Extra-solar Planets via Bayesian Fusion MCMC

This manuscript to appear as Chapter 7 in

'Astrostatistical Challenges for the New Astronomy', Springer Series in

Astrostatistics, Hilbe, J.M (ed), 2012, New York:Springer

Philip C. Gregory

Abstract A Bayesian multi-planet Kepler periodogram has been developed based on a fusion Markov chain Monte Carlo algorithm (FMCMC). FMCMC is a new general purpose tool for nonlinear model fitting. It incorporates parallel tempering, simulated annealing and genetic crossover operations. Each of these features facilitate the detection of a global minimum in chi-squared in a highly multi-modal environment. By combining all three, the algorithm greatly increases the probability of realizing this goal.

The FMCMC is controlled by a unique adaptive control system that automates the tuning of the proposal distributions for efficient exploration of the model parameter space even when the parameters are highly correlated. This controlled statistical fusion approach has the potential to integrate other relevant statistical tools as desired. The FMCMC algorithm is implemented in *Mathematica* using parallized code and run on an 8 core PC. The performance of the algorithm is illustrated with some recent successes in the exoplanet field where it has facilitated the detection of a number of new planets. Bayesian model selection is accomplished with a new method for computing marginal likelihoods called Nested Restricted Monte Carlo (NRMC).

1 Introduction

A remarkable array of new ground based and space based astronomical tools are providing astronomers access to other solar systems. Over 700 planets have been discovered to date, starting from the pioneering work of [1, 2, 3, 4]. One example of the fruits of this work is the detection of a super earth in the habitable zone surrounding Gliese 581 [5]. Fig. 1 illustrates the pace of discovery up to Dec. 2011.

Because a typical star is approximately a billion times brighter than a planet, only a small fraction of the planets have been detected by direct imaging. The majority

P. C. Gregory
Department of Physics and Astronomy, University of British Columbia, Vancouver, BC V6T 1Z1, Canada, e-mail: gregory@phas.ubc.ca

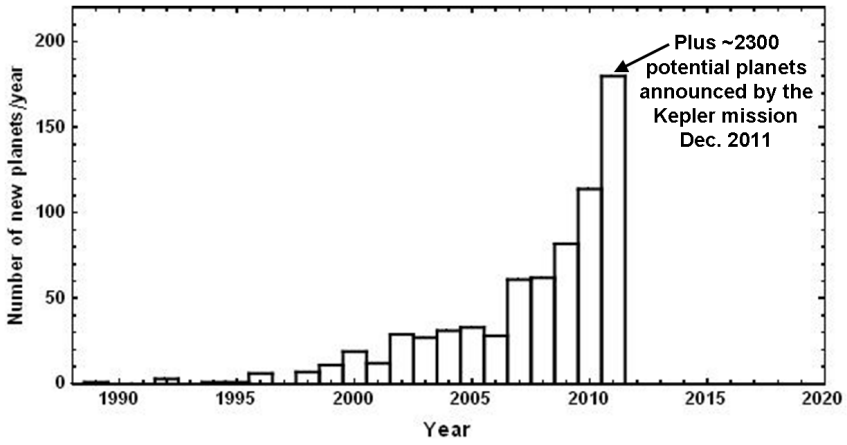

Fig. 1 The pace of exoplanet discoveries.

of planets have been detected by studying the reflex motion of the star caused by the gravitational tug of unseen planets, using precision radial velocity (RV) measurements. There are currently 84 known multiple planet systems, the largest of which has seven planets [6]. Recently the Kepler space mission has detected 2326 planetary candidates using the transit detection method. These are awaiting confirmation using the RV method or transit timing variation analysis. More than thirty of these candidates have a radius $\leq$ the radius of the earth.

These successes on the part of the observers has spurred a significant effort to improve the statistical tools for analyzing data in this field (e.g., [7, 8, 9, 10, 11, 12, 13, 14, 15, 16]). 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. MCMC algorithms provide a powerful means for efficiently computing the required Bayesian integrals in many dimensions (e.g., an 8 planet model has 41 unknown parameters). The output at each iteration of the MCMC is a vector of the model parameters. After an initial burn-in period, the MCMC produces an equilibrium distribution of samples in model parameter space such that the density of samples is proportional to the joint posterior probability distribution of the parameters. The marginal posterior probability density function (PDF) for any single parameter is given by a histogram of that component of the vector for all post burn-in iterations.

Frequently, MCMC algorithms have been augmented with an additional tool such as parallel tempering, simulated annealing or differential evolution depending on the complexity of the problem. My approach [17] has been to fuse together the advantages of all of the above tools together with a genetic crossover operation in a single MCMC algorithm to facilitate the detection of a global minimum in χ^2 (max-

imum posterior probability in the Bayesian context). The FMCMC is controlled by a unique multi-stage adaptive control system that automates the tuning of the proposal distributions for efficient exploration of the model parameter space even when the parameters are highly correlated. The FMCMC algorithm is implemented in *Mathematica* using parallized code and run on an 8 core PC. It is designed to be a very general tool for nonlinear model fitting. When implemented with a multi-planet Kepler model [1], it is able to identify any significant periodic signal component in the data that satisfies Kepler's laws and function as a multi-planet Kepler periodogram [2].

The different components of Fusion MCMC are described in Section 2. In Section 3, the performance of the algorithm is illustrated with some recent successes in the exoplanet field where it has facilitated the detection of a number of new planets. Section 4 deals with the challenges of Bayesian model selection in this arena.

2 Adaptive Fusion MCMC

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

$$p(\mathbf{X}|D,M_i,I) = C\ p(\mathbf{X}|M_i,I) \times p(D|M_i,\mathbf{X},I). \tag{1}$$

where C is the normalization constant which is not required for parameter estimation purposes and $\mathbf{X}$ represent the vector of model parameters. The term, $p(\mathbf{X}|M_i,I)$, is the prior probability distribution of $\mathbf{X}$, prior to the consideration of the current data D. The term, $p(D|\mathbf{X},M_i,I)$, is called the likelihood and it is the probability that we would have obtained the measured data D for this particular choice of parameter vector $\mathbf{X}$, model M_i, and prior information I. At the very least, the prior information, I, must specify the class of alternative models 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 problem see Gregory [11].

An important feature that prevents the fusion MCMC from becoming stuck in a local probability maximum is *parallel tempering* [20] (and re-invented under the name *exchange Monte Carlo* [21]). Multiple MCMC chains are run in parallel. The joint distribution for the parameters of model M_i, for a particular chain, is given by

$$\pi(\mathbf{X}|D,M_i,I,\beta) \propto p(\mathbf{X}|M_i,I) \times p(D|\mathbf{X},M_i,I)^{\beta}. \tag{2}$$

[1] For multiple planet models, there is no analytic expression for the exact radial velocity perturbation. In many cases, the radial velocity perturbation can be well modeled 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 [18, 19]

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

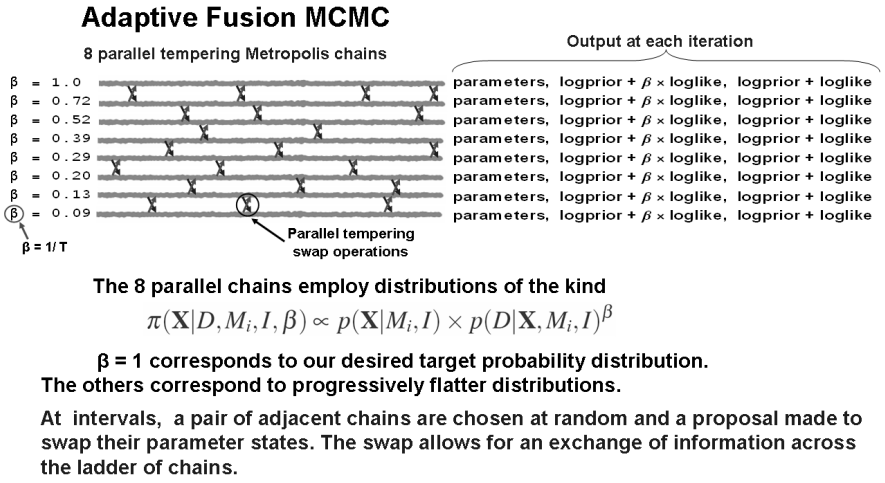

Fig. 2 Parallel tempering schematic.

In equation 2, an exponent $\beta = 0$ yields a joint 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 8 chains were employed. A representative set of β values is shown in Fig. 2. At an interval of 10 to 40 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 [10], 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 $\approx 30\%$ 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%, jumps to adjacent chains will occur at an interval of ~ 230 to 920 iterations while information from more distant chains will diffuse much more slowly. Recently, Atchade et al.

(2010) have shown that under certain conditions, the optimal swap acceptance rate is 0.234. A future goal for fusion MCMC is to extend the control system to automate the selection of an optimal set of β values as well.

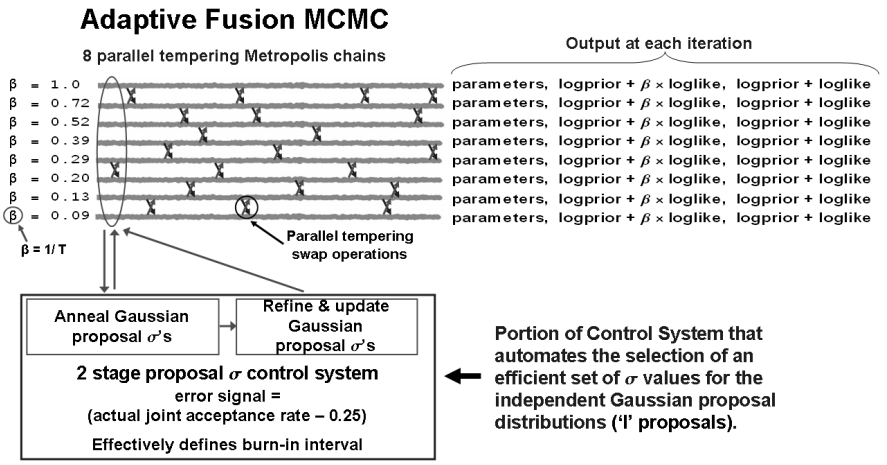

Fig. 3 First two stages of the adaptive control system.

