A BAYESIAN ANALYSIS OF EXTRASOLAR PLANET DATA FOR HD 73526

P. C. GREGORY

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

ABSTRACT

A new Bayesian tool for nonlinear model fitting has been developed that employs a parallel tempering Markov chain Monte Carlo algorithm with a novel statistical control system. The algorithm has been used to reanalyze precision radial velocity data for HD 73526. For a single-planet model, three possible orbits were found with periods of $127.88^{+0.37}_{-0.09}$, $190.4^{+1.8}_{-2.1}$, and $376.2^{+1.4}_{-4.3}$ days. The 128 day orbit, with an eccentricity of 0.56 ± 0.03, has a maximum value of prior × likelihood that is 16 times larger than for the next highest solution at 376 days. However, the 376 day orbit, with an eccentricity of $0.10^{+0.05}_{-0.10}$, is formally more probable because for this sparse data set there is a much larger volume of parameter space with a significant probability density in the vicinity of the 376 day peak. The previously reported orbit (Tinney et al. 2003) of 190.5 ± 3.0 days corresponds to our least probable orbit. The analysis highlights the need for measurements around phase 0.5 for the 376 day period.

Subject headings: methods: data analysis — planetary systems — stars: individual (HD 73526)

Online material: color figure

1. INTRODUCTION

Since the discovery of the extrasolar planet orbiting 51 Peg (Mayor & Queloz 1995), high-precision radial velocity measurements of the reflex radial velocity of the host star have lead to the discovery of over 130 planetary candidate companions to solar-type stars in both single- and multiple-star systems (see, e.g., Eggenberger et al. 2004). Most of these planets have $M \sin i$ comparable to or greater than Jupiter, where M is the mass of the planet and i is the angle between the orbital plane and a plane perpendicular to the line of sight to the planetary system. However, as Doppler precision increases, lower mass planets to the level of Neptune are also being discovered (McArthur et al. 2004; Fisher et al. 2003).

Existing orbital solutions are based on the use of nonlinear least-squares methods that typically require a good initial guess of the parameter values. Frequently, the first indication of a periodic signal comes from a periodogram or epoch-folding analysis of the data. These periodogram methods do not make use of relevant prior information about the known shape of the periodic signal. In the extrasolar planet Kepler problem, the mathematical form of the periodic signal is well known and is readily built into a Bayesian analysis. In this work, we reanalyze recently published high-precision radial velocity measurements using a Bayesian Markov chain Monte Carlo (MCMC) algorithm that is effective for both detecting and characterizing the orbits of extrasolar planets. This approach has the following advantages:

1. Good initial guesses of the parameter values are not required. An MCMC is capable of efficiently exploring all regions of joint prior parameter space having significant probability.[1] There is thus no need to carry out a separate search for the orbital period(s).

2. The method is also capable of fitting a portion of an orbital period, so search periods longer than the duration of the data are possible.

3. The built-in Occam's razor in the Bayesian analysis can save considerable time in deciding whether a detection is be-

lievable. 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. The analysis yields the relative probability of each of the models explored.

4. The analysis yields the full marginal posterior probability density function (PDF) for each model parameter, not just the maximum a posteriori (MAP) values and a Gaussian approximation of their uncertainties.

Tinney et al. (2003) reported the detection of four new planets orbiting HD 2039, HD 30177, HD 73526, and HD 76700. Our results for HD 3029, together with some of the details of the Bayesian MCMC approach, were presented in Gregory (2005, pp. 331–349). In this paper we describe a modified form of the Bayesian MCMC approach and report the results of our reanalysis of the HD 73526 data, which differ significantly from the conclusions of Tinney et al. (2003). For another application of MCMC methods to extrasolar planet detection see Ford (2005).

2. PLANETARY ORBITAL MODELS

There are two questions of interest. Initially, we are interested in whether the reflex motion of the host star provides evidence for 0, 1, 2, 3, . . . planets. Secondly, we are interested in the orbital parameters of each planet detected.

2.1. *Model Space*

The starting point for any Bayesian analysis is Bayes's theorem given by

$$p(M_i|D, I) = \frac{p(M_i|I)p(D|M_i, I)}{p(D|I)}, \qquad (1)$$

where $p(M_i|D, I)$ is the posterior probability that model M_i is true, given the new data D and prior background information I. The logical conjunction D, I is a proposition representing our current state of knowledge that is relevant for computing the probability of the truth of the models in question. A Bayesian analysis is an encoding of our current state of knowledge into a probability distribution concerning the hypothesis space of

---

[1] It may fail to detect pathologically narrow probability peaks in a reasonable number of iterations.

interest. Of course, as our state of knowledge changes (e.g., we acquire new data), so in general will the probabilities. The posterior is proportional to $p(M_i|I)$, the prior probability that M_i is true, times the likelihood function, $p(D|M_i, I)$. The likelihood is the probability that we would have gotten the data D that we did, if M_i is true. The likelihood is often written as $\mathcal{L}(M_i)$.

In a well-posed problem the prior (background) information (I) specifies the hypothesis space of interest (model space) and how to compute the likelihood function. In general, the models have parameters that are discussed below. Table 1 summarizes the hypothesis space of interest. For multiple-planet models, there is no analytic expression for the exact radial velocity perturbation.[2] 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.

In the case of model M_0, the star is moving away or toward the observer with an unknown velocity. Note that the radial velocities published by Tinney et al. (2003) have an arbitrary zero point determined by a template observation, so the V in our analysis is just a constant velocity term. Model M_{0s} assumes that there is an extra noise term in addition to the known measurement uncertainties. The nature of this extra noise term is discussed below. The simplest single-planet model would have a circular orbit with four unknown parameters. Although the circular model is simpler, we focus only on elliptical orbits because the prior probability of an orbit with eccentricity of precisely zero is considered to be extremely small.

We illustrate the calculations for models M_0, M_{0s}, M_1, and M_{1s}. Much of the background material needed to appreciate the current analysis is presented in chapter 12 of Gregory (2005).

We can represent the measured velocities by the equation

$$v_i = f_i + e_i + e_{0i}, \qquad (2)$$

where e_i is the noise component arising from known but unequal measurement errors with an assumed Gaussian distribution. The term e_{0i} accounts for any additional unknown measurement errors[3] plus any real signal in the data that cannot be explained by the model prediction f_i. For example, suppose that the star actually has two planets and the model assumes that only one is present. In regard to the single-planet model, the velocity variations induced by the unknown second planet act like an additional unknown noise term. In addition, researchers have attempted to empirically quantify additional sources of error in the measured velocities, referred to as "jitter," which is due in part to flows and inhomogeneities on the stellar surface (e.g., Wright 2005). Jitter would also contribute to this extra noise term. In general, nature is more complicated than our model and known noise terms. In the absence of detailed knowledge of the sampling distribution for e_{0i}, other than that it has a finite variance, the maximum entropy principle tells us that a Gaussian distribution

---

[2] The deviations from the simple sum of Keplerian models can be divided into two types (Brown 2004): short-period interactions and secular interactions. The magnitude of the short-period interactions is often small compared to the magnitude of the Keplerian perturbation and the observational uncertainties. The secular interactions are typically modeled as changes in the Keplerian orbital parameters. While the secular interactions can have large effects on the observed radial velocities, the timescales are typically much longer than the time span of observations.

[3] In Gregory (2005), we made use of an unknown noise scale parameter b, instead of e_{0i}. The one advantage of using b is that it can allow for the possibility that the measurement errors have been overestimated. A possible disadvantage is that it assumes that relative weight to be assigned to each point is the same as that based only on the known measurement errors alone. For example, if the effective noise is dominated by real signal due to additional planets, which are not accounted for in the model, then we would prefer to weight the points more equally.

TABLE 1
MODEL SPACE

Symbol	Model	Number of Parameters
M_0	Constant velocity V	1
M_{0s}	Constant velocity V + extra noise term	$1 + s$
M_1	V + elliptical orbit	6
M_{1s}	V + elliptical orbit + extra noise term	$6 + s$
M_2	V + 2 elliptical orbits	11
M_{2s}	V + 2 elliptical orbits + extra noise term	$11 + s$
M_3	V + 3 elliptical orbits	16
M_{3s}	V + 3 elliptical orbits + extra noise term	$16 + s$

would be the most conservative choice (i.e., maximally noncommittal about the information we do not have). We assume that the noise variance is finite and adopt a Gaussian distribution for e_{0i} with a variance s^2. Thus, the combination of $e_i + e_{0i}$ has a Gaussian distribution with variance equal to $\sigma_i^2 + s^2$. Marginalizing s has the desirable effect of treating anything in the data that cannot be explained by the model as noise, leading to the most conservative estimates of orbital parameters (for a tutorial demonstration of this point see §§ 9.2.3 and 9.2.4 of Gregory 2005).

The term e_{0i} also plays a significant role in model comparison. To usefully claim the detection of a planet, we want

$$p(M_1|D, I) > \text{both } p(M_0|D, I) \text{ and } p(M_{0s}|D, I). \qquad (3)$$

If $p(M_1|D, I) > p(M_0|D, I)$ but not $> p(M_{0s}|D, I)$, then we would be saying that the addition of an independent Gaussian random component was more successful in accounting for the velocity variations than the reflex motion due to a planet.

Similarly, to claim the detection of a second planet, we want

$$p(M_2|D, I) > \text{both } p(M_1|D, I) \text{ and } p(M_{1s}|D, I). \qquad (4)$$

The predicted model radial velocity, f_i, for one planet is given by the following equation and involves six unknowns:

$$f_i = V + K\{\cos[\theta(t_i + \chi P) + \omega] + e \cos \omega\}, \qquad (5)$$

where V is a constant velocity; K is the velocity amplitude equal to $2\pi P^{-1}(1 - e^2)^{-1/2}a \sin i$; P is the orbital period; a is the semimajor axis of the orbit; e is the orbital eccentricity; i is the inclination of the orbit; ω is the longitude of periastron; χ is the fraction of an orbit, prior to the start of data taking, at which periastron occurred (thus, χP equals the number of days prior to $t_i = 0$ that the star was at periastron, for an orbital period of P days); and $\theta(t_i + \chi P)$ is the angle of the star in its orbit relative to periastron at time t_i, also called the true anomaly.

The dependence of θ on t_i, which follows from the conservation of angular momentum, is given by the solution of

$$\frac{d\theta}{dt} - \frac{2\pi[1 + e \cos \theta(t_i + \chi P)]^2}{P(1 - e^2)^{3/2}} = 0. \qquad (6)$$

To fit equation (5) to the data, we need to specify the six model parameters, P, K, V, e, ω, and χ.

2.2. Model Probabilities

To decide if a planet has been detected, we compare the one-planet model with the constant-velocity model. The equations

that follow pertain to models M_{1s} and M_{0s}. From Bayes's theorem, we can write

$$p(M_{1s}|D,\ I) = \frac{p(M_{1s}|I)p(D|M_{1s},\ I)}{p(D|I)}$$
$$= Cp(M_{1s}|I)p(D|M_{1s},\ I), \qquad (7)$$

where $C = 1/p(D|I)$ is the normalization constant. The quantity $p(D|M_{1s},\ I)$ is called a global likelihood because it is not conditional on any particular choice of the model parameters. The global likelihood is obtained from the conditional likelihood, given by equation (8), by a process known as marginalization. It is equal to the weighted average of the conditional likelihood, weighted by the prior probability distribution of the model parameters and s:

$$p(D|M_{1s},\ I) = \int dP \int dK \int dV \int de \int d\chi \int d\omega$$
$$\int ds\, p(P,\ K,\ V,\ e,\ \chi,\ \omega,\ s|M_{1s},\ I)$$
$$\times p(D|M_{1s},\ P,\ K,\ V,\ e,\ \chi,\ \omega,\ s,\ I), \quad (8)$$

where $p(P,\ K,\ V,\ e,\ \chi,\ \omega,\ s|M_{1s},\ I)$ is the joint prior for the model parameters and $p(D|M_{1s},\ P,\ K,\ V,\ e,\ \chi,\ \omega,\ s,\ I)$ is the conditional likelihood.

