Bias and robustness of eccentricity estimates from radial velocity data

Nathan Hara, G. Boué, J Laskar, J-B Delisle, N Unger

To cite this version:

Nathan Hara, G. Boué, J Laskar, J-B Delisle, N Unger. Bias and robustness of eccentricity estimates from radial velocity data. Monthly Notices of the Royal Astronomical Society, 2019, 489 (1), pp.738-762. ⟨10.1093/mnras/stz1849⟩. ⟨hal-03144362⟩

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **489,** 738–762 (2019)
Advance Access publication 2019 July 8

doi:10.1093/mnras/stz1849

Bias and robustness of eccentricity estimates from radial velocity data

Nathan C. Hara,[1,2★†] G. Boué,[2] J. Laskar,[2] J.-B. Delisle[1] and N. Unger[3]

[1] *Observatoire de Genève, Université de Genève, 51 ch. des Maillettes, CH-1290 Versoix, Switzerland*
[2] *ASD/IMCCE, CNRS-UMR8028, Observatoire de Paris, PSL, UPMC, 77 Avenue Denfert-Rochereau, F-75014 Paris, France*
[3] *Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Intendente Guiraldes 2160 - Ciudad Universitaria, C1428EGA, Argentina*

Accepted 2019 July 1. Received 2019 June 27; in original form 2018 March 26

ABSTRACT

Eccentricity is a parameter of particular interest as it is an informative indicator of the past of planetary systems. It is however not always clear whether the eccentricity fitted on radial velocity data is real or if it is an artefact of an inappropriate modelling. In this work, we address this question in two steps: we first assume that the model used for inference is correct and present interesting features of classical estimators. Secondly, we study whether the eccentricity estimates are to be trusted when the data contain incorrectly modelled signals, such as missed planetary companions, non-Gaussian noises, correlated noises with unknown covariance, etc. Our main conclusion is that data analysis via posterior distributions, with a model including a free error term gives reliable results provided two conditions. First, convergence of the numerical methods needs to be ascertained. Secondly, the noise power spectrum should not have a particularly strong peak at the semiperiod of the planet of interest. As a consequence, it is difficult to determine if the signal of an apparently eccentric planet might be due to another inner companion in 2:1 mean motion resonance. We study the use of Bayes factors to disentangle these cases. Finally, we suggest methods to check if there are hints of an incorrect model in the residuals. We show on simulated data the performance of our methods and comment on the eccentricities of Proxima b and 55 Cnc f.

Key words: methods: data analysis – techniques: radial velocities – planets and satellites: dynamical evolution and stability – planets and satellites: fundamental parameters – methods: analytical.

1 INTRODUCTION

The nearly coplanar and circular orbits of the Solar system have long been an argument in favour of Laplace and Kant's theory of formation of planets in a disc (Swedenborg 1734; Kant 1755; Laplace 1796). The first observations of exoplanets suggested that such low eccentricities were rather the exception than the rule. The 'eccentricity problem', along with the need to envision migration scenarios for hot Jupiters, triggered several theoretical studies which explored migration scenarios after the disc has dissipated. The predictions of these models were compared to measured eccentricities. For instance Jurić & Tremaine (2008) and Petrovich, Tremaine & Rafikov (2014) evaluate the likelihood of formation scenarios of hot and warm Jupiters through their agreement with observed eccentricity distributions.

For a radial velocity data set on a given star, one wants to extract two pieces of information about the eccentricity. First, a best candidate value (the estimation problem) and what are the eccentricity values that are incompatible with the data (the hypothesis testing problem). It is in particular interesting to test if an eccentricity is non-zero. Both problems do not have completely obvious solutions. For instance it is known since Lucy & Sweeney (1971) that when the true eccentricity is small, its least square estimate is biased upwards. Other aspects of the estimation and hypothesis testing problems have been addressed in the exoplanet community.

In Shen & Turner (2008) the bias was found to depend on the signal-to-noise ratio (denoted by SNR) as well as on the time span of the observations. This was confirmed by Zakamska, Pan & Ford (2011), which further showed that the bias depends on the phase coverage, and updated the Lucy & Sweeney (1971) null hypothesis test to determine if a null eccentricity can be rejected or not. They also propose metrics for evaluating the quality of a data set. O'Toole et al. (2009a) showed that error bars on eccentricity from least square are underestimated by a factor 5–10. Brown (2017) shows that there might be orbits with very high eccentricities with similar goodness-of-fit as a low eccentricity one.

★ E-mail: nathan.hara@unige.ch
† NCCR PlanetS CHEOPS Fellow, Switzerland.

Published by Oxford University Press on behalf of the Royal Astronomical Society

Pont et al. (2011), Husnoo et al. (2011), and Husnoo et al. (2012) used Bayesian Information Criterion (BIC) to confirm non-zero eccentricities. More recently, Bonomo et al. (2017a,b) assessed the evidence in favour of eccentric solutions with Bayes factors. A Bayesian test with a physically motivated prior on eccentricity was devised by Lucy (2013). Also, Anglada-Escudé, López-Morales & Chambers (2010), Wittenmyer et al. (2013), and Kürster et al. (2015) note that two planets in 2:1 mean motion resonances can appear as an eccentric planet, and propose ways to disentangle those cases. This problem has also been studied in Boisvert, Nelson & Steffen (2018), Nagel et al. (2019), and Wittenmyer et al. (2019).

The fact that eccentricity estimates can be spuriously high for a given planet gives reasons for concern on the eccentricity distributions. The estimation of those has been tackled in Hogg, Myers & Bovy (2010), which computes the posterior of the eccentricity probability distribution itself. Zakamska et al. (2011) consider the accuracy of the eccentricity catalogues obtained by Bayesian point estimates. They show that for single planet populations contaminated by white noise, estimating the eccentricity via the maximum of the marginalized posterior distribution of eccentricity with a free jitter term gives satisfactory retrieval of the input population. Furthermore, it has been noted by Cumming (2004) that high eccentricity orbits $\gtrsim$0.6 are more difficult to detect at fixed semi-amplitude. For a fixed mass, the detection bias is less strong (Shen & Turner 2008).

We contribute to this series of work by studying in depth the bias and robustness of eccentricity estimates from radial velocity data (note that our analysis also applies to astrometric measurements and to estimates of semi-amplitude and inclination). We proceed in two steps. First, we highlight key properties of classical estimators, in order to have a consistent view of eccentricity estimation. The following questions are then considered: is the eccentricity inference robust to modelling errors? By that, we refer to wrong noise models, planetary companion too small to be detected, etc. If not, how to mitigate the problem? One could encounter a situation similar to the spectroscopic binaries, where proximity effects or gas streams lead to spurious high-significance eccentricities if not properly accounted for (Lucy 2005).

The article is structured as follows. In Section 2 we study the behaviour of eccentricity estimates when the model is correct. The least square estimates as well as Bayesian ones are studied, and it is shown that the latter are less biased at low eccentricity. The problem of spurious local χ^2 minima at high eccentricity is also tackled, in particular through the Proxima b case (Anglada-Escudé et al. 2016). Section 3 is devoted to studying the robustness of the estimates when the numerical method, the model or the prior is poorly chosen. Finally in Section 4, we consider ways to check the validity of a noise model. Our methods are illustrated with the 55 Cnc HIRES data set in Section 5. In Section 6, we conclude with a step-by-step procedure to obtain reliable eccentricities and present perspectives for future work.

2 POINT AND INTERVAL ECCENTRICITY ESTIMATES

2.1 Problem definition

2.1.1 *Point and interval estimates*

Let us first define the problem under study precisely. Some generic symbols, used throughout the text, are summarized in Table 1. Let us consider a time series of N observations, modelled as a vector $\boldsymbol{y} = (y(t_k))_{k=1..N}$, such that

$$\boldsymbol{y}(\boldsymbol{t}) = \boldsymbol{f}(\boldsymbol{t}, \boldsymbol{\theta}) + \boldsymbol{\epsilon} \tag{1}$$

where $\boldsymbol{t} = (t_k)_{k=1..N}$is the vector of measurement times, f is a deterministic model depending on parameters $\boldsymbol{\theta} \in \mathbb{R}^p$ that include the eccentricity e and $\boldsymbol{\epsilon}$ is a random variable modelling the noise. Denoting by $e_{\rm t}$ the true value of e, the estimation problem consists in finding a function $\widehat{e}$ of the data whose output is close to $e_{\rm t}$ in a sense chosen by the data analyst.

Table 1. List of symbols.

$\boldsymbol{\theta}$	Vector of parameters
$\boldsymbol{\theta}_{\rm t}$	True value of the vector of parameters
$\widehat{\boldsymbol{\theta}}$	Estimator of $\boldsymbol{\theta}$
$\boldsymbol{f}(\boldsymbol{t}, \boldsymbol{\theta})$	Deterministic model sampled at times $\boldsymbol{t} = (t_k)_{k=1..N}$ and of parameters $\boldsymbol{\theta}$ (orbital parameters plus possibly offset, trend, etc.)
$p(\boldsymbol{\theta})$	Probability density of $\boldsymbol{\theta}$, $\Pr\{\boldsymbol{\theta} \in \Theta\} = \int_\Theta p(\boldsymbol{\theta})\,{\rm d}\boldsymbol{\theta}$ for some measurable set Θ
$\mathbb{E}\{\boldsymbol{\theta}\}$	Mathematical expectancy of the random variable $\boldsymbol{\theta}$
k, h	$k = e\cos\omega$, $h = e\sin\omega$

If the mean value of $\widehat{e}(\boldsymbol{y})$ ($\mathbb{E}\{\widehat{e}(\boldsymbol{y})\}$)is not equal to, the estimator is said to be biased and

$$b_{\widehat{e}}(\boldsymbol{\theta}_{\rm t}) = \mathbb{E}\{\widehat{e}(\boldsymbol{y})\} - e_{\rm t} \tag{2}$$

is called the bias of the estimator $\widehat{e}$ in $\boldsymbol{\theta}_{\rm t}$, where $\boldsymbol{\theta}_{\rm t}$ is the true value of $\boldsymbol{\theta}$. A common metric for the accuracy of an estimator is the mean squared error (MSE). Denoting by ${\rm Var}\{\widehat{e}(\boldsymbol{y})\}$ the variance of $\widehat{e}(\boldsymbol{y})$, it is defined as

$$\mathrm{MSE} := \mathbb{E}\{(\widehat{e}(\boldsymbol{y}) - e_{\rm t})^2\} = {\rm Var}\{\widehat{e}(\boldsymbol{y})\} + b_{\widehat{e}}(\boldsymbol{\theta}_{\rm t})^2. \tag{3}$$

The other problem we consider is to have a testing procedure to reject or not certain values of the eccentricity. We are now interested in rejecting the hypothesis that $e \in C$ where C is a subset of [0, 1]. More precise definitions are given in the relevant Sections 2.2.4 and 2.3.2.

In the present Section 2, we describe tools for the estimation and hypothesis testing problems and present some of their properties. In Section 3, we study the reliability of these tools when the model is incorrect. By convention, in the following, radial velocity signals are in m s^{-1}. The analysis is unchanged for other units as long as the signal-to-noise ratio is identical.

2.1.2 *Model*

The concern of this work is the estimation of eccentricity from radial velocity data. The model of a radial velocity planetary signal is recalled below

$$\begin{aligned} f(t, e, K, P, \omega, M_0) = K(&\cos(\omega + \nu(t, e, P, \omega, M_0)) \\ &+ e\cos\omega) \end{aligned} \tag{4}$$

$$\cos\nu = \frac{\cos E - e}{1 - e\cos E} \tag{5}$$

$$\sin\nu = \frac{\sqrt{1-e^2}\sin E}{1 - e\cos E} \tag{6}$$

$$E - e\sin E = M_0 + \frac{2\pi}{P}t. \tag{7}$$

The symbols t, e, K, P, ω, M_0 designate respectively the measurement time, eccentricity, semi-amplitude, period, argument of periastron, and mean anomaly at $t = 0$. The symbols E and ν denote the eccentric and true anomalies.

We assume a Gaussian noise model, such that the likelihood function is

$$p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\beta}) = \frac{1}{\sqrt{(2\pi)^N|\mathbf{V}(\boldsymbol{\beta})|}}\,\mathrm{e}^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{t},\boldsymbol{\theta}))^T\mathbf{V}(\boldsymbol{\beta})^{-1}(\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{t},\boldsymbol{\theta}))} \tag{8}$$

where $\boldsymbol{f}(\boldsymbol{t},\boldsymbol{\theta})$ is a sum of Keplerian functions defined as (4) possibly plus some other model features (offset, trend, etc.). The covariance matrix $\mathbf{V}$ is parametrized by $\boldsymbol{\beta}$ and the suffix T denotes the matrix transposition. The explicit expression of $\boldsymbol{f}(\boldsymbol{t},\boldsymbol{\theta})$ and $\mathbf{V}(\boldsymbol{\beta})$ will be given in the relevant sections.

The features of least square and Bayesian estimates are now studied respectively in Sections 2.2 and 2.3.

2.2 Least square estimate

2.2.1 *Bias of the non-linear least square*

A common parameter estimator is the maximum likelihood $\widehat{\theta}_{\mathrm{ML}}$. For the model of equation (8),

$$\widehat{\theta}_{\mathrm{ML}} = \arg \max_{\boldsymbol{\theta}\in\Theta,\boldsymbol{\beta}\in B} p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\beta}). \tag{9}$$

When the parameters of the covariance, $\boldsymbol{\beta}$, are fixed, maximizing the likelihood comes down to the least square problem,

$$\widehat{\theta}_{\mathrm{LS}}(\boldsymbol{y}) = \arg\min_{\boldsymbol{\theta}\in\Theta}(\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{\theta}))^T\mathbf{V}^{-1}(\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{\theta})). \tag{10}$$

When the model $\boldsymbol{f}(\boldsymbol{\theta})$ depends linearly on $\boldsymbol{\theta}$, the least square estimate is unbiased. This is in general untrue when $\boldsymbol{f}(\boldsymbol{\theta})$ is non-linear, which has been noted for instance by Hurwicz (1950) and discussed in Hartley (1964), Bates & Watts (1980), Cook & Witmer (1985), and Firth (1993). The model we are concerned with (equations 4–7) is non-linear, and indeed eccentricities obtained by least square are biased.

2.2.2 *Bias and uncertainty at low eccentricity*

We begin with a numerical experiment. We generate Keplerian signals of eccentricity 0.06, 0.5, and 0.9 with fixed ω, M_0, and $K = 3$ m s^{-1}. The array of time $\boldsymbol{t}$ is the 30 first measurements of GJ 876 (Correia et al. 2010). We generate 1000 realizations of a white Gaussian noise with standard deviation of 1 m s^{-1}. For each realization of the noise, a non-linear Keplerian model and a constant are fitted with a Levenberg–Marquardt algorithm. The minimization step is scaled so that the eccentricity never exceeds 0.999. The values of the estimates ($\widehat{k} := \widehat{e\cos\omega}$, $\widehat{h} := \widehat{e\sin\omega}$) are reported as crosses (red, purple and blue resp. for $e = 0.06$, 0.5, and 0.9) in Fig. 1. The distributions of $\widehat{k}$ and $\widehat{h}$ are fairly isotropic for $e_{\mathrm{t}} = 0.06$ and $e_{\mathrm{t}} = 0.5$. For $e = 0.9$ there seems to be more complicated phenomena at work. The distribution has no circular symmetry and in some cases $\widehat{e}$ is stuck at its maximal value, 0.999 (see Section 2.2.3).

In Appendix A, it is shown that provided e is small enough ($\leq$0.2) and the number of observation is sufficient, $\widehat{k}$ and $\widehat{h}$, follow independent Gaussian laws of same variance. This property allows us to understand the bias qualitatively. In Fig. 2, the distribution of eccentricity estimates is represented. The pair of true values k_{t}, h_{t} is represented by a yellow star. The bold black line delimits the region where 99 per cent of the estimates are located. When the estimate falls in the blue-coloured region, the eccentricity is underestimated. When it falls in the red-coloured region, the eccentricity is overestimated. As the volume of higher eccentricity models is larger in the vicinity of k_{t}, h_{t}, the eccentricity is more probably overestimated. Informally, there are more and more models with eccentricity e as e grows.

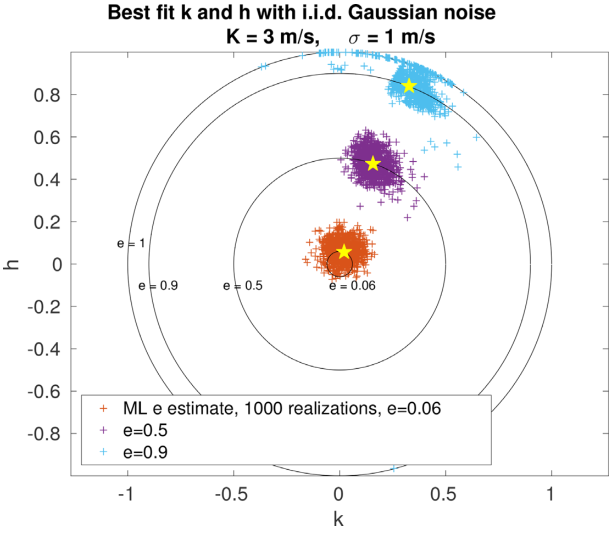

Figure 1. Estimates of $k = e\cos\omega$ and $h = e\sin\omega$ for various true eccentricities. Each cross represents an estimate $\widehat{k}$ and $\widehat{h}$ obtained with a Keplerian model least-squares fit for a given noise realization. The yellow stars represent the true (k, h). Estimates for $e = 0.06$, 0.5, and 0.9 are respectively in red, purple, and blue (3 × 1000 = 3000 estimates in total on each figure). $\sigma = 1$ m s^{-1} and $K = 3$ m s^{-1}.

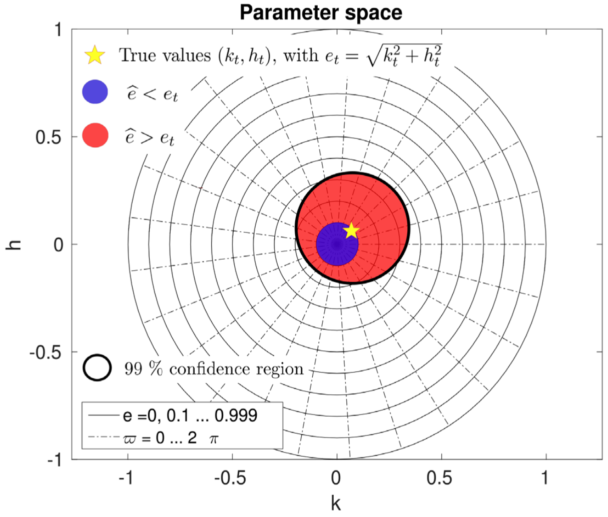

Figure 2. Representation of the estimates $\widehat{k} = \widehat{e}\cos\widehat{\omega}$ and $\widehat{k} = \widehat{e}\sin\widehat{\omega}$ when $\widehat{k}$ and $\widehat{h}$ have a joint Gaussian distributions. The yellow star represents the true value of k and h, the bold black line encircles the region where 99 per cent of the estimates are found. The red and blue regions represent respectively the regions where the eccentricity estimates are overestimated and underestimated.

Furthermore, it is possible to obtain an analytical approximation of the bias. Since $\widehat{k}$ and $\widehat{h}$ approximately follow a joint Gaussian distribution with same variances, $\widehat{e} = (\widehat{k}^2 + \widehat{h}^2)^{1/2}$ follows a Rice distribution, as noted in Shen & Turner (2008). Interestingly enough, the Rice distribution appears as a very good model for the eccentricity densities of the inner planets of the Solar system, resulting from chaotic diffusion (Laskar 2008). Supposing that the measurement noise is white with standard deviation σ and that p parameters are fitted, within our approximation, $\widehat{k}$ and $\widehat{h}$ have a standard deviation $\sigma_k = \sigma_h = \sigma/K_\mathrm{t}\sqrt{2/(N-p)}$. Defining the SNR S as

$$S := \frac{1}{\sigma_k} = \frac{K_\mathrm{t}}{\sigma}\sqrt{\frac{N-p}{2}} \tag{11}$$

and denoting by e_t the true eccentricity, the bias is

$$b(e_\mathrm{t}, S) \approx \frac{1}{S}\sqrt{\frac{\pi}{2}}L_{1/2}\left(-\frac{S^2e_\mathrm{t}^2}{2}\right) - e_\mathrm{t}, \tag{12}$$

where $L_{1/2}$ is the Laguerre polynomial of order 1/2. When $e_\mathrm{t} = 0$, the eccentricity follows a Rayleigh distribution and equation (12) reduces to a very simple expression,

$$b(0, S) \approx \sqrt{\frac{\pi}{2}}\frac{1}{S} = \frac{\sigma}{K_\mathrm{t}}\sqrt{\frac{\pi}{N-p}} = \sqrt{\frac{\pi}{4-\pi}}\sigma_{\widehat{e}}. \tag{13}$$

Equation (13) is identical to equation 18 of Lucy & Sweeney (1971) except that we are able to derive the effect of the correlations between parameters on the SNR through the term $-p$ (see Appendix A for justification).

Formula (13) is useful to see a few trends: the bias is proportional to the uncertainty on k and h, which is proportional to the inverse of K_t and $\sqrt{N-p}$. As a consequence, the bias increases as the SNR decreases, i.e. as σ increases or as K_t or N decrease. This is also found by simulations in Shen & Turner (2008) and Zakamska et al. (2011) for Bayesian estimates. We add that increasing the number of fitted parameters p, increases the bias.

There are particular cases where the correlations between parameters drastically increase the uncertainties on k and h and therefore increase the bias, so that formula (12) should be taken as a lower bound. However, the fact that the bias is approximately proportional to the standard deviation with a factor $\sqrt{\pi/(4-\pi)}$, as in equation (13) stays true. This fact is remarkable because it means that the accuracy of the estimate [seen as the MSE (3)] is proportional to its precision (seen as the standard deviation $\sigma_{\widehat{e}}$). In Appendix A2, we show that poor phase coverage or short observational baseline affect the accuracy of the eccentricity estimate insofar as they decrease the precision of the estimate.

2.2.3 *Local minima at high eccentricities*

As shown in Baluev (2015), the number of local χ^2 minima increases significantly in the high eccentricity region. These minima might lead a local minimization algorithm or a Monte Carlo Markov Chain (MCMC) to be stuck in the wrong region of the parameter space. We here aim at quantifying and understanding this feature. In this section, we provide a summary of our results, whose precise study is in Appendix B. These results are

(i) An incorrect estimation of the noise level can lead to spurious deep local minima at high eccentricities.

(ii) As the SNR degrades, the probability of missing the global minimum by a non-linear fit initialized on a circular orbit increases.