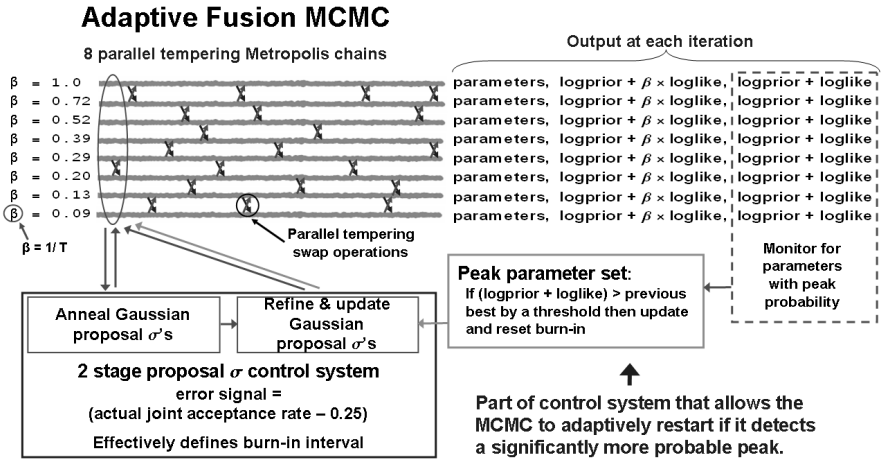

Fig. 4 Schematic illustrating how the second stage of the control system is restarted if a significantly more probable parameter set is detected.

At each iteration, a single joint proposal to jump to a new location in the parameter space is generated from independent Gaussian proposal distributions (centered 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 parallel tempering MCMC, this problem is compounded because of the need for a separate set of Gaussian proposal distributions for each tempering chain. This process is automated by an innovative statistical control system [23, 24] in which the error signal is proportional to the difference between the current joint parameter acceptance rate and a target acceptance rate [25], λ (typically $\lambda \sim 0.25$). A schematic of the first two stages of the adaptive control system (CS) is shown [3] in Fig. 3. Further details on the operation of the control system can be found in Gregory [17] and references therein. A third stage that handles highly correlated parameters is described in Section 2.2.

The adaptive capability of the control system can be appreciated from an examination of Fig. 4. The upper left portion of the figure depicts the FMCMC iterations from the 8 parallel chains, each corresponding to a different tempering level β as indicated on the extreme left. One of the outputs obtained from each chain at every iteration (shown at the far right) is the log prior + log likelihood. This information is continuously fed to 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 CS. If so, the second stage CS is re-run using the new more probable parameter set which is the basic adaptive feature of the existing CS [4]. Fig. 4 illustrates how the second stage of the control system is restarted if a significantly more probable parameter set is detected regardless of which chain it occurs in. This also causes the burn-in phase to be extended.

The control system also includes a genetic algorithm block which is shown in the bottom right of Fig. 5. The current parameter set can be treated as a set of genes. In the present version, one gene consists of the parameter set that specify 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 $\mathbf{X}_{\max}$, and the current most probable

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

[4] *Mathematica* code that implements the version of fusion MCMC shown in Fig. 4 is available on the Cambridge University Press web site for my textbook [10], 'Bayesian Logical data Analysis for the Physical Sciences'. See the 'Additional book examples with *Mathematica* 8 tutorial' in the resource material. There you will find an example entitled, 'Markov chain Monte Carlo powered Kepler periodogram'. Non *Mathematica* users can download a free Wolfram CDF Player to view the resource material.

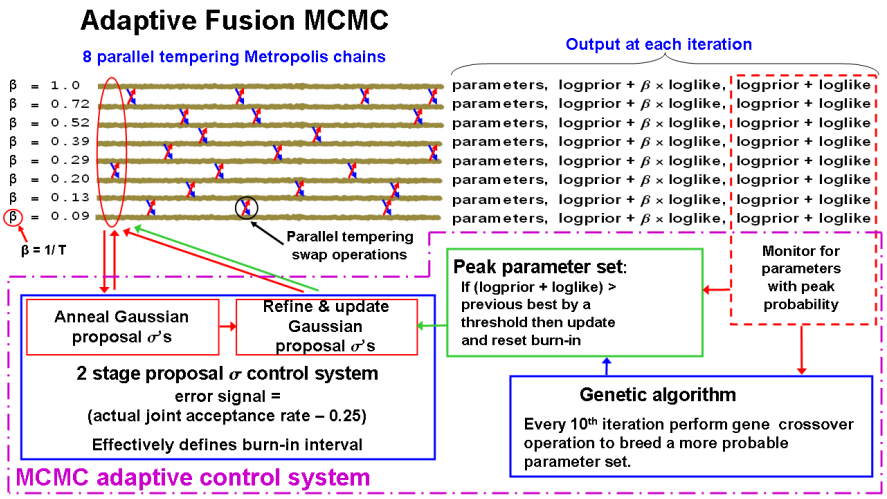

Fig. 5 This schematic shows how the genetic crossover operation is integrated into the adaptive control system.

parameter set of the 8 chains, $\mathbf{X}_{\text{cur}}$. At regular intervals (user specified) each gene from $\mathbf{X}_{\text{cur}}$ is swapped for the corresponding gene in $\mathbf{X}_{\text{max}}$. If either substitution leads to a higher probability it is retained and $\mathbf{X}_{\text{max}}$ 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 $\mathbf{X}_{\text{max}}$ compared to the number of new $\mathbf{X}_{\text{max}}$ arising from the normal parallel tempering MCMC iterations. The gene crossover operations prove to be very effective, and give rise to new $\mathbf{X}_{\text{max}}$ values ≈ 3 times more often than MCMC operations. Of course, most of these swaps lead to very minor changes in probability but occasionally big jumps are created. It turns out that individual gene swaps from $\mathbf{X}_{\text{cur}}$ to $\mathbf{X}_{\text{max}}$ are much more effective (in one test by a factor of 17) than the other way around (reverse swaps). Since it costs just as much time to compute the probability for a swap either way we no longer carry out the reverse swaps. Instead, we have extended this operation to swaps from $\mathbf{X}_{\text{cur2}}$, the parameters of the second most probable current chain, to $\mathbf{X}_{\text{max}}$. This gives rise to new values of $\mathbf{X}_{\text{max}}$ at a rate approximately half that of swaps from $\mathbf{X}_{\text{cur}}$ to $\mathbf{X}_{\text{max}}$. 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.

2.1 Automatic simulated annealing and noise model

The annealing of the proposal σ values occurs while the MCMC is homing in on any significant peaks in the target probability distribution. Concurrent with this,

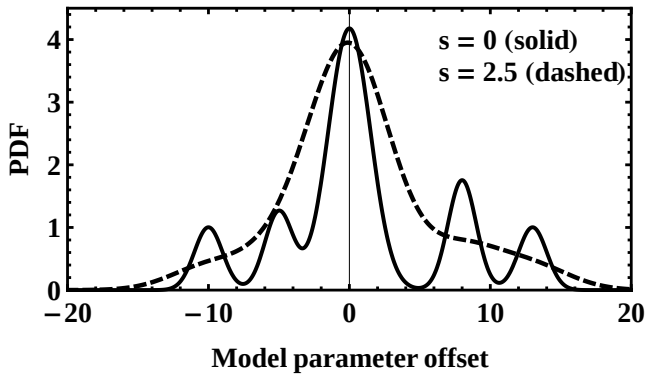

Fig. 6 A simulated toy posterior probability distribution (PDF) for a single parameter model with (dashed) and without (solid) an extra noise term s.

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 fit values. This automatically arises because all the models considered incorporate an extra additive noise [11], 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 [13], allows the Markov chain to explore the large scale structure in parameter space more quickly. This is illustrated in Figure 6 which shows a simulated toy posterior probability distribution (PDF) for a single parameter model with (dashed) and without (solid) an extra noise term s. Figure 7 shows the behavior of Log_{10}[Prior $\times$ Likelihood] and s versus MCMC iteration for a some real data. In the early stages s is inflated to around 38 m s^{-1} and then decays to a value of ≈ 4 m s^{-1} over the first 9,000 iterations as Log_{10}[Prior $\times$ Likelihood] reaches a maximum. This is similar to simulated annealing, but does not require choosing a cooling scheme.