For independent Gaussian errors, the likelihood is a product of N Gaussians, one for each data point:

$$p(D|M_{1s},\ P,\ K,\ V,\ e,\ \chi,\ \omega,\ I) = A \exp\left[-\sum_{i=1}^{N} \frac{(v_i - f_i)^2}{2(\sigma_i^2 + s^2)}\right],$$
$$(9)$$

where

$$A = (2\pi)^{-N/2}\left[\prod_{i=1}^{N}(\sigma_i^2 + s^2)^{-1/2}\right]. \qquad (10)$$

The probability of model M_{0s} is given by

$$p(M_{0s}|D,\ I) = Cp(M_{0s}|I)p(D|M_{0s},\ I). \qquad (11)$$

Since M_{0s} has two parameters V plus s,

$$p(D|M_{0s},\ I) = \int dV \int ds\, p(V|M_{0s},\ I)p(s|M_{0s},\ I)$$
$$\times p(D|M_{0s},\ V,\ s,\ I), \qquad (12)$$

assuming independent priors, and

$$p(D|M_{0s},\ V,\ I) = (2\pi)^{-N/2}\left[\prod_{i=1}^{N}(\sigma_i^2 + s^2)^{-1/2}\right]$$
$$\times \exp\left[-\sum_{i=1}^{N}\frac{(v_i - V)^2}{2(\sigma_i^2 + s^2)}\right]. \qquad (13)$$

We are now in a position to compute the ratio of the posterior probabilities for M_{1s} and M_{0s}. The ratio, $O_{1s0s} = p(M_{1s}|D,\ I)/p(M_{0s}|D,\ I)$, is called the "odds ratio" in favor of model M_{1s}

over the no-planet model M_{0s}. Application of Bayes's theorem leads to

$$O_{1s0s} = \frac{p(M_{1s}|I)}{p(M_{0s}|I)}\frac{p(D|M_{1s},\ I)}{p(D|M_{0s},\ I)} \equiv \frac{p(M_{1s}|I)}{p(M_{0s}|I)}B_{1s0s}, \quad (14)$$

where the first factor is the prior odds ratio and the second factor is called the "Bayes factor." The Bayes factor is simply the ratio of the global likelihoods of the models.

Bayesian model comparison relies on the ratio of global likelihoods, not maximum likelihoods. The marginalization procedure automatically introduces a quantified Occam's factor (one for each parameter) that penalizes the model for wasted prior parameter space that gets ruled out by the data (see Gregory & Loredo 1992; Gregory 2005). Thus, the overall Occam's factor depends not only on the number of model parameters but also on the prior range for each parameter. In other words, the more uncertain we are about the true value of a parameter, the greater is the Occam's penalty. Because the global likelihood is a weighted average of the conditional likelihood, weighted by the prior probability distribution of the model parameters and s, it is important to completely specify these priors, which we do in § 3.1.

If we determine that $O_{1s0s} > 1$, then we are interested in computing the marginal PDF for each of the parameters of M_{1s}. For example, consider the marginal PDF for the orbital period. From Bayes's theorem we can write

$$p(P|D,\ M_{1s},\ I) \propto p(P|M_{1s},\ I)\int dK \int dV \int de \int d\chi \int d\omega$$
$$\int ds\, p(K,\ V,\ e,\ \chi,\ \omega,\ s|M_{1s},\ I)$$
$$\times p(D|M_{1s},\ P,\ K,\ V,\ e,\ \chi,\ \omega,\ s,\ I).$$
$$(15)$$

3. PARAMETER ESTIMATION

3.1. Data and Model Parameter Priors

The radial velocity data of Tinney et al. (2003) for HD 73526 are shown in Figure 1. The sampling is very nonuniform with a minimum sample interval of ≈ 1 day and an average sample interval of 73 days. Our Bayesian analysis does not require uniform sampling and does not introduce a spurious bulge in power at low frequencies (wavelengths comparable to gaps), which is a common experience with schemes that attempt to interpolate missing data. However, the sparseness of the data can result in a multitude of orbital periods that cannot be ruled out with the current data. This point is discussed further in § 6. One advantage of nonuniform sampling is to increase the effective bandwidth (Bretthorst 1988, 2001a), namely, the largest spectral window free of aliases, compared to the same number of uniform samples in the same time window. The aliasing phenomenon returns at a sufficiently high frequency due to one of the following reasons: (1) the sampling times t_i, although nonuniform, are all integer multiples of some small interval δT because the times are all derived from some finite clock interval; or (2) the data are only recorded to a finite number of decimal places, which has the same effect as reason 1. The Nyquist period ($=2\delta T$) corresponding to reason 2 is ~ 17 s. The shortest search period used was 0.5 days.[4] The method is also capable of fitting portions of an

---

[4] The selection of the lower bound on P and upper bound on K was motivated by the fact that for a circular orbit a Jupiter-sized planet would be tidally disrupted for an orbital period of ~ 0.2 days.
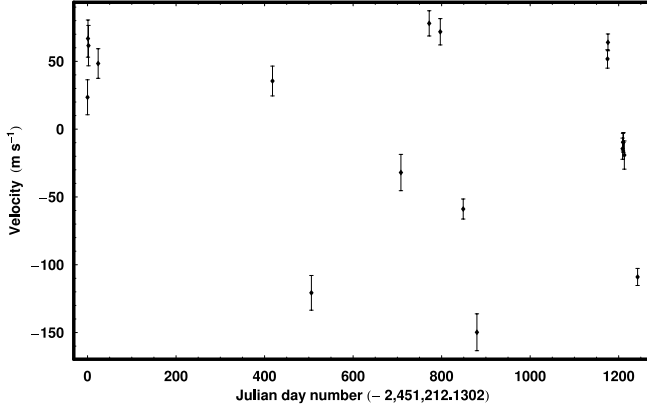
Fig. 1.—Radial velocity data for HD 73526 (Tinney et al. 2003).

TABLE 2
PARAMETER PRIORS USED IN MODELS M_1 AND M_{1s}

Parameter	Prior	Lower Bound	Upper Bound
P (days)	Jeffreys	0.5	3732
K (m s^{-1})	Modified Jeffreys	0 ($K_a = 1$)	400
V (m s^{-1})	Uniform improper prior	All reals	...
e	Uniform	0	1
χ	Uniform	0	1
ω	Uniform	0	2π
s (m s^{-1})	Modified Jeffreys	0 ($s_a = 1$)	100

orbital period so we set the upper limit equal to 3732 days or ~ 3 times the duration of the data (1243 days).

Before we can proceed further, we need to specify the priors for the parameters of M_{1s}. In this analysis, we assume that the prior for each parameter is independent of the prior for any other parameter. Table 2 summarizes our choice of prior and prior ranges for the six model parameters and s. Two of the six model parameters, P and K, as well as s, which characterizes a possible extra noise term, are scale parameters so we used a Jeffreys prior for these. The Jeffreys prior corresponds to assigning equal probability per decade (scale invariance), which is particularly useful if the prior range in a parameter spans several decades. In contrast, a uniform prior would strongly favor the upper decade.[5] Since the prior lower limits for K and s include zero, we used a modified Jeffrey's prior of the form

$$p(K|M_{1s}, I) = \frac{1}{K + K_a} \frac{1}{\ln[(K_a + K_{\max})/K_a]}. \quad (16)$$

For $K \ll K_a$, $p(K|M_{1s}, I)$ behaves like a uniform prior, and for $K \gg K_a$, it behaves like a Jeffreys prior. The $\ln[(K_a + K_{\max})/K_a]$ term in the denominator ensures that the prior is normalized in the interval from 0 to $K_{\max}$. For this analysis, we set K_a and s_a equal to 1 m s^{-1}, which is 1/10 of the typical radial velocity measurement error.

For the V parameter, we use a uniform prior that spans all real values and set $p(V|M_{1s}, I) = 1$. While this is an "improper" (nonnormalized) prior, the normalization cancels when comparing different models, since it appears exactly once in each model.

The joint prior for the model parameters, assuming independence, is given by

$$p(P, K, V, e, \chi, \omega, s|M_{1s}, I)$$
$$= \frac{1}{P \ln(P_{\max}/P_{\min})} \frac{1}{(K + K_a) \ln[(K_a + K_{\max})/K_0]}$$
$$\times \frac{1}{(e_{\max} - e_{\min})} \frac{1}{2\pi} \frac{1}{(s + s_a) \ln[(s_a + s_{\max})/s_a]}. \quad (17)$$

4. MCMC IMPLEMENTATION

To compute $p(D|M_{1s}, I)$, we need to carry out the seven-dimensional integral indicated by equation (8). Similarly, the

computation of any of the marginal posterior distributions involves integration in six dimensions. To handle a two-planet model would require integration over an additional five parameters. In this analysis we utilize an MCMC approach to efficiently evaluate the multidimensional integrals required in our Bayesian analysis of models with many parameters.

In straight Monte Carlo integration, independent samples are randomly drawn from the volume of the parameter space. The price that is paid for independent samples is that too much time is wasted sampling regions where posterior probability density is very small. Suppose in a one-parameter problem the fraction of the time spent sampling regions of high probability is 10^{-1}. Then in an M-parameter problem, this fraction could easily fall to 10^{-M}.

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 target probability distribution for model M_{1s} is given by

$$p(P, K, V, e, \chi, \omega, s|D, M_{1s}, I)$$
$$= C'p(P, K, V, e, \chi, \omega, s|M_{1s}, I)$$
$$\times p(D|M_{1s}, P, K, V, e, \chi, \omega, s, I), \quad (18)$$

where C' is the normalization constant. For the purposes of this discussion, we represent the set of model parameters by X so we can rewrite equation (18) as

$$p(X|D, M_{1s}, I) = C'p(X|M_{1s}, I)p(D|M_{1s}, X). \quad (19)$$

The random walk is accomplished using a Markov chain, whereby the new sample, X_{t+1}, depends on previous sample X_t according to a time-independent entity called the "transition kernel," $p(X_{t+1}|X_t)$. The remarkable property of $p(X_{t+1}|X_t)$ is that after an initial burn-in period (which is discarded) it generates samples of X with a probability density equal to the desired posterior $p(X|D, M_{1s}, I)$ (for details see chap. 12 of Gregory 2005).

This is achieved by proposing a new sample Y from a *proposal distribution*, $q(Y|X_t)$, which is easy to evaluate and is centered on the current sample X_t. The proposal distribution can have almost any form. A common choice for $q(Y|X_t)$ 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 Y is accepted with a probability r given by

$$r = \max \left[1, \frac{p(Y|D, I)}{p(X_t|D, I)} \frac{q(X_t|Y)}{q(Y|X_t)} \right], \quad (20)$$

---

[5] Provided that the upper and lower prior boundaries are well outside the region of parameter space that the likelihood function argues for, the exact choice of prior is not important in most parameter estimation problems. However, for model selection purposes the choice of prior can have a significant effect on the conclusions.
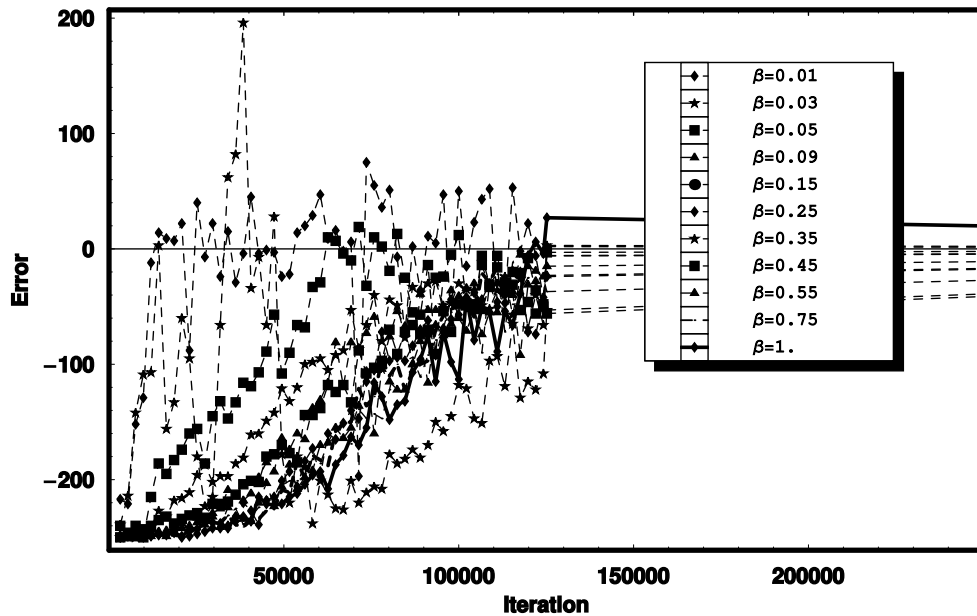
Fig. 2.—A statistical control system automatically steers the Gaussian proposal σ-values to achieve a desirable acceptance rate of $\sim$25%. The evolution of the control system error signal for all 11 parallel chains is illustrated in this figure. [*See the electronic edition of the Journal for a color version of this figure.*]

where $q(X_t|Y) = q(Y|X_t)$ for a symmetrical proposal distribution. If the proposal is not accepted, the current sample value is repeated.

