



Extra-solar Planets via Bayesian model fitting

Phil Gregory

University of British Columbia

March 2011

A rich harvest of over 500 extra-solar planets is motivating a significant effort to improve statistical tools to better understand parameter uncertainties and compute model probabilities.

Recent work has highlighted a Bayesian Markov chain Monte Carlo approach - applicable to many nonlinear model fitting problems.

Outline

- 1. Bayesian primer** 1
- 2. Status of extra-solar planets (exoplanets)** 2
 - Challenge of nonlinear models** 3
- 3. Introduction to Markov chain Monte Carlo (MCMC)** 4
 - Fusion MCMC** 5
 - Multi-planet Kepler periodogram** 6
- 4. Evidence for 3 planets in 47 Ursae Majoris** 7
- 5. Test of habitable zone planet Gliese 581g** 8
- 6. Conclusions** 9

What is Bayesian Probability Theory? (BPT)

BPT = a theory of extended logic

Deductive logic is based on Axiomatic knowledge.

In science we never know any theory of nature is true because our reasoning is based on incomplete information.

Our conclusions are at best probabilities.

Any extension of logic to deal with situations of incomplete information (realm of inductive logic) requires a theory of probability.

A new perception of probability has arisen in recognition that the mathematical rules of probability are not merely rules for manipulating random variables.

They are now recognized as valid principles of logic for conducting inference about any hypothesis of interest.

This view of, ``Probability Theory as Logic'', was championed in the late 20th century by E. T. Jaynes.

**``Probability Theory: The Logic of Science''
Cambridge University Press 2003**

It is also commonly referred to as Bayesian Probability Theory in recognition of the work of the 18th century English clergyman and Mathematician Thomas Bayes.

Logic is concerned with the truth of propositions.

A proposition asserts that something is true.

Examples of propositions:

$A \equiv$ “The newly discovered radio astronomy object is a galaxy.”

$B \equiv$ “The measured redshift of the object is 0.150 ± 0.005 .”

$A \equiv$ “Theory X is correct.”

$\bar{A} \equiv$ “Theory X is not correct.”

$A \equiv$ “The frequency of the signal is between f and $f + df$.”

We will need to consider compound propositions like A, B which asserts that propositions A and B are true

$A, B/C$ asserts that propositions A and B are true given that proposition C is true

Rules for manipulating probabilities

Sum rule : $p(A | C) + p(\bar{A} | C) = 1$

Product rule : $p(A, B | C) = p(A | C) p(B | A, C)$
 $= p(B | C) p(A | B, C)$

Re-arrange the two RH sides of product rule gives

Bayes theorem :

$$p(A | B, C) = \frac{p(A | C) p(B | A, C)}{p(B | C)}$$

How to proceed in a Bayesian data analysis?

Write down Bayes' theorem, identify the terms and solve.

$$p(H_i | D, I) = \frac{p(H_i | I) \times p(D | H_i, I)}{p(D | I)}$$

Posterior probability
that H_i is true, given
the new data D and
prior information I

Normalizing constant

Every item to the right of the
vertical bar $|$ is assumed to be true

The likelihood $p(D | H_i, I)$, also written as $\mathcal{L}(H_i)$, stands for the probability that we would have gotten the data D that we did, if H_i is true.

As a theory of extended logic BPT can be used to find **optimal answers** to well posed scientific questions for a given state of knowledge, in contrast to a numerical recipe approach.

Two basic problems

1. Model selection (discrete hypothesis space)

“Which one of 2 or more models (hypotheses) is most probable given our current state of knowledge?”

e.g.

- Hypothesis or model M_0 asserts that the star has no planets.
- Hypothesis M_1 asserts that the star has 1 planet.
- Hypothesis M_i asserts that the star has i planets.

2. Parameter estimation (continuous hypothesis)

“Assuming the truth of M_1 , solve for the probability density distribution for each of the model parameters based on our current state of knowledge.”

e.g.

- Hypothesis H asserts that the orbital period is between P and $P+dP$.

Significance of this development

Probabilities are commonly quantified by a real number between 0 and 1.



The end-points, corresponding to absolutely false and absolutely true, are simply the extreme limits of this infinity of real numbers.

Bayesian probability theory spans the whole range.

Deductive logic is just a special case of Bayesian probability theory in the idealized limit of complete information.

Calculation of a simple Likelihood $p(D | M, \vec{X}, I)$

Let d_i represent the i^{th} measured data value . We model d_i by,

$$d_i = f_i(\vec{X}) + e_i$$

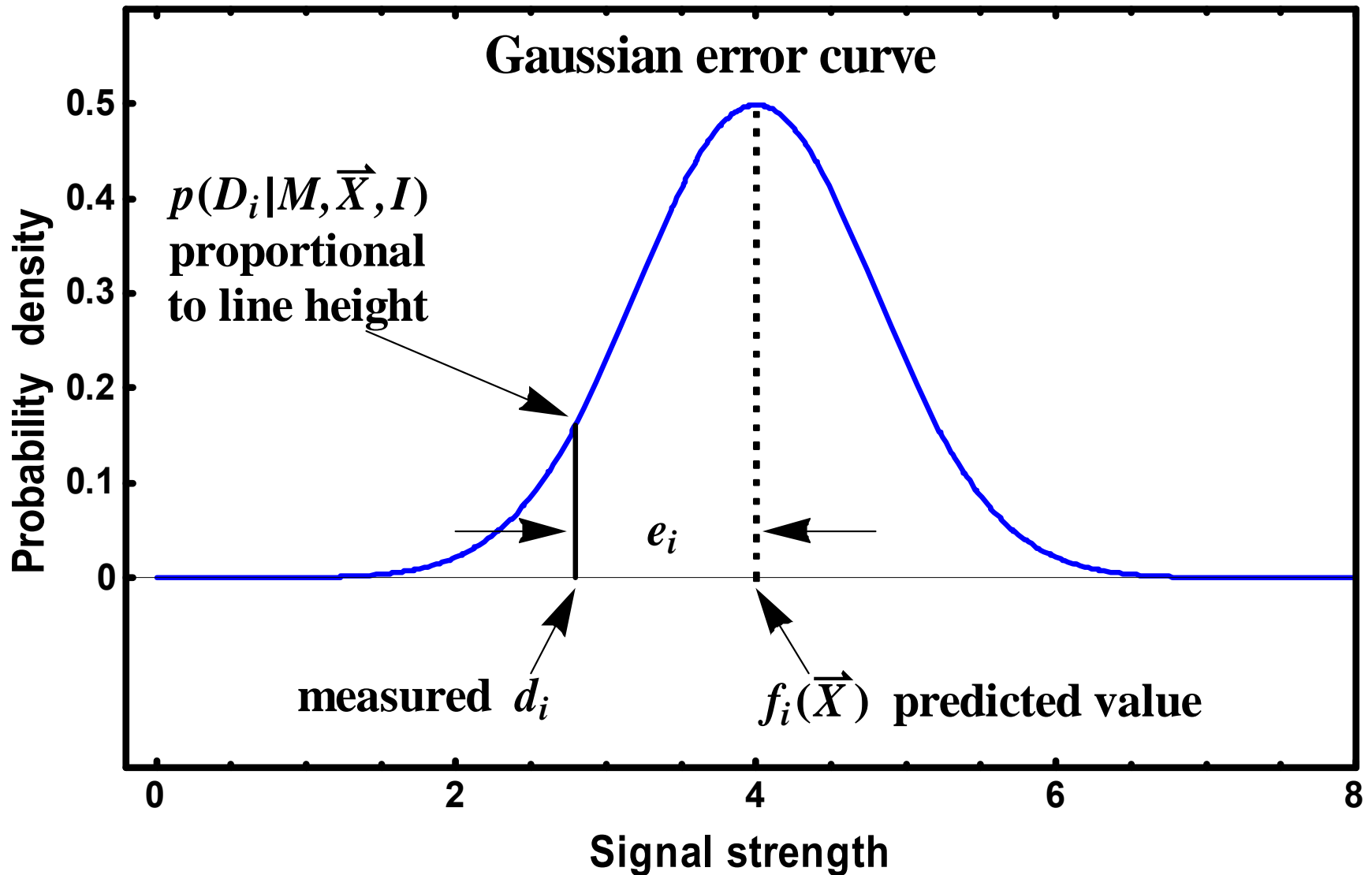
Model prediction for i^{th} data value
for current choice of parameters \vec{X}

where e_i represents the error component in the measurement.

Since M, \vec{X} is assumed to be true, if it were not for the error e_i , d_i would equal the model prediction f_i .

Now suppose prior information I indicates that e_i has a Gaussian probability distribution. Then

$$\begin{aligned} p(D_i | M, \vec{X}, I) &= \frac{1}{\sigma_i \sqrt{2\pi}} \text{Exp}\left[-\frac{e_i^2}{2\sigma_i^2}\right] \\ &= \frac{1}{\sigma_i \sqrt{2\pi}} \text{Exp}\left[-\frac{(d_i - f_i(\vec{X}))^2}{2\sigma_i^2}\right] \end{aligned}$$



Probability of getting a data value d_i a distance e_i away from the predicted value f_i is proportional to the height of the Gaussian error curve at that location.

Calculation of a simple Likelihood $p(D | M, \vec{X}, I)$

For independent data the likelihood for the entire data set $D=(D_1, D_2, \dots, D_N)$ is the product of N Gaussians.

$$p(D | M, \vec{X}, I) = (2\pi)^{-N/2} \left\{ \prod_{i=1}^N \sigma_i^{-1} \right\} \text{Exp} \left[-0.5 \sum_{i=1}^N \frac{(d_i - f_i(\vec{X}))^2}{\sigma_i^2} \right]$$

The familiar χ^2 statistic used in least-squares

Maximizing the likelihood corresponds to minimizing χ^2

Recall: Bayesian posterior \propto prior \times likelihood

Thus, only for a uniform prior will a least-squares analysis yield the same solution as the Bayesian posterior.

Simple example of when not to use a uniform prior

In the exoplanet problem the prior range for the unknown orbital period P is very large from ~ 1 day to 1000 yr (upper limit set by perturbations from neighboring stars).

Suppose we assume a uniform prior probability density for the P parameter. This would imply that we believed that it was $\sim 10^4$ times more probable that the true period was in the upper decade (10^4 to 10^5 d) of the prior range than in the lowest decade from 1 to 10 d.

$$\frac{\int_{10^4}^{10^5} p(P | M, I) dP}{\int_1^{10} p(P | M, I) dP} = 10^4$$

Usually, expressing great uncertainty in some quantity corresponds more closely to a statement of scale invariance or equal probability per decade. The Jeffreys prior has this scale invariant property.

$$p(\ln P | M, I) d \ln P = \frac{d \ln P}{\ln(P_{max} / P_{min})}$$

Integration not minimization

A full Bayesian analysis requires integrating over the model parameter space. Integration is more difficult than minimization.

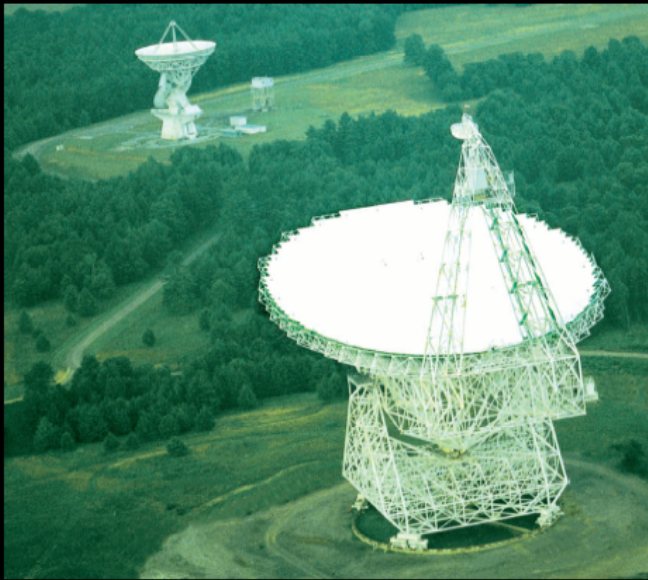
However, the Bayesian solution provides the most accurate information about the parameter errors and correlations without the need for any additional calculations, i.e., Monte Carlo simulations.

**Shortly discuss an efficient method for
Integrating over a large parameter space
called Markov chain Monte Carlo (MCMC).**

PHIL GREGORY

Bayesian Logical Data Analysis for the Physical Sciences

A Comparative Approach with
Mathematica Support



CAMBRIDGE

Chapters

1. Role of probability theory in science
2. Probability theory as extended logic
3. The how-to of Bayesian inference
4. Assigning probabilities
5. **Frequentist statistical inference**
6. **What is a statistic?**
7. **Frequentist hypothesis testing**
8. Maximum entropy probabilities
9. Bayesian inference (Gaussian errors)
10. Linear model fitting (Gaussian errors)
11. Nonlinear model fitting
12. Markov chain Monte Carlo
13. Bayesian spectral analysis
14. Bayesian inference (Poisson sampling)

Introduces statistical inference in the larger context of scientific methods, and includes 55 worked examples and many problem sets.

