

# The HARPS search for southern extra-solar planets – XLV. Two Neptune mass planets orbiting HD 13808: a study of stellar activity modelling’s impact on planet detection

E. Ahrer<sup>1,2★</sup>, D. Queloz<sup>1,3</sup>, V. M. Rajpaul<sup>1</sup>, D. Ségransan<sup>3</sup>, F. Bouchy<sup>3</sup>, R. Hall<sup>1</sup>, W. Handley<sup>1,4</sup>, C. Lovis<sup>3</sup>, M. Mayor<sup>3</sup>, A. Mortier<sup>1,4</sup>, F. Pepe<sup>3</sup>, S. Thompson<sup>1</sup>, S. Udry<sup>3</sup> and N. Unger<sup>3</sup>

<sup>1</sup>*Astrophysics Group, Cavendish Laboratory, JJ Thomson Avenue, CB3 0HE Cambridge, UK*

<sup>2</sup>*Department of Physics, University of Warwick, Gibbet Hill Road, CV4 7AL Coventry, UK*

<sup>3</sup>*Département d’astronomie, Université de Genève, Chemin des Maillettes 51, CH-1290 Versoix, Switzerland*

<sup>4</sup>*Kavli Institute for Cosmology, Cambridge, Madingley Road, CB3 0HA Cambridge, UK*

Accepted 2021 January 27. Received 2020 December 22; in original form 2020 July 21

## ABSTRACT

We present a comprehensive analysis of 10 yr of HARPS radial velocities (RVs) of the K2V dwarf star HD 13808, which has previously been reported to host two unconfirmed planet candidates. We use the state-of-the-art nested sampling algorithm POLYCHORD to compare a wide variety of stellar activity models, including simple models exploiting linear correlations between RVs and stellar activity indicators, harmonic models for the activity signals, and a more sophisticated Gaussian process regression model. We show that the use of overly simplistic stellar activity models that are not well-motivated physically can lead to spurious ‘detections’ of planetary signals that are almost certainly not real. We also reveal some difficulties inherent in parameter and model inference in cases where multiple planetary signals may be present. Our study thus underlines the importance both of exploring a variety of competing models and of understanding the limitations and precision settings of one’s sampling algorithm. We also show that at least in the case of HD 13808, we always arrive at consistent conclusions about two particular signals present in the RV, regardless of the stellar activity model we adopt; these two signals correspond to the previously reported though unconfirmed planet candidate signals. Given the robustness and precision with which we can characterize these two signals, we deem them secure planet detections. In particular, we find two planets orbiting HD 13808 at distances of 0.11, 0.26 au with periods of 14.2, 53.8 d, and minimum masses of 11, 10  $M_{\oplus}$ .

**Key words:** methods: data analysis – methods: statistical – techniques: radial velocities – stars: activity – stars: individual: HD 13808.

## 1 INTRODUCTION

The radial velocity (RV) method has been an important and productive tool for discovering exoplanets ever since it led to the discovery of the first exoplanet orbiting a Sun-like star (Mayor & Queloz 1995). However, the search for small Earth- and Neptune-like planets orbiting Sun-like stars is very challenging. They give rise to relatively small RV signatures of order  $1 \text{ m s}^{-1}$  or less, whilst stellar magnetic activity can induce RV signals that can mimic planetary ones, with amplitudes of order many meters per second (e.g. Queloz et al. 2001).

Thus, with the increasing precision of RV measurements facilitated by upcoming extreme-precision Doppler spectrographs such as ESPRESSO (Pepe et al. 2014), EXPRES (Jurgenson et al. 2016), HARPS3 (Thompson et al. 2016), HIRES (Pasquini et al. 2008), and NEID (Schwab et al. 2016), the characterization of the influence from host stars’ activity on the RV measurements becomes ever more important. Hence, there is a strong interest in developing tools to disentangle stellar and planetary signals in RV data.

A variety of methods exists that attempt to ‘correct’ or model stellar activity signals in RVs, and thus reduce the possibility of false-positive planet detections. Models describing this include pre-whitening and red-noise models (e.g. Hatzes et al. 2010; Feroz & Hobson 2013), as well as studying stellar activity indicators such as the bisector inverse slope (BIS) and full width at half-maximum (FWHM) of the cross-correlation function (CCF) between a target star and a template spectrum, or measurements of chromospheric activity in the target star, such as the  $\log R'_{\text{HK}}$  index (e.g. Boisse et al. 2009; Queloz et al. 2009; Dumusque et al. 2011). The presence of periodic signals in indicators usually suggest an activity-induced signal rather than a planetary one, since a genuine planet would induce a periodic Doppler shift in all spectral lines, but would not produce the same periodic variations in the stellar activity indicators. It is straightforward to use linear correlations (e.g. Queloz et al. 2001) and to include harmonic models which use the rotational period of the star and its harmonics (e.g. Dumusque et al. 2012) to describe stellar activity induced RV variations.

Other, more computationally expensive methods include modelling the stellar surface features directly (e.g. Boisse, Bonfils & Santos 2012) and predicting activity-related RV variability with stel-

\* E-mail: [eva-maria.ahrer@warwick.ac.uk](mailto:eva-maria.ahrer@warwick.ac.uk)

lar activity indicators using the  $FF'$  method and Gaussian processes (GPs) as described in e.g. Aigrain, Pont & Zucker (2012) and Rajpaul et al. (2015); such approaches have already successfully been used to identify false-positive planetary detections as e.g. in Haywood et al. (2014b) or Rajpaul, Aigrain & Roberts (2016).

To quantify the quality of different models for describing observed RV data, Bayesian model comparison has proven to be a robust approach (e.g. Feroz & Hobson 2013; Faria et al. 2016; Hall et al. 2018), also in combination with GP modelling (Faria et al. 2020). However, a significant challenge inherent in this method is very high computing costs when dealing with high-dimensional problems, since computing Bayesian model evidences entails integrating over all model parameter posteriors; a related challenge is degeneracy inherent in combinations of Keplerian and stellar activity models (see Nelson et al. 2020). Nevertheless, it is widely accepted that Bayesian model comparison is a theoretically sound approach to answering questions about competing physical models,<sup>1</sup> and that such computational burdens are therefore a price worth paying (Goodman 1999).

