Mon. Not. R. Astron. Soc. **000**, 1–19 (2011) Printed 3 December 2011 (MN LATEX style file v2.2)

Bayesian Exoplanet tests of a new method for MCMC sampling in highly correlated model parameter spaces

Philip C. Gregory[1⋆]
[1] *Physics and Astronomy Department, University of British Columbia, 6224 Agricultural Rd., Vancouver, BC V6T 1Z1, Canada*

Mon. Not. R. Astron. Soc., Vol. 410, 94-110, 2011
The definitive version is available at www.blackwell-synergy.com

ABSTRACT
The Markov chain Monte Carlo (MCMC) method is a powerful technique for facilitating Bayesian nonlinear model fitting. In many cases the MCMC exploration of the parameter space is very inefficient because the model parameters are highly correlated. Differential evolution MCMC is one technique that addresses this problem by employing multiple parallel chains. We present a new method that automatically achieves efficient MCMC sampling in highly correlated parameter spaces which does not require additional chains to accomplish this. It was designed to work with an existing hybrid Markov chain Monte Carlo algorithm which incorporates parallel tempering, simulated annealing and genetic crossover operations. These features, together with the new correlated parameter sampler, greatly facilitate the detection of a global minimum in χ^2. The new HMCMC algorithm is very general in scope. Two tests of the algorithm are described employing (a) exoplanet precision radial velocity data, and (b) simulated space astrometry data. The latter test explores the accuracy of parameter estimates obtained with the Bayesian HMCMC algorithm on the assumed astrometric noise.

Key words: stars: planetary systems; methods: statistical; methods: numerical; techniques: radial velocities, astrometry.

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 450 planets discovered to date, starting from the pioneering work of Campbell, Walker & Yang (1988), Wolszczan & Frail (1992), Mayor & Queloz (1995), and Marcy & Butler (1996). One example of the fruits of this work is the detection of a super earth in the habitable zone surrounding Gliese 581 (Udry et al. 2007). This and other remarkable successes on the part of the observers has spurred a significant effort to improve the statistical tools for analyzing data in this field (e.g., Loredo & Chernoff 2003, Loredo 2004, Cumming 2004, Gregory 2005a & b, Ford 2005 & 2006, Ford & Gregory 2006, Cumming & Dragomir 2010). Much of the recent 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 MCMC (HMCMC) algorithm that incorporates parallel tempering (PT), simulated annealing and a genetic crossover operation to facilitate the detection of a global minimum in χ^2. This enables the efficient exploration of a large model parameter space starting from a random location. It is able to identify any significant periodic signal component in the data that satisfies Kepler's laws and is able to function as a multi-planet Kepler periodogram [1]. In addition, the Bayesian MCMC algorithm provides full marginal parameters distributions. The algorithm includes an innovative two stage adaptive control system that automates the selection of efficient Gaussian parameter proposal distributions. 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 ~ 10000 days.

Most of the existing applications have been to radial velocity data (Gregory 2005a, b and 2007a, b, c) but the HMCMC algorithm is intended as a general Bayesian nonlinear model fitting tool. 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

⋆ E-mail: gregory@phas.ubc.ca

[1] Following on from the pioneering work on Bayesian periodograms by Jaynes (1987) and Bretthorst (1988)

© 2011 RAS

can be very inefficient. In certain cases, this difficulty can be addressed by a simple transformation to a more orthogonal parameter space (e.g., Ford 2006). Another solution to this problem is Differential Evolution Markov Chain (DE-MC) (TerBraak 2006). DE-MC is a population 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. In DE-MC the proposed jumps are simply a fixed multiple of the differences of two random parameter vectors that are currently in the population. The current HMCMC algorithm already runs parallel tempering chains to avoid becoming trapped in a local probability maximum. To increase the number of chains by a further factor of 15 to 40, to accomplish DE-MC, would not be practical. In this paper, we present a new method in the spirit of DE that automatically achieves efficient MCMC sampling in highly correlated parameter spaces without the need for additional chains.

In the next section the existing HMCMC algorithm and its adaptive control system are reviewed. In Section 3 the new addition to handle highly correlated parameter spaces is presented. Section 4 deals with the first test of the algorithm using precision radial velocity exoplanet data and a single planet model with 7 parameters. Section 5 presents the results of a second test using simulated space astrometry data for a two planet model with 20 parameters. In Section 6 we compare the new method for handling correlated parameter spaces with an alternate approach that employs the covariance matrix of previously accepted MCMC samples.

2 REVIEW OF ADAPTIVE HYBRID MCMC

The HMCMC 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

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

where C is the normalization constant and $\vec{X}$ represent the set of model parameters. The first term on the RHS of the equation, $p(\vec{X}|M_i, I)$, is the prior probability distribution of $\vec{X}$, prior to the consideration of the current data D. The second term, $p(D|\vec{X}, M_i, I)$, 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 $\vec{X}$, 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 problem 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 function (PDF) of the orbital period in a one planet radial velocity model fit is given by

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

where $p(P, K, V, e, \chi, \omega, s|D, M_1, I)$ is the target joint probability distribution of the radial velocity model parameters $(P, K, V, e, \chi, \omega)$ and s is an extra noise parameter which is discussed in Section 5.1. $p(P|M_1, I)$ 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, i.e., Monte Carlo simulations. Bayesian model selection requires integrating over all the model parameters.

In high dimensions, the principle tool for carrying out the integrals is Markov chain Monte Carlo 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 a target posterior density for that region. The random walk is accomplished using a Markov chain, whereby the new sample, $\vec{X}_{t+1}$, depends on previous sample $\vec{X}_t$ according to a time independent entity called the *transition kernel*, $p(\vec{X}_{t+1}|\vec{X}_t)$. The remarkable property of $p(\vec{X}_{t+1}|X_t)$ is that after an initial burn-in period (which is discarded) it generates samples of $\vec{X}$ with a probability density equal to the desired posterior $p(\vec{X}|D, M_1, I)$ (e.g., see Chapter 12 of Gregory 2005 for details).

The transition kernel, $p(\vec{X}_{t+1}|\vec{X}_t)$ is given by

$$p(\vec{X}_{t+1}|\vec{X}_t) = q(\vec{X}_{t+1}|\vec{X}_t)\alpha(\vec{X}_t, \vec{X}_{t+1}), \quad (3)$$

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

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

where $q(\vec{X}_t|\vec{X}_{t+1}) = q(\vec{X}_{t+1}|\vec{X}_t)$ for a symmetrical proposal distribution. If the proposal is not accepted the current sample $\vec{X}_t$ is repeated.

An important feature that prevents the hybrid MCMC from becoming stuck in a local probability maximum is parallel tempering (Geyer (1991) and reinvented by Hukushima & Nemoto (1996) under the name *exchange Monte Carlo*). Multiple MCMC chains are run in parallel. The joint distribution for the parameters ($\vec{X}$) of model M_i, for a particular chain, is given by

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

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, $p(\vec{X}|D, M_i, I)$.

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 8 chains with $\beta = \{0.09, 0.13, 0.20, 0.29, 0.39, 0.52, 0.72, 1.0\}$ were employed. 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. For $\beta \ll 1$, the distribution is much flatter. 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. A swap acceptance rate of $\approx 40\%$ works 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 (centered on the current parameter location), one for each parameter. In general, the σ's of these Gaussian proposal distributions are different because the parameters can be very different entities. If the σ's are chosen too small, successive samples will be highly correlated and will require many iterations to obtain an equilibrium set of samples. If the σ's are too large, then proposed samples will very rarely be accepted. The process of choosing a set of useful proposal σ's 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 σ's for each tempering chain. This process is automated by an innovative two stage statistical control system (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, typically 25% (Roberts et al. 1997). A schematic of the existing adaptive control system (CS) is shown in Fig. 1. An additional module that adds the correlated parameter proposal scheme is described in Section 3.

The first stage CS, which involves annealing the set of Gaussian proposal distribution σ's, was described in Gregory 2005a. An initial set of proposal σ's ($\approx 10\%$ of the prior range for each parameter) are used for each chain. During the major cycles, the joint acceptance rate is measured based on the current proposal σ's and compared to a target acceptance rate. During the minor cycles, each proposal σ is separately perturbed to determine an approximate gradient in the acceptance rate. The σ's are then jointly modified by a small increment in the direction of this gradient. This is done for each of the parallel chains. Proposals to swap parameter values between chains are allowed during major cycles but not within minor cycles.

The annealing of the proposal σ's occurs while the MCMC 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 fit values. This automatically arises because all the models considered incorporate an extra additive noise term (Gregory 2005b), for reasons discussed in Section 5.1, 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-fit parameters. An example of this is shown in Fig. 2 for a two planet fit to simulated astrometry data as discussed in Section 5. In the early stages the extra noise, which is labeled s_A, is inflated to around 1000 μas and then decays to a very small value of ≈ 0.02 μas over the first 14,000 iterations (actually 140,000 iterations since the stored iterations are thinned by a factor of 10). This is similar to simulated annealing, but does not require choosing a cooling scheme. The lower panel of the fit shows the evolution of the two orbital period parameters.

Although the first stage CS achieves the desired joint acceptance rate, it often happens that a subset of the proposal σ's are too small leading to an excessive autocorrelation in the MCMC iterations for these parameters. Part of the second stage CS corrects for this. The goal of the second stage is to achieve a set of proposal σ's that equalizes the MCMC acceptance rates when new parameter values are proposed separately and achieves the desired acceptance rate when they are proposed jointly. Details of the second stage CS were 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 burn-in period occurs within the span of the first stage CS, i.e., the significant peaks in the joint parameter probability distribution are found, and the second stage improves the choice of proposal σ's based on the highest probability