End section 1

Resources and solutions

This title has free
Mathematica based support
software available

Exoplanets planets detected as of 20 Mar. 2011

538 exoplanets planets found in 449 planetary systems

The majority within a distance of 100 ly

55 multiple planet systems (one possible 7 planet system)

Detection method	No. of planets
Radial velocity	493
Transits	124
Microlensing	12
Imaging	21
Timing (includes 4 pulsar planets)	12

The Extrasolar Planets Encyclopaedia
Jean Schneider
 (CNRS-LUTH, Paris Observatory)

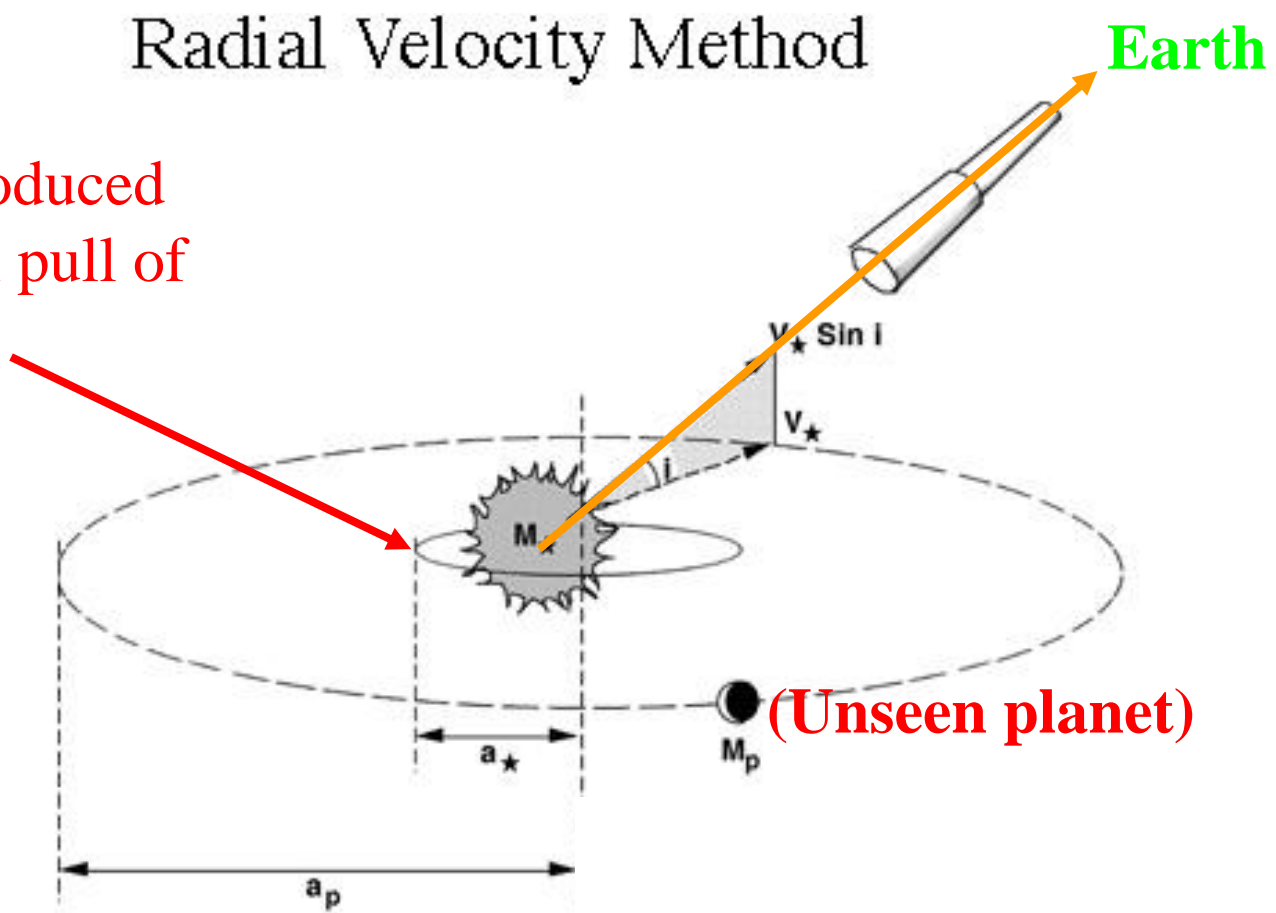
image

This situation is rapidly changing following the launch of the Kepler planet transiting mission. The first 4 months of data yielded over 1200 candidates including 54 in habitable zone, 5 of which are earth sized.

Currently the best spectrograph is the European HARPS (high accuracy radial velocity planet searcher) instrument which is capable of long term sub m/s velocity precision.

Radial Velocity Method

Orbit of star produced by gravitational pull of unseen planet.

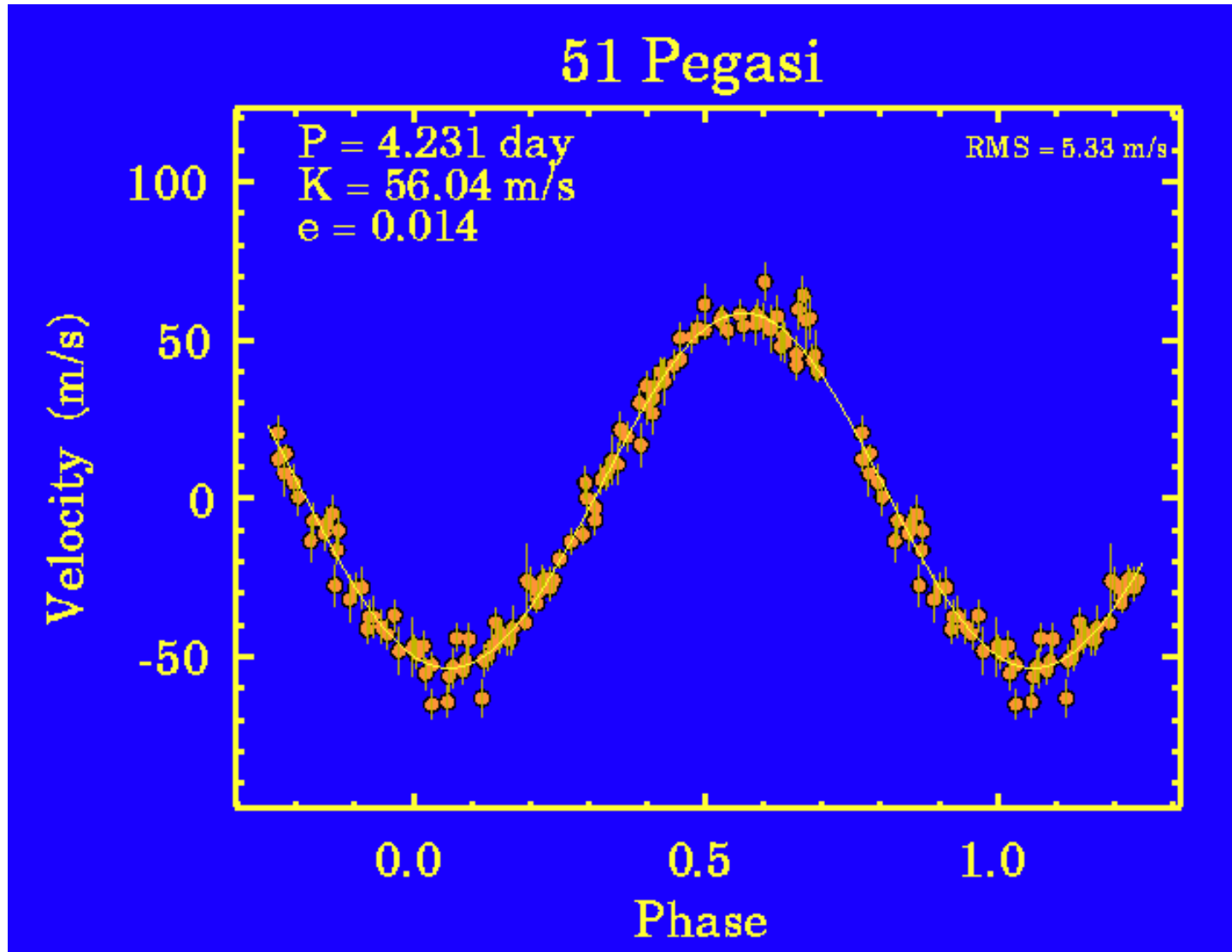


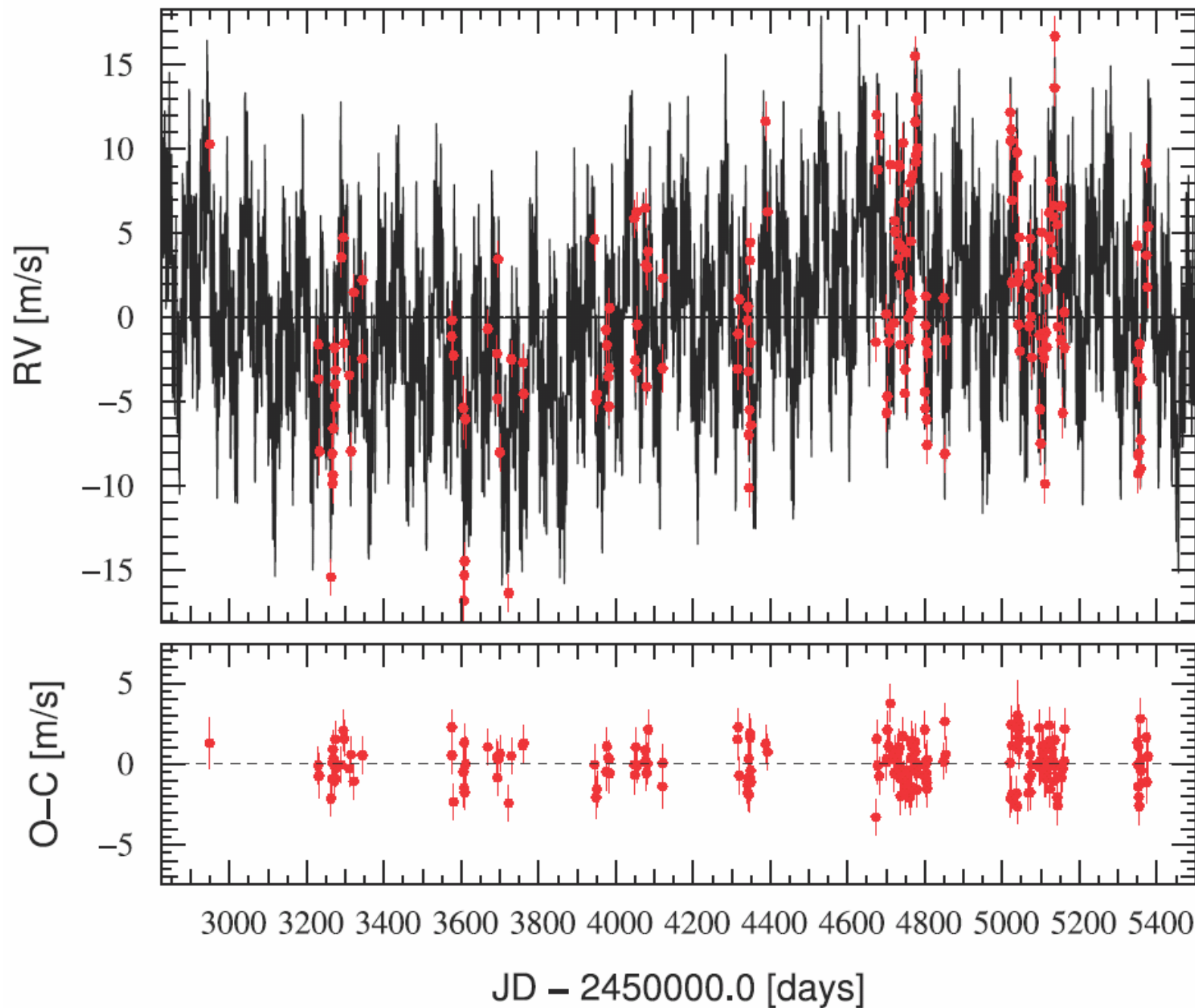
Because the star is a billion times brighter than a planet we are only able to detect a small number of extrasolar planets directly.

Instead we look for the reflex motion of the star due to gravitational tug from the planet.

A velocity (Doppler shift) of 1 m/s corresponds to a displacement of the spectrum by 0.0005 pixels, or about 40 silicon atoms on the CCD detector.

