



<b>Publication Year</b>	2017
<b>Acceptance in OA @INAF</b>	2020-10-02T10:32:51Z
<b>Title</b>	Proxima Centauri reloaded: Unravelling the stellar noise in radial velocities
<b>Authors</b>	Damasso, Mario; Del Sordo, F.
<b>DOI</b>	10.1051/0004-6361/201630050
<b>Handle</b>	<a href="http://hdl.handle.net/20.500.12386/27549">http://hdl.handle.net/20.500.12386/27549</a>
<b>Journal</b>	ASTRONOMY & ASTROPHYSICS
<b>Number</b>	599

# Proxima Centauri reloaded: Unravelling the stellar noise in radial velocities

M. Damasso<sup>1</sup> and F. Del Sordo<sup>2,3</sup>

<sup>1</sup> INAF–Osservatorio Astrofisico di Torino, via Osservatorio 20, 10025 Pino Torinese, Italy  
e-mail: damasso@oato.inaf.it

<sup>2</sup> Department of Geology & Geophysics, and Department of Astronomy, Yale University, 06511 New Haven, CT, USA  
e-mail: fabio.delsordo@yale.edu

<sup>3</sup> Nordita, Royal Institute of Technology and Stockholm University, 10691 Stockholm, Sweden

Received 12 November 2016 / Accepted 10 December 2016

## ABSTRACT

**Context.** The detection and characterisation of Earth-like planets with Doppler signals of the order of  $1 \text{ m s}^{-1}$  currently represent one of the greatest challenges for extrasolar-planet hunters. As results for such findings are often controversial, it is desirable to provide independent confirmations of the discoveries. Testing different models for the suppression of non-Keplerian stellar signals usually plaguing radial velocity data is essential to ensuring findings are robust and reproducible.

**Aims.** Using an alternative treatment of the stellar noise to that discussed in the discovery paper, we re-analyse the radial velocity dataset that led to the detection of a candidate terrestrial planet orbiting the star Proxima Centauri. We aim to confirm the existence of this outstanding planet, and test the existence of a second planetary signal.

**Methods.** Our technique jointly modelled Keplerian signals and residual correlated signals in radial velocities using Gaussian processes. We analysed only radial velocity measurements without including other ancillary data in the fitting procedure. In a second step, we have compared our outputs with results coming from photometry, to provide a consistent physical interpretation. Our analysis was performed in a Bayesian framework to quantify the robustness of our findings.

**Results.** We show that the correlated noise can be successfully modelled as a Gaussian process regression, and contains a periodic term modulated on the stellar rotation period and characterised by an evolutionary timescale of the order of one year. Both findings appear to be robust when compared with results obtained from archival photometry, thus providing a reliable description of the noise properties. We confirm the existence of a coherent signal described by a Keplerian orbit equation that can be attributed to the planet Proxima b, and provide an independent estimate of the planetary parameters. Our Bayesian analysis dismisses the existence of a second planetary signal in the present dataset.

**Key words.** planetary systems – stars: individual: Proxima Centauri – techniques: radial velocities

## 1. Introduction

The discovery of a potentially terrestrial planet in an approximately 11-day orbit around the M dwarf Proxima Centauri, announced by [Anglada-Escudé et al. \(2016\)](#); hereafter AE16), represents a major breakthrough in the extrasolar-planet science. Proxima is the closest known star to the Sun, and the planet Proxima b might have an equilibrium surface temperature suitable for liquid water to exist. Intrinsic stellar activity can mask planetary signals measured in radial velocity time series, especially for M dwarf systems. This stellar activity signal can be of the order of a few  $\text{m s}^{-1}$ , larger than the semi-amplitude of the planetary signal previously measured for the Proxima system. To date, several techniques have been developed to mitigate this effect. [Dumusque et al. \(2017\)](#) devised a blind test, based on simulated RV measurements including realistic stellar activity noise ([Dumusque 2016](#)), to compare the performances of different methods. To derive orbital and physical properties of Proxima b from high-precision RV measurements it is therefore necessary a proper treatment of the stellar activity (the “noise”) contribution. To do so, AE16 used a Bayesian framework consisting of a moving average term and

linear correlations with activity indexes. The technique used by AE16 was tested on the RV datasets of [Dumusque et al. \(2017\)](#) and found to be very effective for recovering signals induced by low-mass planets. A different Bayesian technique applied to the test which performed similarly well is based on modelling the correlated noise using Gaussian processes (GPs) (see e.g. [Rasmussen & Williams 2006](#); and [Roberts et al. 2012](#), for a general description of GPs). GPs are a powerful tool to mitigate the correlated noise in a set of measurements, such as the stellar activity signature in RV data. This technique has been used successfully in several recent works dealing with real RV data ([Haywood et al. 2014](#); [Grunblatt et al. 2015](#); [Rajpaul et al. 2015](#); [Affer et al. 2016](#); [Faria et al. 2016](#); [Lopez-Morales et al. 2016](#)).

In this article we have applied the GP technique to the radial velocity observations of Proxima published in AE16, taken with the UVES and HARPS spectrographs. HARPS data are divided in a pre-2016 and 2016 (PRD: Pale Red Dot campaign) subsamples. We aim to confirm and characterise the signal of Proxima b in an independent way through an alternative and robust modelling of the stellar noise. We also consider the existence of a second planetary signal, by testing the hypothesis in a Bayesian framework.

## 2. Description of the model and technique

We model the correlated noise by adopting a quasi-periodic covariance function, similarly to what was done, for example, in [Affer et al. \(2016\)](#). This function is described by parameters, called hyperparameters, which try to model in a simple fashion some of the physical phenomena underlying the stellar noise. The quasi-periodic kernel is described by the covariance matrix

$$k(t, t') = h^2 \cdot \exp \left[ -\frac{(t-t')^2}{2\lambda^2} - \frac{\sin^2 \left( \frac{\pi(t-t')}{\theta} \right)}{2w^2} \right] + (\sigma_{\text{instr,RV}}^2(t) + \sigma_{\text{instr,jit}}^2) \cdot \delta_{t,t'}, \quad (1)$$

where  $t$  and  $t'$  indicate two different epochs. This kernel is composed by a periodic term coupled to an exponential decay one, in order to model a recurrent signal linked to stellar rotation and taking into account the size-evolution of finite-lifetime active regions. Such approach is therefore particularly suitable to model the stellar noise on short-term timescales, as we consider that modulated by the stellar rotation period. The hyperparameter  $h$  of the covariance function represents the amplitude of the correlations;  $\theta$  is related to the rotation period of the star;  $w$  is the length scale of the periodic component, linked to the size evolution of the active regions;  $\lambda$  is the correlation decay timescale, that we assume to be related to the active regions lifetime. Here,  $\sigma_{\text{instr,RV}}(t)$  is the RV internal error at time  $t$  for a given instrument;  $\sigma_{\text{instr,jit}}$  is the additional uncorrelated ‘‘jitter’’ term, one for each instrument, that we add in quadrature to the internal errors to take into account instrumental effects and other  $\lesssim 1 \text{ m s}^{-1}$  noise sources included neither in  $\sigma_{\text{instr,RV}}(t)$  nor in our stellar activity framework;  $\delta_{t,t'}$  is the Dirac delta function.