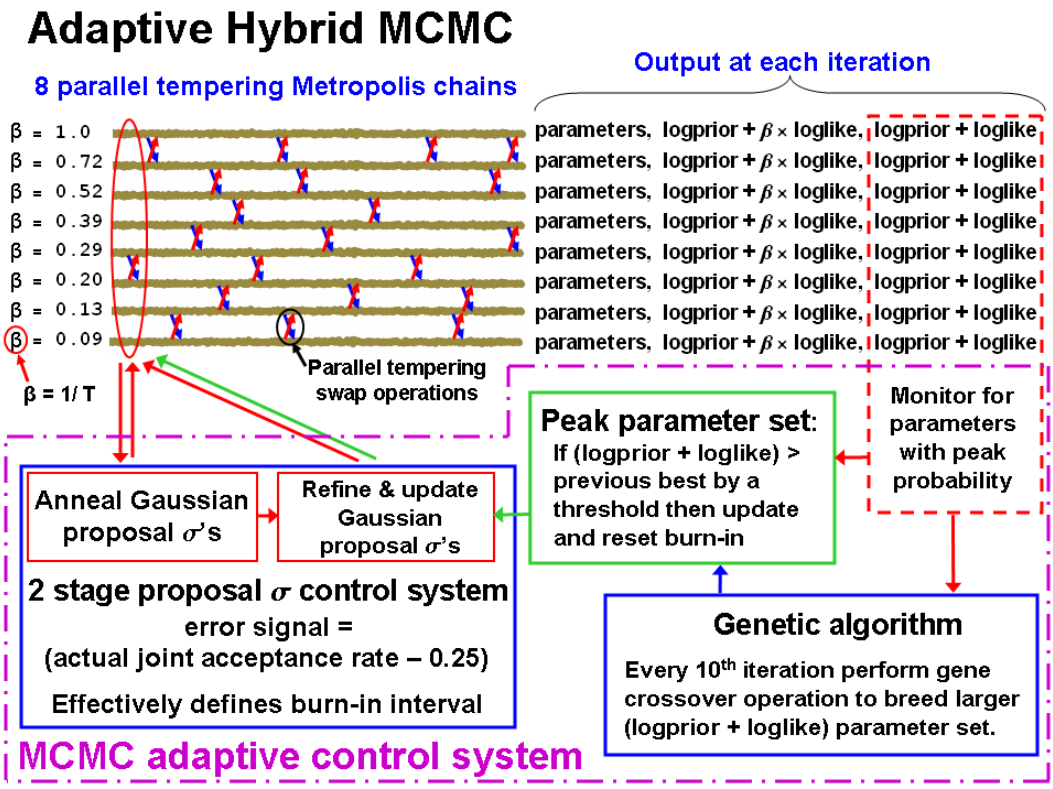

Figure 1. Schematic of the operation of the existing adaptive hybrid MCMC algorithm. The new correlated sampler proposal system is an additional module that is described in the Section 3.

parameter set. Occasionally, a new higher (by a user specified threshold) target probability parameter set emerges after the first two stages of the CS are completed. The control system 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 control system is reset. The requirement that the transition kernel be time independent means that $q(\vec{X}_{t+1}|\vec{X}_t)$ be time independent, so useful MCMC simulation data are obtained only after the CS is switched off.

The adaptive capability of the control system can be appreciated from an examination of Fig. 1. The upper left portion of the figure depicts the MCMC 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.

The CS also includes a genetic algorithm block which is shown in the bottom right 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 $\vec{X}_{\rm max}$, and the current most probable parameter set of the 8 chains, $\vec{X}_{\rm cur}$. At regular intervals (user specified) each gene from $\vec{X}_{\rm cur}$ is swapped for the corresponding gene in $\vec{X}_{\rm max}$. If either substitution leads to a higher probability it is retained and $\vec{X}_{\rm max}$ 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 $\vec{X}_{\rm max}$ compared to the number of new $\vec{X}_{\rm max}$ arising from the normal parallel tempering MCMC iterations. The gene crossover operations prove to be very effective, and give rise to new $\vec{X}_{\rm max}$ 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.

Gene swaps from $\vec{X}_{\rm cur2}$, the parameters of the second most probable current chain, to $\vec{X}_{\rm max}$ are also utilized. This gives rise to new values of $\vec{X}_{\rm max}$ at a rate approximately half that of swaps from $\vec{X}_{\rm cur}$ to $\vec{X}_{\rm max}$. 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.

3 ADDITION TO HYBRID MCMC TO HANDLE CORRELATED PARAMETERS

In the original adaptive hybrid MCMC algorithm, new parameter moves were jointly proposed based on independent Gaussian proposal distributions, one for each parameter.

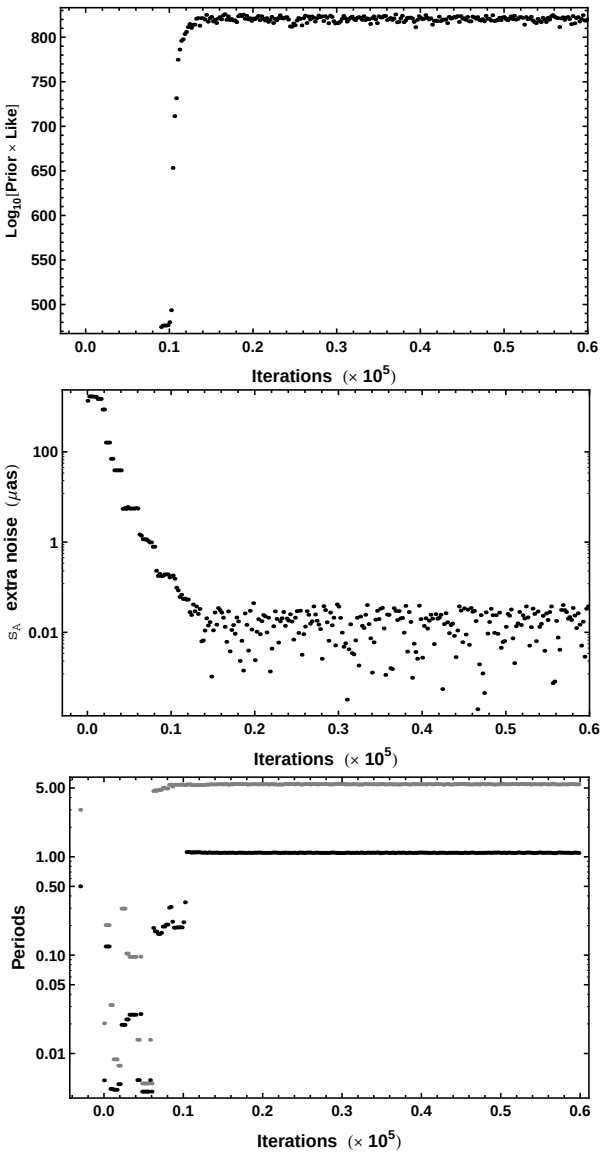

Figure 2. The upper panel is a plot of the Log_{10}[Prior × Likelihood] versus MCMC iteration for a 2 planet fit of the simulated astrometry data discussed in Section 5. The middle panel is a similar plot for the extra noise term s_A. Initially s_A is inflated and then rapidly decays to a much lower level as the best fit parameter values are approached. The lower plot shows the evolution of the two period parameters of the fit. The two starting periods are shown on the left hand side of the plot at a negative iteration number.

This will be referred to as the 'I' proposal scheme. In some problems the model parameter estimates are highly correlated and the 'I' proposal scheme can result in a very inefficient exploration of the parameter space. An example of this can be seen in the upper panel of Fig. 3, which shows the joint marginal distribution of the parameters, χ and ω, used in a one planet model of the precision radial velocity data for HD 88133. The star is known to have a planet with a very low eccentricity (Fischer et al. 2005). For a low eccentricity orbit, periastron is not well defined and so χ and ω are separately not well defined.

To achieve a 25% acceptance rate with the 'I' proposals,

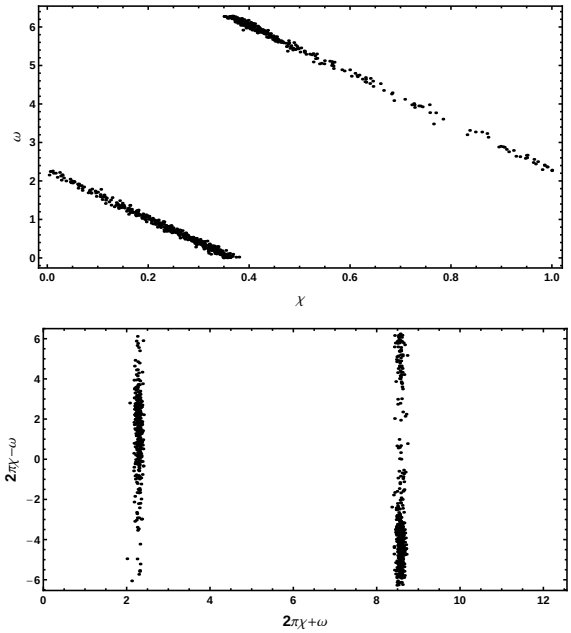

Figure 3. The upper panel shows the strong correlation present in the joint marginal distribution of two of the parameters, χ and ω, used in a one planet model of the radial velocity measurements of HD 88133. A simple transformation to two other parameters $\psi = 2\pi\chi + \omega$ and $\phi = 2\pi\chi - \omega$ eliminates this correlation as shown in the lower panel.

the Gaussian proposal σ's for the χ and ω parameters must be very small necessitating very slow movement along the correlation diagonal. One solution commonly adopted is to transform χ and ω to two more orthogonal parameters

$$\begin{aligned} \psi &= 2\pi\chi + \omega, \\ \phi &= 2\pi\chi - \omega. \end{aligned} \quad (6)$$

The distribution of the transformed parameters is shown in the lower panel of Fig. 3. In this case ψ occupies two narrow regions. ϕ is not well determined but is at least orthogonal to ψ. In other cases the correlation between parameters can be much more banana shaped so a simple transformation is not as powerful.

In DE-MC (TerBraak 2006) the proposed jumps are simply a fixed multiple of the differences of two random parameter vectors that are currently in a population generated from a set of independent MCMC chains. Here we describe a modification to the hybrid MCMC algorithm that automatically achieves efficient MCMC sampling in highly correlated parameter spaces. This is achieved by an additional control system module that generates a second 'C' proposal distribution (utilized 50% of the time), which like DE-MC reflects the appropriate scale and orientation for the jumping distribution but does not require multiple chains to be run in parallel. In fact we do employ multiple parallel chains but that is to accomplish the parallel tempering function. A schematic of the full adaptive control system (CS) is shown in Fig. 4. The existing 'I' proposal scheme is represented by the module at the lower left of the panel. The 'C' proposal scheme is the next module to the right, and the rest of the schematic illustrates how the correlated proposal distribution is generated.

