Bayesian model fitting through Controlled Statistical Fusion with Exoplanet examples

Phil Gregory

Physics and Astronomy
Univ. of British Columbia

See papers for full details at http://www.physics.ubc.ca/~gregory/gregory.html

Abstract

Many different statistical techniques have been developed to facilitate the detection of a global minimum in chi-squared in the multi-modal environment commonly encountered in nonlinear model fitting. This is particularly challenging as the number of model parameters increase necessitating a MCMC approach.

This paper describes progress in the development of a new fusion Markov chain Monte Carlo algorithm which incorporates the advantages of parallel tempering, simulated annealing and the genetic algorithm. This statistical fusion approach has been achieved through the development of a unique multi-stage adaptive control system, hence the term controlled statistical fusion. Among other things the control system automates the tuning of the proposal distributions for efficient exploration of the model parameter space even when the parameters are highly correlated.

The fusion MCMC algorithm is implemented in Mathematica using parallelized code and run on an 8 core PC.

Fusion MCMC

The latest version of FMCMC for nonlinear model fitting incorporates:

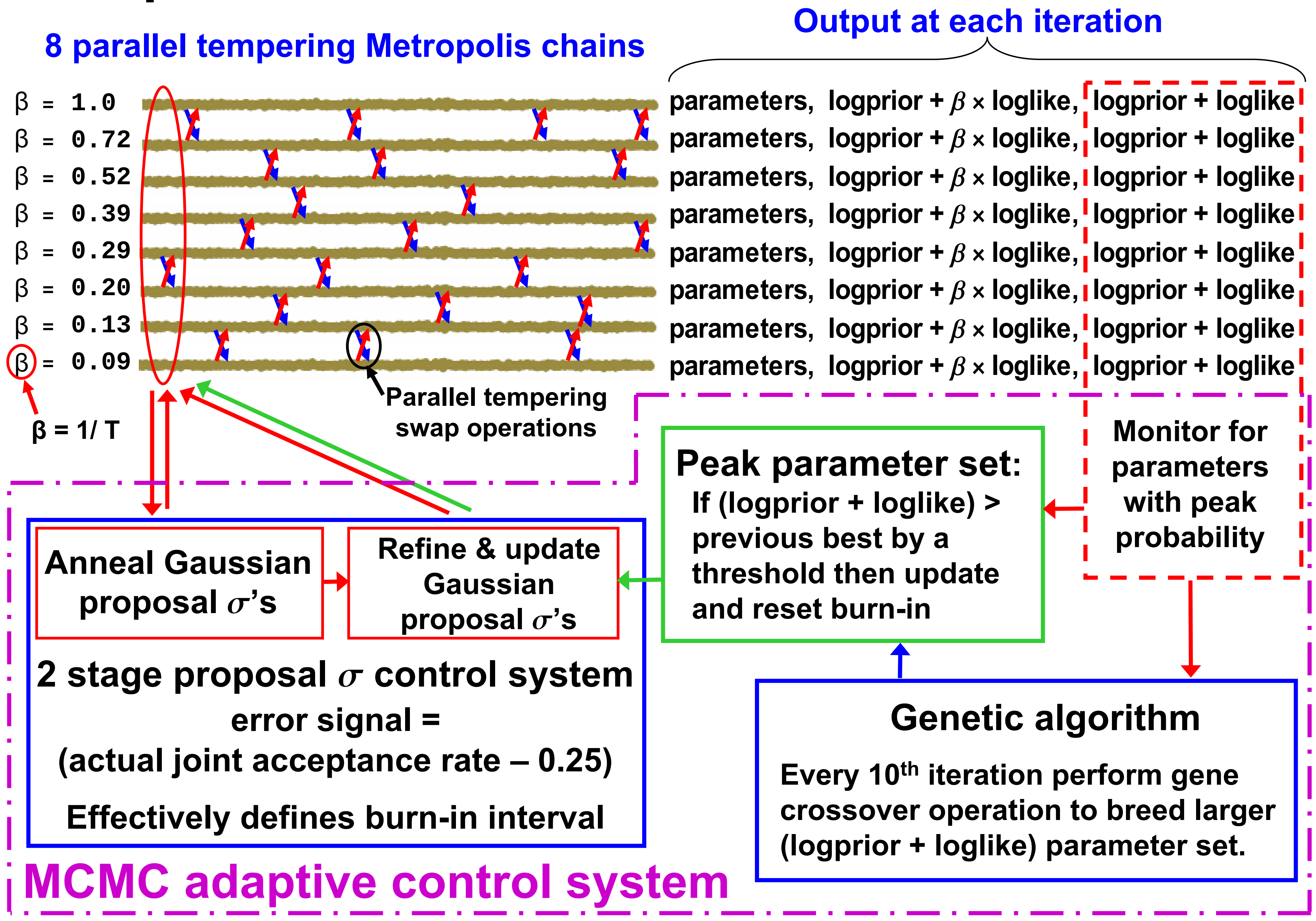
Parallel tempering
Simulated annealing
Genetic algorithm

Each of these methods was designed to facilitate the detection of a global minimum in χ^2
(in a Bayesian context: a global maximum of probability).
By combining all three we greatly increase the probability of realizing this goal.

This fusion has only been possible through the development of a unique adaptive control system. Among other things it automates the choice of an efficient set of MCMC proposal distributions even if the parameters are highly correlated.

Note: FMCMC was previously named hybrid MCMC (HMCMC) but I discovered this term already had a different meaning in the literature.

Adaptive Fusion MCMC



Data Model Prior

Bayes Theorem

$$p\left(\vec{X} \mid D, M, I\right)$$

Target posterior

If you input a Kepler model the fusion MCMC becomes

A Kepler periodogram

Optimum for finding Kepler orbits and evaluating their probabilities.

Capable of simultaneously fitting multiple planet models.

A multi-planet Kepler periodogram

The radial velocity equation, a nonlinear model

$f_i =$ model prediction $= V + K\,(\cos(\theta[t_i + \chi P] + \omega) + e\cos\omega)$

V = systematic velocity

K = velocity amplitude $= \dfrac{2\pi a \sin i}{P\sqrt{1-e^2}}$;

(a = semi – major axis, i = inclination)

e = orbital eccentricity

ω = longitude of periastron

χ = fraction of orbit prior to data reference time that periastron occurred at

$\theta(t_i + \chi P)$ = true anomaly

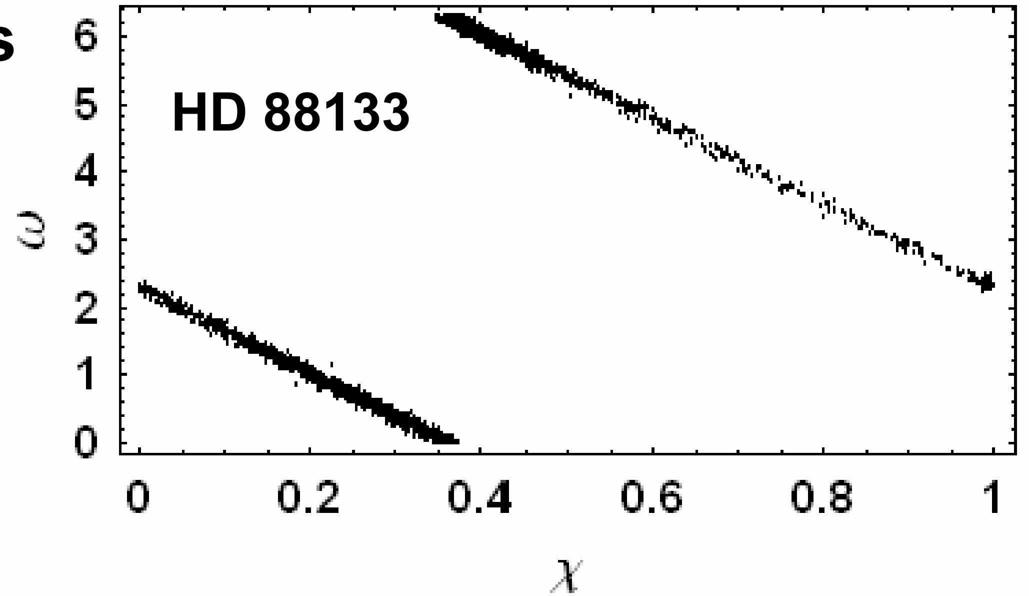
= angle of star in orbit at time t_i relative to periastron

θ_i and t_i are related by the conservation of angular momentum equation

$$\partial_t\,\theta[t] - \frac{2\pi}{P\left(1-e^2\right)^{3/2}}\,(1 + e\,\mathrm{Cos}[\theta])^2 = 0$$

