{acc}}) \\ &\quad - \sin \omega (\sin \theta_{\text{int}} - \sin \theta_{\text{acc}}). \end{aligned} \quad (\text{A3})$$

$$\text{Let } \theta_0 = \frac{(\theta_{\text{int}} + \theta_{\text{acc}})}{2} \text{ and } \delta\theta = (\theta_{\text{int}} - \theta_{\text{acc}}), \quad (\text{A4})$$

then

$$\begin{aligned} \frac{1}{K}(\text{RV}_{\text{int}} - \text{RV}_{\text{acc}}) &= -2 \sin \frac{\delta\theta}{2} \sin \theta_0 \cos \omega \\ &\quad - 2 \sin \frac{\delta\theta}{2} \cos \theta_0 \sin \omega \\ &= -2 \sin \frac{\delta\theta}{2} \sin(\theta_0 + \omega) \\ &\approx -\delta\theta \sin(\theta_0 + \omega). \end{aligned} \quad (\text{A5})$$

For any θ_0 , the fractional model RV error has a maximum positive or negative value for $\sin(\theta_0 + \omega) = \pm 1$. Fig. A1 shows plots of the \log_{10} of the absolute magnitude of the fractional error versus q for three different values of eccentricity. The figure assumes a worst case value for $|\sin(\theta_0 + \omega)| = 1$. More realistically, these errors should be reduced by a factor of $2/\pi$ which is the expectation value of $|\sin(\theta_0 + \omega)|$. Even for $e = 0.99$, the fractional error is $< 10^{-5}$ over most of the q range, only rising above this towards the very end of the interpolation interval. The bottom panel shows the maximum value of the fractional error versus eccentricity. Based on this analysis, the maximum error in the MATHEMATICA-derived model RVs, expressed as a fraction of the K parameter, is $\leq 2.2 \times 10^{-5}$ for e values in the range 0–0.8. The situation degrades progressively for larger values of e but is still $\leq 2.8 \times 10^{-3}$ for $e = 0.98$, rising to 1.2×10^{-2} for $e = 0.99$.

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