A Bayesian re-analysis of HD 11964: evidence for three planets

P. C. Gregory

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

Published in 'Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 27th International Workshop', Kevin H. Knuth, Ariel Caticha, Julian L. Center, Adom Giffin, Carlos C. Rodríguez, eds, AIP Conference Proceedings, **954**, *307-314*

Abstract. Astronomers searching for the small signals induced by planets inevitably face significant statistical challenges. Bayesian inference has the potential of improving the interpretation of existing observations, the planning of future observations and ultimately inferences concerning the overall population of planets. This paper illustrates how a re-analysis of published radial velocity data sets with a Bayesian multi-planet Kepler periodogram is providing strong evidence for additional planetary candidates. The periodogram is implemented with a Markov chain Monte Carlo (MCMC) algorithm that employs an automated adaptive control system. For HD 11964, the data has been re-analyzed using 1, 2, 3 and 4 planet models. The most probable model exhibits three periods of $38.02^{+0.06}_{-0.05}$, 360^{+4}_{-4} and 1924^{+44}_{-43} d, and eccentricities of $0.22^{+0.11}_{-0.22}$, $0.63^{+0.34}_{-0.17}$ and $0.05^{+0.03}_{-0.05}$, respectively. Assuming the three signals (each one consistent with a Keplerian orbit) are caused by planets, the corresponding limits on planetary mass ($M \sin i$) and semi-major axis are $(0.090^{+0.15}_{-0.14} M_J, 0.253^{+0.009}_{-0.009}\mathrm{au})$, $(0.21^{+0.06}_{-0.07}\mathrm{M_J}, 1.13^{+0.04}_{-0.04}\mathrm{au})$, $(0.77^{+0.08}_{-0.08}\mathrm{M_J}, 3.46^{+0.13}_{-0.13}\mathrm{au})$, respectively.

Keywords: Extrasolar planets, Bayesian methods, Markov chain Monte Carlo, time series analysis, periodogram, HD 11964

INTRODUCTION

Improvements in precision radial velocity measurements and continued monitoring are permitting the detection of lower amplitude planetary signatures. One example of the fruits of this work is the detection of a super earth in the habital zone surrounding Gliese 581 by Udry et al. [19]. This and other remarkable successes on the part of the observers is motivating a significant effort to improve the statistical tools for analyzing radial velocity data, e.g., [7, 6, 5, 9, 3, 15, 14]. Much of the recent work has highlighted a Bayesian MCMC approach as a way to better understand parameter uncertainties and degeneracies and to compute model probabilities.

Gregory [8, 9, 10, 11, 12] presented a Bayesian MCMC algorithm that makes use of parallel tempering to efficiently explore the full range of a large model parameter space starting from a random location. The prior information insures that any periodic signal detected satisfies Kepler's laws and thus the algorithm functions as a Kepler periodogram [1]. In addition, the Bayesian MCMC algorithm provides full marginal parameters distributions for all the orbital elements that can be determined from radial velocity data. The samples from the parallel chains can also be used to compute the marginal likelihood for a given model [8] for use in computing the Bayes factor that is needed to compare models with different numbers of planets. The parallel tempering MCMC algorithm employed in this work includes an innovative two stage adaptive control system that automates the selection of efficient Gaussian parameter proposal distributions through an annealing operation. This feature coupled with parallel tempering makes it practical to attempt a blind search for multiple planets simultaneously. This was done for the analysis of the current data set and for the analysis of the HD 208487 reported earlier [11]. Of course, there is no guarantee that the algorithm will discover all modes in a multiple mode problem. More discussion of the control system is given below.

This paper illustrates how a Bayesian re-analysis of the 87 precision radial velocity measurements for HD 11964 published by Butler et al. [2] is providing strong evidence for additional planetary candidates. A more detailed account of many aspects of this analysis can be found in Gregory [12]. Butler et al. [2] reported the detection of a single planet with a period of 2110 ± 270d after removing a trend in the data.