This paper presents a comprehensive analysis of a series of 246 spectroscopic measurements on HD 13808 carried out with HARPS from 2003 to 2014. This star belongs to a sample of bright stars selected for their low level of RV ‘jitter’ to maximize the sensitivity of the survey to low-mass exoplanets. Series of stellar spectrum of HD 13808 with a signal-to-noise ratio (S/N) large enough to reach the photon noise RV measurement error of order  $50 \text{ cm s}^{-1}$  are available. In addition, the exposure time was at least 15 min to minimize the effect of acoustic mode stellar oscillations (e.g. Dumusque et al. 2011; Chaplin et al. 2019). Early in the survey of HD 13808, small amplitude RV variations were detected, suggesting a combination of time variable signal due to a multiplanet systems and moderate stellar activity. We present a comprehensive analysis of HARPS RVs on that star using Bayesian inference to combine various stellar activity models with a set of independent Keplerian orbit solutions.

With this paper, we aim to examine the importance of comparing systematically a number of competing physical models, since the usefulness of Bayesian model comparison is limited by the quality of the models considered (favouring one inappropriate model over another does not imply that either model is in any sense ‘correct’). We show that if an inadequate stellar activity model is considered, one may be led astray and end up with spurious conclusions about the number of planets present in one’s system. Models we consider in our analysis include linear correlations between RV data and stellar activity indicators; sinusoidal models; and joint GP regression of RV measurements and activity. Moreover, we present a thorough test of the nested sampling algorithm POLYCHORD, a state-of-the-art sampler, and discuss its limitations and the importance of the precision settings in exoplanet applications. The HD 13808 system is used as a test case for this study as (i) the host star shows stellar activity variability, and (ii) it has been suggested in the literature that the system hosts at least two-planet candidates, though the existence of these planets was never securely confirmed (see Mayor et al. 2011; Gillon et al. 2017; at the time of writing this paper, the NASA Exoplanet Archive<sup>2</sup> lists HD 13808 as having no confirmed planets). With the analysis in this paper, the status of these candidates elevates to confirmed planets.

<sup>1</sup>The same cannot be said for various other commonly used approaches to model selection, such as residual minimization, Bayesian information criterion testing, etc. (Gelman et al. 2013).

<sup>2</sup>Available online at [exoplanetarchive.ipac.caltech.edu](http://exoplanetarchive.ipac.caltech.edu).

**Table 1.** Description of the HARPS RV data of HD 13808.

HD13808	
Number of observations	246
Time span (d)	4051
mean (RV) ( $\text{m s}^{-1}$ )	41 095
rms (RV) ( $\text{m s}^{-1}$ )	3.94
mean ( $\sigma_{\text{RV}}$ ) ( $\text{m s}^{-1}$ )	0.76
median ( $\log R'_{\text{HK}}$ )	− 4.90

This paper is structured as follows. We describe the HARPS observations in Section 2, followed by the modelling of Keplerian orbits in Section 3. We introduce the stellar activity models applied in this work in Section 4 and outline model comparison using Bayesian inference and POLYCHORD in Section 5. Afterwards, we present the results of our analysis of HD 13808 RVs in Section 6 and conclude with our discussion in Section 7.

## 2 HARPS OBSERVATIONS OF HD 13808

Spectra of HD 13808 were measured with HARPS, a fibre-fed spectrograph installed in a vacuum vessel, mounted on the 3.6m ESO telescope of La Silla in Chile (Mayor et al. 2003). Observations were made using the most precise observing mode that utilizes a simultaneous calibration by the ThAr calibration lamp. In this configuration, HARPS achieves a long-term RV precision better than  $1 \text{ m s}^{-1}$ , allowing us to detect small stellar RV variations of this order (Lovis et al. 2008).

The data obtained by HARPS are automatically processed on site by a data reduction software – the HARPS DRS – that extracts the spectra, calibrates it and eventually computes a CCF with a stellar template. The stellar RV is measured from the CCF as well as other parameters like the FWHM or the BIS (Queloz et al. 2001; Santos et al. 2002). In addition for each HARPS spectra, the value of the Calcium S activity index is estimated from the chromospheric re-emission in the Ca II H and K lines and converted in standard  $\log R'_{\text{HK}}$  index.

Since the release of the first version of the HARPS DRS in 2003, a series of successive versions with improved algorithms have been developed, leading to a steady gain in the RV precision. Details about the historical changes in the HARPS DRS algorithms may be found in following series of papers: Baranne et al. (1996), Pepe et al. (2002), Lovis & Pepe (2007), and Mayor et al. (2009a, b).

Measurements used in this paper were reprocessed with version 3.5 of the HARPS DRS.<sup>3</sup>

A description of the HARPS data used in this study is given in Table 1, while HD 13808’s stellar parameters are summarized in Table 2. HD 13808 has been observed by HARPS 246 times over a span of more than 10 yr, achieving consistent quality of  $\text{mean}(\sigma_{\text{RV}}) = 0.76 \text{ m s}^{-1}$ .

## 3 MODELLING RADIAL VELOCITIES

To model the observed RV of a star at time  $t_i$  with  $N_p$  planets, we sum the planets’ individual Keplerian terms, neglecting planet–planet interactions; following Feroz, Balan & Hobson (2011) and the

<sup>3</sup>All data used in the analysis will be available via VizieR at CDS (Ochsenbein, Bauer & Marcout 2000).

**Table 2.** HD 13808: stellar properties, with corresponding references (1) Gray et al. (2006); (2) Kharchenko (2001); (3) Skrutskie et al. (2006); (4) Santos et al. (2013); (5) Delgado Mena et al. (2019); (6) GAIA DR2 Bailer-Jones et al. (2018).

	HD13808	Reference
Spectral type	K2V	(1)
$V$ (mag)	$8.38 \pm 0.01$	(2)
$K$ (mag)	$6.25 \pm 0.02$	(3)
$T_{\text{eff}}$ (K)	$5035 \pm 50$	(4)
Fe/H (dex)	$-0.21 \pm 0.02$	(4)
$M$ ( $M_{\odot}$ )	$0.771 \pm 0.022$	(5)
Age (Gyr)	$7.2 \pm 4.8$	(5)
$R_S$ ( $R_{\odot}$ )	$0.781^{+0.017}_{-0.022}$	(6)
$d$ (pc)	$28.2535 \pm 0.0256$	(6)

**Table 3.** Priors for the Keplerian, RV trend and additive white noise ('jitter') parameters used in all of our models and analyses.