(iii) There is a geometrical interpretation of the numerous local minima at high eccentricity: the set of models with fixed eccentricity explore more and more dimensions of the N-dimensional sample space as eccentricity grows.

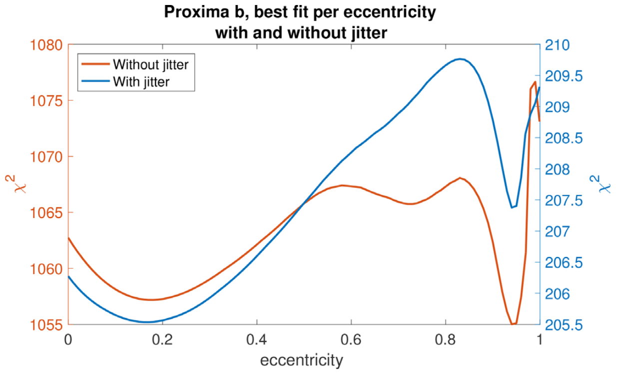

Figure 3. χ^2 of the residuals of a Keplerian fit as a function of the eccentricity on Proxima b (Anglada-Escudé et al. 2016).

Let us illustrate the first point on Proxima b. Brown (2017) re-analyses the data of Proxima b, a ≈1.27 $M_\oplus$ planet orbiting the M star Proxima Centauri with a period of 11.186 d (Anglada-Escudé et al. 2016), and finds that there are local minima at eccentricity 0.75 and 0.95, the 0.95 eccentricity being the global least-squares fit.

We compute a Keplerian periodogram (O'Toole et al. 2009b; Zechmeister & Kürster 2009; Baluev 2015) in the vicinity of the 11.186 period. That is Keplerian models (equations 4–7) are fitted for a grid of periods, argument of periastron, and eccentricity. We then represent per eccentricity the χ^2 minimized over all other parameters. Using the nominal uncertainties, a single Keplerian model plus one offset per instrument, a linear and quadratic trend, we obtain the red curve in Fig. 3.

The curve displays three local minima, the deepest being at eccentricity 0.92. However, let us note that the best fit gives χ^2 of 1057, while there are only 214 measurements. This gives a reduced χ^2 of 5.16, which is unrealistic. We here simply add a constant jitter term in quadrature with the nominal error bars to obtain a reduced χ^2 of 1 at the best fit. The minimum χ^2 as a function of eccentricity so obtained is represented in blue in Fig. 3. The global minimum now occurs at $e = 0.17$. We interpret the global minimum at eccentricity 0.92 as an artefact of an incorrect estimation of the error bars.

The same Keplerian periodogram calculations can be done on simulated data sets with different noise levels and different numbers of measurements. We simulate such systems with eccentricity sampled from a uniform distribution on [0, 0.999], and count how many of them that have an SNR between 0 and 5, 5 and 10, etc. have $k = 1, 2, 3$, etc. local minima. The histogram of Fig. 4 is obtained. Also, for each bin of SNR, we compute the proportion of systems where the global minimum is not the closest to 0 (as in the case in Fig. 3, red curve) and therefore a local minimization should miss the global minimum. It appears that as the SNR increases, the fraction of cases where the global minimum is missed is decreasing, though not reaching zero.

Let us now briefly comment on the geometrical interpretation of the higher number of local minima at high eccentricities. Finding the best-fitting model amounts to finding the closest model to the observation in a geometrical sense. We consider the figure drawn in $\mathbb{R}^N$ by all the models that have an eccentricity e and a period P, denoted by $\mathcal{M}_{e,P}$. This figure might explore more or less dimensions. For instance, if it is close to a plane, it is nearly confined to a two-dimensional space. Otherwise, exploring many dimensions traduces a 'rough' surface, which increases the chances of finding

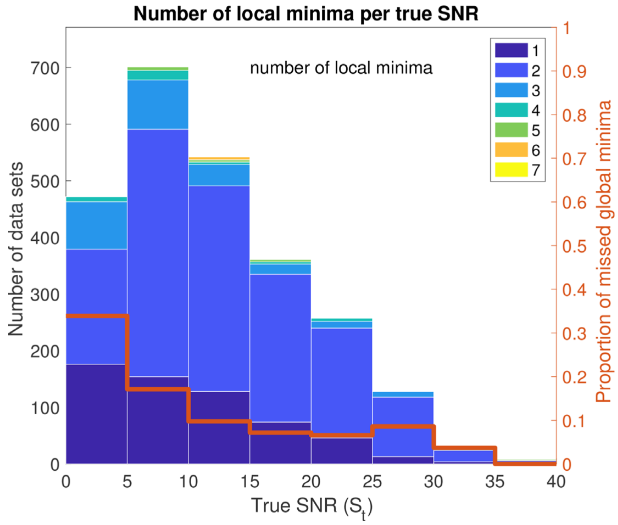

Figure 4. Blue bins: binned values of the number of systems that have a condition number lower than 10^7 with 1, 2, 3, 4, 5, or 6 local minima, with a bin size in fitted SNR S_{fit} of 5. Red curve: fraction of the binned systems where the global minimum is not the one obtained with a linear fit.

a local minimum of distance to the data. By a procedure based on singular value decomposition, detailed in Appendix B and in Hara 2017 (PhD thesis), it is possible to obtain an approximate number of dimensions explored by $\mathcal{M}_{e,P}$ as a function of e. Table 2 shows such calculations for the 214 measurement times of GJ 876 (Correia et al. 2010). As eccentricity increases, $\mathcal{M}_{e,P}$ explores more and more dimensions.

2.2.4 *Interval estimates*

As said in Section 1, one does not only want to obtain a value of the eccentricity with error bars, but also to test whether a certain value of the eccentricity is compatible with the data. This can be done in the frequentist setting with interval estimates. Since the focus will be put on Bayesian estimates in Section 3, we simply here give their definition and refer the reader to Appendix C for their derivation and detailed study.

The hypothesis that the eccentricity has a certain value e is rejected with a confidence level α if the likelihood ratio LR_e satisfies

$$\mathrm{LR}_e := \frac{\max\limits_{\boldsymbol{\theta}\in\Theta_e} p(\boldsymbol{y}|\boldsymbol{\theta})}{\max\limits_{\boldsymbol{\theta}\in\Theta} p(\boldsymbol{y}|\boldsymbol{\theta})} \leqslant \mathrm{e}^{-\frac{1}{2}\beta}, \qquad \text{where} \tag{14}$$

$$\beta := F^{-1}_{\chi^2_\rho}(1-\alpha) \tag{15}$$

$$\rho := 2 + 2S'^2\frac{e^2}{1+e^2} - \frac{\pi e}{1+e^2} L_{\frac{1}{2}}\left(-\frac{S'^2}{2}\right) L_{\frac{1}{2}}\left(-\frac{e^2 S'^2}{2}\right). \tag{16}$$

where Θ_e is the set of parameters that have all eccentricity e, $p(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood, $F^{-1}_{\chi^2_\rho}$ is the inverse cumulative distribution function of a χ^2 law with ρ degrees of freedom, $S' = (\sigma/K_{\mathrm{t}})\sqrt{2/N}$, and $L_{\frac{1}{2}}$ is the Laguerre polynomial of order 1/2. We also fit a free jitter term so that the reduced χ^2 equals one. Conversely, for a certain measured LR_e the FAP(e) of the hypothesis $e_{\mathrm{t}} = e$ is defined as

$$\mathrm{FAP}(e) = 1 - F^2_{\chi_\rho}(-2\ln \mathrm{LR}_e). \tag{17}$$

The confidence interval of eccentricity is the set of e with FAP(e) greater than a certain threshold (for instance 0.05).

2.3 Posterior distributions

2.3.1 *Point estimators*

The previous sections are devoted to the study the least square eccentricity estimate. However, the standard practice in the exoplanet community is rather to compute the posterior probability $p(\boldsymbol{\theta}|\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})/p(\boldsymbol{y})$ of the orbital elements $\boldsymbol{\theta}$ using MCMCs algorithms (e.g. Ford 2005, 2006).

From such posterior distributions, one can compute the orbital elements corresponding to the maximum a posteriori (MAP)

$$\widehat{\theta}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{\theta}\in\Theta} p(\boldsymbol{\theta}|\boldsymbol{y}). \tag{18}$$

For a simple Keplerian fit $\widehat{\theta}_{\mathrm{MAP}} = (K, e, P, \omega, M_0)^{\mathrm{MAP}}$, an eccentricity estimate is then obtained by e^{MAP}. Alternately, one can compute the marginal distribution

$$p(e|\boldsymbol{y}) = \int_{\Theta_e} p(e, \tilde{\boldsymbol{\theta}}|\boldsymbol{y})\,\mathrm{d}\tilde{\boldsymbol{\theta}} = \frac{1}{p(\boldsymbol{y})}\int_{\Theta_e} p(\boldsymbol{y}|e, \tilde{\boldsymbol{\theta}})p(\tilde{\boldsymbol{\theta}})\,\mathrm{d}\tilde{\boldsymbol{\theta}} \tag{19}$$

and its mode, mean, and median

$$e_{\max} = \arg\max_{e\in[0,1]} p(e|\boldsymbol{y}) \;\; ; \;\; e_{\mathrm{mean}} = \mathrm{mean}(p(e|\boldsymbol{y}))$$

$$e_{\mathrm{med}} = \mathrm{median}(p(e|\boldsymbol{y})), \tag{20}$$

where $\tilde{\boldsymbol{\theta}}$ is the vector of parameters that are not eccentricity, and Θ_e is the space of parameters where the eccentricity is equal to e. A standard result is that e_{mean} is the estimator that minimizes the MSE (see equation 3). Also, e_{med} minimizes the mean absolute error (MAE), defined as

$$\mathrm{MAE} := \mathbb{E}\{|\widehat{e} - e_{\mathrm{t}}|\}. \tag{21}$$

The estimators (20) are in general less than the maximum likelihood or the maximum a posteriori. This is shown for e_{mean} with a numerical experiment. We generate a circular planet of semi-amplitude 3.5 $\mathrm{m\,s^{-1}}$ and 100 realizations of Gaussian white noise at 2 $\mathrm{m\,s^{-1}}$. The estimates e_{ML} and e_{mean} (equation 20) are computed with a Monte-Carlo Markov Chain (MCMC) algorithm. The sampler is the same as in Delisle et al. (2018), based on the adaptive Metropolis sampler of Haario, Saksman & Tamminen (2001). The model consists of a Keplerian, an offset, and a free jitter term, with uniform priors on all parameters. Fig. 5 shows experimental distribution function of $e_{\mathrm{ML}} - e_{\mathrm{mean}}$. The condition $e_{\mathrm{ML}} > e_{\mathrm{mean}}$ is verified in 79 cases out of 100, with a mean value of $e_{\mathrm{ML}} - e_{\mathrm{mean}}$ equal to 0.0944.

The efficiency of the estimates (20) is understandable in terms of trade-off between model simplicity and agreement with the data. In Sections 2.2.2 and 2.2.3, it appeared that for eccentricities $e_{\mathrm{t}} < 0.2$, in the vicinity of the true model, there is a larger volume of models with eccentricities $e > e_{\mathrm{t}}$ than $e < e_{\mathrm{t}}$. The integration over the domain Θ_e in equation (19) penalizes the models with high complexity, which here are the high-eccentricity models. This is comparable to the penalization of models with too many planets by the marginal likelihood in the context of exoplanets detection (e.g. Nelson et al. 2018). In Section 2.3.2, we argue in favour of reporting e_{mean} and/or e_{med}, as $e_{\max}$ is too biased towards low eccentricities.

Table 2. Dimension of the models with fixed eccentricity as a function of the eccentricity, GJ876 measurement times.

Eccentricity	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.999
Dimension of $\mathcal{M}_{e,P}$	3	4	4	6	8	10	14	24	46	91

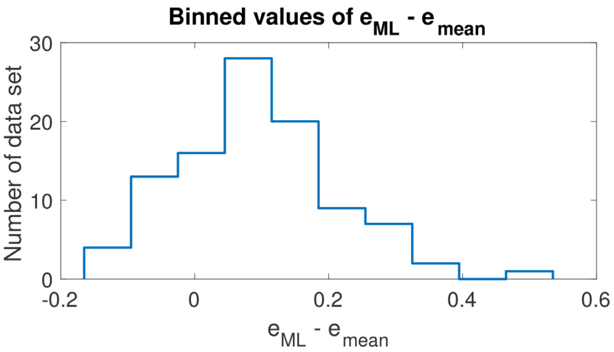

Figure 5. Binned values of the difference of maximum likelihood and posterior mean estimates, $e_{\mathrm{ML}} - e_{\mathrm{mean}}$ for 100 realization of white noise and $K/\sigma = 3.5$.

2.3.2 *Hypothesis testing*

Several of the works cited in Section 1 address the question of whether an eccentricity should be set to zero or not, which is a model selection problem. It can be addressed by computing the ratio of posterior likelihood, or odds ratio, of the models (e.g. Kass & Raftery 1995),

$$R = \frac{\Pr\{e \neq 0|\boldsymbol{y}\}}{\Pr\{e = 0|\boldsymbol{y}\}} = \frac{\Pr\{\boldsymbol{y}|e \neq 0\}\ \Pr\{e \neq 0\}}{\Pr\{\boldsymbol{y}|e = 0\}\ \Pr\{e = 0\}}, \tag{22}$$

where $\Pr\{\boldsymbol{y}|e \neq 0\} = \int_{\boldsymbol{\theta}\in\Theta} p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}$. If this ratio is greater than a certain value, then one favours $e \neq 0$ over $e = 0$. This methodology has been used in Bonomo et al. (2017a,b). As the number of samples N tends to infinity, assuming $\Pr\{e \neq 0\} = \Pr\{e = 0\}$, the odds ratio is equivalent to the BIC (Schwarz 1978), as used by Pont et al. (2011) and Husnoo et al. (2011, 2012).

More generally, one can compute a credible set, that is a set of e, denoted by $C \subset [0, 1]$ such that

$$\Pr\{e \in C|\boldsymbol{y}\} = \int_C p(e|\boldsymbol{y})\,\mathrm{d}e = \alpha, \tag{23}$$

where $\alpha \in [0, 1]$ is a probability. The set C is in general taken as an interval but this need not be the case. Let us note that this approach is also a ratio of posterior likelihood as in equation (22). If $e = 0$ is given a non-null probability, then the prior probability takes the form $p(e) = p(0)\delta(0) + p(e)$ where δ is the Dirac function. We now focus on credible intervals, (equation 23) since these are more widely used and give finer information.

3 ROBUSTNESS OF ECCENTRICITY ESTIMATES

3.1 Problem statement

In Section 2, several tools to make inferences on eccentricity were presented. We now study whether these are reliable even if the adopted model is incorrect. What we call a model is a couple of prior and likelihood functions. We assume that the orbital elements are distributed according to a true prior $p_{\mathrm{t}}(\boldsymbol{\theta})$ and the observations have a true likelihood $p_{\mathrm{t}}(\boldsymbol{y}|\boldsymbol{\theta})$ ($\mathcal{M}_{\mathrm{t}}$) while the analysis is made with the model $\mathcal{M}$: $p(\boldsymbol{\theta})$, $p(\boldsymbol{y}|\boldsymbol{\theta})$. The model $\mathcal{M}$ can be too simple: missed planetary signal, non-modelled correlated noise or too complicated: for instance Gaussian processes are known to be very flexible, possibly too much.

The case where the model is too complicated will not be treated in detail. We simply point out that from Section 2.2, we expect that the bias is higher than for a correct model. Indeed, as the model grows in complexity the correlation between parameters increases, therefore the error on eccentricity σ_e increases. This might be problematic as the bias is proportional to σ_e (see equation 13). However, the error bars broaden, so that having too complex a model is unlikely to produce spurious conclusions. On the contrary, as we shall see, simplistic models can be problematic.

In the following sections, we study the effect of the noise level estimate (Section 3.2), numerical effects (Section 3.3), incorrect noise models (Section 3.4), priors (Section 3.5), and the comparison of two models: one eccentric planet or two planets in 2:1 mean motion resonance (Section 3.6).

3.2 Noise level

In Sections 2.2.2 and 2.2.3, it appeared that an incorrect estimate of the noise norm leads to an underestimated bias and to spurious local minima at high eccentricity. As a consequence, it is key, as is standard practice, to adjust at least an extra jitter term σ_J in the likelihood,

$$p(\boldsymbol{y}|\boldsymbol{\theta}, \sigma_J) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{V_0} + \sigma_J^2\mathbf{I}|}}\,\mathrm{e}^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{\theta}))^T(\mathbf{V_0}+\sigma_J^2\mathbf{I})^{-1}(\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{\theta}))}, \tag{24}$$

where $\mathbf{V_0}$ is the nominal covariance, $\mathbf{I}$ the identity matrix, and $\boldsymbol{f}(\boldsymbol{\theta})$ is the signal model, containing Keplerians and possibly other features. All the following analyses are made with the model (24).

3.3 Numerical effects

It appeared in Section 2.2.2 that uncertainties on the eccentricity estimates increase the biases. This is also valid for the uncertainties stemming from the numerical methods used to compute the orbital elements.

As noted by Eastman, Gaudi & Agol (2013), there is a specific error in the implementation of the Metropolis–Hastings algorithm that worsens the bias, when the true eccentricity is close to zero and when the parameter space is parametrized by (e, ω) instead of (k, h) or $(\tilde{k}, \tilde{h})$. This error consists in not recording the value of a proposed parameter in the chain if it is rejected.

More generally, credible intervals or Bayes factors can be unreliable if the numerical schemes have not converged. Hogg & Foreman-Mackey (2018) gives several ways to check for convergence of MCMC algorithms. In the following analyses, the convergence diagnostic is the number of effective samples, as computed in appendix A of Delisle et al. (2018) (see also Sokal 1997). This number, N_{eff} is interpretable as an equivalent number of independent samples from the posterior distribution. Then, for instance, the numerical uncertainty on the mean of the marginal posterior $p(\theta_0|\boldsymbol{y})$ of a parameter θ_0 scales as $\sigma_p/\sqrt{N_{\mathrm{eff}}}$, where σ_p is the standard deviation of $p(\theta_0|\boldsymbol{y})$.

The first result we show is that at low SNR, the convergence is slower. This is likely due to the existence of local minima at high eccentricity (see Fig. 4, Section 2.2.3). We simulate a one planet system on CoRoT-9 28 measurements (Bonomo et al. 2017b), the eccentricity is generated with a Beta distribution ($a = 0.867$, $b = 3.03$), angles are uniformly distributed and the period is fixed to 95 d. The nominal measurement errors are normalized to obtain a mean variance of 1. A Gaussian white noise following the normalized nominal errors plus a 1 $\mathrm{m\,s^{-1}}$ white noise is added. A hundred data sets with orbital elements and noise sampled according to their distributions are created. This is done for $K = 2.5\ \mathrm{m\,s^{-1}}$ and $K = 5\ \mathrm{m\,s^{-1}}$, which corresponds to SNR 4.2 and 8.5 (as defined in equation 11), so to obtain $2 \times 100 = 200$ data sets. Finally, an MCMC is performed on each of them with model (24) with 1100 000 samples and a burn-in phase of a fourth of the total number of samples. The average number of independent samples is 7300 versus 12 000, and the chain does not reach a 1000 efficient samples in 28 and 5 cases, respectively. In conclusion, low SNR signals should be treated with particular attention since the number of independent samples of the MCMC is smaller, while chains are initialized with a high eccentricity due to the higher bias of least square estimates at low SNR (see equation 13).

As an example of the problem associated with not having enough samples, this experiment is repeated with an injected eccentricity of 0. We consider that the 'small eccentricity' hypothesis is rejected if $p(e \in [0, 0.05]) < 0.05$. Let us consider the samples obtained for the experiment described above with $K = 2.5\ \mathrm{m\,s^{-1}}$. Depending on whether we take the first fourth of the samples or all of them, the fraction of rejection of the small eccentricity hypothesis goes respectively from 12 to 4 per cent.

In the following sections, simulations are taken into consideration if they reach an effective number of independent samples greater than 1000, so that the mean of the eccentricity posterior (of variance σ_e) is known with at least a $\approx \sigma_e/\sqrt{N_{\mathrm{eff}}} \approx \sigma_e \times 3$ per cent accuracy. Since we compare different types of noise, we do not require an extremely good precision on the posterior distribution (also the following result stay approximately identical if we take runs with at least 2000, 5000 efficient samples). In practice, it is safer to ensure that 10 000 independent samples are reached to obtain a $\sigma_e/\sqrt{10\,000} = \sigma_e \times 1$ per cent accuracy on the posterior mean.

3.4 Incorrect noise model

3.4.1 *Non-Gaussian noise*

The first question we address is whether, when the noise is non-Gaussian, using the model (24) leads to spurious inferences on eccentricity. For the sake of brevity, we here only report the results of the analysis done in Appendix D. The non-Gaussianity of the noise can only lead to slightly underestimating or overestimating the error bars, when the noise distribution is very heavy-tailed. We found that the estimates of the eccentricity and the error bars are mostly sensitive to the covariance of the noise and the following sections are focused on this aspect.

3.4.2 *Incorrect covariance: simulations*

To study the effect of the noise covariance on eccentricity estimates, we proceed as follows. We consider the 28 measurement times of CoRoT-9, spanning on 1700 d, and generate a circular planet at 95 d, denoted by $y_{\mathrm{planet}}(t)$. We then generate 100 realizations of white noise, $y_{\mathrm{noise}}(t)$, and for each of them, a non-modelled noise is added. Six different types of such noises, described below, are considered. In total, $6 \times 100 = 600$ data sets are obtained. The signal generated is then of the form $y(t) = y_{\mathrm{planet}}(t) + y_{\mathrm{noise}}(t) + y_k(t)$ with $k = 0..5$. $y_{\mathrm{noise}}(t)$ is generated according to the nominal uncertainties, which are CoRoT-9 uncertainties scaled so that their root mean square is equal to 1 $\mathrm{m\,s^{-1}}$. Such signals are generated in three contexts: with a semi-amplitude of the planet at 95 d of 2.5, 3.5, or 5 $\mathrm{m\,s^{-1}}$. The methodology described aims at evaluating if, when the noise model is incorrect, circular planets tend to appear as eccentric. This experiment is also done with the eccentricity of the planet drawn from the same Beta distribution as the prior, to evaluate the impact of the noise on the estimates. The simulation where the eccentricity is fixed to zero and the one where it follows a beta distribution are respectively called S_0 and S_e.

In each simulation, on each of the 3×600 data sets, the posterior distribution of the orbital elements is computed using the model (24) and priors given in Table 4. 1100 000 samples are computed, the first fourth being the burn-in phase. The algorithm is an adaptive Metropolis algorithm as in Delisle et al. (2018), and the convergence is checked by calculating the effective number of independent samples (see Delisle et al. 2018, Appendix A).