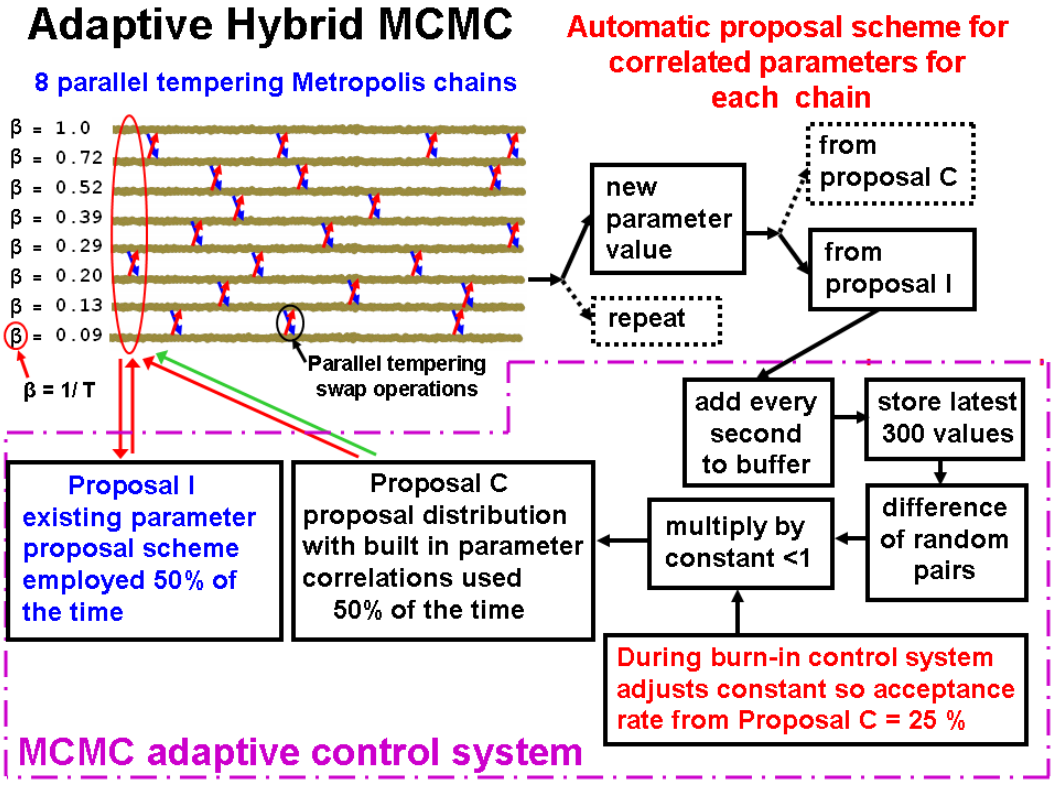

Figure 4. Schematic of the modified adaptive hybrid MCMC algorithm that includes the new correlated parameter proposal scheme.

A separate 'C' proposal distribution is generated for each parallel chain in the following way. At each MCMC iteration and in each chain, a proposal is made to jump to a new parameter set. A flag indicates if the proposal is accepted ($\sim 25\%$ of the time) and whether the proposed move was taken from the 'I' or 'C' proposal distribution. Initially, only the 'I' proposal system is used. It is clear that if there are strong correlations between the parameters, the accepted 'I' proposals will generally lie along the correlation path. Every second accepted 'I' proposal is added to a buffer called the correlated sample buffer. Only the *nmov* most recent additions to the buffer are retained. Typically $nmov = 300$. Once the buffer contains *nmov* accepted 'I' proposals, a pair of randomly selected parameter vectors can be drawn from the buffer and the difference between them used in a 'C' proposal after multiplication by a constant, *corscale*. Actually, *corscale* is a vector of constants, one for each parallel tempering chain. *corscale* is updated from an initial default value of 0.2 by another block of the control system which is designed to achieve a 'C' proposal acceptance rate of ~ 0.25 in each chain [2].

[2] This rescaling operation makes use of another vector named *acorsub* which keeps track of the fraction of the 'C' proposals that are accepted in each chain.

$$acorsub = (nacor - nacorsub)/(npropP - npropPsub), \qquad (7)$$

where *nacor* = the total number of accepted 'C' proposals, *nacorsub* = the number of accepted 'C' proposals immediately prior to the beginning of the latest *corscale* rescaling operation, *npropP* = the total number of 'C' proposals, and *npropPsub* = the number of 'C' proposals made immediately prior to the beginning of the latest *corscale* rescaling operation. During a rescaling operation, *acrosub* and *corscale* are computed after *nmov* additional accepted 'I' samples are added to the correlation buffer. The *corscale* vector is rescaled according to equation (8).

$$corscale = \text{previous } corscale \times \left[\frac{(acorsub + \Delta)}{0.25} \times \frac{0.75}{(1 - acorsub + \Delta)}\right]^{1/4}, \qquad (8)$$

where we use a $\Delta = 0.01$.

If $acorsub = 0.25$, then equation (8) leaves the proposal *corscale* unchanged except for the small effect of the Δ term. The Δ term is there to handle the extremes of $acorsub = 0$ and 1 gracefully. If $acorsub > 0.25$, equation (8) will cause *corscale* to increase which reduces *acorsub*, the 'C' proposal acceptance rate. Similarly if $acorsub < 0.25$ equation (8) will cause *corscale* to decrease which increases *acorsub*. In practice we iterate the rescaling operation until $0.22 < acorsub < 0.28$. After this no further rescaling occurs unless the second stage control system is re-started.

Although the buffer contains recent samples from the same chain, we are not using the samples directly. Only the differences [3] of randomly selected pairs of buffer samples are employed to provide for a scale and direction for proposed

[3] A complication arises in taking the differences of angular parameters since they are wrap around continuous. For example, if the peak of the parameter distribution occurs near the top end, the tail of this distribution will wrap around into the lower end. Thus one of the pair of buffer samples to be differenced can land at one end of the range and the other fall at the other end. The two samples can actually be very close in non wrapped space but are far apart in the wrapped space. It is important to use the

future jumps. This is illustrated in Fig. 5 which shows the relationship between two of 20 model parameters employed in the astrometry test described in Section 5. The upper panel shows a plot of an eccentricity parameter, e_2, versus orbital frequency, f_2, for the *nmov* samples in the correlated sample buffer. The lower panel displays the correlation between these two parameters when we plot the distribution of the differences of 1000 randomly selected pairs of samples from the buffer after multiplication by *corscale*. Each point in this plot gives the increment $(\Delta e_2, \Delta f_2)$ in each parameter corresponding to a potential 'C' proposal. A plot like the one shown in the lower panel in Fig. 5 will be referred to as an increment 'C' proposal correlation plot.

The final proposal distribution is a random selection of 'I' and 'C' proposals such that each is employed 50% 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 MCMC 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. This helps to greatly reduce the burn-in period. Once the first two stages of the control system have been turned off, the third stage continues until a minimum of an additional *nmov* post burn-in accepted 'I' proposals have been added to the buffer and the 'C' proposal acceptance rate is within the range $\geqslant 0.22$ and $\leqslant 0.28$. At this point further additions to the buffer are terminated. Because of the way the 'C' proposal are generated the resulting 'C' proposal distribution is necessarily symmetric and reversible. One way to satisfy the requirement that the transition kernel (equation 3) be time independent is to only make use of MCMC samples taken after we have stopped adding to the 'C' proposal system buffer. The need to stop adding to the 'C' proposal buffer may however be overly restrictive. Once burn-in has been achieved, there should be no statistical difference between 'C' proposals derived from correlated sample buffer values collected during different post burn-in time intervals. We explore this question further in the algorithm testing sections below.

Because of the adaptive nature of the control system, if the parallel tempering, genetic crossover operation or peak finding routine leads to a new parameter set which is significantly more probable than the current most probable set, the control system switches on again (as described in the previous section) and new accepted 'I' samples are added to the 'C' sample buffer as outlined above.

4 TEST OF ALGORITHM WITH RADIAL VELOCITY DATA

In one test of the new algorithm we analyzed a sample of seventeen HD 88133 precision radial velocity measurements (Fischer et al. 2005) using a single planet model in three different ways. In case (1) the analysis employed the original

smaller of these two possible differences when constructing 'C' proposal moves.

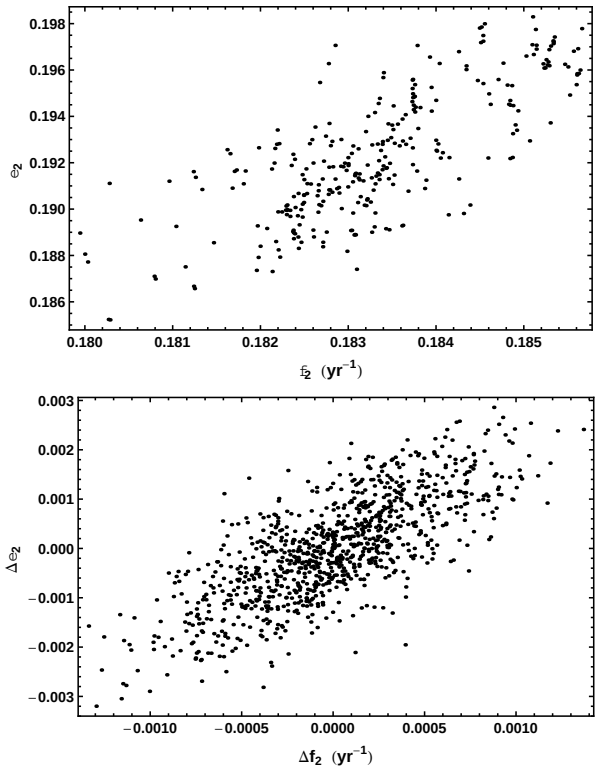

Figure 5. The upper panel shows a plot of eccentricity e_2 versus orbital frequency f_2 for the *nmov* accepted samples in the correlated sample buffer. The lower panel displays the distribution of the differences of 1000 randomly selected pairs of samples from the buffer after multiplication by a constant.

highly correlated χ and ω parameters using only 'I' proposals. In case (2), both 'I' and 'C' proposal were used as described above. In case (3), the search was carried out using the orthogonal transformed parameters $\psi = 2\pi\chi + \omega$ and $\phi = 2\pi\chi - \omega$ with only 'I' proposals.

For the one planet model the predicted radial velocity is given by

$$v(t_i) = V + K[\cos\{\theta(t_i + \chi P) + \omega\} + e\cos\omega], \tag{9}$$

and involves the 6 unknown parameters

V = a constant velocity.
K = velocity semi-amplitude.
P = the orbital period.
e = the orbital eccentricity.
ω = the longitude of periastron.
χ = the fraction of an orbit, prior to data reference epoch, 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.

For details concerning the implementation strategy and parameter priors see Gregory & Fischer (2010). All the calculations were implemented in *Mathematica* using parallized code run on an 8 core PC.