Parameter	Prior	Lower bound	Upper bound
$P$ (d) <sup>a</sup>	Log uniform	5	100
$K$ ( $\text{m s}^{-1}$ )	Log uniform	0.1	10
$\sqrt{e} \sin(\omega)^b$	Uniform	-1	1
$\sqrt{e} \cos(\omega)^b$	Uniform	-1	1
$\lambda$ (rad)	Uniform	0	$2\pi$
$V$ ( $\text{m s}^{-1}$ ) <sup>c</sup>	Uniform	$-RV_{\text{max}}$	$RV_{\text{max}}$
$\sigma_{\text{RV}}^+$ ( $\text{m s}^{-1}$ )	Uniform	0	20

Notes. <sup>a</sup>We also require for the periods to be sorted, i.e.  $P_b < P_c$  etc.

<sup>b</sup>We also require that the corresponding eccentricity  $e < 1$ .

<sup>c</sup>This prior applies to all polynomial terms.

formalism given in Balan & Lahav (2009), we arrive at

$$RV(t_i) = V_i + \sum_{p=1}^{N_p} K_p [\cos(f_{i,p} + \omega_p) + e_p \cos(\omega_p)], \quad (1)$$

where  $V_i$  is the systemic velocity,  $K_p$  is the RV semi-amplitude,  $f_{i,p}$  is the true anomaly,  $e_p$  is the orbital eccentricity, and  $\omega_p$  is the argument of periastron of the  $p$ th planet, respectively. In addition, the introduction of the mean longitude  $\lambda_p$  of the  $p$ th planet is necessary as part of the computation of  $f_{i,p}$  that requires  $e_p$  and the period of the planet  $P_p$  as well.

A planet's semi-amplitude  $K_p$  is related to its mass  $M_p$ , period and orbital eccentricity  $e_p$ , as well as to the mass of the parent star  $M_S$  and the inclination of the system  $i$  relative to the observer:

$$K_p = \frac{28.4329 \text{ m s}^{-1}}{\sqrt{1 - e_p^2}} \left( \frac{P_p}{1 \text{ yr}} \right)^{-1/3} \frac{M_p \sin(i)}{M_{\text{Jup}}} \left( \frac{M_S}{M_{\odot}} \right)^{-2/3}. \quad (2)$$

In summary, we have five free parameters per planet ( $K$ ,  $\omega$ ,  $e$ ,  $P$ , and  $\lambda$ ) plus a white-noise 'jitter' term  $\sigma_{\text{RV}}^+$  which is added to the observational error in quadrature (see Section 5.2) and any terms for describing the systemic velocity,  $V$  – typically a constant offset plus a linear or possibly quadratic polynomial term. For the analysis in this paper, a quadratic polynomial term was used for all models. Additionally, our models for describing stellar activity contributions to RVs contain anything from 1 to 16 free parameters; these models are described in detail in Section 4.

The uninformative priors for our five free parameters per planet are listed in Table 3, as well as the priors for the polynomial terms and for the additive white-noise term. Note that the lower and upper boundary for planet periods was set to 5 and 100 d for computational efficiency. This choice was supported, in the first instance, by the

**Table 4.** Priors for the parameters of the various stellar activity models described in Section 4.

Parameter	Prior	Lower bound	Upper bound
$\alpha$	Uniform	$-RV_{\text{max}}$	$+RV_{\text{max}}$
$P_{\text{rot}}$	Uniform	30 d	42 d
$P_{\text{magn}}$	Uniform	500 d	4500 d
$\phi$	Uniform	0	$2\pi$
$C_{\text{BIS}}$	Uniform	$-BIS_{\text{max}}$	$+BIS_{\text{max}}$
$C_{\text{magn}}$	Uniform	$-\log R'_{\text{HK,max}}$	$+\log R'_{\text{HK,max}}$
$\beta$	Uniform	$-RV_{\text{max}}$	$+RV_{\text{max}}$
$\Psi_0$	Uniform	$\text{FWHM}_{\text{max}}$	$\text{FWHM}_{\text{max}} + 15\sigma^b$
$\delta V_{cK}$	Uniform	0	10 000
$R^a$	Uniform	$R_S - 5\sigma_S a$	$R_S + 5\sigma_S a$
$V_c$	Uniform	$-\text{rms}(\text{RV})$	$+\text{rms}(\text{RV})$
$V_r$	Uniform	$-\text{rms}(\text{RV})$	$+\text{rms}(\text{RV})$
$L_c$	Uniform	$-\text{rms}(\log R'_{\text{HK}})$	$+\text{rms}(\log R'_{\text{HK}})$
$B_c$	Uniform	$-\text{rms}(\text{BIS})$	$+\text{rms}(\text{BIS})$
$B_r$	Uniform	$-\text{rms}(\text{BIS})$	$+\text{rms}(\text{BIS})$
$P_{\text{GP}}$	Uniform	10 d	100 d
$\lambda_p$	Jeffreys	0.01	10
$\lambda_e$	Jeffreys	10 d	400 d

Notes. <sup>a</sup> $R_S$  is the radius of the star and  $\sigma_S$  is the estimated error.

<sup>b</sup> $\sigma$  is the observed standard deviation of the FWHM.

fact that the Lomb–Scargle power spectrum of RVs revealed no significant periodicities above about 60 d (see Fig. 3). Moreover, models allowing planets with periods longer than 100 d or shorter than 5 d always had lower Bayesian evidences than models without such planets (when using a GP stellar activity model – cf. Section 6), and the inferred RV semi-amplitudes of these short- or long-period planets was always consistent with zero.<sup>4</sup>

The priors used for stellar activity model parameters were generally also chosen to be uninformative, and are discussed in Section 4 and summarized in Table 4.

## 4 MODELLING STELLAR ACTIVITY

In this section, we introduce the multiple ways of modelling stellar activity applied in this work. Simple models like linear dependencies are considered, as well as harmonic modelling of the rotation period of the star, combining BIS and RV measurements in a simultaneous fit. We also discuss more sophisticated approaches to modelling activity-induced RV variations, namely the  $FF'$  approach and simultaneous GP regression over multiple activity indicators.