Since our goal is to test the effect of the noise nature, and not its level, we impose that the norm of the non-modelled noise is such that $\|\boldsymbol{y}_0\| = \|\boldsymbol{y}_1\| = \dots\|\boldsymbol{y}_5\| = \gamma$. For each of the 100 realizations of white noises, we draw γ from a χ^2 law with $N = 28$ degrees of freedom. The y_k are defined as follows:

y_0: white, Gaussian noise identically distributed.

y_1: also white Gaussian noise but with different variances. The variances are drawn from a Laplace law so to obtain a wide range of value.

y_2: A circular planet that is too small to be fitted, the period of the added planet P is drawn from a log-normal distribution until a period is found such that P differs from 95 d of at least 20 per cent. This value is chosen to avoid the lowest probability region of period ratios of planet pairs found by Kepler (Steffen & Hwang 2015).

y_3: planet in resonance with the injected planet, in 1:2 or 3:2 resonance, inner or outer with probability 1/2.

y_4: a Gaussian correlated noise with covariance κ

$$\kappa_{i,j} = \alpha^2 \exp\left[-\frac{1}{2}\left\{\frac{\sin^2[\pi(t_i - t_j)/\tau]}{\lambda_{\mathrm{p}}^2} + \frac{(t_i - t_j)^2}{\lambda_{\mathrm{e}}^2}\right\}\right], \qquad (25)$$

as in Haywood et al. (2014). We use the values of the Evidence Challenge (Nelson et al. 2018), $\alpha = \sqrt{3}\ \mathrm{m\,s^{-1}}$, $\lambda_{\mathrm{e}} = 50.0$ d, $\lambda_{\mathrm{p}} = 0.5$ (unitless), and $\tau = 20.0$ d.

y_5: same as y_4 but with values $\alpha = \sqrt{3}\ \mathrm{m\,s^{-1}}$, $\lambda_{\mathrm{e}} = 50.0$ d, $\lambda_{\mathrm{p}} = 0.3$ (unitless), and $\tau = 30.0$ d.

The rationale behind taking y_1 as such is to emulate the effect of mild outliers, not obvious enough to be completely discarded. As the Laplace distribution has heavy tails, it generates values of the variances that are very different from each other. y_2 and y_3 are chosen as such because the strongest resonances found in Kepler data are the 3:2 and 1:2 (or close to) resonances. In first approximation, the real period ratio distribution is a combination of these two artificial distributions. Finally, y_4 and y_5 are two types of stellar noises, the second one having a slightly stronger periodic component in the covariance.

For the simulation where the eccentricity is fixed to zero, S_0, on each simulation we compute the probability $\Pr\{e \in [0, 0.05]\}$. We report the number of simulations where this quantity is below a threshold $\alpha = 0.05$ and that have at least a thousand effective

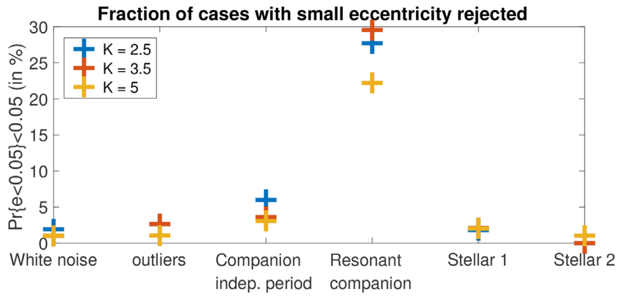

Figure 6. Fraction of cases where the posterior probability of $e \in [0, 0.05]$ is below 0.05 for an injected circular signal and different noises described in Section 3.4.2. The blue, red, and yellow points correspond to the experiment for $K = 2.5$, 3.5, and 5 m s^{-1}, respectively. The posterior distribution is computed with the model (24).

Table 3. Root mean squared error ($\sqrt{\mathrm{MSE}}$) and MAE averaged over all types of noises of several estimators: maximum likelihood (e_{ML}, max. a posteriori e_{MAP}, mode, mean and median of the eccentricity posterior $e_{\max}$, e_{mean}, and e_{med} for the $K = 3.5$ m s^{-1} simulation).

Estimator	e_{ML}	e_{MAP}	$e_{\max}$	e_{mean}	e_{med}
$\sqrt{\mathrm{MSE}}$	0.1384	0.1362	0.1613	0.1088	0.1117
MAE	0.1071	0.1062	0.1239	0.0842	0.0849

samples. The results are shown in Fig. 6 for different values of the semi-amplitude of the input planet ($K = 2.5$, 3.5, or 5 m s^{-1}, respectively blue, red, and yellow markers). These values of K were chosen to be close to relatively low SNRs. The small eccentricity rejection rate is approximately constant for the different noises except for a resonant companion, where the non-zero eccentricity is, on average, rejected in 26 per cent of the cases, that is 13 times more often than for the other noise models. A non-resonant companion also might induce an increased rate of rejected zero eccentricity (4.2 per cent on average versus 1.3 per cent for white noise). The power of the stellar noise is stronger around 20 and 30 d, which is not on a harmonic of the period of the planet (95 d), and therefore they lead to an even lower rejection of the low eccentricity scenario (see Section 3.4.3).

In the simulation S_e, we compute the absolute value of the difference between the estimated eccentricity and the true one for the three estimators (20). In all cases, we find that the median and mean of the posterior distribution of eccentricity are very similar and largely outperform the other estimates in terms of MSE (equation 3) and MAE (equation 21). We have seen in Section 2.3 that for a correct model, the mean and median have respectively minimal MSE and MAE, this is indeed the case on the simulations with the white noise. The MSE and MAE over the six types of noise are reported in Table 3. The estimator e_{mean} and e_{med} are more accurate when the noise model is incorrect both for MSE and MAE, as a consequence we deem them as the best ones overall. An argument in favour of e_{mean} is that it has minimal MSE (equation 3). The MSE penalizes the square of the difference between the estimated and true value. Multiplying by two this difference multiplies by four the cost of the error. Therefore, estimators with small MSE are less likely to produce large errors.

This result seems in contradiction with Zakamska et al. (2011), which finds the mode of the posterior to be less biased. However, they consider cases where the eccentricity is small. We also find that for small eccentricities, $e_{\max}$ is less than e_{med} and e_{mean}. Due to the Beta prior in eccentricity, it happens that the posterior is bi-modal and $e_{\max} = 0$. When e follows the prior distribution, e_{med} and e_{mean} are more accurate, at least in terms of MSE and MAE.

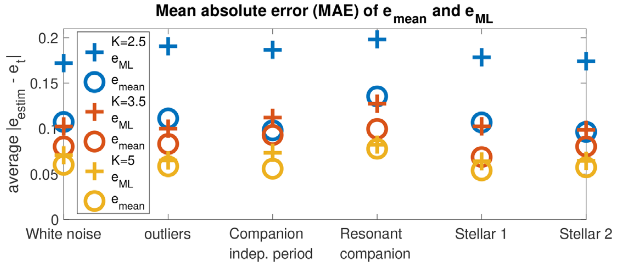

Figure 7. MAE on eccentricity of the posterior mean (e_{mean}, equation 20) and maximum likelihood (e_{ML}) eccentricity estimator. The values of K are given in m s^{-1}.

We now only present in Fig. 7 the performance in terms of MAE of the estimator e_{mean} (circles) and the maximum likelihood estimator e_{ML} (crosses), for comparison (MSE behaves similarly). In all cases e_{mean} shows better performance. The MAE does not heavily depend on the type of noise. However, we do observe a slight increase of the error for the unseen resonant companion. Let us also note that as the signal semi-amplitude increases, the difference of performance between the maximum likelihood and e_{mean} becomes less clear.

3.4.3 Interpretation: noise power at the planet semiperiod

In Section 3.4.1 we showed that the error on eccentricity is mainly determined by the true covariance of the noise. In Section 3.4.2 however, the simulated stellar noises did not yield particularly high errors on the eccentricity. We now show that the property of correlated noises most impacting eccentricity estimates is their power at the semiperiod of the planet of interest.

Let us consider a signal $\boldsymbol{y} = \boldsymbol{y}_0 + \boldsymbol{\epsilon}$, where $\boldsymbol{y}_0$ is a circular orbit of period P, and $\boldsymbol{\epsilon}$ is an unknown stochastic signal, that the data analyst supposes to have covariance $\mathbf{V}$. Denoting by ω_0: $=2\pi/P$, we define the $N \times 2$ matrix $\mathbf{M}_k = [\cos(k\omega_0\boldsymbol{t}), \sin(k\omega_0\boldsymbol{t})]$ and $\mathbf{P} := \mathbf{M}_k(\mathbf{M}_k^T\mathbf{V}^{-1}\mathbf{M}_k)^{-1}\mathbf{M}_k^T\mathbf{V}^{-1}$ the projection matrix on to the space spanned by the columns of $\mathbf{M}_k$. More generally, for a projection matrix $\mathbf{P}$ on to a vector space $\mathbf{M}$ we define

$$R_{\epsilon}(\mathbf{M}) = \frac{(\mathbf{P}\boldsymbol{\epsilon})^T\mathbf{V}^{-1}(\mathbf{P}\boldsymbol{\epsilon})}{\boldsymbol{\epsilon}^T\mathbf{V}^{-1}\boldsymbol{\epsilon}}\frac{N}{2} \tag{26}$$

$$Q_{\epsilon}(\mathbf{M}) = \mathbb{E}\left\{Q_{\epsilon}\right\}. \tag{27}$$

The rationale of defining these quantities is to identify if, assuming a covariance $\mathbf{V}$ a noise is more correlated to a certain space than to its orthogonal. If $\boldsymbol{\epsilon}$ is a white noise, then it is not particularly correlated to any particular space, so that $R = 1$. In the limit cases where $\boldsymbol{\epsilon}$ lies in, or is orthogonal to the space spanned by $\mathbf{M}_k$, then $R = N/2$ resp. 0.

To test the influence of $Q_{\epsilon}(\mathbf{M}_2)$ on the eccentricity error we proceed as follows. We consider an array of measurement times from a real system and generate a circular signal y_0 plus a noise with a certain true covariance. In order to obtain noises with very different spectral contents, we proceed as follows. For a given frequency ω, we draw thirty realizations of $\boldsymbol{\epsilon} = \cos(\omega\boldsymbol{t} + \phi)$, ϕ following a uniform distribution on $[0, 2\pi]$. For each of them we compute R_{ϵ}, the least square estimate of the orbital elements $\widehat{\boldsymbol{\theta}}$, the estimate of

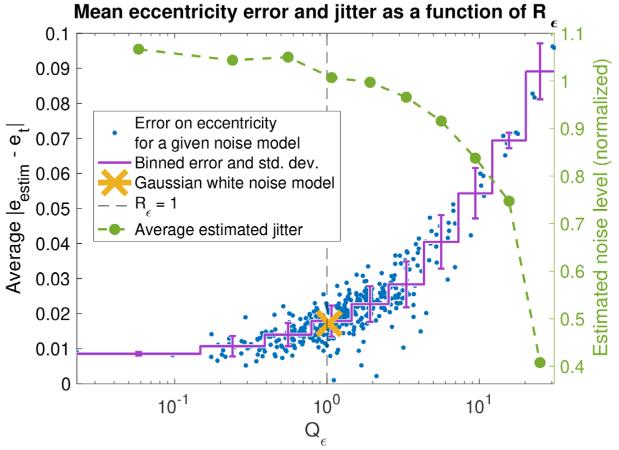

Figure 8. Error on eccentricity plotted against the estimated Q_ϵ (see equation 27) for different noise types (blue points) and Gaussian white noise (yellow cross). Same quantity averaged per interval of Q_ϵ with standard deviations (purple stair curve). The estimates of noise level, averaged per bin, are represented in green.

the noise level $\|\boldsymbol{y} - \boldsymbol{f}(\widehat{\boldsymbol{\theta}})\|/\sqrt{N}$. The R_ϵ are averaged to have an estimate $\widehat{Q_\epsilon}$ of Q_ϵ and the average error on eccentricity $\langle|\widehat{e} - e_t|\rangle$. For each type of noise $\boldsymbol{\epsilon}$, we plot $(\widehat{Q_\epsilon}, \langle|\widehat{e} - e_t|\rangle)$, which corresponds to a blue point in Fig. 8. The point obtained with $\boldsymbol{\epsilon}$ being a white Gaussian noise model is represented with a yellow cross. We then bin the values with a constant step in $\log Q_\epsilon$ and compute the average error, as well as its standard deviation (purple stair curve). We also compute the mean value of the estimated jitter, and divide it by the value of the jitter estimated for $Q_\epsilon = 1$. The normalized jitter so obtained is represented in green, with its scale on the right y-axis.

Fig. 8 is obtained with the time array of Gl 96 SOPHIE measurements (67 measurements), $e_t = 0$, an assumed covariance matrix $\mathbf{V}$ equal to identity, a period of 40 d and fixed ratio of the norm of the Keplerian signal and the input noise of 10 (which corresponds to a very high SNR = 78), to have as little influence as possible of the noise level. As the noise becomes more correlated with the $\mathbf{M}_2$ space (Q_ϵ increases) it is absorbed in the fit and the RMS of the residual decreases. Indeed in Fig. 8 it is apparent that as Q_ϵ increases the error on eccentricity grows while the estimated level of the noise decreases. The error on e for $e_t = 0.9$ is consistently higher, which is likely due to local minima at high eccentricity (see Section 2.2.3).

One can then wonder how the SNR S affects the bias on eccentricity for correlated noises. Defining the noise 'quality factor' as $q := \sqrt{Q_\epsilon(\mathbf{M}_2)}$, in the linear approximation used for (11), the uncertainty on k and h becomes $q\sigma_k$. With notations of equation (13), the bias at $e_t = 0$ is therefore approximately

$$b(0, S) \approx \sqrt{\frac{\pi}{2}\frac{q}{S}}. \tag{28}$$

For a given q, as the SNR increases, the bias decreases.

To check if the power at semiperiod is also correlated with the error on eccentricity if the true eccentricity is high, we perform the same experiment with a value of the eccentricity equal to 0, 0.3, 0.5, 0.7, 0.9. Again, the ratio of the norm of the Keplerian signal and the input noise is fixed to 10. For each eccentricity the experiment is done 20 times with periods randomly drawn with a lognormal law. The results are shown in Fig. 9. We first remark that as e increases, the bias, and therefore the error on eccentricity, decreases. Secondly, from $e = 0.5$, Q_ϵ is less relevant to predict the effect of the noise on eccentricity estimates. This result is in accordance with Wittenmyer et al. (2019), who found that planets with e>0.5 are very unlikely to be mistaken for a two circular planet model. As eccentricity increases, the harmonics of order greater than two increase, so that a noise localized at the semiperiod cannot mimic a higher eccentricity.

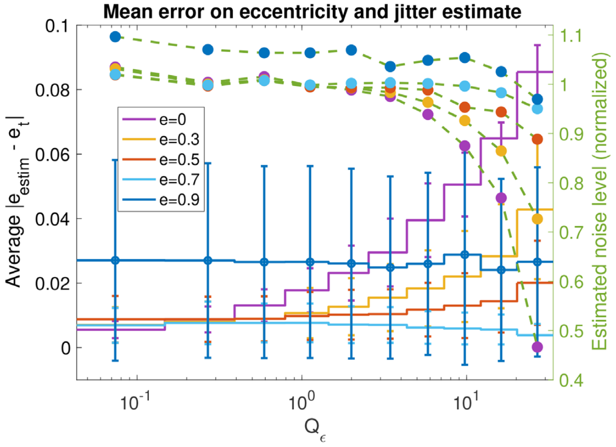

Figure 9. Error on eccentricity plotted against the estimated Q_ϵ (see equation 27) averaged for 10 different periods. The average error per Q_ϵ and the estimated jitter for the different values of the true eccentricity are represented: $e = 0$, 0.3, 0.5, 0.7, 0.9 (resp. purple, yellow, red, light blue, dark blue).

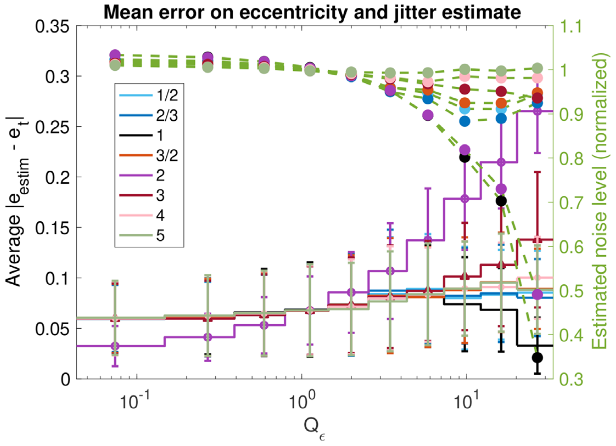

Figure 10. Erroron eccentricity (solid stairs) and estimated error (round markers) plotted for a true circular orbit against the estimated $Q_\epsilon(\mathbf{P}_k)$ (see equation 27) where $\mathbf{M}_k$ are the vector spaces = $(\cos k\omega_0\boldsymbol{t}, \sin k\omega_0\boldsymbol{t})$ for $k = 0.5\ldots5$ (see legend for colour code) and ω_0 is the frequency of the input planet.

The same experiment is performed with a ratio of norm of the signal and the noise of three and an injected circular orbit. We compute $Q_\epsilon(\mathbf{M}_k)$ with $k = 1/2, 2/3, 1, 3/2, 2, 3, 4, 5$. The results are plotted in Fig. 10, with a colour code for each k. It appears that a strong component of the noise on the harmonic 2, 3, 4, or 3/2, and 5 leads to an increased error on eccentricity (in decreasing order of effect). The noise level is notably underestimated in the $k = 2$ case. On the contrary, a strong component of the noise on $k = 0$ leads

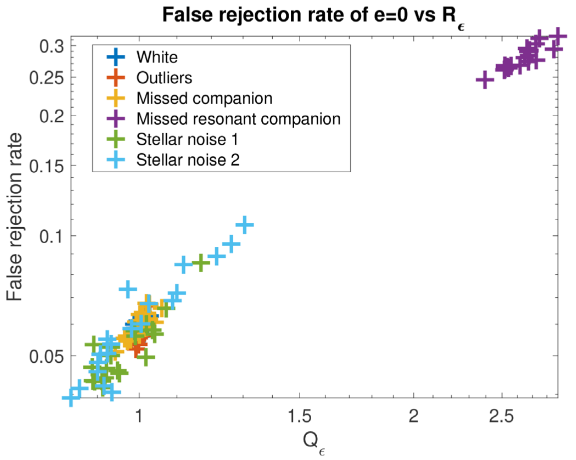

Figure 11. False rejection rate estimated with formula (15) as a function of Q_ϵ as defined in equation (27) for the six noises generated (y_0, ...y_5) and period of the true planet equal to 5..5..100 d.

to a smaller error, which is easy to understand. Indeed, since the eccentricity of the injected signal is zero, the noise reinforces the signal.

The results of Section 3.4.2 are interpretable with the analysis above. The metric Q_ϵ defined in equation (27) is computed [here $Q_\epsilon(\mathbf{M}_2)$] for the noises of Section 3.4.2, that are noises generated with nominal uncertainties plus y_0, ...y_5. For K = 2.5, 3.5, and 5 m s^{-1}, we generate 10 000 realizations of these noises, and inject a circular planet at period P with a random phase. We do this simulation for P = 5–100 d per step of 5 d. in each of the 6 × 10 000 × 20 simulations, we compute the FAP associated with $e = 0$ with formula (17), and adopt as a convention that $e = 0$ is rejected if the p-value is below 0.05. We also compute $R_\epsilon(\mathbf{M}_2)$ as defined in equation (26). For the six noises and 20 periods, we average the values of the R_ϵ to obtain Q_ϵ (27). Fig. 11 shows the proportion of false eccentricity rejected as a function of Q_ϵ for K = 5 m s^{-1}. It clearly appears that there is a strong correlation with the power of the noise at the semiperiod.

Obviously, one can test on a given system if a specific type of noise has a particular impact on a planet with injected parameters, the goal of this section was to identify some generic properties of the noise that lead to spurious inferences.

3.5 Robustness to prior changes

3.5.1 Simulations

In the previous sections we have studied the impact of having a wrong likelihood function. We now turn to the sensitivity of the estimate on the prior probability, here with a numerical experiment. With the formalism of Section 3.1, the data are generated with prior and likelihood $p_\mathrm{t}(\boldsymbol{\theta})$, $p_\mathrm{t}(\mathbf{y}|\boldsymbol{\theta})$, and the analysis is done with $p(\boldsymbol{\theta})$, $p_\mathrm{t}(\mathbf{y}|\boldsymbol{\theta})$. The likelihood is correct, but the prior is not, which corresponds to having an incorrect idea of the population distribution. Note that the prior probability of all parameters has an effect on the eccentricity estimate, but we focus on the prior probability chosen for the eccentricity.

Table 4. Priors used for the numerical experiments.

Parameter	Prior
K	Uniform on [0, 10 000]
P	$1/P$ uniform on [0, 20]
e	Beta ($a = 0.867$, $b = 3.03$) as in Kipping (2014)
ω	Uniform on [0, 2π]
M_0	Uniform on [0, 2π]

Table 5. Mean absolute and mean square error (MAE and MSE) of the estimate when the eccentricities are generated with a Beta distributions, and the analysis is done with the same Beta distribution as prior (correct model) or done with a uniform prior (incorrect model).

Estimator	Prior	$K = 2$ m s^{-1}	$K = 3.5$ m s^{-1}	$K = 5$ m s^{-1}
$\sqrt{\mathrm{MSE}}$	Correct	0.1325	0.0994	0.0791
$\sqrt{\mathrm{MSE}}$	Incorrect	0.1530	0.1118	0.0899
MAE	Correct	0.1073	0.0803	0.0601
MAE	Incorrect	0.1188	0.0851	0.0713

Two distributions are considered. We generate eccentricities according to the distribution Beta ($a = 0.867$, $b = 3.03$) and compute posterior probabilities with a uniform prior. Otherwise the priors are taken as in Table 4 and the data are generated as in Section 3.4. The results are shown in Table 5 as a function of the input amplitude. The errors, measured by MSE and MAE, are systematically worse when using the incorrect prior, however with an extra error not exceeding 15 per cent. It seems like an error of ≈0.01 on e could be counted as uncertainty on the prior distribution.

The uncertainty on the prior seems not to be a major concern, at least for the estimation of eccentricity. However, in some cases, one might want to recompute the credible interval with another prior, which is the object of the next section.