In the GP framework, the log-likelihood function to be maximised by the Markov chain Monte Carlo (MCMC) procedure is

$$\ln \mathcal{L} = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln(\det \mathbf{K}) - \frac{1}{2} \bar{r}^T \cdot \mathbf{K}^{-1} \cdot \bar{r}, \quad (2)$$

where  $n$  is the number of the data points,  $\mathbf{K}$  is the covariance matrix built from the covariance function in Eq. (1), and  $\bar{r}$  represents the RV residuals, obtained by subtracting the Keplerian signal(s) from the original RV dataset.

The general form for the models that we tested in this work is given by the equation

$$\begin{aligned} \Delta \text{RV}(t) &= \gamma_{\text{instr}} + \sum_{j=1}^{n_{\text{planet}}} \Delta \text{RV}_{\text{Kep},j}(t) + \Delta \text{RV} \dot{V}(t) \\ &+ \Delta \text{RV}(t)_{(\text{activity, short-term})} \\ &= \gamma_{\text{instr}} + \sum_{j=1}^{n_{\text{planet}}} K_j [\cos(\nu_j(t, e_j, T_{c,j}, P_j) + \omega_j) + e_j \cos(\omega_j)] \\ &+ \Delta \text{RV} \dot{V}(t) + \text{GP}, \end{aligned} \quad (3)$$

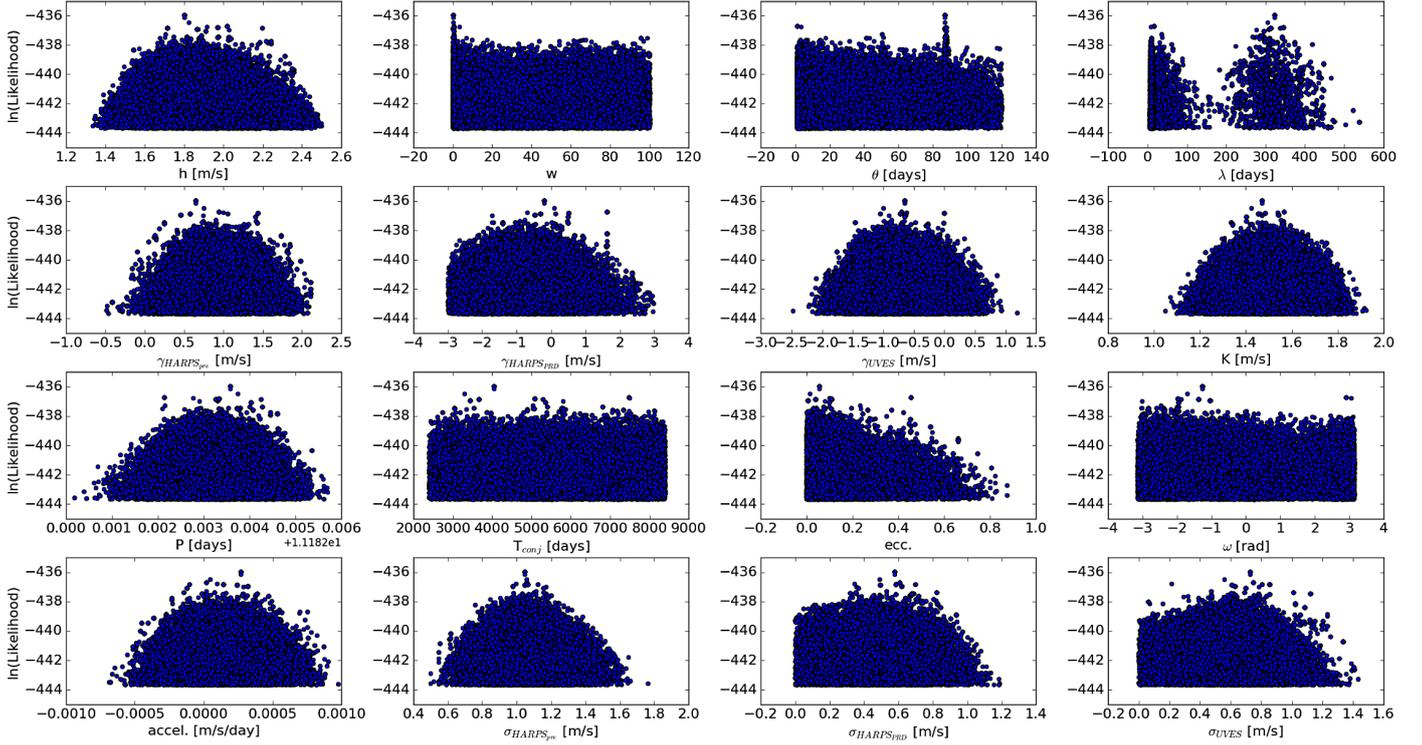
where  $n_{\text{planet}} = 1, 2$ ;  $\nu$  is function of time  $t$ , time of the inferior conjunction  $T_{c,j}$ , orbital period  $P_j$ , eccentricity  $e$  and argument of periaapsis  $\omega_j$ ;  $\gamma_{\text{instr}}$  is the RV offset, one for each instrument;  $\Delta \text{RV} \dot{V}(t)$  is the secular acceleration; GP is the stellar noise modelled with the Gaussian process. Instead of fitting separately  $e_j$  and  $\omega_j$ , we use the auxiliary parameters  $C_j = \sqrt{e_j} \cdot \cos \omega_j$  and  $S_j = \sqrt{e_j} \cdot \sin \omega_j$  to reduce the covariance between  $e_j$  and  $\omega_j$ .