1995, first detection of planet orbiting a sun like star by Mayor and Queloz ($M \sin i \sim \frac{1}{2} M_J$)





Gliese 581 the star with two possible habitable zone planets

History

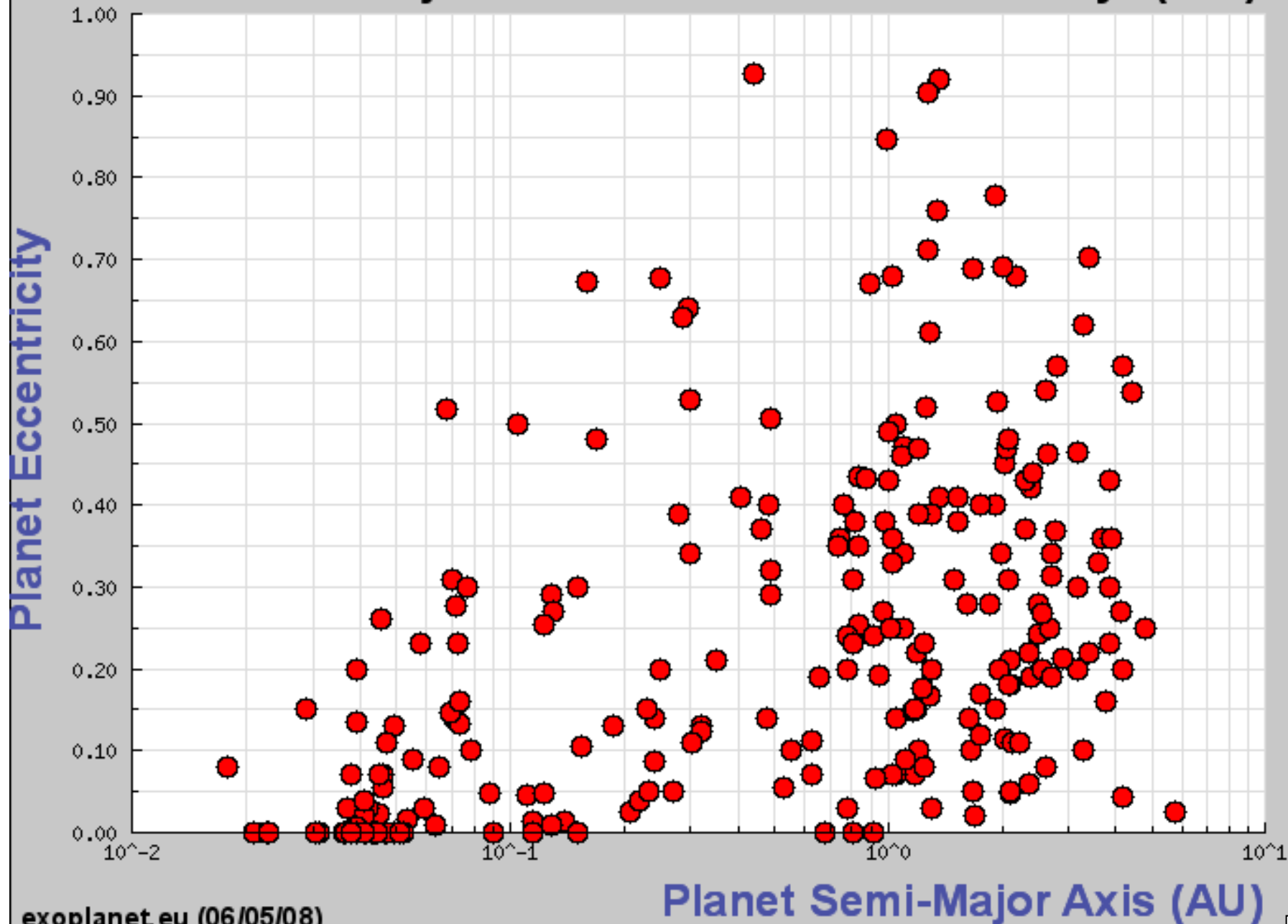
- 1) 2005 to 2009,
 - Planet e is 1.9 Earth mass
 - Planet b is 16 Earth mass
 - Planet c is 5 Earth mass
 - Planet d is 7 Earth mass (edge of habitable zone)

Latest paper: [M. Mayor et al., A&A, 507, p. 487, Nov. 2009](#)

- 2) 2010,
 - Planet f is 7 Earth mass
 - Planet g is 3.1 Earth mass (middle of habitable zone)

[Steven Vogt et al., ApJ, 723, p. 954, 2011](#)

"Planet Semi-Major Axis" vs "Planet Eccentricity" (245)



The radial velocity equation, a nonlinear model

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

V = systematic velocity;

K = velocity amplitude = $\frac{2 \pi a \sin i}{P \sqrt{1-e^2}}$; (a = semi-major axis; i = inclination of orbit)

P = orbital period

χ = fraction of orbit prior to data taking that periastron occurred at
(χP has units of time)

e = orbital eccentricity

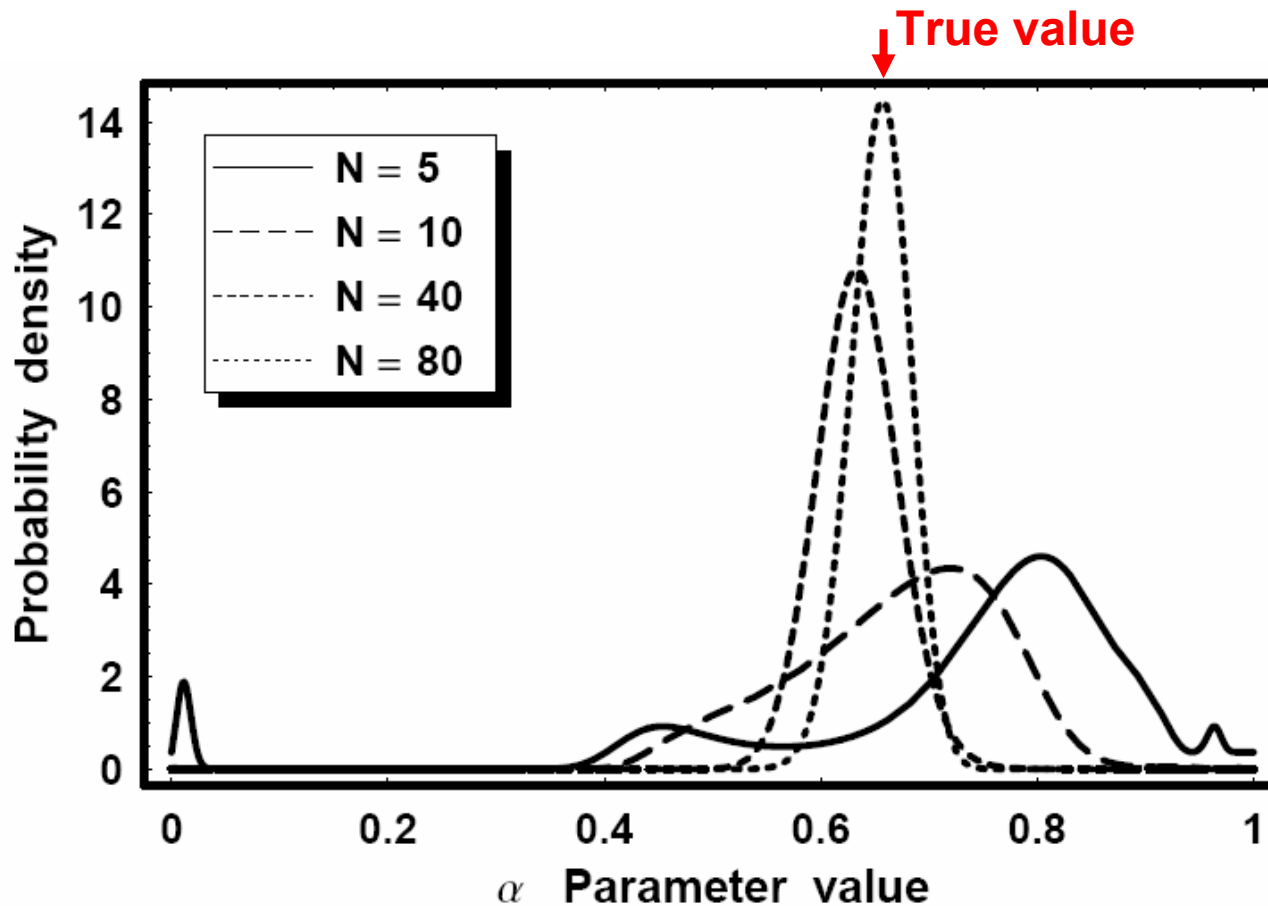
ω = Longitude of periastron

$\theta(t_i + \chi P)$ = True anomaly =
angle of star in orbit at time t_i measured with respect to periastron

solve

$$\frac{d\theta[t]}{dt} - \frac{2*\pi}{P(1-e^2)^{3/2}} (1 + e \text{Cos}[\theta])^2 = 0$$

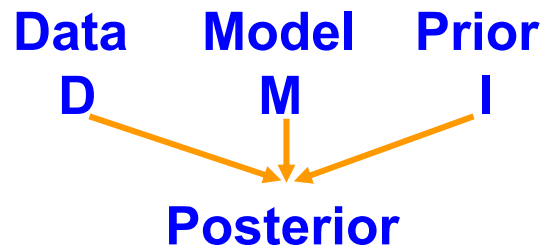
The challenge of nonlinear models, multiple peaks



The Bayesian posterior density for a nonlinear model with a single parameter, α , for 4 simulated data sets of different size ranging from $N = 5$ to $N = 80$. The $N = 5$ case has the broadest distribution and exhibits 4 maxima.

Asymptotic theory says that the maximum likelihood estimator becomes more unbiased, more normally distributed and of smaller variance as the sample size becomes larger.

Numerical tools for Bayesian model fitting



Linear models (uniform priors)

Posterior has a single peak
(multi-dimensional Gaussian)

Parameters given
by the normal equations
of linear least-squares

No integration required
solution very fast
using linear algebra

(chapter 10)

Nonlinear models

+ linear models (non-uniform priors)

Posterior may have multiple peaks

Brute force
integration

For some
parameters
analytic
integration
sometimes
possible

Asymptotic
approx.'s

peak finding
algorithms
(1) Levenberg-
Marquardt
(2) Simulated
annealing
(3) Genetic
algorithm
|
Laplace
approx.'s

(chapter 11)

Moderate
dimensions

quadrature
|
randomized
quadrature
|
adaptive
quadrature

High
dimensions

MCMC

(chapter 12)

Why is integration important for Bayesian analysis?

Parameter estimation: to find the marginal posterior probability density function (PDF) for the orbital period P , we need to integrate the joint posterior over all the other parameters.

$$\underbrace{p(P | D, M_1, s, I)}_{\text{Marginal PDF for } P} = \int dK dV d\chi d\mathbf{e} d\omega ds \underbrace{p(P, K, V, \mathbf{e}, \phi, \omega, \mathbf{s} | D, M_1, s, I)}_{\text{Joint posterior probability density function (PDF) for the parameters}}$$

Marginal PDF
for P

Joint posterior probability
density function (PDF) for
the parameters

Markov chain Monte Carlo (MCMC) algorithms provide a powerful means for efficiently computing integrals in many dimensions to within a constant factor. This factor is not required for parameter estimation.

After an initial burn-in period, the MCMC produces an equilibrium distribution of samples in parameter space such that the density of samples is **proportional** to the **joint posterior PDF**.

It is very efficient because, unlike straight Monte Carlo integration, it doesn't waste time exploring regions where the joint posterior is very small.

Starting point: Metropolis-Hastings MCMC algorithm

$P(X|D,M,I)$ = target posterior probability distribution
 (X represents the set of model parameters)

1. Choose X_0 an initial location in the parameter space . Set $t = 0$.

2. Repeat {

– Obtain a new sample Y from a proposal distribution $q(Y|X_t)$ that is easy to evaluate . $q(Y|X_t)$ can have almost any form.

I use a Gaussian proposal distribution. i.e., Normal distribution $N(X_t, \sigma)$

– Sample a Uniform (0, 1) random variable U .

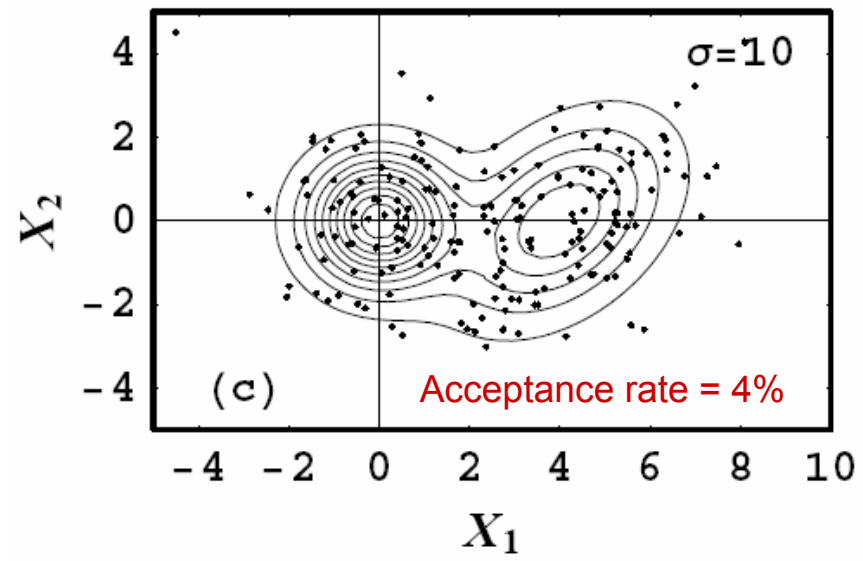
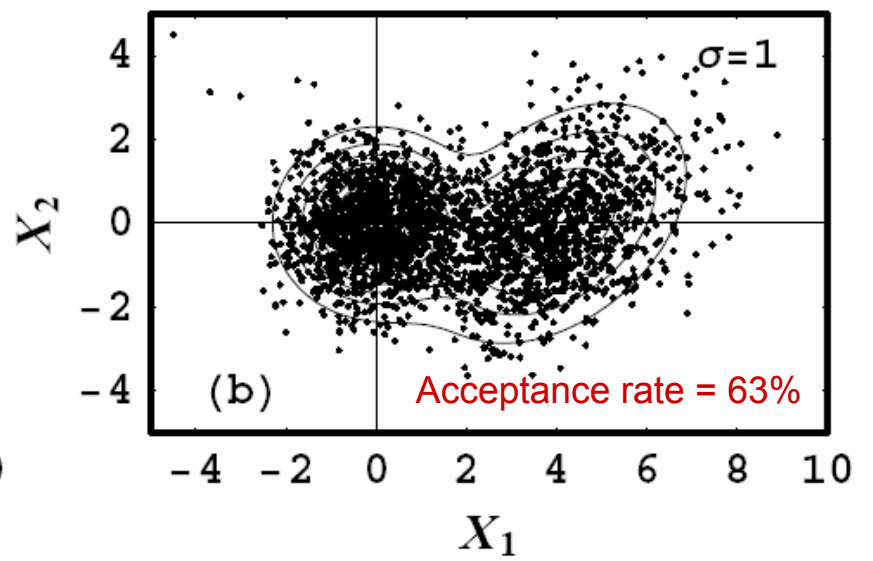
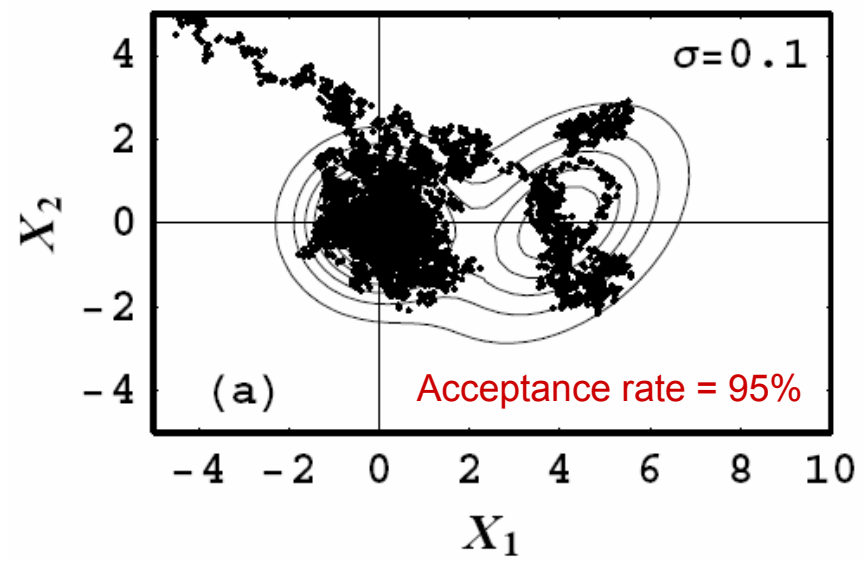
– If $U \leq \frac{p(Y|D, I)}{p(X_t|D, I)} \times \frac{q(X_t|Y)}{q(Y|X_t)}$ then set $X_{t+1} = Y$