3.5.2 Recomputing the posterior without new sampling

The most straightforward way to explore the dependence of the posterior on the prior is to recompute it with another prior distribution. However, this might be lengthy to do systematically on several systems. We here propose an alternative which consists in multiplying the prior by a constant on a subset of its domain of definition, and to scale it elsewhere. In so doing, the output of the posterior sampler can be used straightforwardly without doing any sampling. In the following we illustrate the process with the prior on eccentricity.

We consider a measurable subset D of [0, 1] and its complement $\bar{D}$ in [0, 1]. For instance an interval $D = [0, e_0]$ for some arbitrary $e_0 \in (0, 1]$. Let us denote by $\tilde{\boldsymbol{\theta}}$ the model parameters other than eccentricity, so $\boldsymbol{\theta} = (e, \tilde{\boldsymbol{\theta}})$. We define a new prior p' such that for $e \in D$, $p'(e, \tilde{\boldsymbol{\theta}}) = xp(e, \tilde{\boldsymbol{\theta}})$. To ensure that $\int_0^1 \int p'(e, \tilde{\boldsymbol{\theta}})\,\mathrm{d}\tilde{\boldsymbol{\theta}}\,\mathrm{d}e = 1$, we take

$$a = \int_D \int p(e, \tilde{\boldsymbol{\theta}})\,\mathrm{d}\tilde{\boldsymbol{\theta}}\,\mathrm{d}e. \tag{29}$$

$$x \in [0, (1-a)/a] \tag{30}$$

$$p'(e, \tilde{\boldsymbol{\theta}}) = (1-ax)/(1-a)p(e, \tilde{\boldsymbol{\theta}}) \quad \text{for} \quad e \in \bar{D} \tag{31}$$

We now want to compute the probability that $e \in C$ for a prior distribution $p'(e, \tilde{\boldsymbol{\theta}})$. This one is given by replacing p by p' in equation (23). With the notations

$$z := (1-ax)/(1-a) \tag{32}$$

$$I_E := \int_{e\in E}\int_{\tilde{\boldsymbol{\theta}}} \frac{p(\mathbf{y}|e,\tilde{\boldsymbol{\theta}})p(e,\tilde{\boldsymbol{\theta}})}{p(y)}\,\mathrm{d}\tilde{\boldsymbol{\theta}}\,\mathrm{d}e \tag{33}$$

for $E\subset[0, 1]$, we compute

$$\begin{aligned}\Pr\{e\in C|\mathbf{y}, p'\} &= \int_C\int_{\tilde{\boldsymbol{\theta}}} p'(e|\mathbf{y})\,\mathrm{d}\tilde{\boldsymbol{\theta}}\,\mathrm{d}e\\ &= \frac{x\ I_{C\cap D} + z\ I_{C\cap\bar{D}}}{x\ I_D + z\ I_{\bar{D}}},\end{aligned} \tag{34}$$

where $p'(e|\mathbf{y})$ is the posterior distribution when the prior is $p^{'}(e)$ and $p(\mathbf{y}|e)$ is the likelihood marginalized on all parameters but eccentricity. When $x = 0$, all the prior probability goes to the complementary of D and $\Pr\{e \in D|\mathbf{y}\}$ goes to 0. If $x = (1 - a)/a$, e is certainly in D therefore $\Pr\{e \in D|\mathbf{y}\} = 1$.

The advantage of this calculation is that the integrals can be computed from the posterior samples. Denoting, $|E|$ the number of the MCMC samples that are such that $e \in E \subset [0, 1]$ and N_{tot} the total number of samples, an estimate of I_E is $\widehat{I_E} = |E|/N_{\mathrm{tot}}$, so that an estimate of (34) is

$$\widehat{\Pr}\{e\in C|\mathbf{y}, p'\} = \frac{x\ |C\cap D| + z\ |C\cap\bar{D}|}{x\ |D| + z|\bar{D}|}. \tag{35}$$

The reasoning can be extended straightforwardly to credible regions D and C in the parameter space, and to prior region subdivisions in $D_1 ... D_q$ with disjoint $(D_i)_{i=1..q}$ whose union is the whole parameter space.

Apart from the Markov chain samples, the only quantity needed to use (35) is a as given by (29). This expression might be difficult to compute in general, but in the case where $p(e,\tilde{\boldsymbol{\theta}}) = p(e)p(\tilde{\boldsymbol{\theta}})$, $a = \int_D p(e)\,\mathrm{d}e$, which is a one-dimensional integral. Analytic expressions might exist and a Riemann integration is always possible.

Since the integrals to be evaluated from posterior samples are random variables, it must be ensured that they have a controlled uncertainty. When breaking the posterior in many domains $D_1 ... D_q$, the procedure outlined may become unreliable if there are not enough independent samples in each D_k. One can easily compute the effective number of samples in each D_k, $N_{\mathrm{eff},k}$ by counting how many samples are in that region and dividing by the correlation time-scale. A number of effective samples greater than n gives an accuracy of $\approx 1/\sqrt{n} \times 100$ per cent on the probability $\Pr\{e \in D_k|\mathbf{y}\}$. Further investigation is left for future work.

3.6 Model comparison: one eccentric planet or 2:1 mean motion resonance

A system of two planets in 2:1 mean motion resonance can be mistaken for one eccentric planet, and vice versa. We here study the possibility to disentangle the two cases via Bayes factor as a function of the SNR. Two models are considered, $\mathcal{M}_e$ and $\mathcal{M}_{1:2}$, respectively an eccentric planet and two circular planets in mean motion resonance:

$$\mathcal{M}_e: \quad y(t) = K(\cos(\nu+\omega) + e\cos(\omega)) + g(\tilde{\boldsymbol{\theta}}) + \epsilon \tag{36}$$

$$\begin{aligned}\mathcal{M}_{1:2}: \quad y(t) = K_1\cos\left(\frac{2\pi}{P_1}t + \phi_{01}\right) + K_2\cos\left(\frac{2\pi}{P_2}t + \phi_{02}\right)\\ + g(\tilde{\boldsymbol{\theta}}) + \epsilon.\end{aligned} \tag{37}$$

Table 6. Priors used for the Bayes factor of the eccentric and 2:1 mean motion resonance models. The symbol T_{obs} denotes the observation time, $P_0 = 60$ d and $\alpha = 0.1$.

Parameter	Prior
σ_J^2	Uniform on [0, 100] m s^{-1}
offset	Uniform on $[-100, 100]$
K, K_1	Uniform in $\ln K$ on $[-1, 9]$
P, P_1	$1/P$ uniform on $[1/P_0 - 1/T_{\mathrm{obs}}, 1/P_0 + 1/T_{\mathrm{obs}}]$
e	Uniform on [0, 1]
ω	Uniform on $[0, 2\pi]$
M_0	Uniform on $[0, 2\pi]$
K_2	Uniform in $\ln K$ on $[-1, 9]$
P_2	$1/P$ uniform on $[(1-\alpha)\frac{2}{P_0}, (1+\alpha)\frac{2}{P_0}]$
ϕ_1	Uniform on $[0, 2\pi]$
ϕ_2	Uniform on $[0, 2\pi]$

Denoting by $g(\tilde{\boldsymbol{\theta}})$ a deterministic model encapsulating other planets, offsets, trends, etc. We also let vary a jitter term σ_J as in equation (8). The Bayes factor of the two models is defined as

$$B = \frac{p(\mathbf{y}|\mathcal{M}_e)}{p(\mathbf{y}|\mathcal{M}_{2:1})}, \tag{38}$$

where

$$p(\mathbf{y}|\mathcal{M}) = \int p(\mathbf{y}|\boldsymbol{\theta})p(\theta)\,\mathrm{d}\boldsymbol{\theta} \tag{39}$$

with $\boldsymbol{\theta} = (K, k, h, P, M_0, \tilde{\boldsymbol{\theta}}, \sigma_J)$ or $\boldsymbol{\theta} = (K_1, P_1, M_{01}, K_2, P_2, M_{02}, \tilde{\boldsymbol{\theta}}, \sigma_J)$ for $\mathcal{M} = \mathcal{M}_e$ and $\mathcal{M} = \mathcal{M}_{2:1}$, respectively. We expect the two models to be distinguishable if the amplitude of the second harmonic of the signal can be resolved (Anglada-Escudé et al. 2010, see equation 5).

In order to determine at which SNR the Bayes factor allows to disentangle resonant planets and eccentric ones, we perform a numerical experiment. We select a semi-amplitude K and a period P, then generate a Keplerian signal with $e = 0.25$, random M_0 and ω. K is chosen on a grid (2, 5, 8, 11, 14 m s^{-1}), to see how the ability to disentangle scenarios evolves with the true SNR.

On the other hand, we generate a two planet system. The outer planet has period P with random phase and semi-amplitude $K_1 = K$. The inner planet is circular with $K_2 = K_1/4$, so that $K_2/K_1 = e$ of the single planet. The phase of the second planet is chosen uniformly. The simulation is performed in two different settings. In the first one, the period ratio is fixed to $P_2 = P_1/2$. In the second, $1/P_2$ is chosen uniformly between $[(1 - \alpha)2/P_1, (1 + \alpha)2/P_1]$ with $\alpha = 0.1$. The rationale behind the choice of P_2 is that the period ratios of Kepler planets are located within a neighbourhood of 2:1 (Steffen & Hwang 2015).

The priors chosen to compute the Bayes factor are summarized in Table 6. The Bayes factor is computed with the nested sampling algorithm PolyChord (Handley, Hobson & Lasenby 2015a,b). The performance of the algorithm was checked on the data sets of the Evidence Challenge (Nelson et al. 2018). For each data set, the algorithm is ran at least five times. The $\ln\mathcal{Z}$ estimate is taken as the median of the different runs and the error bars are given by the variance of the empirical median as provided by Kenney & Keeping (1962). The error on the log Bayes factor $\ln\mathrm{BF} = \ln\mathcal{Z}_2 - \ln\mathcal{Z}_1$ is taken as $(\sigma^2_{\mathcal{Z}_1} + \sigma^2_{\mathcal{Z}_2})^{1/2}$.

Figs 12 and 13 show the results of the simulation, where the measurement time arrays are those of CoRoT-9 (Bonomo et al. 2017b) and Gl 96 (Hobson et al. 2018). The ln of the Bayes

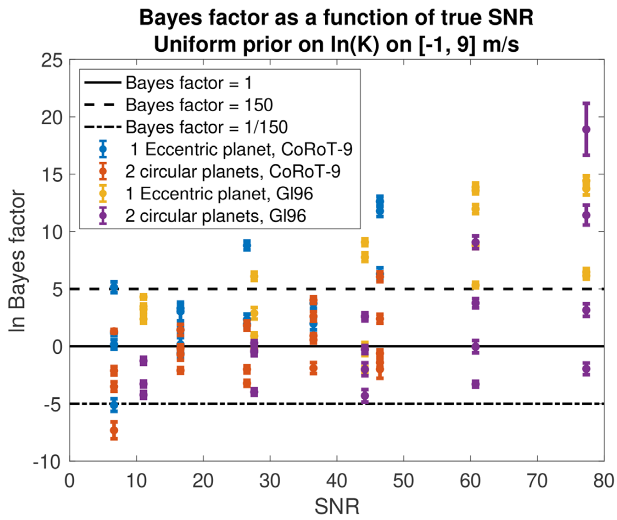

Figure 12. Difference of the log evidence of the correct and incorrect models. In the two planet case, the inner planet is generated with period exactly half of the outer planet. The colour code corresponds to the description of the true data set (one eccentric or two circular planets) and the array of measurement times used (CoRoT-9 or Gl96).

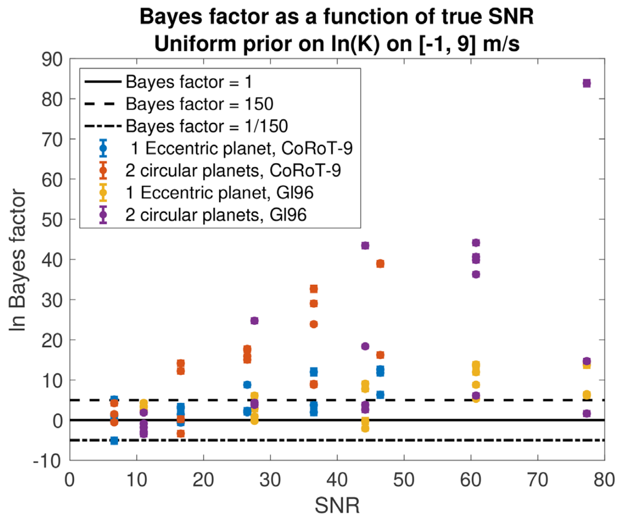

Figure 13. Difference of the log evidence of the correct and incorrect models, when the inner planet is generated with a frequency uniformly drawn on the same interval as the prior (Table 6).

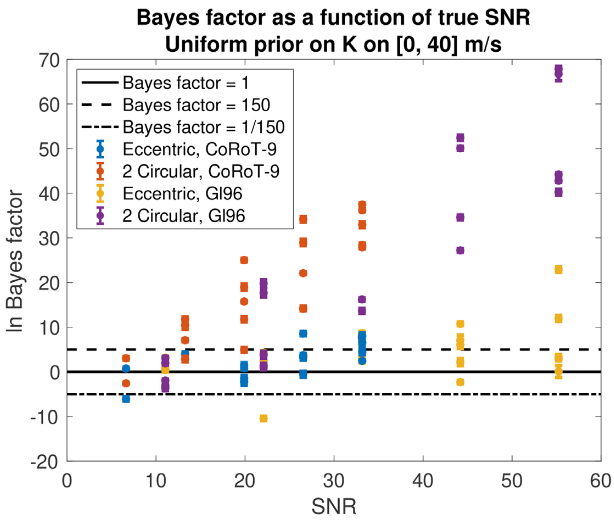

Figure 14. Difference of the log evidence of the correct and incorrect models, when the inner planet is generated with a frequency uniformly drawn on the same interval as the prior (Table 6) except for the semi-amplitude, where the prior is uniform on [0, 40] m s^{-1}.

factor is represented, such that the correct model is always at the numerator. The black dashed lines indicate a Bayes factor equal to 150 and 1/150. Bayes factors above 150 and below 1/150 correspond respectively to very strong evidence in favour or against the correct model.

Figs 12 and 13 show that distinguishing the resonant and eccentric models is possible, especially if the period of the inner planet varies, which is the case in Kepler data. None the less, at low SNR, there are cases of decisive Bayes factor against the correct model. In the low SNR regime, the evidence is dominated by the prior, and the parameter space is hard to explore. Both effects can account for these spurious results.

We point out that the values reported in Figs 12 and 13 are very sensitive to the prior on K. The narrower it is, the 'cheaper' it is to add a planet, such that the two planets model is be favoured. The same experiment as above is done with a flat prior on K on [0, 40] m s^{-1}. In that case, the two planet model is systematically favoured, as shown in Fig. 14.

As a conclusion, it seems good practice to check the influence of the prior on K. Secondly, since we expect that as the period ratio of resonant planets not to be exactly one half, if the two planet hypothesis is true then it will be strongly favoured by the Bayes factor. It therefore seems reasonable to consider the eccentric planet as the null hypothesis and not to reject it if there is no strong evidence for the two planet model.

4 DETECTING MODEL ERRORS: RESIDUAL ANALYSIS

4.1 Objective

In the previous section, we have seen that adjusting a jitter term is satisfactory in most cases, however we did not envision all possible errors. It is good practice to check if the models considered are plausible at all. One approach to take into account model uncertainty is to consider many models (Jones et al. 2017) and rank them via cross validation, Akaike Information Criterion or BIC, or even Bayes factor if possible. Alternately, we can test the hypothesis that the residuals are consistent with the model in an absolute sense. This problem is sometimes referred to as the goodness-of-fit problem, and is in general difficult (see Lehmann & Romano 2005, chap. 14).

We reason as follows: if the set of models is appropriate to describe the data, then the residuals of the best fit must verify certain properties. If they do not, then we reject the hypothesis that there is one acceptable model that explains entirely the data set, among the set of models considered. Furthermore, we would like to obtain hints on the origin of a model misspecification. We expect outliers to change the distribution of the residuals, and astrophysical or instrumental noise to introduce correlations. As a consequence, we consider two types of checks: is the distribution of the residuals approximately Gaussian? and: is there remaining time correlations in the residuals?

4.2 Distribution of the residuals

As in Section 2.2.2, we first examine the linear case and show that the results are still helpful in the non-linear setting. Let us suppose that we have a linear model $\boldsymbol{y} = \mathbf{A}\boldsymbol{\theta} + \boldsymbol{\epsilon}$ where $\mathbf{A}$ is a $N \times p$ matrix and $\boldsymbol{\epsilon}$ is a Gaussian noise of covariance matrix $\mathbf{V} =: \mathbf{W}^{-1}$. Let us denote by $\widehat{\boldsymbol{y}}$ the least-squares fit model, and suppose the model $(\mathbf{A}, \mathbf{V})$ is known. Then the weighted residual

$$\boldsymbol{r}_W := \mathbf{W}^{1/2}(\boldsymbol{y} - \widehat{\boldsymbol{y}}) \tag{40}$$

is a vector of N random variables that are approximately independent, Gaussian of null mean and variance one. To obtain a weighted residual that is a vector of independent Gaussian variables, let us define $\mathbf{Q}$, the matrix such that $\mathbf{J} = \mathbf{Q}^T(\mathbf{I}_N - \mathbf{W}^{1/2}\mathbf{A}^T(\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}^{1/2})\mathbf{Q}$ is diagonal (it exists). Then the reweighted residual $\boldsymbol{r}'_{QW} = \mathbf{Q}^T\mathbf{W}^{1/2}(\boldsymbol{y} - \widehat{\boldsymbol{y}})$ has p null components. The $N - p$ others are Gaussian variables of mean 0 and variance 1. In what follows, we denote by $\boldsymbol{r}_{QW}$ the vector made of the $N - p$ components of $\boldsymbol{r}'_{QW}$. These two results are proven in Appendix E.

In practice, $\mathbf{A}$ and $\mathbf{V}$ are unknown, and we choose models $\mathbf{A}'$ and $\mathbf{V}'$. The two above properties can be used to test if $(\mathbf{A}, \mathbf{V}) = (\mathbf{A}', \mathbf{V}')$ because if so, then the weighted residuals $\boldsymbol{r}_W$ and $\boldsymbol{r}_{QW}$ have a known distribution.

We compute an experimental cumulative distribution function (CDF) of $\boldsymbol{r}_W$ and $\boldsymbol{r}_{QW}$. If our model is correct, then it should be close to the CDF of a Gaussian variable of mean zero and variance one.

4.3 Correlations in the residuals

The test suggested in the previous section is relevant to check the distribution of the residuals without temporal information, and thus is not most adapted to spot correlations. We here adapt the variogram (Matheron 1963) for unevenly sampled time series, similarly to Baluev (2013a). The quantity $d(t_i, t_j) = \boldsymbol{r}_W(t_i) - \boldsymbol{r}_W(t_j)$ is plotted as a function of $t_i - t_j$ for $t_i > t_j$. If $\boldsymbol{r}_W$ is indeed independent and Gaussian, $d(t_i, t_j)$ should not depend on the time interval.

Secondly, we consider n time bins with constant spacing in log t. For each bin B, we compute the sample variance of the $d(t_i, t_j)$ such that $t_i - t_j \in B$. We expect that if there are correlations, these variances should grow as $t_i - t_j$ increases. To obtain an error bar on the variances, we add an independent Gaussian noise of mean 0 and variance 1 to r_W and re-compute the variances for the same time bins. One can alternately add a Gaussian noise of covariance $\mathbf{V}$ to the data, and re-compute the residuals. The error bars are taken as $\pm\sigma$ where σ is the standard deviation of the variances estimates per bin.

4.4 Example

Let us now show how it can be used in practice. We take the 214 measurement times of Proxima b Anglada-Escudé et al. (2016). $\mathbf{A}$ is made of six columns as defined in Appendix A and fix $\boldsymbol{x}_t$. We then generate three series of a thousand realization of $\boldsymbol{y} = \mathbf{A}\boldsymbol{x}_t + \boldsymbol{\epsilon}$. The covariance matrix of the noise has a kernel $\mathrm{e}^{-|\Delta t|/\tau}$ where Δt is the duration between two samples. The three series are generated with a noise time-scale $\tau = 0$, 10 and 100 d. For each of the 3 × 1000 signals generated, we compute the least-squares fit with the correct matrix $\mathbf{A}$, but with a weight matrix $\mathbf{W}$ equal to identity, so our model is entirely correct only in the first case. First, we pick randomly one realization among the 1000 available in each series, and perform the first test whose result is plotted in Fig. 15. One clearly sees a pattern: the higher the correlation, the smaller is the difference between residuals. We then consider six time bins, and compute the variances of the data and their uncertainties within each bin. The results of these calculations are represented by the stair curves. For correlated noises, the variance increases with the time interval, while it stays compatible with a constant for the white noise. Fig. 16 shows the 1000 empirical CDFs in the three cases.

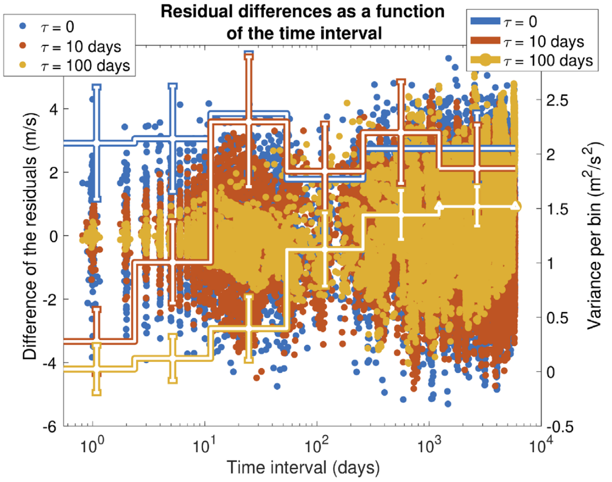

Figure 15. Dots: difference between the residuals at two different time as a function of the time interval between them in three cases: when the noise has a time-scale of 0, 10, and 100 d, units on the left y-axis. Stair curves: variance of the data points per time bin, with uncertainty, units on the right y-axis.