4.1. *Parallel Tempering*

A simple Metropolis-Hastings MCMC algorithm can run into difficulties if the target probability distribution is multimodal with widely separated peaks. It can fail to fully explore all peaks that contain significant probability, especially if some of the peaks are very narrow. This problem is very similar to the one encountered in finding a global minimum in a nonlinear model fitting problem. One solution to finding a global minimum is to use simulated annealing by introducing a temperature parameter $\mathcal{T}$ that is gradually decreased. In this work we use a related process called "tempering." For tempering we create a discrete set of progressively flatter versions of the target distribution using a temperature parameter. For $\mathcal{T} = 1$, the distribution is the desired target distribution, which is referred to as the cold sampler. For $\mathcal{T} \gg 1$, the distribution is much flatter. The basic idea is that by repeatedly heating up the distribution (making it flatter), the new sampler can escape from local modes and increase its chance of reaching all regions of the target distribution that contain significant probability. Typical inference is based on samples drawn from the cold sampler. The samples from the hotter distributions are in principle useful in evaluating the Bayes factors needed for model selection, as discussed in § 5.3.

Mathematically, we can describe these tempering distributions by adapting equation (19), which we repeat here:

$$p(X|D, M_1, I) = C'p(X|M_1, I)p(D|M_1, X, I),$$

where the normalization constant, C', is unimportant and will be dropped. We can construct other flatter distributions as follows:

$$\pi(X|D, \beta, M_1, I) = p(X|M_1, I)p(D|M_1, X, I)^{\beta}$$
$$= p(X|M_1, I)\exp(\beta\ln[p(D|M_1, X, I)])$$
$$(21)$$

for $0 < \beta < 1$. Rather than use a temperature that varies from 1 to infinity, we use its reciprocal, which we label β and refer to as the tempering parameter. Thus, β varies from 1 to 0.

In parallel tempering, multiple copies of the simulation are run in parallel, each at a different temperature (i.e., a different $\beta = 1/\mathcal{T}$). One of the simulations, corresponding to $\beta = 1/\mathcal{T} = 1$, is the desired target probability distribution. The other simulations correspond to a ladder of higher temperature distributions indexed by i. Let n_β equal the number of parallel MCMC simulations. At intervals, a pair of adjacent simulations on this 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.

This swap allows for an exchange of information across the population of parallel simulations. In the higher temperature simulations, radically different configurations can arise, whereas in higher β (lower temperature) states, a configuration is given the chance to refine itself. By making exchanges, we can capture and improve the higher probability configurations generated by the population by putting them into higher β simulations. Some experimentation is needed to refine suitable choices of β_i values. Adjacent simulations need to have some overlap to achieve a sufficient acceptance probability for an exchange operation. For the current analysis we ran $n_\beta = 11$ parallel simulations (chains) with $\beta = \{0.01, 0.03, 0.05, 0.09, 0.15, 0.25, 0.35, 0.45, 0.55, 0.75, 1.0\}$, which achieved a typical acceptance rate between adjacent levels of $\sim$50%. The mean number of iterations between swap proposals was set equal to 10.

4.2. *Choosing the σ-Values for the Gaussian Proposal Distributions*

In this analysis, we used independent Gaussian proposal distributions, which requires specifying a σ for each parameter. In general, the σ-values will be different because the parameters can be very different entities. Also, if the σ-values are chosen too small, successive samples will be highly correlated and it will require many iterations to obtain an equilibrium set of samples. If the σ-values are too large, then proposed samples will very rarely be accepted. Based on empirical studies, Roberts et al.

(1997) recommend calibrating the acceptance rate to about 25% for a high-dimensional model and to about 50% for models of one or two dimensions.

The process of choosing a set of useful proposal σ-values when dealing with a large number of different parameters can be very time consuming. The current analysis employs an automated version of this process based on a statistical control system (Gregory 2005) in which the error signal is proportional to the difference between the current acceptance rate and a target acceptance rate, typically 25%. The control system steers the proposal σ-values to desirable values during the burn-in stage of a single MCMC run. The initial set of Gaussian proposal σ-values (one for each parameter) are set $\approx$20% of the prior range for the parameter. In the latest implementation of the algorithm, the control system steers the proposal σ-values for each of the 11 parallel chains independently, in order to achieve an acceptance rate of 25% in each chain. Figure 2 shows an example of the evolution of the control system error signal for all 11 chains. The control system is turned off as soon as the error signals for all chains are less than a measure of the statistical fluctuations expected for a zero mean error, which in this example occurred in the first $\sim$125,000 iterations of a 600,000 iteration run.

Although inclusion of the control system may result in a somewhat longer effective burn-in period, there is a huge savings in time because it eliminates many trial runs to manually establish a suitable set of proposal σ-values. In practice, it is important to repeat the operation for a few different choices of initial parameter values, to ensure that the MCMC results converge.

One of the difficulties encountered by any integration scheme is how to deal with extremely narrow probability peaks. This is particularly a problem for the orbital period parameter, P, which spans the greatest number of decades. In general, the sharpness of the peak depends in part on how many periods fit within the duration of the data. It proved useful to use a proposal distribution for P that was a Gaussian in the logarithmic of P. This resulted in a constant fractional period resolution instead of a fixed absolute resolution, increasing the probability of detecting narrow spectral peaks at smaller values of P.

5. RESULTS FOR MODEL M_1

5.1. Results with $s = 0$

We first show results from MCMC runs carried out with the extra noise parameter, s, set equal to zero. Five MCMC runs were carried out with durations ranging for 600,000 to 900,000 iterations,[6] with different starting points in parameter space. Three of the runs exhibited three probability peaks at $P = 128$, 190, and 376 days. For two runs only the 190 and 376 day peaks appeared. As we demonstrate below, the 128 day probability peak was very much narrower than the other two, which is probably why it was not detected on two MCMC runs. However, the fact that the three periods were not found on all five runs clearly indicates that we have not converged to a stationary or equilibrium distribution. For the moment we focus on the three runs that yielded all three period peaks.

Figure 3 shows the parameter values for every 500th iteration for the $\beta = 1.0$ chain of an MCMC run of 600,000 iterations. The top left panel shows the value of the prior times the likelihood versus iteration number, and the top right panel shows a blowup of the y-axis, neglecting the first 5000 iterations. We

---

[6] All calculations were performed using *Mathematica* on a Pentium M processor with a clock speed of 1.7 GHz and 1.5 GB of memory. A 600,000 iteration MCMC run with 11 values of β required 3.4 hr of computation.

---

used a burn-in period of 125,000 iterations for this run. The large jumps in P-values signify multiple peaks of probability.

It is instructive to examine a plot of $\log_{10}[\text{prior} \times (\text{likelihood})^\beta]$ versus P, for the $\beta = 1$ chain. This is shown in Figure 4. For this plot, the prior $\times$ likelihood values have been scaled so that the maximum $\log_{10}(\text{prior} \times \text{likelihood}) = 0$. The individual points correspond to every 100th MCMC iteration with a value of the scaled $\log_{10}(\text{prior} \times \text{likelihood}) \geq -5.0$. The upper envelope of the points is the projection of the joint posterior for all the parameters onto the period–$\log_{10}(\text{prior} \times \text{likelihood})$ plane. Three peaks are clearly present at $\sim$128, 190, and 376 days. To determine if the three peaks are harmonically related, solid vertical lines are shown at periods corresponding to $\frac{1}{2}$, 1, 3/2, 2, and 3 times the period of the highest peak. The resolution of the peaks indicates that the three orbital solutions are not exact harmonics. We also note that all three orbital solutions have very different eccentricities (see Table 4). From a comparison of the 376 day peak with the dashed vertical line at 365.25 days, it appears that we can rule out the possibility that the 376 day period is an artifact of the Earth's orbital motion.

The peak at 128 days has the highest value of projected probability (and smallest fit residuals) and is clearly very narrow. One way to roughly estimate the relative volume occupied by each peak in parameter space is to compute the quantity

$$\Delta_{128} = \int_{\Delta_{128}} dP\, dK\, dV\, de\, d\chi\, d\omega\, p(P,\, K,\, V,\, e,\, \chi,\, \omega,\, |M_1,\, I)$$
$$\times\, p(D|M_1,\, P,\, K,\, V,\, e,\, \chi,\, \omega,\, I)$$
$$\times\, [p_{\max}(D|M_1,\, P,\, K,\, V,\, e,\, \chi,\, \omega,\, I)]^{-1}, \qquad (22)$$

where the terms in the first three lines are the contribution to the global likelihood due to the 128 day peak and the term in the last line is the maximum likelihood for the 128 day peak. The value of the terms in the first three lines can be obtained from Table 5, and the value of the term in the last line comes from Table 3. Then Δ_{128}:Δ_{190}:Δ_{376} are in the ratio 1:59:330. Since the 128 day probability peak was only detected on three out of five runs, it almost qualifies as a pathologically narrow peak.

In their analysis, Tinney et al. (2003) reported an orbital period of 190 days, which corresponds to the weakest of our three peaks. Table 3 demonstrates that even though we did not converge on a stationary distribution, the probability distribution maxima for the three runs that did detect all three peaks agree very closely. The table gives the maximum value of the prior $\times$ likelihood for each peak and for each run. We also give the peak values for the 190 and 376 day peaks for the other two runs. It is apparent that the peak values are quite robust. Although the projected probability distribution is instructive, the relative probability of the three periods in a Bayesian analysis is based on marginalization (integration) rather than projection, which allows for the volume of probability in parameter space. This is treated in § 5.3.

Figures 5, 6, and 7 show the MAP model radial velocity curves versus phase and time for the three periods, together with the residuals. Actually, these three figures are the MAP fits based on the results obtained after marginalizing s, which is discussed in § 5.2, but they are not significantly different from the results obtained with $s = 0$.

The rms residuals from the MAP fits for the 128, 190, and 376 day periods are 10.6, 14.7, and 13.7 m s^{-1}, respectively.