Our analysis is performed in two steps through two separate MCMC runs. First, we investigate the existence of the signal attributed to Proxima b by including one Keplerian in the model together with the GP noise term. Through our GP analysis, we can constrain the values of our model hyperparameters, and thus constrain the stellar properties we assume these hyperparameters are related to. Here we examine the existence of a second planetary signal in the RV data, whose presence could not be ruled out at long orbital periods, as suggested by AE16. Finally, we evaluate the Bayesian evidences (marginal likelihoods) for the two models to assess whether the present RV dataset supports the two-planet scenario over the other. We use the publicly available emcee algorithm ([Foreman-Mackey et al. 2013](#)) to perform the MCMC analyses of the RV data, and the publicly available GEORGE Python library to perform the GP fitting within the MCMC framework ([Ambikasaran et al. 2014](#)). We used 150 random walkers for each MCMC run, and to derive the parameter posterior distributions we performed a burn-in by removing from the samples those with  $\ln \mathcal{L}$  lower than the median of the whole  $\ln \mathcal{L}$  dataset. The strategy of our MCMC analysis is to put as little constraint as possible to the model parameters. Following such a philosophy, the starting points of the walkers in the parameter space were randomly chosen from normal distributions with  $\sigma$  sufficiently large to allow a broad exploration of the parameter space around guess values. We also used uninformative priors for almost all the jump parameters. The MCMC convergence is checked using the diagnostic proposed by [Ford \(2006\)](#), and the best fit values and uncertainties for each jump parameter are calculated as the median of the marginal posterior distributions and the 16% and 84% quantiles. We quantify the relative statistical significance of the two tested models by deriving their Bayesian evidences  $\mathcal{Z}_1$  and  $\mathcal{Z}_2$ , according to Bayes’ theorem. The selection of the model better describing the data is made by calculating the ratio  $\mathcal{Z}_1/\mathcal{Z}_2$ . In fact, by assuming that there is not any a priori reason to favour one model over the other, the  $\mathcal{Z}$ -ratio represents the Bayes factor, that is, the figure of merit encoding all the support the data give to one model over the other (see, e.g. [Díaz et al. 2016](#), for a detailed discussion in the framework of the extrasolar-planets detection). We estimate the natural logarithm of each evidence  $\mathcal{Z}_i$  by using two analytic approximations due to [Chib & Jeliazkov \(2001\)](#) and [Perrakis et al. \(2014\)](#); hereafter indicated as the CJ01 and Perr14 estimators). For the statistical interpretation we follow here a conventional empirical scale, the so-called Jeffreys’ scale, which states that the model with the highest  $\mathcal{Z}$  is strongly favoured over the other when  $\Delta \ln \mathcal{Z} \geq 5$ , while  $2.5 < \Delta \ln \mathcal{Z} < 5$  denotes moderate evidence,  $1 < \Delta \ln \mathcal{Z} < 2.5$  weak evidence, and  $\Delta \ln \mathcal{Z} < 1$  corresponds to inconclusive evidence.

## 3. One-planet model

In order to minimally constrain our model, we have chosen Bayesian prior probability distributions as described in Table 1. All are uniform (uninformative) priors, except for the orbital eccentricity: for this we adopted a Beta distribution, because it is well-known that the noise tends otherwise to favour higher eccentricities (e.g. see Appendix B.4 in [Gregory 2010](#), April 2016 supplement<sup>1</sup>). We considered the GP periodic hyperparameter  $\theta$  and the orbital period  $P$  as scale invariant parameters, for which

<sup>1</sup> Available at <http://www.cambridge.org/pl/academic/subjects/statistics-probability/statistics-physical-sciences-and-engineering/bayesian-logical-data-analysis-physical-sciences-comparative-approach-mathematica-support>



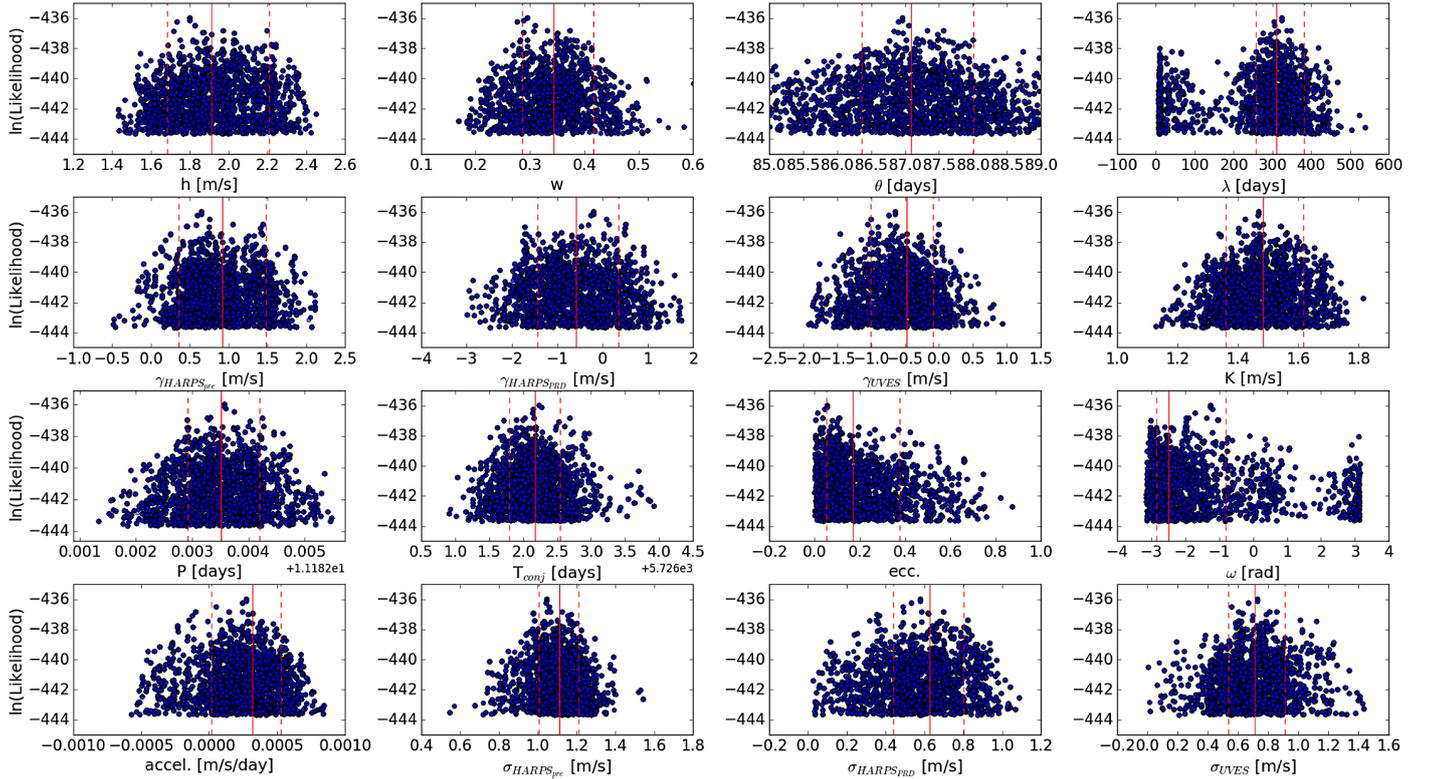
**Fig. 1.** Distributions of total MCMC values for the jump parameters of the one-planet model as function of the natural logarithm of the likelihood function.