RE-ANALYSIS OF HD 11964

The Bayesian multi-planet Kepler periodogram utilizes a parallel tempering Markov chain Monte Carlo algorithm which yields samples of the joint probability density distribution of the model parameters and permits a direct comparison of the probabilities of models with differing numbers of planets. In parallel tempering, multiple MCMC chains are run in parallel with each chain corresponding to a different temperature. We parameterize the temperature by its reciprocal, $\beta = 1/T$ which varies from zero to 1. The joint probability density distribution for the parameters ($\vec{X}$) of model M_i for a particular chain is given by

$$p(\vec{X}|D,M_i,I,\beta) = P(\vec{X}|M_i,I) \times p(D|\vec{X}M_i,I)^{\beta} \tag{1}$$

For parameter estimation purposes 12 chains
($\beta = \{0.05, 0.1, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.70, 0.80, 0.90, 1.0\}$) were employed. At intervals, a pair of adjacent chains on the tempering ladder are chosen at random and a proposal made to swap their parameter states. The mean number of iterations between swap proposals was set $= 8$. 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. The final samples are drawn from the $\beta = 1$ chain, which corresponds to the desired target probability distribution. For $\beta \ll 1$, the distribution is much flatter. The choice of β values can be checked by computing the

---

[1] Following on from the pioneering work on Bayesian periodograms by Jaynes [13] and Bretthorst [1].

swap acceptance rate. When they are too far apart the swap rate drops to very low values. For the β values employed the swap rate was $\sim 50\%$. The lowest β value was chosen to achieve a broad sampling of the prior parameter range. A more common strategy is to propose a swap after each iteration and use fewer more widely spaced chains that achieve a swap rate of $\sim 25\%$. During the early development phase of the algorithm, this latter strategy appeared not to be quite as satisfactory, but we plan to re-visit this issue.

The samples from hotter simulations can also used to evaluate the marginal (global) likelihood needed for model selection, following Section 12.7 of Gregory [8] and Gregory [12]. Marginal likelihoods estimated in this way require many more parallel simulations. For HD 11964, 40 β levels were used spanning the range $\beta = 10^{-8}$ to 1.0.

For a one planet model the predicted radial velocity is given

$$v(t_i) = V + K[\cos\{\theta(t_i + \chi P) + \omega\} + e\cos\omega], \tag{2}$$

and involves the 6 unknown parameters

$V =$ a constant velocity.

$K =$ velocity semi-amplitude.

$P =$ the orbital period.

$e =$ the orbital eccentricity.

$\omega =$ the longitude of periastron.

$\chi =$ the fraction of an orbit, prior to the start of data taking, that periastron occurred at. Thus, $\chi P =$ the number of days prior to $t_i = 0$ that the star was at periastron, for an orbital period of P days.

$\theta(t_i + \chi P) =$ the angle of the star in its orbit relative to periastron at time t_i measured with the focus of the orbital ellipse as the origin, also called the true anomaly.

We utilize this form of the equation because we obtain the dependence of θ on t_i by solving the conservation of angular momentum equation. Gregory [11] describes the advantage of this approach.

In a Bayesian analysis we need to specify a suitable prior for each parameter. The priors used in the current analysis are given in Table 1 of Gregory [12]. Following Gregory [9], all of the models considered in this paper incorporate an extra additive noise whose probability distribution is Gaussian with zero mean and standard deviation s. Marginalizing s has the desirable effect of treating anything in the data that can't be explained by the model and known measurement errors (e.g., stellar jitter) as noise, leading to conservative estimates of orbital parameters. Following Gregory [11], we employed a modified Jeffrey's prior for s with a knee, $s_0 = 1\text{m s}^{-1}$.

MCMC ADAPTIVE CONTROL SYSTEM

The process of choosing a set of useful proposal σ's when dealing with a large number of different parameters can be very time consuming. In parallel tempering MCMC, the problem is compounded because of the need for a separate set of proposal σ's for each chain. We have automated this process using an innovative two stage statistical control system (CS) in which the error signal is proportional to the difference between the current joint parameter acceptance rate and a target acceptance rate, typically 25% [17].

In the first stage, an initial set of proposal σ's ($\approx 10\%$ of the prior range for each parameter) are used for each chain. During the major cycles, the joint acceptance rate is measured based on the current proposal σ's. During the minor cycles, each proposal σ is separately perturbed to determine an approximate gradient in the acceptance rate for that parameter. The σ's are then jointly modified by a small increment in the direction of this gradient. This is done for each of the parallel simulations or chains as they are sometimes called. Proposals to swap parameter values between tempering levels are allowed during major cycles but not within minor cycles.

Although the first stage CS achieves the desired joint acceptance rate, it often happens that a subset of the proposal σ's are too small leading to an excessive autocorrelation in the MCMC iterations for these parameters. Part of the second stage CS corrects for this as follows.