5.2. Results with $s \neq 0$

We repeated the calculations described in § 5.1 only this time assuming a prior range for $s = 0-100$ m s^{-1}. As discussed in
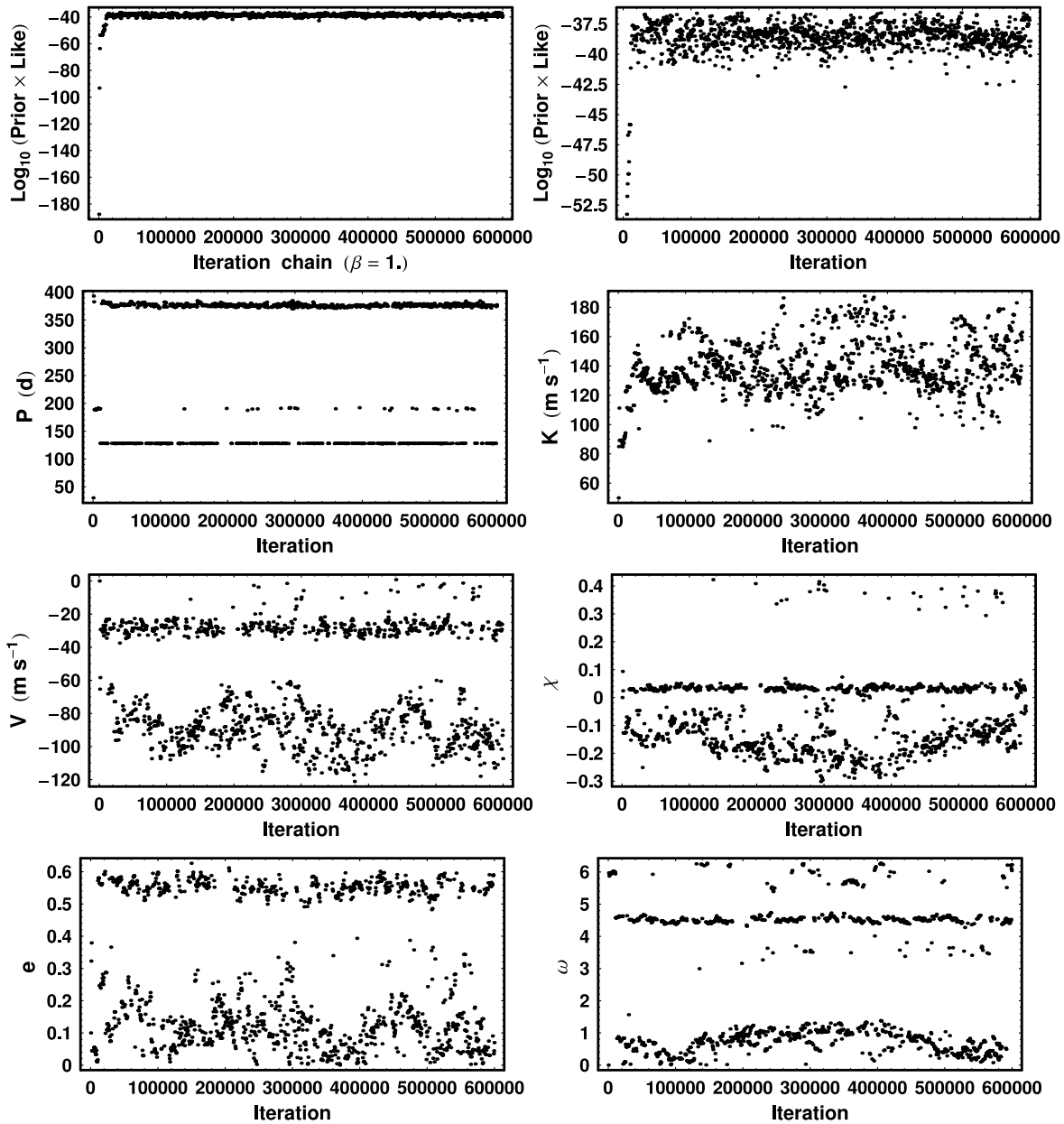
Fig. 3.—MCMC parameter iterations for a 600,000 iteration run. The top left panel shows prior × likelihood vs. iteration number, and the top right panel shows a blowup of the y-axis.

§ 2.1, marginalizing over the extra noise parameter, s, has the desirable effect of treating anything in the data that cannot be explained by the model as noise, leading to the most conservative estimates of orbital parameters. However, with the inclusion of the extra s parameter, we were unable to detect the narrow 128 day peak in repeated MCMC runs when the search range for P spanned the full range of 0.5–3732 days. Thus, for runs that incorporate the whole prior parameter range and include s as an additional free parameter, the 128 day period solution is a real-life example of a pathologically narrow peak. Instead, separate MCMC runs were executed by restricting the period search range to a fractional period interval of 0.47 centered on each of the three period peaks. Of course, the prior parameter ranges used in the calculation of the target probability distribution within each of these three search spaces are those given in Table 2.

The run for each period was repeated several times, using different starting parameters, to check for convergence to the
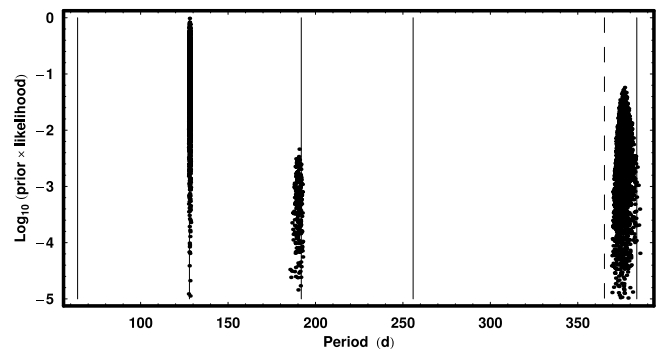

Fig. 4.—Plot of $\log_{10}$(prior × likelihood) vs. orbital period P for the $\beta = 1$ simulation. The upper envelope of the points is the projection of the joint posterior for the parameters onto the period–$\log_{10}$(prior × likelihood) plane. The solid vertical lines are located at $\frac{1}{2}$, 1, 3/2, 2, and 3 times the period of the highest peak. The dashed vertical line corresponds to the Earth's orbital period.

		PRIOR × LIKELIHOOD		
RUN	ITERATIONS	$P = 128$ days	$P = 190$ days	$P = 376$ days
1..................	900000	3.8×10^{-37}	1.9×10^{-39}	2.3×10^{-38}
2..................	600000	3.7×10^{-37}	1.8×10^{-39}	2.3×10^{-38}
3..................	700000	3.9×10^{-37}	1.9×10^{-39}	2.3×10^{-38}
4..................	600000	...	2.0×10^{-39}	2.2×10^{-38}
5..................	600000	...	1.8×10^{-39}	2.3×10^{-38}

same equilibrium sampling distribution. Several tests of convergence were carried. The first was a simple graphical comparison of the parameter iterations from different runs. Figure 8 shows one example for two 128 day period runs. We also used the Gelman-Rubin (GR) convergence statistic (Gelman & Rubin 1992), which tests if the variance across the runs is the same as the variance within the runs. The GR statistic should be close to 1.0 for convergence. The GR statistic was computed for repeated runs on all three period ranges and in all cases the GR statistic was ≤1.02.
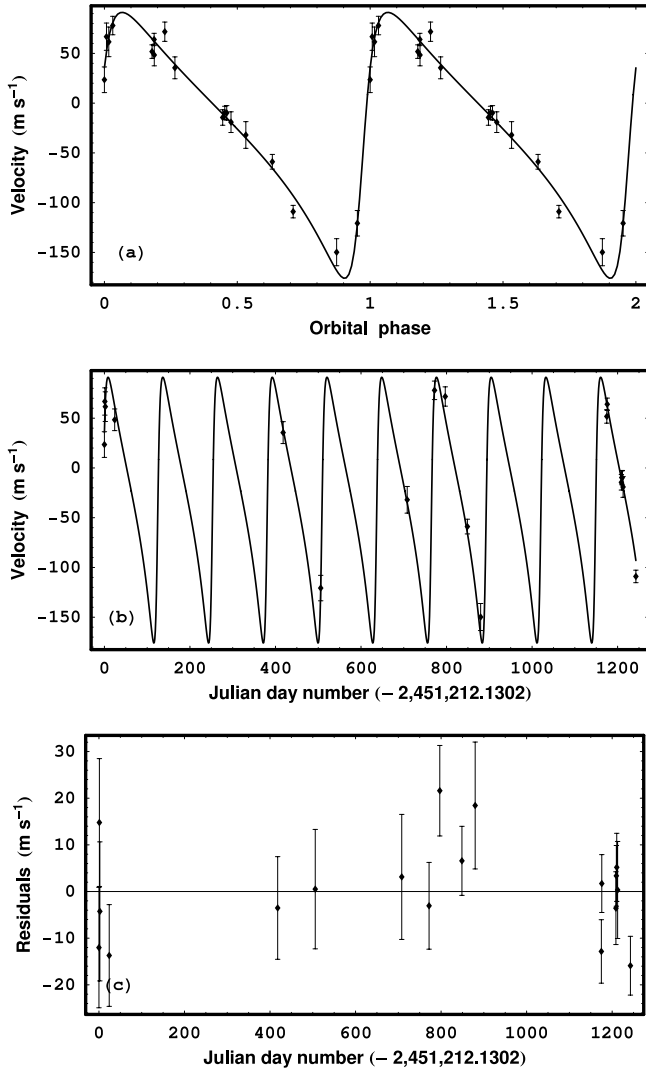

FIG. 5.—(a) Best (MAP) fit 128 day orbit vs. orbital phase with the raw data overlaid. The second cycle is a repeat of the first. (b) Best-fit orbit vs. Julian Day number. (c) Residuals from the MAP fit.
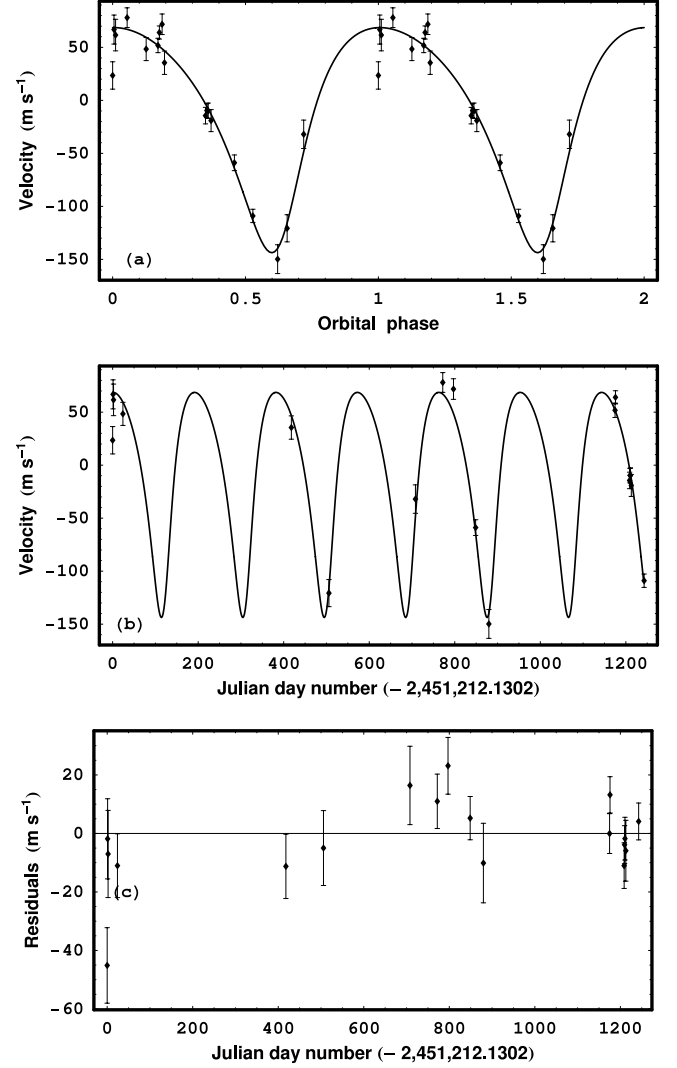

FIG. 6.—(a) Best (MAP) fit 190 day orbit vs. orbital phase with the raw data overlaid. The second cycle is a repeat of the first. (b) Best-fit orbit vs. Julian Day number. (c) Residuals from the MAP fit.

Figures 9, 10, and 11 show the marginal posteriors for the six model parameters and s for the three periods. The maximum values of the prior × likelihood are 6.1×10^{-38}, 3.9×10^{-40}, and 3.9×10^{-39} for the 128, 190, and 376 day orbital solutions, respectively. Thus, the maximum value of the prior × likelihood is 16 times larger for the 128 day orbit than for the next highest solution at 376 days.