2.2 Highly correlated parameters

For some models the data is such that the resulting estimates of the model parameters are highly correlated and the MCMC exploration of the parameter space can be very inefficient. Fig. 8 shows an example of two highly correlated parameters and possible ways of dealing with this issue which includes a transformation to more orthogonal parameter set. It would be highly desirable to employ a method that automatically samples correlated parameters efficiently. One potential solution in the literature is Differential Evolution Markov Chain (DE-MC) [26]. DE-MC is a popu-

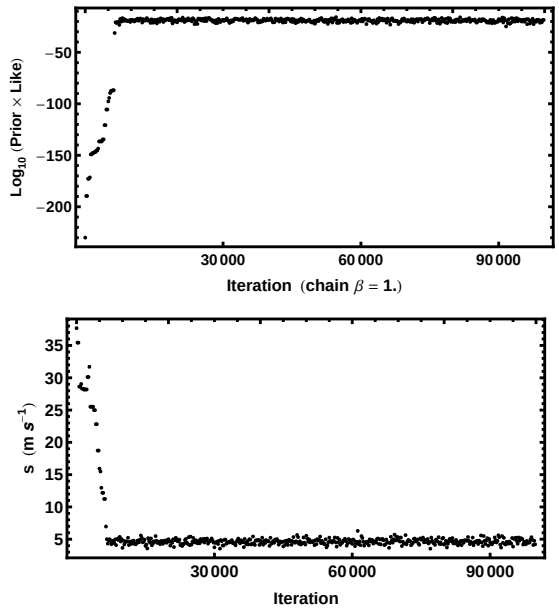

Fig. 7 The upper panel is a plot of the Log_{10}[Prior $\times$ Likelihood] versus MCMC iteration. The lower panel is a similar plot for the extra noise term s. Initially s is inflated and then rapidly decays to a much lower level as the best fit parameter values are approached.

lation MCMC algorithm, in which multiple chains are run in parallel, typically from 15 to 40. DE-MC solves an important problem in MCMC, namely that of choosing an appropriate scale and orientation for the jumping distribution.

For the fusion MCMC algorithm, I developed and tested a new method [27], in the spirit of DE, that automatically achieves efficient MCMC sampling in highly correlated parameter spaces without the need for additional chains. The block in the lower left panel of Fig. 9 automates the selection of efficient proposal distributions when working with model parameters that are independent 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 the σ values 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 control system every second [5] accepted 'I' proposal is appended to a correlated sample buffer. There is a separate buffer for each parallel tempering level. 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 corre-

[5] Thinning by a factor of 10 has already occurred meaning only every tenth iteration is recorded.

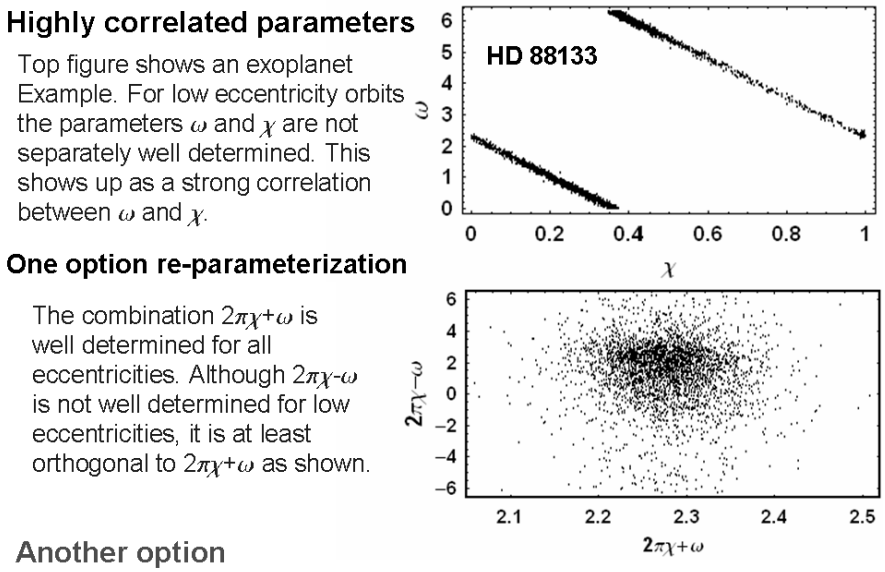

Fig. 8 An example of two highly correlated parameters and possible ways of dealing with this issue which includes a transformation to more orthogonal parameter set.

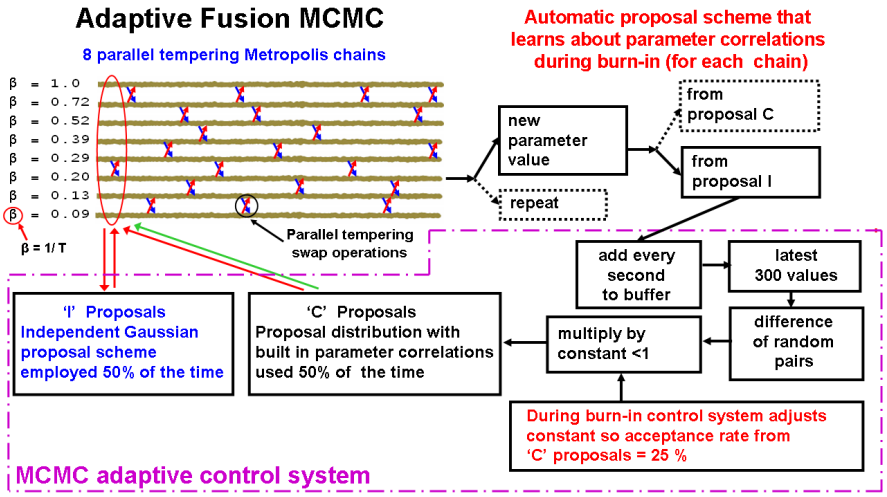

Fig. 9 This schematic illustrates the automatic proposal scheme for handling correlated ('C') parameters.

lated sample buffer for that tempering level, after multiplication by a constant. The value of this constant (for each tempering level) is computed automatically [27] by another control system module which ensures that the 'C' proposal acceptance rate is close to 25%. With very little computational overhead, the 'C' proposals provide the scale and direction for efficient jumps in a correlated parameter space.

The final proposal distribution is a random selection of 'I' and 'C' proposals such that each is employed 50% of the time. The combination ensures that the whole parameter space can be reached and that the FMCMC chain is aperiodic. The parallel tempering 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 control system 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.

2.2.1 Tests of the 'C' proposal scheme

Gregory [27] carried out two tests of the 'C' proposal scheme using (a) simulated exoplanet astrometry data, and (b) a sample of real radial velocity data. In the latter test we analyzed a sample of seventeen HD 88133 precision radial velocity measurements [28] using a single planet model in three different ways. Fig. 10 shows a comparison of the resulting post burn-in marginal distributions for two correlated parameters χ and ω together with a comparison of the autocorrelation functions. The black trace corresponds to a search in χ and ω using only 'I' proposals. The red trace corresponds to a search in χ and ω with 'C' proposals turned on. The green trace corresponds to a search in the transformed orthogonal coordinates $\psi = 2\pi\chi + \omega$ and $\phi = 2\pi\chi - \omega$ using only 'I' proposals. It is clear that a search in χ and ω with 'C' proposals turned on achieves the same excellent results as a search in the transformed orthogonal coordinates ψ and ϕ using only 'I' proposals.

3 Exoplanet applications

As previously mentioned the FMCMC algorithm is designed to be a very general tool for nonlinear model fitting. When implemented with a multi-planet Kepler model 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 multi-planet Kepler periodogram.

In this section we describe the model fitting equations and the selection of priors for the model parameters. For a one planet model the predicted radial velocity is given by

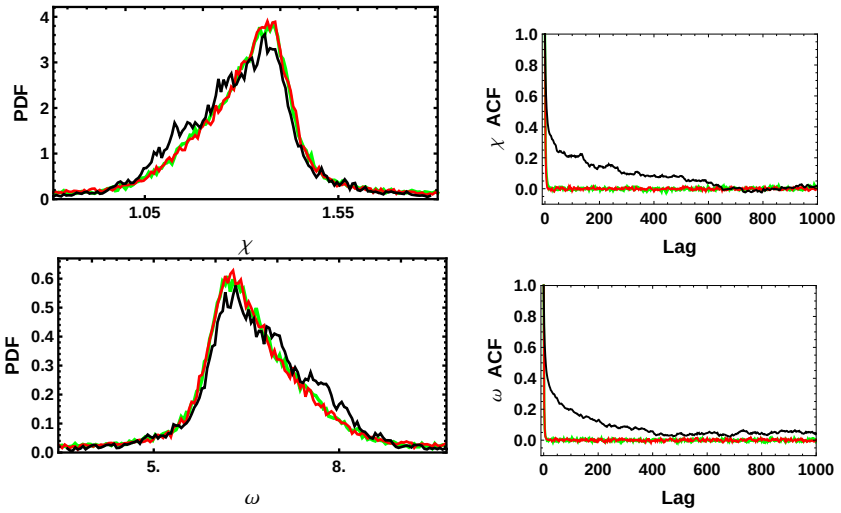

Fig. 10 The two panels on the left show a comparison of the post burn-in marginal distributions for χ and ω. The two panels on the right show a comparison of their MCMC autocorrelation functions. The black trace corresponds to a search in χ and ω using only 'I' proposals. The red trace corresponds to a search in χ and ω with 'C' proposals turned on. The green trace corresponds to a search in the transformed orthogonal coordinates $\psi = 2\pi\chi + \omega$ and $\phi = 2\pi\chi - \omega$ using only 'I' proposals.

$$v(t_i) = V + K[\cos\{\theta(t_i + \chi P) + \omega\} + e\cos\omega], \tag{3}$$

and involves the 6 unknown parameters

$V =$ a constant velocity.
$K =$ velocity semi-amplitude.
$P =$ the orbital period.
$e =$ the orbital eccentricity.
$\omega =$ the longitude of periastron.
$\chi =$ the fraction of an orbit, prior to the start of data taking, that periastron occurred at. Thus, $\chi P =$ the number of days prior to $t_i = 0$ that the star was at periastron, for an orbital period of P days.
$\theta(t_i + \chi P) =$ 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

$$\frac{d\theta}{dt} - \frac{2\pi[1 + e\cos\theta(t_i + \chi\, P)]^2}{P(1-e^2)^{3/2}} = 0. \tag{4}$$

Our algorithm is implemented in *Mathematica* and it proves faster for *Mathematica* to solve this differential equation than solve the equations relating the true anomaly to the mean anomaly via the eccentric anomaly. *Mathematica* generates an accurate

interpolating function between t and 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 4 is implemented are given in the Appendix of [17].