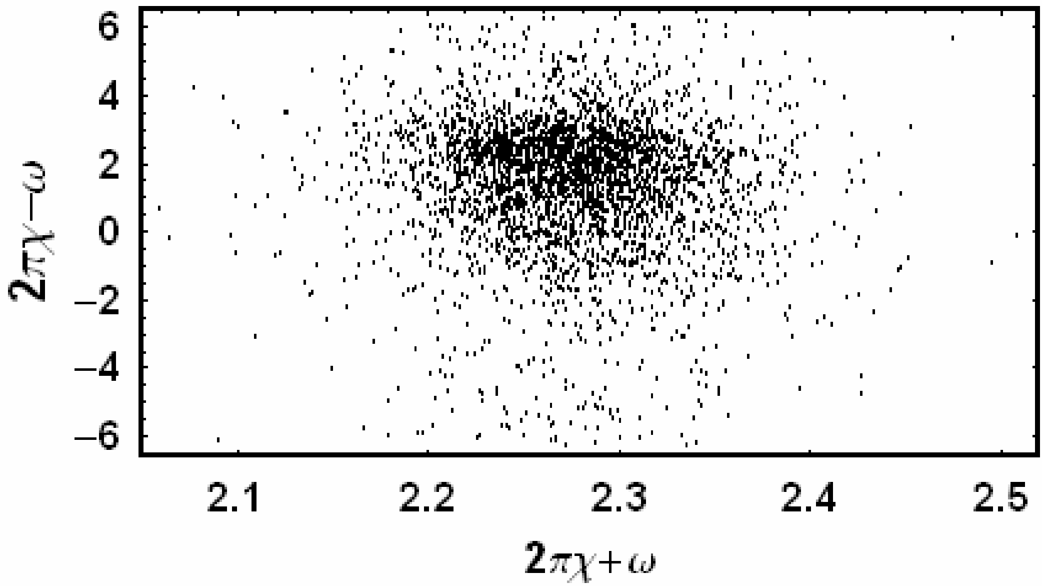
Highly correlated parameters

Exoplanet example: for low eccentricity orbits the parameters ω and χ are not separately well determined. This shows up as a strong correlation between ω and χ as shown on the right.



One option re-parameterization

The combination $2\pi\chi+\omega$ is well determined for all eccentricities. Although $2\pi\chi-\omega$ is not well determined for low eccentricities, it is at least orthogonal to $2\pi\chi+\omega$ as shown.

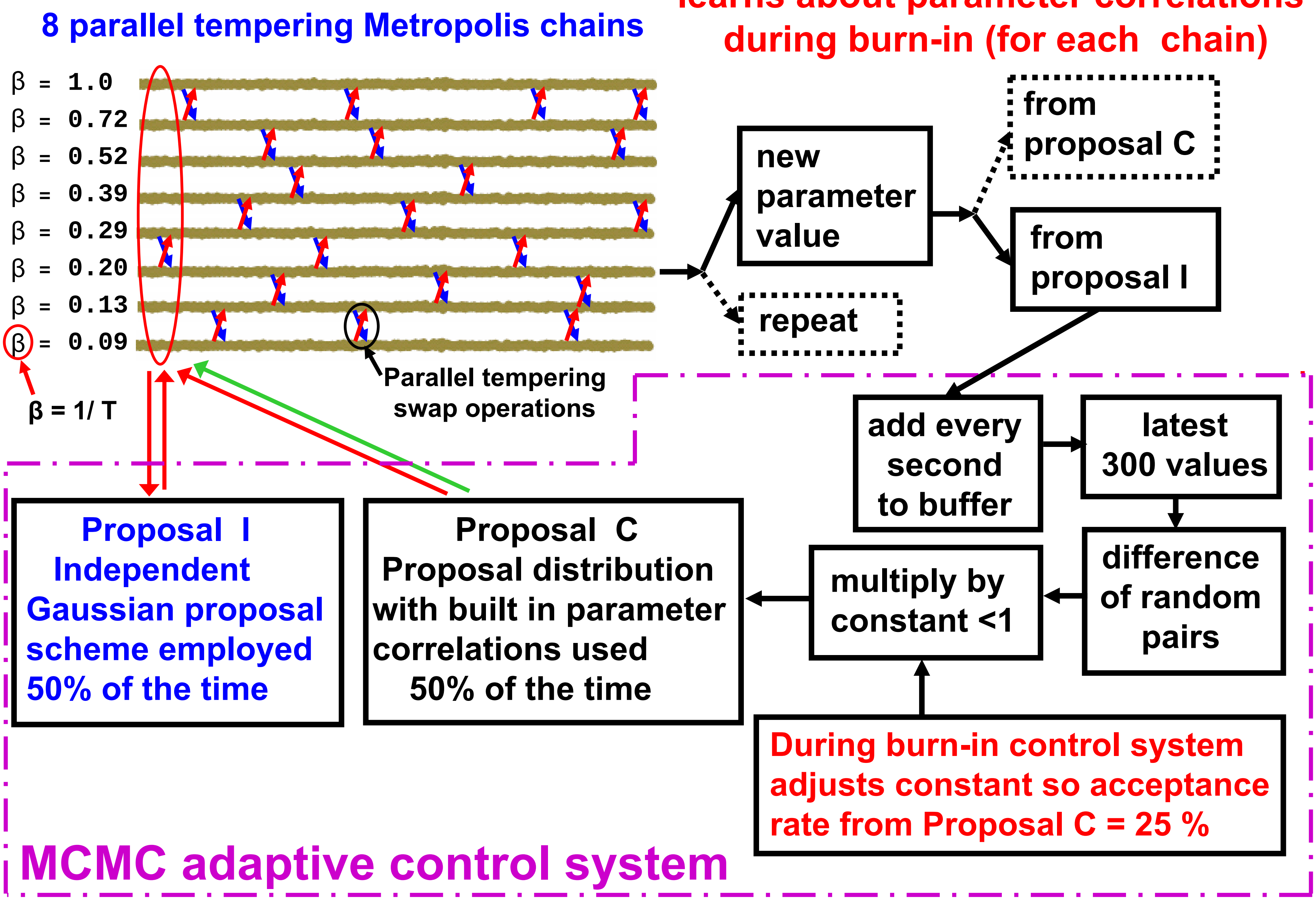


Another option

Algorithm learns about the parameter correlations during the burn-in and generates proposals with these statistical correlations.

Adaptive Fusion MCMC

Automatic proposal scheme that learns about parameter correlations during burn-in (for each chain)



How to deal with highly correlated parameters

Using only independent Gaussian proposals the ('I' scheme) the σ's need to be very small for any proposal to be accepted and consequently convergence is very slow.

Learn about parameter correlations during burn-in

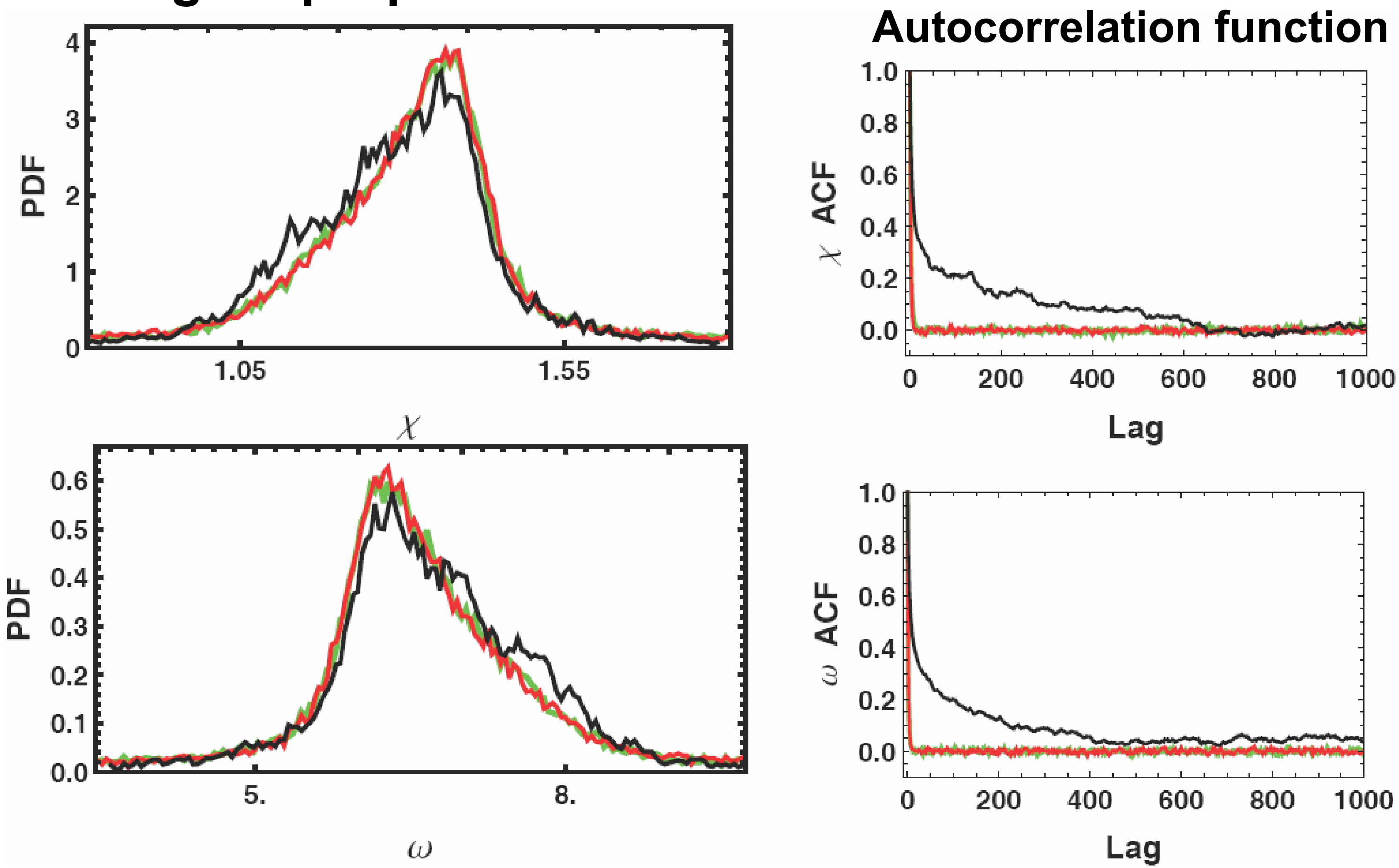
The accepted 'I' proposals will generally cluster along the correlation path so every 2nd accepted 'I' proposal is appended to a correlated sample buffer (separate buffer for each tempering level).