The goal of the second stage is to achieve a set of proposal σ's that equalizes the MCMC acceptance rates when new parameter values are proposed separately and achieves the desired acceptance rate when they are proposed jointly. Let $\mathrm{acc}(1)$ equal the acceptance for single parameter proposals and $\mathrm{acc}(m)$ the desired acceptance rate (typically 0.25) for m parameter joint proposals. We make use of the following relationship between $\mathrm{acc}(1)$ and $\mathrm{acc}(m)$

$$\mathrm{acc}(1) = \mathrm{acc}(m)^{1/m^{k\alpha}}, \tag{3}$$

where α is given by

$$\alpha = 0.8061 - 1.1205 \times 10^{-2} m + 3.1233 \times 10^{-4} m^2 - 3.0357 \times 10^{-6} m^3, \tag{4}$$

and $k = 0.85$ is an empirical determined quantity. Equ. (3) was arrived at in the following way. An MCMC simulation was run on an m parameter multivariate normal target probability distribution with a mean for each parameter of zero and a covariance matrix equal to an identity matrix. New parameters were proposed using another multivariate normal with mean zero and a covariance matrix equal to γ^2 times the identity matrix. Thus, γ is the ratio of the proposal σ to the target distribution σ for each parameter. For each choice of γ in the range 0.4 to 1.1, the MCMC acceptance rate for joint parameter proposals was determined as a function of m. For each γ the acceptance rate was well fit by a function of the form

$$\mathrm{acc}(m) = \mathrm{acc}(1)^{m^{\alpha_\gamma}}, \tag{5}$$

and the value of $m = m_\gamma$ at which $\mathrm{acc}(m) = 0.25$ was noted. For γ ranging from 0.4 to 1.1, m_γ varied from 34 to 5.4 and α_γ from 0.667 to 0.755. A cubic polynomial was fit to the $(m_\gamma, \alpha_\gamma)$ pairs yielding Equ. (3) without the k value. Of course, the actual Kepler target distribution is not a multivariate normal but with the inclusion of the empirically determined k value, Equ. (3) provides a useful scaling relationship.

The next step is to adjust the individual parameter proposal σ's to achieve an acceptance of $\mathrm{acc}(1)$ given by Equ. (3). Using the proposal σ's obtained in the first stage CS, each parameter is allowed to vary one at a time during a minor cycle and the acceptance rate measured. Let $acc_1 =$ the measured acceptance rate when the proposal σ for the parameter in question was σ_1. We then update the proposal σ for this parameter to σ_2 according to

$$\sigma_2 = \sigma_1 \sqrt{\frac{(acc_1 + \Delta)}{\mathrm{acc}(1)} \frac{(1 - \mathrm{acc}(1))}{(1 - acc_1 + \Delta)}}, \tag{6}$$

where we use a $\Delta = 0.01$.

If $acc_1 = \mathrm{acc}(1)$, then Equ. (6) leaves the proposal σ unchanged except for the small effect of the Δ term. The Δ term is there to handle the extremes of $acc_1 = 0$ and 1 gracefully. If $acc_1 = 1$, then we want to increase the proposal σ for that parameter. From Equ. (6) and $m = 17$ parameters, $\sigma_2/\sigma_1 = 6.7$. If on the other hand acc_1 is too low, say $acc_1 = 0.25$, we want to decrease the size of the proposal distribution. In this case, Equ. (6) yields $\sigma_2/\sigma_1 = 0.39$. Equ. (6) can be iterated for each parameter to achieve a final set of proposal σ's that achieve equal acceptance rates and a final joint acceptance rate of $\mathrm{acc}(m)$. In practice we iterate Equ. (6) twice for each parameter. Other forms of Equ. (6) could also achieve the same goal in an iterative fashion.

In general, the burn-in period occurs within the span of the first stage CS, i.e., the significant peaks in the joint parameter probability distribution are found, and the second stage improves the choice of proposal σ's for the highest probability parameter set. Occasionally, a new higher (by a user specified threshold) target probability parameter set emerges after the first two stages of the CS are completed. The control system has the ability to detect this and re-activating the second stage. In this sense the CS is adaptive. If this happens the iteration corresponding to the end of the control system is reset. The useful MCMC simulation data is obtained after the first two stages of the CS are switched off.