Fig. 6 shows a comparison of the resulting post burn-in marginal distributions for χ and ω together with a comparison of the autocorrelation functions. The black trace corre-

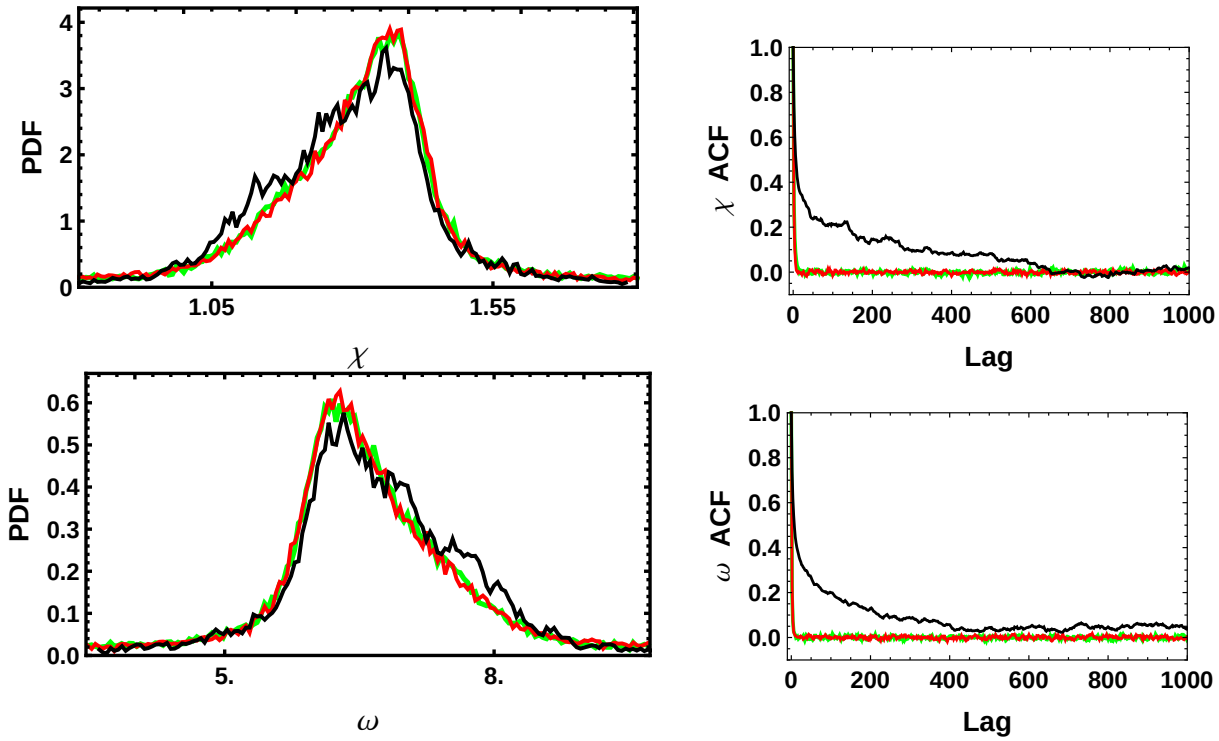

Figure 6. The two panels on the left show a comparison of the post burn-in marginal distributions for χ and ω. The two panels on the right show a comparison of their MCMC autocorrelation functions. The black trace corresponds to a search in χ and ω using only 'I' proposals. The red trace corresponds 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.

sponds 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.

Once burn-in has been achieved, there should be no statistical difference between 'C' proposals derived from correlated sample buffer values collected during different post burn-in time intervals. We tried a run on HD 88133 data in which we continued to add to the correlated sample buffer throughout the entire run and obtained parameter marginals and autocorrelation results that appeared identical with those obtained when the data taking occurred only after we stopped adding to the correlated sample buffer.

5 TEST OF ALGORITHM WITH SIMULATED ASTROMETRY DATA

In 2008, a NASA JPL simulation study (Traub et al. 2009) was conducted to see how well Earth-like planets (i.e., terrestrial masses, habitable-zone periods) in multi-planet systems could be detected using astrometric and radial velocity (RV) observations. An additional goal was to see what accuracy of SIM Lite is needed to detect Earth-like planets. Six teams of scientists competed in the double blind analysis of the data. The author had intended to participate using the Bayesian hybrid MCMC (HMCMC) model fitting algorithm but due to strong correlations between a number of the astrometric model parameters the algorithm took far too long for convergence to be useful in its existing form. The challenge to deal with multi-planet astrometric data was one of the main reasons for the development of the new 'C' proposal scheme described in Section 3. In this section we illustrate the performance of the improved HMCMC which incorporates the 'C' proposal scheme.

An initial noise free simulation was used to test that the model astrometry equations (which included the stars proper motion, parallax and positional offset as free parameters) corresponded to the simulation. Next a variant of another simulation was constructed that contained two planets with masses of 0.922 and 12.124 M_E and periods of 1.09705 and 5.43505 yr, respectively. No simulated RV data was used in these test.

The full list of simulated model parameters [4] is given in Table 1. Three different versions of this astrometry simulation were used each containing a different amount of additive Gaussian noise with a $\sigma = 0.1, 0.5, 0.8\mu$as, respectively. These noise levels were added to both astrometry coordinates. The simulated astrometry data, with parallax and proper motion removed, together with our final HM-

[4] As explained by Traub et al. (2009), a number of astrophysical effects were explicitly ignored in the simulations because they anticipate these can be removed from the data with essentially perfect accuracy: relativistic effects in the orbit of the astrometric spacecraft, aberration of light, deflection of light by Jupiter and other bodies. These effects were also ignored in the present work.

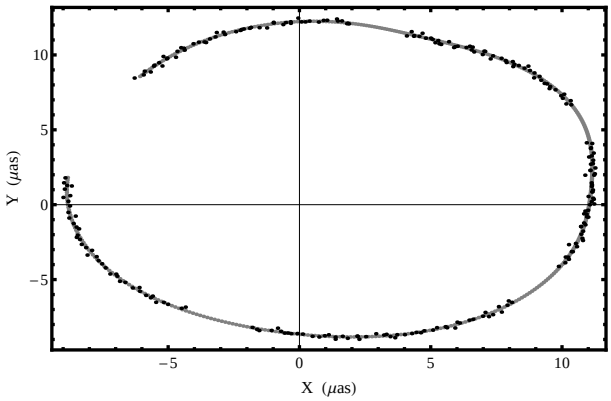

Figure 7. The astrometry data, with parallax and proper motion removed, together with our final HMCMC mean fit for the 0.1 μas noise level.

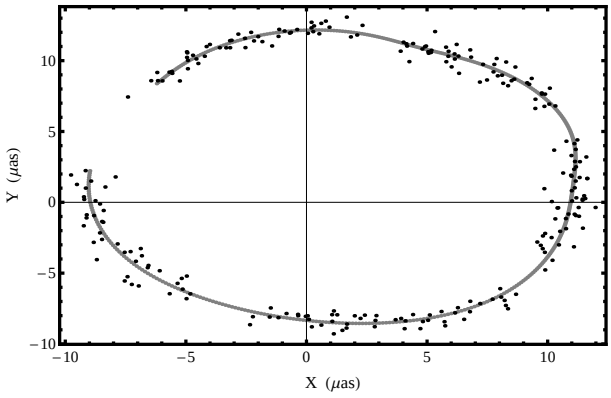

Figure 8. The astrometry data, with parallax and proper motion removed, together with our final HMCMC mean fit for the 0.5 μas noise level.

CMC mean fit (employing the 'I' and 'C' proposal scheme) is shown in Figures 7 and 8 for the two noise levels of 0.1 and 0.5 μas, respectively. The data consist of 250 observations over 5 years with gaps as shown. The time interval between observations is a constant plus or minus a random component. The main purpose of these simulations is to test the effectiveness of the new combined 'I' and 'C' proposal scheme. As a by product, the analysis also yields information about the Bayesian marginal parameter distributions achievable for different noise levels based on the analysis of only the astrometry data.

The ecliptic longitude and latitude of the target star was $\lambda = 1.0344$ and $\beta = 0.53465$(rad), respectively. The astrometric displacement caused by the proper motion and parallax are $\sim 10^5$ times larger than the reflex motion caused by the planets. As a first step in the analysis, a first order removal of the proper motion, parallax and offsets X_0, Y_0 was carried by a Nelder-Mead Nelder-Mead (1965) fit of these terms. In real data, the offset terms will generally differ from zero and therefore we included them as additional parameters. Their inclusion can introduce correlations with other parameters which can significantly affect the final parameter error estimates.

The next phase was to use the Bayesian HMCMC to fit a two planet astrometric model which also included residual proper motion, residual parallax and residual offset parameters, $\Delta\mu_x, \Delta\mu_y, \Delta\varpi, \Delta X, \Delta Y$, respectively. In keeping with our previous radial velocity analysis (e.g., Gregory & Fischer (2010)) we introduce the parameter χ = the fraction of an orbit, prior to the reference data time (mean of observation times), that periastron occurred at.

$$\begin{aligned} \chi_1 &= \text{FractionalPart}[(\text{Mean}[\text{time}] - T_{0_1})/P_1], \\ \chi_2 &= \text{FractionalPart}[(\text{Mean}[\text{time}] - T_{0_2})/P_2]. \end{aligned} \quad (10)$$

As in Section 4, the HMCMC search was carried out using the transformed parameters $\psi = 2\pi\chi + \omega$ and $\phi = 2\pi\chi - \omega$. 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). The big (ψ, ϕ) square holds two copies of the probability patch in (χ, ω) which doesn't matter. What matters is that the prior is now wraparound continuous in (ψ, ϕ). For the purpose of plotting results, ψ and ϕ were transformed back to χ and ω.

5.1 Priors

A list of the parameter priors adopted for this analysis is given in Table 2.

For the Kepler model with sparse data, the target probability distribution can be very spiky. This is particularly a problem for the orbital period parameters which span 4 decades. Instead of orbital period the search is carried out in frequency space for the following reasons. In a Bayesian analysis, the width of a spectral peak, which reflects the accuracy of the frequency estimate, is determined by the duration of the data, the signal-to-noise (S/N) ratio and the number of data points. More precisely (Gregory 2005a, Bretthorst 1988), for a sinusoidal signal model, the standard deviation of the spectral peak, δf, for a S/N > 1, is given by

$$\delta f \approx \left(1.6 \frac{\text{S}}{\text{N}} \mathcal{T}\sqrt{N}\right)^{-1} \quad \text{Hz}, \quad (12)$$

where $\mathcal{T}$ = the data duration in s, and N = the number of data points in $\mathcal{T}$. The thing to notice is that the width of any peak is independent of the frequency of the peak. Thus the same frequency proposal distribution will be efficient for all frequency peaks. This is not the case for a period search where the width of a spectral peak is $\propto P^2$. Not only is the width of the peak independent of f, but the spacing of peaks is constant in frequency (roughly $\Delta f \sim 1/T$), which is another motivation for searching in frequency space (e.g., Scargle 1982, Cumming 2004).

Gregory (2007a) discussed two different strategies to search the orbital frequency parameter space for a multi-planet model: (i) an upper bound on $f_1 \leqslant f_2 \leqslant \cdots \leqslant f_n$ 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-labeled 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.

Table 1. Model parameter values.

Parameter	value	Parameter	value	Meaning
M_1 (M_E)	0.922	M_2 (M_E)	12.124	planetary masses
P_1 (yr)	1.09705	P_2 (yr)	5.43505	planetary periods
a_1 (au)	1.0637	a_2 (au)	3.0913	planetary semi-major axes
e_1	0.053791	e_2	0.194098	orbital eccentricities
i_1 (rad)	2.53204	i_2 (rad)	2.82563	orbital inclinations
Ω_1 (rad)	0.738587	Ω_2 (rad)	6.04406	longitudes of ascending nodes
ω_1 (rad)	3.89274	ω_2 (rad)	2.18914	longitudes of periastron
T_{0_1}	2018.37	T_{0_2}	2018.52	periastron passage times
X_0 (mas)	0.0	Y_0 (mas)	0.0	systematic astrometry offset errors
μ_x (mas/yr)	1169.86174	μ_x (mas/yr)	-482.90480	proper motion
ϖ (mas)	94.9			parallax
t_0	2020.33			epoch of proper motion = mean of observation times
M_* (M$\odot$)	1.0			stellar mass

The analysis also includes an extra noise parameter, s_A, 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_A^2. Thus, the combination of the known errors and extra noise has a Gaussian distribution with variance $= \sigma_i^2 + s_A^2$, where σ_i is the standard deviation of the known noise for i$^{\rm th}$ data point. For example, suppose that the star actually has three planets, and the model assumes only two are present. In regard to the two planet model, the astrometric variations induced by the unknown third planet acts like an additional unknown noise term. Other factors like star spots and chromospheric activity can also contribute to this extra noise term. In general, nature is more complicated than our model and known noise terms. Marginalizing s_A 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 Gregory (2005a) for a tutorial demonstration of this point. If there is no extra noise then the posterior probability distribution for s_A will peak at $s_A = 0$. We employed a modified Jeffrey's prior (Gregory 2005b and caption of Table 2) with a knee, $s_0 = 1\mu$as. Finally, as was pointed out in Section 2, the presence of the extra Gaussian noise term with unknown σ automatically gives rise to a very useful annealing operation when the Markov chain is started from a location in parameter space that is far from the best fit values.