It is a simple matter to compute the PDF for any derived quantity that depends on one or more of the model parameters. Examples of these derived quantities are the planet's projected semimajor axis, $a \sin i$, and $M \sin i$, the mass of the planet times the sine of the angle of inclination:

$$a \sin i (\mathrm{km}) = 1.38 \times 10^5 K P \sqrt{1 - e^2}, \qquad (23)$$

where K is in units of m s^{-1} and P is in days, and

$$M \sin i = 4.91 \times 10^{-3} (M_*)^{2/3} K P^{1/3} \sqrt{1 - e^2}, \qquad (24)$$

where M is the mass of the planet measured in Jupiter masses and M_* is the mass of the star in units of solar masses. Equation (24)
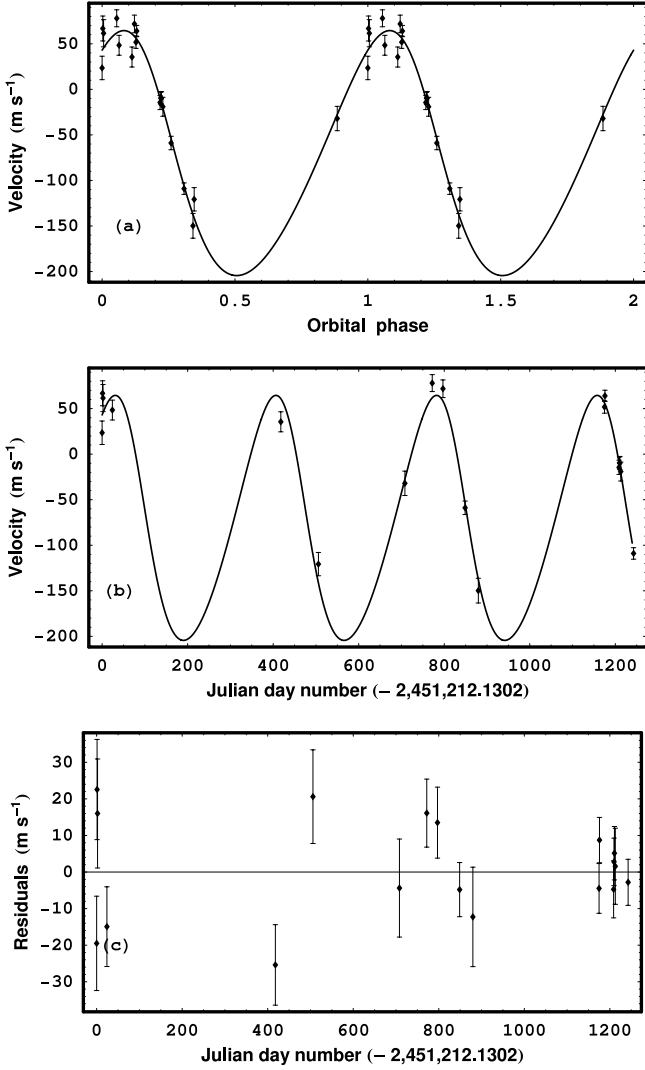
Fig. 7.—(a) Best (MAP) fit 376 day orbit vs. orbital phase with the raw data overlaid. The second cycle is a repeat of the first. (b) Best-fit orbit vs. Julian Day number. (c) Residuals from the MAP fit.

assumes $M \ll M_*$. For example, to obtain the posterior PDF for $a \sin i$, one simply evaluates equation (23) for every set of K, P, e parameter values obtained from the MCMC iterations and then plots a histogram of the $a \sin i$ values. The posterior PDF for $a \sin i$ is just this histogram, normalized to have unit area. Figure 12 shows the posterior PDF for these two derived quantities for the 128 day orbital solution.

The final model parameter values are given in Table 4, along with values of $a \sin i$, $M \sin i$, and the Julian date of periastron passage, which were derived from the parameter values. We summarize the results for each parameter by specifying the MAP value and upper and lower bounds of the 68% credible region. Note that the systematic velocities are relative to a zero point defined by a template observation that was not included in the paper by Tinney et al. (2003). The last line of Table 4 gives our final probability assignments for the three one-planet orbital solutions as derived in § 5.3.

5.3. Model Selection Calculation

In principle, the samples from all (n_β) simulations can be used to calculate the global likelihood of a model, which is needed to evaluate the Bayes factors (eq. [14]) used in model selection.

The global likelihood of the model is given by (see, e.g., Gregory 2005)

$$\ln \left[p(D|M_{1s}, I) \right]$$
$$= \int d\beta \langle \ln[p(D|M_{1s}, P, K, V, \chi, e, \omega, s, I)] \rangle_\beta. \quad (25)$$

In words, for each of the n_β parallel simulations, compute the expectation value (average) of the natural logarithm of the likelihood for post burn-in MCMC samples. Of course, to do this, it is necessary to save the value of $\ln [p(D|M_{1s}, P, K, V, \chi, e, \omega, s, I)]$, along with the parameter values, for each iteration. This procedure requires us to employ a sufficient number of tempering levels that we can estimate the above integral by interpolating values of

$$\langle \ln[p(D|M_{1s}, P, K, V, \chi, e, \omega, s, I)] \rangle_\beta$$
$$= \frac{1}{n} \sum_t \ln [p(D|M_{1s}, P, K, V, \chi, e, \omega, s, I)]_\beta, \quad (26)$$

in the interval from $\beta = 0$ to 1, from the finite set. There are at least two difficulties with this approach. First, any errors in $\ln [p(D|M_{1s}, I)]$ are then magnified when we compute the desired global likelihood $p(D|M_{1s}, I)$ by exponentiating $\ln [p(D|M_{1s}, I)]$. This may require many more than the 11 tempering levels used in the current analysis. Second, it is necessary to establish that the samples for each tempering level have converged to a stationary distribution.

From the results of § 5.1, it is clear that we did not achieve convergence with MCMC runs that explored the full prior range so we are not able to use the above method for model selection calculations. However, we were able to identify three possible orbital solutions with periods of 128, 190, and 376 days. Subsequent MCMC runs explored these three peak regions and defined ranges of parameter space where the target probability distribution has a significant probability density. For each of the three ranges we can compute the contribution to the global likelihood by straight Monte Carlo integration. As discussed in § 4, Monte Carlo integration can be very inefficient in exploring the whole prior parameter range, but once we have established the significant regions of parameter space with the MCMC results, this is no longer the case.

The outer borders of the MCMC marginal parameter distributions (for each peak) were used to delineate the boundaries of the volume of parameter space to be used in the Monte Carlo integration. Again, the prior parameter ranges used in the calculation of the target probability distribution within each of these three volumes are those given in Table 2.

Table 5 summarizes our model selection calculations. Since models M_0 and M_{0s} involved only one and two unknown parameters, respectively, the required integrals were performed numerically. The global likelihood, $p(D|M_1, I)$, is just the sum of the contributions from the three period peaks,[7] and similarly for $p(D|M_{1s}, I)$. Assuming equal prior probabilities for models M_0, M_{0s}, M_1, and M_{1s}, the probability of model M_{1s} is given by

$$p(M_{1s}|D, I) = p(D|M_{1s}, I)[p(D|M_0, I) + p(D|M_{0s}, I)$$
$$+ p(D|M_1, I) + p(D|M_{1s}, I)]^{-1}. \quad (27)$$

---

[7] This assumes that there are no other solutions within the prior parameter space that were missed by the MCMC algorithm, perhaps because they are pathologically narrow.
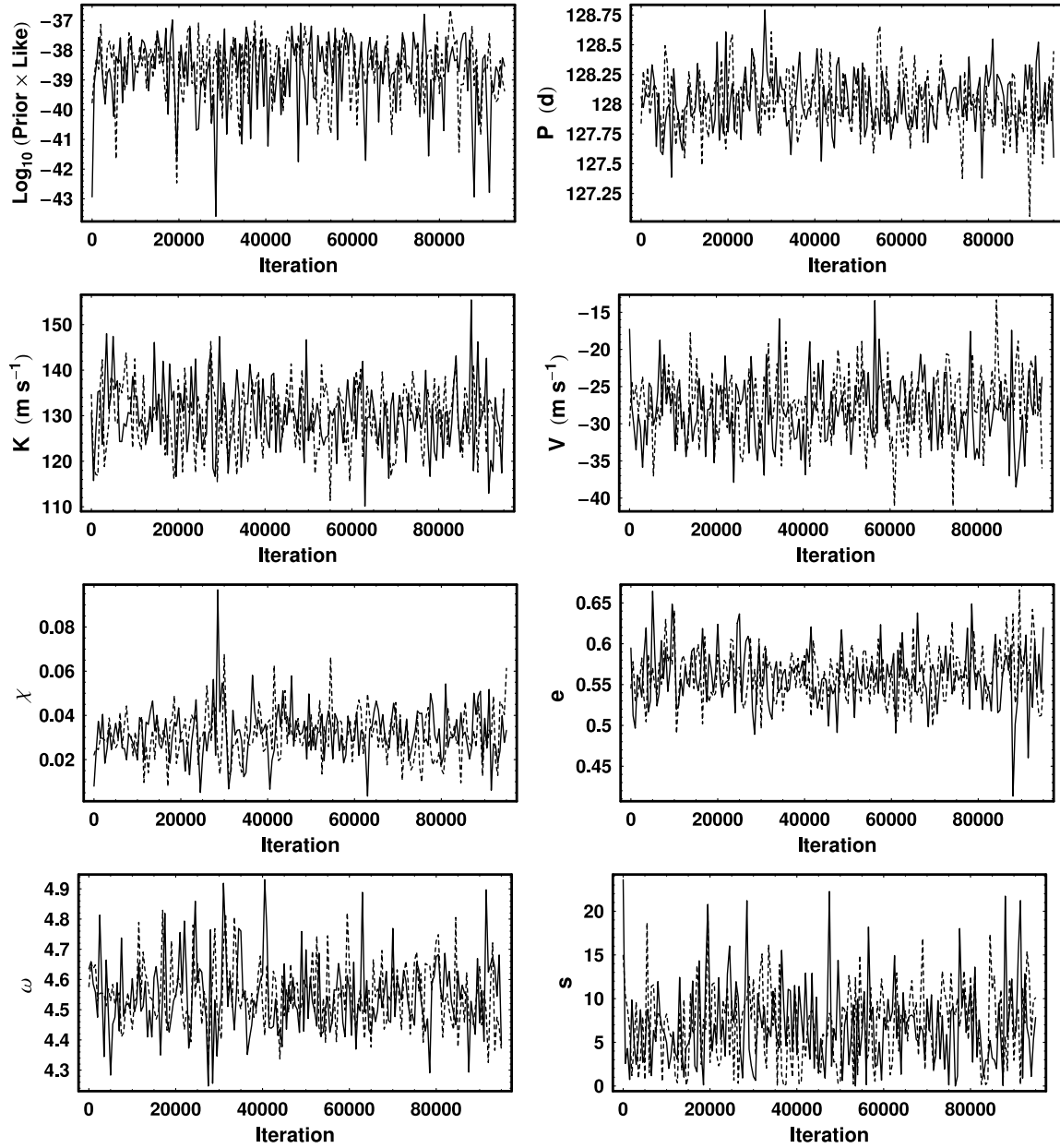
Fig. 8.—Comparison of MCMC parameter iterations for two different runs for model M_{1s} after removal of the burn-in period. The top left panel shows prior × likelihood vs. iteration number.
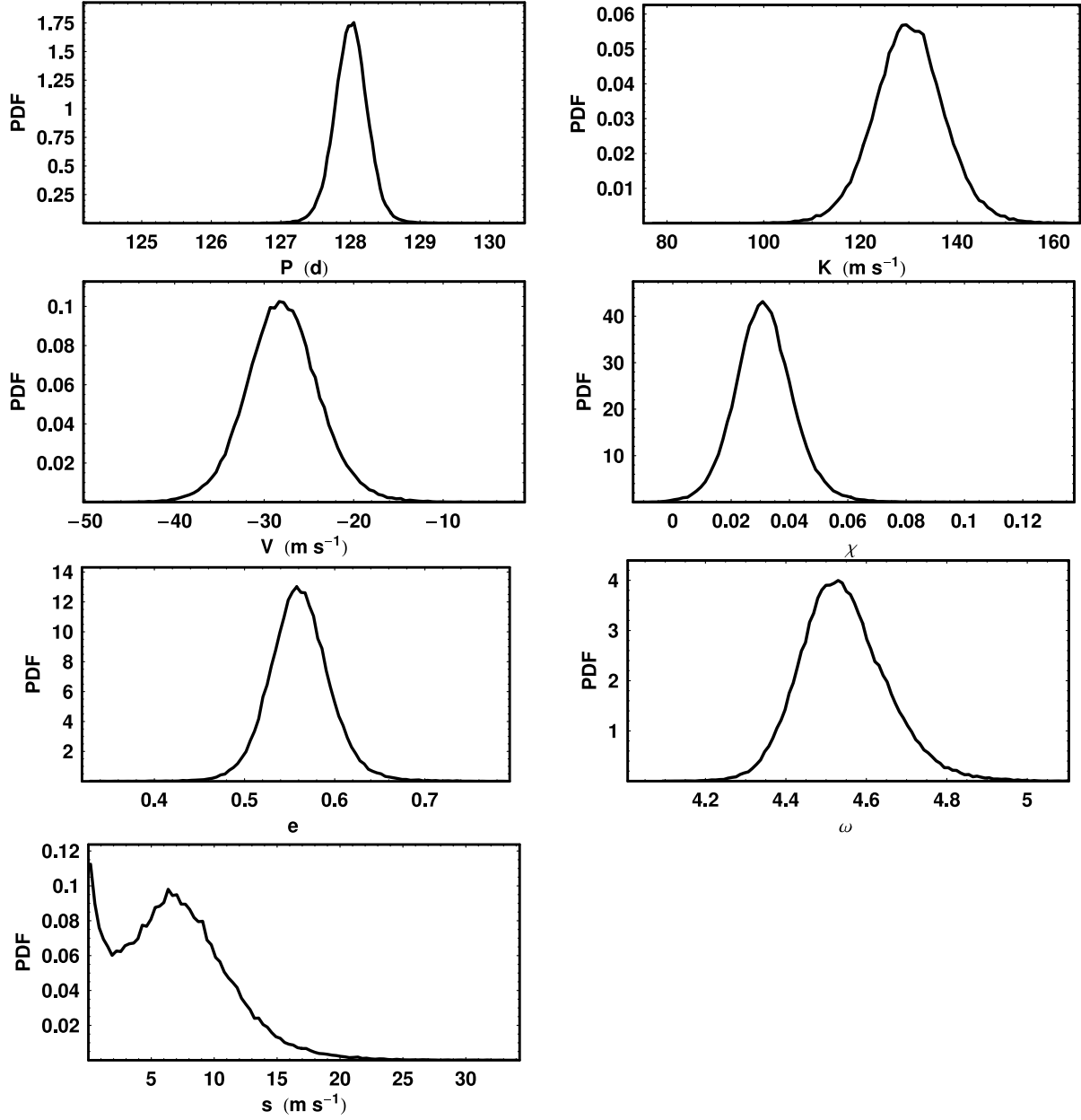
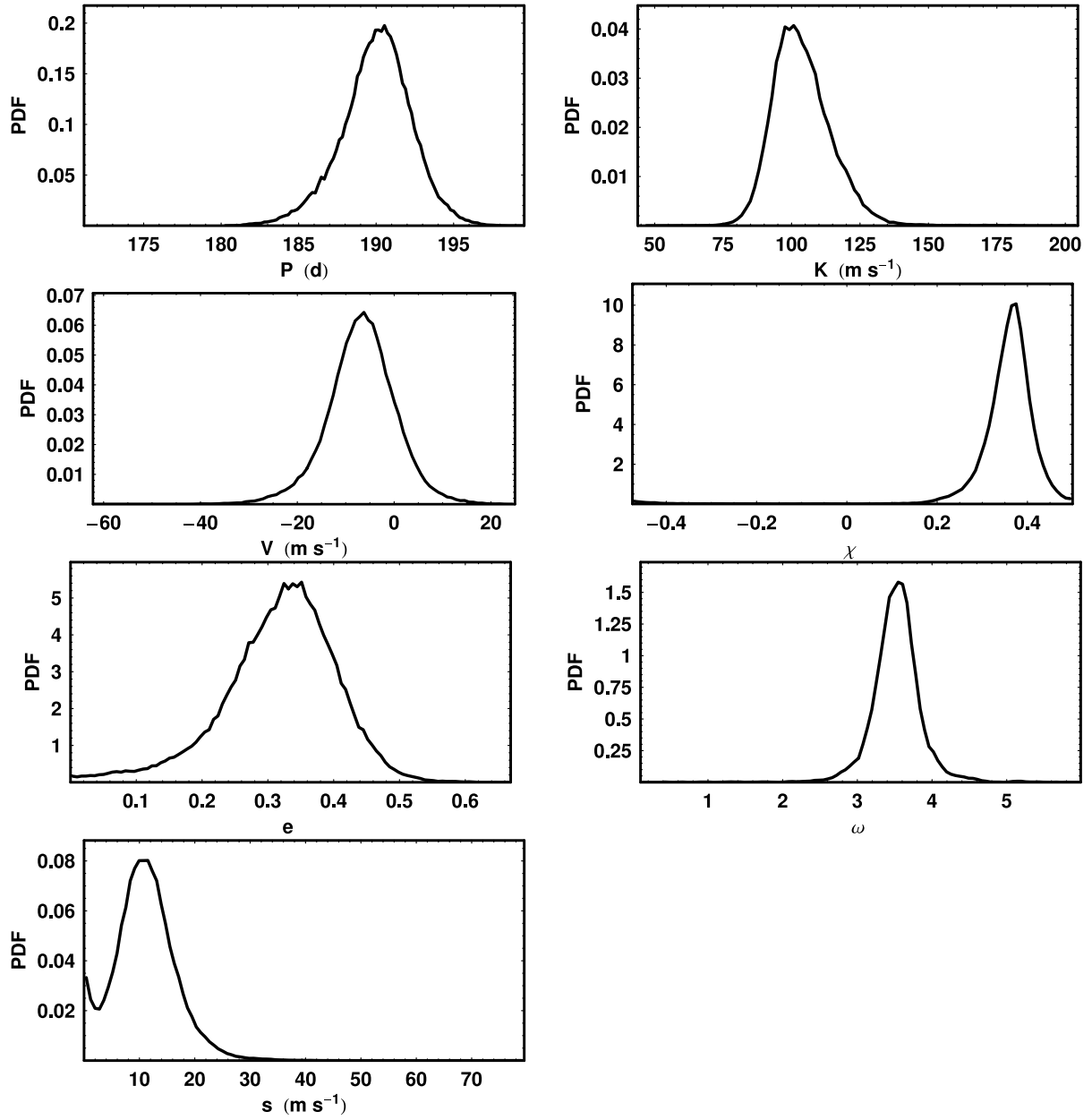Fig. 9.—Marginal PDFs for model M_{1s} parameters corresponding to the 128 day peak.