### 4.1 Linear activity model

A linear relation between RVs and BIS was considered as seen in equation (3), where  $RV_{\text{Kepler}}(t)$  are the RV signatures of the planets at time  $t$  as described previously in equation (1);  $BIS(t)$  represents the BIS measurements taken at time  $t$  and  $\alpha$  describes the free parameter for the linear relation. A quadratic polynomial term including a constant offset is represented by  $V$ .

$$RV_{\text{total}}(t) = RV_{\text{Kepler}}(t) + \alpha BIS(t) + V. \quad (3)$$

This method has already been successful in identifying false positives (e.g. Queloz et al. 2001). However, note that it has been

<sup>4</sup>However, the computational cost of evaluating models with expanded planet-period priors was typically an order of magnitude greater than when using the more restricted prior, and posterior multimodality was far more pronounced.

shown that there can be a temporal offset between activity and associated RV variations on time-scales of few days causing a ‘blur’ of any linear correlation between those two parameters, making a linear model a choice with a major caveat (e.g. Santos et al. 2014; Collier Cameron et al. 2019). The prior distribution for  $\alpha$  is shown with the other parameters related to stellar activity modelling in Table 4.

In a similar way, a linear dependence on the  $\log R'_{\text{HK}}$  activity indicator was considered as we hoped to capture any possible linear long-term relationship with

$$\text{RV}_{\text{total}}(t) = \text{RV}_{\text{Kepler}}(t) + \beta \log R'_{\text{HK}}(t) + V, \quad (4)$$

where  $\beta$  represents the linear factor and  $\log R'_{\text{HK}}(t)$  the  $\log R'_{\text{HK}}$  measurements at time  $t$ . The prior distribution for  $\beta$  can be found in Table 4.

#### 4.2 Harmonic activity model

Our second activity model entailed fitting a sinusoid to both the RVs and BIS time series, enforcing an identical period but allowing different phases and amplitudes between the two. We assume that this captures solar spots and other stellar features which are sensitive to the rotation of the star. In addition, it accounts for a likely time shift which is a problem when considering linear correlations as mentioned in Section 4.1.

Equations (5) and (6) show the two models with the harmonics of the rotation period simultaneously fit to the RV and BIS measurements, respectively. The parameter  $P_{\text{rot}}$  is the putative rotation period of the star fit to both models, while the other parameters  $K_i$ ,  $\phi_{\text{RV}}$ ,  $K_k$ ,  $\phi_{\text{BIS}}$ , and  $C_{\text{BIS}}$  describe the sinusoidal signal of the  $i$ th and  $k$ th harmonic for the  $\text{RV}(t)$  and  $\text{BIS}(t)$  model, respectively, with  $N_h$  being the number of harmonics. As before, the parameter  $V$  represents a quadratic polynomial contribution:

$$\text{RV}_{\text{total}}(t) = \text{RV}_{\text{Kepler}}(t) + \sum_{i=1}^{N_h} K_i \sin\left(\frac{2\pi i t}{P_{\text{rot}}} + \phi_{\text{RV}}\right) + V, \quad (5)$$

$$\text{BIS}(t) = \sum_{k=1}^{N_h} K_k \sin\left(\frac{2\pi k t}{P_{\text{rot}}} + \phi_{\text{BIS}}\right) + C_{\text{BIS}}. \quad (6)$$

The prior distribution for  $K$  is as in Table 3; the priors for  $\phi$ ,  $C_{\text{BIS}}$  and  $P_{\text{rot}}$  are displayed in Table 4. Note that the prior on the rotation period  $P_{\text{rot}}$  was chosen to cover a narrow range based on the following: (i) the rotation period for HD 13808 was estimated before to be  $\sim 40$  d based on the average of  $\log R'_{\text{HK}}$  measurements by Lovis et al. (2011); (ii) signals likely corresponding to the harmonics of the rotation period are detected in the periodograms of the stellar activity indicators (see Section 6.1; Fig. 3). Further motivation was provided by the fact that (iii) when considering zero or one-planet models, the rotation period of the star became locked on to the period of one of the two planets, and that (iv) during preliminary runs, the MAP value for the rotation period never went above 40 d. By limiting the period to below 42 d, we reduced the computational burden of our modelling.

#### 4.3 Long-term magnetic activity model

In a similar fashion to the harmonic activity model, the ‘magn. cycle’ model fits a sinusoid to two data sets, this time to the RVs and the  $\log R'_{\text{HK}}$  time series, enforcing an identical period but allowing different phases and amplitudes. We assume that this captures stellar activity corresponding to a long-term magnetic activity cycle.

Equations (7) and (8) show the two models with the cycle period  $P_{\text{magn}}$  simultaneously fit to the RV and  $\log R'_{\text{HK}}$  measurements, respectively. The parameters  $K_{\text{magn,RV}}$ ,  $\phi_{\text{magn,RV}}$  and  $K_{\text{magn}}$ ,  $\phi_{\text{magn}}$ ,  $C_{\text{magn}}$  describe the  $\text{RV}(t)$  and  $\log R'_{\text{HK}}(t)$  model, respectively;  $V$  represents a constant offset and a quadratic polynomial term.

$$\text{RV}_{\text{total}}(t) = \text{RV}_{\text{Kepler}}(t) + K_{\text{magn,RV}} \sin\left(\frac{2\pi t}{P_{\text{magn}}} + \phi_{\text{magn,RV}}\right) + V \quad (7)$$

$$\log R'_{\text{HK}}(t) = K_{\text{magn}} \sin\left(\frac{2\pi t}{P_{\text{magn}}} + \phi_{\text{magn}}\right) + C_{\text{magn}}. \quad (8)$$

The prior distribution for  $K$  is as in Table 3; the priors for  $\phi$ ,  $C_{\text{magn}}$  and  $P_{\text{magn}}$  are displayed in Table 4. The upper and lower prior limits for  $P_{\text{magn}}$  were determined by a broad peak around 1000–4000 d in the periodogram of  $\log R'_{\text{HK}}$  as well as by the obvious periodic signal in the  $\log R'_{\text{HK}}$  measurements (see Fig. 2).