5.2 Results

Prior to incorporating the 'C' proposals, we managed to achieve MCMC convergence for the 0.5 μas noise level astrometry data after a long burn-in period, but the parameter traces exhibited very large autocorrelation times. With the inclusion of both 'I' and 'C' proposals the modified HMCMC algorithm quickly converged on the correct orbital periods of 1.1 and 5.4 yr for the simulations with 0.1 and 0.5 μas noise levels, starting from initial periods of 0.5 and 3 yr. The upper panel in Fig. 2 shows the Log$_{10}$[Prior $\times$ Likelihood] versus MCMC iteration for the 2 planet fit of the 0.1 μas noise level simulation. The middle panel shows the extra noise term s_A which is initially inflated and then rapidly decays to a much lower level as the best fit parameter values are approached. The result indicates no extra noise which is consistent with the simulation. The lower plot shows the evolution of the two period parameters of the fit starting from initial periods of $P_1 = 0.5$ and $P_2 = 3$ yr.

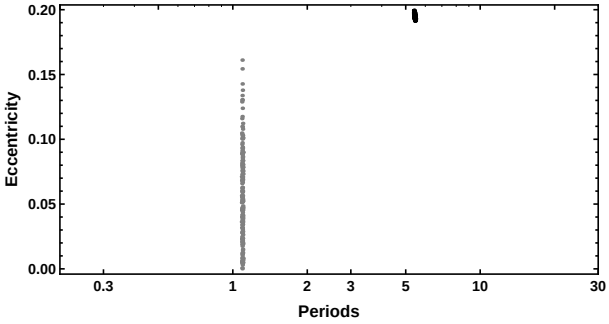

Figure 9. A plot of eccentricity versus period for the 2 planet fit for a simulation noise level of 0.1 μas.

Figures 9, 10 and 11 show post burn-in eccentricity versus period plots for the 2 planet fits for the 3 noise levels. For the simulations with noise levels of 0.1 and 0.5 μas, the two correct orbital periods were readily detected. For the 0.8 μas noise level simulation, the solution yielded the expected value of P_2 but with much larger uncertainty. P_1 exhibited a number of peaks. One of these peaks was close to the expected P_1 of 1.1 yr, but the period with the highest value of log$_{10}$[prior $\times$ likelihood] was 2.37 yr or close to the second harmonic. As expected the spread in eccentricity and period values decreased as the noise level was reduced.

Fig. 12 shows the HMCMC post burn-in parameter iterations for the 0.1 μas noise level simulation. For each parallel chain, 10^6 iterations were executed but only every tenth iteration was stored. In this run, accepted 'I' samples were continually added to the correlated sample buffer up to stored iteration number 50000 in the figure. After iteration 50000 no new samples were added to the correlated sample buffer.

[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, i.e., the maximum entropy distribution (Jaynes 1957, Gregory 2005a section 8.7.4.)

Table 2. Prior parameter probability distributions.

Parameter	Prior	Lower bound	Upper bound
Orbital frequency	$p(\ln f_1, \ln f_2, \cdots \ln f_n \vert M_n, I) = \frac{n!}{[\ln(f_H/f_L)]^n}$ (n =number of planets)	1.1/365.25 yr	30 yr
e_i Eccentricity	Uniform	0	1
a_{star_i} Star reflex motion	Modified Jeffreys [a] $\frac{(a_{\mathrm{star}}+a_0)^{-1}}{\ln\left(1+\frac{a_{\mathrm{star}_{\mathrm{max}}}}{a_0}\right)}$	0 (au) $a_0 = 3\times10^{-6}$ (au)	0.1 (au)
i_i Inclination	Uniform	0	π
χ_i (equ. 10)	uniform	0	1
ω_i Longitude of periastron	Uniform	0	2π
ψ_i (equ. 6)	uniform	0	4π
ϕ_i (equ. 6)	uniform	-2π	2π
Ω_i Longitude of ascending node	Uniform	0	2π
$\Delta\mu_x$ residual proper (mas/yr)	Uniform motion	-10300.0	10300.0
$\Delta\mu_y$ residual proper (mas/yr)	Uniform motion	-10300.0	10300.0
$\Delta\varpi$ residual parallax (mas)	Uniform	-20	20.0
ΔX (X offset) (mas)	Uniform	-10.0	10.0
ΔY (Y offset) (mas)	Uniform	-10.0	10.0
s_A Extra noise (μas)	$\frac{(s+s_0)^{-1}}{\ln\left(1+\frac{s_{\mathrm{max}}}{s_0}\right)}$	0 ($s_0 = 1$)	3000

[a] Since the prior lower limits for a_{star} and s_A include zero, we used a modified Jeffreys prior of the form

$$p(X|M,I) = \frac{1}{X+X_0}\,\frac{1}{\ln\left(1+\frac{X_{\mathrm{max}}}{X_0}\right)} \tag{11}$$

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 $\ln\left(1+\frac{X_{\mathrm{max}}}{X_0}\right)$ term in the denominator ensures that the prior is normalized in the interval 0 to X_{max}.

There is no apparent change in the behavior of the traces after iteration 50000.

In parallel tempering MCMC, exchanges can occur between independent chains each having a different tempering parameter β. Roughly every 200 iterations the $\beta = 1$ simulation accepts a swap proposal from its neighboring simulation. The final $\beta = 1$ simulation is thus an average of a very large number of independent $\beta = 1$ simulations. The Gelman-Rubin (1992) statistic is typically used to test for convergence of the parameter distributions. We divided the $\beta = 1$ iterations into 10 equal time intervals and intercompared the 10 independent average distributions for each parameter using a Gelman-Rubin test. Five of these time intervals span the first 50000 post burn-in iterations and the other five span the last 50000 iterations. For all of the model parameters the computed Gelman-Rubin statistic was $\leqslant 1.018$.

Fig. 13 shows a comparison of the parameter marginal distributions before (black) and after (gray) iteration number 50000 for the 0.1 μas noise level simulation. The true parameter value is indicated by the dashed vertical line. Very minor differences are discernable for a few parameters but they were judged not to be significant. In astrometry work the ω_i and Ω_i parameters are often strongly correlated

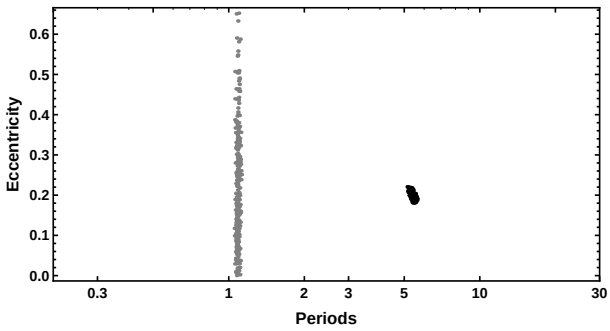

Figure 10. A plot of eccentricity versus period for the 2 planet fit for a simulation noise level of 0.5 μas. The gray and black points correspond to the two different period parameters.

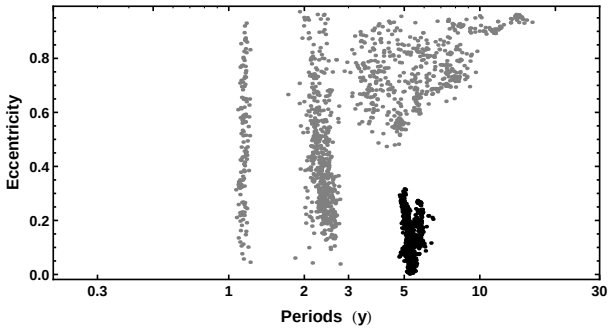

Figure 11. A plot of eccentricity versus period for the 2 planet fit for a simulation noise level of 0.8 μas.

and typically only their sum is well determined. We have included two additional plots for $\omega_1 + \Omega_1$ and $\omega_2 + \Omega_2$.

Table 3 compares the true parameter values to those detected in the HMCMC analysis. The measured parameter value is the median of the marginal probability distribution for the parameter in question and the error bars identify the boundaries of the 68.3% credible region. The value immediately below in parentheses is the maximum *a posteriori* (MAP) value. In the case of the parallax, proper motion and offset parameters, the measured values are the sum of two components: (a) the values obtained in the initial first order fit and (b) the residuals obtained from the HMCMC fit. The only parameters that are significantly different from the true values are ω_2 and Ω_2. As mentioned above these two parameters are often strongly correlated and typically only their sum is well determined. It is clear from the table that their sum is in close agreement with the sum of ω_2 and Ω_2 used in the simulation.

Fig. 14 shows parameter marginal distributions for the 2 planet fit to the simulation with a 0.5 μas noise level. In this case both ω_2 and Ω_2 exhibit two peaks with the stronger peak in each case corresponding to the true value. Again, the distribution of the sum $\omega_2 + \Omega_2$ has a single peaked distribution with the expected value of the peak of 8.2.

Fig. 15 shows a comparison of the semi-major axis and mass probability distributions for the lower mass planet derived from the HMCMC results for the 0.1 (gray) and 0.5 μas (black) simulations. As expected the 0.1 μas case has

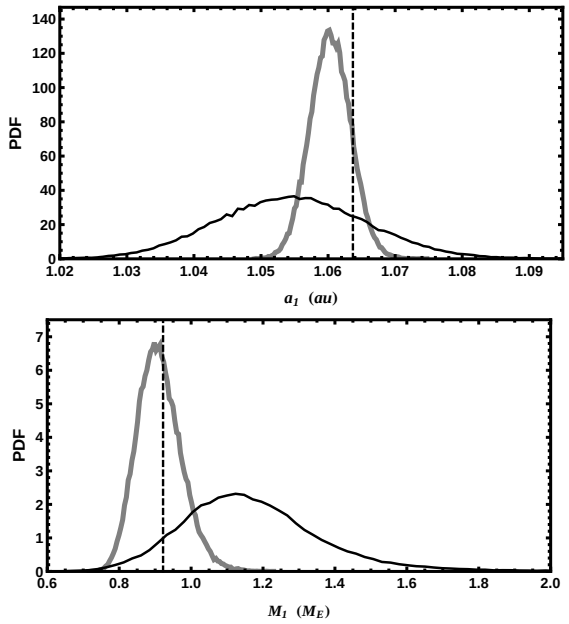

Figure 15. A comparison of the semi-major axis and mass probability density distributions for the lower mass planet derived from the 0.1 (gray) and 0.5 μas (black) simulations. The true parameter value is indicated by the dashed vertical line.

a much narrower probability distribution centered closer to the true value.