FIG. 10.—Marginal PDFs for model M_{1s} parameters corresponding to the 190 day peak.
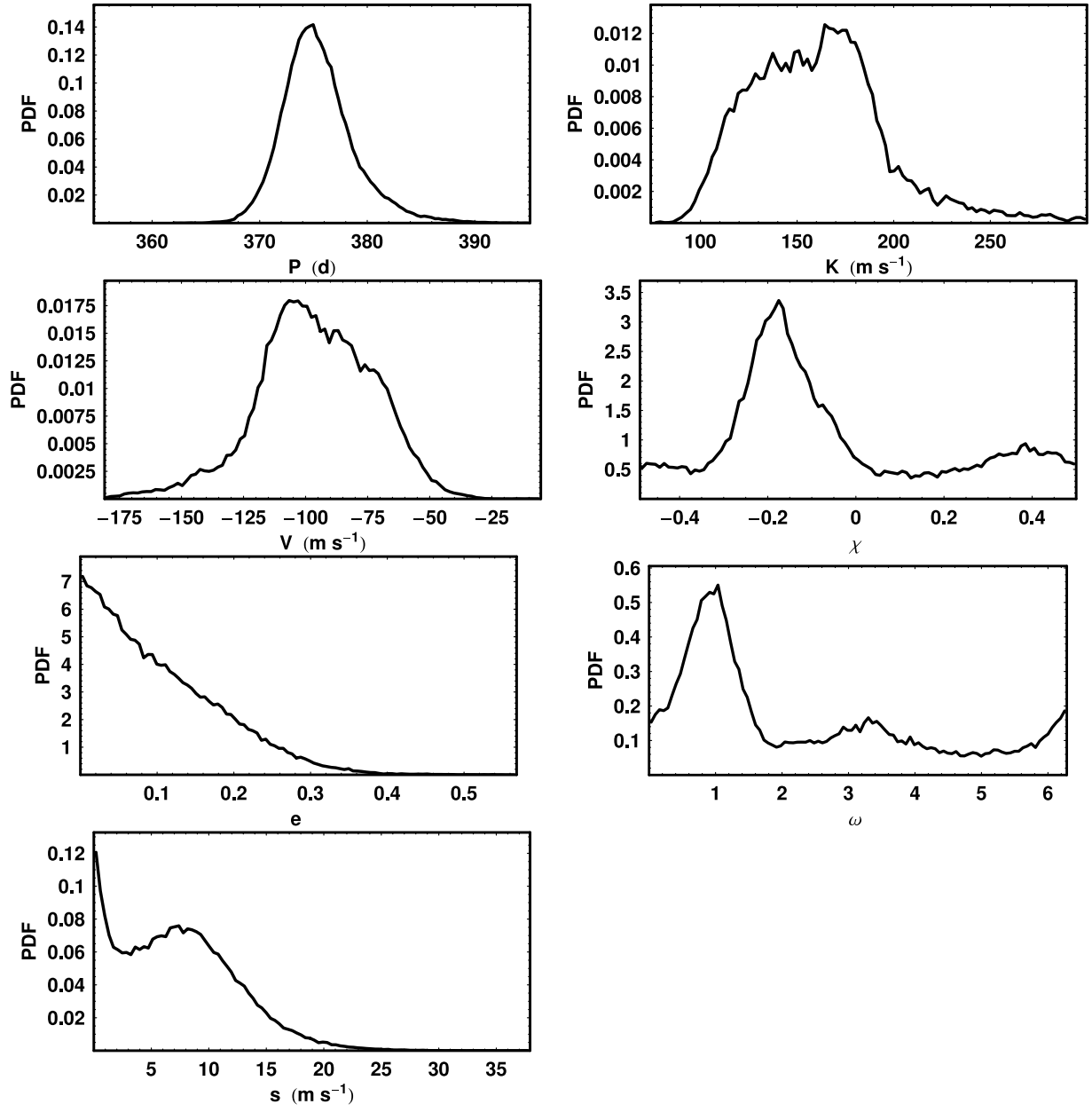
FIG. 11.—Marginal PDFs for model M_{1s} parameters corresponding to the 376 day peak.
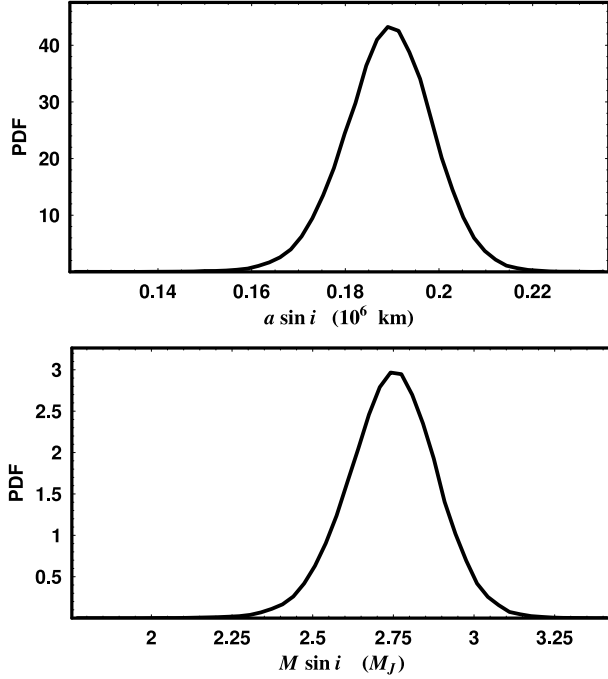
Fig. 12.—Posterior PDFs for $a \sin i$ and $M \sin i$ for the 128 day orbital solution.

The probabilities of the other three models are calculated in a similar fashion. The Bayes factor is just the ratio of the global likelihood of a particular model to model M_{0s}, taken as a reference.

Examination of Table 5 indicates that a one-planet model is strongly favored over M_0 and M_{0s}. Thus, there is very strong evidence for the existence of one planet. In spite of an additional Occam penalty associated with the s parameter, $p(D|M_{1s})$ is slightly greater than $p(D|M_1)$, which suggests that there is an additional source of variability in the radial velocity data. One possible explanation is that there is an additional planet(s) tugging on the star.