#### 4.4 $FF'$ method

The FWHM measurements were used for applying the  $FF'$  method that was introduced by Aigrain et al. (2012), as a method for relating the photometric brightness and RV variations of a star. It uses the flux of the star  $\Psi(t)$  and its derivative  $\dot{\Psi}(t)$  as an indicator for spot coverage and predicts RV variations. These variations include the RV perturbation  $\Delta\text{RV}_{\text{rot}}(t)$  due to the presence of spots on the rotating photosphere and their effect of the suppression of convective blueshift  $\Delta\text{RV}_{\text{conv}}(t)$ .

Aigrain et al. (2012) describes these with the following two equations, where  $f$  represents the drop in flux produced by a spot at the centre of the stellar disc,  $R_*$  is the stellar radius,  $\delta V_c$  is the difference between the convective blueshift in the unspotted photosphere and that within the magnetized area and  $\kappa$  is the ratio of this area to the spot surface:

$$\Delta\text{RV}_{\text{rot}}(t) = -\frac{\dot{\Psi}(t)}{\Psi_0} \left[ 1 - \frac{\Psi(t)}{\Psi_0} \right] \frac{R_*}{f}, \quad (9)$$

$$\Delta\text{RV}_{\text{conv}}(t) = \left[ 1 - \frac{\Psi(t)}{\Psi_0} \right]^2 \frac{\delta V_c \kappa}{f}. \quad (10)$$

The total RV variation  $\Delta\text{RV}_{\text{activity}}$  created by stellar activity is then the sum of both terms:

$$\Delta\text{RV}_{\text{activity}} = \Delta\text{RV}_{\text{rot}}(t) + \Delta\text{RV}_{\text{conv}}(t). \quad (11)$$

The parameter  $f$  is approximated as

$$f \approx \frac{\Psi_0 - \Phi_{\text{min}}}{\Psi_0}, \quad (12)$$

with  $\Phi_{\text{min}}$  being the minimum observed flux.

It has been argued that CCF FWHM and  $\log R'_{\text{HK}}$  measurements should both behave, to first order, as the convective blueshift suppression term in the  $FF'$  method, which in turn is a close proxy for the integrated active region coverage on the visible stellar hemisphere (Rajpaul et al. 2015); indeed, it has been shown in practice that the CCF FWHM can be a good tracer of photometric flux (e.g. Suárez Mascareño et al. 2020). Thus, as we did not have photometric measurements of HD 13808 available, we chose to interpret the FWHM as a proxy for the flux  $\Psi(t)$ . The derivative of the ‘flux’ was computed numerically by modelling the FWHM with a GP using the CELERITE algorithm by Foreman-Mackey et al. (2017). In total, there were three free parameters in this stellar activity model,  $\delta V_c \kappa$ ,  $R_*$  and  $\Psi_0$  – their respective priors appear in Table 4.

**Table 5.** Jeffreys’ scale, as introduced by Jeffreys (1983), for interpreting differences in Bayesian evidences. This scale interprets the strength of evidence favouring one model over another.

$ \ln R $	Odds	Probability	Remark
$<1.0$	$\lesssim 3:1$	$<0.750$	Inconclusive
$1.0$	$\sim 3:1$	$0.750$	Weak evidence
$2.5$	$\sim 12:1$	$0.923$	Moderate evidence
$5.0$	$\sim 150:1$	$0.993$	Strong evidence

#### 4.5 GP regression model

We used the GP framework developed by Rajpaul et al. (2015, hereafter R15) to model RVs simultaneously with  $\log R'_{\text{HK}}$  and BIS observations. In short, this framework assumes that all observed stellar activity signals are generated by some underlying latent function  $G(t)$  and its derivatives; this function, which is not observed directly, is modelled with a Gaussian process (Rasmussen & Williams 2006; Roberts et al. 2013).

Following R15, activity variability in the RV,  $\log R'_{\text{HK}}$  and BIS time series can be modelled as

$$\Delta\text{RV} = V_c G(t) + V_r \dot{G}(t), \quad (13)$$

$$\log R'_{\text{HK}} = L_c G(t), \text{ and} \quad (14)$$

$$\text{BIS} = B_c G(t) + B_r \dot{G}(t), \quad (15)$$

respectively. The coefficients  $V_c$ ,  $V_r$ ,  $L_c$ ,  $B_c$ , and  $B_r$  are free parameters relating the individual observations to the unobserved Gaussian process  $G(t)$ . In R15’s framework,  $G(t)$  itself can be loosely interpreted as representing the projected area of the visible stellar disc covered in active regions at a given time; the GP describing  $G(t)$  is assumed to have zero mean and covariance matrix  $\mathbf{K}$ , where  $K_{ij} = \gamma(t_i, t_j)$ . As in R15, we adopt the following quasi-periodic covariance kernel function:

$$\gamma(t_i, t_j) = \exp \left[ -\frac{\sin^2 [\pi(t_i - t_j)/P_{\text{GP}}]}{2\lambda_p^2} - \frac{(t_i - t_j)^2}{2\lambda_e^2} \right], \quad (16)$$

where  $P_{\text{GP}}$  is the period of the quasi-periodic activity signal,  $\lambda_p$  is the inverse harmonic complexity of the signal (such that signals become sinusoidal for large values of  $\lambda_p$ , and show increasing complexity/harmonic content for small values of  $\lambda_p$ ), and  $\lambda_e$  is the time-scale over which activity signals evolve. This quasi-periodic covariance kernel has been widely used to model stellar activity signals in both photometry and RVs (e.g. Haywood et al. 2014a; Grunblatt, Howard & Haywood 2015; Rajpaul et al. 2015; Bonfils et al. 2018). The full expressions for the covariance between the three different observables models are given in R15.

We expect, in principle at least, that this sophisticated GP-based approach to modelling RVs jointly with activity indicators should enable more reliable planet characterization for several reasons. First, by modelling multiple activity-sensitive time series simultaneously (e.g.  $\log R'_{\text{HK}}$ , BIS, FWHM, etc., or some subset of these), more information can be gleaned on activity signals in RVs, compared to exploiting only simple correlations between RVs and (typically) one of these time series. Additionally, the framework uses GP draws and derivatives thereof as basic functions for modelling available time series, rather than e.g. sinusoids or other simple parametric models, the inappropriate use of which could easily lead to the introduction of correlated signals into model residuals. The GP basis functions could in principle take any form, although in the GP framework their properties are constrained to some extent by the data itself, and

from reasonable prior assumptions about the quasi-periodic nature of stellar activity signals. The GP framework also incorporates the  $FF'$  formalism directly as a special case; the former approach may be thought of as a generalization of the latter.