otherwise set $X_{t+1} = X_t$

This factor = 1
 for a symmetric proposal
 distribution like a Gaussian

– Increment t }

Toy MCMC simulations: the efficiency depends on tuning proposal distribution σ 's. Can be a very difficult challenge for many parameters.



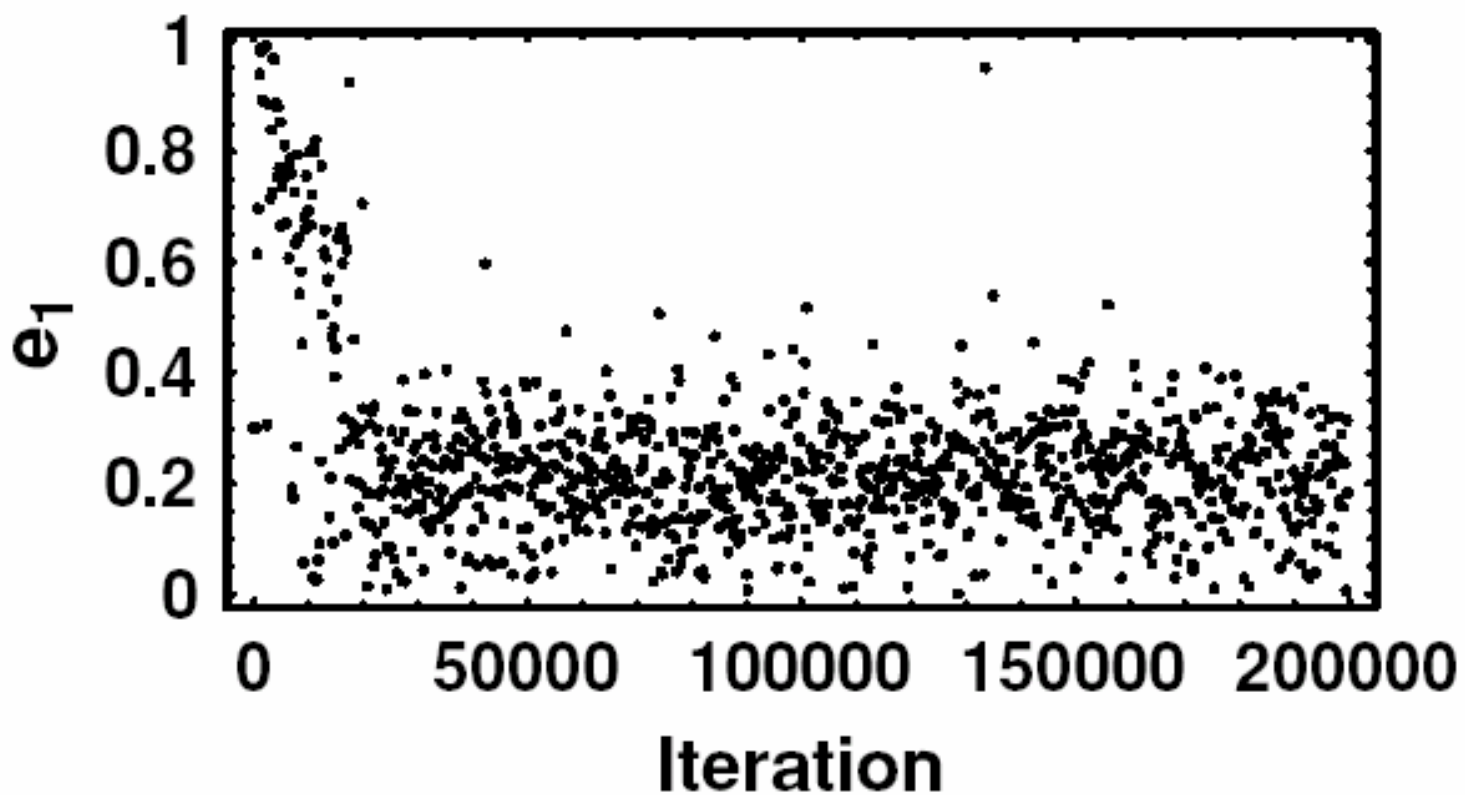
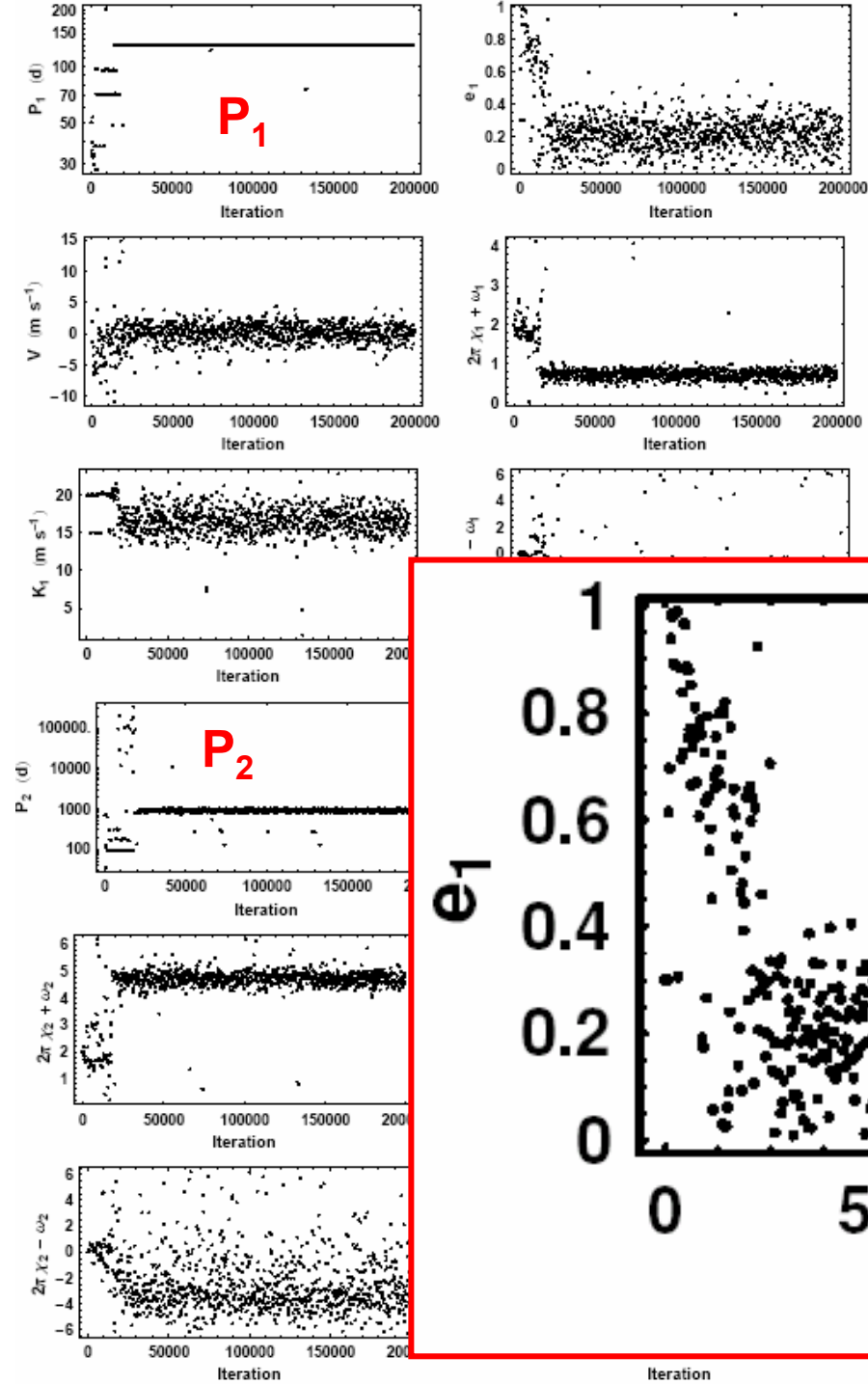
In this example the posterior probability distribution consists of two 2 dimensional Gaussians indicated by the contours

MCMC parameter samples for 2 planet model.

MNRAS, 374, 1321, 2007

P. C. Gregory

Title: **A Bayesian Kepler
Periodogram Detects a
Second Planet in HD 208487**



Parallel tempering MCMC

The simple Metropolis-Hastings MCMC algorithm can run into difficulties if the probability distribution is multi-modal with widely separated peaks. It can fail to fully explore all peaks which contain significant probability, especially if some of the peaks are very narrow.

One solution is to run multiple Metropolis-Hastings simulations in parallel, employing probability distributions of the kind

$$\pi(\mathbf{X} | \mathbf{D}, \mathbf{M}, \beta, \mathbf{I}) = p(\mathbf{X} | \mathbf{M}, \mathbf{I}) p(\mathbf{D} | \mathbf{X}, \mathbf{M}, \mathbf{I})^\beta \quad (0 < \beta \leq 1)$$

Typical set of β values = 0.09, 0.15, 0.22, 0.35, 0.48, 0.61, 0.78, 1.0
 $\beta = 1$ corresponds to our desired target distribution. The others correspond to progressively flatter probability distributions.

At intervals, a pair of adjacent simulations are chosen at random and a proposal made to swap their parameter states. The swap allows for an exchange of information across the ladder of simulations.

In the low β simulations, radically different configurations can arise, whereas at higher β , a configuration is given the chance to refine itself.

Final results are based on samples from the $\beta = 1$ simulation. Samples from the other simulations provide one way to evaluate the Bayes Factor in model selection problems.

Bayesian Logical Data Analysis for the Physical Sciences

A Comparative Approach with
Mathematica Support



CAMBRIDGE

My involvement: since 2002, ongoing development of a general Bayesian Nonlinear model fitting program.

My latest fusion Markov chain Monte Carlo (FMCMC) nonlinear model fitting algorithm incorporates:

- Parallel tempering
- Simulated annealing
- Genetic algorithm
- Unique control system automates the tuning of the proposal distributions for efficient exploration of the parameter space even for highly correlated parameters.**

Current extra-solar planet applications:

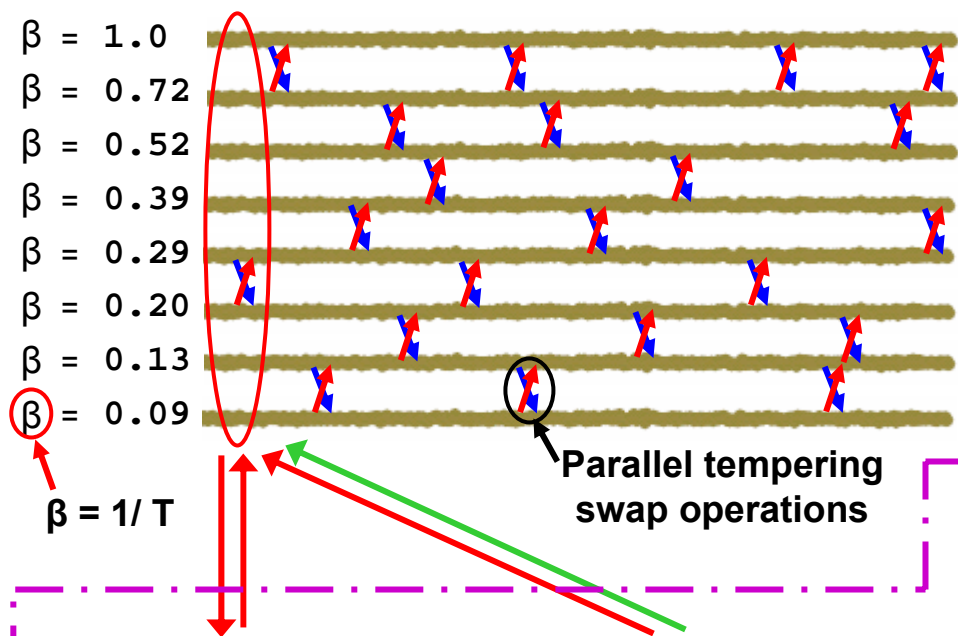
- precision radial velocity data – (4 new planets published to date)
- pulsar planets from timing residuals of NGC 6440C
- NASA stellar interferometry mission astrometry testing

Submillimeter radio spectroscopy of galactic center methanol lines

Code is implemented in *Mathematica* which provides an easy route to parallel computing. On an 8 core PC the speed-up is 7 times.