We employed a re-parameterization 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 $\phi = 2\pi\chi - \omega$. 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 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-parameterization it is necessary to multiply the Bayesian integrals by a factor of $(4\pi)^{-nplan}$, where $nplan =$ 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 2.2 which typically saves about 25% in 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 2 is the product of the individual parameter priors. Detailed arguments for the choice of each prior were given in [29, 24].

As mentioned in Section 2.1, all of the models considered in this paper incorporate an extra noise parameter, s, that can allow for any additional noise beyond the known measurement uncertainties[6]. 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 i$^{\text{th}}$ data point. For example, suppose that the star actually has two planets, and the model assumes only one is present. In regard to the single planet model, the velocity variations induced by the unknown second planet acts 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. 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 can't 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 [10] 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 to $K_{\max}$. We employed a modified Jeffrey's (scale invariant) prior for s with a knee, $s_0 = 1\text{m s}^{-1}$.

Fig. 11 shows sample MCMC traces for a two planet fit [29] to a set of HD 208487 radial velocity data [31].

[6] 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, i.e., the maximum entropy distribution [30] and [10] (section 8.7.4.)

Table 1 Prior parameter probability distributions.

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

[a] Since the prior lower limits for K and s include zero, we used a modified scale invariant prior of the form

$$p(X|M,I) = \frac{1}{X+X_0}\,\frac{1}{\ln\left(1+\frac{X_{\max}}{X_0}\right)} \tag{5}$$

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

3.1 47 Ursae Majoris

47 Ursae Majoris (47 UMa) is a solar twin at a distance of 46 light years. A brief history of the analysis of the radial velocity data together with the recent 3 planet FMCMC fit of the Lick Observatory data [24] is shown in Fig. 12.

Five different models assuming 0, 1, 2, 3, & 4 planets were explored. Our Bayesian model selection analysis indicates that the 3 planet model is significantly [24] more probable than the others. Fig. 13 shows some of the FMCMC results for the preferred 3 planet model. The upper left graph shows the trace of the Log_{10}[Prior $\times$ Likelihood] versus iteration for the Lick telescope data. The lower left graph shows the corresponding trace of the three period parameters. The starting parameter values are indicated by the three arrows. The top right graph shows a plot of eccentricity versus period for the same run. There is clear evidence for three

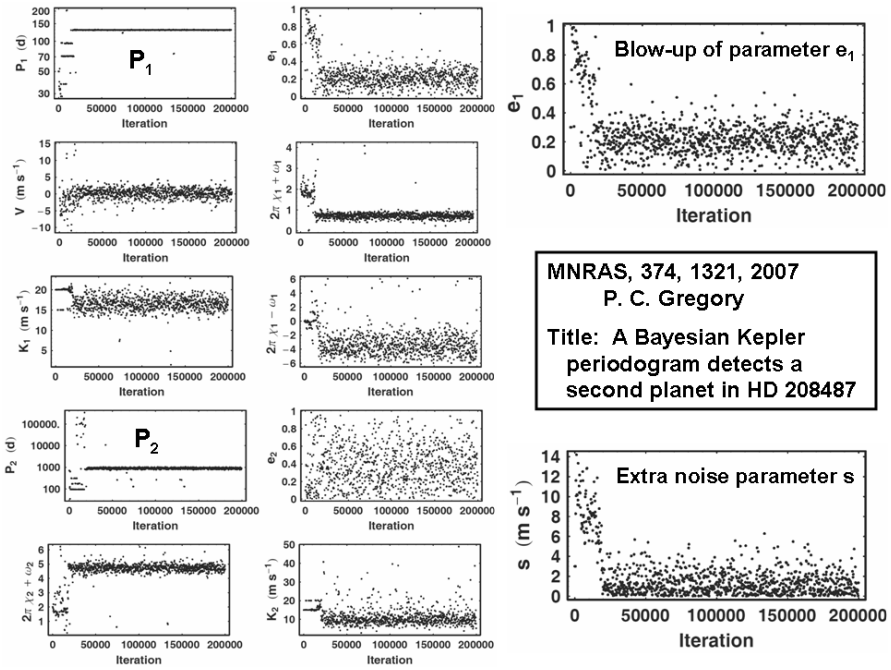

Fig. 11 Sample FMCMC traces for a two planet fit to HD 208487 radial velocity data.

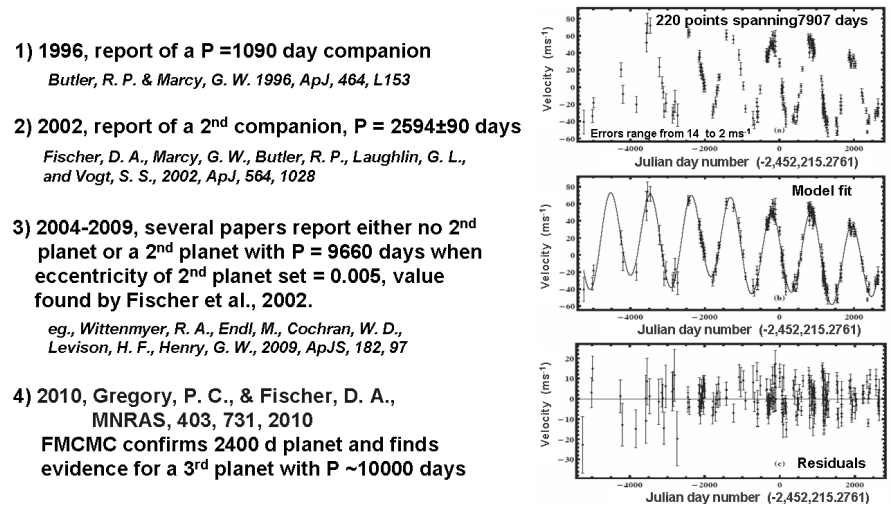

Fig. 12 Brief history of 47 UMa radial velocity fits.

signals including one with a period of $\sim$ 2400d. The longest period of $\sim$ 10000d is not as well defined mainly because it corresponds to periods longer than the data duration of 7907d. Previous experience with the FMCMC periodogram indicates that it is capable of finding a global peak in a blind search of parameter space for a three planet model. The lower right graph shows the trace of the three period parameters for a second run with very different starting parameter periods of 5, 20, 100d. The algorithm readily finds the same set of final periods in both cases.

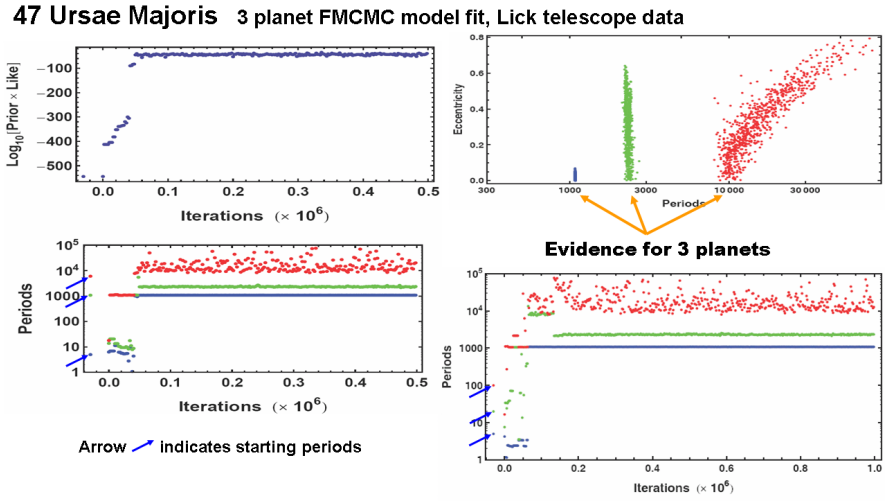

Fig. 13 The upper left graph shows the trace of the Log_{10}[Prior $\times$ Likelihood] versus iteration for the 3 planet FMCMC Kepler periodogram of the Lick telescope data for 47 UMa. The lower left graph shows the corresponding trace of the three period parameters. The starting parameter values are indicated by the three arrows. The top right graph shows a plot of eccentricity versus period for the same run. The lower right graph shows the trace of the three period parameters for a second run with very different starting parameter values.

To test the evidence for three planets the analysis was repeated with the Lick data combined with the data [32] from the 9.2 m Hobby-Eberly Telescope (HET) and 2.7 m Harlam J. Smith (HJS) telescopes of the McDonald Observatory. Figure 14 shows a plot of eccentricity versus period for our 3 planet FMCMC fit to the combined data set. The same three periods appear as before but with the extra data the results now favor low eccentricity orbits for all three periods. This is a particularly pleasing result as low eccentricity orbits are more likely to exhibit long term stability than high eccentricity orbits.