## 5 BAYESIAN INFERENCE AND POLYCHORD

### 5.1 Bayesian model comparison

As we shall use Bayesian inference to evaluate the relative posterior probabilities of different models, we summarize briefly here the relevant formalism. First, Bayes’ Theorem, given in equation (17), is used to relate (i) the posterior probability  $\Pr(\Theta|D, M) = P(\Theta)$  of the parameters  $\Theta$  given data  $D$  and a model  $M$  to (ii) the prior distribution  $\Pr(\Theta|M) = \pi(\Theta)$  of  $\Theta$  given  $M$ , (iii) the likelihood  $\Pr(D|\Theta, M) = L(\Theta)$  of  $D$  given  $\Theta$  and  $M$  and the Bayesian evidence  $\Pr(D|M) = Z$  of  $D$  given  $M$ . Following the notation used by Feroz, Hobson & Bridges (2009):

$$\Pr(\Theta|D, M) = \frac{\Pr(D|\Theta, M) \Pr(\Theta|M)}{\Pr(D|M)}, \quad (17)$$

or simply

$$P(\Theta) = \frac{L(\Theta) \pi(\Theta)}{Z}. \quad (18)$$

For model selection, one can compare two models  $M_1$  and  $M_2$ , given data  $D$ , by computing the ratio of their respective posterior probabilities; this ratio is also known as the Bayes factor  $R$ :

$$R = \frac{\Pr(M_1|D)}{\Pr(M_2|D)} = \frac{\Pr(D|M_1) \Pr(M_1)}{\Pr(D|M_2) \Pr(M_2)} = \frac{Z_1 \Pr(M_1)}{Z_2 \Pr(M_2)}. \quad (19)$$

Note that  $\frac{\Pr(M_1)}{\Pr(M_2)}$  is the relative a priori probability between the two models, which is usually set to one.

To decide whether the evidence difference is significant to favour one model over the other, we make use of the Jeffreys scale as given in Table 5, where the logarithmic scale of the evidence difference is used. This simplifies equation (19) to

$$\ln(R) = \ln \left( \frac{Z_1}{Z_2} \right) = \ln Z_1 - \ln Z_2. \quad (20)$$

### 5.2 Likelihood function

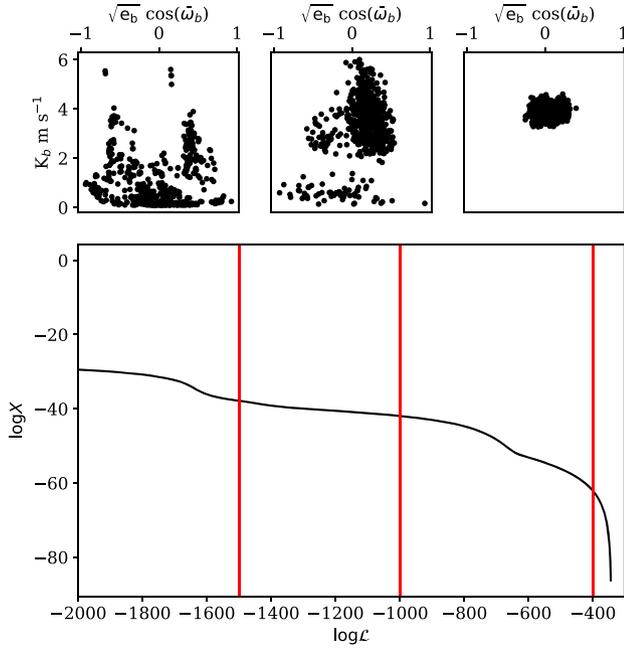
It is commonly assumed that an additive white Gaussian noise (AWGN) model is sufficient for describing observational noise, as in probability theory the central limit theorem states that the sum of independent random variables tends towards a Gaussian distribution, even if the original ones are not normally distributed (Fischer 2011). In this case, the likelihood  $L(\Theta)$  of parameters  $\Theta$  can be written as

$$L(\Theta) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left( -\frac{[v(t_i; \Theta) - v_i]^2}{2\sigma_i^2} \right), \quad (21)$$

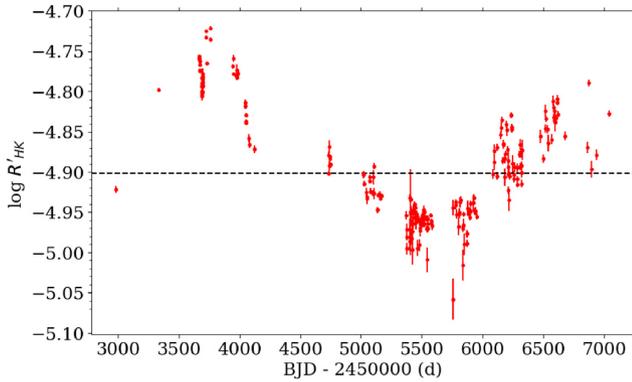
where  $v(t_i; \Theta)$  describes the model’s predicted RV for parameters  $\Theta$ , while  $v_i$  describes the RV observed at time  $t_i$ , with corresponding error estimate  $\sigma_i$  which contains the observational error and the additive noise ‘jitter’ term  $\sigma_{\text{RV}}^+$  added in quadrature. The logarithmic likelihood, which is more convenient to work with, is computed as

$$\ln L(\Theta) = \sum_{i=1}^N -\ln \sqrt{2\pi\sigma_i^2} - \frac{1}{2\sigma_i^2} [v(t_i; \Theta) - v_i]^2. \quad (22)$$

Note that the above formalism applies to all of our models *except* for the GP model, which explicitly generalises the AWGN model by



**Figure 1.** A typical run with POLYCHORD. The bottom graph shows the logarithmic likelihood over the course of the run, with  $\log X$  corresponding to the prior volume i.e. at the start of the run  $\log X$  is at its highest value. Two phase transitions are clearly visible as ‘knees’ in the  $\log L$ – $\log X$  curve. On the top, the live-point distributions of two parameters ( $K_b$  and  $\sqrt{e_b} \cos(\omega_b)$ ) within a given likelihood contour (indicated by the red vertical lines) are shown, demonstrating the transitions between different phases. Note that if the precision criterion is not set low enough, one runs the risk of stopping nested sampling at too early a  $\log X$  value (i.e. at one of the lower knees), and this is indeed what we observed in preliminary tests.



