

## Signals embedded in the radial velocity noise\*

Mikko Tuomi<sup>1,2,a</sup>

<sup>1</sup>*University of Hertfordshire, Centre for Astrophysics Research, Science and Technology Research Institute, College Lane, AL10 9AB Hatfield, UK*

<sup>2</sup>*University of Turku, Tuorla Observatory, Department of Physics and Astronomy, Väisäläntie 20, 21500 Piikkiö, Finland*

**Abstract.** Bayesian data analysis techniques, together with suitable statistical models, can be used to obtain much more information from noisy data than the traditional frequentist methods. For instance, when searching for periodic signals in noisy data, the Bayesian techniques can be used to define exact detection criteria for low-amplitude signals – the most interesting signals that might correspond to habitable planets. We present an overview of Bayesian techniques and present detailed analyses of the HARPS-TERRA velocities of HD 40307, a nearby star observed to host a candidate habitable planet, to demonstrate in practice the applicability of Bayes' rule to astronomical data.

### 1. INTRODUCTION

The Bayes' rule of conditional probabilities has a long and controversial history in astronomy and has been used widely in some parts of the astronomical community, such as among cosmologists [15], but has only recently been applied to transit or radial velocity planet search data [12, 16, 18, 19]. The principles of Bayesian inference are remarkably simple but they still enable one to perform data analyses to be with little assumptions that might be limiting the quality of the inferred results when using more traditional *frequentist* data analysis methods.

The Bayes' rule can be written simply as

$$\pi(\theta|m) = \frac{l(m|\theta)\pi(\theta)}{P(m)}, P(m) > 0, \quad (1)$$

where random variable  $m$  denotes measurements and random variable  $\theta$  represents all the unknowns that are of principal interest to the researcher, i.e. the model parameters. In the above equations,  $\pi(\theta|m)$  is the posterior density of the parameter given the measurements;  $\pi(\theta)$  is the prior density, the information on  $\theta$  before the measurement  $m$  was made;  $l(m|\theta)$  is the likelihood function of the measurement, i.e. the statistical model or likelihood model; and  $P(m)$  is a constant, usually called the marginal integral, that makes sure that the posterior is a proper probability density in the sense that the product of the likelihood and prior integrates to unity over the parameter space  $\Theta$ . Thus, the marginal integral can be written as

$$P(m) = \int_{\theta \in \Theta} l(m|\theta)\pi(\theta)d\theta. \quad (2)$$

---

\*Readers are referred to an extended version of this article available at: [arxiv.org/abs/1301.3438](https://arxiv.org/abs/1301.3438).

<sup>a</sup>The author acknowledges the RoPACS (Rocky Planets Around Cool Stars), a Marie Curie Initial Training Network funded by the European Commission's Seventh Framework Programme. Author [mikko.tuomi@utu.fi](mailto:mikko.tuomi@utu.fi)

This is an Open Access article distributed under the terms of the Creative Commons Attribution License 2.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The integral in Eq. (2) is a key feature in the Bayesian model comparison framework. However, we do not discuss model selection further in this article and simply refer to some recent papers discussing the caveats of estimating  $P(m)$ , and methods of overcoming them and using it in practice [5, 6, 10, 17]

The Bayes' rule works the same way regardless of the number of measurements (or sets of measurements) available. For independent measurements  $m_i, i = 1, \dots, N$ , we can generalise Eq. (1) and write

$$\pi(\theta|m_1, \dots, m_N) = \frac{l(m_1, \dots, m_N|\theta)\pi(\theta)}{P(m_1, \dots, m_N)} = \frac{\pi(\theta) \prod_{i=1}^N l_i(m_i|\theta)}{P(m_1, \dots, m_N)}, \quad (3)$$

where the last equality holds only if the measurements are independent and thus  $l(m_1, \dots, m_N|\theta) = \prod_i l_i(m_i|\theta)$  holds.