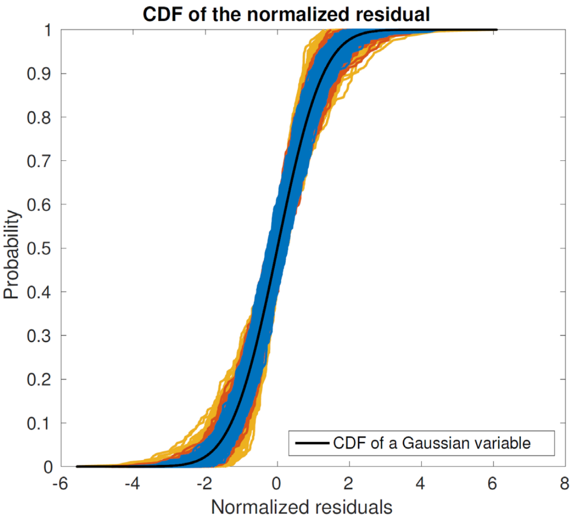

Figure 16. A thousand realizations of the cumulative distribution functions of the normalized residual in three cases: when the noise has a time-scale of 0, 10, and 100 d.

The plots 15 and 16 are useful indicators of remnant correlations in the residuals and non-Gaussianity. However, they do not constitute metrics with known statistical properties. One can potentially test the hypothesis that $\mathbf{r}_{QW}$ is a realization of such a law with a Kolmogorov–Smirnov test or other metrics such as Anderson & Darling (1954), Shapiro & Wilk (1965), etc. We have tested the Anderson–Darling metric which did not show a high statistical power (see Hara 2017, thesis) and that is not discussed further. The visual inspection, though less quantifiable, seems more accurate.

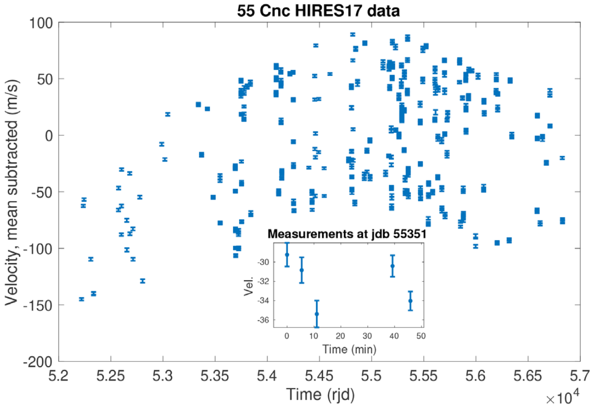

Figure 17. 55 Cnc HIRES17 data with nominal error bars. The box on the bottom is a zoom on measurements taken at rjd 55351.

Study of correlations in RV residuals have already been undertaken for instance by Baluev (2011, 2013b) with a smoothed residual periodogram (Baluev 2009). The systematic comparison of these statistics is left for future work.

5 APPLICATION: 55 CANCRI

To illustrate the methods above, the estimation of the eccentricities of the 55 Cnc system is discussed. This system has been extensively studied and many measurements, from several spectrographs, are publicly available: Hamilton (Marcy et al. 2002; Fischer et al. 2008), ELODIE (Naef et al. 2004), HRS (McArthur et al. 2004; Endl et al. 2012), HIRES (Butler et al. 2017). We here focus on the HIRES (Butler et al. 2017) data set, from now on denoted by HIRES17, which illustrates the importance of the noise model choice for reliable eccentricities.

The HIRES17 data set is made of 607 velocity measurements, with on average four data points per night. Fig. 17 shows these data with nominal error bars. The inset is a zoom on five points taken within a time interval of 50 min at reduced Julian day (RJD) 55351. We first consider the raw data. The posterior distributions of the orbital elements is computed with a model including the five known planets plus a free jitter term as defined in Section 3.2, and priors defined as in Table 4. We now focus on 55 Cnc f, orbiting at 260.9 d with a minimum mass of $0.1503^{+0.0076}_{-0.0076}$ M_J and a period of $259.88^{+0.29}_{-0.29}$ d (Bourrier et al. 2018). In Fig. 18, the posterior distribution of the eccentricity is represented in yellow. The estimate is 0.5 with a 68 per cent credible interval equal to [0.44, 0.58]. This estimate is very different from the one in Bourrier et al. (2018), $e = 0.08^{+0.05}_{-0.0.04}$, where orbital elements estimates are based on all the available radial velocity data.

To check for correlations in the residuals, these are studied with the method of Section 4.3. We define $\mathbf{W} = (\mathbf{V} + \widehat{\sigma}^2_{\rm ML})^{-1}$, where $\mathbf{V}$ is the nominal covariance matrix, and $\widehat{\sigma}^2_{\rm ML}$ is the maximum likelihood estimate of the jitter. The weighted residuals $r_{\rm W}$ as defined in equation (40) are computed. For all combinations of measurement times $t_i > t_j$ we represent $d_{ij} := r_{\rm W}(t_j) - r_{\rm W}(t_i)$ as a function of $t_j - t_i$. We then compute the variance of the d_{ij} such that $t_j - t_i$ is in a certain time bin. Fifteen such intervals are considered, with a constant length in log scale. The uncertainty on this variance σ^2 is estimated by bootstrap. The results are shown in Fig. 19, where the blue points represent a couple $(t_j - t_i, d_{ij})$, and the red stair curve represents the variance per bin and its uncertainty. As the time difference grows, the residuals are more and more dispersed, which is indicative of correlated noise .

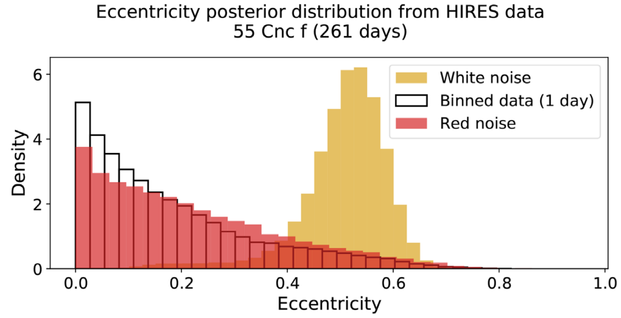

Figure 18. Posterior distributions and confidence intervals of the eccentricity of 55 Cnc f for different models of noise, raw and binned data, HIRES17 data. The histograms represent the posterior distributions obtained with different noise models: white noise and red + white noise for the raw data (respectively in yellow and red), and a white noise model on the binned data (in white).

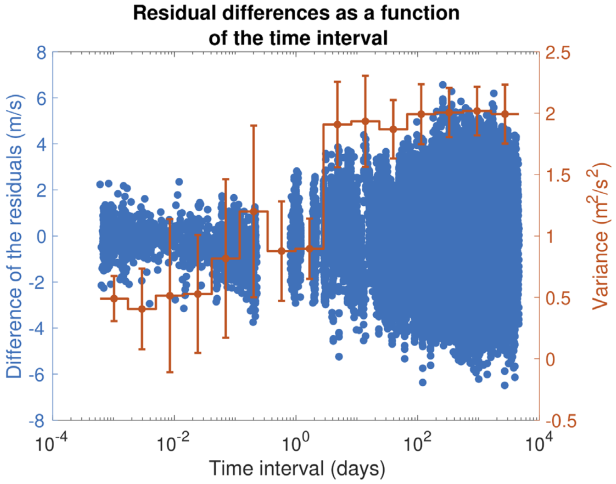

Figure 19. Difference between all couples of residuals with a five planets + offset and white noise fit on the HIRES17 raw data as a function of the time difference between residuals (see Section 4.3). The red stair curves represent the variance of the residuals on a constant step in log time with an estimate of the error on this variance.

To account for this correlated noise, we adopt two strategies. First, we include a red noise term in the model of Section 3.2. This one is assumed Gaussian with a covariance kernel

$$k(t, t') = \sigma_R^2 \, \mathrm{e}^{-\frac{(t-t')^2}{2\tau^2}}, \tag{41}$$

and σ_R and τ are included in the posterior sampling with a normal prior on $\ln \tau$ of mean and variance 1. The resulting eccentricity posterior for 55 Cnc f is shown in red in Fig. 18. The posterior median of τ is 2.3 d, with a 68 per cent confidence interval of [0.5, 9.6] d. The second strategy we adopt is to bin the velocity measurements per day. Such a strategy is actually what likely lead to the measurement pattern chosen, as binning data per night averages out stellar oscillations (Dumusque et al. 2011). The posterior so obtained is represented by the white histogram in Fig. 18. In both cases it appears that that the eccentricity is in fact much less constrained. Note that the posterior distribution favours a null eccentricity. We attribute this to the choice of the beta prior, which favours small eccentricities, but also to the fact described in

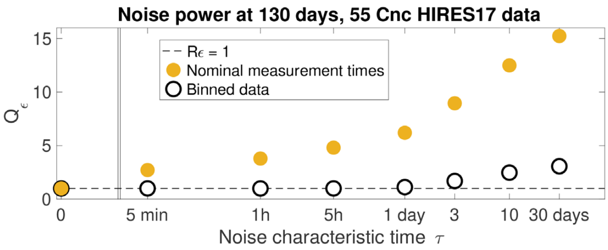

Figure 20. Q_ϵ as defined in equation (27) at 130 d for the measurement times of HIRES17 data, as a function of the assumed noise characteristic time. The yellow dots correspond the nominal measurements, and the white dots correspond to the data binned per day.

Section 2.3: high eccentricity models represent a larger volume of models, and are therefore be penalized by Bayesian estimates.

In Section 3.4.3, we argued that the main characteristic of a noise that influences eccentricity estimation is, for a particular time sampling, its power at the semiperiod of the planet. The results above are indeed interpretable in this term. The quantity $Q_\epsilon(\mathbf{M_2})$ defined in equation (27) is computed for noises with a kernel defined in equation (41). The calculation is made with the nominal time sampling and the one corresponding to a one day binning. Fig. 20 shows the result of these calculations. With the nominal times, the power at 130 d is much more sensitive to correlated noise, even at time-scales as short as 5 min, which results in a clear eccentricity overestimation. This example illustrates that due to the sampling, some noises might create spurious eccentricities if not properly accounted for.

As a remark, some measurement dates correspond to the observation of the 55 Cnc e transit. As noted in Bourrier et al. (2018), the effect of the Rossiter–McLaughlin effect has been put in question in López-Morales et al. (2014), and should not exceed 0.5 m s^{-1}, which is well below HIRES precision and does not impact the analysis above.

6 CONCLUSION

6.1 Summary and step-by-step method

In Section 2, we have seen that at low eccentricity, the bias of the least square estimate can be approximated analytically (equation 12). This equation as well as numerical simulations show that the bias is approximately proportional to the uncertainty on eccentricity. We show that an incorrect estimate of the noise level can create spurious global minima at high eccentricity through the Proxima b example, and that at high SNR it is less likely, though not impossible, to miss the global minimum by doing a least-squares fit initialized at $e = 0$. In Section 2.3, it appeared that the maximum of the marginal distribution of eccentricity (equation 20) is less biased than the maximum likelihood.

In Section 3, we explored the reliability of the maximum likelihood and posterior distributions when the noise model is incorrect, along with the sensitivity of the inference to numerical methods. In summary, we recommend the following data analysis method for robust inference of eccentricities.

(i) Computing the posterior distribution of eccentricity leads to correct inference in general (see Section 2.3.2), if one includes in the noise model at least a free jitter term (equation 24). Without such a term the inference is very likely to be spurious. The mean and median of the marginal posterior distribution of eccentricity, defined in Section 2.3, constitute good point estimates (see Section 3.4.2). Credible intervals (equation 23) allow reliable hypothesis testing, confidence intervals (see Section 2.2.4) are an alternative for probing the low eccentricity region especially.

(ii) Using a white, Gaussian noise model might however lead to spurious inferences in some cases. Eccentricity estimates are weakly sensitive to non-Gaussianity of the noise, but are impacted by its true covariance, especially by the noise power at $P/2$ and to a lesser extent at $P/3$ where P is the period of the planet of interest (see Section 3.4.3). Stellar signals might have a modulation at the rotation period, but as shown in Section 5, correlated noise combined with the sampling can also create spurious eccentricities. The effects on the bias of the noise level and nature are summarized in equation (28). Inappropriate models leave signatures in the residuals, these ones can be studied with techniques presented in Section 4.

(iii) It is key to check the convergence of the MCMC used for posterior calculation, for instance with the effective number of independent samples (we recommend at least 10 000). An unreliable numerical method increases the bias (see Section 3.3).

(iv) At low SNR, posteriors are dominated by the prior and the parameter space is harder to explore. This might lead to spurious inferences (e.g. in Section 3.6). The influence of the prior can be assessed with the method of Section 3.5.2.

(v) The most degenerate case is an inner planet in 2:1 resonance. As discussed in Section 3.6, Bayes factors can disentangle those cases. However, the value of the Bayes factor strongly depends on the prior used for the semi-amplitude. We suggest to select a log prior on semi-amplitude and to consider the eccentric scenario as the null hypothesis.

(vi) In multiplanetary systems, checking the system stability might also help ruling out some values of the eccentricity (e.g. Hébrard et al. 2016; Delisle et al. 2018).

(vii) As more measurements are obtained, least square and Bayesian estimates get closer, so that equation (13) can also be used to approximate the number of measurements $N(\epsilon_e)$ needed to have an average bias on eccentricity ϵ_e for a planet with semi-amplitude K, p parameters fitted in total and a measurement root mean square RMS,

$$N(\epsilon_e) = p + \pi \frac{\mathrm{RMS}^2}{K^2 \epsilon_e^2}. \tag{42}$$

This formula is for $\epsilon_e \leq 0.05$ and subtends that there is a good phase coverage.

This procedure should be familiar to observers, since it is mainly a formalized and tested version of common practices. It is applicable to other purposes, especially the methods of Section 4, since checking the model validity improves the inference robustness in general.

6.2 Perspectives

The main point of this work is that modelling errors might have systematic impacts on the estimates of orbital elements, and thus on our understanding of planetary systems. It is yet to be determined in which extent this has been the case in past studies.

In Section 4, we presented tools to measure the absolute adequacy between a model and the data, which is complementary to comparing models to one another. There remain many such adequacy metrics to explore. These ones could prove useful in the context of exoplanets characterization but also exoplanets detection.

ACKNOWLEDGEMENTS

NH and JD acknowledge the financial support of the National Centre for Competence in Research PlanetS of the Swiss National Science Foundation (SNSF). NH thanks J. J. Zanazzi for his interesting inputs. We thank the anonymous referee for his/her insightful suggestions.

REFERENCES

Anderson T. W., Darling D. A., 1954, J. Am. Stat. Assoc., 49, 765
Anglada-Escudé G., López-Morales M., Chambers J. E., 2010, ApJ, 709, 168
Anglada-Escudé G. et al., 2016, Nature, 536, 437
Baluev R. V., 2009, MNRAS, 393, 969
Baluev R. V., 2011, Celest. Mech. Dyn. Astron., 111, 235
Baluev R. V., 2013a, MNRAS, 429, 2052
Baluev R. V., 2013b, MNRAS, 436, 807
Baluev R. V., 2015, MNRAS, 446, 1478
Bates D. M., Watts D. G., 1980, J. R. Stat. Soc. B, 42, 1
Boisvert J. H., Nelson B. E., Steffen J. H., 2018, MNRAS, 480, 2846
Bonomo A. S. et al., 2017a, A&A, 602, A107
Bonomo A. S. et al., 2017b, A&A, 603, A43
Bourrier V. et al., 2018, A&A, 619, A1
Brown R. A., 2017, ApJ, 844, 100
Butler R. P. et al., 2017, AJ, 153, 208
Casella G., Berger R., 2001, Statistical Inference. Duxbury Resource Center, CA
Cook R. D., Witmer J. A., 1985, J. Am. Stat. Assoc., 80, 872
Correia A. C. M. et al., 2010, A&A, 511, A21
Cumming A., 2004, MNRAS, 354, 1165
Delisle J.-B. et al., 2018, A&A, 614, A133
Dumusque X., Udry S., Lovis C., Santos N. C., Monteiro M. J. P. F. G., 2011, A&A, 525, A140
Eastman J., Gaudi B. S., Agol E., 2013, PASP, 125, 83
Endl M. et al., 2012, ApJ, 759, 19
Firth D., 1993, Biometrika, 80, 27
Fischer D. A. et al., 2008, ApJ, 675, 790
Ford E. B., 2005, AJ, 129, 1706
Ford E. B., 2006, ApJ, 642, 505
Haario H., Saksman E., Tamminen J., 2001, Bernoulli, 7, 223
Handley W. J., Hobson M. P., Lasenby A. N., 2015a, MNRAS, 450, L61
Handley W. J., Hobson M. P., Lasenby A. N., 2015b, MNRAS, 453, 4384
Hara N. C., 2017, PhD thesis, Observatoire de Paris
Hartkopf W. I., McAlister H. A., Franz O. G., 1989, AJ, 98, 1014
Hartley H. O., 1964, Biometrika, 51, 347
Haywood R. D. et al., 2014, MNRAS, 443, 2517
Hébrard G. et al., 2016, A&A, 588, A145
Hobson M. J. et al., 2018, A&A, 618, A103
Hogg D. W., Foreman-Mackey D., 2018, ApJS, 236, 11
Hogg D. W., Myers A. D., Bovy J., 2010, ApJ, 725, 2166
Hurwicz L., 1950, in Koopmans T., ed., Statistical Inference in Dynamic Economic Models. Wiley, New York, p. 365
Husnoo N., Pont F., Mazeh T., Fabrycky D., Hébrard G., Moutou C., 2011, in Sozzetti A., Lattanzi M. G., Boss A. P., eds, Proc. IAU Symp. 276, The Astrophysics of Planetary Systems: Formation, Structure, and Dynamical Evolution. p. 243
Husnoo N., Pont F., Mazeh T., Fabrycky D., Hébrard G., Bouchy F., Shporer A., 2012, MNRAS, 422, 3151
Jensen J. L. W. V., 1906, Acta Math., 30, 175
Jones D. E., Stenning D. C., Ford E. B., Wolpert R. L., Loredo T. J., Dumusque X., 2017, preprint (arXiv:1711.01318)
Jurić M., Tremaine S., 2008, ApJ, 686, 603
Kant E., 1755, Allgemeine Naturgeschichte und Theorie des Himmels, Available at: https://web.archive.org/web/20081216021207/http://records.viu.ca/~johnstoi/kant/kant2e.htm
Kass R. E., Raftery A. E., 1995, J. Am. Stat. Assoc., 90, 773
Kenney J. F., Keeping E. S., 1962, 'The Median,' 'Relation Between Mean, Median, and Mode,' 'Relative Merits of Mean, Median, and Mode,' and 'The Median', 3rdedn. Princeton Univ. Press, Princeton, NJ, p. 32
Kipping D. M., 2014, MNRAS, 444, 2263
Kürster M., Trifonov T., Reffert S., Kostogryz N. M., Rodler F., 2015, A&A, 577, A103
Laplace P.-S., 1796, Exposition du système du monde. Imprimerie du Cercle-Social, Available at: https://fr.wikisource.org/wiki/Exposition_du_syst%C3%A8me_du_monde
Laskar J., 2008, Icarus, 196, 1
Lehmann E. L., Romano J. P., 2005, Testing Statistical Hypotheses, 3rdedn. Springer Texts in Statistics, Springer, New York
Levenberg K., 1944, Q. Appl. Math., 2, 164
López-Morales M. et al., 2014, ApJ, 792, L31
Lovis C. et al., 2006, Nature, 441, 305
Lucy L. B., 2005, A&A, 439, 663
Lucy L. B., 2013, A&A, 551, A47
Lucy L. B., 2014, A&A, 565, A37
Lucy L. B., Sweeney M. A., 1971, AJ, 76, 544
Marcy G. W., Butler R. P., Fischer D. A., Laughlin G., Vogt S. S., Henry G. W., Pourbaix D., 2002, ApJ, 581, 1375
Marquardt D. W., 1963, SIAM J. Appl. Math., 11, 431
Matheron G., 1963, Econ. Geol., 58, 1246
Mayor M. et al., 2009, A&A, 493, 639
McArthur B. E. et al., 2004, ApJ, 614, L81
Naef D., Mayor M., Beuzit J. L., Perrier C., Queloz D., Sivan J. P., Udry S., 2004, A&A, 414, 351
Nagel E. et al., 2019, A&A, 622, A153
Nelson B. E., Ford E. B., Buchner J., Cloutier R., Díaz R. F., Faria J. P., Rajpaul V. M., Rukdee S., 2018, preprint (arXiv:1806.04683)
O'Toole S. J., Tinney C. G., Jones H. R. A., Butler R. P., Marcy G. W., Carter B., Bailey J., 2009a, MNRAS, 392, 641
O'Toole S. J., Tinney C. G., Jones H. R. A., Butler R. P., Marcy G. W., Carter B., Bailey J., 2009b, MNRAS, 392, 641
Pelat D., 2013, Bases et méthodes pour le traitement de données, Observatoire de Paris. Available at: https://media4.obspm.fr/public/M2R/supports/Dea45beta.pdf
Petrovich C., Tremaine S., Rafikov R., 2014, ApJ, 786, 101
Pont F., Husnoo N., Mazeh T., Fabrycky D., 2011, MNRAS, 414, 1278
Satterthwaite F. E., 1946, Biometrics Bull., 2, 110
Schwarz G., 1978, Ann. Stat., 6, 461
Shapiro S. S., Wilk M. B., 1965, Biometrika, 52, 591
Shen Y., Turner E. L., 2008, ApJ, 685, 553
Sokal A., 1997, Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms. Springer, Boston, MA
Steffen J. H., Hwang J. A., 2015, MNRAS, 448, 1956
Swedenborg E., 1734, Opera Philosophica et Mineralia.
Welch B. L., 1947, Biometrika, 34, 28
Wittenmyer R. A. et al., 2013, ApJS, 208, 2
Wittenmyer R. A., Bergmann C., Horner J., Clark J., Kane S. R., 2019, MNRAS, 484, 4230
Zakamska N. L., Pan M., Ford E. B., 2011, MNRAS, 410, 1895
Zechmeister M., Kürster M., 2009, A&A, 496, 577

APPENDIX A: FIRST-ORDER APPROXIMATION

In this section, we developp the Keplerian model to first order in eccentricity to obtain an analytical expression of the bias. Within this approximation, the distribution of the least-squares fit knowing that $e = 0$ is given in Lucy & Sweeney (1971). This section extends their formula to small e, and takes into account the number of fitted parameters. First, we develop (4) to order one in e, obtaining

$$y(\lambda, K, P, e, \omega) = K(\cos(\lambda) + e\cos(2\lambda - \omega)), \tag{A1}$$

where $\lambda = nt + \omega + M_0 = \lambda_0 + 2\pi t/P$ is the mean longitude, λ_0 being its value at $t = 0$. Denoting by $n = 2\pi/P$ the mean motion, the above expression can be re-written

$$y^{(1)}(t, A, B, C, D, n) = A\cos nt + B\sin nt + C\cos 2nt + D\sin 2nt, \tag{A2}$$

where $A = K\cos\lambda_0$, $B = -K\sin\lambda_0$, $C = Ke\cos(2\lambda_0 - \omega)$, $D = -Ke\sin(2\lambda_0 - \omega)$. When other parameters are fitted, the uncertainties on A, B, C, D increase as well. To quantify this effect, we consider the problem of fitting the period and a constant.