Adaptive fusion MCMC

8 parallel tempering Metropolis chains



Output at each iteration

parameters, logprior + $\beta \times$ loglike, logprior + loglike
 parameters, logprior + $\beta \times$ loglike, logprior + loglike
 parameters, logprior + $\beta \times$ loglike, logprior + loglike
 parameters, logprior + $\beta \times$ loglike, logprior + loglike
 parameters, logprior + $\beta \times$ loglike, logprior + loglike
 parameters, logprior + $\beta \times$ loglike, logprior + loglike
 parameters, logprior + $\beta \times$ loglike, logprior + loglike
 parameters, logprior + $\beta \times$ loglike, logprior + loglike

Anneal Gaussian proposal σ 's

Refine & update Gaussian proposal σ 's

2 stage proposal σ control system
 error signal =
 (actual joint acceptance rate - 0.25)
 Effectively defines burn-in interval

Peak parameter set:
 If (logprior + loglike) > previous best by a threshold then update and reset burn-in

Monitor for parameters with peak probability

Genetic algorithm
 Every 10th iteration perform gene crossover operation to breed larger (logprior + loglike) parameter set.

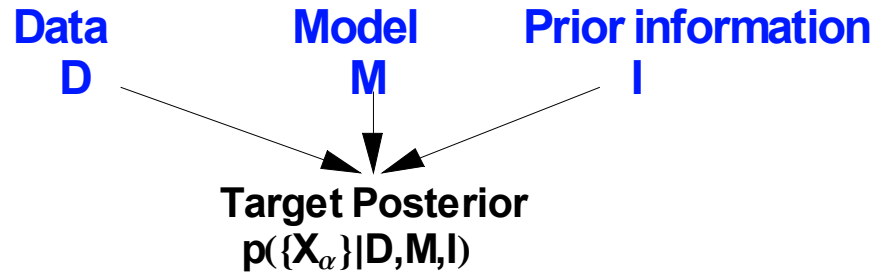
MCMC adaptive control system

Blind searches with fusion MCMC

Parallel tempering
Simulated annealing
Genetic algorithm

Each of these methods was designed to facilitate the detection of a global minimum in χ^2 . By combining all three in a fusion MCMC we greatly increase the probability of realizing this goal.

Experience with the fusion MCMC to date indicates that a blind search for at least three planets (17 parameters) is very practical in typical sparse RV data sets.



If you input a Kepler model the fusion MCMC becomes

A Kepler periodogram

Optimum for finding all Kepler orbits and evaluating their probabilities.

Capable of simultaneously fitting multiple planet models.

A multi-planet Kepler periodogram

Model space considered

Symbol	Model	# of parameters
M_{0s}	Constant velocity V + extra noise term s	$1+s$
M_{1s}	V + elliptical orbit + extra noise term s	$6+s$
M_{2s}	V + 2 elliptical orbits + extra noise term s	$11+s$
M_{3s}	V + 3 elliptical orbits + extra noise term s	$16+s$
M_{js}	V + j elliptical orbits + extra noise term s	$5j+1+s$

An extra noise term allows for an intrinsic stellar variability (“jitter”) that we model as an additional source of uncorrelated Gaussian noise with variance s^2 and add to the measurement uncertainties in quadrature. s becomes an additional parameter to marginalize over.

Note: some forms of stellar jitter (e.g., star spots) can produce Keplerian-like radial velocity variations.

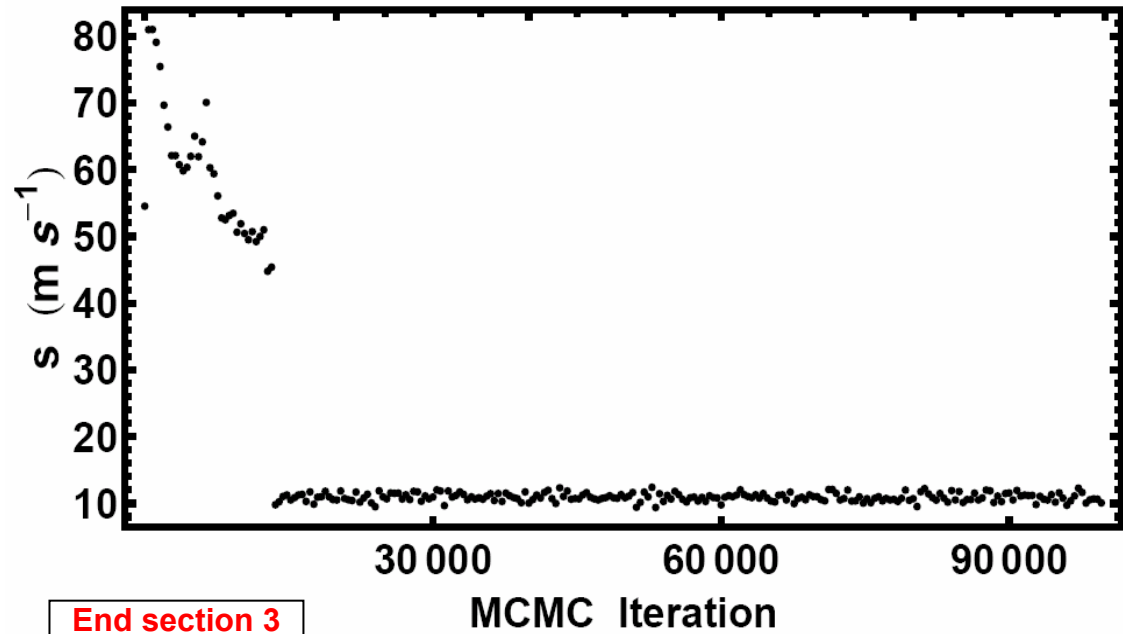
Annealing due to extra noise term, s

The inclusion of an extra noise term of unknown magnitude also gives rise to an annealing operation when the Markov chain is started far from the best-fit values.

If only known observational errors are included, the posterior probability distribution is often very “rough” with many local maxima throughout parameter space.

When s is included, 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 posterior surface and allows the Markov chain to explore the parameter space more quickly. The chain begins to decrease the value of s as it settles in near the best-fit parameters. **This behavior is similar to simulated annealing, but does not require choosing a cooling scheme.**



47 Ursae Majoris is a solar twin ~46 light years away in Ursa Major

History

- 1) 1996, report of a $P = 1090$ day companion

Butler, R. P. & Marcy, G. W. 1996, ApJ, 464, L153

- 2) 2002, report of a 2nd companion,
 $P = 2594 \pm 90$ days

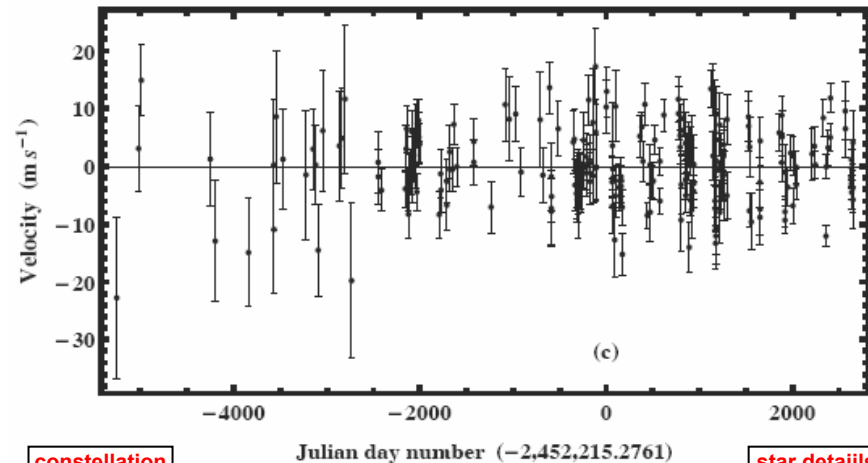
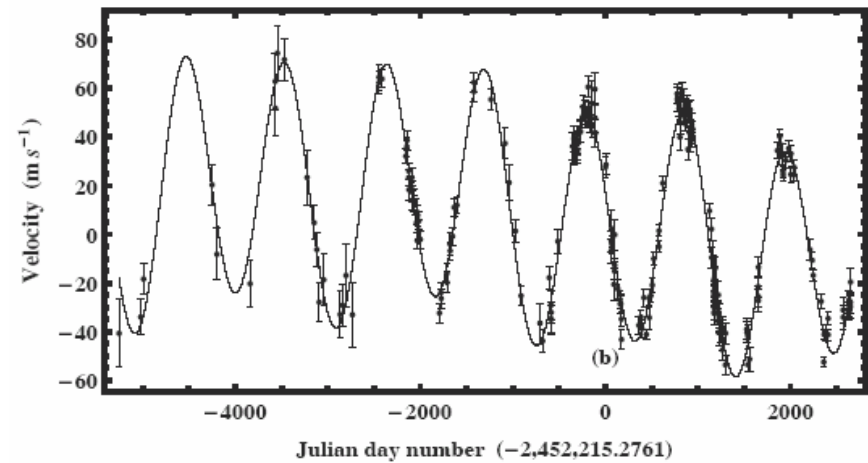
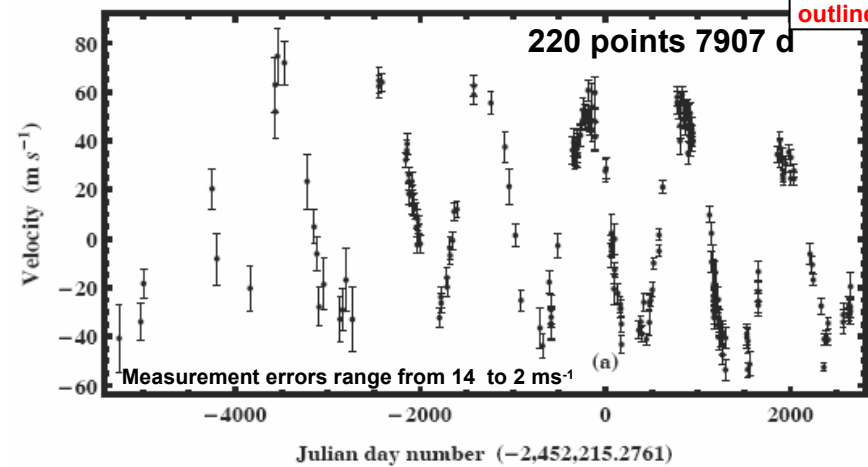
Fischer, D. A., Marcy, G. W., Butler, R. P., Laughlin, G. L., and Vogt, S. S., 2002, ApJ, 564, 1028

- 3) 2004-2009, several papers report either no 2nd planet or a 2nd planet with $P = 9660$ days when ecc. of 2nd planet set = 0.005, value found by Fischer et al., 2002.

eg., Wittenmyer, R. A., Endl, M., Cochran, W. D., Levison, H. F., Henry, G. W., 2009, ApJS, 182, 97

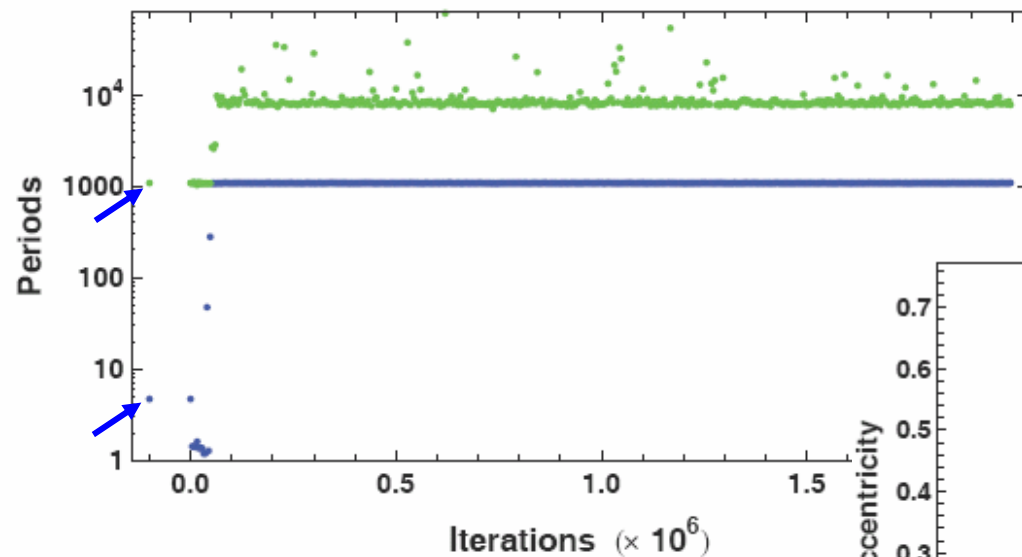
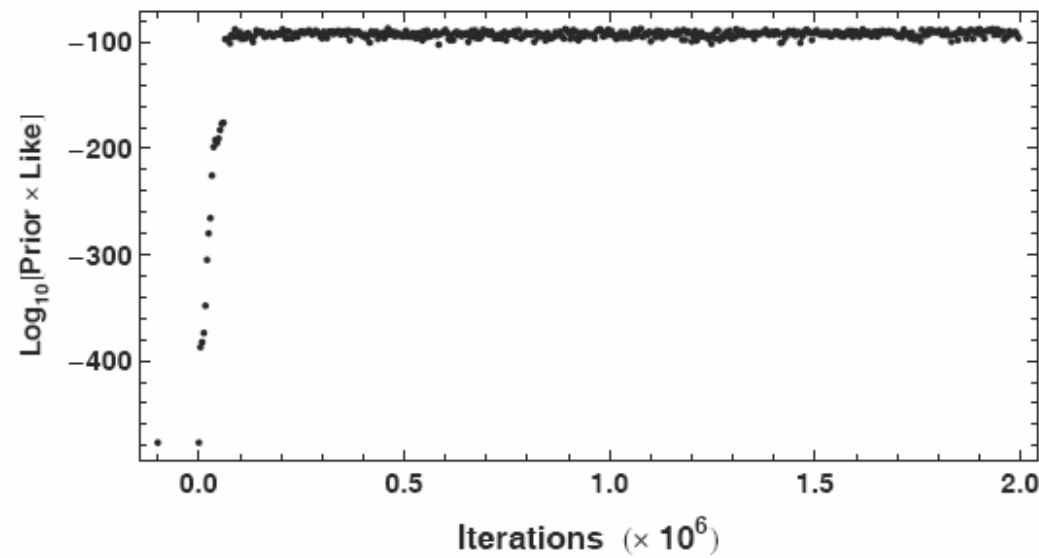
- 4) 2010, Gregory, P. C., & Fischer, D. A.,
MNRAS, 403, 731, 2010

FMCMC confirms 2400 d planet and finds evidence for a 3rd planet of ~10000 d.