Although inclusion of the control system may result in a somewhat longer effective burn-in period, there is a huge saving in time because it eliminates many trial runs to manually establish a suitable set of proposal σ's. When the σ's are large all the MCMC chains explore broadly the prior distribution and locate significant probability peaks in the joint parameter space. As the proposal σ's are refined these peaks are more efficiently explored, especially in the higher β chains. This annealing of the proposal σ's typically takes place over the first 5,000 to 150,000 (unthinned) iterations for one planet and first 5,000 to 300,000 iterations for three planets. This may seem like an excessive number of iterations but keep in mind that (a) we are dealing with sparse data sets that can have multiple, widely separated probability peaks, (b) the typical start location in parameter space is far from the target posterior peak, and (c) we want the MCMC to locate the most significant probability peak before finalizing the choice of proposal σ's. Within each chain, the CS corresponds to an annealing operation. Taken together with the parallel tempering, the two operations enhance the chances of detecting peaks in the target posterior compared to just implementing either one.

RESULTS

Panel (a) of Figure 1 shows the precision radial velocity data for HD 11964 from Butler et al. [2] who reported a single planet with $M \sin i = 0.61 \pm 0.10$ in a 2110 ± 270 day

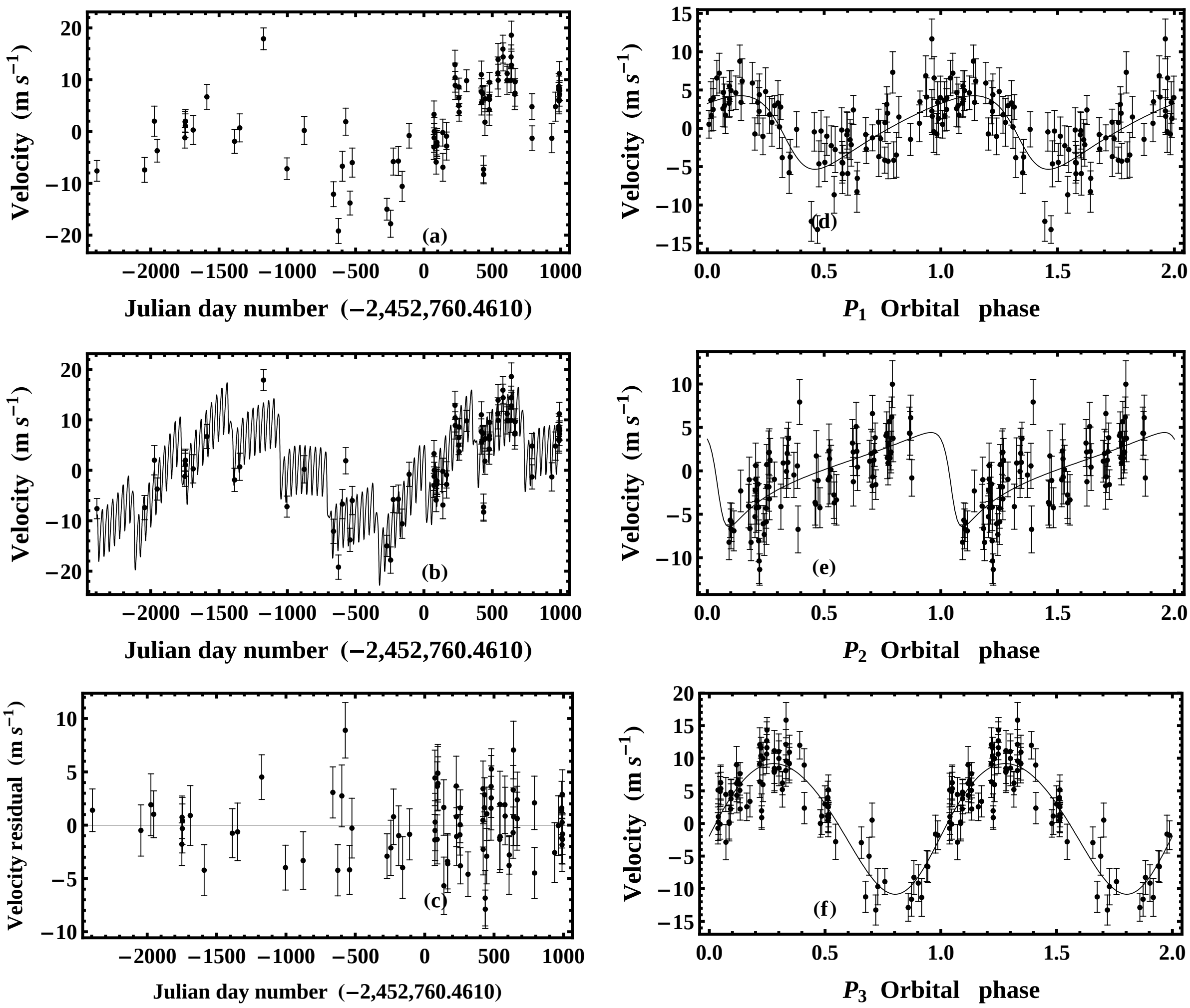

FIGURE 1. The data is shown in panel (a), the best fitting three planet ($P_1 = 38$, $P_2 = 357$, & $P_3 = 1928$ days) model versus time in (b), and the residuals in (c). Panel (d) shows the data plotted versus P_1 phase for two cycles, with the best fitting P_2 and P_3 orbits removed. Panel (e) shows the data plotted versus P_2 phase with the best fitting P_1 and P_3 orbits removed. Panel (f) shows the data plotted versus P_3 phase with the best fitting P_1 and P_2 orbits removed.

orbit with an eccentricity of 0.06 ± 0.17. Panels (b) and (c) show our best fitting three planet velocity curve and residuals. The initial starting location in period parameter space that was used for the Kepler periodogram ($P_1 = 10, P_2 = 500$ and $P_3 = 2300$d) was significantly different from the best location the algorithm found. Similar results were obtained with other different starting positions.

