Additional Keplerian Signals in the HARPS data for Gliese 667C from a Bayesian re-analysis

Phil Gregory, Samantha Lawler, Brett Gladman

Physics and Astronomy
Univ. of British Columbia
Abstract

A re-analysis of Gliese 667C HARPS radial velocity data was carried out with a Bayesian multi-planet Kepler periodogram (from 0 to 7 planets) based on a controlled fusion MCMC algorithm. The most probable number of signals detected is 6 with a Bayesian false alarm probability of 0.012. The 6 signals detected include two previously reported with periods of 7.2 and 28.1 days, plus additional periods of 30.8, 38.8, 53.2, and 91.3 days suggesting the possibility of further planets. Stellar diagnostic information suggest that the 53 day signal is the second harmonic of the stellar rotation period and is likely the result of surface activity.

N-body simulations are underway to determine which of the remaining signals are consistent with a stable planetary system. At present, we have not found a long term (10^7 yr) stable 5 planet system consistent with the data. If we assume the 30.8 d period is a spectral artifact, we are able to identify a long term stable system with periods of 7.2, 28.1, 38.8, and 91.3 d. The corresponding M sin i values are 5.4, 4.8, 2.4, and 5.4 M_E and the semi-major axes inferred for the 28 and 38.8 d signals place them in the central region of the habitable zone. Further analysis is underway to define the probability bubble of stable orbits corresponding to this candidate 4 planet system.
Prior information: the signals satisfy Kepler’s laws

The radial velocity equation, a nonlinear model

\[ f_i = \text{model prediction} = V + K \left( \cos \left( \theta [t_i + \chi P] + \omega \right) + e \cos \omega \right) \]

\[ V = \text{systematic velocity} \]
\[ K = \text{velocity amplitude} = \frac{2 \pi a \sin i}{P \sqrt{1 - e^2}}; \]
\[ (a = \text{semi-major axis}, \ i = \text{inclination}) \]
\[ e = \text{orbital eccentricity} \]
\[ \omega = \text{longitude of periastron} \]
\[ \chi = \text{fraction of orbit prior to data reference time that periastron occurred at} \]
\[ \theta (t_i + \chi P) = \text{true anomaly} \]
\[ = \text{angle of star in orbit at time } t_i \text{ relative to periastron} \]

\[ \theta_i \text{ and } t_i \text{ are related by the conservation of angular momentum equation} \]

\[ \frac{\partial}{\partial t} \theta [t] = -\frac{2 \pi}{P \left( 1 - e^2 \right)^{3/2}} \left( 1 + e \cos \theta \right)^2 = 0 \]
Controlled Fusion MCMC

The combination of nonlinear model, sparse sampling and huge prior period range of 0.5 d to 1000 yr, yields a highly multi-modal target distribution which is a problem for a straight Metropolis MCMC.

To deal with this, Gregory has developed a new Markov chain Monte Carlo algorithm called controlled fusion MCMC. It incorporates:

a) parallel tempering*,
   b) simulated annealing and,
   c) genetic crossover operations.

Each of these features facilitate the detection of a global minimum in chi-squared in a multi-modal environment. By combining all three, the algorithm greatly increases the probability of realizing this goal.

Fusion MCMC is a very general nonlinear model fitting method applicable to many other problems.

* Also known as Exchange Monte Carlo (Hukushima & Nemoto 1996)
This fusion MCMC approach has been achieved through the development of a unique multi-stage adaptive control system, hence the term controlled fusion MCMC.

Among other things the control system automates the tuning of MCMC proposal distributions for efficient exploration of the model parameter space even when the parameters are highly correlated.

At each iteration, a single joint proposal to jump to a new location in the parameter space is generated.

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

8 parallel tempering Metropolis chains

Output at each iteration

- Parameters, logprior + β \times \text{like}, \log\text{prior} + \log\text{like}
- Parameters, logprior + β \times \text{like}, \log\text{prior} + \log\text{like}
- Parameters, logprior + β \times \text{like}, \log\text{prior} + \log\text{like}
- Parameters, logprior + β \times \text{like}, \log\text{prior} + \log\text{like}
- Parameters, logprior + β \times \text{like}, \log\text{prior} + \log\text{like}
- Parameters, logprior + β \times \text{like}, \log\text{prior} + \log\text{like}
- Parameters, logprior + β \times \text{like}, \log\text{prior} + \log\text{like}
- Parameters, logprior + β \times \text{like}, \log\text{prior} + \log\text{like}

β values

Parallel tempering swap operations

Anneal independent Gaussian proposal σ’s

2 stage proposal σ control system error signal = (actual joint acceptance rate – 0.25)

Effectively defines burn-in interval

Refine & update Gaussian proposal σ’s

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

MCMC adaptive control system

Monitor for parameters with peak probability

Genetic algorithm

Every 40th iteration perform gene crossover operation to breed a more probable parameter set.
Controlled Fusion MCMC

8 parallel tempering Metropolis chains

\[ \beta = 1.0 \]
\[ \beta = 0.72 \]
\[ \beta = 0.52 \]
\[ \beta = 0.39 \]
\[ \beta = 0.29 \]
\[ \beta = 0.20 \]
\[ \beta = 0.13 \]
\[ \beta = 0.09 \]