Only the 300 most recent additions to the buffer are retained.

A 'C' proposal is generated using the difference between a pair of randomly selected samples drawn from the correlated sample buffer (for that tempering level), after multiplication by a constant.

Value of constant is computed automatically by another control system module which ensures that the 'C' proposal acceptance rate is close to 25%.

Testing ‘C’ proposal scheme



Left panels show marginal distributions for two orbital parameters χ and ω. The black trace corresponds to a search in χ and ω using only ‘I’ proposals. Red trace corresponds to search with ‘C’ proposals turned on. Green trace, to an ‘I’ search using transformed orthogonal coordinates.

Right panels show a comparison of the MCMC autocorrelation functions.

Noise model

It is useful to incorporate an extra noise parameter, s, of unknown magnitude, added in quadrature to the known measurement uncertainties σ_i .

$$\text{variance}_i = \sigma_i^2 + s^2$$

In general, nature is more complicated than our model and known noise terms. Marginalizing *s* has the desirable effect of treating anything in the data that can't be explained by the model and known measurement errors as noise, leading to more conservative estimates of the parameters.

In the absence of detailed knowledge of the sampling distribution for the extra noise, pick an independent Gaussian model because for any given finite noise variance it is the distribution with the largest uncertainty as measured by the entropy, i.e., the maximum entropy distribution.

If there is no extra noise then the posterior probability distribution for s will peak at s = 0.

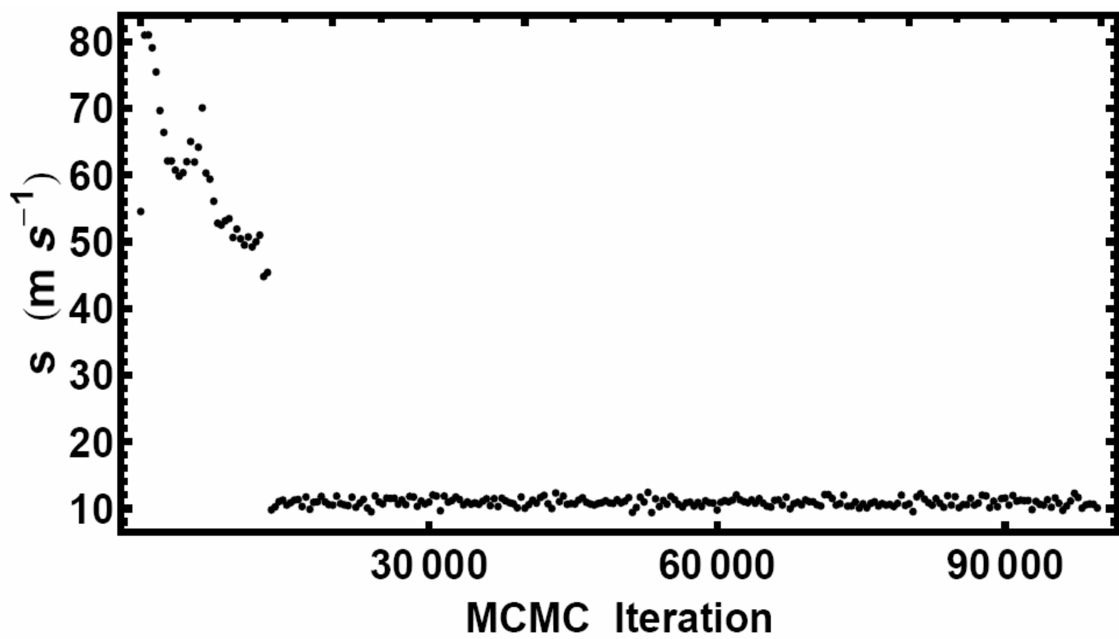
Annealing due to extra noise term, s

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±90 days

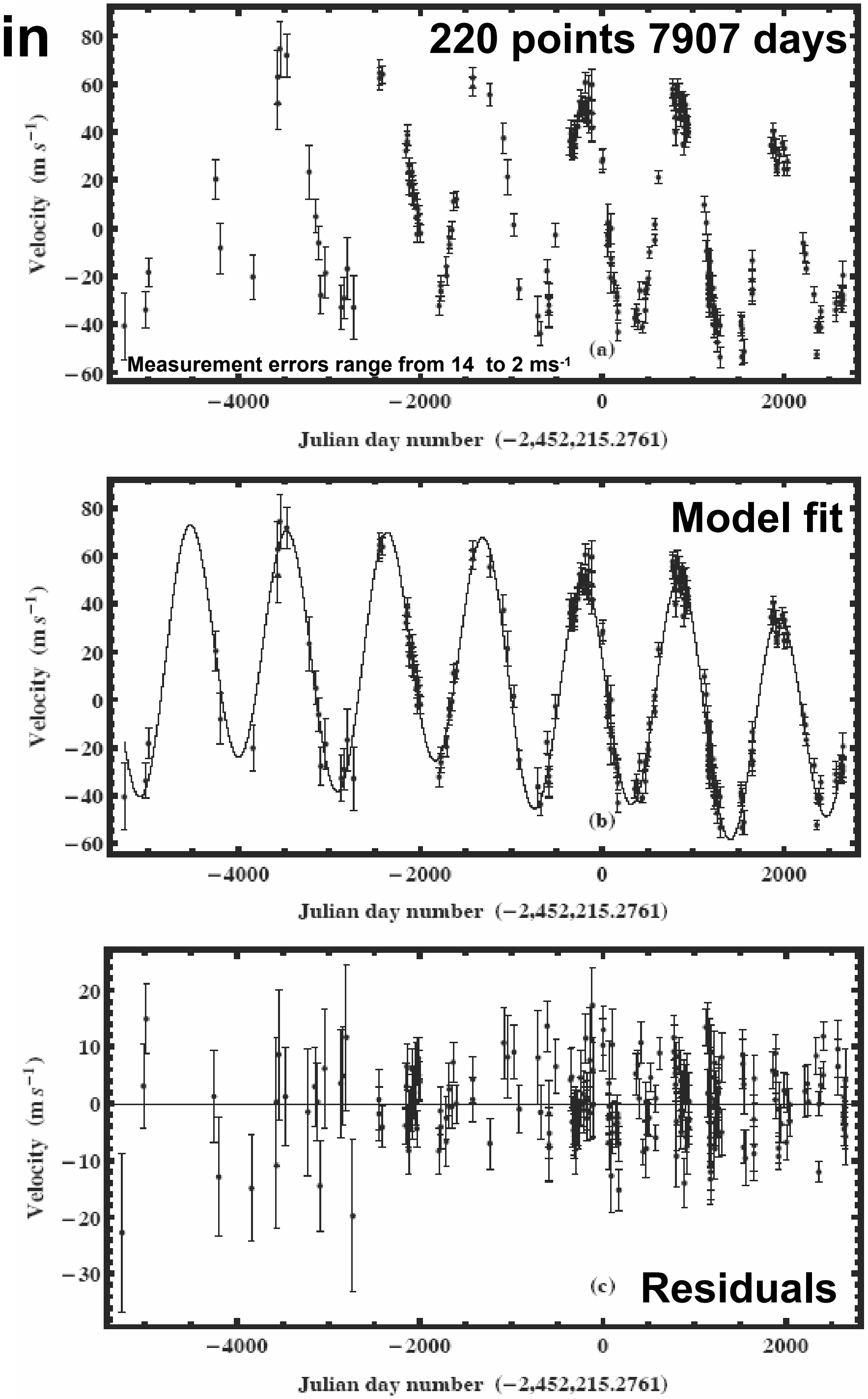
Fischer,D. A., Marcy, G. W., Butler, R. P., Laughlin, G. L., and Vogt, S. S., 2002, ApJ, 564, 1028

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.