**Figure 2.** Measurement of the  $\log R'_{\text{HK}}$  activity index on the series of spectra of HD 13808. The black dashed line represents the median value of  $-4.90$  which was used as the condition for separating the data set into two parts.

**Table 6.** An overview of the HARPS RV data subsets of HD 13808 referred to as low and high activity when  $\log R'_{\text{HK}}$  is  $< -4.90$  and  $> -4.90$ , respectively.

	Low Activity	High Activity
$\log R'_{\text{HK}}$	$< -4.90$	$> -4.90$
$N$ data	123	123
$\Delta T$ (d)	3332	3701
rms (RV) ( $\text{m s}^{-1}$ )	3.53	3.78
mean ( $\sigma_{\text{RV}}$ ) ( $\text{m s}^{-1}$ )	0.72	0.78
mean ( $\log R'_{\text{HK}}$ )	$-4.95$	$-4.83$

allowing for red (correlated) noise. In this case, the log likelihood may be computed via the more general expression

$$\ln L(\Theta) = -\frac{N}{2} \ln 2\pi - \frac{1}{2} \ln \det \mathbf{K} - \frac{1}{2} \mathbf{r}^T \mathbf{K}^{-1} \mathbf{r}, \quad (23)$$

where  $\mathbf{K}$  is the matrix defining the covariance between all pairs of observations (see R15), and  $\mathbf{r}$  is a vector of residuals with the  $i$ th element given by  $v(t_i; \Theta) - v_i$ . Note that in the special case where  $\mathbf{K}$  is a diagonal matrix with the  $i$ th element given by  $\sigma_i^2$ , i.e. where the noise is assumed to be white, equation (23) reduces to equation (22).

As the Keplerian model requires the Kepler equation to be solved, we made use of the efficient CORDIC-like method introduced by Zechmeister (2018) where double precision is obtained within 55 iterations.

### 5.2.1 Avoiding unstable Keplerian orbits

While computing likelihoods we checked if each pair of planets would be stable following the criterion introduced by Gladman (1993):  $\Delta > 2\sqrt{3} R_H(i, j)$  where  $\Delta = a_j - a_i$  is the difference between the semimajor axis of the  $i$ th and  $j$ th planet and  $R_H(i, j)$  the planets’ mutual Hill radius. This has been used before e.g. by Malavolta et al. (2017).

If the parameters drawn resulted in a planet system which was considered unstable, the likelihood was set to zero, i.e.  $Z_0 = 0$ ; in our case, as we worked with log likelihoods, this was approximated numerically by  $\ln Z_0 = -1 \times 10^{30}$ .