$$\boldsymbol{y}^{(2)}(t, A, B, C, D, E, F) = A\cos nt + B\sin nt + C\cos 2nt + D\sin 2nt + E\frac{\partial y}{\partial n}(t) + F. \tag{A3}$$

which in a matrix form gives

$$\boldsymbol{y}^{(2)}(\boldsymbol{t}, A, B, C, D, E, F) = \mathbf{M}(P)\boldsymbol{x}. \tag{A4}$$

Let us assume that the observations are $\boldsymbol{y}(t) = \mathbf{M}(P)x_{\rm t} + \epsilon$, where ϵ is a Gaussian noise, independent and identically distributed with variance σ^2. The least square estimate of $\boldsymbol{x}$ is $\widehat{\boldsymbol{x}} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\boldsymbol{y}$, and the estimate of eccentricity is

$$\widehat{e} = \sqrt{\frac{\widehat{C}^2 + \widehat{D}^2}{\widehat{A}^2 + \widehat{B}^2}} = \frac{\sqrt{\widehat{C}^2 + \widehat{D}^2}}{K_{\rm t}}\left(\frac{\sqrt{\widehat{A}^2 + \widehat{B}^2}}{K_{\rm t}}\right)^{-1}, \tag{A5}$$

where $K_{\rm t}$ is the true semi-amplitude. By change of random variable we can obtain the law followed by $\widehat{e}$. If we assume that N is large enough then the columns of $\mathbf{M}(P)$ are approximately orthogonal, the components of $\widehat{x}$ are independent Gaussian variables. Since the modulus of a sum of independent Gaussian variables follows a Rice distribution,

$$U \equiv \frac{\sqrt{\widehat{C}^2 + \widehat{D}^2}}{K_{\rm t}} \sim g(u) = S^2 u\, {\rm e}^{-\frac{S^2}{2}(u^2 + e_{\rm t}^2)}\, I_0(S^2 eu) \tag{A6}$$

$$W \equiv \frac{\sqrt{\widehat{A}^2 + \widehat{B}^2}}{K_{\rm t}} \sim h(w) = S^2 w\, {\rm e}^{-\frac{S^2}{2}(w^2 + 1)}\, I_0(S^2 w), \tag{A7}$$

where I_0 is a modified Bessel function of first kind, $S = K_{\rm t}/\sigma$ is the SNR, where σ is the standard deviation of $\widehat{A}$, $\widehat{B}$, $\widehat{C}$, and $\widehat{D}$. If K is sufficiently large, W is close to 1 and $g(u)$ gives a good approximation of the law followed by the eccentricity fitted. Within this approximation, one can obtain analytical formula for the bias b of the eccentricity that only depends on the true eccentricity and the SNR,

$$b(e_{\rm t}, S, n) = \frac{1}{S}\sqrt{\frac{\pi}{2}}L_{1/2}\left(\frac{S^2 e_{\rm t}^2}{2}\right) - e_{\rm t}, \tag{A8}$$

where $L_{1/2}$ is the Laguerre polynomial of order 1/2. In case $K_{\rm t}$ is small, one must use the formula for the law followed by the quotient of two random variables:

$$\widehat{e} = \frac{U}{W} \sim f(e) = \int_{-\infty}^{+\infty} g(u)h(ue)|u|\,{\rm d}u \tag{A9}$$

but no simple analytical expression was found.

When fitting model (A4) to $\boldsymbol{y}(t)$, the estimate $\widehat{\theta}$ have a covariance matrix $\boldsymbol{\Sigma}^{-1}$ where $\boldsymbol{\Sigma} = \sigma^2(\mathbf{M}(P)^T\mathbf{M}(P))$ [this is a classical statistical result, see for example Pelat (2013)]. The variances of the components of $\widehat{x}$ are given by the diagonal elements of $\boldsymbol{\Sigma}^{-1}$. Their approximate calculation is the object of the next section.

A1 Average error

First we consider the estimation of the error on A, B, C, D when averaging over the mean motion n. At little cost, we can generalize our claim to the fitting of model (A3) plus fitting other linearized Keplerian model. This approximately corresponds to fitting a multiplanetary system starting closely from the correct local minimum of χ^2. Again, the model can be written as a linear one, $\boldsymbol{y} = \mathbf{M}\boldsymbol{x}$ but where $\mathbf{M}$ has $p = 6 + 5k$ columns, k being the number of additional planets. To facilitate the discussion, we normalize the columns of $\mathbf{M}$. To have an expression of the model of the form (A4), we have multiplied the kth component of θ by the norm of the kth column of M. The variances of these new model parameters are still given by the diagonal elements of $\sigma^2\boldsymbol{\Sigma}^{-1}$ where $\boldsymbol{\Sigma} = (\mathbf{M}^T\mathbf{M})$, but now $\boldsymbol{\Sigma}$ has only ones on its diagonal.

Calculating precisely the uncertainty on A, B, C, D averaged over n and the phase of the signal as a function of the instant of observations t is complex since it requires the inversion of σ which is a $6 + 5k \times 6 + 5k$ matrix. Instead, we use an approximation that grasps the effect we want to estimate: how the uncertainty worsens as more parameters are added to the model. We consider that the elements of $\mathbf{M}$ are drawn from independent Gaussian laws that have a variance $1/N$. To avoid confusion with the true model, the so defined random matrix is denoted by $\tilde{\mathbf{M}}$ and its covariance matrix by $\tilde{\boldsymbol{\Sigma}}$.

This approximation seems to be rough at first but turns out to be surprisingly accurate as a lower bound in practice. A few arguments to justify that it is a reasonable guess are listed below.

(i) The variances of the entries were chosen such that the expectancy of a squared norm of a column is one, which is the value of $\boldsymbol{\Sigma}$ diagonal elements.

(ii) The columns are cosines and sines, which are approximately orthogonal, and in the Gaussian case decorrelation implies independence. Furthermore, the average of the spectral window is equal to the expected value of a correlation between two random Gaussian variables.

(iii) The normed vectors $\cos\nu t$ and $\sin\nu t$ are approximately distributed uniformly on the sphere of $\mathbb{R}^N$ when ν is distributed uniformly between 0 and $2\pi/T_{\rm obs}$.

The expected value of the variance of any parameter is the expected value of any diagonal element of $\boldsymbol{\Sigma}^{-1}$, since all the columns of $\tilde{\mathbf{M}}$ follow the same law. To tackle that problem, we rewrite $\tilde{\boldsymbol{\Sigma}}$ as

$$\tilde{\boldsymbol{\Sigma}} = \sigma^2\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_1^T \\ \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_c\end{pmatrix},$$

where $\boldsymbol{\Sigma}_{11}$ is $\tilde{\boldsymbol{\Sigma}}$ element at first row and first column and $\boldsymbol{\Sigma}_1$ is a column vector with $N - 1$ entries. We now have

$$\mathbb{E}\left\{\tilde{\boldsymbol{\Sigma}}_{11}^{-1}\right\} = \frac{1}{\sigma^2}\mathbb{E}\left\{\frac{1}{\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_1^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\Sigma}_1}\right\}.$$

By Jensen inequality (Jensen 1906), since $x \rightarrow 1/x$ is convex,

$$\mathbb{E}\left\{\tilde{\boldsymbol{\Sigma}}_{11}^{-1}\right\} \leqslant \frac{1}{\sigma^2}\frac{1}{\mathbb{E}\left\{\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_1^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\Sigma}_1\right\}}.$$

Now since for two independent variables X and Y, $\mathbb{E}\{XY\} = \mathbb{E}\{X\}\mathbb{E}\{Y\}$,

$$\boldsymbol{\Sigma}_1^T \boldsymbol{\Sigma}_c^{-1} \boldsymbol{\Sigma}_1 = \sum_{k=2}^{p} \mathbb{E}\left\{\boldsymbol{\Sigma}_{1k}^2\right\} \mathbb{E}\left\{\boldsymbol{\Sigma}_{c,kk}^{-1}\right\} \leqslant \sum_{k=2}^{p} \mathbb{E}\left\{\boldsymbol{\Sigma}_{1k}^2\right\}$$

$$= 1 - \frac{p-1}{N}.$$

As by construction $\mathbb{E}\left\{\boldsymbol{\Sigma}_{11}\right\} = 1$, we finally obtain

$$\mathbb{E}\left\{\tilde{\boldsymbol{\Sigma}}_{11}^{-1}\right\} \leqslant \frac{1}{\sigma^2} \frac{1}{1 - \frac{p-1}{N}},$$

where the inequality follows again from Jensen's inequality applied to matrix inversion. Finally, the standard deviation on $I = A$, B, C, D is

$$\sigma_I \geqslant \sigma \sqrt{\frac{1}{1 - \frac{p-1}{N}}}. \tag{A10}$$

With the approximation $\| \cos \nu \boldsymbol{t} \| \approx \| \sin \nu \boldsymbol{t} \| \approx \sqrt{N/2}$, the errors on $k = C/\sqrt{A^2 + B^2}$ and $k = D/\sqrt{A^2 + B^2}$ then verify

$$\sigma_k \gtrsim \frac{\sigma}{K_{\mathrm{t}}} \sqrt{\frac{2}{N}} \sigma_I = \frac{\sigma}{K_{\mathrm{t}}} \sqrt{\frac{2}{N - p + 1}} \approx \frac{\sigma}{K_{\mathrm{t}}} \sqrt{\frac{2}{N - p}} =: 1/S. \tag{A11}$$

As k and h approximately follow a Gaussian law, $e = \sqrt{k^2 + h^2}$ follows a Rice distribution, whose mean is given by

$$\mathbb{E}\{\widehat{e}\} = \frac{1}{S} \sqrt{\frac{\pi}{2}} L_{1/2} \left(-\frac{S^2 e_{\mathrm{t}}^2}{2} \right). \tag{A12}$$

$$\mathbb{E}\{\widehat{e} | e_{\mathrm{t}} = 0\} = \frac{\sigma}{K_{\mathrm{t}}} \sqrt{\frac{\pi}{N - p}}. \tag{A13}$$

$L_{1/2}$ being the Laguerre polynomial of degree 1/2. The relevance of formula (A12) is checked on numerical examples next section. As we shall see, the lower bound is tight when p does not exceeds $\approx N/2$.

Let us finally stress that formula (A12) approximates the bias averaged on the mean motion, that is the frequency of the orbit, not the period. Averaging on the period would give more weight to the bias at low frequencies, which is high, and would therefore lead to a greater average value of the bias.

A2 Precision and accuracy

Equation (3) expresses the MSE as a function of the bias and the standard deviation. Assuming the model is correct, the MSE is an accuracy metric (dispersion about the true value), while the variance captures the precision of the estimate (dispersion of the estimate about its mean value). Shen & Turner (2008); Zakamska et al. (2011) have shown that the estimates accuracy degrades – all other parameters being fixed – as the SNR decreases, as the period of the planet is longer, and as phase coverage degrades.

In this appendix, we argue that these effects can be seen as degrading the precision of the eccentricity estimates. More precisely, we show that the bias is proportional to the standard deviation of the estimate, so that the MSE is also proportional to the standard deviation.

We proceed with a numerical simulation. We consider the 74 measurement times of HD 69830 (Lovis et al. 2006), spanning on 800 d, as they are spaced in a typical manner. We then inject a simulated planet in circular orbit and a white, Gaussian noise of standard deviation 1 m s^{-1}. In all the following simulations, the phase is uniformly random. By default, the semi-amplitude of the planet is $K = 3$ m s^{-1} and the period is 31.56 d (like HD 69830 c).

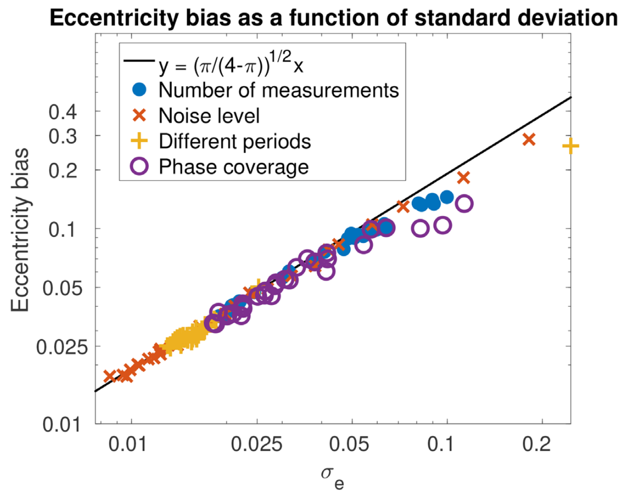

Figure A1. Bias as a function of the standard deviation of the eccentricity estimates in different configurations.

(i) Data points are taken off two by two, from 74 to 14.

(ii) The semi-amplitude of the planet varies from 0.5 to 6.3 m s^{-1} by a step of 0.2 m s^{-1}.

(iii) The period is drawn from a lognormal law, where $\log_{10} P \sim G(1, 1)$. Thirty different periods are drawn.

(iv) The phase coverage is degraded. We consider the 25 times $t_k = k \times P$ where $P = 31.56$ is the planet period. For each t_k, three epochs of measurements are drawn uniformly between t_k and $t_k + \Delta t$. We choose thirty different lengths for Δt, equispaced from $P/4$ to P. The rationale is to generate observations more or less localized around the same time when folded in phase at P.

Each of the four simulations is made with 30 different parametrizations. For each of the 30 × 4 of these, we generate 500 realizations of white noise and report the average value of eccentricity (the bias) and the standard deviation of the estimate. The results are reported in Fig. A1, where it appears that the bias is proportional to the standard deviation. The points obtained with simulations 1 to 4 described above correspond to the blue, red, yellow, and purple points on the graph. The analytical approximation (13) suggests that the bias should be proportional to the standard deviation of the estimate with a factor $\sqrt{\pi/(4-\pi)}$. The black line, which represents $y = \sqrt{\pi/(4-\pi)} x$ is in close agreement with the scatter observed.

The σ_e reported in Fig. A1 is computed as the standard deviation of the estimate. Note that the bias is also proportional to the uncertainty on eccentricity computed from the correlation matrix, obtained from the least-squares fit.

APPENDIX B: LOCAL χ^2 MINIMA AT HIGH ECCENTRICITIES

As shown in Baluev (2015), the number of local minima increases significantly in the high eccentricities region. These minima might lead a local minimization algorithm or an MCMC to be stuck in the wrong region of the parameter space. We here aim at quantifying and understanding this feature.

In Section 2.2.2, we defined a notion of SNR (equation 11). Interestingly enough, one can adapt this notion to predict the number of local minima at high eccentricity, provided the value of the SNR is checked to be reliable.

B1 Number of local minima per SNR

We explore through simulations how many local minima should be expected, and how frequently the local minimum given by a non-linear fit starting at a circular orbit is not the global minimum. Our simulation is structured as follows. We consider the measurement times of CoRoT-9 (Bonomo et al. 2017b), Gl96 (Hobson et al. 2018), of HD69830 (Lovis et al. 2006), HD40307 (Mayor et al. 2009), which have respectively 28, 67, 74, and 129 observations and generate a planet with uniform distribution in e, ω, M_0, and K (the latter on [0,3]). The period is drawn from a log normal distribution ($\log_{10}P \sim G(1.5,\,1)$). The uncertainties on the measurements are taken as the nominal ones that are normalized to obtain a mean variance of 1. We then obtain the fitted uncertainty $\sigma_{\rm fit}^2$.

For each realization of the true planet and the noise, we perform first a local minimization, initialized at a circular orbit. We then compute three quantities at this local minimum: the condition number of the Fisher matrix **F**, as well as the SNR, similarly to equation (11),

$$c_F = \frac{\max_{i=1..N}\lambda_i}{\min_{i=1..N}\lambda_i} \quad \text{with} \quad (\lambda_i)_{i=1..N} = \text{eigenvalues}(\mathbf{F}) \tag{B1}$$

$$S_{\rm t} = \frac{K_{\rm t}}{\sigma_{\rm t}}\sqrt{\frac{N-p}{2}} \tag{B2}$$

$$S_{\rm fit} = \frac{K_{\rm fit}}{\sigma_{\rm fit}}\sqrt{\frac{N-p}{2}}, \tag{B3}$$

where $K_{\rm t}$ and $K_{\rm fit}$ are respectively the true and fitted values of the semi-amplitude, N is the number of measurements, p the number of fitted parameters. $\sigma_{\rm fit}$ is defined as follows. We adjust a term σ_J^2 so that the reduced χ^2 is equal to one and take $\sigma_{\rm fit}$ as the mean of $(\sigma_J^2 + \sigma_k^2)_{k=1..N}$. The rationale of taking this definition instead of the true SNR is to have a quantity that does not require to know the true orbital elements and can be computed on a real data set.

Secondly, we perform a Keplerian periodogram on a closely spaced grid of $e \in [0,\,1]$, $\omega \in [0,\,2\pi)$, and $n \in [n_{\rm t} - 1/T_{\rm obs},\, n_{\rm t} + 1/T_{\rm obs}]$, where $n_{\rm t}$ is the true mean motion. We compute the value of the eccentricity where the global minimum of χ^2 is attained, as well as the number of local minima. For instance, in Fig. 3, the red curve displays four local minima at $e = 0.17$, 0.75, 0.92, and 1, and the global minimum is attained at $e = 0.92$. In that case, a local minimization would not give the global minimum.

For both definitions of the SNR, we compute the number of systems that have an SNR between $5k$ and $5(k+1)$, $k = 0..7$ and greater than 35. In each bin, we compute the number of systems with 1, 2, 3, 4, 5, 6, or 7 local minima. The results are represented in Figs B1 and B2 for $S_{\rm t}$ (equation B2) resp. $S_{\rm fit}$ (equation B3) by the blue histograms. We then compute in each bin the fraction of systems where the global minimum is not the local minimum closest to 0, which means that most likely, a local minimization does not yield the global minimum (red stair curve, see the right y-axis scale). For instance, in Fig. B1, we see that out of the 2500 systems simulated, 472 had a S_t between 0 and 5, 176 of which had only one local minimum, 203 had 2, 84 had 3, 9 had 4, and none had more. Among those 472 systems, 34 per cent of them had a local minimum that is not the global minimum.

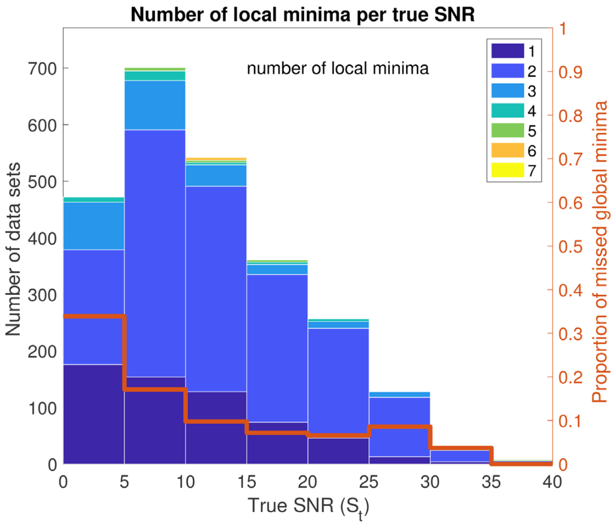

Figure B1. Blue bins: binned values of the number of systems with 1, 2, 3, 4, 5, 6, or 7 local minima, with a bin size in true SNR $S_{\rm t}$ of 5. Red curve: fraction of the binned systems where the global minimum is not attained at the one obtained with a linear fit.

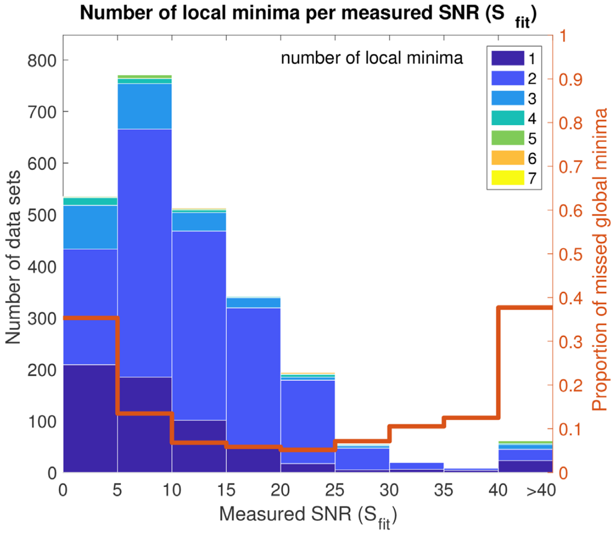

Figure B2. Blue bins: binned values of the number of systems with 1, 2, 3, 4, 5, 6, or 7 local minima, with a bin size in fitted SNR $S_{\rm fit}$ of 5. Red curve: fraction of the binned systems where the global minimum is not attained at the one obtained with a linear fit.

Note that in Fig. B2, the systems with $S_{\rm fit} > 35$ present the highest proportion of missed global minima. This is due to the fact that no systems with very high SNR where generated. As a consequence, all the very high values of $S_{\rm fit}$ result from datasets with very low $S_{\rm t}$ where K was very overestimated. To obtain a more reliable diagnostic, we need to determine if a fitted SNR can be trusted. The criterion we used is to select only the data sets where $c_{\rm F} < 10^7$. The Fig. B3 shows the number of local minima per bin of $S_{\rm fit}$ so obtained. The predominance of systems with two local minima is due to the fact that at very high eccentricity, there is in general a decrease of χ^2. As a consequence, the second local minima is attained at the maximum eccentricity used for the calculation of the

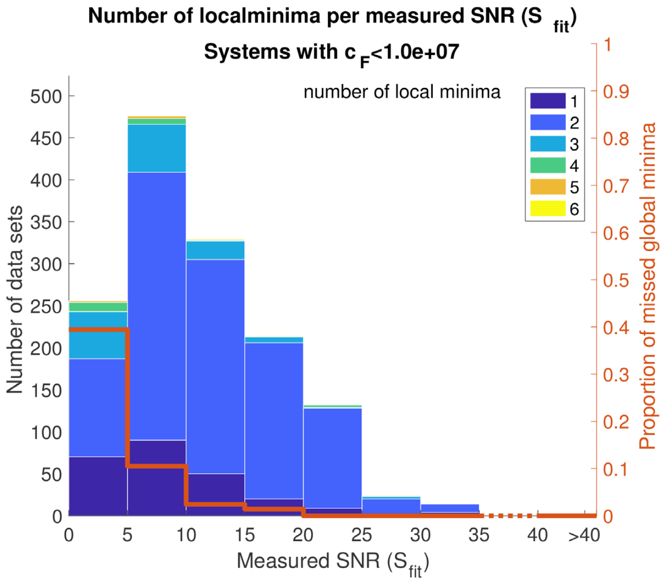

Figure B3. Blue bins: binned values of the number of systems that have a condition number lower than 10^7 with 1, 2, 3, 4, 5, or 6 local minima, with a bin size in fitted SNR S_{fit} of 5. Red curve: fraction of the binned systems where the global minimum is not attained at the one obtained with a linear fit.

global periodogram. For instance in the case of Proxima b (Fig. 3), there is such a minimum at $e = 0.999$. Note if we do the same analysis per data set (CoRoT-9, Gl96, HD69830, HD40307), we obtain very similar figures, which shows that indeed the SNR is a reliable metric for the number of local minima. As a conclusion, we expect that the exploration of the parameter space to be more difficult at low SNR, since there are more local minima to expect.