The expression in Eq. (3) holds with remarkable generality in practice and its applicability is practically without limits. For instance, no assumptions are needed on the nature of the likelihood function, sometimes called the likelihood model, as long as it can be expressed mathematically. Furthermore, the measurements can be anything: from different sources such as radial velocity, transit, or astrometric surveys in the context of planet searches; time-series, individual numbers, matrices such as digital images, etc.; integers, real numbers, or even boolean operators. We also note that no assumptions are required about the nature of the probability densities of model parameters  $\theta$  – in fact these densities are the product of Bayesian inference in the sense that finding out their properties and estimating them is one of the primary goals of Bayesian analyses.

The prior probability density of the model parameters  $\pi(\theta)$  can also have a variety of shapes and natures but ideally it should be kept uninformative<sup>1</sup> w.r.t. the likelihood function in the sense that the likelihood is a narrower probability density (has e.g. greater Fischer information) in the parameter space. This means that the likelihood *overwhelms* the prior, though sometimes it is necessary to reverse the situation by using informative priors that in turn overwhelm the likelihood – especially if the likelihood function does not contain information on some aspects of the modelled system.

Priors are the only subjective part of Bayesian analyses – the rest consists of computations and is largely neglected throughout this article with few exceptions, though sometimes the computational difficulties are considerable problems on their own. However, some assumptions have to be made regarding prior probability densities. This is evident because a flat, i.e. uniform, prior is still a prior density (one that all frequentists using likelihoods assume). Similarly, fixing parameters (e.g. fixing eccentricity such that  $e = 0$  in planet searches) corresponds to a delta-function prior of the form  $\pi(e) = \delta(e)$  that has all the density at zero and none elsewhere. Furthermore, prior probabilities of different models do not have to be equal but some models might appear more realistic *a priori* and it would then be desirable to assign them greater prior probabilities in practice. Finally, the collection of candidate models is also selected *a priori* and because comparisons of these models is actually indistinguishable from comparisons of different prior models<sup>2</sup>, the choice of candidate models is of very high importance in Bayesian analyses of scientific data.

## 2. BAYESIAN INFERENCE WITH DYNAMICAL INFORMATION

In case of detections of exoplanets in systems of multiple planets, there is a very useful additional source of information – the Newtonian (or post-Newtonian if necessary) laws of motion. Because we

---

<sup>1</sup> Sometimes “uninformative” is used to denote a probability density that all scientists can agree on really being an uninformative one, i.e. representing “maximum amount of ignorance”. However, it remains a subjective issue and we do not define what we mean by uninformative here other than that the obtained results depend more on the likelihood than the prior in practice.

<sup>2</sup> For instance, comparison of two models containing  $k$  and  $k + 1$  Keplerian signals is equivalent to comparing two priors that are otherwise the same, but the other one has a delta-function form for the amplitude of the  $k + 1$ th signal such that  $\pi(K_{k+1}) = \delta(K_{k+1})$  that peaks at zero.

cannot expect to detect an unstable planetary system in practice – for the very reason that because of instability it would not exist in the first place unless we were observing it exactly at the time when its chaotic nature causes collisions or ejections of the bodies in the system – we can say that the prior probability of detecting something unstable is almost negligible.

In the Bayesian context, this additional information from dynamical analyses of such planetary systems can be written as

$$\pi(\theta|m, \mathcal{S}) = \frac{l(m, \mathcal{S}|\theta)\pi(\theta)}{P(m, \mathcal{S})} = \frac{l(m|\mathcal{S}, \theta)l(\mathcal{S}|\theta)\pi(\theta)}{P(m, \mathcal{S})}, \quad (4)$$

where  $\mathcal{S}$  denotes this “dynamical information” in terms of stability constraints of the system. Because Newton’s laws, or more generally the laws of gravity, do not depend on the measurements we have possibly obtained but the results we obtain must depend on them (unless we have found something that violates them – a significant discovery indeed), we have written the last equality in the above equation. In particular, the likelihood containing the dynamical information,  $l(\mathcal{S}|\theta)$ , needs to be defined in some suitable way to be able to use the Eq. (4).