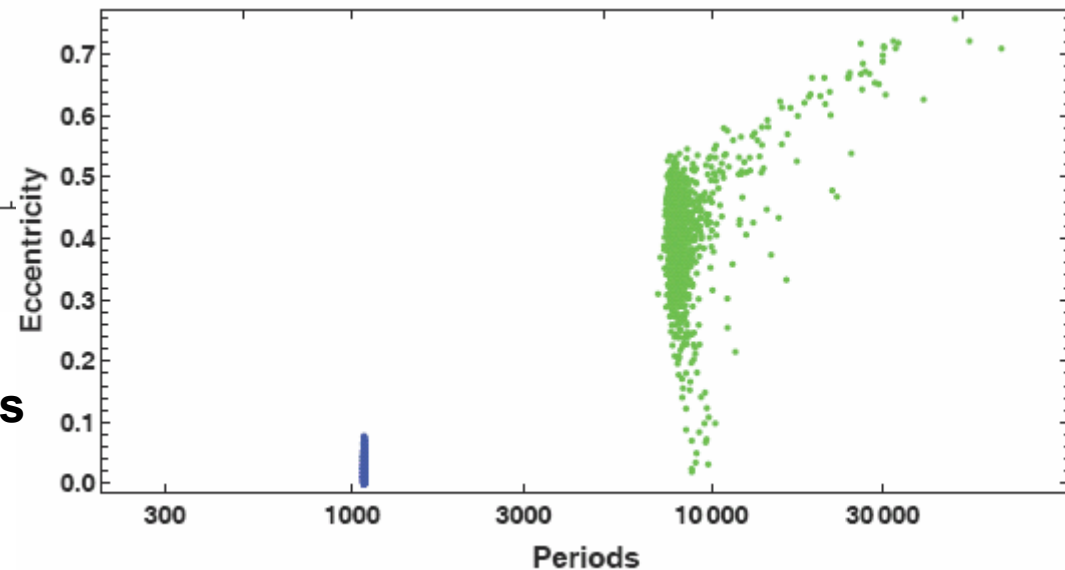
47 Ursae Majoris

2 planet FCMC model fit

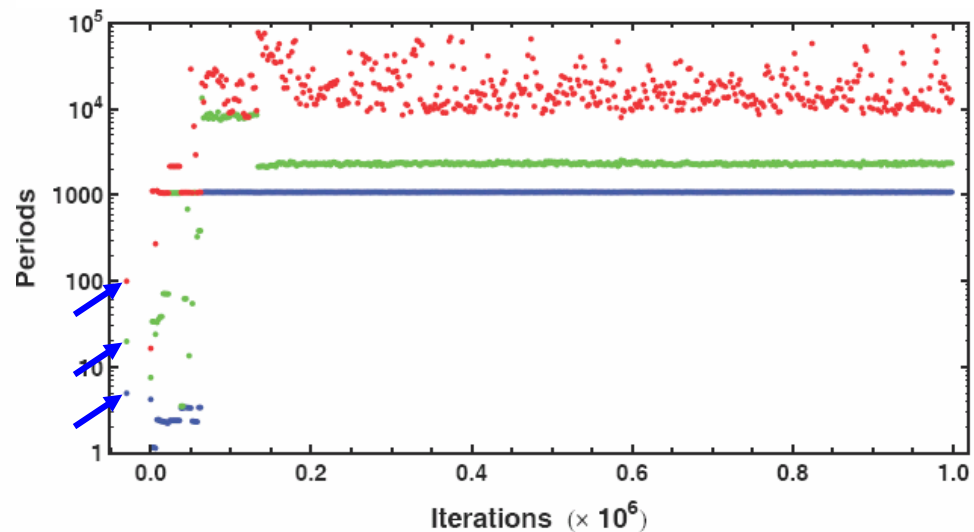
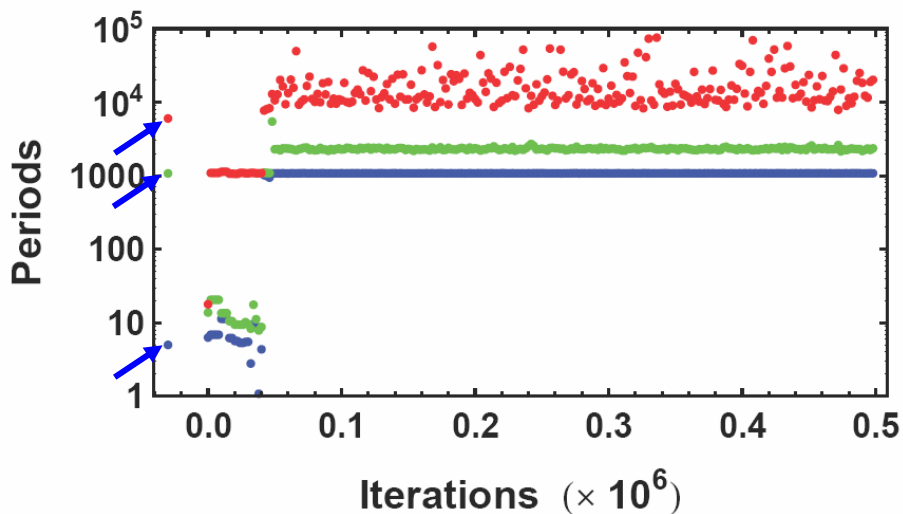
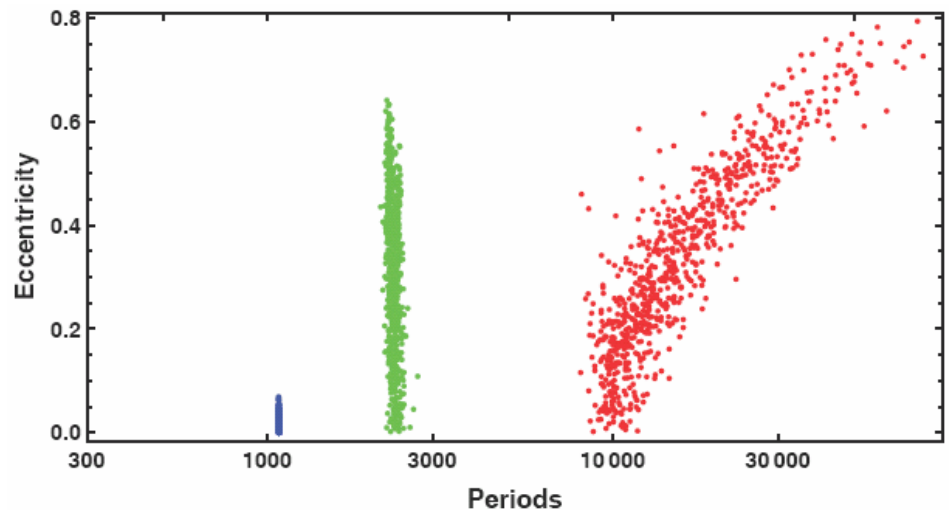
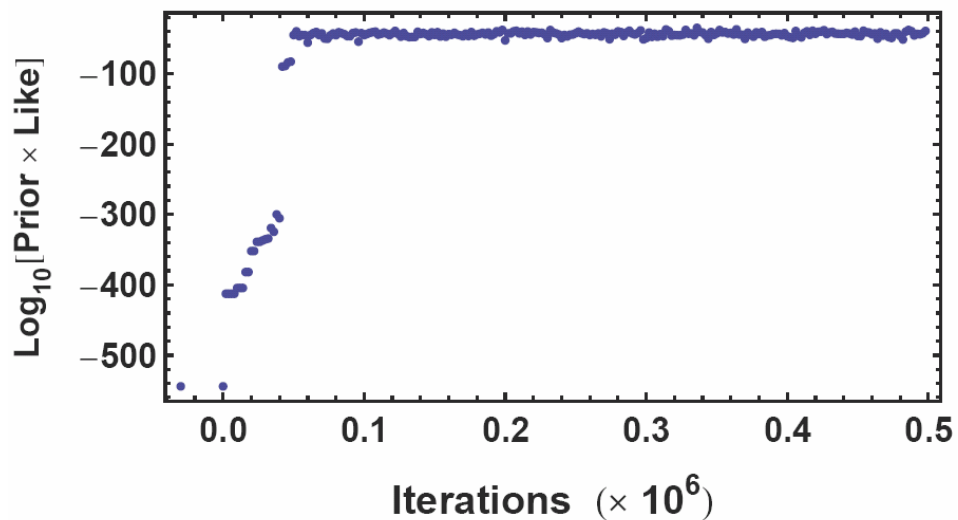


For two planet model we do not detect 2400 d period!

Arrow  indicates starting periods



47 Ursae Majoris 3 planet FCMC model fit, Lick telescope data



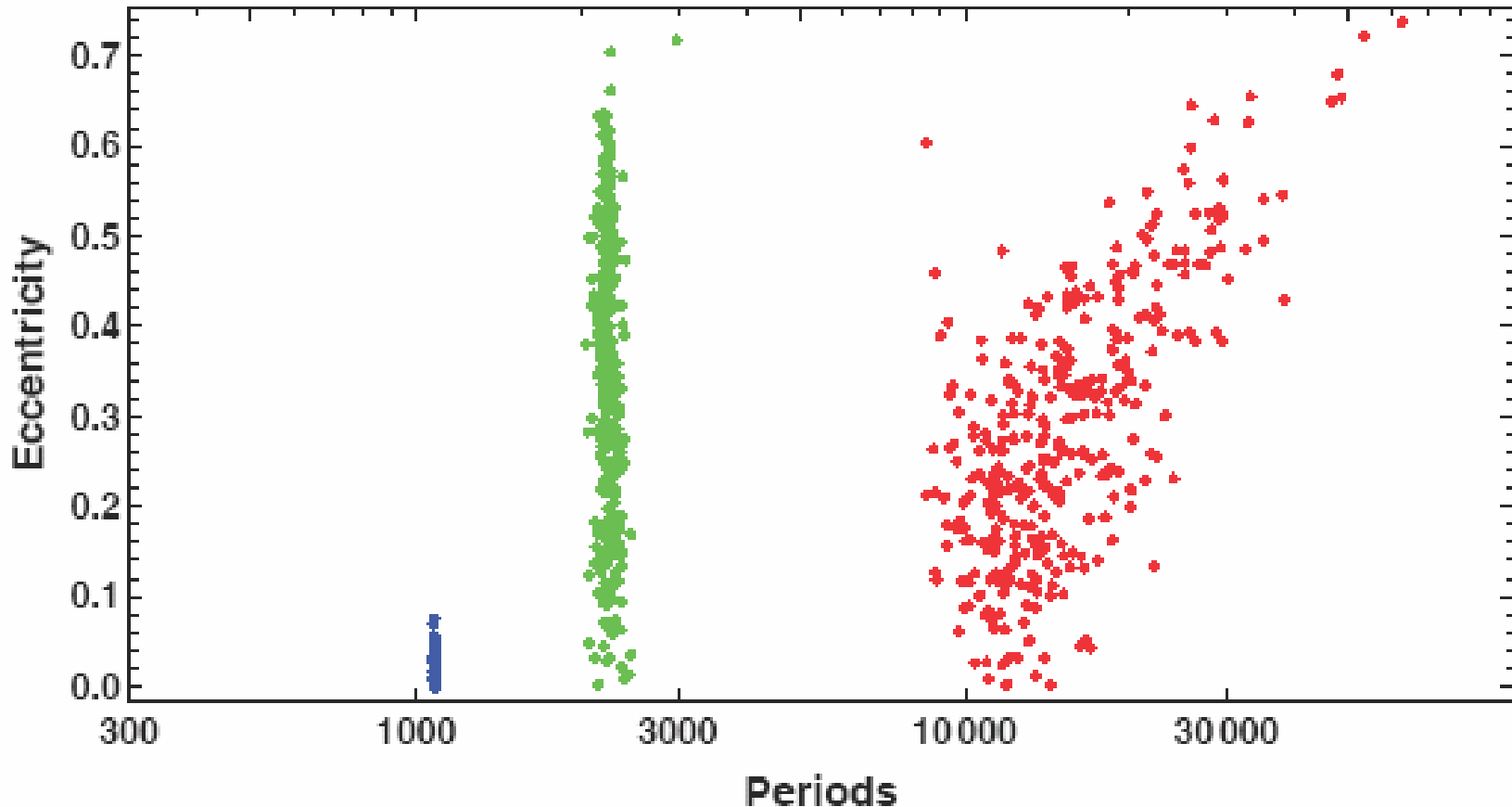
Arrow  indicates starting periods

Allowance for systematic residual offset velocities associated with detector dewar changes