Figure 15 shows the final marginal distributions for the parameters for the three telescope combined analysis. One of the advantages of the Bayesian approach is the ability to deal with nuisance parameters such as unknown systematic residual velocity offsets when combining data from different observatories, and different detector dewars on the same telescope. The marginal distributions for these nuisance parameters are included Figure 15 together with the distribution of the extra noise

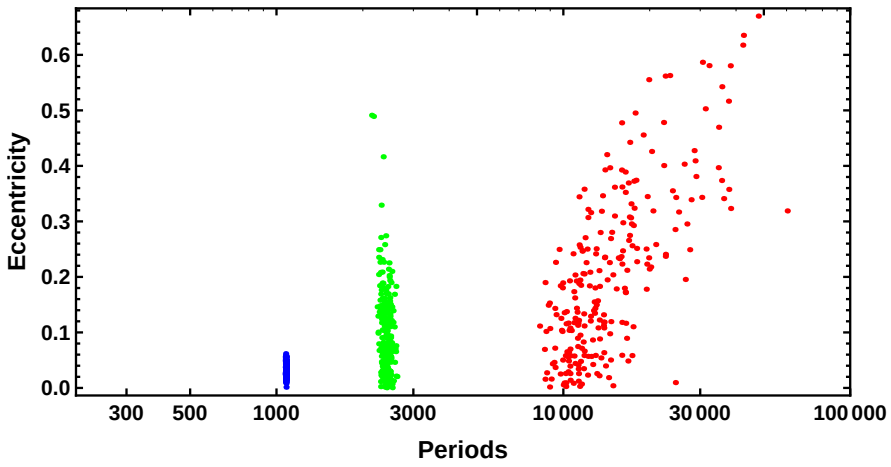

Fig. 14 A plot of eccentricity versus period for a 3 planet FMCMC fit [24] of the combined Lick, Hobby-Eberly, and Harlam J. Smith telescope data set for 47 UMa.

parameter s. The systematic residual offset velocity parameters, relative to the Lick dewar number 24, are designated V_j, where $j = 6, 8, 39, 18$ correspond to the other Lick dewars and subscripts HET and HJS refer to the Hobby-Eberly Telescope and Harlam J. Smith telescopes.

3.2 Gliese 581

Gliese 581 (Gl 581) is an M dwarf with a mass of 0.31 times the mass of the sun at a distance of 20 light years which has received a lot of attention because of the possibility of two super-earths in the habitable zone where liquid water could exist. A brief history of the analysis of the radial velocity data together with our recent [17] 5 planet FMCMC fit of the HARPS [33] data is shown in Fig. 16.

We carried out a Bayesian re-analysis of the HARPS [33] and HIRES data [34]. Our analysis of the HARPS data found significant evidence for 5 planets, the four reported by [33] and a 399^{+14}_{-16}d period ($6.6^{+2.0}_{-2.7}$M$_\oplus$) planet similar to the 433 ± 13d period reported by Vogt et al. [34]. Fig. 17 shows the 5 planet Kepler periodogram results for the HARPS data. The best set of parameters from the 4 planet fit were used as start parameters. The starting period for the fifth period was set $= 300$d and the most probable period found to be ~ 400d. As illustrated in this example, the parallel tempering feature identifies not only the strongest peak but other potential interesting ones as well. The fifth period parameter (orange points) shows 3 peaks but the 400d period has a maximum value of prior $\times$ likelihood that is almost 1000 times larger than the next strongest which is a second harmonic at 200d. A much

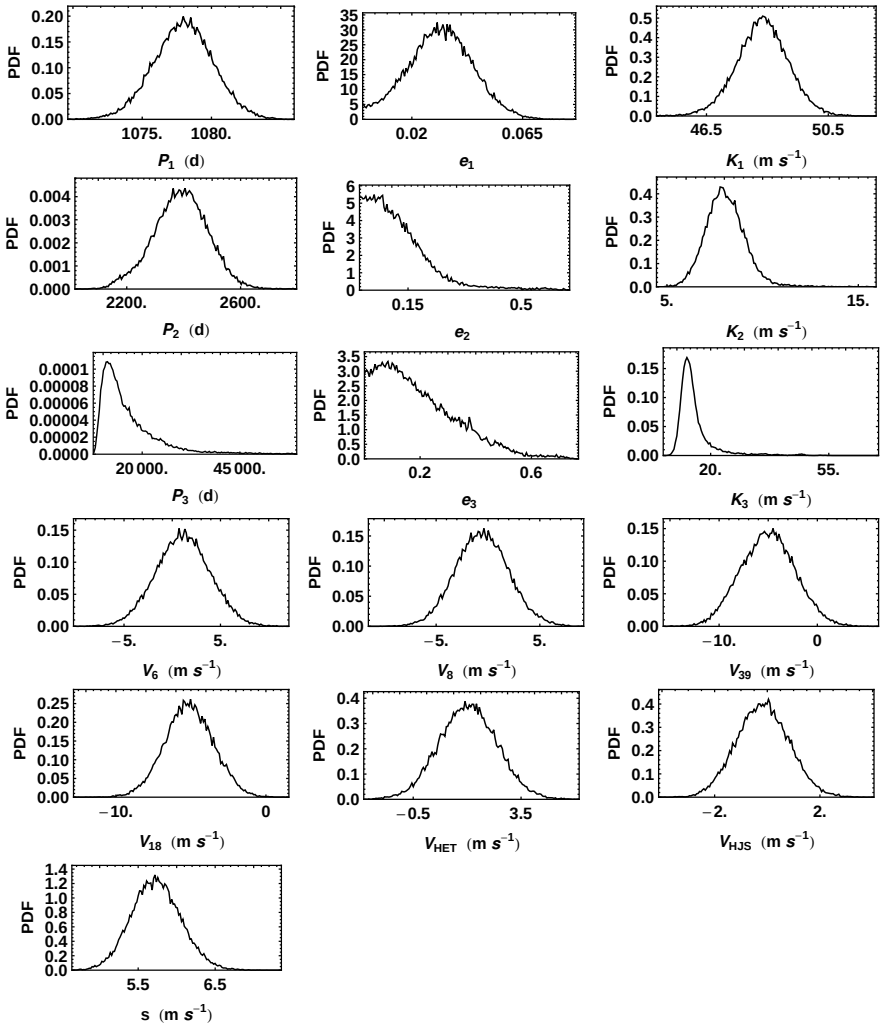

Fig. 15 A plot of parameter marginal distributions for a 3 planet FMCMC of the combined Lick, HET, and HJS telescope data set for 47 UMa. In addition to the extra noise parameter s distributions are also shown for 6 additional nuisance parameters. They are systematic residual offset velocity parameters relative to the Lick dewar 24. They are designated V_j, where $j = 6, 8, 39, 18$ correspond to the other Lick dewars and subscripts HET and HJS refer to the Hobby-Eberly Telescope and Harlam J. Smith telescopes.

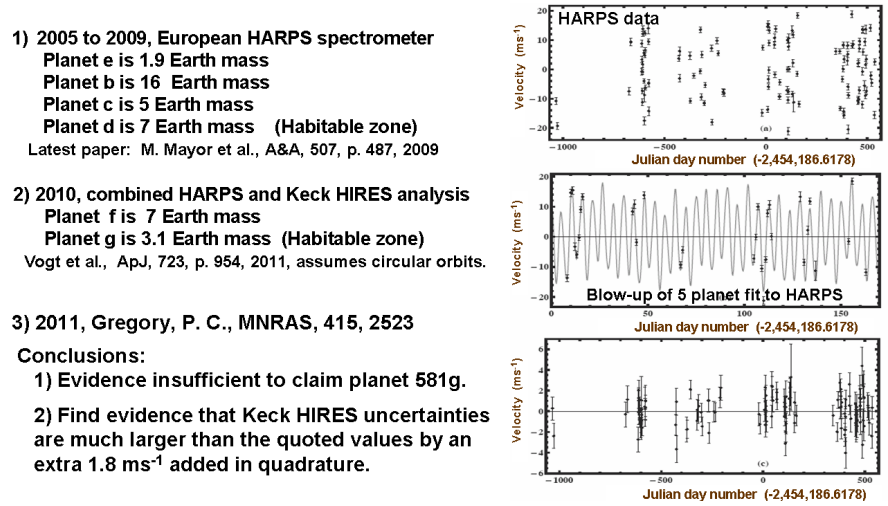

Fig. 16 Brief history of Gliese 581 radial velocity fits.

weaker peak shows up with a 34.4d period. In a Bayesian analysis the relative importance of the three peaks is in proportion to the number of MCMC samples in each peak which is in the ratio of 1.0:0.35:0.04 for the 400, 200, and 34.4d peaks, respectively.

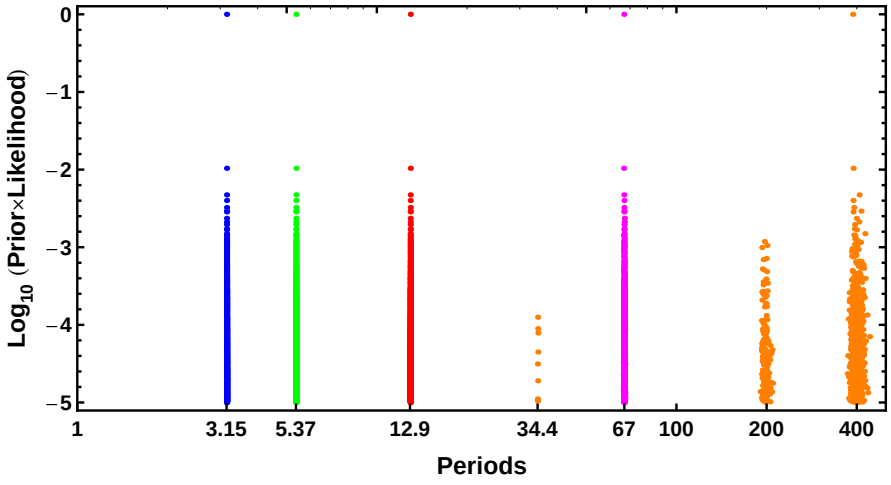

Fig. 17 A plot of the 5 period parameter values versus a normalized value of Log_{10}[Prior $\times$ Likelihood] for the 5 planet FMCMC Kepler periodogram of the HARPS data for Gliese 581. The fifth period parameter points are shown in orange.