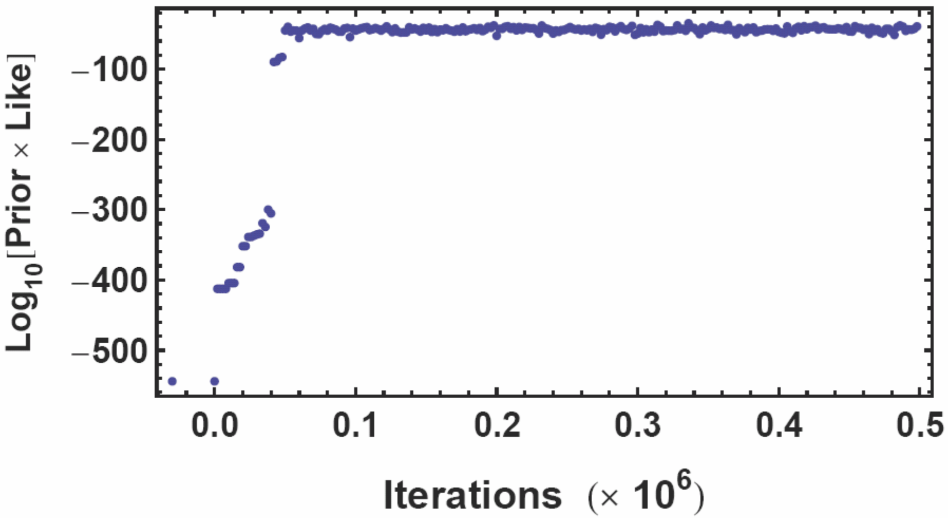
eg., Wittenmyer, R. A., Endl, M., Cochran, W. D., Levison, H. F., Henry, G. W., 2009, ApJS, 182, 97

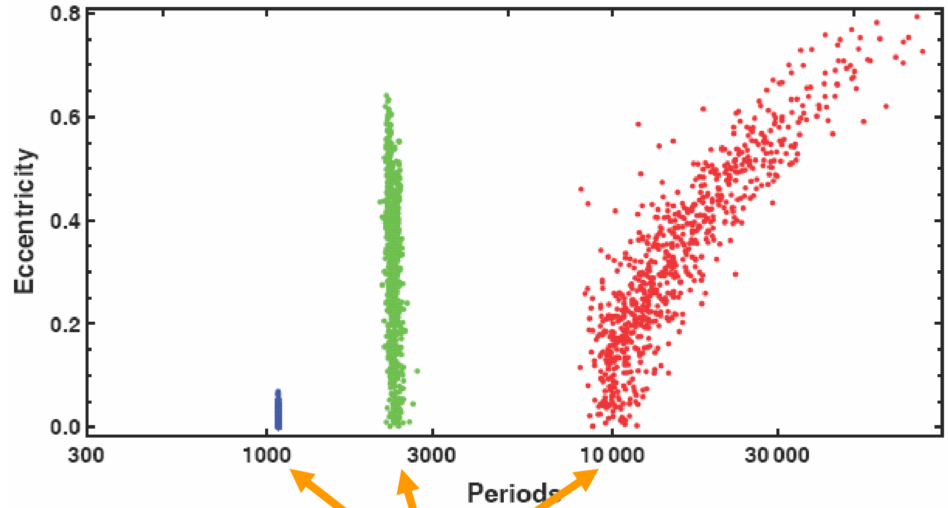
4) 2010, Gregory, P. C., & Fischer, D. A., MNRAS, 403, 731, 2010

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



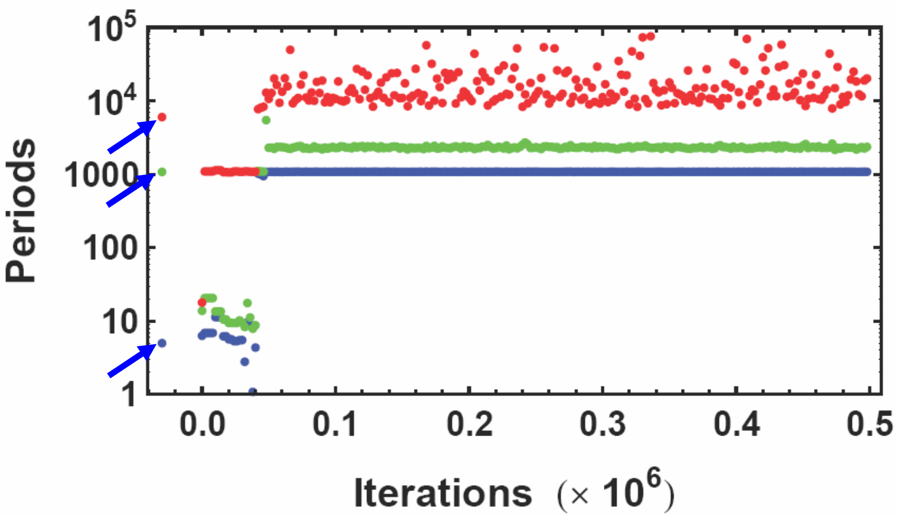
47 Ursae Majoris 3 planet FMCMC model fit, Lick telescope data