## 2.1 Analytical criteria for dynamical stability

The dynamical likelihood in Eq. (4) can be written simply by using some simple stability criteria and by defining the likelihood in such a way that it is negligible for systems not satisfying these criteria while it has a non-negligible constant value for systems that do satisfy the criteria.

One simple such criterion is that the planets in a given system should not experience orbital crossings. Assuming co-planarity of the orbits, this leads to an approximate criterion for the eccentricities and semi-major axes ( $a$ ) of the planets  $i$  and  $j$  such that  $a_i(1 + e_i) < a_j(1 - e_j)$  when the  $i$ th planet is the inner one and the  $j$ th is the outer one. This criterion can further be improved by requiring, for instance, that the planets stay out of each others Hill spheres at all times. However, as a very rough criterion, this expression can only be used as a starting point and cannot be expected to provide much additional constraints in practice unless the proposed planetary system is clearly unstable and one of some of the corresponding signals are actually caused by something else but planets.

A slightly more useful criterion is e.g. the approximated Lagrange stability criterion for two subsequent planets defined as [2]

$$\alpha^{-3} \left( \mu_1 - \frac{\mu_2}{\delta^2} \right) (\mu_1 \gamma_1 + \mu_2 \gamma_2 \delta)^2 > 1 + \mu_1 \mu_2 \left( \frac{3}{\alpha} \right)^{4/3}, \quad (5)$$

where  $\mu_i = m_i M^{-1}$ ,  $\alpha = \mu_1 + \mu_2$ ,  $\gamma_i = \sqrt{1 - e_i^2}$ ,  $\delta = \sqrt{a_2/a_1}$ ,  $M = m_\star + m_1 + m_2$ ,  $m_i$  is the planetary mass, and  $m_\star$  is stellar mass. Again, we simply set  $l(\mathcal{S}|\theta) = c$  if the criterion is satisfied and  $l(\mathcal{S}|\theta) = 0$  otherwise (we call the set of stable orbits in the parameter space  $B \subset \Theta$ ). This criterion can also be used to estimate briefly whether a given observed system is indeed stable or not [16, 18, 19].

However, the above criteria do not take into account e.g. orbital resonances and are only very approximate criteria in practice. Therefore, better methods are needed to take into account the dynamical information in practice.

## 2.2 Dynamical information from posterior samplings

If the shape of the posterior density of the model parameters is already known and a sample has been drawn from it e.g. using one of the various posterior sampling algorithms [8, 9, 13], we can use this sample in assessing the shape of the dynamical likelihood function  $l(\mathcal{S}|\theta)$ . We assume that such a statistically representative sample has been drawn from the posterior density and that it exists in a form of a Markov chain such that  $\theta_i \sim \pi(\theta|m)$ ,  $i = 1, \dots, K$ . These  $K$  parameter vectors can now be used as initial states of direct numerical integrations of planetary orbits.

As performing numerical integrations of planetary orbits is a simple task with suitable algorithms, such as the Bulirsch-Stoer algorithm [4], we assume that each  $K$  initial states of the system have been integrated from  $t_0 = 0$  to  $t_N = T$ , where the duration of these integrations,  $T$ , is typically set to millions to hundred million years. Also, some criteria have been chosen to determine whether a system is stable or not (i.e. collisions, increasing  $a_i$  without limit for some  $i$ , etc.). We divide the duration  $T$  into  $N$  intervals such that  $t_i = iT/N$ . With this notation, and with  $\theta_i$  as an initial state of orbital integrations for all  $i = 1, \dots, K$ , we have  $K$  chains of  $N$  vectors that we denote as  $\theta_i^j$ ,  $j = 1, \dots, N$ , where  $\theta_i^j = \theta_i(t_j)$ . Each of these chains of vectors represent the orbital evolution of the system as a function of time.