6 DISCUSSION

The main purpose of this paper was to describe and report test results for a new method that automatically achieves efficient MCMC sampling in highly correlated parameter spaces and which does not require additional MCMC chains. It was designed to work with an existing hybrid Markov chain Monte Carlo (HMCMC) algorithm which incorporates parallel tempering, simulated annealing and genetic crossover operations. In the original hybrid MCMC algorithm, new parameter values were jointly proposed based on independent Gaussian proposal distributions ('I' scheme), one for each parameter. The HMCMC incorporates a control system that automates the selection of an efficient set of Gaussian proposal σ's. Initially, only this 'I' proposal system is used and it is clear that if there are strong correlations between any parameters the σ's 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 new addition 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 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

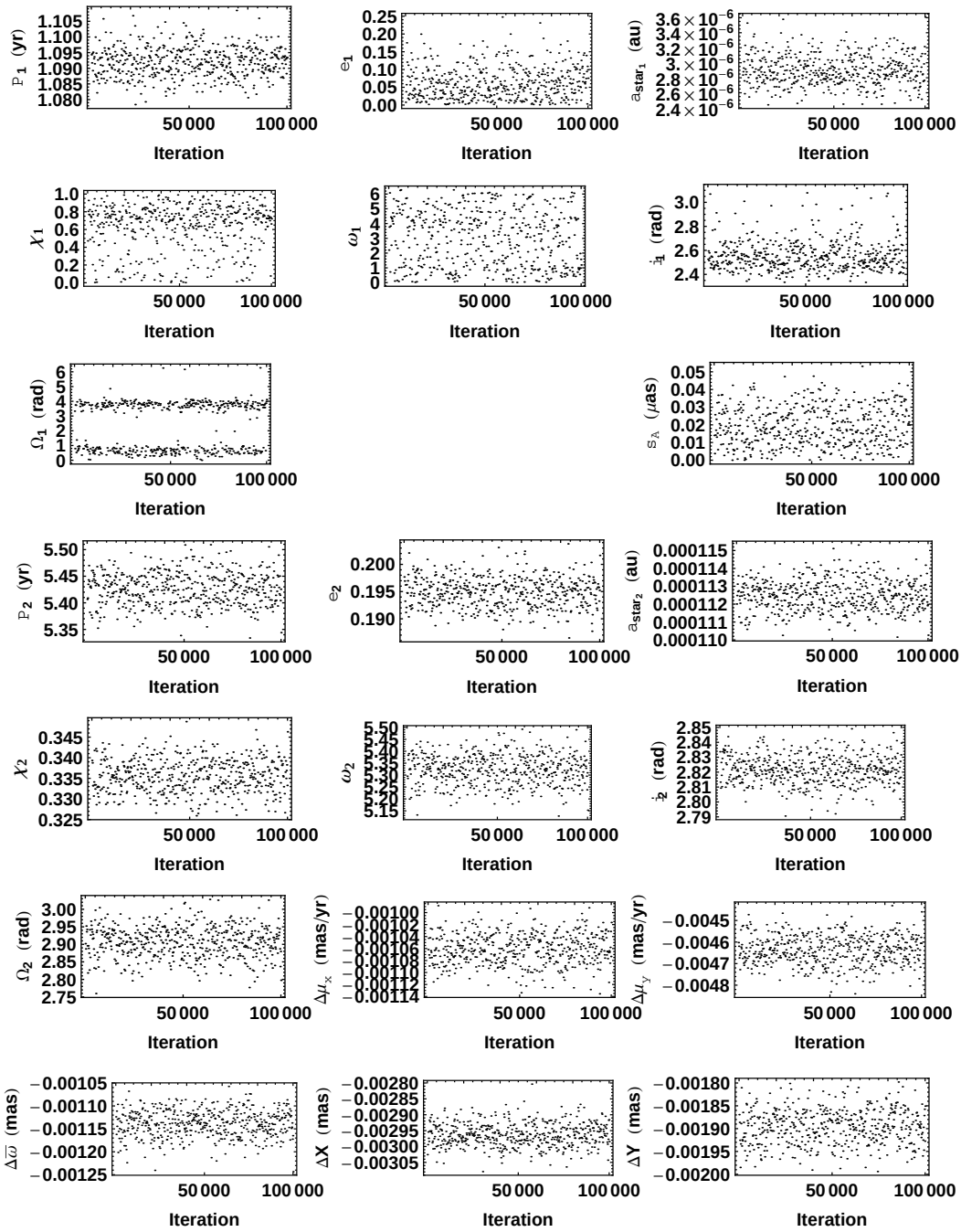

Figure 12. A plot of post burn-in iterations for the 2 planet fit for a simulation noise level of 0.1 μas. Every hundredth iteration is displayed. After iteration 50000 no new samples were added to the correlated sample buffer.

and direction for efficient jumps in a correlated parameter space.

In Section 3 we introduced the concept of an increment 'C' proposal correlation plot and showed an example of one in the lower panel of Fig 5. Other examples of increment 'C' proposal correlation plots are shown in Fig. 16 with $\Delta(\Delta Y)$ as a common axis. ΔY is the residual offset parameter in the y coordinate. They provide an indication of the extent of the correlation in parameter jumps in the astrometry problem.

The final proposal distribution is a random selection of 'I' and 'C' proposals such that each is employed 50% of the time. The existing features of the HMCMC together with the new 'C' correlated parameter proposal system, greatly facilitate the detection of a global minimum in χ^2.

The first test of the algorithm employed a simple precision radial velocity exoplanet data set known to contain a single low eccentricity planet. Two of the parameters were highly correlated and the combined 'I' and 'C' proposals were found to achieve the same excellent results as an analysis which eliminated the parameter correlations by transforming to orthogonal parameters. This was tested by comparing the marginals and autocorrelation functions of the

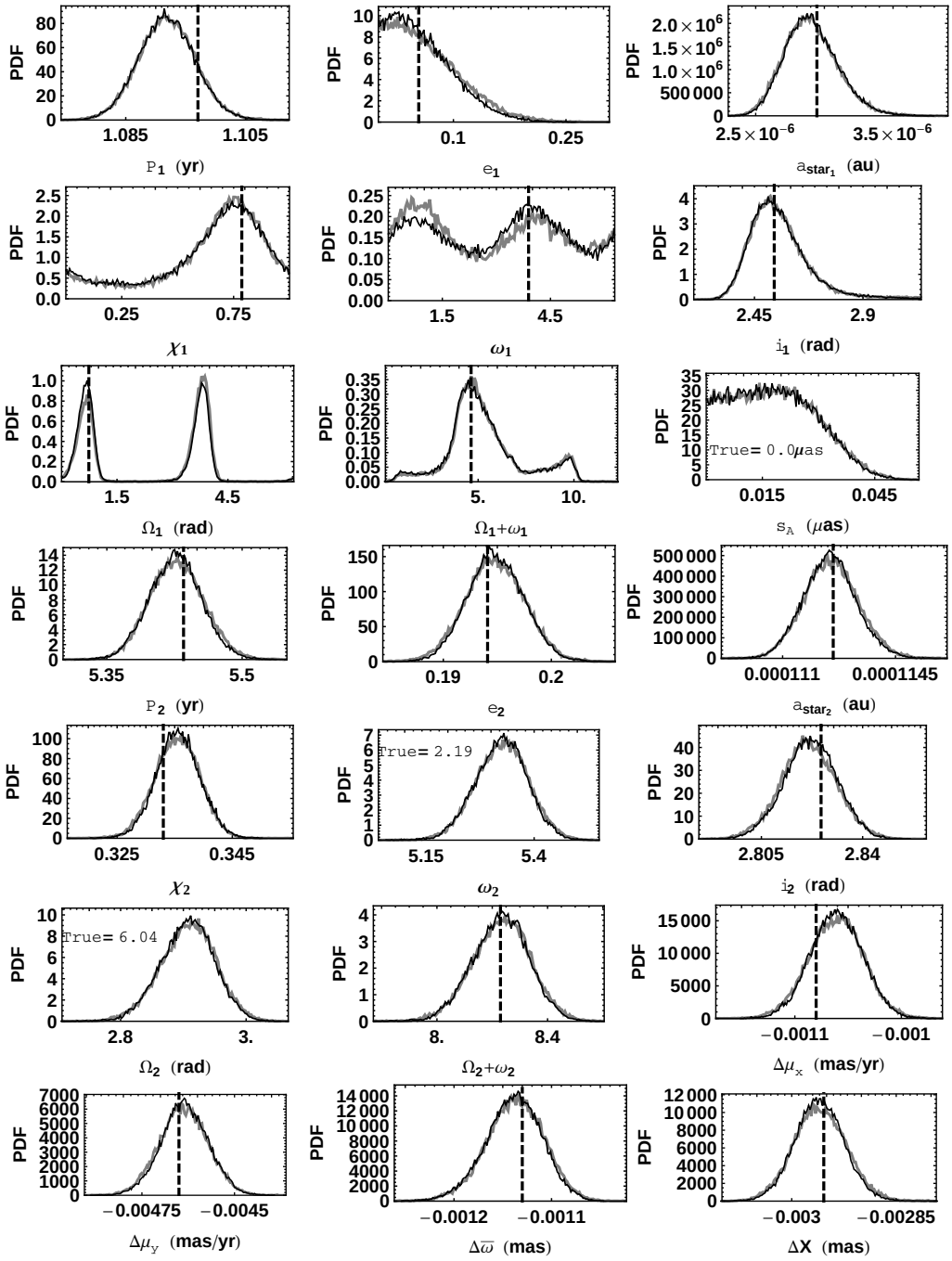

Figure 13. A comparison of the marginal parameter distributions before (black) and after (gray) iteration number 50000 for the 2 planet fit for a simulation noise level of 0.1 μas. After iteration 50000 no new samples were added to the correlated sample buffer. The true parameter value is indicated by the dashed vertical line.

two parameters in question. The execution time for 'I' only proposals was 41 minutes for 600000 iterations on an 8 core PC running *gridMathematica 7*. The execution time increased to 1.3 hr with the inclusion of the 'C' proposal scheme.

The second test employed simulated astrometry data for a two planet system together with associated stellar parallax and proper motion. Prior to incorporating the 'C' proposals we managed to achieve MCMC convergence for the astrometry data after a long burn-in period, but the parameter traces exhibited very large autocorrelation times. Instead of seeking a re-parameterization of the problem, to lessen parameter correlations, we sought to develop a method to automate efficient jumps in correlated parameter spaces like this one. The new 'C' proposal scheme very efficiently accomplished this goal. The execution time for 'I' only proposals was 5.6 hours for 600000 iterations. The execution time increased to 6.8 hr with the inclusion of the 'C' proposal scheme. Relative to the years required to acquire real data sets of this kind, these execution times are very acceptable.