Within the M_{1s} model, the probabilities of the three orbital periods are proportional to their contribution to the global likelihood. If we designate $p(M_{1s(P128)}|D, I)$ as the probability of the 128 day orbit, then

$$
p\big(M_{1s(P128)}|D, I\big) = p\big(D|M_{1s(P128)}, I\big)\Big[p\big(D|M_{1s(P128)}, I\big)
$$
$$
+ p\big(D|M_{1s(P190)}, I\big) + p\big(D|M_{1s(P376)}, I\big)\Big]^{-1}. \quad (28)
$$

TABLE 5
MODEL PROBABILITIES AND BAYES FACTORS

Model	Period (days)	Mean s (m s^{-1})	$p(D|$Model, $I)$	Bayes Factor	Probability
M_0........	...	0	9.6×10^{-232}	3.4×10^{-188}	2×10^{-149}
M_{0s}.......	...	69	2.8×10^{-44}	1.0	6×10^{-7}
M_1........	128	0	$(7.5 \pm 0.03) \times 10^{-40}$	...	...
	190	0	$(2.2 \pm 0.02) \times 10^{-40}$	...	...
	376	0	$(1.5 \pm 0.2) \times 10^{-38}$	...	...
	All	0	$(1.6 \pm 0.2) \times 10^{-38}$	5.7×10^5	0.36
M_{1s}.......	128	7	$(1.2 \pm 0.03) \times 10^{-39}$	...	...
	190	11	$(1.1 \pm 0.01) \times 10^{-39}$	...	...
	376	8	$(2.6 \pm 0.3) \times 10^{-38}$	...	...
	All	8	$(2.8 \pm 0.3) \times 10^{-38}$	9.3×10^5	0.64

The probabilities for the other two orbital solutions were computed in a similar fashion.

Similarly, for model M_1 we can write

$$
p\big(M_{1(P128)}|D, I\big) = p\big(D|M_{1(P128)}, I\big)\Big[p\big(D|M_{1(P128)}, I\big)
$$
$$
+ p\big(D|M_{1(P190)}, I\big) + p\big(D|M_{1(P376)}, I\big)\Big]^{-1}. \quad (29)
$$

Since the relative probabilities of M_1 and M_{1s} are not very different, we base our final probability of the 128 day orbital solution, designated $p(M_{128}|D, I)$, on a weighted sum of equations (28) and (29). Thus, $p(M_{128}|D, I)$ is given by

$$
p(M_{128}|D, I) = p(M_{1s}|D, I)p\big(M_{1s(P128)}|D, I\big)
$$
$$
+ p(M_1|D, I)p\big(M_{1(P128)}|D, I\big). \quad (30)
$$

The last line of Table 4 gives our final probability assignments for the three one-planet orbital solutions based on equations (28)–(30).

6. DISCUSSION

In their paper, Tinney et al. (2003) report that they used a nonlinear least-squares analysis, which typically requires a good initial guess of the parameters to avoid becoming stuck in a local minimum of χ^2. No mention is made of what periodogram was used to obtain an initial estimate of the orbital period. In an attempt to understand the period discrepancy, we computed the Lomb-Scargle (LS) periodogram for the same data with the mean velocity subtracted, which is shown in the top panel of Figure 13. This is a widely used periodogram for dealing with sparsely

TABLE 4
PARAMETER VALUES AND 68% CREDIBLE REGIONS FOR M_{1s} PARAMETERS

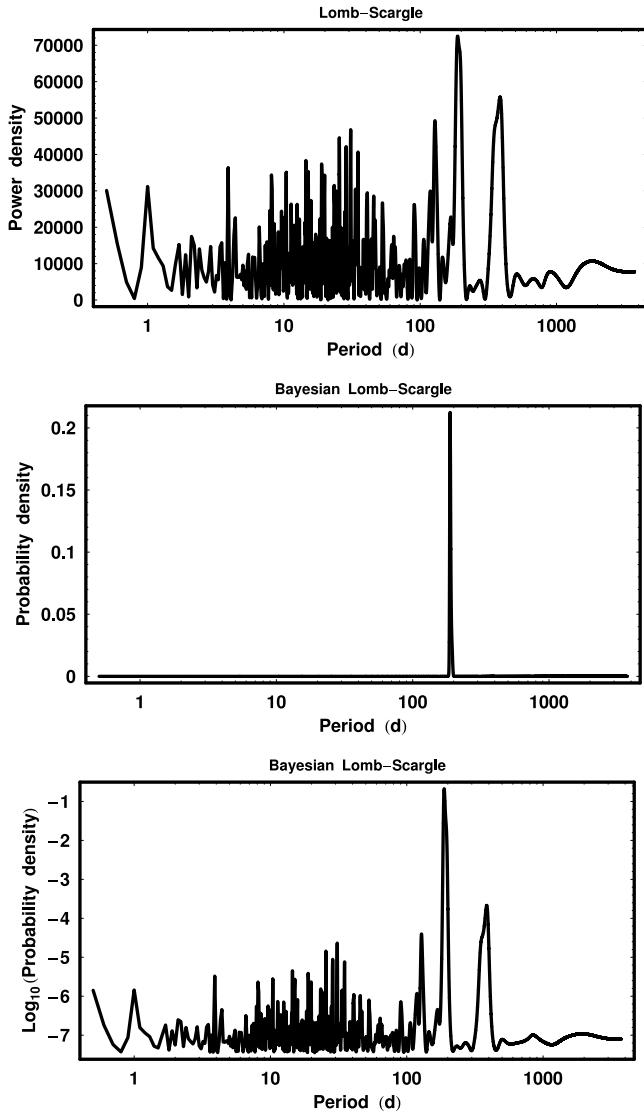| Parameter	Tinney et al. (2003)	$P = 128$ days	$P = 190$ days	$P = 376$ days
P (days).................................	190.5 ± 3.0	$127.88^{+0.37}_{-0.09}$	$190.4^{+1.8}_{-2.1}$	$376.2^{+1.4}_{-4.3}$
K (m s^{-1}).............................	108 ± 8.0	131^{+5}_{-7}	106.2^{+6}_{-13}	132^{+56}_{-7}
V (m s^{-1}).............................	...	$-26.3^{+1.7}_{-5.3}$	$-6.2^{+4.1}_{-6.8}$	-77^{+6}_{-39}
e...	0.34 ± 0.08	$0.56^{+0.03}_{-0.03}$	$0.33^{+0.06}_{-0.07}$	$0.10^{+0.05}_{-0.1}$
χ...	...	$0.030^{+0.009}_{-0.009}$	$0.37^{+0.03}_{-0.05}$	$-0.17^{+0.064}_{-0.30}$
ω (deg)................................	207 ± 30	259.2^{+6}_{-6}	205^{+10}_{-17}	49^{+309}_{-46}
$a \sin i$ (10^6 km).....................	0.265 ± 0.01	$0.191^{+0.007}_{-0.011}$	$0.263^{+0.013}_{-0.032}$	$0.67^{+0.29}_{-0.04}$
$M \sin i$...................................	3.0 ± 0.3	$2.78^{+0.09}_{-0.16}$	$2.93^{+0.09}_{-0.34}$	$4.9^{+2.0}_{-0.5}$
Periastron passage (JD $-2,450,000$).............	951 ± 12	$1079.9^{+1.1}_{-1.3}$	952^{+9}_{-7}	905^{+50}_{-231}
rms residuals (m s^{-1})...................	18	10.6	14.7	13.7
Probability................................	...	0.044	0.029	0.926

FIG. 13.—*Top*: LS periodogram for the HD 73526 data. *Middle*: Probability density for the period computed from a Bayesian generalization of the LS periodogram due to Bretthorst (2001b). *Bottom*: $\log_{10}$ (probability density).

sampled time series. All three peaks are present in the LS periodogram together with a multitude of smaller peaks at shorter periods. The 190 day period is clearly the biggest. In this paper we have been concerned with probabilities and probability density distributions that are not directly comparable to the power spectrum given in Figure 13. However, there is a useful Bayesian generalization of the LS periodogram due to Bretthorst (2001b) that yields the PDF for the period. This is shown in the middle panel of Figure 13 and only serves to emphasize the 190 day period. The logarithm of the PDF is shown in the bottom panel and is roughly similar in appearance to the LS periodogram in the top panel.

Why do the two approaches, (1) the method outlined in this paper and (2) the LS periodogram plus nonlinear least squares, lead to such different conclusions? First, the LS periodogram makes no use of the strong prior information that we have about the mathematical form of the periodic signal to be expected from a Keplerian orbit. This knowledge is built into our MCMC analysis, while the LS periodogram assumes that the periodic signal is sinusoidal. A second possible contributing reason is that the common implementation of the LS algorithm weights all points the same, but in fact the errors range from 14.9 to 6.3 m s^{-1}.

Although our 128 day period has the lowest MAP fit residuals, the global likelihood favors the 376 day period. At first sight, it appears surprising that the 68% credible regions for the parameters like K, V, and eccentricity are so much broader for the 376 day peak than for the other two less probable peaks. It is a simple matter to take the MCMC samples and look for correlations between parameter estimates.

Figure 14 demonstrates that there are strong correlations between velocity amplitude K, eccentricity, and V for this period. Examination of the MAP fit versus phase in Figure 7 suggests a possible reason for the correlations. There are no samples that provide a strong bound on the region of the minimum velocity excursion, which results in a wide range of orbital parameters that are consistent with the available data. Clearly, the addition of even one extra point in this region could significantly reduce the range of viable parameter space (see Loredo & Chernoff [2003] for a Bayesian strategy for adding extra points). Since HD 73526 is not circumpolar at the Anglo-Australian Telescope site, there will inevitably be a window of phase that is unobservable in any given year. Because of the difference between the orbital period and 1 yr, the 376 day orbital phase will shift by approximately 0.029 per year so it will be possible to fill in the region around the predicted velocity minimum within ∼7 yr.

Of course, only one of the three periods is correct, and some readers may wonder if there is some way to discern, without the
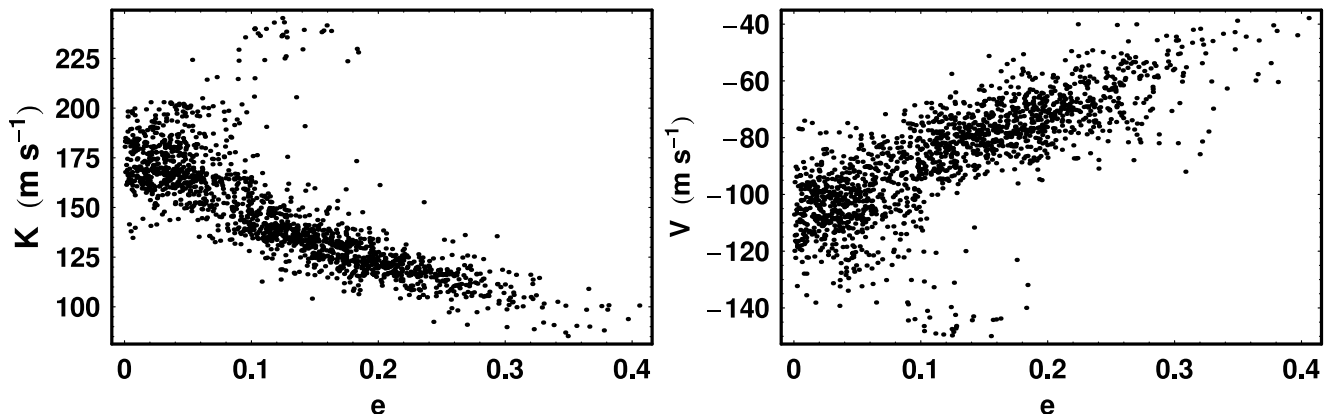