**Table 1.** Prior probability distributions for the one-planet model parameters.

Jump parameter	Prior	Lower bound	Upper bound
$h$ [ $\text{m s}^{-1}$ ]	Uniform	0	10
$\lambda$ [days]	Uniform	0	6000
$w$	Uniform	0	100
$\ln \theta$ [days]	Uniform	0 ( $\theta = 1$ )	$\ln(120)$
$K_1$ [ $\text{m s}^{-1}$ ]	Uniform	1	10
$\ln P_1$ [days]	Uniform	0 ( $P_1 = 1$ )	$\ln 6000$
$T_{1,c}$ [JD-2 400 000]	Uniform	51 500	57 500
$C_1 = \sqrt{e_1} \cdot \cos \omega_1$	Uniform	-1	1
$S_1 = \sqrt{e_1} \cdot \sin \omega_1$	Uniform	-1	1
$e_1 (= C_1^2 + S_1^2)$	$3.1 \cdot (1 - e_1)^{2.1}$	0	1
$\omega_1$ [rad]	Uniform	0	$2\pi$
$dV_r/dt$ [ $\text{m s}^{-1} \text{ day}^{-1}$ ]	Uniform	-0.001	+0.001
$\gamma_{\text{HARPS}_{\text{pre-2016}}}$ [ $\text{m s}^{-1}$ ]	Uniform	-3	+3
$\gamma_{\text{HARPS}_{\text{PRD}}}$ [ $\text{m s}^{-1}$ ]	Uniform	-3	+3
$\gamma_{\text{UVES}}$ [ $\text{m s}^{-1}$ ]	Uniform	-3	+3
$\sigma_{\text{jit,HARPS}_{\text{pre-2016}}}$ [ $\text{m s}^{-1}$ ]	Uniform	0	10
$\sigma_{\text{jit,HARPS}_{\text{PRD}}}$ [ $\text{m s}^{-1}$ ]	Uniform	0	10
$\sigma_{\text{jit,UVES}}$ [ $\text{m s}^{-1}$ ]	Uniform	0	10

an uninformative prior is one that is uniform in  $\ln P$ . We have chosen a very large range for the GP hyperparameter  $w$ , which usually is constrained to values less than one when there is clear evidence of a stellar-rotation-linked periodic component in the correlated noise (e.g. [Lopez-Morales et al. 2016](#)). Our MCMC run stopped at  $\sim 200\,000$  steps: the median of  $\ln \mathcal{L}$  calculated over the 150 chains maintained a nearly constant value over  $\sim 140\,000$  steps. After the burn-in of the first 100 000 steps, we have performed an additional burn-in in the  $\ln \mathcal{L}$  space by deriving the marginal posterior distributions of samples with  $\ln \mathcal{L} > \text{median}(\ln \mathcal{L})$  (see Fig. 1). The posterior distribution of the GP hyperparameter  $\theta$  shows a concentrated and symmetrical

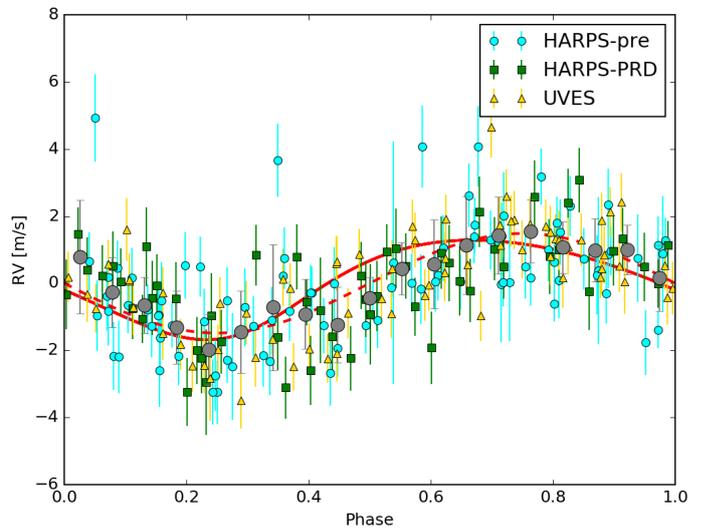
sub-sample between 85 and 89 days with the highest values of  $\ln \mathcal{L}$ , which is the range where the known rotation period of Proxima falls in. Considering that  $\theta$  is expected to represent the stellar rotation period, this evidence is particularly relevant because it reveals that the correlated noise does contain a periodic term which is modulated by the rotation of the star. The rotation period of Proxima has been robustly estimated from ASAS V-band photometry over an interval spanning up to 15 yr (e.g. see Fig. 3 in [Robertson et al. 2016](#), or Fig. 3 in [Wargelin et al. 2017](#)). In particular, [Wargelin et al. \(2017\)](#) present the Lomb-Scargle periodogram of the ASAS-3 and ASAS-4 photometry covering 15 yr, which shows a very significant and sharp peak at 83.1 days. The same authors show evidence for differential rotation in the light curve by analysing periodograms of different observing seasons, finding significant peaks ranging from  $P_{\text{rot}} = 77$  to  $P_{\text{rot}} = 90$  days. Based on this robust result from photometry, we can naturally restrict our analysis to the 697 510 posterior samples for which  $\theta$  falls in the range 85–89 days, ensuring the periodic features of our model are tied to the stellar rotation. The corresponding posterior distributions for all the model parameters are showed in Fig. 2. For the GP hyperparameter  $w$  we get a symmetric distribution with all the values less than one, which is what is expected, as described, for example, in [Lopez-Morales et al. \(2016\)](#). Moreover, the distribution of the GP hyperparameter  $\lambda$  has a median of 311 days and is nearly symmetric (excluding the points with  $\lambda < 200$  days that are relatively few compared to the whole sample, as indicated by the value of the 16th percentile). This long evolutionary timescale appears particularly interesting when compared with photometric results. [Wargelin et al. \(2017\)](#) show the evidence for the rotational phasing to remain remarkably constant over the  $\sim 15$  yr timespan at a fixed  $P_{\text{rot}} = 83.1$  days, and for the modulation amplitude not to change significantly over two years, which could be indicative of the presence of persistent active longitudes on



**Fig. 2.** Distributions as in Fig. 1 after constraining the GP hyperparameter  $\theta$  in the range [85, 89] days. The vertical red lines represent the median values (solid) and the 16th and 84th percentiles (dotted) of these distributions, which are listed in Table 2.

the stellar photosphere. For instance, the light curve corresponding to epochs between 2010 and 2012, which overlap with the timespan of the RV data, can be well described by a sinusoidal function with constant semi-amplitude and phase (see Fig. 3 in Wargelin et al. 2017), suggesting that the photospheric structures responsible for the correlations observed in the RV residuals could indeed have a decay timescale of the order of approximately one year. The best-fit estimates for all the model parameters are showed in Table 2, along with values for the planet orbital solution reported by AE16. The residuals of our global model have an rms of  $1.3 \text{ m s}^{-1}$ . In Fig. 3 we show the RV residuals, after removing the stellar signal described by the GP, folded at our best-fit estimate of the orbital period (Fig. 4 shows the same plot for each dataset taken separately). It can be seen that a circular model well describes the data, in agreement with our finding that the eccentricity is consistent with zero within  $\sim 1.4\sigma$ , whilst usually at least a  $2.5\sigma$  level is required for claiming a significant eccentricity. Figure 5 shows the RV residuals, after subtracting the Keplerian solution, with our best-fit model for the correlated stellar noise superposed. We note that the model describes the second half of the HARPS-PRD dataset particularly well.