Fig. 18 shows a plot of a subset of the FMCMC parameter marginal distributions for the 5 planet fit of the HARPS data after filtering out the post burn-in FMCMC iterations that correspond to the 5 dominant period peaks at 3.15, 5.37, 12.9, 66.9, and 400d. The median value of the extra noise parameter $s = 1.16\text{m s}^{-1}$.

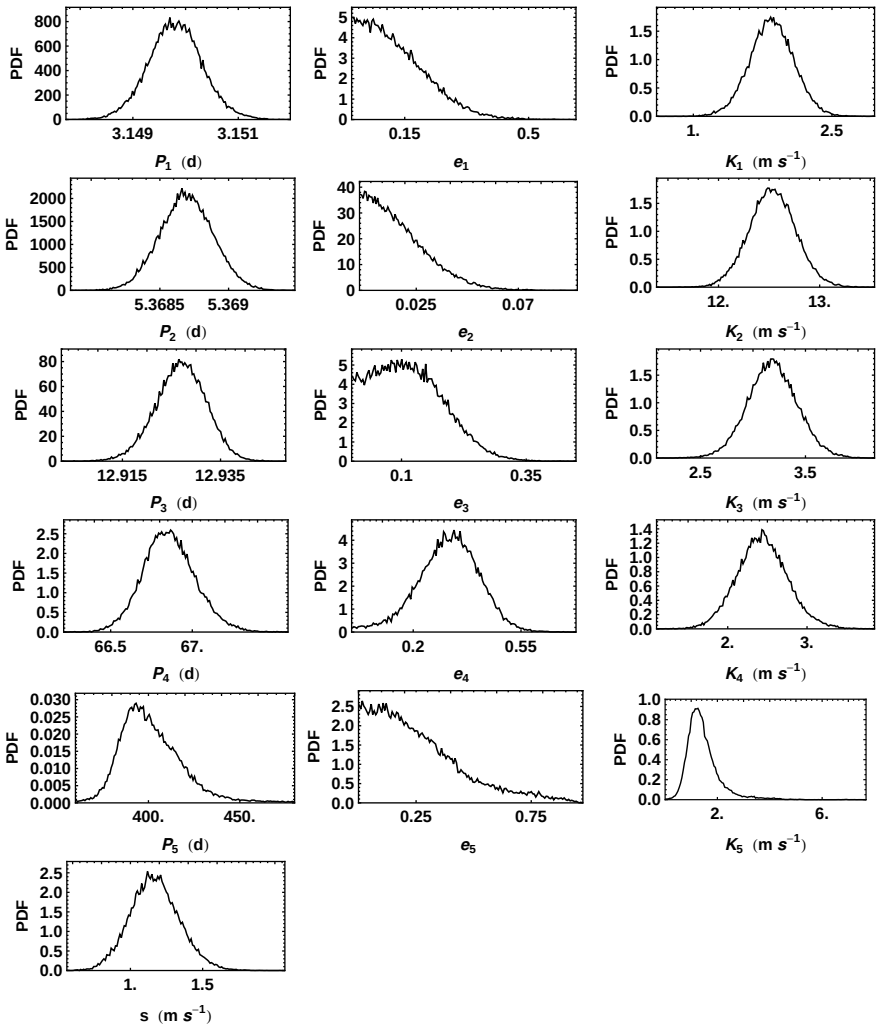

Fig. 18 A plot of a subset of the FMCMC parameter marginal distributions for a 5 planet fit of the HARPS data for Gliese 581.

A six planet model fit found multiple periods for a sixth planet candidate, the strongest of which had a period of 34.4 ± 0.1d but our Bayesian false alarm probability (discussed in Section 4) was much too high to consider it significant. Vogt et al. found a period of 36.562 ± 0.052d for Gl 581g.

The analysis of the HIRES data set yielded a reliable detection of only the strongest 5.37 and 12.9 day periods. For a two planet fit the value of the extra noise term s (stellar jitter) was higher for the HIRES data than for the HARPS data suggesting the possibility the HIRES measurement uncertainties were underestimated. The analysis of the combined HIRES/HARPS data again only reliably detected the 5.37 and 12.9d periods. Detection of 4 planetary signals with periods of 3.15, 5.37, 12.9, and 66.9d was only achieved by including an additional unknown but parameterized Gaussian error term added in quadrature to the HIRES quoted errors. The marginal probability density of the sigma for this additional HIRES Gaussian noise term has a well defined peak at $1.84^{+0.35}_{-0.33}$m s^{-1}.

The two conclusions of our GL 581 analysis are: (1) The current evidence is insufficient to warrant a claim for Gl 581g. (2) The quoted errors for the HIRES data are significantly underestimated equivalent to an effective $1.8^{0.35}_{0.33}$ ms^{-1} Gaussian error added in quadrature with the quoted uncertainties.

4 Model Selection

One of the great strengths of Bayesian analysis is the built-in Occam's razor. More complicated models contain larger numbers of parameters and thus incur a larger Occam penalty, which is automatically incorporated in a Bayesian model selection analysis in a quantitative fashion (see for example, Gregory [10], p. 45). The analysis yields the relative probability of each of the models explored.

To compare the posterior probability of the i$^{\text{th}}$ planet model to 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,

$$O_{i4} = \frac{p(M_i|I)}{p(M_4|I)} \, \frac{p(D|M_i,I)}{p(D|M_4,I)} \equiv \frac{p(M_i|I)}{p(M_4|I)} \, B_{i4} \tag{6}$$

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

$$p(D|M_i,I) = \int d\mathbf{X} p(\mathbf{X}|M_i,I) \times p(D|\mathbf{X},M_i,I). \tag{7}$$

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 (e.g., see Gregory [10], page 48). The Bayes factor will favor 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 ΔX, and a centrally peaked likelihood function with characteristic width δX, the Occam factor is $\approx \delta X/\Delta X$. If the data is 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 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 the 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 scale invariant prior (e.g., see Gregory [10], page 61). 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 the Bayes factors. Because the Occam factor arises automatically in the marginalization procedure, its effect will be present in any model selection calculation. Note: 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. [35] reviewed the state of techniques for model selection from a statistical perspective and Ford and Gregory [14] have evaluated the performance of a variety of marginal likelihood estimators in the exoplanet context. Other techniques that have recently been proposed include: Nested Restricted Monte Carlo (NRMC) [24], MultiNest [36], Annealing Adaptive Importance Sampling (AAIS) [37], and Reversible Jump Monte Carlo using a kD-tree [38]. For one planet models with 7 parameters, a wide range of techniques perform satisfactorily. The challenge is to find techniques that handle high dimensions. A six planet model has 32 parameters and one needs to develop and test methods of handling at least 8 planets with 42 parameters. At present there is no widely accepted method to deal with this challenge.

For the 47UMa and Gliese 581 data discussed in Sections 3.1 and 3.2 the author employed Nested Restricted Monte Carlo (NRMC) to estimate the marginal likelihoods. Monte Carlo (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 increase. For example, if the fractional volume with significant probability is 0.1 in one dimension then in 32 dimensions the fraction might be of order 10^{-32}. In restricted MC integration (RMC) this problem is reduced because the volume of 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. However, in high dimensions most of the MC samples will fall near the outer boundaries of that volume and so the sampling could easily under sample interior regions of high probability.

In NRMC integration, multiple boundaries are constructed based on credible regions ranging from 30% to $\geq 99\%$, 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% credible regions, we generate random parameter samples within the 60% region and reject any sample that falls within the 30% region. Using the remaining samples we can compute the contribution to the NRMC integral from that interval.

The left panel of Fig. 19 shows the NRMC contributions to the marginal likelihood from the individual intervals for five repeats of a 3 planet fit to the HARPS data [33]. The right panel shows the summation of the individual contributions versus the volume of the credible region. The credible region listed as 9995% is defined as follows. Let X_{U99} and X_{L99} correspond to the upper and lower boundaries of the 99% credible region, respectively, for any of the parameters. Similarly, X_{U95} and X_{L95} are the upper and lower boundaries of the 95% credible region for the parameter. Then $X_{U9995} = X_{U99} + (X_{U99} - X_{U95})$ and $X_{L9995} = X_{L99} + (X_{L99} - X_{L95})$. Similarly [7], $X_{U9984} = X_{U99} + (X_{U99} - X_{U84})$. For the 3 planet fit the spread in results is within $\pm 23\%$ of the mean. For each credible region interval approximately 320,000 MC samples were used. The mean value of the prior $\times$ likelihood within the 30% credible region is a factor of 2×10^5 larger than the mean in the shell between the 97 and 99% credible regions. However, the volume of parameter space in the shell between the 97 and 99% credible regions is a factor of 8×10^{11} larger than the volume within the 30% credible region so the contribution from the latter to the marginal likelihood is negligible.

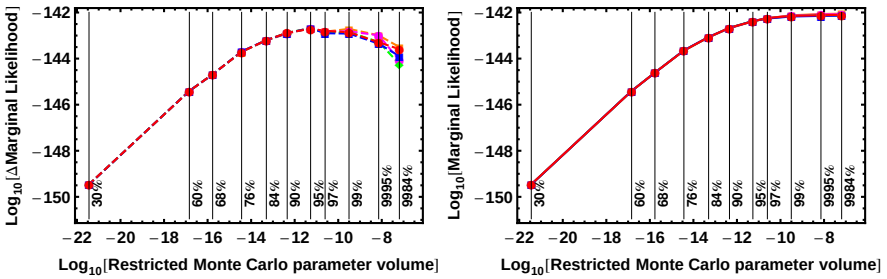

Fig. 19 Left panel shows the contribution of the individual nested intervals to the NRMC marginal likelihood for the 3 planet model for five repeats. The right panel shows the sum of these contributions versus the parameter volume of the credible region.