Table 1 gives our Bayesian three planet orbital parameter values and their marginal uncertainties. The parameter values given for our analysis are the median of the marginal probability distribution for the parameter in question and the error bars identify the boundaries of the 68.3% marginal credible regions. The value immediately below in square brackets is the maximum *a posteriori* (MAP) value determined using the Nelder-

TABLE 1. Three planet model parameter estimates.

Parameter	**planet 1**	**planet 2**	**planet 3**
P (d)	$38.02^{+0.6}_{-0.5}$ [38.07](38.08)	360^{+4}_{-4} [357](356)	1925^{+44}_{-44} [1928](1914)
K (m s^{-1})	$4.3^{+0.7}_{-0.7}$ [4.8](5.4)	$6.1^{+3.0}_{-3.3}$ [5.4](5.8)	$9.7^{+0.8}_{-0.8}$ [10.0](9.3)
e	$0.23^{+.10}_{-.22}$ [0.31](0.34)	$0.63^{+.35}_{-.13}$ [0.63](0.60)	$0.05^{+.03}_{-.05}$ [0.09](0.07)
ω (deg)	123^{+41}_{-48} [111](108)	103^{+38}_{-34} [107](90)	195^{+80}_{-74} [205](208)
a (au)	$0.2527^{+.0085}_{-.0085}$ [0.253](0.253)	$1.132^{+.039}_{-.039}$ [1.124](1.123)	$3.46^{+.13}_{-.13}$ [3.46](3.45)
$M \sin i$ (M_J)	$0.090^{+.014}_{-.015}$ [0.098](0.098)	$0.213^{+.058}_{-.067}$ [0.191](0.209)	$0.77^{+.08}_{-.08}$ [0.795](0.735)
Periastron passage (JD - 2,440,000)	12737^{+6}_{-3} [12736](12736)	12397^{+35}_{-32} [12421](12370)	10535^{+401}_{-414} [10564](10598)

Mead [16] downhill simplex method. Next to this, in parenthesis, is the MCMC parameter value corresponding to the largest joint posterior probability density, which is an approximation to the MAP value. The values derived for the semi-major axis and $M \sin i$, and their errors, are based on the assumed mass of the star $= 1.49 \pm 0.15$ M$_\odot$ [18]. Butler et al. [2] assumed a mass of $= 1.12$ M$_\odot$ but also quote Valenti & Fischer [18] as the reference. Panel (d) of Figure 1 shows the data with the best fitting P_2 and P_3 orbits subtracted, for two cycles of P_1 phase with the best fitting P_1 orbit overlaid. Panel (e) shows the data plotted versus P_2 phase with the best fitting P_1 and P_3 orbits removed. Panel (f) shows the data plotted versus P_3 phase with the best fitting P_1 and P_2 orbits removed.