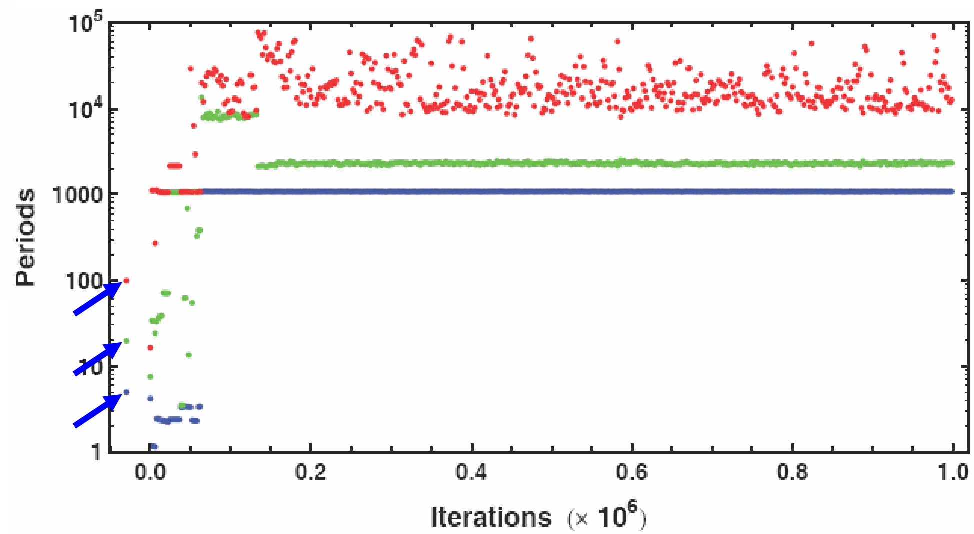


Evidence for 3 planets

Longest period not well constrained because period is longer that data span



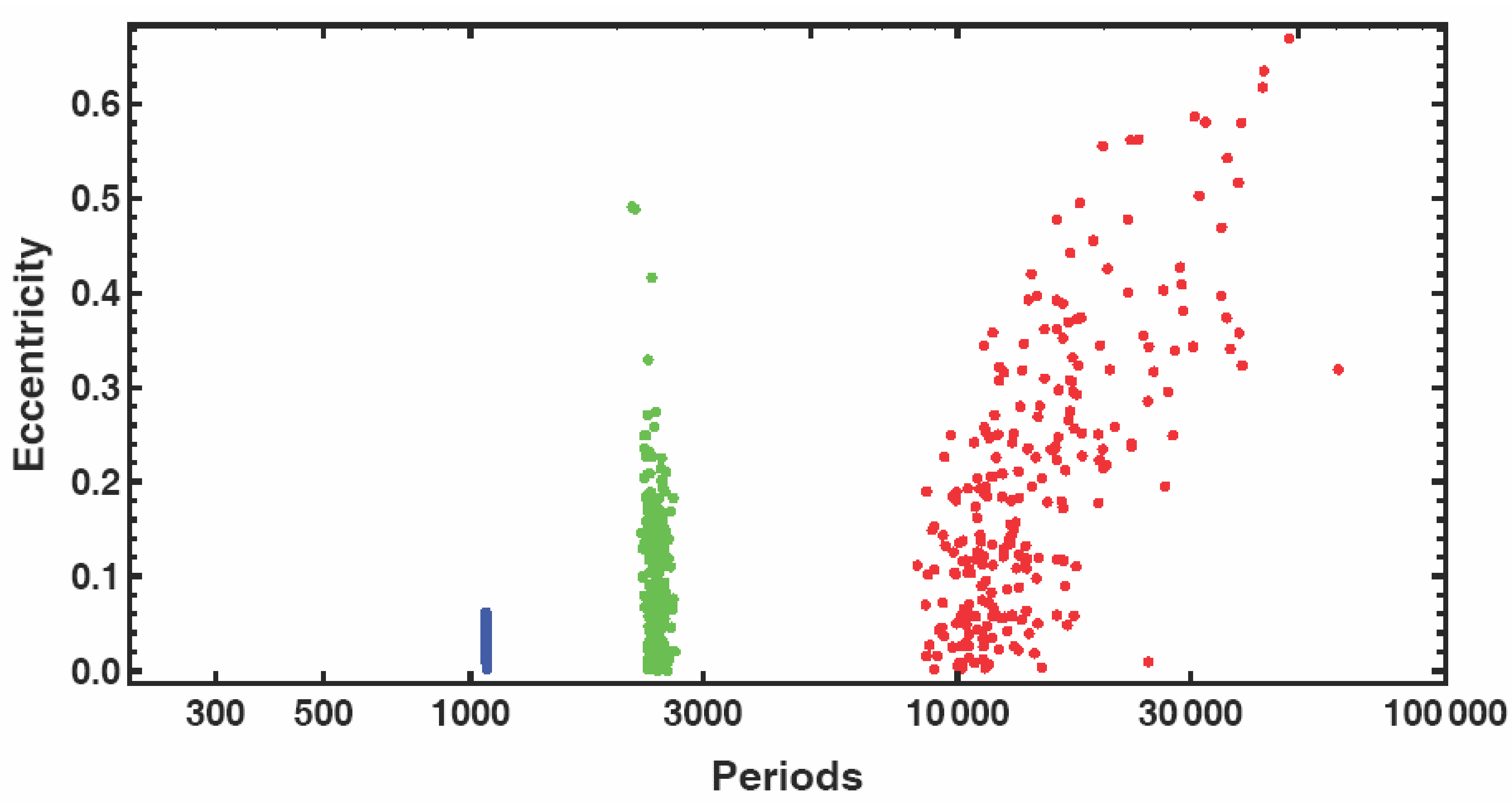
Arrow indicates starting periods



47 Ursae Majoris **3 planet FMCMC model fit**

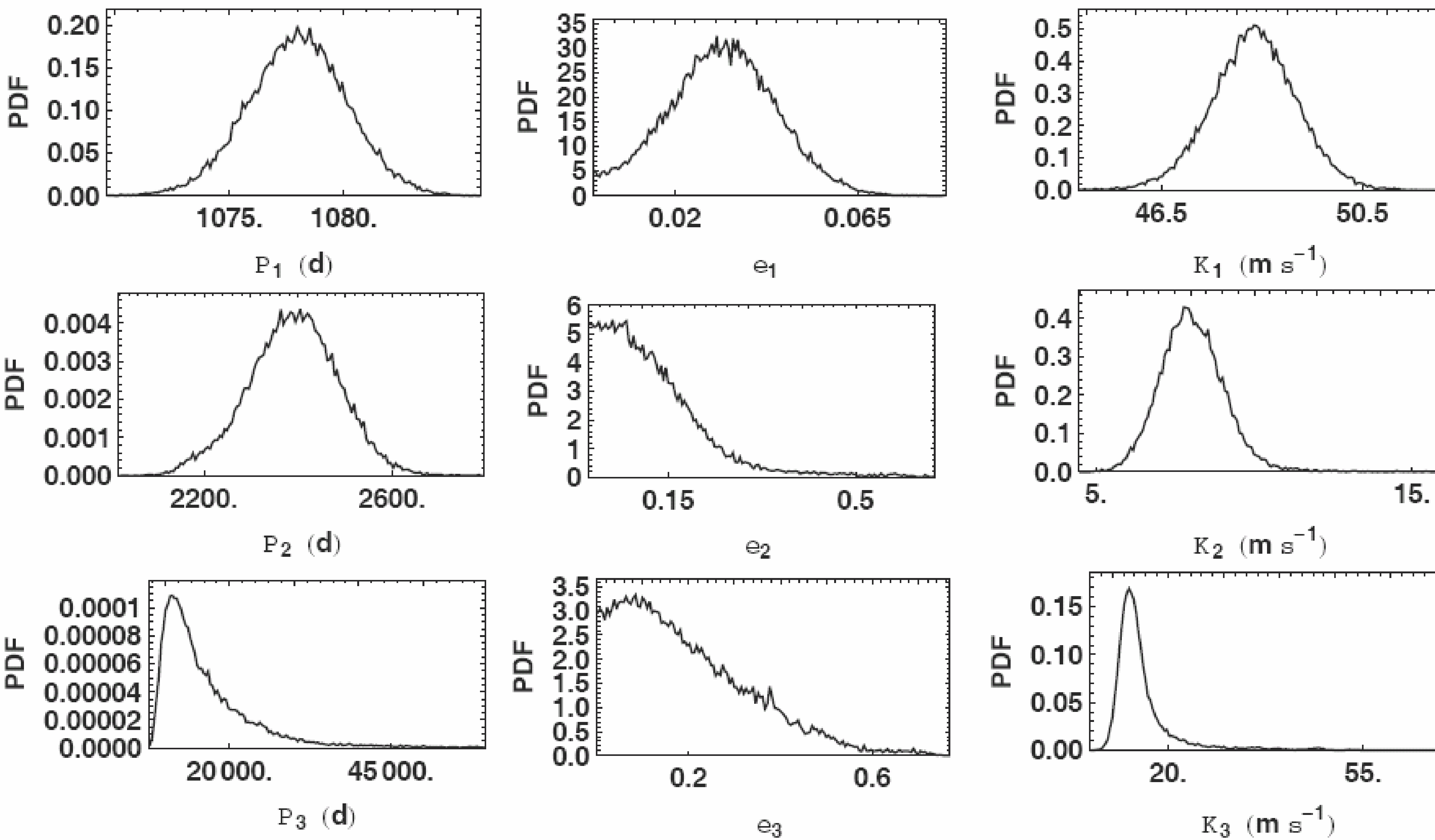
Lick + Hobby-Eberly + Harlem J Smith telescope data.

Total of 413 velocity measurements



47 Ursae Majoris

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



47 Ursae Majoris Model selection

$$\text{Odds ratio} = O_{i2} = \frac{p(M_i \mid D, I)}{p(M_2 \mid D, I)} = \frac{p(M_i \mid I)}{p(M_2 \mid I)} \times \frac{p(D \mid M_i, I)}{p(D \mid M_2, I)} = \frac{p(M_i \mid I)}{p(M_2 \mid I)} \times B_{i2}$$

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

Model	Bayes factor B_{i2}	False Alarm Probability (assuming a priori all models equally probable)
M_{0s}	1.6×10^{-141}	
M_{1s}	2.0×10^{-28}	8×10^{-114}
M_{2s}	1.0	2×10^{-28}
M_{3s}	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.