The left panel of Fig. 20 shows the contributions from the individual credible region intervals for five repeats of the NRMC marginal likelihood estimate for a 5 planet fit to the HARPS data [33]. The right panel shows the summation of the individual contributions versus the volume of the credible region. In this case the

[7] Test that the extended credible region (like 9930) for one period parameter does not overlap the credible region of an adjacent period parameter in a multiple planet fit.

spread in five NRMC marginal likelihood estimates extends from $2.1\times$ the mean to $0.65\times$ the mean.

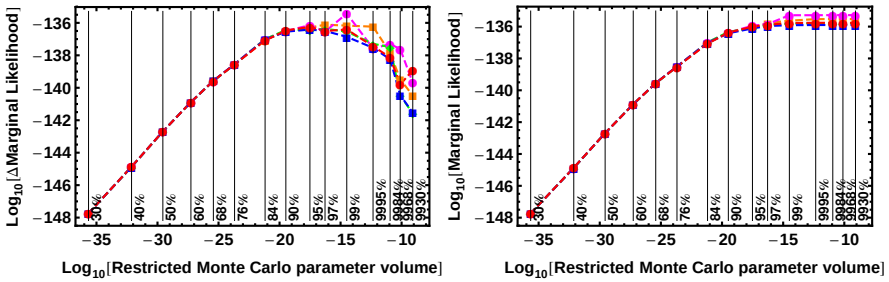

Fig. 20 Left panel shows the contribution of the individual nested intervals to the NRMC marginal likelihood for the 5 planet model for five repeats. The right panel shows the integral of these contributions versus the parameter volume of the credible region.

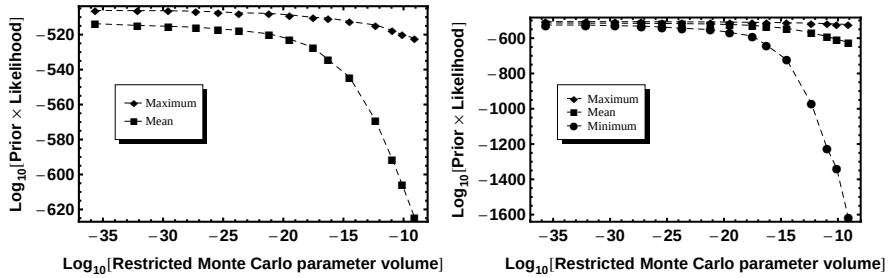

Fig. 21 Left panel shows the maximum and mean values of the Log_{10}[prior $\times$ likelihood] for each interval of credible region versus parameter volume for the 5 planet fit. The right panel shows the maximum, mean, and minimum values of the Log_{10}[prior $\times$ likelihood] versus the parameter volume.

The biggest contribution to the spread in NRMC marginal likelihood estimates for the 5 planet fit comes from the outer credible region intervals starting around 99%. The reason for the increased scatter in the $\text{Log}_{10}[\Delta$ Marginal Likelihood] is apparent when we examine the MC samples for one of the five repeats. Fig 21 shows plots of the maximum and mean values (Left) and maximum, mean, and minimum values (Right) of the MC samples of the Log_{10}[prior $\times$ likelihood] for each interval of credible region, versus the parameter volume. The range of Log_{10}[prior $\times$ likelihood] values increase rapidly with increasing parameter volume starting around the 99% credible region boundary. This makes the Monte Carlo evaluation of the mean value more difficult in these outer intervals. We can compute the fraction of our total marginal likelihood estimate that arises for the intervals beyond the 99% credible region. Averaging over the 5 repeats the mean fraction is $6.6 \pm 3.4\%$. This

mean fraction increases to $36 \pm 12\%$ for all intervals beyond the 97% credible region and to $56 \pm 9\%$ for the all intervals beyond the 95% credible region.

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, i.e. increasing number of planets [39]. When we conclude, as we do, that the NRMC computed odds in favor of a five planet model for Gliese 581 compared to the four planet model is $\sim 10^2$ (see Table 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 that was discussed above. A recent comparison by Gregory [40] of the NRMC method with a second method, the Ratio Estimator (RE), described in Ford and Gregory [14], indicates the two methods agree within 25% for a 3 planet model fit (17 parameters) to three different exoplanet data sets including Gliese 581. However, the agreement between the two methods breaks down at the 4 planet fit level (22 parameters) with the value of the RE estimate 62 times the NRMC estimate for the Gliese 581 HARPS data. Unlike the NRMC method, the RE method has the potential to pay too much attention to the mode and is expected to overestimate the marginal likelihood [41] at sufficiently high dimensions.

We can readily convert the Bayes factors to a Bayesian False Alarm Probability (FAP) which we define in equation 8. For example, in the context of claiming the detection of m planets the FAP_m is the probability that there are actually fewer than m planets, i.e., $m-1$ or less.

$$\mathrm{FAP}_m = \sum_{i=0}^{m-1} (\text{prob.of i planets}) \tag{8}$$

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

$$p(M_i \mid D, I) = \frac{B_{i4}}{\sum_{j=0}^{N} B_{j4}} \tag{9}$$

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 8 gives

$$\mathrm{FAP}_5 = \frac{(B_{04} + B_{14} + B_{24} + B_{34} + B_{44})}{\sum_{j=0}^{5} B_{j4}} \approx 10^{-2} \tag{10}$$

For the 5 planet model we obtain a low FAP $\approx 10^{-2}$.

Table 2 gives the NRMC marginal likelihood estimates [8], Bayes factors and false alarm probabilities for 0, 1, 2, 3, 4, 5 and 6 planet models which are designated

[8] Table 1 gives two different choices of prior for the eccentricity parameter. The marginal likelihoods listed in the Table 2 correspond to the eccentricity noise bias prior. Marginal likelihood values assuming a uniform eccenticity prior were systematically lower. For example, for a 3 planet fit using the uniform eccenticity prior the marginal likelihood was a factor of 3 smaller.

$M_0, \cdots, M_6$. For each model the NRMC calculation was repeated 5 times and the quoted errors give the spread in the results, not the standard deviation. The Bayes factors that appear in the third column are all calculated relative to model 4. Based on the HARPS [33] data, the Bayes factor favors a 5 planet model.

Table 2 Marginal likelihood estimates [17], Bayes factors relative to model 4, and false alarm probabilities. The quoted errors are the spread in the results for 5 repeats, not the standard deviation.

Model	Periods (d)	Marginal Likelihood	Bayes factor nominal	False Alarm Probability
M_0		6.10×10^{-197}	2.0×10^{-59}	
M_1	(5.37)	$(4.221 \pm 0.003) \times 10^{-155}$	1.4×10^{-17}	1.4×10^{-42}
M_2	$(5.37, 12.9)$	$(1.94 \pm 0.01) \times 10^{-145}$	6.5×10^{-8}	2.2×10^{-10}
M_3	$(5.37, 12.9, 66.9)$	$(3.0^{+0.7}_{-0.5}) \times 10^{-142}$	10^{-4}	6.5×10^{-4}
M_4	$(3.15, 5.37, 12.9, 66.9)$	$(3.0^{+1.1}_{-0.6}) \times 10^{-138}$	1.0	10^{-4}
M_5	$(3.15, 5.37, 12.9, 66.9, 399)$	$(3.0^{\times 2.1}_{\times 0.65}) \times 10^{-136}$	10^2	0.01
M_6	$(3.15, 5.37, 12.9, 34.4, 66.9, 399)$	$(6.7^{\times 2.4}_{\times 1/3}) \times 10^{-141}$	2.2×10^{-3}	0.999978

5 Conclusions

The main focus of this chapter has been on a new fusion MCMC approach to Bayesian nonlinear model fitting. In fusion MCMC the goal has been to develop an automated MCMC algorithm which is well suited to exploring multi-modal probability distributions such as those that occur in the arena of exoplanet research. This has been accomplished by the fusion of a number of different statistical tools. At the heart of this development is a sophisticated control system that automates the selection of efficient MCMC proposal distributions (including for highly correlated parameters) in a parallel tempering environment. It also adapts to any new significant parameter set that is detected in any of the parallel chains or is bred by a genetic crossover operation. This controlled statistical fusion approach has the potential to integrate other relevant statistical tools as required. A future goal is to automate the selection of an efficient set of β values used in the parallel tempering.

For some special applications it is possible to develop a faster more specialized MCMC algorithm, perhaps for dealing with real time analysis situations. In the development of fusion MCMC, the primary concern has not been speed but rather to see how powerful a general purpose MCMC algorithm we could develop and automate. In real life applications to challenging multi-modal exoplanet data fusion MCMC is proving to be a powerful tool. One can anticipate that this approach might

also allow for the joint analysis of different types of data (e.g., radial velocity, astrometry, and transit information) giving rise to statistical fusion and data fusion algorithms.

On the Bayesian model selection front, a wide variety of marginal likelihood estimators perform satisfactorily for models with ~ 7 parameters and the author has achieved satisfactory agreement between two different Bayesian estimators at the 17 parameter level. One of these, the nested restricted Mont Carlo (NRMC) method, has been investigated for up to 32 model parameters. More research aimed at the development and testing of reliable and efficient estimators that work in high dimensions is an important priority.