B2 Interpretation

We now give a geometrical interpretation of why the high eccentricity region is prone to having local minima.

Finding the best-fitting model amounts to finding the model closest to the observation in a geometrical sense. We consider the figure drawn in $\mathbb{R}^N$ by all the models that have an eccentricity e and a period P, $\mathcal{M}_{e,P}$. This figure might explore more or less dimensions. For instance, if it is close to a plane, it is nearly confined to a two-dimensional space. Otherwise, exploring many dimensions traduces a ‘rough’ surface, which increases the chances of finding a local minimum of distance to the data.

By a procedure based on singular value decomposition, detailed in Hara (2017) Section 4.3, we obtain an approximate number of dimensions explored by $\mathcal{M}_{e,P}$ as a function of e. We here provide a brief description of our methodology. The measured velocity can be expressed as a linear combination of the velocity components in the orbital frame $\dot{X}$, $\dot{Y}$. With these variables, equation (4) becomes

$$y(t, A, B, \omega, e, P) = A\dot{X}(t, \omega, e, P) + B\dot{Y}(t, \omega, e, P). \tag{B4}$$

The components $\dot{\boldsymbol{X}}(P, e, \omega) = (\dot{X}(t_k, P, e, \omega))_{k=1..N}$, $\dot{\boldsymbol{Y}}(P, e, \omega) = (\dot{Y}(t_k, P, e, \omega))_{k=1..N}$ are computed for a grid of ω, $(\omega_k)_{k=1..n}$. Those vectors are concatenated to form a matrix $M(e, P) = [\dot{\boldsymbol{X}}(P, e, \omega_1)..\dot{\boldsymbol{X}}(P, e, \omega_n)\dot{\boldsymbol{Y}}(P, e, \omega_1)..\dot{\boldsymbol{Y}}(P, e, \omega_n)]$, whose columns are normalized to obtain $\tilde{M}(e, P)$. We compute the number of singular values of $\tilde{M}(e, P)$ that are greater than a tenth of the maximum singular value. This constitutes a proxy for the number of dimensions explored by models with e, P fixed.

As an example, we perform this calculation on the 214 measurement times of GJ 876 (Correia et al. 2010). The number of explored dimensions are shown in Table 2. As eccentricity increases, the models explore more dimensions. Since the models at high eccentricity occupy a very large volume in many dimensions, there will often be at least one high eccentricity model closely fitting the data.

APPENDIX C: FREQUENTIST METHODOLOGIES

C1 Presentation

Testing possible eccentricities can also be done in a frequentist framework. This one offers confidence intervals, which are not as easy to interpret as Bayesian credible intervals but have the advantage of being quicker to compute. Furthermore, the associated algorithms have clearer convergence tests.

So far frequentist inferences for eccentricities have been done in several ways. Lucy & Sweeney (1971) and Husnoo et al. (2012) respectively used p-values and BIC to test the hypothesis that eccentricity is non-zero. More precisely, Lucy & Sweeney (1971) compute the probability distribution of the eccentricity estimate under the hypothesis that the eccentricity is null and find a Rayleigh distribution whose variance depends on the SNR (which we also obtain as a special case of our analysis Section 2.2.2 with $p = 0$). For a given measured eccentricity $\widehat{e}$, they measure the probability that the Rayleigh distribution is higher than $\widehat{e}$ and report an eccentric orbit if this probability is lower than a certain threshold.

Husnoo et al. (2012) computes

$$\mathrm{BIC}(\mathcal{M}) = \chi^2_{\min}(\mathcal{M}) + p \ln N + \ln(2\pi|\mathbf{V}|), \tag{C1}$$

where $\chi^2_{\min}$ is the minimum χ^2 obtained when minimizing the distance between the data and model $\mathcal{M}$, p is the number of degrees of freedom of $\mathcal{M}$ (three for a sine model and five for a Keplerian one), and $|\mathbf{V}|$ is the determinant of the correlation matrix. The orbit is said to be eccentric if $\mathrm{BIC}(\mathcal{M}_{\mathrm{ecc}}) \geqslant \mathrm{BIC}(\mathcal{M}_{\mathrm{circ}})$, where $\mathcal{M}_{\mathrm{ecc}}$ and $\mathcal{M}_{\mathrm{circ}}$ are respectively eccentric and circular models.

Though reasonable, these techniques can be improved. First, they both consider the alternative e is zero or non-zero, and do not allow to test if a given value of eccentricity is compatible with the data or not. Secondly, the analytical approximation of the eccentricity distribution is not always accurate. Finally, the BIC (C1) gives equal weight to all parameters, only their number k appears. This approximation is valid in the limit of a large number of observations.

C2 New methods

Our aim is to overcome as much as possible these limitations. It turns out that the procedure to construct confidence intervals outlined in Casella & Berger (2001), chapter 9, allows us to test the hypothesis that the true eccentricity is equal to a certain value e for all e. The idea is to reject the hypothesis that eccentricity is equal to e if all models with eccentricity e have a likelihood lower than a fraction of the maximum likelihood. The following criterion is computed in Appendix C4. We reject the hypothesis that the eccentricity has a certain value e with a confidence level α if

$$\mathrm{LR} := \frac{\max\limits_{\boldsymbol{\theta}\in\Theta_e} f(\boldsymbol{y}|\boldsymbol{\theta})}{\max\limits_{\boldsymbol{\theta}\in\Theta} f(\boldsymbol{y}|\boldsymbol{\theta})} \leqslant \mathrm{e}^{-\frac{1}{2}\beta} \tag{C2}$$

$$\beta = F^{-1}_{\chi^2_\rho}(1-\alpha) \tag{C3}$$

$$\rho = 2 + 2S'^2\frac{e^2}{1+e^2} - \frac{\pi e}{1+e^2}L_{\frac{1}{2}}\left(-\frac{S'^2}{2}\right)L_{\frac{1}{2}}\left(-\frac{e^2S'^2}{2}\right), \tag{C4}$$

where Θ_e is the set of parameters that have all eccentricity e, $f(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood, $F^{-1}_{\chi^2_\rho}$ is the inverse cumulative distribution function of a χ^2 law with ρ degrees of freedom, $S' = (\sigma/K_\mathrm{t})\sqrt{2/N}$ and $L_{\frac{1}{2}}$ is the Laguerre polynomial of order 1/2. The quantity (C2) is simply the ratio of the maximum likelihood obtained by restriction to the models with fixed eccentricity divided by the maximum likelihood on all models. The condition states that if all models that have eccentricity e have too low a likelihood, then e is rejected. The following equations give the value of that threshold, which is obtained by calculating the law followed by the random variable LR under the hypothesis that the true eccentricity is e ($LR|(e_\mathrm{t} = e)$). It is in fact easier to compute the law followed by the logarithm of LR, to obtain a χ^2 law whose degree depends on a definition of the SNR S and on the eccentricity under study, but is always smaller than 2. Our computations, detailed in Appendix C4, also make use of simplifying assumptions, but these are checked to give satisfactory results on simulated signals.

One of the problems of that expression is that it depends on the true value of the semi-amplitude, K_t, which is unknown. There are two ways to circumvent this issue: either by assuming that $\rho = 2$ for all e, which is the maximum value ρ can take, of K_t, or by approximating K_t by the semi-amplitude of a circular orbit fitted at the period of the signal. The first option can be used to obtain conservative intervals to ensure that e is non-zero. The second one gives a more realistic criterion to reject an eccentricity if no extra care is needed. Let us note that $\rho = 2$ is obtained for $e = 0$. This has a simple interpretation: the model can be approximated by a linear one in $k = e\cos\omega$ and $h = e\sin\omega$. When $e = 0$, both k and h are set to zero, which blocks two degrees of freedom. Denoting by $\boldsymbol{y}_e$ the model with fixed eccentricity e that has maximum likelihood and $\boldsymbol{y}^\star$ the model with maximum likelihood, all parameters free,

$$0.5\ln(LR) = \|\mathbf{W}(\boldsymbol{y}-\boldsymbol{y}_e)\|^2 - \|\mathbf{W}(\boldsymbol{y}-\boldsymbol{y}^\star)\|^2 \tag{C5}$$

behaves then as a χ^2 law with two degrees of freedom.

To offer a point of comparison of the confidence interval computed equation (C4), two other metrics are tested. First, we simply consider

$$F = \frac{N-p}{\rho}\frac{\|\mathbf{W}(\boldsymbol{y}-\boldsymbol{y}_e)\|^2 - \|\mathbf{W}(\boldsymbol{y}-\boldsymbol{y}^\star)\|^2}{\|\mathbf{W}(\boldsymbol{y}-\boldsymbol{y}^\star)\|^2} \tag{C6}$$

which is basically equation (C5), normalized by $\|\mathbf{W}(\boldsymbol{y}-\boldsymbol{y}^\star)\|^2$ so that it depends less on the noise level assumption. The quantity (C6) is assumed to follow a F distribution with ρ and $N - p$ degrees of freedom.

Secondly, we generalize the test suggested by Lucy & Sweeney. Let us denote by $e^\star$ the estimate of eccentricity obtained by maximum likelihood when all parameters are free. For eccentricity e, we fit a Keplerian model that has an eccentricity fixed at e. We then compute the probability

$$\Pr\{|\widehat{e}-e| > |e^\star - e| \,|\, e, \mathbf{V}, \widehat{e} \sim \mathrm{Rice}(e,\eta^2)\} \tag{C7}$$

that is the probability that an eccentricity estimated by maximum likelihood $\widehat{e}$ is at least as far from its assumed value e than the distance between e and the best fit actually observed, assuming the noise model is Gaussian with known covariance matrix $\mathbf{V}$. We also assume that e follows a Rice distribution as in appendix A. A Rice distribution can be seen as the modulus of a vector with two independent Gaussian variables that have the same variance, $X \sim G(a, \eta^2)$ and $Y \sim G(b, \eta^2)$ where k and h are the means of these variables. To specify the distribution, we need therefore two scalars: the variance of both random variables η^2 and the modulus of the mean of these two variables, $r = \sqrt{a^2+b^2}$. Here $a = k$ and $b = h$, so $r = e$. Then η^2 is the variance of the estimates of k or h, which under the hypotheses of Section 2.2.2 have the same variance $\eta^2 = (\sigma^2_\mathrm{RV}/K^2_e)(\pi/(N-p))$. Then the quantity (C7) can easily be evaluated by the cumulative distribution function of the Rice distribution, which is a Marcum Q-function.

Computing (C5), (C6), or (C7) necessitates to compute the minimum distance between the observations and a model with fixed eccentricity. To do so, we exploit the fact that Keplerian models are partly linear, $\boldsymbol{y}(\boldsymbol{t},\boldsymbol{\theta}) = A\dot{\boldsymbol{X}}(P, e, \omega) + B\dot{\boldsymbol{Y}}(P, e, \omega) + C$, where $\dot{\boldsymbol{X}}$ and $\dot{\boldsymbol{Y}}$ are the components of the velocity on the orbital plane. For each couple e, ω, we can minimize $\|\boldsymbol{y}-\boldsymbol{y}(\boldsymbol{t},\boldsymbol{\theta})\|$ on A, B, C, and P, which are respectively three linear parameters and one non-linear parameter. Such a problem is fast to solve with, for instance, a Levenberg–Marquardt algorithm (Levenberg 1944; Marquardt 1963). If the period is already known (which is supposed here), obtaining an array of χ^2 on a fine grid of e and ω (60 values each) takes only up to one minute. Let us finally note that the idea of restricting the global χ^2 minimization to a grid of non-linear parameters is not new (Hartkopf, McAlister & Franz 1989; Lucy 2014). There are even further resemblances of our interval calculation with Lucy (2014), where confidence intervals on orbital parameters are computed in a similar way. However, Lucy (2014) uses a degree of freedom $\rho = 1$ for all parameters. This is correct only if the model is linear in all the parameters or approximately linear in the vicinity of the best fit and unimodal.

C3 tests

The formula (C3) and (C4) have been derived with simplifying assumptions. To test and compare them to other options, we proceed as follows. We define the acceptable interval as the set of e where $\mathrm{LR}_e \leqslant \exp(-0.5F_{\chi^2_\rho}(1-\alpha))$.

(i) We generate a population of exoplanets according to a certain prior density on the orbital elements $p(K, e, P, M_0, \omega)$. The measurement times are taken from existing data sets. The noise generated according to a Gaussian density of covariance matrix $\mathbf{V}$.

(ii) For each system, we compute the set of eccentricity that are not rejected, we check that the true eccentricity belongs to this set and compute the measure of its complement in [0, 1], that is the measure of the set of rejected eccentricities.

(iii) The results are summarized in two plots. First, the fraction of cases where the true eccentricity is not in the acceptable interval as a function of α. Secondly, the curve drawn when α goes from 0 to 1 by a point whose ordinate is the measure of the complement of the set of acceptable eccentricity and whose abscissa is the fraction of cases where the true eccentricity is not in the acceptable interval.

Such tests were carried out with the following inputs: the measurement times are those of CoRoT-9 (Bonomo et al. 2017b). The angles ω and M_0 are chosen uniformly, e follows an uniform distribution. In Fig. C1(a) and (b) we plot the result of the experiment for a period fixed at 95 d the semi-amplitude is fixed to $K = 3.5\sigma$ where σ is the RMS of the errors. These are the parameters of the detected Jupiter in the system. In Fig. C1(c) and (d), we let the period vary uniformly in $\log P$ and compute the same quantities.

Plots (a) and (c) of Fig. C1 are labelled 'ROC-like' curve as a reference to receiver–operator characteristic. These ones are defined when the data are used to decide between two hypothesis. The ROC curve represents the fraction of true positives as a function of the fraction of false positives for a given decision rule. We adapt this notion to our case, where there is an infinity of hypotheses (each e in [0, 1] is a hypothesis). For a given rate of true eccentricity that is not in the acceptable interval (false negatives), the y-axis gives the precision on the estimate. The more eccentricities are rejected, the more precise the estimate. The closer such a curve is to the upper left corner the better: regardless of the value of α, the fraction of true e rejected is zero (no false negatives) and almost all other eccentricities are rejected: the estimation is very precise.

Interestingly enough, the ROC curve (left) is very similar for all the metric considered with a slight advantage for the F-ratio and the likelihood ratio tests (formulae C5 and C6), which have a better precision (more eccentricities rejected) when the fraction of true e rejected is low. We now need to set the level of true e rejected. As expected, the curve obtained for $\rho = 2$ gives an overestimated error rate for a given α. For the three other tests, the correspondence seems appropriate. Overall, the F-ratio and the likelihood ratio tests seem to perform best.

One advantage of the frequentist method is that it relies only on local minimization algorithms, therefore it is fast and convergence is ensured. We have also shown that the parameter α (see equation C3) allows us to directly control the confidence intervals meaning. Let us mention that we observed some peculiar behaviour of the estimates for some periods where the matrix of the linearized model is ill-conditioned, that we wish to investigate into more depth in future studies. In those cases, the hypotheses allowing to compute formula (C4) are not verified and Bayesian analysis or more sophisticated formula would be required. On the other hand, it seems unlikely that someone would want to prove a non-zero eccentricity of a planet particularly poorly sampled.

C4 Confidence interval calculation

In this section, we outline the calculation of the confidence intervals for eccentricity. Such an interval is constructed as a set of eccentricities that are not rejected by a hypothesis test. We choose the likelihood ratio test:

$$e \text{ is rejected if} \quad R := \frac{\max_{\boldsymbol{\theta}\in\Theta_e} f(\boldsymbol{y}|\boldsymbol{\theta})}{\max_{\boldsymbol{\theta}\in\Theta} f(\boldsymbol{y}|\boldsymbol{\theta})} \leqslant \beta,$$

where $\boldsymbol{y}$ denotes the actual observations, $f(\boldsymbol{y}|\boldsymbol{\theta})$ denotes the likelihood, Θ_e is the set of parameters that have eccentricity e, and β is a constant which will be made explicit later. Our aim is to compute the distribution of R under the assumption that the random variable giving the observations is $Y = \boldsymbol{y}_\mathrm{t} + \epsilon$, ϵ being a Gaussian noise. We further assume the noise is independent and identically distributed, the condition translates to

$$e \text{ is rejected if} \quad D := \|\boldsymbol{y} - \boldsymbol{y}(\boldsymbol{\theta}_e)\|^2 - \|\boldsymbol{y} - \boldsymbol{y}^\star\|^2 \geqslant -2\sigma^2 \ln\beta, \tag{C8}$$

where $\boldsymbol{\theta}_e = \arg\min_{\boldsymbol{\theta}\in\Theta_e} \|\boldsymbol{y} - \boldsymbol{y}(\boldsymbol{\theta})\|^2$, σ^2 is the variance of the observations, and $\boldsymbol{y}^\star$ is the global minimum. We now compute the law followed by D, so that we can select a β that corresponds to a false alarm probability. Since D is defined implicitly, the calculation of its distribution is difficult. We make two simplifying assumptions that allow us to obtain an analytical expression. The expression will then be tested on real cases through numerical simulations.

Let us first consider the linear approximation $\boldsymbol{y} = \mathbf{M}\boldsymbol{x}_\mathrm{t} + \boldsymbol{\epsilon}$ where $\mathbf{M}$ is defined as in (A3) and (A4). We further suppose that the columns of $\mathbf{M}$ are orthonormal. Since the columns are originally of the form $\cos \boldsymbol{n}\boldsymbol{t}$, $\sin \boldsymbol{n}\boldsymbol{t}$, $\cos 2\boldsymbol{n}\boldsymbol{t}$, $\sin 2\boldsymbol{n}\boldsymbol{t}$, they must be multiplied by $\sqrt{2/N}$ and the amplitude of the signal is no K_t but $K_\mathrm{t}\sqrt{N/2}$. We look for the solution $\hat{\theta}_e$ defined as

$$\hat{\theta}_e = \arg\min_{x\in\mathbb{R}^p} \|\boldsymbol{y} - \mathbf{M}x\| \quad \text{subject to} \quad \sqrt{\frac{x_3^2 + x_4^2}{x_1^2 + x_2^2}} = e. \tag{C9}$$

Thanks to the Lagrange multipliers theorem, $\hat{\theta}_e$ satisfies the conditions

$$\frac{\partial L}{\partial x} = 0, \quad \frac{\partial L}{\partial \lambda} = 0, \quad \text{where} \tag{C10}$$

$$L(x, \lambda) = \frac{1}{2}\|\boldsymbol{y} - \mathbf{M}x\|^2 + \frac{\lambda}{2}x^T\mathbf{E}x \tag{C11}$$

with

$$\mathbf{E} = \mathrm{diag}\left(-e^2, -e^2, 1, 1, 0, \ldots, 0\right). \tag{C12}$$

The condition $\partial L/\partial x = 0$ leads to

$$\left(\mathbf{M}^T\mathbf{M} + \lambda\mathbf{E}\right)x = \mathbf{M}^T\boldsymbol{y}. \tag{C13}$$

Since the columns of $\mathbf{M}$ are orthonormal, $\mathbf{M}^T\mathbf{M}$ is the identity, thus

$$x_1 = \frac{u_1}{1 - \lambda e^2}, \quad x_2 = \frac{u_2}{1 - \lambda e^2}, \quad x_3 = \frac{u_3}{1 + \lambda},$$

$$x_4 = \frac{u_4}{1 + \lambda}, \tag{C14}$$

and $x_j = u_j, \forall j \geq 5$, where we have defined $u_i = {M_i}^T\boldsymbol{y}$, M_i being the i-th column of $\mathbf{M}$. The first four components of x are also constrained by $\partial L/\partial\lambda = 0$. Let $U = u_1^2 + u_2^2$ and $V = u_3^2 + u_4^2$. We get

$$\frac{-e^2}{(1 - \lambda e^2)^2}U + \frac{1}{(1 + \lambda)^2}V = 0, \tag{C15}$$

or, equivalently,

$$e^2(e^2V - U)\lambda^2 - 2e^2(V + U)\lambda + V - e^2U = 0, \tag{C16}$$

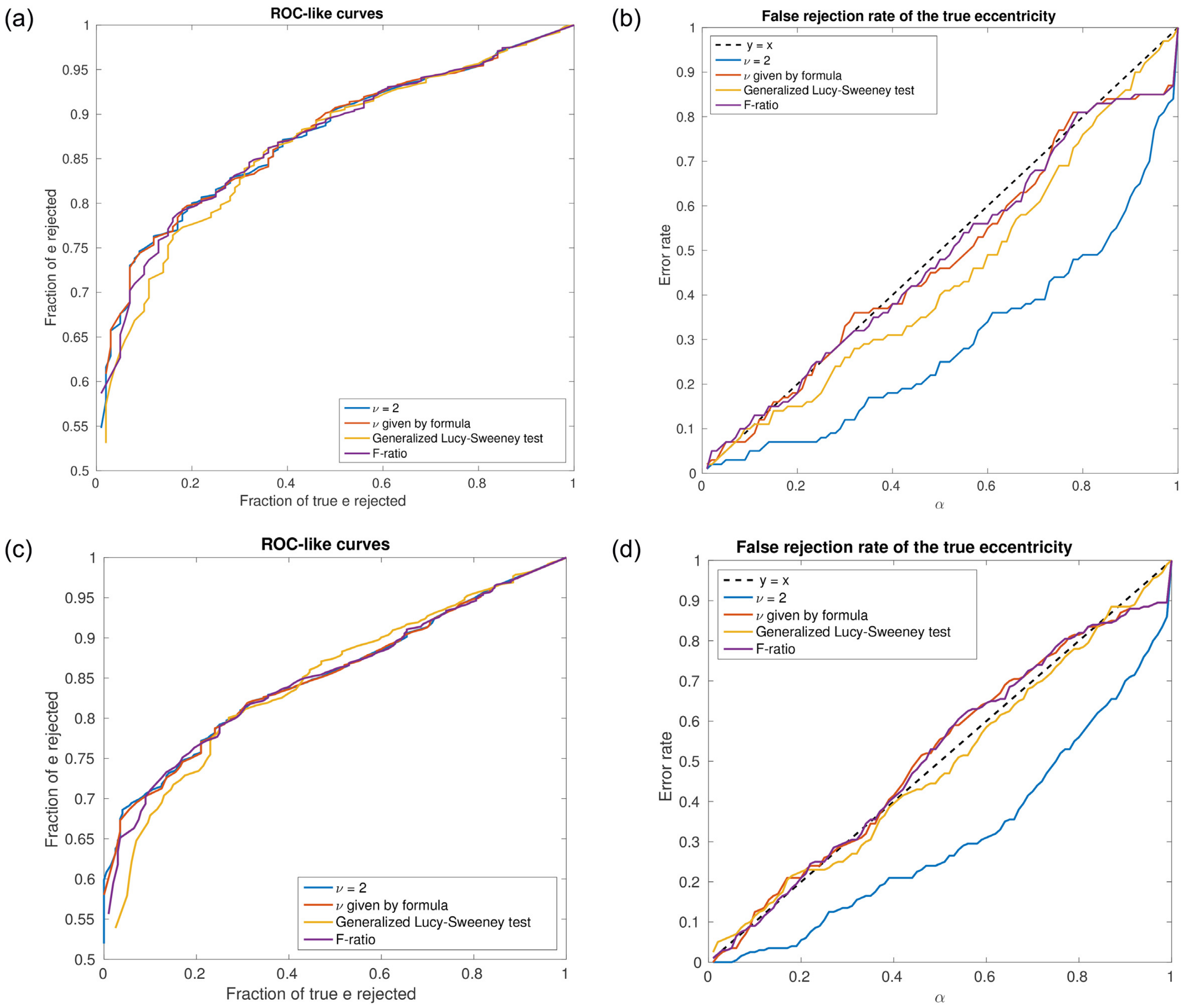

Figure C1. Results of simulations described by steps (i)–(iii) of Section C3. Here $K = 3.5\,\sigma$, ω, and M_0 are chosen uniformly in $[0, 2\pi]$, e is chosen uniformly in [0, 0.99]. For figures (a) and (b), the period is fixed to 95 d, while it is chosen uniformly in $\log P$ for plots (c) and (d). Figures (a) and (c) represent the fraction of eccentricities rejected (or equivalently, the measure of the set of rejected eccentricities) as a function of the rate of rejection of true eccentricity. Figures (b) and (c) represent the rate of rejection as a function of α. The blue, red, yellow, and purple curves are respectively obtained with rejection criteria given by equations (C2)–(C3) with $\rho = 2$, equations (C2)–(C3) with ρ given by equations (C4), (C5), and (C6) with ρ given by equation (C4). In the figure legends, ν refers to ρ.