AE16 show (extended data Fig. 2 therein) that their likelihood-ratio periodogram of the HARPS pre-2016 RV dataset has the highest peak at  $P = 215$  days ( $p$ -value = 1%), but they could not conclude about the nature of this signal. To investigate in more detail the possible cause of its existence, we analyse the same dataset with the generalised Lomb-Scargle algorithm (GLS; Zechmeister & Kürster 2009). We find that the highest power density peak appears at  $P \sim 57$  days (upper plot in Fig. 6), while at  $P = 215$  days there is no significant power ( $p$ -value  $> 1\%$ , estimated through a bootstrap with re-sampling



**Fig. 3.** Radial velocity residuals, after subtracting the best-fit GP correlated noise model, folded at the best-fit orbital period  $P = 11.1855$  days. Different symbols and colours are used for each dataset. Grey dots represent average values for 20 phase bins. Superposed are the best-fit eccentric model (continuous red curve) and the simpler circular model (dashed red curve).

analysis on 10000 fake datasets). Then, we run GLS on the RV residuals of our global one-planet plus noise model. We do not find power excess in the periodogram at low frequencies (lower plot in Fig. 6), indicating that our correlated noise model has suppressed any long-term modulation in the data. The impact of the GP model on these residuals is not trivial to be characterised, and could prevent the detection of additional

**Table 2.** Percentiles of the distributions in Fig. 2, referred to the cut for  $\theta$  in the range 85–89 days, compared with values found by AE16.

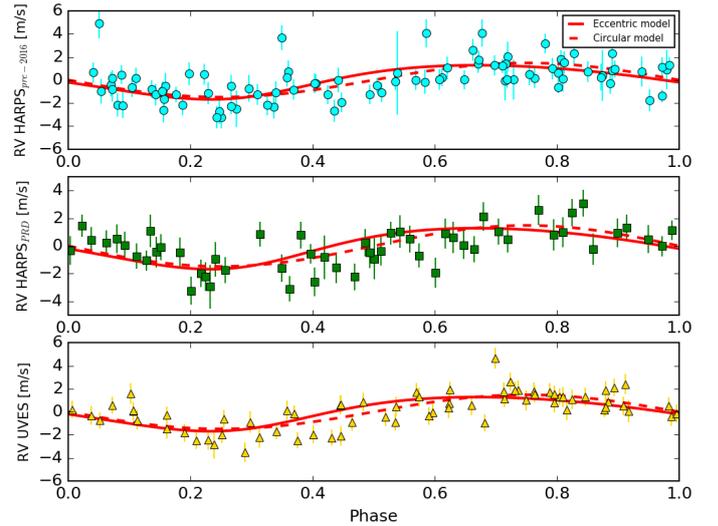
Jump parameter	Value (This work)	Value (Anglada-Escudé et al. 2016)
$h$ [ $\text{m s}^{-1}$ ]	$1.91^{+0.30}_{-0.23}$	
$\lambda$ [days]	$311^{+71}_{-54}$	
$w$	$0.34^{+0.07}_{-0.06}$	
$\theta$ [days]	$87.1^{+0.9}_{-0.7}$	
$\gamma_{\text{HARPS}_{\text{pre-2016}}}$ [ $\text{m s}^{-1}$ ]	$0.92 \pm 0.56$	
$\gamma_{\text{HARPS}_{\text{PRD}}}$ [ $\text{m s}^{-1}$ ]	$-0.58^{+0.94}_{-0.85}$	
$\gamma_{\text{UVES}}$ [ $\text{m s}^{-1}$ ]	$-0.48^{+0.39}_{-0.52}$	
$K_1$ [ $\text{m s}^{-1}$ ]	$1.48^{+0.13}_{-0.12}$	$1.38 \pm 0.21$
$P_1$ [days]	$11.1855^{+0.0007}_{-0.0006}$	$11.186^{+0.001}_{-0.002}$
$T_{c,1}$ [JD-2 400 000]	$57383.71^{+0.24}_{-0.21}$	
$e_1$	$0.17^{+0.21}_{-0.12}$	$<0.35$
$\omega_1$ [rad]	$-2.49^{+1.49}_{-0.35}$	$5.41$ (unconstrained)
$dV_r/dt$ ( $\cdot 10^{-4}$ ) [ $\text{m s}^{-1} \text{ day}^{-1}$ ]	$3.2^{+2.1}_{-3.0}$	$2.3 \pm 8.4$
$\sigma_{\text{jit,HARPS}_{\text{pre-2016}}}$ [ $\text{m s}^{-1}$ ]	$1.11 \pm 0.10$	$1.76^{+0.60}_{-0.54}$
$\sigma_{\text{jit,HARPS}_{\text{PRD}}}$ [ $\text{m s}^{-1}$ ]	$0.63^{+0.17}_{-0.19}$	$1.14^{+0.70}_{-0.57}$
$\sigma_{\text{jit,UVES}}$ [ $\text{m s}^{-1}$ ]	$0.71^{+0.20}_{-0.18}$	$1.69^{+0.64}_{-0.47}$
Minimum mass, $m_p \sin i$ ( $M_{\oplus}$ ) <sup>a</sup>	$1.21 \pm 0.16$	$1.27^{+0.19}_{-0.17}$
Orbital semi-major axis, $a$ (AU) <sup>a</sup>	$0.048 \pm 0.002$	$0.0485^{+0.0041}_{-0.0051}$

**Notes.** They represent our adopted solution for the one-planet model. Our analysis makes use of 13 parameters as well as three values  $\gamma$ s for the offsets, while AE16 employed 26 parameters. <sup>(a)</sup> Derived quantities from the posterior distributions. We assumed  $M_s = 0.120 \pm 0.015 M_{\odot}$  as the mass of Proxima, and we used the following equations (assuming  $M_s + m_p \cong M_s$ ):  $m_p \sin i \cong (K_1 \cdot M_s^{\frac{2}{3}} \cdot \sqrt{1 - e^2} \cdot P_1^{\frac{1}{3}}) / (2\pi G)^{\frac{1}{3}}$ ;  $a \cong [(M_s \cdot G)^{\frac{1}{3}} \cdot P_1^{\frac{2}{3}}] / (2\pi)^{\frac{2}{3}}$ , where  $G$  is the gravitational constant.