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

References

1. Campbell, B., Walker, G. A. H., & Yang, S.: A Search for Substellar Companions to Solar-type Stars, ApJ, **331**, 902-921 (1988)
2. Wolszczan, A., & Frail, D.: A Planetary System Around the Millisecond Pulsar PSR1257 + 12, Nature, **355**, 145-147 (1992)
3. Mayor M., Queloz D.: A Jupiter-mass companion to a solar-type star. Nature, **378**, 355-359 (1995)
4. Marcy G. W., Butler R. P.: A Planetary Companion to 70 Virginis, Astrophys. J., **464**, L147-151 (1996)
5. Udry, S., Bonfils, X., Delfosse, X., Forveille, T., Mayor, M., Perrier, C., Bouchy, F., Lovis, C., Pepe, F., Queloz, D., and Bertaux, J.-L.: The HARPS search for southern extra-solar planets. XI. Super-Earths (5 and 8 M) in a 3-planet system. Astron. Astrophys., **469**, L43L47 (2007)
6. Lovis, C., Ségransan, D., Mayor, M., Udry, S., Benz, W., Bertaux, J.-L., Bouchy, F., Correia, A.C.M., Laskar, J., Lo Curto, G., Mordasini, C., Pepe, F., Queloz, D., Santos, N.C.: The HARPS search for southern extra-solar planets. XXVIII. Up to seven planets orbiting HD 10180: probing the architecture of low-mass planetary systems. Astron. Astrophys. **528**, 112L (2011)
7. Loredo, T. L. and Chernoff, D.D.: Bayesian Adaptive Exploration. In: Feigelson, E.D., Babu, G.J. (eds.) Statistical Challenges inModern Astronomy III, pp. 5769. Springer-Verlag, New York (2003)
8. Loredo, T.: Bayesian Adaptive Exploration. In: Erickson, G.J., Zhai, Y. (eds.) Bayesian Inference And Maximum Entropy Methods in Science and Engineering: 23rd International Workshop, pp. 330346. AIP Conf. Proc., Vol. 707 (2004)
9. Cumming, A.:Detectability of extrasolar planets in radial velocity surveys , MNRAS, **354**, 1165-1176 (2004)
10. Gregory, P. C.: Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with *Mathematica* Support, Cambridge University Press (2005)
11. Gregory, P. C.: A Bayesian Analysis of Extra-solar Planet Data for HD 73526. Astrophys. J. **631**, 1198-1214 (2005)
12. Ford, E. B.: Quantifying the Uncertainty in the Orbits of Extrasolar Planets, Astron. J. **129**, 1706-1717 (2005)
13. Ford, E. B.: Improving the Efficiency of Markov Chain Monte Carlo for Analzing the Orbits of Extrasolar Planets. Astrophys. J. **642**, 505-522 (2006)

14. Ford, E. B., Gregory, P. C.: Bayesian Model Selection and Extrasolar Planet Detection. In: Babu, G.J., Feigelson, E.D. (eds.) Statistical Challenges in Modern Astronomy IV, pp. 189204. ASP Conf. Ser., Vol. 371 (2007)
15. Cumming, A., Dragomir, D.: An Integrated Analysis of Radial Velocities in Planet Searches, Mon. Not. R. Astron. Soc., **401**, 1029-1042 (2010)
16. Dawson, R. I., & Fabrycky, D. C.: Radial Velocity Planets De-aliased: A New, Short Period for Super-Earth 55 Cnc e, Astrophys. J. **722**, 937-953 (2010)
17. Gregory, P. C.: Bayesian Re-analysis of the Gliese 581 Exoplanet System. Mon. Not. R. Astron. Soc., **415**, 2523-2545 (2011)
18. Jaynes, E.T.: Bayesian Spectrum and Chirp Analysis. In: Smith, C.R., Erickson, G.L. (eds.) Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems, pp. 1-37, D. Reidel, Dordrecht (1987)
19. Bretthorst, G. L.: Bayesian Spectrum Analysis and Parameter Estimation, Springer-Verlag, New York (1988)
20. Geyer, C. J.: Markov Chain Monte Carlo. In: Keramidas, E.M. (ed.) Computing Science and Statistics, pp. 156163. Proceedings of the 23rd Symposium on the Interface, Interface Foundation, Fairfax Station (1991)
21. Hukushima, K., Nemoto, K.: Exchange Monte Carlo Method and Application to Spin Glass Simulations. Journal of the Physical Society of Japan **65**(4), 16041608 (1996)
22. Atchadé, Y. F., Roberts, G. O., & Rosenthal, J. S.: Towards optimal scaling of metropolis-coupled Markov chain Monte Carlo, Stat. Comput., **21**(4), 555-568 (2011)
23. Gregory, P. C.: A Bayesian Re-analysis of HD 11964: Evidence for Three Planets. In: Knuth, K.H., Caticha, A., Center, J.L., Giffin, A., Rodrguez, C.C. (eds.) Bayesian Inference and-Maximum Entropy Methods in Science and Engineering: 27th International Workshop, pp. 307314. AIP Conference Proceedings, Vol. 954 (2007)
24. Gregory, P. C., and Fischer, D. A.: A Bayesian Periodogram Finds Evidence for Three Planets in 47 Ursae Majoris. Mon. Not. R. Astron. Soc., **403**, 731-747, (2010)
25. Roberts, G. O., Gelman, A. and Gilks, W. R.: Weak convergence and optimal scaling of random walk Metropolis algorithms. Ann. Appl. Probab. **7**, 110120 (1997)
26. Ter Braak, C. J. F.: A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. Stat. Comput. 16, 239249 (2006)
27. Gregory, P. C.: Bayesian Exoplanet Tests of a New Method for MCMC Sampling in Highly Correlated Parameter Spaces. Mon. Not. R. Astron. Soc., **410**, 94-110 (2011)
28. Fischer,D. A., Laughlin, G. L., Butler, R. P., Marcy, G. W., Johnson, J., Henry, G.,Valenti,J., Vogt, S. S., Ammons, M., Robinson, S., Spear, G., Strader, J., Driscoll, P., Fuller, A., Johnson, T., Manrao, E., McCarthy, C., Munõz, M., Tah, K. L., Wright, J., Ida, S., Sato, B., Toyota, E., and Minniyi, D.: The N2K Consortium. I. A Hot Saturn Planet Orbiting HD 88133. Astrophys. J. **620**, 481486 (2005)
29. Gregory, P. C.: A Bayesian Kepler Periodogram Detects a Second Planet in HD 208487. Mon. Not. R. Astron. Soc., **374**, 1321-1333 (2007)
30. Jaynes, E. T.: How Does the Brain Do Plausible Reasoning?, Stanford University Microwave Laboratory Report 421, 1957, Reprinted in Erickson, G.J., Smith, C.R. (eds.) Maximum Entropy and Bayesian Methods in Science and Engineering, pp. 1-29. Kluwer Academic Press, Dordrecht (1988)
31. Tinney, C. G., Butler, R. P., Marcy, G. W., Jones, H. R. A., Penny, A. J., McCarthy, C., Carter, B. D., Fischer, D. A.: Three Low-Mass Planets from the Anglo-Australian Planet Search, Astrophys. J. **623**, 11711179 (2005)
32. Wittenmyer, R. A., Endl, M., Cochran, W. D., Levison, H. F., Henry, G. W.: A Search for Multi-Planet Systems Using the Hobby-Eberly Telescope. Astrophys. J. Suppl. **182**, 97119 (2009)
33. Mayor, M., Bonfils, X., Forveille, T., Delfosse2, X, Udry, S., Bertaux,J.-L., Beust, H., Bouchy, F., Lovis, C., Pepe, F., Perrier, C., Queloz, D., and Santos, N. C.: The HARPS search for southern extra-solar planets. XVIII. An earth-mass planet in the GJ 581 planetary system. Astron. Astrophys. **507**, 487494 (2009)

34. Vogt, S.S, Butler, R. P., Rivera, E. J., Haghighipour, N., Henry, G. W., and Williamson, M. H.: The Lick-Carnegie Exoplanet Survey: A 3.1 Earth Mass Planet in the Habitable Zone of the Nearby M3V Star Gliese 581. Astrophys. J. **723**, 954965 (2010)
35. Clyde, M. A., Berger, J. O., Bullard, F., Ford, E. B., Jeffreys, W. H., Luo, R., Paulo, R., Loredo, T.: Current Challenges in Bayesian Model Choice. In: Babu, G.J., Feigelson, E.D. (eds.) Statistical Challenges in Modern Astronomy IV, pp. 224240. ASP Conf. Ser., Vol. 371 (2007)
36. Feroz, F., Balan, S. T., Hobson, M. P.: Detecting extrasolar planets from stellar radial velocities using Bayesian evidence. Mon. Not. R. Astron. Soc. **415**, 34623472 (2011)
37. Loredo, T. L., Berger, J. O., Chernoff, D. F., Clyde, M. A., Liu, B.: Bayesian Methods for Analysis and Adaptive Scheduling of Exoplanet Observations. Stat. Meth. **9**, 101114 (2011)
38. Farr, W.M., Sravan, N., Cantrell, A., Kreidberg, L., Bailyn, C. D., Mandel, I., and Kalogera, V.: The Mass Distribution of Stellar-Mass Black Holes. Astrophys. J. **741**, 103122 (2011)
39. Gregory, P. C.: A Bayesian Periodogram Finds Evidence for Three Planets in HD 11964. Mon. Not. R. Astron. Soc. **381**, 16071619 (2007)
40. Gregory, P. C.: Discussion on paper by Martin Weinberg regarding Bayesian Model Selection and Parameter Estimation. In: Feigelson, E.D., Babu, G.J. (eds.) Statistical Challenges in Modern Astronomy V, Springer-Verlag (2012, in press)
41. Jefferys, W. H.: Discusion on Current Challenges in Bayesian Model Choice by Clyde et al. In: Babu, G.J., Feigelson, E.D. (eds.) Statistical Challenges in Modern Astronomy IV, pp. 241244. ASP Conf. Ser., Vol. 371, (2007)