Hence, we can approximate the posterior probability of finding  $\theta \in I_l \subset \Theta$  of dynamical information and data for each  $n$ -interval  $I_l$  as

$$\begin{aligned} P(\theta \in I_l | \mathcal{S}, m) &\approx \frac{1}{K} \sum_{i=1}^K P(\theta \in I_l | \mathcal{S}, \theta_i) \\ &\approx \frac{1}{KN} \sum_{i=1}^K \sum_{j=1}^N \mathbf{1}_l(\theta_i^j) \mathbf{1}(\theta_i^j), \end{aligned} \quad (6)$$

where

$$\mathbf{1}_l(\theta) = \begin{cases} 1 & \text{if } \theta \in I_l \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \mathbf{1}(\theta) = \begin{cases} 1 & \text{if } \theta \in B \\ 0 & \text{otherwise} \end{cases}. \quad (7)$$

In the above expression, the second indicator function denotes whether a given chain of vectors corresponds to an unstable configuration or not, i.e. whether any member of the chain is outside the set of stable orbits  $B$ . The first indicator function denotes the probability of finding the parameter vector in the given  $n$ -interval and weights each part of the set  $B$  with this probability. Thus, this expression sets negligible weight to unstable configurations and assigns nonzero probabilities for the stable ones and can be interpreted as the desired posterior probability density of the model parameters given the measurements and the dynamical information in Eq. (4). Applying this equation is not difficult in practice given that one can increase  $K$ ,  $N$ , and  $T$  such that the dynamical analyses help in ruling out unstable areas of the parameter space and the results become statistically representative [18].

### 3. PLANET DETECTION CRITERIA

When the goal is to find planetary signatures from noisy data, the criteria for detecting confidently such signatures are important in order to extract weak signals hidden in the noise from the data and to avoid detecting false positives or signals arising from other periodic or quasi-periodic sources such as the movement of the celestial bodies of the Solar System; stellar rotation coupled with magnetic activity, starspots, and other activity-induced phenomena of the stellar surface; of instrumental source as a consequence of lack of sufficient instrumental stability or problems in calibration.

The first part of these criteria can be summarised briefly as follows [16]. First, it is required that the posterior probability of a model with  $k + 1$  signals is significantly greater than that of a model with only  $k$  signals. This can be expressed as  $P(\mathcal{M}_k | m) > \alpha P(\mathcal{M}_{k-1} | m)$  for a selected threshold  $\alpha > 1$ . Several authors use  $\alpha = 150$  that has been interpreted as corresponding to “significant evidence” [7, 10, 16] but the exact choice of this threshold remains subjective and depends only on how confident one wants to be when making decisions based on model comparisons.

The second criterion is that, in case of radial velocities that will be used as examples below, the radial velocity amplitudes of all signals are statistically distinguishable from zero, i.e. their Bayesian

credibility sets (BCSs)<sup>3</sup> do not overlap with zero for a selected threshold  $\delta \in [0, 1]$ . In practice this means that the signals are constrained from below in a sense that they cannot be said to be consistent with zero, which in turn would imply that they are not significantly present in the data. In practice there are several examples of signals that do not satisfy this criterion or satisfy it only just and it should be taken into account when detecting periodic variations of planetary origin [16, 18, 19].

In addition to having a well-constrained amplitude, the period of a signal also has to be well-constrained in the sense that it has clear lower and upper limits in the parameter space (with suitable credibility). If this was not the case, a signal could not be said to be periodic and could not be readily interpreted as being caused by the Doppler effect of planetary origin. However, this one, and the other two above criteria are only mathematical ones applicable to any detection problem of periodic signals. There are also other, physical, criteria that have to be satisfied to be able to claim that a given signal is of planetary origin. However, we do not discuss these criteria further here for brevity.