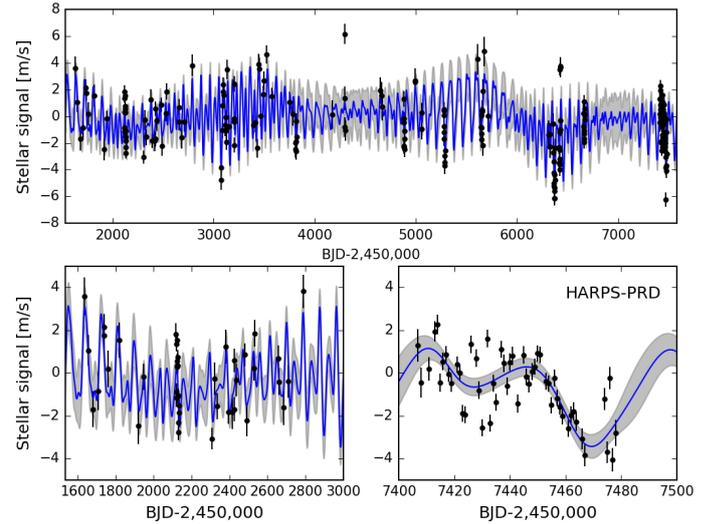
planetary signals in the periodogram. Therefore, we explore an additional two-planet plus stellar activity model to test whether this signal is more effectively modelled as a planet or stellar activity.

#### 4. Two-planet model

We tested the two-planet scenario using Gaussian priors for all the GP hyperparameters, except for  $h$ , based on the physically reliable results obtained for the stellar noise representation for the one-planet model. In modelling the RVs we ignored the mutual gravitational perturbations between the two planets. Taking further advantage of the one-planet model results to speed up the analysis, for Proxima b we fixed the upper limit for RV semi-amplitude to  $3 \text{ m s}^{-1}$ , that of the orbital period to 20 days and the eccentricity to zero. We explored the parameter space for the orbital period of the second Keplerian starting from 100 days, on the basis that no clear modulation with a period equal or less than that is observed in the HARPS-PRD RV dataset (see AE16). We ran both a model where the orbit of the outer planet is assumed circular, and one where the eccentricity is treated as a free parameter. The list of the adopted priors is showed in Table 3. The marginal posterior distributions for the simpler circular model are showed in Fig. 7. For this case, we note that the orbital period of the outer planet appears to be multi-modal, and the semi-amplitude of the Keplerian is below the average value of the internal RV uncertainties ( $\langle \sigma_{\text{RV}} \rangle = 0.94 \text{ m s}^{-1}$ ). Moreover, the rms of the residuals remains the same between the one and two-planet model ( $1.3$  and  $1.2 \text{ m s}^{-1}$ , respectively). These evidences alone do not favour the existence of a second Keplerian signal in



**Fig. 4.** Radial velocity residuals for the one-planet model, after subtracting the best-fit GP correlated noise model, folded at the best-fit orbital period  $P = 11.1855$  days. Each dataset is showed here separately for clarity. Superposed are the best-fit eccentric model (continuous red curve) and the simpler circular model (dashed red curve).

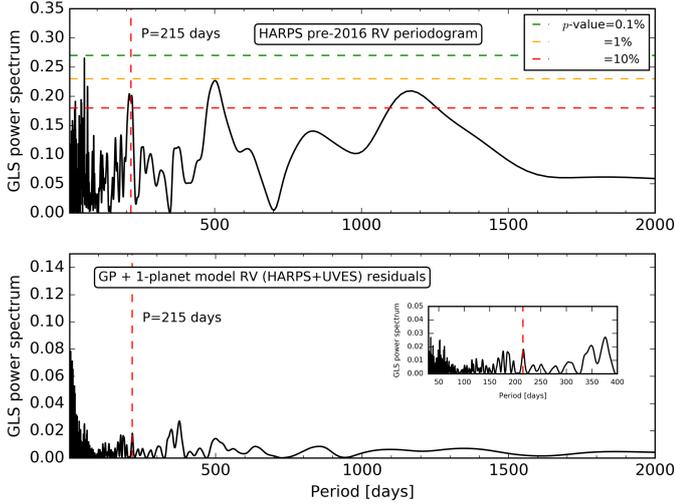


**Fig. 5.** Radial velocity residuals time series (black dots), after subtracting our best-fit orbital solution for Proxima b. The blue line with grey shaded  $1\sigma$  regions represents our best-fit GP quasi-periodic model for the correlated stellar noise. The *upper plot* shows the complete dataset, while the *two plots in the second row* show selected epochs, to easier visualize the agreement between the model and the data.

the data. A similar conclusion can be drawn for the model with the eccentricity as free parameter. Here, the eccentricity appears to be insignificant within  $2\sigma$ .

#### 5. Summary and conclusions

We used a Gaussian process framework to allay the stellar correlated noise in the radial velocity timeseries of Proxima Centauri published by Anglada-Escudé et al. (2016). In our fitting procedure we considered only radial velocity measurements, without including other ancillary data. We then compared the outputs of our procedure with those coming from photometry, to provide a consistent physical interpretation of the results. We adopted a quasi-periodic kernel to describe the correlated noise: this is



**Fig. 6.** *Upper plot:* GLS periodogram of the HARPS pre-2016 RV data. The vertical red dashed line marks  $P = 215$  days, the period found by AE16 in their likelihood-ratio periodogram, for which they could not give a physical interpretation. The  $p$ -value levels have been derived through a bootstrap analysis. *Lower plot:* GLS periodogram of the complete RV datates residuals after subtracting the GP plus one-planet model. The *inset plot* shows the same periodogram around  $P = 215$  days, for more clarity.