JD-2440000	RV m s ⁻¹	Δ RV m s ⁻¹	Dewar	JD-2440000	RV m s ⁻¹	Δ RV m s ⁻¹	Dewar	JD-2440000	RV m s ⁻¹	Δ RV m s ⁻¹	Dewar
6959.7372	-40.70	14.00	1	11607.9163	-17.77	4.51	18	12722.8295	-20.88	3.13	24
7194.9122	-33.96	7.49	6	11626.7707	-34.76	6.65	18	12737.7703	-10.01	2.47	24
7223.7982	-18.31	6.14	6	11627.7539	-29.07	5.87	18	12793.7298	1.53	2.41	24
7964.8927	20.40	8.19	6	11628.7275	-34.86	5.71	18	12794.7134	-5.06	2.20	24
8017.7302	-8.18	10.57	6	11629.8320	-32.06	4.48	18	12834.6981	21.08	2.83	24
8374.7707	-20.25	9.37	6	11700.6937	-2.83	4.80	18	12991.0537	57.90	3.94	24
8647.8971	62.95	11.41	8	11861.0498	36.20	5.53	18	12992.0732	55.57	4.69	24
8648.9100	51.93	11.02	8	11874.0684	39.39	5.34	18	13009.0525	53.57	2.70	24
8670.8777	74.56	11.45	8	11881.0443	32.79	4.41	18	13009.9546	51.65	2.88	24
8745.6907	71.89	8.76	8	11895.0663	33.89	4.28	18	13018.9971	55.32	4.48	24
8992.0612	23.42	11.21	8	11906.0148	34.69	3.91	18	13020.9531	39.96	5.42	24
9067.7708	4.86	7.00	8	11907.0112	37.74	4.24	18	13022.0027	46.17	5.15	24
9096.7339	-6.19	6.79	8	11909.0420	39.07	3.76	18	13044.9198	58.89	3.33	24
9122.6909	-27.90	7.91	8	11910.9537	36.96	4.13	18	13068.8447	54.81	5.38	24
9172.6855	-18.68	10.55	8	11914.0674	34.35	5.17	18	13069.8323	48.36	3.34	24
9349.9122	-32.93	9.52	8	11915.0473	41.14	3.72	18	13072.8875	45.63	2.93	24
9374.9638	-29.14	8.67	8	11916.0335	40.99	3.47	18	13078.8069	52.75	3.30	24
9411.8387	-16.88	12.81	8	11939.9703	42.47	4.72	18	13079.8275	52.69	3.18	24
9481.7197	-33.01	13.40	8	11946.9598	42.21	4.19	18	13080.7919	52.88	3.27	24
9767.9184	64.68	5.34	39	11969.9024	48.36	4.29	18	13081.8171	48.72	2.99	24
9768.9072	62.32	4.79	39	11971.8934	52.56	4.80	18	13100.8148	53.34	3.83	24
9802.7911	63.99	3.61	39	11998.7785	49.07	3.81	18	13107.7773	35.01	4.43	24

Introduced 4 additional nuisance parameters relating the residual velocity offsets of dewars 6, 8, 39, 18 relative to 24. Each assumed to have a Gaussian prior with $\sigma = 3$ m s⁻¹.

47 Ursae Majoris 3 planet HMCMC model fit incorporating 4 additional nuisance parameters relating the residual velocity offsets of dewars 6, 8, 39, 18 relative to 24. Dewar 1 had only one measurement so not used. Data set shorter by 235 days.



Model M_3 was found to be 2×10^5 more probable than M_2

Final residual velocities: $V_6 = 0.07^{+2.7}_{-2.6}$, $V_8 = 1.7^{+3.0}_{-2.3}$,
 $V_{39} = -3.2^{+2.5}_{-2.4}$, $V_{18} = -1.1 \pm 2.0 \text{ m s}^{-1}$

47 Ursae Majoris Model selection

$$\text{Odds ratio} = \mathbf{O}_{i2} = \frac{p(M_i | D, I)}{p(M_2 | D, I)} = \frac{p(M_i | I)}{p(M_2 | I)} \times \frac{p(D | M_i, I)}{p(D | M_2, I)} = \frac{p(M_i | I)}{p(M_2 | I)} \times \mathbf{B}_{i2}$$

$$p(D | M_i, I) = \int d \vec{X} p(\vec{X} | M_i, I) \times p(D | \vec{X}, M_i, I)$$

Model	Bayes factor B_{i2}	False Alarm Probability
M_0	1.6×10^{-141}	
M_1	2.0×10^{-28}	8×10^{-114}
M_2	1.0	2×10^{-28}
M_3	2.0×10^5	5×10^{-6}

False Alarm Probability (FAP): in the context of claiming the detection of a 3 planet model the FAP is the probability that there are actually 2 or less planets.

$$\text{FAP} = \sum_{i=0}^2 (\text{prob. of } i \text{ planets}) \quad p(M_i | D, I) = \frac{B_{i2}}{\sum_{j=0}^{N_{\text{mod}}} B_{j2}}$$

$$\text{FAP} = \frac{(B_{02} + B_{12} + B_{22})}{\sum_{j=0}^3 B_{j2}} = 5 \times 10^{-6}$$

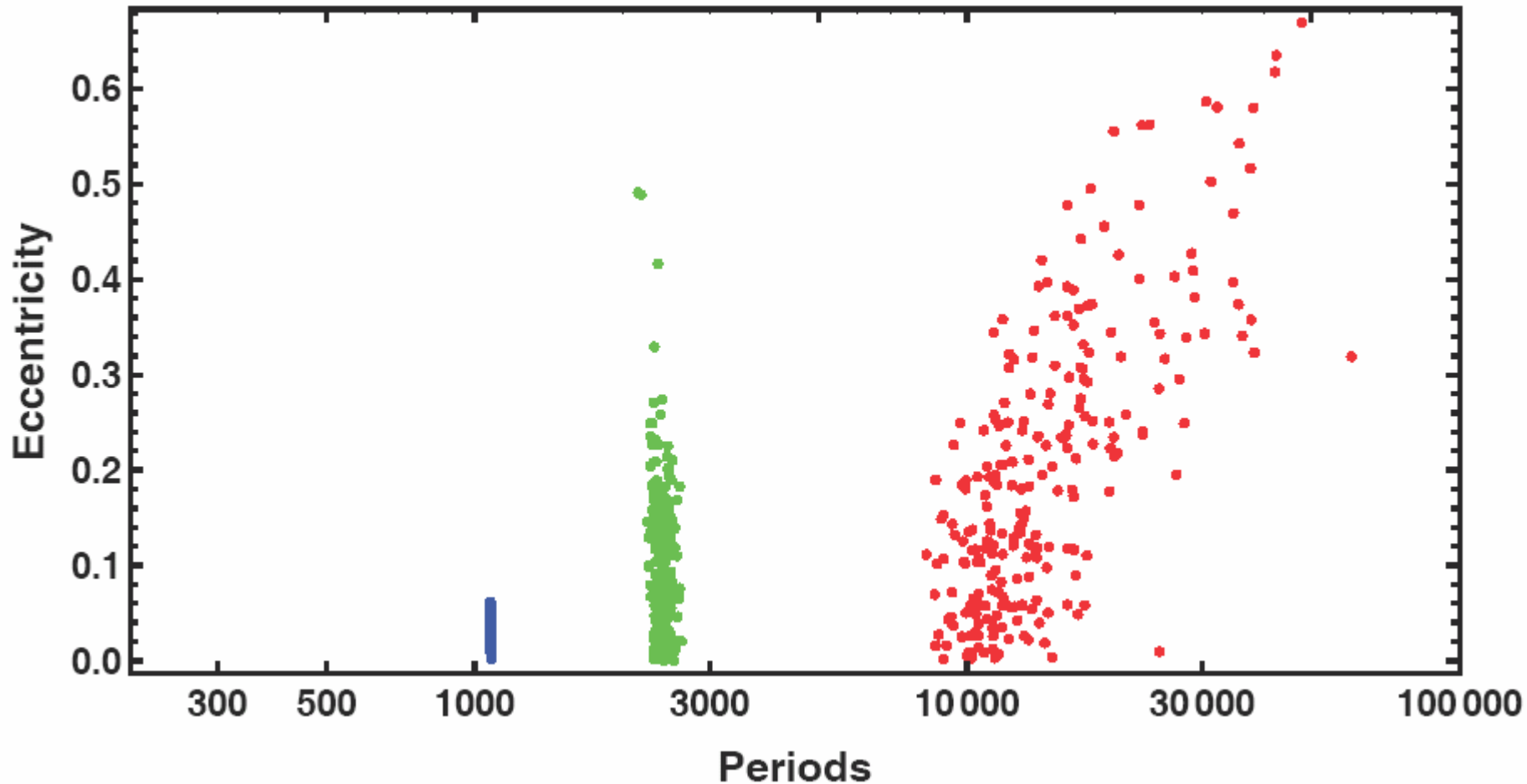
47 Ursae Majoris

3 planet FMCMC model fit

Lick + Hobby-Eberly + Harlem J Smith telescope data.

Included 2 additional residual offset velocity parameters.

Total of 413 velocity measurements.

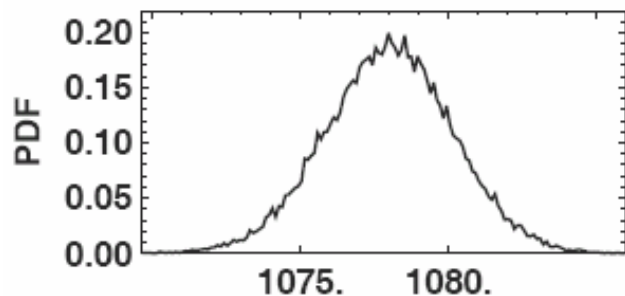


Reference for Hobby-Eberly telescope and Harlam J. Smith telescope data

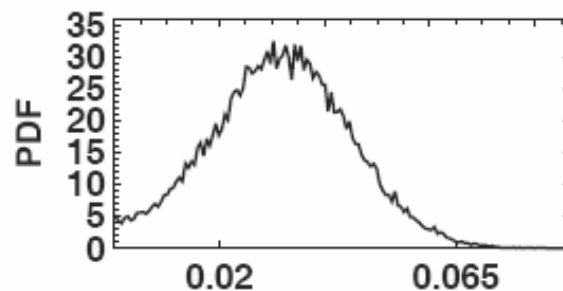
Wittenmyer, R. A., Endl, M., Cochran, W. D., Levison, H.F., Henry, G. W., 2009, ApJS, 182, 97

47 Ursae Majoris

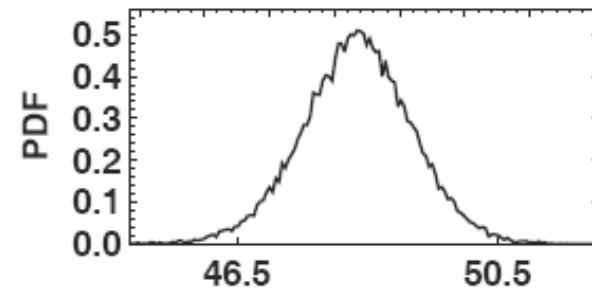
Parameter marginal distributions for a 3 planet HMCMC of the combined Lick, HET, and HJS telescope data set.