Table 3. Comparison of actual and measured model parameter values.

Parameter	True value	Measured value	Parameter	True value	Measured value
M_1 (M_E)	0.922	$0.908^{+0.057}_{-0.062}$ (0.915)	M_2 (M_E)	12.124	$12.127^{+0.047}_{-0.047}$ (12.131)
P_1 (yr)	1.09705	$1.092^{+0.005}_{-0.005}$ (1.093)	P_2 (yr)	5.43505	$5.427^{+0.030}_{-0.028}$ (5.415)
a_1 (au)	1.0637	$1.060^{+0.003}_{-0.003}$ (1.061)	a_2 (au)	3.0913	$3.088^{+0.011}_{-0.010}$ (3.084)
e_1	0.053791	$0.054^{+0.027}_{-0.053}$ (0.060)	e_2	0.194098	$0.195^{+0.003}_{-0.003}$ (0.196)
i_1 (rad)	2.53204	$2.53^{+0.09}_{-0.11}$ (2.528)	i_2 (rad)	2.82563	$2.822^{+0.009}_{-0.009}$ (2.821)
Ω_1 (rad)	0.738587	2 peaks (0.592)	Ω_2 (rad)	6.04406	$2.908^{+0.044}_{-0.041}$ (2.896)
ω_1 (rad)	3.89274	2 peaks (3.861)	ω_2 (rad)	2.18914	$5.326^{+0.063}_{-0.057}$ (5.304)
$\Omega_1 + \omega_1$ (rad)	4.62	$4.6^{+0.6}_{-0.6}$	$\Omega_2 + \omega_2$ (rad)	8.23	$8.24^{+0.07}_{-0.10}$
T_{0_1}	2018.37	$2.018.48^{+0.17}_{-0.25}$ (2018.40)	T_{0_2}	2018.52	$2018.51^{+0.01}_{-0.01}$ (2018.51)
X_0 (mas)	0.0	$0.00000^{+0.00004}_{-0.00003}$ (-0.00001)	Y_0 (mas)	0.0	$0.00002^{+0.00004}_{-0.00004}$ (0.00001)
μ_x (mas/yr)	1169.86174	$1169.86176^{+0.00003}_{-0.00002}$ (1169.86175)	μ_y (mas/yr)	-482.90480	$-482.90479^{+0.00006}_{-.00006}$ (-482.90477)
ϖ (mas)	94.90000	$94.90000^{+0.00003}_{-0.00003}$ (94.89998)			

Also keep in mind that the HMCMC is executing a blind search in a complex 20 parameter space.

One condition that is required for an MCMC to achieve an equilibrium distribution is that the transition kernel (equation 3) be time independent. This can be ensured by only collecting MCMC samples after we have stopped adding to the 'C' proposal system buffer. The need to stop adding to the 'C' proposal buffer may however be overly restrictive. Once burn-in has been achieved, there should be no statistical difference between 'C' proposals derived from correlated sample buffer values collected during different post burn-in time intervals. This assumption was born out by our tests with both the radial velocity and astrometry data sets. We found no significant change in the behavior of the HMCMC parameter traces or the marginals before and after terminating the additions to the 'C' proposal buffer. One advantage of continuing to add to the 'C' buffer is to enhance the possibility of finding a significantly improved model solution, should it exist.

One additional complexity associated with the correlated sample buffer arises when more than one planet is being fit at a time. The individual orbital frequencies, f_i, are allowed to roam over the entire frequency range and the parameters re-labeled afterwards. This was found to be significantly more successful at converging on the highest posterior probability peak, in fewer iterations, during repeated blind frequency searches, than constraining $f_1 \leqslant f_2 \leqslant \cdots \leqslant f_n$. Thus during the execution of the MCMC the various frequency parameters can exchange meaning. For example, in a 2 planet model f_1 might start off designating the planet with the higher orbital frequency but change to mean the lower frequency orbit and vice versa. This is especially true during the burn-in period when the orbital frequencies make large excursions. Since the parameter correlations will in general be different for the two different planets, it is necessary to keep track of the frequency order of each sample. This is done with a tag that is appended at the end of each sample vector. If the current sample has a tag '21', indicating that the first period parameter is the higher frequency, then to propose a new 'C' move, only buffer samples with the same '21' tag are randomly selected in pairs to construct a difference vector for the current 'C' proposal.

6.1 Alternate 'C' proposal scheme

Another way to encapsulate the correlations between parameters is to compute the covariance matrix of the accepted 'I' samples stored in the 'C' proposal buffer and then propose future steps along the eigenvectors of the covariance matrix. This did not prove successful and the explanation was traced to the wrap around angular parameters (see footnote 3 earlier). Instead the covariance matrix of 100 differences (following footnote 3) of randomly selected pairs of buffer samples was computed. The alternate 'C' proposal move was then drawn from a multivariate Normal distribution that employed this covariance matrix. When this alternate 'C' proposal approach was tested on the simulated astrometry data the results were essentially identical to those obtained using our simpler differencing 'C' proposal method. Fig. 17 shows a comparison of the marginal distributions obtained with the two different 'C' proposal methods for the 0.1 μas noise level simulation. The execution time for the alternate 'C' proposal method was approximately 20% longer.

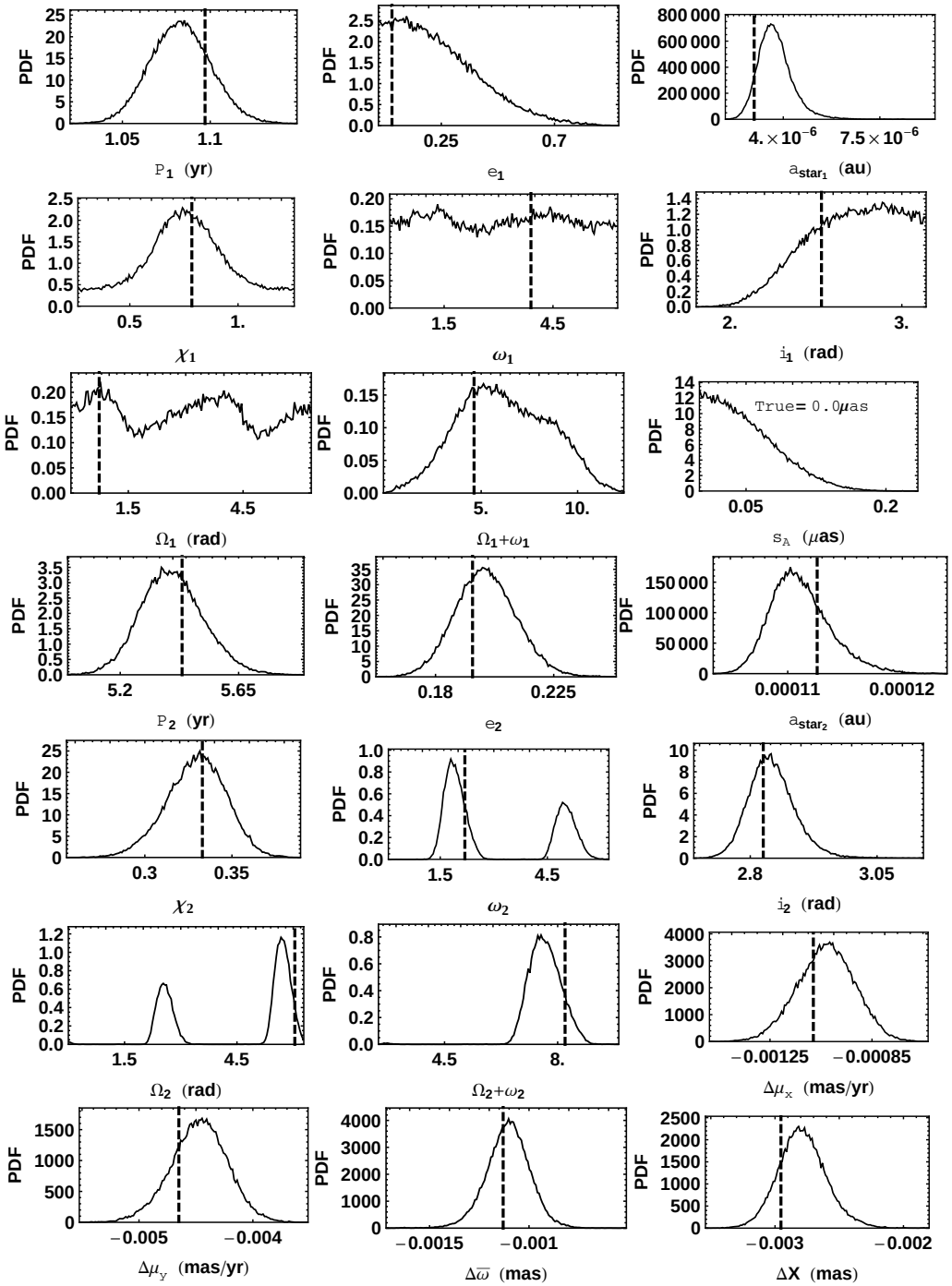

Figure 14. The marginal parameter distributions for the 2 planet fit for a simulation noise level of 0.5 μas. The true parameter value is indicated by the dashed vertical line.

7 CONCLUSIONS

This paper has presented a new method to automatically achieve efficient MCMC sampling in highly correlated parameter spaces. Unlike Differential Evolution Markov Chain (DE-MC) (TerBraak (2006)) it does not require additional MCMC chains. Like DE-MC, it provides a means of choosing an appropriate scale and orientation for the jumping distribution. It was designed to work with an existing hybrid Markov chain Monte Carlo (HMCMC) algorithm (Gregory 2009 and Gregory & Fischer 2010) which incorporates parallel tempering, simulated annealing and genetic crossover operations. The computational penalty associated with new 'C' correlated parameter proposal system amounted to 23% for a 20 parameter astrometry model fitting exercise. Two tests of the algorithm were described employing (a) exoplanet precision radial velocity data, and (b) simulated space astrometry data. The latter test explores the accuracy of parameter estimates obtained with the Bayesian HMCMC algorithm on the assumed astrometric noise.