**Table 3.** Prior probability distributions for the two-planet model parameters.

Jump parameter	Prior	Lower bound	Upper bound	Mean	$\sigma$
$h$ [m s <sup>-1</sup> ]	Uniform	0	10		
$\lambda$ [days]	Gaussian	–	–	311	71
$w$	Gaussian	–	–	0.34	0.07
$\theta$ [days]	Gaussian	–	–	87.1	0.9
$K_1$ [m s <sup>-1</sup> ]	Uniform	1	3		
$\ln P_1$ [days]	Uniform	0 ( $P_1 = 1$ )	$\ln 20$		
$T_{c,1}$ [JD-2 400 000]	Uniform	51 500	57 500		
$K_2$ [m s <sup>-1</sup> ]	Uniform	0	100		
$\ln P_2$ [days]	Uniform	4.6 ( $P_2 = 100$ )	$\ln 6000$		
$T_{c,2}$ [JD-2 400 000]	Uniform	51 500	57 500		
$C_2 = \sqrt{e_2} \cdot \cos \omega_2$	Uniform	-1	1		
$S_2 = \sqrt{e_2} \cdot \sin \omega_2$	Uniform	-1	1		
$e_2 (=C_2^2 + S_2^2)^{1/2}$	3.1 · (1 - $e_2$ ) <sup>2.1</sup>	0	1		
$\omega_2$ [rad]	Uniform	0	$2\pi$		
$dV_r/dt$ [m s <sup>-1</sup> day <sup>-1</sup> ]	Uniform	-0.001	+0.001		
$\gamma_{\text{jit,HARPSpre-2016}}$ [m s <sup>-1</sup> ]	Uniform	-3	+3		
$\gamma_{\text{HARPSpre-2016}}$ [m s <sup>-1</sup> ]	Uniform	-3	+3		
$\gamma_{\text{HARSPRD}}$ [m s <sup>-1</sup> ]	Uniform	-3	+3		
$\gamma_{\text{UVES}}$ [m s <sup>-1</sup> ]	Uniform	-3	+3		
$\sigma_{\text{jit,HARPSpre-2016}}$ [m s <sup>-1</sup> ]	Uniform	0	5		
$\sigma_{\text{jit,HARSPRD}}$ [m s <sup>-1</sup> ]	Uniform	0	5		
$\sigma_{\text{jit,UVES}}$ [m s <sup>-1</sup> ]	Uniform	0	5		

**Notes.** The orbit of the inner planet is assumed circular.

a widely used function that represents the covariance between measurements at different epochs in terms of parameters that can be related to some physical properties of the star. The philosophy behind our approach – but not the analysis tools – is similar to that of Faria et al. (2016), which used GPs to model the stellar noise in RVs of the host-star CoRoT-7 and were able to retrieve planets CoRoT-7 b and CoRoT-7 c without analysing any other additional dataset.

In our study, we confirm the detection of a coherent signal well described by a Keplerian orbit equation that can be attributed to the planet Proxima b. We find system parameters to be in good agreement with those of Anglada-Escudé et al. (2016). We note that our estimates of the uncorrelated jitter

**Table 4.** Two-planet model parameter estimates, obtained by fixing the eccentricity of the second Keplerian to zero, and also keeping it as a free parameter.

Jump parameter	Value	
	Circular orbit	Eccentric orbit
$h$ [m s <sup>-1</sup> ]	$1.87 \pm 0.18$	$1.81^{+0.18}_{-0.07}$
$\lambda$ [days]	$317^{+35}_{-21}$	$337^{+25}_{-34}$
$w$	$0.33 \pm 0.03$	$0.33^{+0.03}_{-0.04}$
$\theta$ [days]	$87.2 \pm 0.5$	$87.1^{+0.5}_{-0.7}$
$\gamma_{\text{HARPSpre-2016}}$ [m s <sup>-1</sup> ]	$0.95^{+0.34}_{-0.41}$	$0.88^{+0.38}_{-0.36}$
$\gamma_{\text{HARSPRD}}$ [m s <sup>-1</sup> ]	$-0.65^{+0.97}_{-0.92}$	$0.35^{+0.46}_{-1.04}$
$\gamma_{\text{UVES}}$ [m s <sup>-1</sup> ]	$0.51 \pm 0.47$	$-0.70^{+0.63}_{-0.38}$
$K_1$ [m s <sup>-1</sup> ]	$1.45^{+0.10}_{-0.09}$	$1.45^{+0.09}_{-0.07}$
$P_1$ [days]	$11.1854^{+0.0006}_{-0.0005}$	$11.1854^{+0.0005}_{-0.0007}$
$T_{c,1}$ [JD-2 400 000]	$57\,383.57 \pm 0.19$	$57\,383.59 \pm 0.20$
$K_2$ [m s <sup>-1</sup> ]	$0.87 \pm 0.22$	$1.17^{+0.74}_{-0.30}$
$P_2$ [days]	$243^{+123}_{-68}$	$344^{+22}_{-99}$
$T_{c,2}$ [JD-2 400 000]	$54\,791^{+109}_{-119}$	$56\,040^{+123}_{-110}$
$e_2$	0	$0.46 \pm 0.26$
$\omega_2$ [rad]	0	$-0.15^{+1.1}_{-2.3}$
$dV_r/dt$ [m s <sup>-1</sup> day <sup>-1</sup> ]	$1.4 \pm 2.6 \times 10^{-4}$	$3^{+2.1}_{-1.7} (\times 10^{-5})$
$\sigma_{\text{jit,HARPSpre-2016}}$ [m s <sup>-1</sup> ]	$1.04 \pm 0.13$	$0.98^{+0.17}_{-0.08}$
$\sigma_{\text{jit,HARSPRD}}$ [m s <sup>-1</sup> ]	$0.45 \pm 0.17$	$0.56^{+0.12}_{-0.23}$
$\sigma_{\text{jit,UVES}}$ [m s <sup>-1</sup> ]	$0.78^{+0.14}_{-0.16}$	$0.67^{+0.18}_{-0.16}$

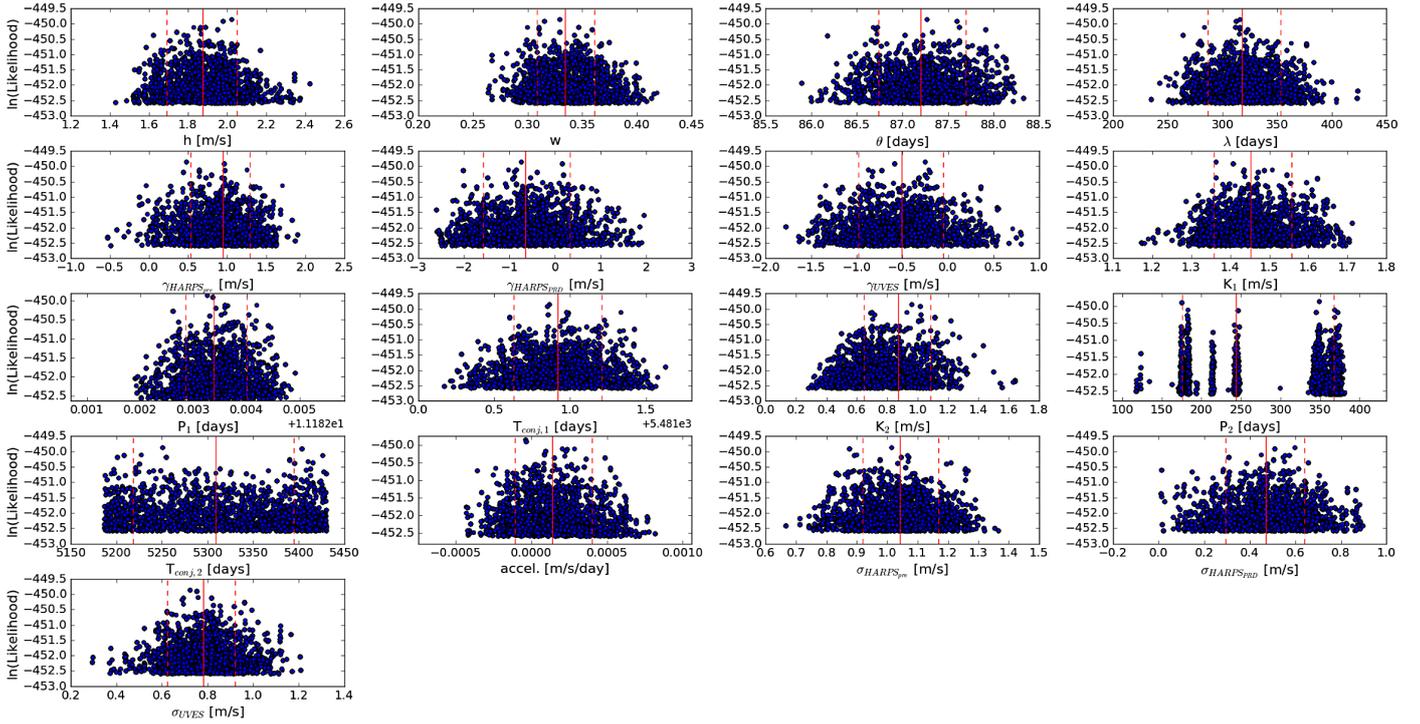
**Table 5.** Bayesian evidences for the tested models.