#### 4. ANALYSIS OF HD 40307 RADIAL VELOCITIES

HD 40307 is known to be a system of three super-Earths with orbital periods of 4.31, 9.62, and 20.44 days and minimum masses of 4.3, 7.0, and 10.5  $M_{\oplus}$ , respectively [11, 18]. Obtaining the signals of these three super-Earths from HARPS-TERRA velocities [1, 18] is rather easy. However, the three planetary companions were reported based on analyses of binned HARPS-CCF velocities [11]. Therefore, binning the data, and using a sub-optimal spectral reduction method, i.e. the cross-correlations function (CCF) method [1], might disable a reliable detection of additional low-amplitude signals of planetary origin that might still be present in the data in addition to the three clear signatures of super-Earths.

Indeed, moving sequentially by subtracting the signals corresponding to the greatest posterior probabilities (and strongest periodogram powers) from the data, and modelling the noise carefully with an MA(3) model enables the detections of three new signals in the HD 40307 velocities [18].

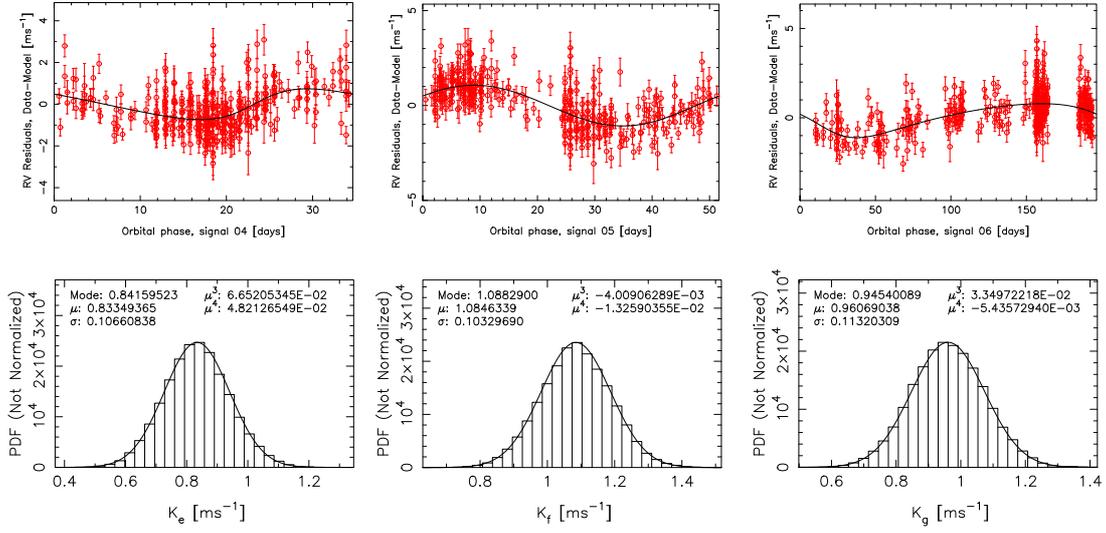
The residual periodograms of the three-Keplerian model for the unbinned velocities indicates that there are two powers exceeding the 10% FAP in the data. However, without binning, and after actually throwing part of the spectral information away in the sense that the velocities are calculated only from the redmost part of the HARPS spectra likely containing less activity-induced variation [18], the significances of the powers increase and indicate clearly that e.g. the 51-day peak is a significant periodicity in the data.

According to Bayesian analyses of the velocities of HD 40307, there are actually three additional periodicities in the data that satisfy all the detection criteria presented in Section 3. This implies that there are actually six planets orbiting HD 40307 [18]. We demonstrate this by showing the phase-folded signals of these candidates and the posterior probability densities of their radial velocity amplitudes in Fig. 1. Interestingly, the outermost one of these planet candidates is in the liquid water habitable zone of the star [14] and appears to be one of the best candidate for a habitable world out of those detected.