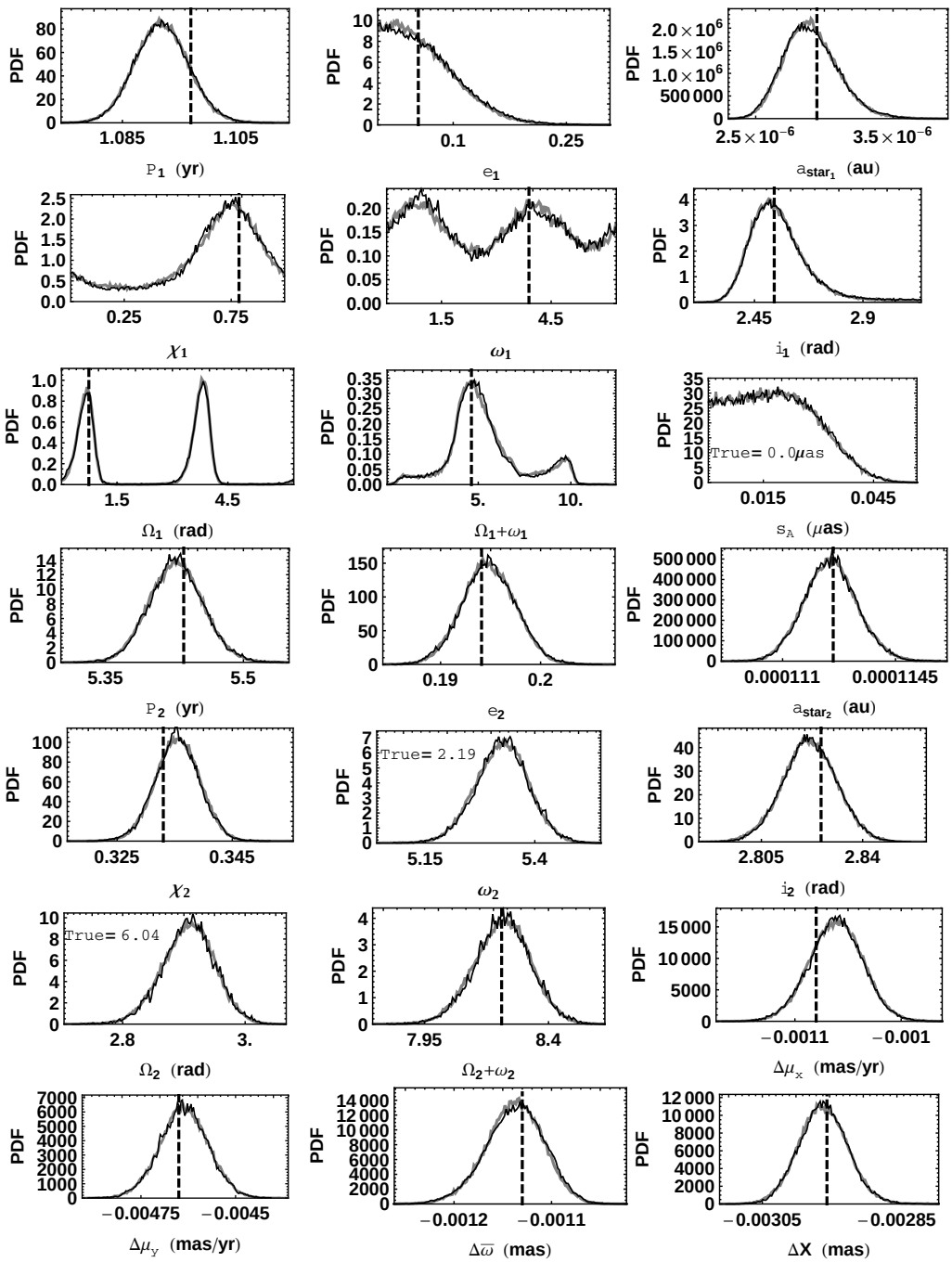

Figure 17. A comparison of astrometry parameter marginal distributions obtained with the two different 'C' proposal methods for the 0.1 μas noise level simulation. The black curve was obtained with the alternate 'C' proposal method and the gray curve with the simpler differencing 'C' proposal method. The true parameter value is indicated by the dashed vertical line.

ACKNOWLEDGMENTS

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

8 BIBLIOGRAPHY

REFERENCES

Ter Braak, C. J. F., 2006, Statistical Computing, 16, 239

Bretthorst, G. L., 1988, Bayesian Spectrum Analysis and Parameter Estimation, New York: Springer-Verlag

Campbell, B., Walker, G. A. H., & Yang, S., 1988, ApJ, 331, 902

Cumming, A., 2004, MNRAS, 354, 1165

Cumming, A., Dragomir, D., 2010, MNRAS, 401, 1029

Fischer,D. A., Marcy, G. W., Butler, R. P., Laughlin, G. L., and Vogt, S. S., 2002, ApJ, 564, 1028

Fischer,D. A., Laughlin, G. L., Butler, R. P., Marcy, G. W., Johnson, J., Henry, G.,Valenti,J., Vogt, S. S., Am-

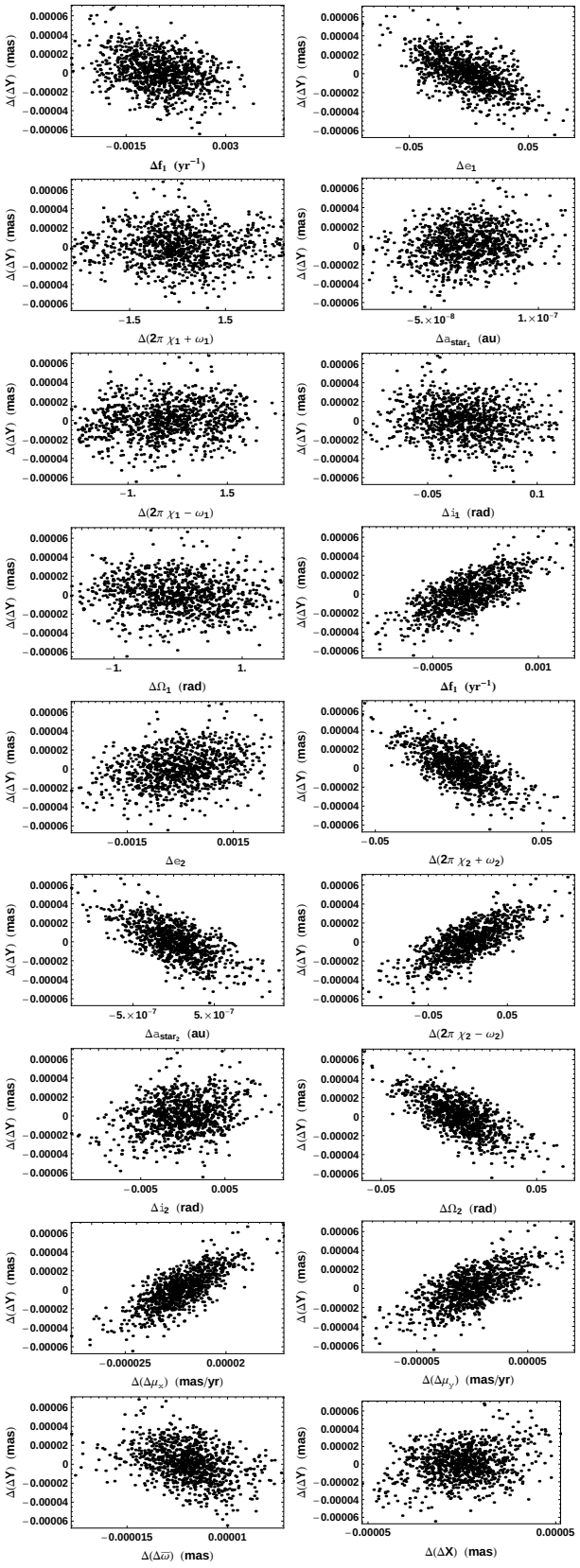

Figure 16. Other examples of increment 'C' proposal correlation plots, which indicate the extent of the correlation in parameter jumps in the astrometry problem. $\Delta(\Delta Y)$ as a common axis in these plots where ΔY is the residual offset parameter in the y coordinate

mons, 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., 2005, ApJ, 620, 486

Ford, E. B., 2005, AJ, 129, 1706

Ford, E. B., 2006, ApJ, 620, 481

Ford, E. B., & Gregory, P. C., 2006, in 'Statistical Challenges in Modern Astronomy IV,' G. J. Babu and E. D. Feigelson (eds.), ASP Conf. Ser., 371, 189

Gelman, A., & Rubin, D. B., 1992, Statistical Science 7, 457

Geyer, C. J., 1991, in 'Computing Science and Statistics: 23rd symposium on the interface, American Statistical Association, New York, p. 156

Gregory, P. C., 2005a, 'Bayesian Logical Data Analysis for the Physical Sciences: A Comparative approach with *Mathematica* Support', Cambridge University Press

Gregory, P. C., 2005b, ApJ, 631, 1198

Gregory, P. C., 2007a, MNRAS, 374, 1321

Gregory, P. C., 2007b, in 'Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 27th International Workshop', Saratoga Springs, eds. K. H. Knuth, A. Caticha, J. L. Center, A,Ġiffin, C. C. Rodrguez, AIP Conference Proceedings, 954, 307

Gregory, P. C., 2007c, MNRAS, 381, 1607

Gregory, P. C., 2008, JSM Proceedings, Denver, American Statistical Association, arXiv:0902.2014v1 [astro-ph.EP]

Gregory, P. C., and Fischer, D. A., 2010, MNRAS, 403, 731

Hukushima, K., and Nemoto, K., 1996, Journal of the Physical Society of Japan, 65(4), 1604

Jaynes, E. T., 1957, Stanford University Microwave Laboratory Report 421, Reprinted in 'Maximum Entropy and Bayesian Methods in Science and Engineering', G. J. Erickson and C. R. Smith, eds, (1988) Dordrecht: Kluwer Academic Press, p.1

Jaynes, E.T. (1987), 'Bayesian Spectrum & Chirp Analysis,' in *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems*, ed. C.R. Smith and G.L. Erickson, D. Reidel, Dordrecht, p. 1

Loredo, T., 2004, in 'Bayesian Inference And Maximum Entropy Methods in Science and Engineering: 23rd International Workshop', G.J. Erickson & Y. Zhai, eds, AIP Conf. Proc. 707, 330 (astro-ph/0409386)

Loredo, T. L. and Chernoff, D., 2003, in 'Statistical Challenges in Modern Astronomy III', E. D. Feigelson and G. J. Babu (eds) , Springer, New York, p. 57

Marcy G. W., Butler R. P., 1996, ApJ, 464, L147

Mayor M., Queloz D., 1995, Nature, 378, 355

Nelder, J. A. and Mead, R., Comput. J. 7, 308

Roberts, G. O., Gelman, A. and Gilks, W. R., 1997, Annals of Applied Probability, 7, 110

Scargle, J. D. 1982, ApJ, 263, 835

Traub, W. A. and 23 other authors, 2010, European Astonomy Society publication series, Vol. 42, p.191, arXiv:0904.0822v1 [astro-ph]

Wolszczan, A., & Frail, D., 1992, Nature,355, 145

Udry, S., Bonfils, X., Delfosse, X., Forveille, T., Mayor, M., Perrier, C., Bouchy, F., Lovis, C., Pepe, F., Queloz, D., and Bertaux, J.-L., 2007, A&A, 469, L43

This paper has been typeset from a TEX/ LATEX file prepared by the author.