Following Gregory [11], the marginal likelihoods and their uncertainties for the 1,2,3 and 4 planet models were computed for the HD 11964 data set. Assuming that all the models are equally probable *a priori*, the three planet model was found to be ≥ 600 times more probable than the next most probable model which is a two planet model. A detailed comparison of the different marginal likelihood estimates is given in a Gregory [12].

For the most probable three planet model, the estimated stellar jitter based on the MAP value of the s parameter is 1.9m s^{-1}.

CONCLUSIONS

In this paper, we provided further details of the innovative adaptive control system employed by our automated parallel tempering MCMC nonlinear model fitting algorithm. This has been applied to the analysis of precision radial velocities for HD 11964 using 1, 2, 3 and 4 planet models. Assuming that all the models are equally probable *a priori*, the three planet model was found to be ≥ 600 times more probable than the next most probable model which is a two planet model. The most probable model exhibits three periods of $38.02^{+0.06}_{-0.05}$, 360^{+4}_{-4} and 1924^{+44}_{-43} d. The small difference (1.3σ) between the 360 day period and one year suggests that it might be worth investigating the barycentric correction for the HD 11964 data. Based on our three planet model results, the remaining unaccounted for stellar jitter parameter is ~ 1.9m s^{-1}.

This research was supported in part by grants from the Canadian Natural Sciences and Engineering Research Council at the University of British Columbia.

BIBLIOGRAPHY

REFERENCES

1. Bretthorst, G. L., 1988, Bayesian Spectrum Analysis and Parameter Estimation, New York: Springer-Verlag
2. Butler, R. P., Wright, J. T., Marcy, G. W., Fischer, D. A., Vogt, S. S., Tinney, C. G., Jones, H. R. A., Carter, B. D., Johnson, J. A., McCarthy, C., and Penny, A. J., 2006, ApJ, 646, 505
3. Cumming, A. 2004, MNRAS, 354, 1165
4. Clyde, M., in ‘Statistical Challenges in Modern Astronomy IV,’ G. J. Babu and E. D. Feigelson (eds.), San Francisco:Astron. Soc. Pacific (in press 2006)
5. Ford, E. B., 2005, AJ, 129, 1706
6. Ford, E. B., 2006, ApJ, 620, 481
7. Ford, E. B., & Gregory, P. C., in ‘Statistical Challenges in Modern Astronomy IV’, G. J. Babu and E. D. Feigelson, eds, San Francisco:Astron. Soc. Pacific (in press 2006) astro-ph 0608328
8. Gregory, P. C., 2005a, ‘Bayesian Logical Data Analysis for the Physical Sciences: A Comparative approach with *Mathematica* Support’, Cambridge University Press
9. Gregory, P. C., 2005b, ApJ, 631, 1198
10. Gregory, P. C.,2005c, in ‘Bayesian Inference and Maximum Entropy Methods in Science and Engineering’, San Jose, eds. A. E. Abbas, R. D. Morris, J. P.Castle, AIP Conference Proceedings, 803, 139
11. Gregory, P. C., 2007, MNRAS, 374, 1321
12. Gregory, P. C., MNRAS (in press 2007), astro-ph/0709.0970
13. Jaynes, E.T. (1987), ‘Bayesian Spectrum & Chirp Analysis,’ in *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems*, ed. C.R. Smith and G.L. Erickson, D. Reidel, Dordrecht, p. 1.
14. Loredo, T., 2003, in ‘Bayesian Inference And Maximum Entropy Methods in Science and Engineering: 23rd International Workshop’, G.J. Erickson & Y. Zhai, eds, AIP Conf. Proc. 707, 330 (astro-ph/0409386).
15. Loredo, T. L. and Chernoff, D., 2003, in ‘Statistical Challenges in Modern Astronomy III’, E. D. Feigelson and G. J. Babu, eds , p. 57
16. Nelder, J. A. and Mead, R., Comput. J. 7, 308
17. Roberts, G. O., Gelman, A. and Gilks, W. R., 1997, Annals of Applied Probability, 7, 110
18. Valenti, J. A., & Fischer, D. A. 2005, ApJS, 159, 141
19. Udry, S., Bonfils, X., Delfosse, X., Forveille, T., Mayor, M., Perrier, C., Bouchy, F., Lovis, C., Pepe, F., Queloz, D., and Bertaux, J.-L., A&A in press 2007.