FIG. 14.—The left panel demonstrates a strong correlation between the radial velocity amplitude K and the eccentricity for MCMC points associated with the 376 day period. The right panel demonstrates the correlation between V and eccentricity.
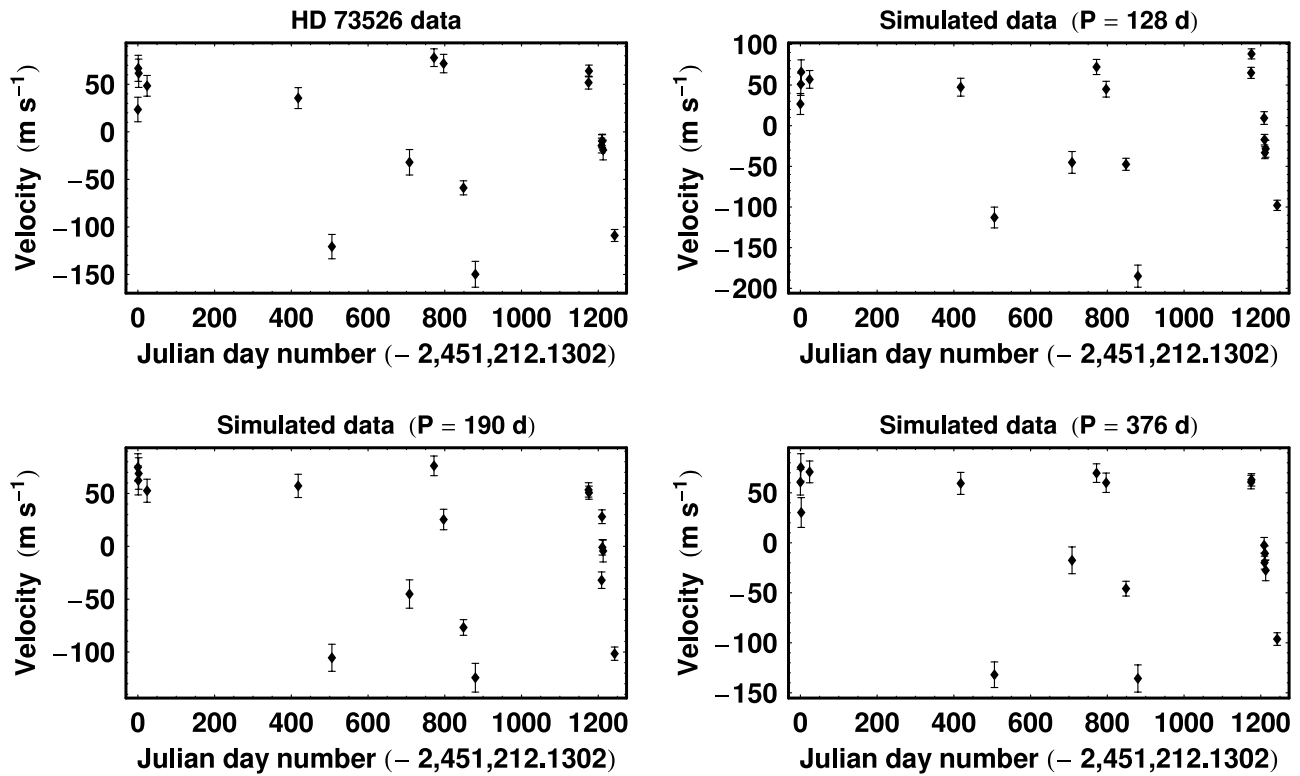
Fig. 15.—Comparison of the three simulated data sets, one for each orbital solution, to the HD 73526 data set.

use of additional data, whether any of the solutions is clearly an alias that can be discounted. To answer this question, we simulated three data sets using the MAP parameters found for the 128, 190, and 376 day solutions. Each simulated radial velocity curve was sampled at the same time as the HD 73526 data set, and independent Gaussian random noise was added corresponding to the quadrature sum of the HD 73526 measurement errors and

the mean value of s determined for that solution. The four panels in Figure 15 compare the simulated data sets to the HD 73526 data set. The Bayesian generalization (Bretthorst 2001b) of the LS periodogram was run on each simulated data set, and the results are shown in Figure 16 compared to the result for the HD 73526 data. All four periodograms look very similar. They all show peaks at 128, 190, and 376 days (plus others at lower
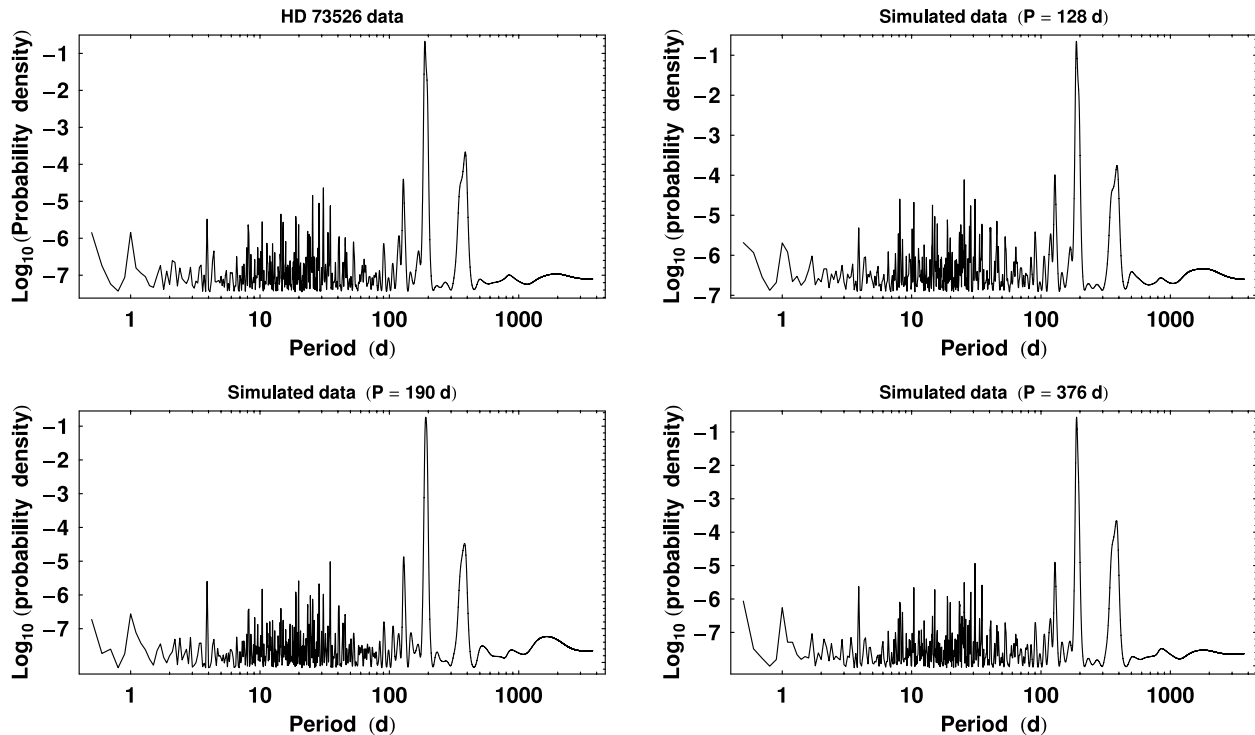

Fig. 16.—Comparison of the logarithm of the probability density for the three simulated data sets, one for each orbital solution, and the HD 73526 data set.

periods), and the relative heights of the peaks are very similar. Both the Bayesian MCMC method and the LS method involve fitting nonlinear signal models to a sparse data set (with no interpolation); the signal models just happen to be different. It is well known that nonlinear models can yield multiple solutions (peaks in probability). Any judgment of which solution is correct is based on a measure of the probability associated with each peak. As the simulation demonstrates, we have no other basis for judging which is the correct solution. In the Bayesian case, the final probabilities of the different solutions are simply an encoding of our current state of knowledge into a probability distribution about the hypothesis space of interest, which in this case is a hypothesis space of orbital periods. As our state of knowledge changes (more data are acquired), so in general will the encoded probability distribution.

7. CONCLUSIONS

In this paper we outlined a Bayesian approach to the analysis of precision radial velocity data for use in detecting extrasolar planets and provided an independent analysis of the data of Tinney et al. (2003). The analysis was based on a parallel tempering MCMC algorithm that incorporates as prior information the mathematical form of any possible Keplerian planetary signal embedded in the host star's reflex motion. The algorithm is capable of efficiently exploring all regions of the joint prior parameter space having significant probability, with the exception of pathologically narrow peaks, starting from an arbitrary location in the prior parameter space. This is in contrast to nonlinear least-squares analysis, which requires a good initial estimate of the starting parameter values. The MCMC analysis also yields the full marginal posterior PDF for each model parameter and any derived quantity like $M_j \sin i$, not just the MAP values and a Gaussian approximation of their uncertainties. The built-in Occam's razor in the Bayesian analysis provides a more rigorous approach to deciding whether a detection is believable. The method is capable of fitting a portion of an orbital period, so search periods longer than the duration of the data are possible. The MCMC algorithm incorporates a novel statistical control system that automates the selection of suitable Gaussian proposal distribution σ-values. This can result in huge savings in time because it eliminates many trial runs to manually establish a suitable set of proposal σ-values.

Based on our prior information and the data reported in Tinney et al. (2003), the one-planet model M_{1s} has the highest probability. Three possible orbits were found with periods of 128, 190, and 376 days. For model M_{1s}, the maximum value of prior $\times$ likelihood is 16 times larger for the 128 day orbit than for the next highest solution at 376 days. Although the 128 day orbit with an eccentricity of 0.56 ± 0.03 has the smallest residuals, formally the most probable orbit has a period of $376.2^{+1.4}_{-4.3}$ days and an eccentricity of $0.10^{+0.05}_{-0.10}$. This is because there is a much greater volume of parameter space with a significant probability density in the vicinity of the 376 day peak. This arises because there are no measurements that provide a strong bound on the minimum velocity excursion for the 376 day period, which results in a wide variety of possible Keplerian orbits that fit the existing data well. The addition of even one extra point in this region could significantly reduce the range of viable parameter space. The current Bayesian probabilities for the 376, 128, and 190 day orbital solutions are 0.93, 0.04, and 0.03, respectively. The previously reported orbit (Tinney et al. 2003) of 190.5 ± 3.0 days corresponds to our least probable orbit.

For all three orbital solutions there is evidence for an excess velocity residual in the range 7–11 km s^{-1}. This may indicate the presence of additional planets. In an attempt to identify a second planet, we reran our one-planet Bayesian MCMC algorithm on the MAP residuals. We also repeated the analysis for a two-planet model, M_2, using the original data. In all cases, the Bayes factor was less than 1 when compared to model M_{1s}. On the basis of our current state of knowledge, we are unable to demonstrate the existence of a second planet.

Mathematica versions of the MCMC algorithm, LS algorithm, and Bayesian generalization of LS, together with other useful tools for data analysis and plotting, are included with the software that accompanies the book "Bayesian Logical Data Analysis for the Physical Sciences," published by Cambridge University Press (Gregory 2005). Although this paper focuses on extrasolar planet detection, the software can be easily modified to deal with other nonlinear model fitting problems.

REFERENCES

Bretthorst, G. L. 1988, Bayesian Spectrum Analysis and Parameter Estimation (New York: Springer)
———. 2001a, in AIP Conf. Proc. 567, Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 19th Internat. Workshop, ed. J. Rychert, G. Erickson, & C. R. Smith (New York: AIP), 1
———. 2001b, in AIP Conf. Proc. 568, Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 20th Internat. Workshop, ed. A. Mohammad-Djafari (New York: AIP), 241
Brown, R. 2004, ApJ, 610, 1079
Eggenberger, A., Udry, S., & Mayor, M. 2004, A&A, 417, 353
Fisher, D. A., Butler, R. P., Marcy, G. W., Vogt, S. S., & Henry, G. W. 2003, ApJ, 590, 1081
Ford, E. B. 2005, AJ, 129, 1706
Gelman, A., & Rubin, D. B. 1992, Stat. Sci., 7, 457

Gregory, P. C. 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with *Mathematica* Support (Cambridge: Cambridge Univ. Press)
Gregory, P. C., & Loredo, T. J. 1992, in Maximum Entropy and Bayesian Methods, ed. A. Mohammad-Djafari & G. Demoment (Dordrecht: Kluwer), 225
Loredo, T. J., & Chernoff, D. F. 2003, in Statistical Challenges in Modern Astronomy III, ed. E. D. Feigelson & G. J. Babu (New York: Springer), 57
Mayor, M., & Queloz, D. 1995, Nature, 378, 355
McArthur, B. E., et al. 2004, ApJ, 614, L81
Roberts, G. O., Gelman, A., & Gilks, W. R. 1997, Ann. Appl. Probability, 7, 110
Tinney, C. G., Butler, R. P., Marcy, G. W., Jones, H. R. A., Penny, A. J., McCarthy, C., Carter, B. D., & Bond, J. 2003, ApJ, 587, 423
Wright, J. T. 2005, PASP, 117, 657