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

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

Gliese 581 a 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

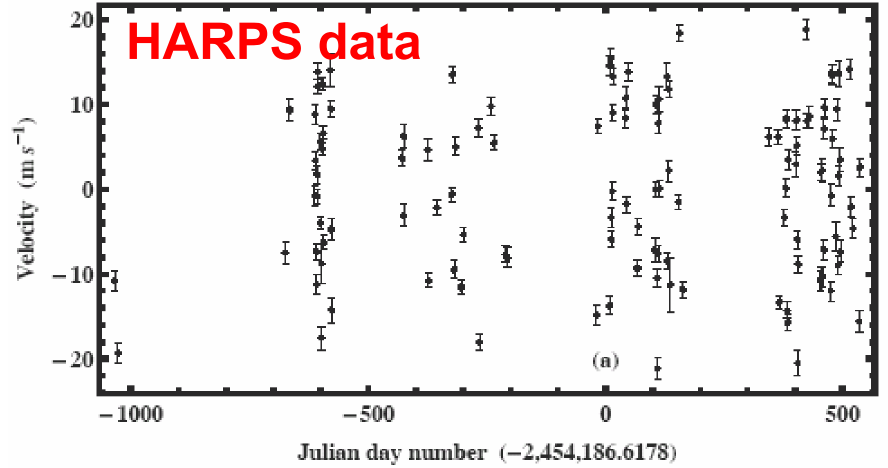
2) **2010,**
   - Planet f is 7 Earth mass
   - Planet g is 3.1 Earth mass

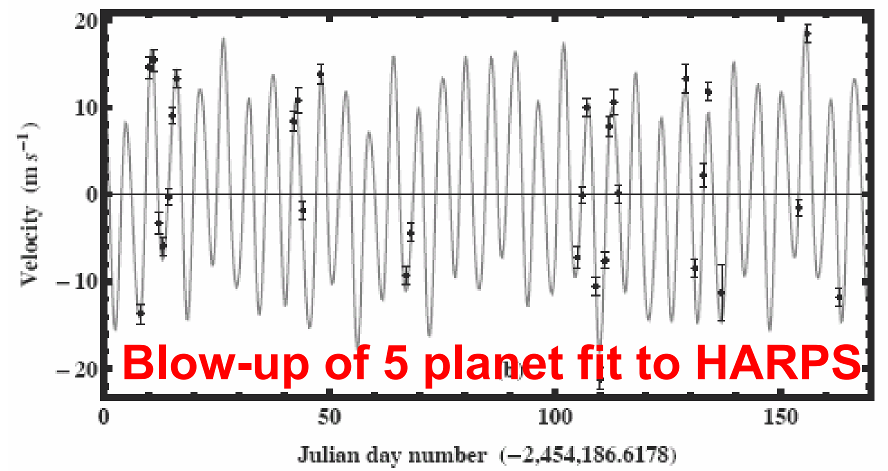
Steven Vogt et al., ApJ, 723, p. 954, 2011, their analysis assumes all circular orbits.

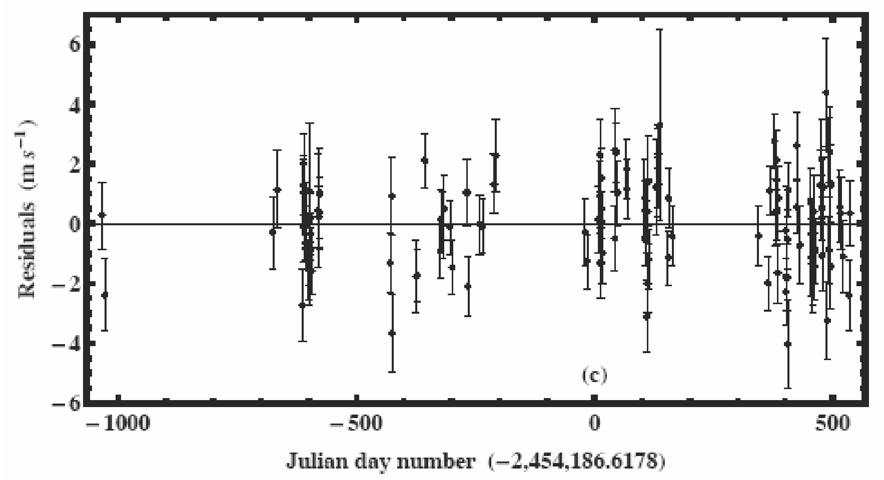
3) **2011, Gregory, P. C., MNRAS in press**
http://www.physics.ubc.ca/~gregory/gregory.html

Don't support claim for planet 581g.

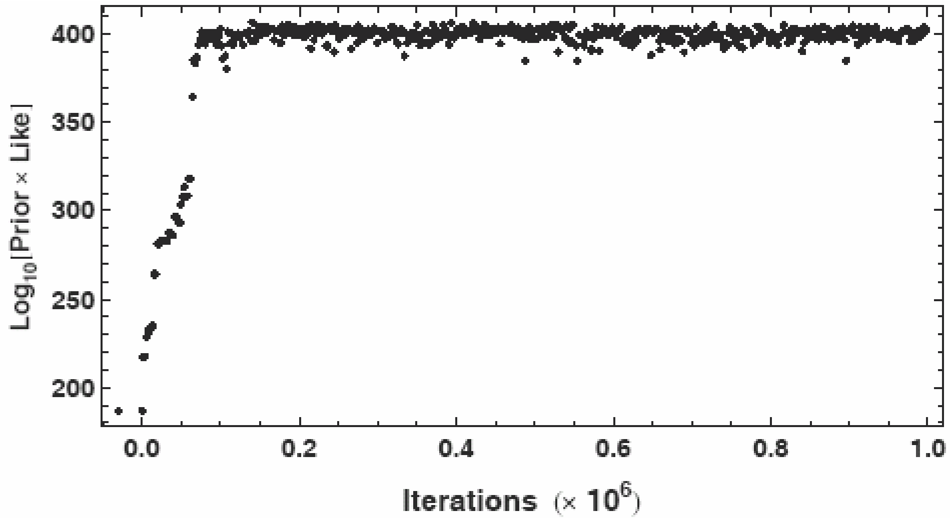
Find evidence that HIRES uncertainties are much larger than the quoted values, an extra 1.8 m s^{-1} added in quadrature.