\( \beta \) values
Parallel tempering swap operations

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

new parameter value
from proposal C
repeat
from proposal I

Add every 2\(^{nd}\) to a buffer
latest 300 values
difference of random pairs
multiply by constant <1

During burn-in control system adjusts constant so acceptance rate from C proposals = 25 %

‘I’ proposals
Independent Gaussian proposal scheme employed 50% of the time

‘C’ proposals
Proposal distribution with built in parameter correlations used 50% of the time

MCMC adaptive control system
Bayes Theorem

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

Target posterior

If we 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**

Multiple Planets model = sum of multiple independent Keplerian orbits.
Gliese 667C 3rd isolated member of triple star system
distance = 22.1 ly

History

1) 2011,
   Planet b with M sin i = 5.9 M_{Earth}
   + two other interesting periods at 90 & 28d (habitable zone orbit)
   Bonfils et al., arXiv:1111.5019v2

2) 2012, Anglada et al. & Delfosse et al.,
   Confirm planet b & report planet c 28d, 4.3 M_{Earth} (HZ)

Likely explanation of the slope is the orbital motion (P ~ 3900 yr) of star about the CM of the Gl 667ABC triple system.
Gliese 667C

2 Keplerian FMCMC model fit

Arrow indicates starting periods

2nd period exhibits multiple peaks
Gliese 667C

2 Keplerian FMCMC model fit

Arrow ↗ indicates starting periods
Gliese 667C    3 Keplerian signal  FMCMC model fit

Additional periods detected in other 3 signal FMCMC runs 38.8, 87, 106, & 128 d. In all cases signals were detected at 7.2 and 28.1 d.
Questions and complications

How to decide when to stop adding more Keplerian signals?

For a Bayesian -> use model selection and compute the Bayes factor.

=> Favors the 6 signal model over nearest rival by a factor of 137

Star spots can also give rise to a Keplerian–like signal.

Complication: If there are additional signals beyond what is assumed in the model, they will lead to correlated or colored noise.

Can investigate this by computing the autocorrelation function of the model residuals.

=> 6 signal model residuals consistent with white noise while models with fewer signals exhibit correlated noise.
Autocorrelation function of residuals

\[ \rho(j) = \frac{\sum_{\text{overlap}} (x_i - \bar{x})(x_{i+j} - \bar{x})}{\sqrt{\sum_{\text{overlap}} (x_i - \bar{x})^2} \times \sqrt{\sum_{\text{overlap}} (x_{i+j} - \bar{x})^2}} \]

where \( x_i \) is the \( i^{th} \) residual, \( j \) is the lag and \( x \) is the mean of the samples in the overlap region. Because the data are not uniformly sampled, for each lag all sample pairs that differed in time by this lag \( \pm 0.1 \) d were utilized.

The solid red curve in the 1 signal residuals is the average autocorrelation generated from 400 simulated data sets of a 5 signal model (28.1, 30.8, 38.8, 53.2, & 91 d periods) together with the quoted measurement errors.
Gliese 667C  6 Keplerian signal  FMCMC model fit
A generalized Lomb-Scargle (GLS) periodogram (Zechmeister & Kurster 2009) for the maximum a posteriori (MAP) parameter values of the 6 signal fit residuals. The GLS allows for a floating offset and weights.

The dashed horizontal lines correspond to peak periodogram levels for which the frequentist false alarm probability (FAP) would $= 0.1$ & $0.01$.

No evidence of any significant peaks or red noise in these residuals.
GI 667C Subset of HARPS data (1st 143 velocities)

Includes burn-in phase to show aliases

Upper and lower one year alias of 38.8d.
Are any of our 6 signals spectral artifacts?

Delfosse et al. (2012) conclude the star’s rotation $P \sim 105$ days, from a periodogram of a stellar activity diagnostic.

At various stages in our RV analysis a $P = 106$ d was detected. In our preferred 6 signal model, a $P = 53$ d (2$^{nd}$ harmonic) signal was detected. Consider this signal an artifact of surface activity.

Periodogram of a stellar activity diagnostic, the FWHM of the cross correlation function (Delfosse et al. 2012).
Further analysis in progress

1. Modified likelihood in Fusion MCMC to incorporate n-body simulation for an n planet model (for n = 3,4,5) plus m additional Keplerian-like artifact signals (m = 3,2,1)
   
   Uses a 4th order Hermite n-body integrator with automatic step size.

2. Check the long term stability \((10^7 \text{ yr})\) of n planet model (for n = 3,4,5) with FMCMC parameters that provides a good fit to the data.

   => 4 of the signals (7.2, 28.1, 38.8, 91 d) are consistent with a stable candidate 4 planet system.
Conclusions

1. Our results suggest there are 6 interesting signals in the HARPS data for Gliese 667C. One is likely an artifact of surface activity.

2. The 6 signal residuals are consistent with white noise but residuals with fewer than 6 signals are correlated (colored noise).

3. Further analysis is underway to determine which of the remaining signals are consistent with a stable planetary system.

4. Our 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 by Gregory of a “Controlled Fusion MCMC” nonlinear model fitting algorithm, which can be applied to many other problems.
Gliese 667C    4 planet FMCMC model fit

Additional periods detected in other FMCMC runs 38.8 & 91 d.
Additional periods detected in other FMCMC runs were 30.8, 38.8, 87, 106, 128, & 184 d.

In all 5 signal runs $P = 7.2, 28.1, 53, & 91$ d were detected.