Model	$\ln \mathcal{Z}$	
	[CJ14]	[Perr14]
One-planet	$-444.6 \pm 0.3$	$-440.6 \pm 0.3$
Two-planet (circular)	$-455.7 \pm 0.2$	$-458.67 \pm 0.03$
Two-planet (free eccentricity)	$-454.6 \pm 0.6$	$-457.25 \pm 0.05$
$\Delta(\ln \mathcal{Z})$ , one-planet minus two-planet (circular)	+11	+18
$\Delta(\ln \mathcal{Z})$ , one-planet minus two-planet (free eccentricity)	+10	$\sim +17$

terms for each independent RV dataset are lower than in Anglada-Escudé et al. (2016, Table 2), suggesting that our model removes the stellar noise more efficiently.

As a major outcome, the correlated noise has indeed a periodic component modulated by the stellar rotation period, and we can describe it through a simple GP-based regression. The best-fit estimates of all the hyperparameters assume realistic physical values, as the stellar rotation period or the typical lifetime of active regions, because we have a measured or deduced counterpart from an independent photometric dataset. Therefore our model is physically trustworthy as well as simpler, because it makes use of fewer free parameters. This is likely the reason because we find smaller errors for the Doppler semi-amplitude and the orbital period. We note that Anglada-Escudé et al. (2016) do not mention the existence in their RV periodograms of a significant signal with a frequency close to the stellar rotation period, but in light of our results the rotation of Proxima could be considered the main cause of the clear trend visible in the high-sampling HARPS-PRD dataset.

Once constrained the stellar noise on the results found for the one-planet model, we tested the presence of a second companion by adding a Keplerian to the model. Table 5 shows the Bayesian evidences for the considered models from two



**Fig. 7.** Marginal posterior distributions for the two-planet model (circular option).

different estimators. Both strongly favour the one-planet scenario, with  $\Delta(\ln \mathcal{Z})$  between 10 and 18. Our analysis thus shows that, with our proposed treatment of the stellar noise, the present RV dataset does not show unambiguous evidence for the existence of an additional, low-mass planetary companion to Proxima on an outer orbit with an orbital period of between 100 to 6000 days.

*Acknowledgements.* This work was conceived and partially carried out during the “Approaching the Stellar Astrophysical Limits of Exoplanet Detection: Getting to  $10 \text{ cm s}^{-1}$ ” workshop held at the Aspen Center for Physics on Sept. 2016. The Aspen Center for Physics is supported by National Science Foundation grant PHY-1066293. M.D. was granted by the Simons Foundation to attend the workshop, and he acknowledges funding from Progetto Premiale INAF “Way to Other Worlds” (WoW), decreto 35/2014. F.D.S. acknowledges the Swedish Research Council International Postdoc fellowship for support. We thank A. Sozzetti, A.S. Bonomo and R. Haywood for valuable discussions, and the referee for useful comments. We finally thank M. Boldi and C. De Sica for some pleasant hours spent in Aspen, twenty years in the past.

## References

Affer, L., Micela, G., Damasso, M., et al. 2016, *A&A*, **593**, A117  
Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O’Neil, M. 2014, ArXiv e-prints [arXiv:1403.6015]

Anglada-Escudé, G., Amado, P. J., Barnes, J., et al. 2016, *Nature*, **536**, 437  
Chib, S., & Jeliazkov, I. 2001, *J. Am. Stat. Assoc.*, **96**, 270  
Dumusque, X. 2016, *A&A*, **593**, A5  
Dumusque, X., Borsa, F., Damasso, M., et al. 2017, *A&A*, **598**, A133  
Díaz, R. F., Ségransan, D., Udry, S., et al. 2016, *A&A*, **585**, A134  
Faria, J. P., Haywood, R. D., Brewer, B. J., et al. 2016, *A&A*, **588**, A31  
Ford, E. B. 2006, *ApJ*, **642**, 505  
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306  
Gregory, P. 2010, Bayesian Logical Data Analysis for the Physical Sciences (CUP)  
Grünblatt, S. K., Howard, A. W., & Haywood, R. D. 2015, *ApJ*, **808**, 127  
Haywood, R. D., Collier Cameron, A., Queloz, D., et al. 2014, *MNRAS*, **443**, 2517  
Lopez-Morales, M., Haywood, R. D., Coughlin, J. L., et al. 2016, *AJ*, **152**, 204  
Perrakis, K., Ntzioufras, I., & Tsionas, E. G. 2014, *Comput. Stat. Data Analysis*, **77**, 54  
Rajpaul, V., Aigrain, S., Osborne, M. A., Reece, S., & Roberts, S. 2015, *MNRAS*, **452**, 2269  
Rasmussen, C. E., & Williams, C. 2006, Gaussian Processes for Machine Learning – The MIT press  
Roberts, S., Osborne, M., Ebden, M., et al. 2012, *Phil. Trans. Roy. Soc. London A*, **371**  
Robertson, P., Bender, C., Mahadevan, S., Roy, A., & Ramsey, L. W. 2016, *ApJ*, **832**, 112  
Wargelin, B. J., Saar, S. H., Pojmański, G., Drake, J. J., & Kashyap, V. L. 2017, *MNRAS*, **464**, 3281  
Zechmeister, M., & Kürster, M. 2009, *A&A*, **496**, 577