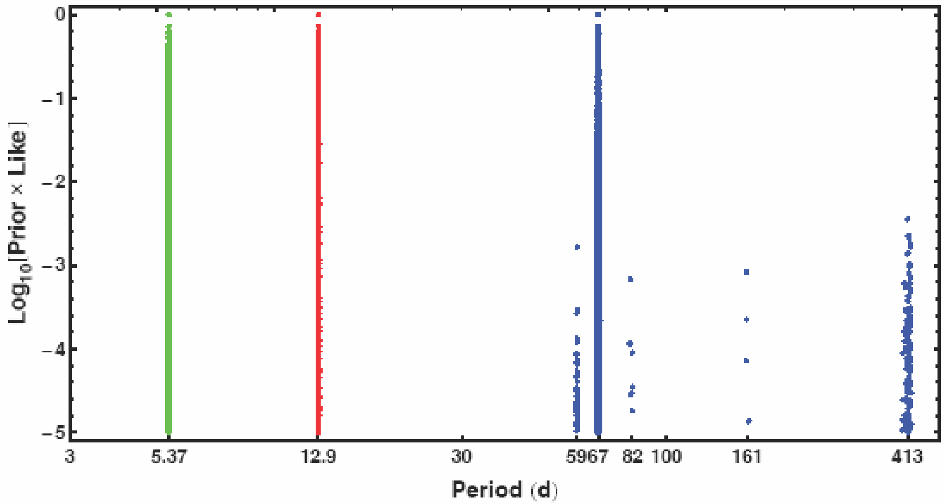




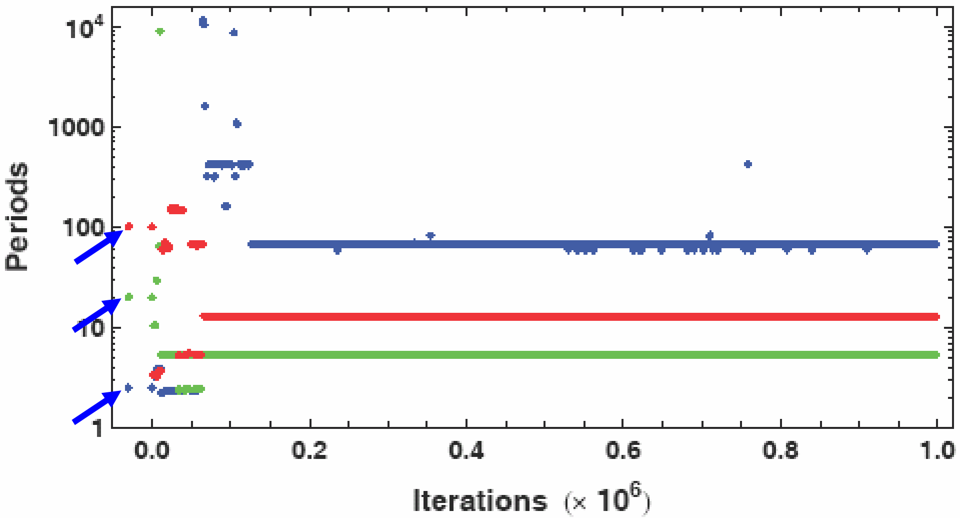


Gliese 581 three planet fit to HARPS radial velocity data

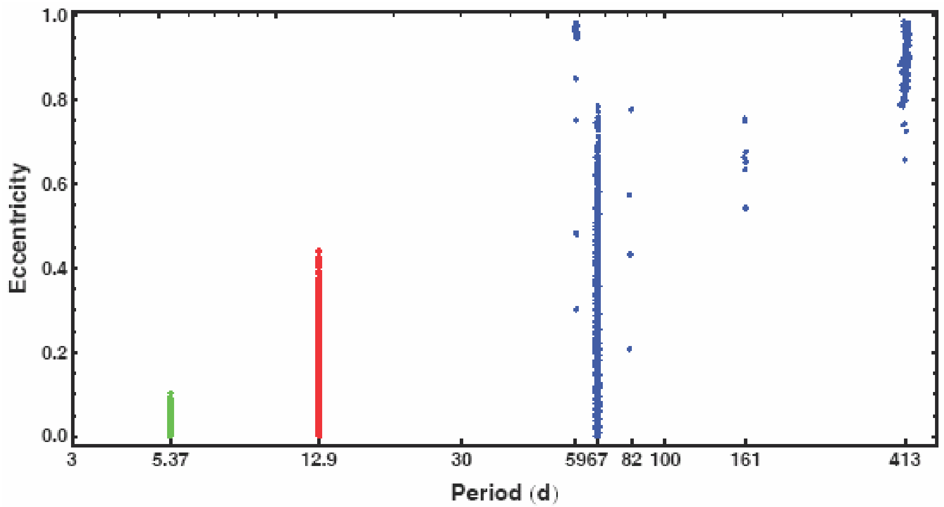




3 planet Kepler periodogram



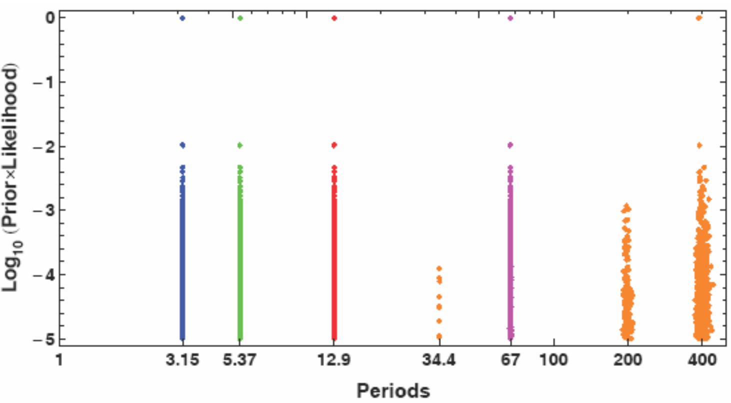
Arrow ↗ indicates starting periods



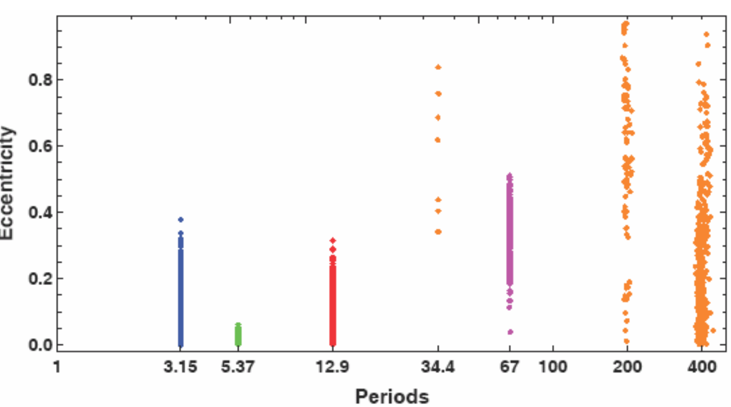
Eccentricity versus period

Gliese 581

5 planet fit to HARPS data

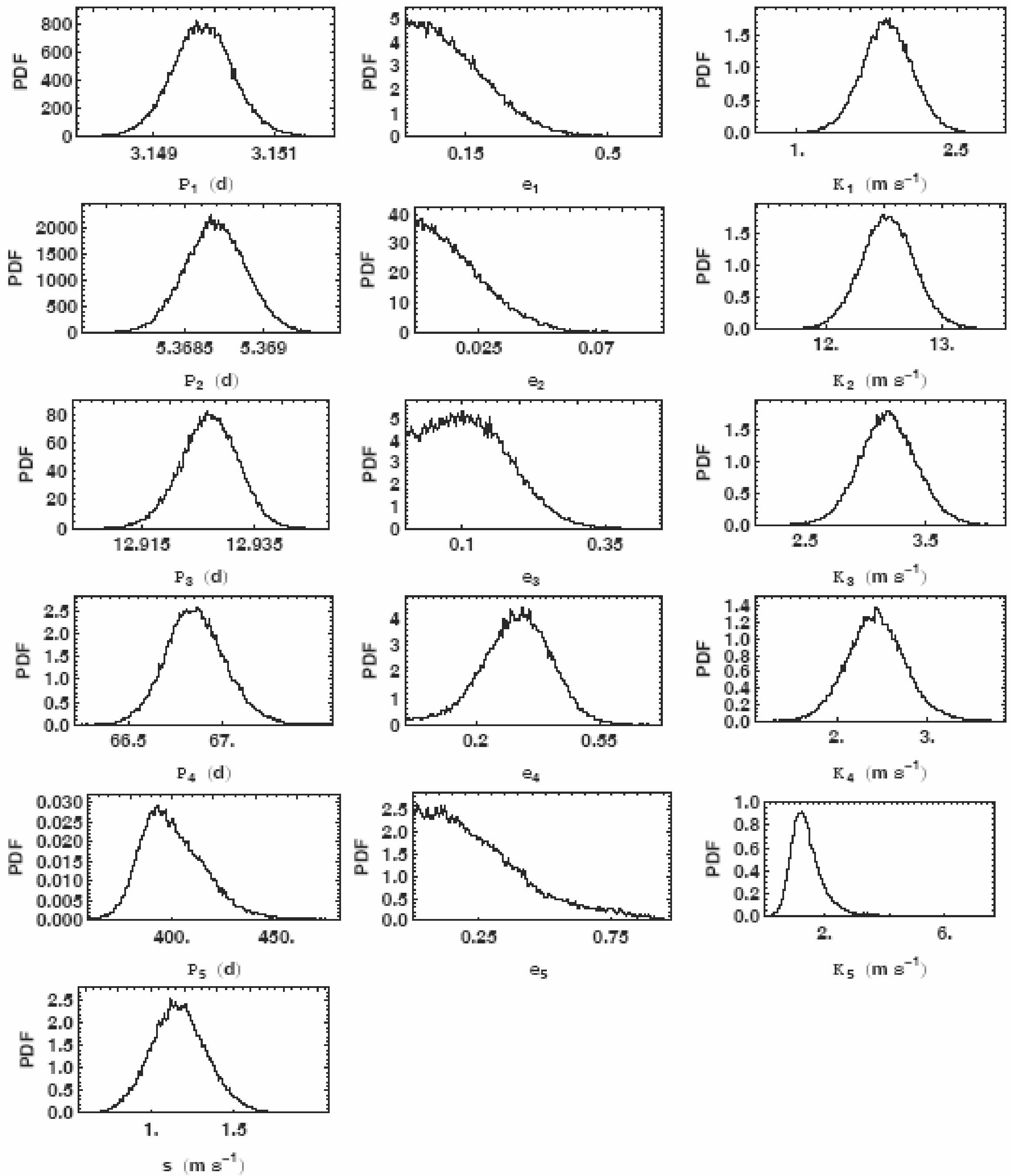


5 planet Kepler periodogram



Eccentricity versus period

Marginal parameter distributions



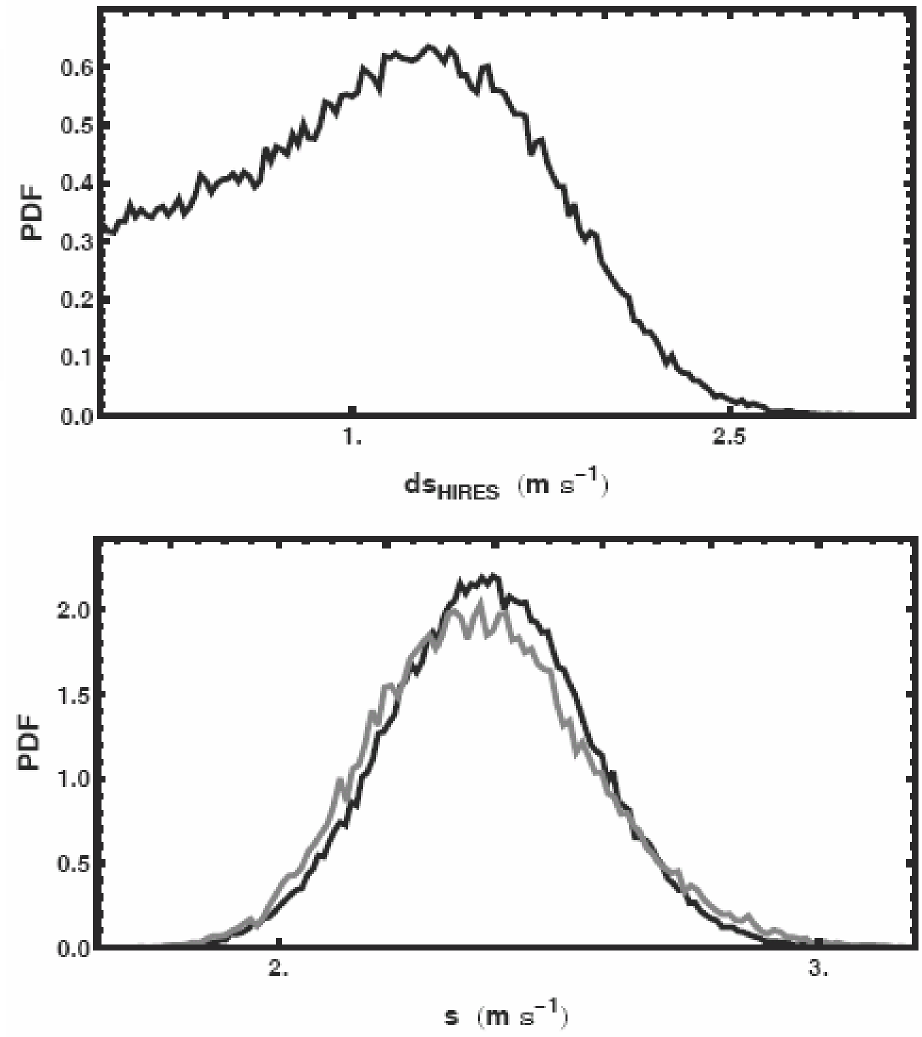
Gliese 581: The availability of multi-observatory data (HARPS and HIRES) permits a Bayesian analysis to distinguish and quantify any common extra noise (stellar jitter) from other unrecognized measurement uncertainties.

Final Noise model

$$\sigma_{\mathrm{HARPS}_j} = \sqrt{\sigma^2_{\mathrm{quoted}_j} + s^2}$$

$$\sigma_{\mathrm{HIRES}_j} = \sqrt{\sigma^2_{\mathrm{quoted}_j} + ds^2_{\mathrm{HIRES}} + s^2}$$

2 planet fit to HIRES/HARPS



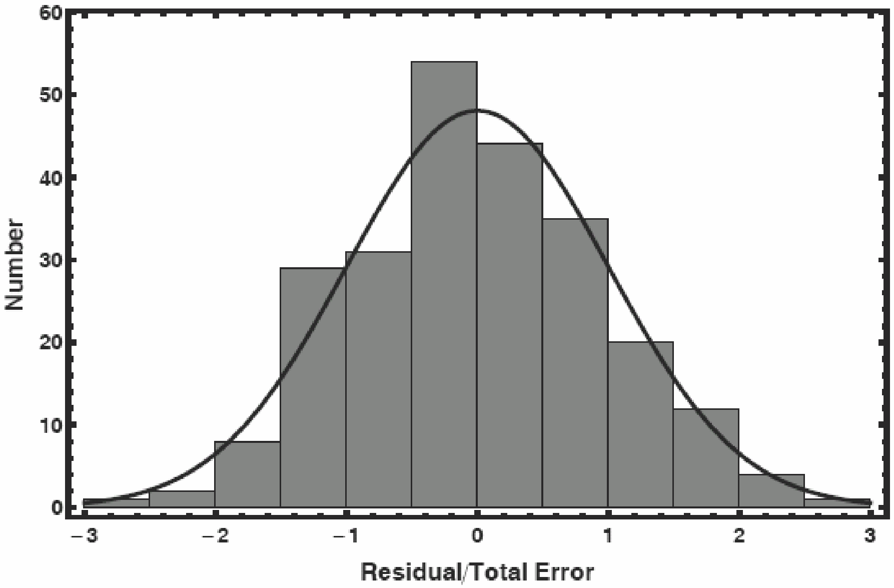
The figure shows the resulting marginal probability distributions for ds_{HIRES} and s obtained from a two planet fit to the combined HARPS/HIRES data set. The gray curve in the lower panel is the marginal for s from fit to HARPS only data.

Gliese 581: Find evidence that HIRES uncertainties are much larger than the quoted values by an extra 1.8 m s^{-1} added in quadrature.

Final Noise model

$$\sigma_{\mathrm{HARPS}_j} = \sqrt{\sigma^2_{\mathrm{quoted}_j} + s^2}$$