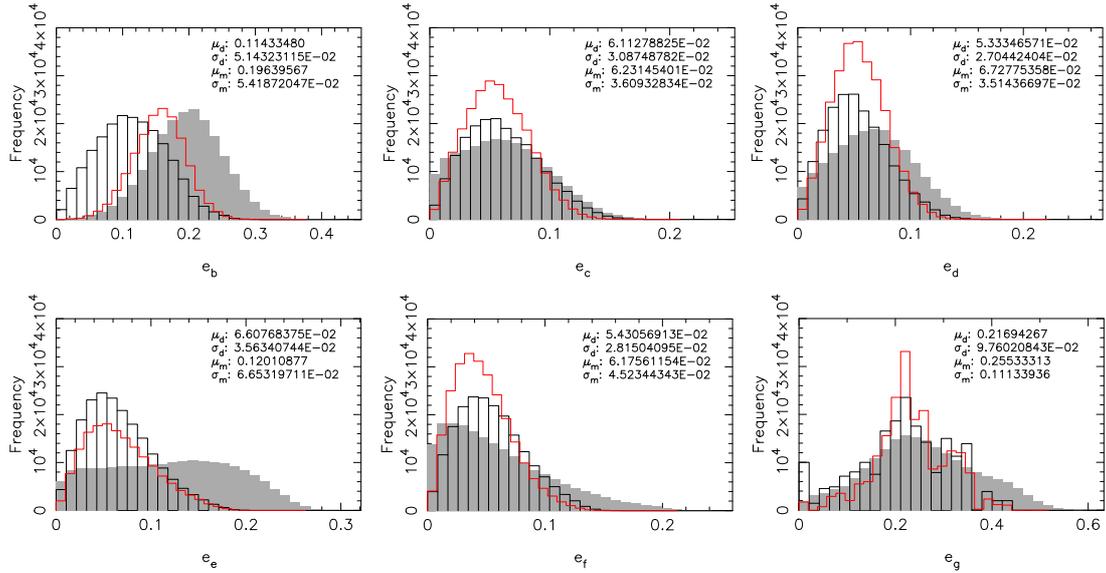
When analysing the HARPS-TERRA radial velocities of HD 40307, we did not use uninformative (in any practical meaning of uninformative) prior densities for the orbital eccentricities. Instead, we estimated that close-circular orbits of planets in any given system are more probable *a priori* than more eccentric ones, and that extremely eccentric orbits, with e.g.  $e > 0.6$ , are very improbable in practice because they would certainly make such systems of low-mass planets unstable due to orbital crossings. Furthermore, it is a well-known (but rather poorly understood) fact that analyses of radial velocity measurements containing signals with low eccentricities, tend to yield overestimates for the eccentricities [20]. For these reasons, we chose the prior such that  $\pi(e) \propto \mathcal{N}(0, 0.3^2)$  [16, 18].

---

<sup>3</sup> The BCS, a subset of the parameter space  $\Theta$ , of threshold  $\delta$  is defined for the posterior density  $\pi(\theta|m)$  as  $\mathcal{D}_{\delta} = \{\mathcal{D}_{\delta} \subset \Theta : \int_{\theta \in \mathcal{D}_{\delta}} \pi(\theta|m) = \delta, \pi(\theta|m)|_{\theta \in \partial \mathcal{D}_{\delta}} = c\}$ .



**Figure 1.** Phase-folded orbits and velocity amplitudes of the three new candidates around HD 40307.



**Figure 2.** Posterior densities of the eccentricities (white histograms) of the six planet candidates orbiting HD 40307, dynamical likelihoods (grey histograms), and their combination (red curve).

Furthermore, additional constraints for eccentricities can be readily obtained from dynamical analyses of the planetary orbits. Using the method described briefly in Eq. (6), we plotted the resulting eccentricity posteriors, dynamical likelihoods, and their combinations in Fig. 2. This corresponds to Bayesian updating of the posterior density from the data by the dynamical likelihood (grey histograms in Fig. 2) and obtaining an updated posterior given both data and dynamical information (red curves in Fig. 2).