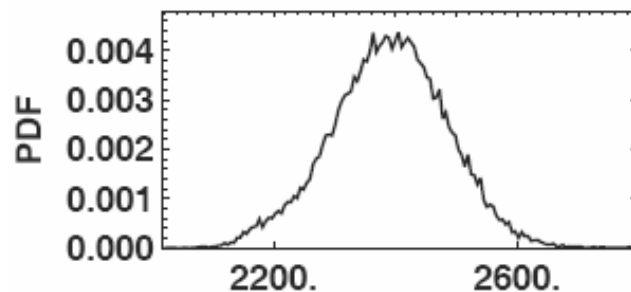
P_1 (d)



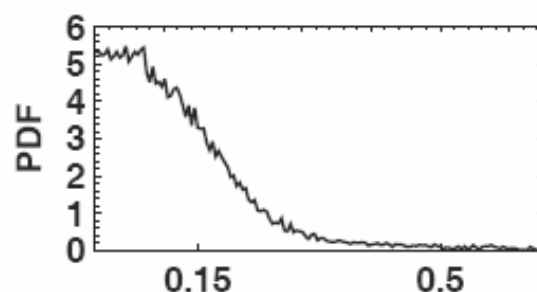
e_1



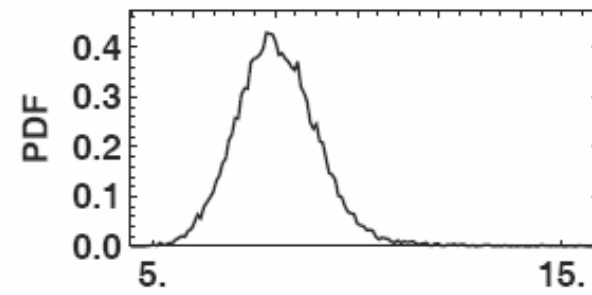
K_1 (m s^{-1})



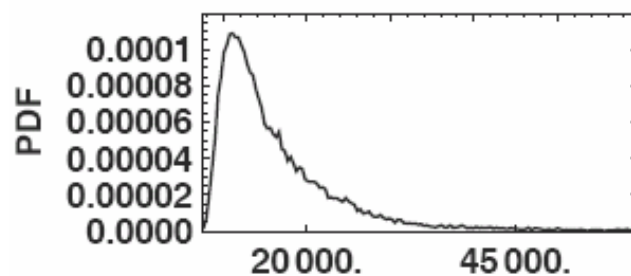
P_2 (d)



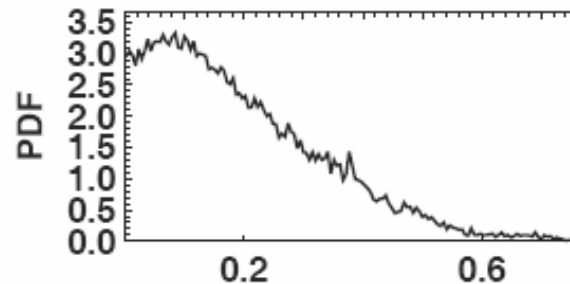
e_2



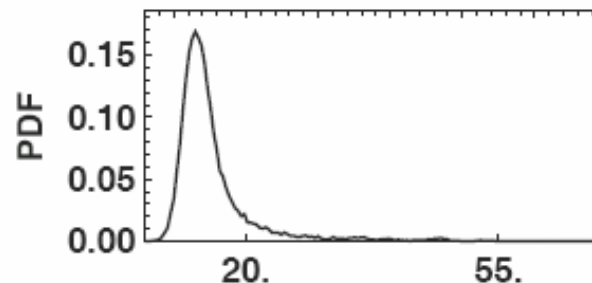
K_2 (m s^{-1})



P_3 (d)



e_3



K_3 (m s^{-1})

Gliese 581 the star with two possible habitable zone planets

History

- 1) 2005 to 2009,
Planet e is 1.9 Earth mass
Planet b is 16 Earth mass
Planet c is 5 Earth mass
Planet d is 7 Earth mass

Latest paper:

M. Mayor et al., *A&A*, 507, p. 487, Nov. 2009

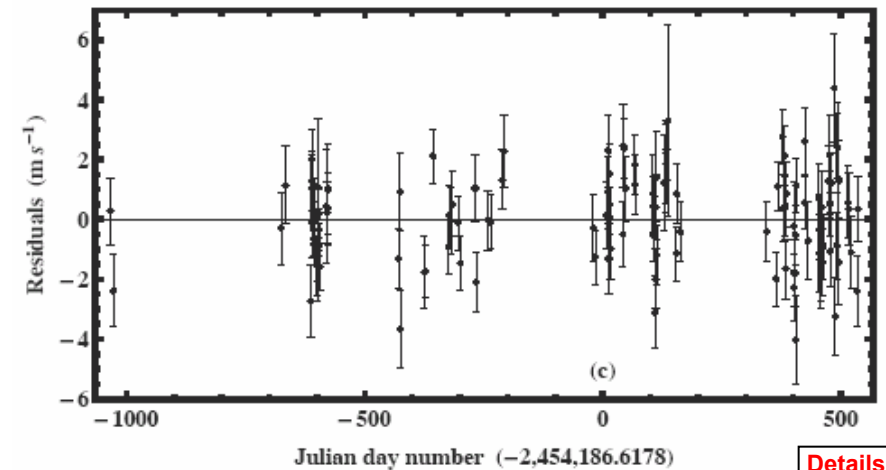
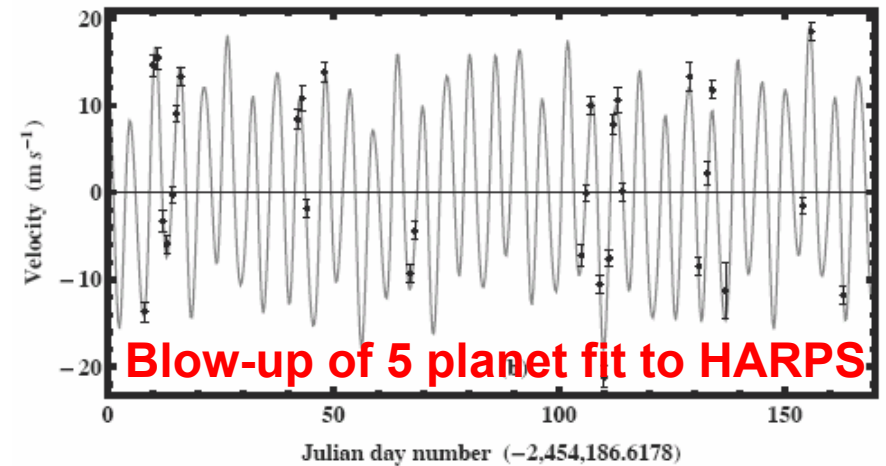
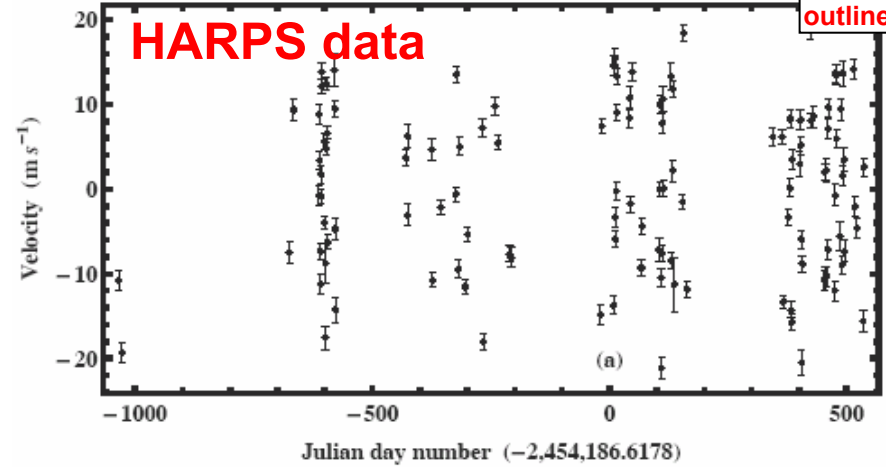
- 2) 2010,
Planet f is 7 Earth mass
Planet g is 3.1 Earth mass

Steven Vogt et al., *ApJ*, 723, p. 954, 2011,
their analysis assumes all circular orbits.

- 3) 2011, Gregory, P. C.,
<http://arxiv.org/abs/1101.0800>

Don't support claim for planet 581g.

Find evidence that HIRES uncertainties are much larger than the quoted values, an extra 1.8 m s^{-1} added in quadrature.



Conclusions

1. A Bayesian analysis of the latest radial velocity data sets for 47 Ursae Majoris confirms the existence of a 2nd planet with $M \sin i = 0.54 M_J$ and $P = 2400\text{d}$, and finds evidence for a 3rd planet with a $P \sim 10000\text{ d}$ and $M \sin i \sim 1.6 M_J$.
2. The analysis yields the full marginal posterior probability density function (PDF) for each model parameter, which allows for rigorous error estimates even for a planet with $P >$ data duration.
3. This built in Occam's razor allows for a direct comparison of the probabilities of models with different numbers of planets.
4. This Bayesian approach is helping to raise the bar on how much useful information can be extracted from the radial velocity data.
5. Experience gained while working with the exoplanet data sets has led to the development of an "Adaptive fusion MCMC" nonlinear model fitting algorithm, which can be applied to many other problems.

For copies of my papers please Google Phil Gregory

The rewards of data analysis:

**‘The universe is full of magical things,
patiently waiting for our wits to grow
sharper.’**

Eden Philpotts (1862-1960)

Author and playwright

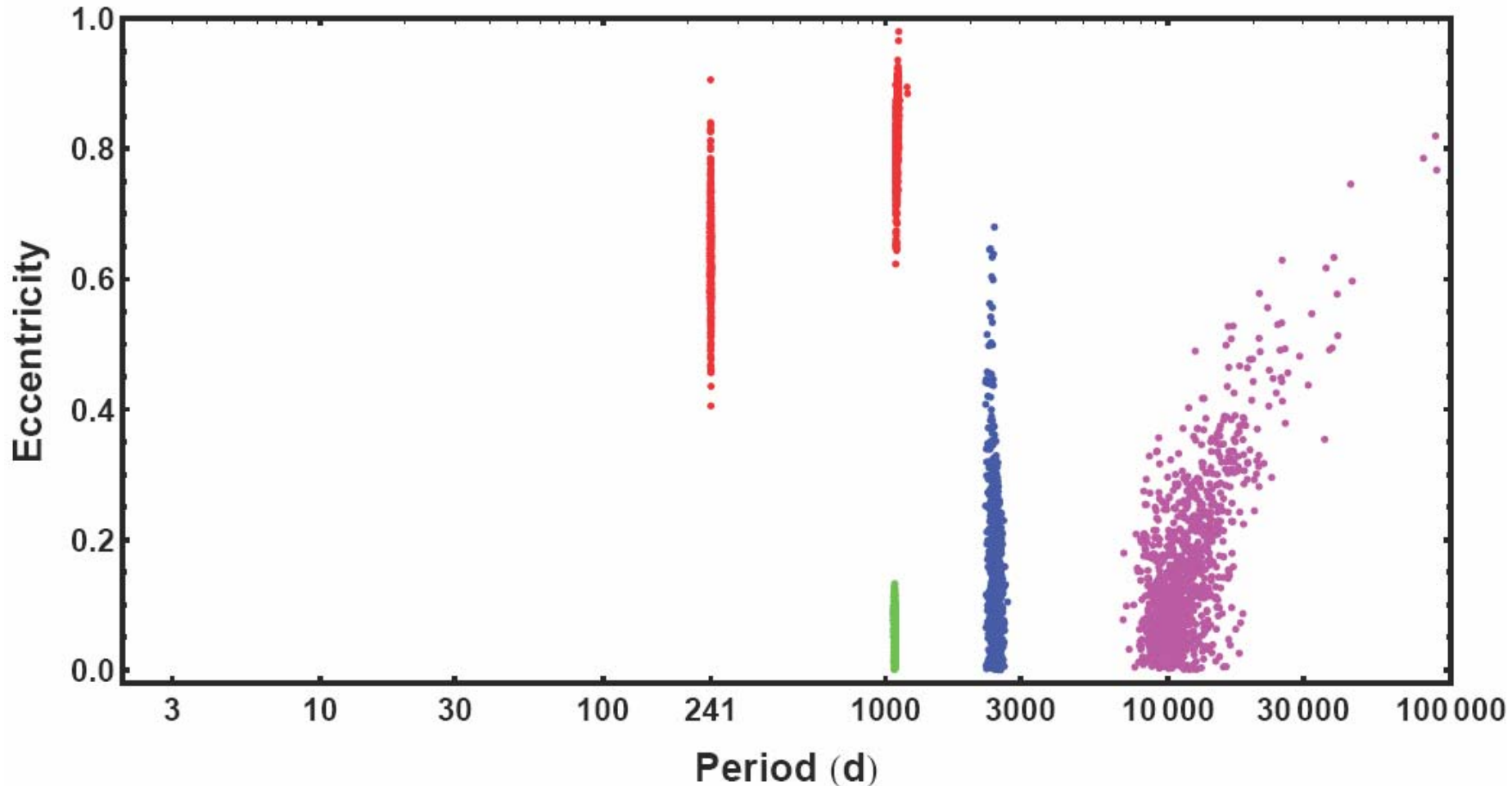
47 Ursae Majoris

4 planet HMCMC model fit

Lick + Hobby-Eberly + Harlem J Smith telescope data.

Included 2 additional residual offset velocity parameters.

Total of 413 velocity measurements.



Reference for Hobby-Eberly telescope and Harlam J. Smith telescope data

Wittenmyer, R. A., Endl, M., Cochran, W. D., Levison, H.F., Henry, G. W., 2009, ApJS, 182, 97

Parameter priors

Parameter	Prior	Lower bound	Upper bound
Orbital frequency	$p(\ln f_1, \ln f_2, \dots, \ln f_n M_n, I) = \frac{n!}{[\ln(f_H/f_L)]^n}$ (n =number of planets)	1/1.5 d	1/1000 yr
Velocity K_i (m s^{-1})	Modified Jeffreys ^a $\frac{(K+K_0)^{-1}}{\ln \left[1 + \frac{K_{\max}}{K_0} \left(\frac{P_{\min}}{P_i} \right)^{1/3} \frac{1}{\sqrt{1-e_i^2}} \right]}$	0 ($K_0 = 1$)	$K_{\max} \left(\frac{P_{\min}}{P_i} \right)^{1/3} \frac{1}{\sqrt{1-e_i^2}}$ $K_{\max} = 2129$ K_{\max} corresponds to a max. planet-star mass ratio = 0.01
V (m s^{-1})	Uniform	$-K_{\max}$	K_{\max}
e_i Eccentricity	a) Uniform b) Ecc. noise bias correction filter	0 0	1 0.99
ω_i Longitude of periastron	Uniform	0	2π
s Extra noise (m s^{-1})	$\frac{(s+s_0)^{-1}}{\ln \left(1 + \frac{s_{\max}}{s_0} \right)}$	0 ($s_0 = 1$)	K_{\max}

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

$$p(X|M, I) = \frac{1}{X + X_0} \frac{1}{\ln \left(1 + \frac{X_{\max}}{X_0} \right)} \quad (4)$$

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