$$\sigma_{\mathrm{HIRES}_j} = \sqrt{\sigma^2_{\mathrm{quoted}_j} + ds^2_{\mathrm{HIRES}} + s^2}$$

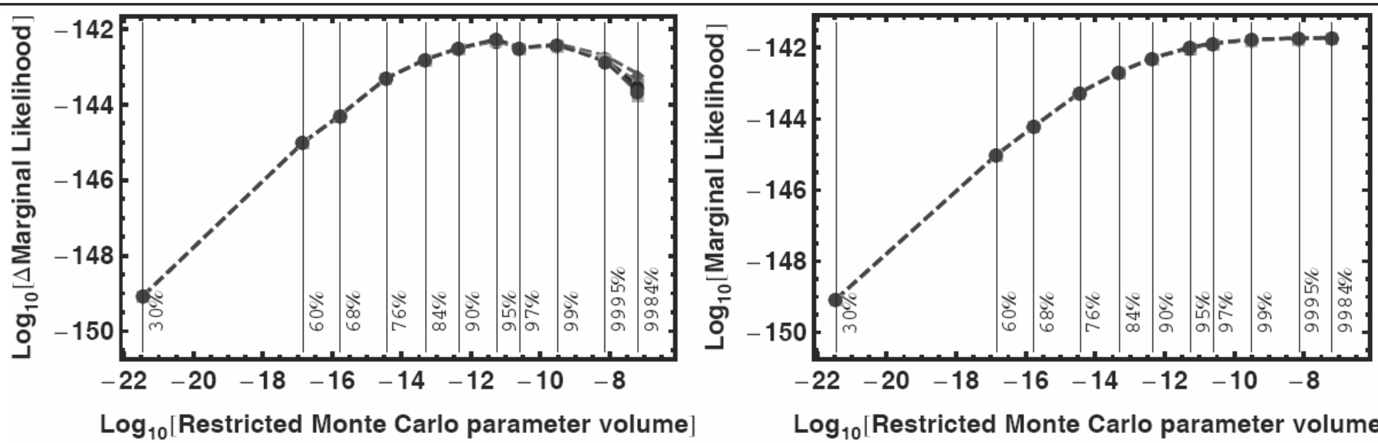


Extra noise parameters from fit to combined HIRES/HARPS data

Parameter	2 planet full data	3 planet full data	4 planet full data	4 planet partial
ds_{HIRES}	$1.15^{+1.05}_{-0.35}$	$1.55^{+0.58}_{-0.47}$	$1.84^{+0.35}_{-0.33}$	$1.76^{+0.45}_{-0.39}$
s	$2.39^{+0.25}_{-0.11}$	$2.0^{+0.2}_{-0.2}$	$1.45^{+.17}_{-.17}$	$1.39^{+.15}_{-.18}$

Gliese 581 Model Selection: Marginal likelihoods computed using Nested Restricted Monte Carlo (NRMC) method

Model	Periods (d)	Marginal Likelihood	Bayes factor nominal	False Alarm Probability	s ($\mathrm{m\,s^{-1}}$)	RMS residual ($\mathrm{m\,s^{-1}}$)
M_0		6.10×10^{-197}	2.0×10^{-59}			9.8
M_1	(5.37)	$(4.221 \pm 0.003) \times 10^{-155}$	1.4×10^{-17}	1.4×10^{-42}	3.5	3.6
M_2	(5.37, 12.9)	$(1.94 \pm 0.01) \times 10^{-145}$	6.5×10^{-8}	2.2×10^{-10}	2.4	2.6
M_3	(5.37, 12.9, 66.9)	$(3.0^{+0.7}_{-0.5}) \times 10^{-142}$	10^{-4}	6.5×10^{-4}	1.9	2.2
M_4	(3.15, 5.37, 12.9, 66.9)	$(3.0^{+1.1}_{-0.6}) \times 10^{-138}$	1.0	10^{-4}	1.4	1.7
M_5	(3.15, 5.37, 12.9, 66.9, 399)	$(3.0^{\times 2.1}_{\times 0.65}) \times 10^{-136}$	10^2	0.01	1.2	1.5
M_6	(3.15, 5.37, 12.9, 34.4, 66.9, 399)	$(6.7^{\times 2.4}_{\times 1/3}) \times 10^{-141}$	2.2×10^{-3}	0.999978	1.0	1.4



Left panel shows the contribution of the individual nested intervals to the NRMC marginal likelihood for the 3 planet model (5 repeats). The right panel shows the integral of these contributions versus the parameter volume of the credible region.