whose solutions are

$$\lambda_\pm = \frac{e^2(U+V) \pm e(1+e^2)\sqrt{UV}}{e^2(e^2V-U)}. \qquad \text{(C17)}$$

For the solution $\hat{\theta}_e$ to actually be a minimum of L, all its eigenvalues must be positive, i.e., λ must verify $-1 < \lambda < 1/e^2$. Only λ_- fulfils this criterion, thus

$$\lambda = \frac{e^2(U+V) - e(1+e^2)\sqrt{UV}}{e^2(e^2V-U)}, \qquad \text{(C18)}$$

and

$$x_1 = \frac{1+e_0^2}{1+e^2}u_1, \quad x_2 = \frac{1+e_0^2}{1+e^2}u_2, \quad x_3 = \frac{e^2}{e_0^2}\frac{1+e_0^2}{1+e^2}u_3,$$

$$x_4 = \frac{e^2}{e_0^2}\frac{1+e_0^2}{1+e^2}u_4, \qquad \text{(C19)}$$

with $e_0^4 = e^2V/U$. After a few calculation, we show that

$$D = \sum_{k=1}^{4}(u_k - x_k)^2 = \frac{\left(e\sqrt{u_1^2+u_2^2} - \sqrt{u_3^2+u_4^2}\right)^2}{1+e^2}. \qquad \text{(C20)}$$

Let $x = e\sqrt{u_1^2+u_2^2}/K_\mathrm{t}$ and $y = \sqrt{u_3^2+u_4^2}/K_\mathrm{t}$. These two random variables follow Rice distributions with parameters

$$\rho_x = e\sqrt{\frac{N}{2}}, \quad \sigma_x = \frac{e\sigma}{K_\mathrm{t}}, \quad \rho_y = e\sqrt{\frac{N}{2}}, \quad \sigma_y = \frac{\sigma}{K_\mathrm{t}}. \qquad \text{(C21)}$$

An expansion of the product term shows that D behaves approximately as a weighted sum of variables following a χ^2 distribution. We can then use the Welch–Satterthwaite approximation (Satterthwaite 1946; Welch 1947): D approximately follows a χ^2 distribution whose number of degrees of freedom ρ is given by $\mathbb{E}\{D\}$. In the following, we denote by $S' = \frac{K_\mathrm{t}}{\sigma}\sqrt{\frac{N}{2}}$ the

SNR. The expected value of D is

$$\begin{aligned}\mathbb{E}\{D\} &= \frac{K_{\mathrm{t}}^2}{1+e^2}\int_0^\infty\int_0^\infty (x-y)^2 f(x|\rho_x,\sigma_x) f(y|\rho_y,\sigma_y)\,\mathrm{d}x\mathrm{d}y,\\ &= \frac{K_{\mathrm{t}}^2}{1+e^2}\Bigg[2\sigma_x^2+\rho_x^2+2\sigma_y^2+\rho_y^2-\pi\sigma_x\sigma_y L_{\frac{1}{2}}\\ &\quad\times\left(-\frac{\rho_x^2}{2\sigma_x^2}\right)L_{\frac{1}{2}}\left(-\frac{\rho_y^2}{2\sigma_y^2}\right)\Bigg]\\ &= \frac{K_{\mathrm{t}}^2}{1+e^2}\Bigg[2\frac{\sigma^2}{K_{\mathrm{t}}^2}(1+e^2)+Ne^2-e\pi\frac{\sigma^2}{K_{\mathrm{t}}^2}L_{\frac{1}{2}}\\ &\quad\times\left(-\frac{S'^2}{2}\right)L_{\frac{1}{2}}\left(-\frac{e^2S'^2}{2}\right)\Bigg].\end{aligned} \tag{C22}$$

With $\rho = \mathbb{E}\{D\}/\sigma^2$, we get

$$\rho = 2 + 2S'^2\frac{e^2}{1+e^2} - \frac{\pi e}{1+e^2}L_{\frac{1}{2}}\left(-\frac{S'^2}{2}\right)L_{\frac{1}{2}}\left(-\frac{e^2S'^2}{2}\right). \tag{C23}$$

To obtain a confidence level α, then we need to take $-2\ln\beta = F^{-1}_{\chi^2_\rho}(1-\alpha)$ where $F^{-1}_{\chi^2_\rho}$ is the inverse cumulative distribution function of a χ^2 distribution with ρ degrees of freedom. Conversely, it is possible to convert a measured D to a probability simply by computing $\alpha_e = 1 - F_{\chi^2_\rho}(D)$. The hypothesis $e_{\mathrm{t}} = e$ is rejected if α_e is below a certain threshold.

This formula was tested numerically. It is in very good agreements with the simulations as soon as S' is above ≈20. As it decreases, the average of estimated eccentricity increases (which is exactly saying that the bias increases) therefore the approximation of low eccentricities does not hold any more. The value of S' can be evaluated keeping in mind that when the linearized model at $e = 0$ is poorly conditioned, (matrix $\mathbf{M}$, as defined equations A3 and A4), then the uncertainty on k and h is higher than given by the simple formula (A11) and the S' analytical approximation is inoperative.

APPENDIX D: NON-GAUSSIAN NOISE

D1 simulations

In this appendix, we show that the non-Gaussianity of the noise has a small impact on the quality of the eccentricity estimates.

We generate 1000 realizations of six different types of noises. These have a null mean, are independent, identically distributed and scaled to have a standard deviation $\sigma = 1$ (their covariance matrix is the identity). We consider noises that are Gaussian, Student T with 3 and 4 degrees of freedom, uniform, exponential, and Poisson. We inject a circular planet on the measurement times of Gl 96 with $K = 4\times\sigma$ and uniform ω, M_0. Ten periods are chosen randomly according to a log normal distribution on the 5–500 d interval.

For each type of noise, each of the 10 × 1000 noise realizations chosen, we compute the error on eccentricity $|\widehat{e} - e_{\mathrm{t}}|$, e_{t} being the true eccentricity and $\widehat{e}$ the estimate), and the root mean square (RMS) of the residuals. The RMS is as a proxy for the estimated noise level, and thus the width of the error bars. The latter is the maximum likelihood estimate of the noise level for an i.i.d. noise model.

The average values of those on the 10 times 1000 realizations are reported in Table D1. Note that we take the square root of the MSE. We find that the mean error on eccentricity for non-Gaussian noise is within 2 per cent of the value of the mean error for a Gaussian noise, and the estimated jitter is within 5 per cent of the value of the mean jitter for a Gaussian noise. Only the MSE varies by 10 per cent between the Gaussian and Student T distribution with 3 degrees of freedom. Such results are remarkable, since for instance the Poisson noise takes discrete values and is non-symmetrical.

The cumulative distribution function of the error and the estimated noise levels (the standard deviation of the residuals) are shown in Fig. D1. Note that the jitter estimates have a slightly greater dispersion for the Student distributions, but as in the case of the average values we do not see striking differences. We simply note that for noise distributions with heavy tails (here Student), as expected, there is a higher fraction of cases where the noise level is severely underestimated or overestimated. We conclude that non-Gaussianity does not play a significant role, except for an increased variability of the noise level estimation when the noise distribution has heavy tails.

D2 Distribution of the bias and the jitter

In this section, we state and prove some mathematical results on the estimate of eccentricity obtained via a linear model. We assume as in Section 2.2.2 that the eccentricity estimate is given by the linear model, $\boldsymbol{y} = \mathbf{M}\boldsymbol{x} + \boldsymbol{\epsilon}$ where the first four columns of $\mathbf{M}$ are such that the i-th line is evaluated at observation time t_i, $\mathbf{M}_i = (\cos(nt_i)\ \sin(nt_i)\ \cos(2nt_i)\ \sin(2nt_i))$ and $\boldsymbol{x} = (A\ B\ C\ D)^T$. The remaining columns $\mathbf{M}$ are such that $\mathbf{M}$ is of maximal rank. The eccentricity estimate is

$$\widehat{e} = \sqrt{\frac{\widehat{C}^2+\widehat{D}^2}{\widehat{A}^2+\widehat{B}^2}}. \tag{D1}$$

We denote by $\widehat{K} = \sqrt{\widehat{A}^2+\widehat{B}^2}$. The only assumption on the noise $\boldsymbol{\epsilon}$ is that a vanishing mean and a non-degenerate covariance matrix $\mathbf{V}$.

The model just described is supposed to be the true model, which is unknown by the data analyst. We assume that the model with which the analysis is done is $\boldsymbol{y} = \mathbf{M}\boldsymbol{x} + \boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon}$ is a Gaussian noise identically distributed of variance σ^2, that is a free parameter. In that model, for a data set $\boldsymbol{y}_0$ the maximum likelihood estimates of $\boldsymbol{x}$ and σ are respectively

$$\widehat{\boldsymbol{x}} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\boldsymbol{y}_0 \tag{D2}$$

$$\widehat{\sigma^2} = \frac{\|\boldsymbol{y}_0 - \mathbf{M}\widehat{\boldsymbol{x}}\|^2}{N} = \frac{\|(\mathbf{I}-\mathbf{M}(\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T)\boldsymbol{y}_0\|^2}{N} = \frac{\|\mathbf{Q}\boldsymbol{y}_0\|^2}{N}, \tag{D3}$$

where we denote by $\mathbf{I}$ the $N\times N$ identity matrix and by $\mathbf{Q} = \mathbf{I} - \mathbf{M}(\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T$.

The question we address is the dependence of the estimate (D1) on the noise nature. We show that several quantities relevant to our purposes are only determined by the covariance of the noise. Assertion *iv* below shows that the bias on e^2 depends only on the covariance structure of the noise to order 2 in $1/K$. More precisely,

THEOREM 1.

(i) $\mathbb{E}\{\widehat{\sigma^2}\} = \frac{1}{N}\mathrm{tr}(\mathbf{QVQ}^T)$ *where tr is the sum of the diagonal terms of a matrix (the trace).*

(ii) $\mathbb{E}\{\widehat{\boldsymbol{x}}\} = \boldsymbol{x}_{\mathrm{t}}$

(iii) $\mathrm{Cov}\{\widehat{\boldsymbol{x}}\} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\mathbf{VM}(\mathbf{M}^T\mathbf{M})^{-1}$

(iv) $\mathbb{E}\{\widehat{e}^2\} - e_{\mathrm{t}}^2 = f(\mathbf{V}) + o\left(\frac{1}{\widehat{K}^3}\right)$.

Table D1. Error, MSE, and noise level estimate of the eccentricity estimates for Gaussian and non-Gaussian noises.

Noise type	Gaussian	Student T 3	Student T 4	Uniform	Exponential	Poisson
Mean error	0.067	0.066	0.0672	0.067	0.066	0.067
$\sqrt{\text{MSE}}$	0.084	0.094	0.090	0.085	0.083	0.084
Noise level	0.951	0.904	0.938	0.952	0.940	0.947

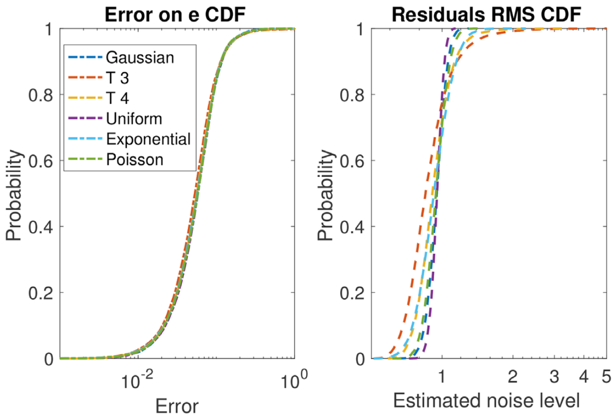

Figure D1. Left: cumulative distribution function of the error on eccentricity for six different types of noise: Gaussian, Student *T* with 3 and 4 degrees of freedom, uniform, exponential, and Poisson. Right: standard deviation of the residuals after a Keplerian fit.

Proof.

(i) As $\boldsymbol{y}_0 = \boldsymbol{y}_\mathrm{t} + \boldsymbol{\epsilon}$ and $\boldsymbol{y}_\mathrm{t}$ is in the image space of $\mathbf{M}$, we have $\mathbf{Q}\boldsymbol{y}_0 = \mathbf{Q}\boldsymbol{\epsilon}$.
The estimate of the noise level is given by equation (D3). Since $\|\mathbf{Q}\boldsymbol{y}_0\|^2 = \boldsymbol{y}_0^T\mathbf{Q}^T\mathbf{Q}\boldsymbol{y}_0$ and $\mathbf{Q}\boldsymbol{y}_0 = \mathbf{Q}\boldsymbol{\epsilon}$, $\mathbb{E}\{\widehat{\sigma^2}\} = \sum_{i,j}(QQ^T)_{ij}\mathbb{E}\{\epsilon_i\epsilon_j\}/N = \sum_{i,j}(QQ^T)_{ij}V_{ij}/N = \frac{1}{N}\mathrm{tr}(\mathbf{QVQ}^T)$.

(ii) We have $\mathbb{E}\{\widehat{\boldsymbol{x}}\} = \mathbb{E}\{(\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\boldsymbol{y}_0\} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T(\boldsymbol{y}_\mathrm{t} + \mathbb{E}\{\boldsymbol{\epsilon}\}$ and by hypothesis, $\mathbb{E}\{\epsilon\} = 0$, hence the result.

(iii) $\mathrm{Cov}\{\widehat{\boldsymbol{x}}\} = \mathrm{Cov}\{(\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\boldsymbol{y}_0\} = \mathrm{Cov}\{(\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\boldsymbol{\epsilon}\}$

(iv) We pose $\widehat{A} = \bar{A} + a$, $\widehat{B} = \bar{B} + b$, $\widehat{C} = \bar{C} + a$, $\widehat{D} = \bar{D} + d$. Denoting by $\bar{K} = \sqrt{\bar{C} + \bar{D}}$. A Taylor expansion of the denominator $\widehat{e^2}$ about $\bar{A}$, $\bar{B}$, $\bar{C}$, $\bar{D}$ at order two in $1/\bar{K}$ is

$$\widehat{e^2} = \frac{\bar{C}^2 + \bar{D}^2 + 2\bar{C}c + 2\bar{D}d + c^2 + d^2}{\bar{K}^2}$$

$$\times\left(1 - 2\frac{\bar{A}a + \bar{B}b}{\bar{K}^2} - \frac{a^2 + b^2}{\bar{K}^2} - 4\frac{(\bar{A}a + \bar{B}b)^2}{\bar{K}^4} + \ldots\right)$$

$$\mathbb{E}\{\widehat{e^2}\} - e_\mathrm{t}^2 = \sum_{i=1}^{\infty}\frac{\gamma_i}{\bar{K}^i}. \tag{D4}$$

A simple development shows that $\gamma_1 = 0$ and that $\gamma_2 = f(\boldsymbol{\mu}^2)$ where $\boldsymbol{\mu}^2$ is the vector of moments of order 2 of $\mathbf{x}$, which is a function of the noise covariance as shown in *(iii)*. For completeness, we give the explicit expression $\gamma_2 = \sigma_c^2 + \sigma_d^2 + 4e_\mathrm{t}(\cos\psi\cos\phi C_{ac} + \sin\psi\cos\phi C_{cb} + \cos\psi\sin\phi C_{ad} + \sin\psi\sin\phi C_{db}) + e_\mathrm{t}^2(\sigma_a^2 + \sigma_b^2)$, where C_{ij} is the covariance of $\widehat{I}$ and $\widehat{J}$ and $\bar{A} = \bar{K}\cos\phi$, $\bar{B} = \bar{K}\sin\phi$ $\bar{C} = \bar{e}\bar{K}\cos\phi$, $\bar{C} = \bar{e}\bar{K}\sin\phi$.

□

APPENDIX E: RESIDUAL ANALYSIS

In this section we compute the law followed by the residuals of a linear least-squares fit. Let us suppose that we have a model

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim G(0, \mathbf{V}),$$

where $\boldsymbol{y}$ is a vector of N observations, modelled as a linear combination of the column of the $N \times p$ matrix $\mathbf{A}$, and $\boldsymbol{\epsilon}$ is a Gaussian noise of covariance matrix $\mathbf{V} =: \mathbf{W}^{-1}$. Assuming $\mathbf{V}$ and $\mathbf{A}$ are known, the least square estimate of $\boldsymbol{y}$ is $\widehat{\boldsymbol{y}} = \mathbf{A}(\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}\boldsymbol{y}$ Therefore

$$\begin{aligned}\mathbf{W}^{1/2}(\boldsymbol{y} - \widehat{\boldsymbol{y}}) &= \mathbf{W}^{1/2}\left(\mathbf{A}\boldsymbol{x} + \boldsymbol{\epsilon} - \mathbf{A}(\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}(\mathbf{A}\boldsymbol{x} + \boldsymbol{\epsilon})\right)\\ &= \mathbf{W}^{1/2}(\mathbf{I}_N - \mathbf{B})\boldsymbol{\epsilon}\\ &=: \boldsymbol{r}_W,\end{aligned}$$

where $\mathbf{I}_N$ is the identity matrix of size N and $\mathbf{B} := \mathbf{A}\ (\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}$. The quantity r_W, being a product of a matrix $(\mathbf{W}^{1/2}(\mathbf{I}_n - \mathbf{B}))$ with a Gaussian random variable of covariance $\mathbf{V}$ has a covariance $\mathbf{U}$

$$\begin{aligned}\mathbf{U} &= \mathbf{W}^{1/2}(\mathbf{I}_N - \mathbf{B}^T)\mathbf{V}(\mathbf{I}_N - \mathbf{B})\mathbf{W}^{1/2}\\ &= \mathbf{W}^{1/2}(\mathbf{V} - \mathbf{BV} - \mathbf{VB}^T + \mathbf{BVB}^T)\mathbf{W}^{1/2}\end{aligned}$$

since $\mathbf{W}^{1/2}\mathbf{VB}^T = \mathbf{W}^{-1/2}\mathbf{B}^T\mathbf{W}^{1/2} = \mathbf{W}^{1/2}\mathbf{BVB}^T\mathbf{W}^{1/2}$,

$$\begin{aligned}\mathbf{U} &= \mathbf{I}_N - \mathbf{W}^{1/2}\mathbf{B}^T\mathbf{W}^{-1/2}\\ &= \mathbf{I}_N - \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T,\end{aligned}$$

where $\mathbf{C} = \mathbf{W}^{1/2}\mathbf{A}$. This notation is convenient because it shows clearly that $\mathbf{P} = \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T$ is a projection matrix on the space generated by the columns of $\mathbf{C}$. Finally

$$\mathbf{U} = \mathbf{I}_N - \mathbf{P} \tag{E1}$$

is a projection on the space orthogonal to the one generated by $\mathbf{C}$ columns. Therefore, there exists an orthonormal matrix $\mathbf{Q}$ such that

$$\mathbf{Q}^T\mathbf{U}\mathbf{Q} = \mathbf{J}_p$$

where $\mathbf{J}_p$ is a diagonal matrix whose first p elements are zero and the others are equal to one. Finally, let us remark that $\boldsymbol{r}_{QW} := \mathbf{Q}^T\boldsymbol{r}_W$ has a covariance matrix $\mathbf{Q}^T\mathbf{U}\mathbf{Q} = \mathbf{J}_p$, which shows the claim of Section 4, $\mathbf{Q}^T\mathbf{W}^{1/2}(\boldsymbol{y} - \widehat{\boldsymbol{y}})$ has p vanishing components and the others are Gaussian variables of mean 0 and variance 1.

Let us finally note that the covariance matrix $\mathbf{U} = \mathbf{I}_N - \mathbf{P}$ of $\boldsymbol{r}_W = \mathbf{W}^{1/2}(\boldsymbol{y} - \widehat{\boldsymbol{y}})$ will be close to identity if there are many more observations than parameters. This explains why the weighted residuals $\boldsymbol{r}_W$ almost behaves like independent Gaussian variables and, for instance, why plotting $\boldsymbol{r}_W(t_i) - \boldsymbol{r}_W(t_j)$ as a function of $t_i - t_j$ gives hints on the correlations.

This paper has been typeset from a TEX/LATEX file prepared by the author.