It can be seen in Fig. 2 that the highest eccentricities of all the candidates in the HD 40307 system are basically penalised by the dynamical likelihood—because they correspond to unstable orbital configurations. This example shows that information from dynamical analyses can be readily incorporated into Bayesian analyses of radial velocity data and can be used to better constrain the planetary orbits.

## 5. CONCLUSIONS

The Bayes' rule of conditional probabilities is an efficient and very practical way of inferring important information from noisy data. We have presented the typical formulations of Bayesian data analysis problems and showed some examples demonstrating the effectiveness of the combination of Bayes' rule and realistic noise models in practice.

In particular, when detecting periodic signals in noisy data, Bayesian formulation allows the definition of exact detection criteria that have to be satisfied to be able to conclude that a signal exists beyond reasonable doubt. Moreover, these criteria [16] can be expressed in a natural way and have intuitive probabilistic interpretations that leave few questions unanswered. When applying these criteria, the greatest caveat is the ability to assess whether the statistical models used to analyse the data, i.e. the likelihood and prior models, contain a sufficiently representative collection of models. If the detection criteria are satisfied for a variety of such models, there is little doubt about the existence of the signals. However, if this is the case for only one or few models, the physical origin of noise has to be understood better in order to make conclusions regarding the number of significant signals in the data.

The example of HD 40307 (Section 4) shows that a combination of priors, suitable likelihoods, and information from dynamical analyses can provide much more information on a given planetary system than any traditional periodogram-based methods. Furthermore, the combination of these sources of information is a natural application of the Bayes' rule and enables the combinations of any other sources of data in addition to these as well. Therefore, as the only logically consistent framework of data analysis [3], the applications of Bayes' rule are plenty and it should be added to the basic toolbox of all astronomers in practice.

## References

- [1] Anglada-Escudé, G. & Butler, R. P. 2012, *ApJS*, **200**, 15
- [2] Barnes, R. & Greenberg, R. 2006, *ApJ*, **647**, L163
- [3] Berger, J.O. 1980, *Statistical Decision Theory and Bayesian Analysis* (Springer)
- [4] Bulirsch, R. & Stoer, J. 1966, *Numer. Math.*, **8**, 1
- [5] Chib S. & Jeliazkov I. 2001, *J. Am. Stat. Ass.*, **96**, 270
- [6] Clyde, M.A., Berger, J.O., Bullard, F., et al. 2007, *Statistical Challenges in Modern Astronomy IV*, Babu, G.J. & Feigelson, E.D. (eds.), *ASP Conf. Ser.*, **371**, 224
- [7] Feroz, F., Balan, S.T., & Hobson, M.P. 2011, *MNRAS*, **415**, 3462
- [8] Haario, H., Saksman, E., & Tamminen, J. 2001, *Bernoulli*, **7**, 223
- [9] Hastings, W. 1970, *Biometrika* **57**, 97
- [10] Kass, R.E. & Raftery, A.E. 1995, *J. Am. Stat. Ass.*, **90**, 773
- [11] Mayor, M., Udry, S., Lovis, C., et al. 2009, *A&A*, **493**, 639
- [12] Parviainen, H., Deeg, H., Belmonte, J.A. 2012, *A&A*, accepted
- [13] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., et al. 1953, *J. Chem. Phys.*, **21**, 1087
- [14] Selsis, F., Kasting, J.F., Levrard, B., et al. 2007, *A&A*, **476**, 1373
- [15] Trotta, R. 2008, *Contemporary Physics*, **49**, 71
- [16] Tuomi, M. 2012, *A&A*, **543**, A52

- [17] Tuomi, M. & Jones, H.R.A. 2012, A&A, **544**, A116
- [18] Tuomi, M., Anglada-Escudé, G., Gerlach, E., et al. 2013a A&A, **549**, A48
- [19] Tuomi, M., Jones, H.R.A., Jenkins, J.S., et al. 2013b, A&A, accepted (arXiv:1212.4277 [astro.ph-EP])
- [20] Zakamska, N.L., Pan, M., & Ford, E.B. 2011, MNRAS, **410**, 1895