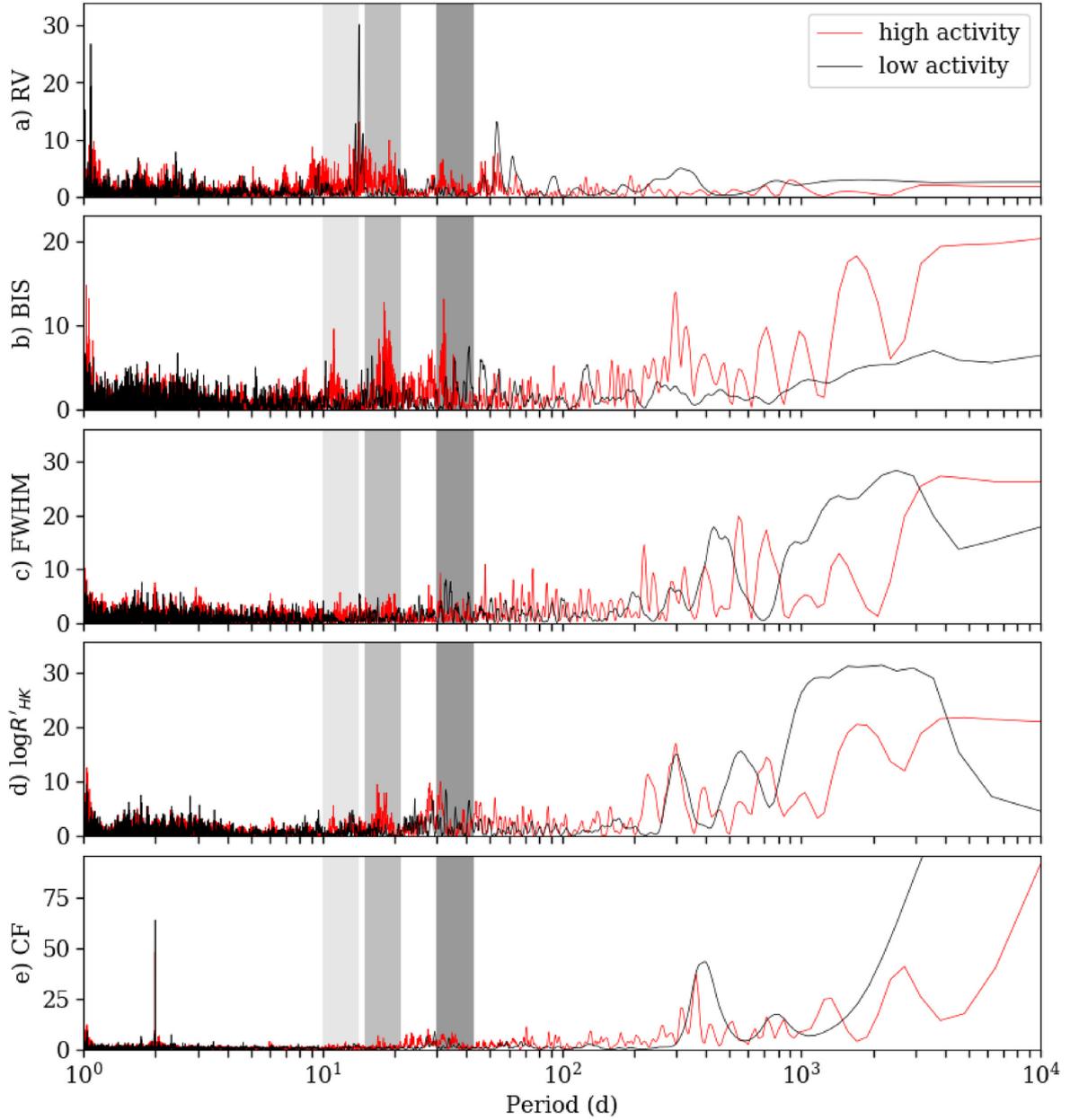
## 5.3 POLYCHORD

We used POLYCHORD as it is a state-of-the-art nested sampling algorithm, designed to work with very high dimensional parameter spaces (Handley, Hobson & Lasenby 2015). A short discussion comparing it with MULTINEST (Feroz et al. 2009) can be found in Hall et al. (2018). It was developed using C++ and FORTRAN and can also be called as a PYTHON package.

In order to test the reliability of our algorithm with POLYCHORD, we used two sets of simulated data. One data set contained random uncorrelated Gaussian noise (with a standard deviation of  $1 \text{ m s}^{-1}$ ) without any planetary signals, while the other one consisted of two planets on elliptical orbits with the same Gaussian noise. Both data sets also included an offset of a few  $\text{m s}^{-1}$  and linear trend of  $10^{-5} \text{ m s}^{-1}$  per day. Models of 0–3 planet signals with long-term trends in the form of second-order polynomials were fitted and their evidences compared. Both elliptical and circular orbital solutions were computed. We note here that in general the evidence uncertainty reported by POLYCHORD is an underestimate, and it is better practice to estimate it via the range of scatter in the evidence across multiple runs (Higson et al. 2018; Nelson et al. 2020); we used the latter approach throughout our analyses. (It is probable that this behaviour would also be ameliorated by increasing the number of live points beyond the computational resources available at the time of writing.)

For the first data set without planetary signals, the algorithm successfully favoured the no-planet model consisting of a second-order polynomial. Every other model computed showed lower evidences, with at least  $\ln R \approx 2$  for the zero-planet model versus one sinusoidal signal or  $\ln R \approx 3$  for the zero-planet model versus one elliptical orbital signal; evidences decreased with the number of signals fitted.

Similarly, the two planets with eccentric orbits in the second test were also recovered: the preferred model featured two planets, with



**Figure 3.** BGLS periodograms of, from top to bottom, the RV, BIS, FWHM, and  $\log R'_{\text{HK}}$  measurements and of a constant function (CF); that is, a data set with the same timestamps and RV errors but assuming constant RV. The colours red and black indicate the high-activity and low-activity phases. The dark grey shaded box represents the range of the putative rotation period with its second (grey) and third harmonic (light grey).

**Table 7.** HD 13808: Overview of the models used for the study of the HD 13808.

Model name	Description
Circular	Circular orbit modelling
Kepler	Eccentric orbit modelling
+ magn. cycle	Kepler + long-term magnetic activity cycle; Section 4.3
+ linear BIS	Kepler + linear dependency on the BIS; Section 4.1, equation (3)
+ BIS $P_{\text{rot}}$ 1st harm.	Kepler + simultaneous fit of 1st harmonic of $P_{\text{rot}}$ to the BIS and RVs; Section 4.2
+ BIS $P_{\text{rot}}$ 2nd harm.	Kepler + simultaneous fit of 1st and 2nd harmonic of $P_{\text{rot}}$ to the BIS and RVs; Section 4.2
+ BIS $P_{\text{rot}}$ 3rd harm.	Kepler + simultaneous fit of 1st, 2nd, and 3rd harmonic of $P_{\text{rot}}$ to the BIS and RVs; Section 4.2
+ linear $\log R'_{\text{HK}}$	Kepler + linear dependency on the $\log R'_{\text{HK}}$ ; Section 4.1, equation (4)
+ $FF'$	Kepler + $FF'$ method with FWHM as a proxy for flux; Section 4.4
Gaussian process	Simultaneous GP modelling of RVs and activity indicators; Keplerian terms for RVs; Section 4.5

**Table 8.** HD 13808: Relative Bayesian evidences for different number of planets for each model and their scatter – as well as the residual RV (O–C) rms and median absolute deviation (MAD) – for the two-planet model (or 1-planet model, if favoured) computed using the respective MAP values (see Table 10). The model with two planetary signals for each model type was set to be zero. The errors correspond to the standard error on the mean evidence across multiple runs or to the highest individual run error provided by POLYCHORD, whichever of the two is greater.

Model	No planets $\ln(R)$	1 planet $\ln(R)$	2 planets $\ln(R)$	3 planets $\ln(R)$	4 planets $\ln(R)$	RV residual (rms m s <sup>-1</sup> )	RV residual MAD m s <sup>-1</sup>
Circular	-65 ± 2	-15 ± 2	0 ± 3	+2 ± 3	+6 ± 2	4.09	2.79
Kepler	-56 ± 3	-6 ± 4	0 ± 4	+6 ± 4	+6 ± 4	3.23	2.31
+ magn. cycle	-78 ± 3	-21 ± 3	0 ± 3	+10 ± 4	-8 ± 13	3.89	2.41
+ linear BIS	-62 ± 2	-12 ± 2	0 ± 3	-2 ± 2	0 ± 2	2.97	2.11
+ BIS $P_{\text{rot}}$ 1st harm.	-51 ± 4	-7 ± 4	0 ± 4	-2 ± 8	-10 ± 6	2.62	1.76
+ BIS $P_{\text{rot}}$ 2nd harm.	-52 ± 2	-7 ± 3	0 ± 3	-3 ± 6	-6 ± 4	2.92	1.99
+ BIS $P_{\text{rot}}$ 3rd harm.	-71 ± 4	-28 ± 4	0 ± 3	-9 ± 4	-15 ± 4	2.96	1.73
+ magn. cycle	-61 ± 6	-7 ± 6	0 ± 6	-7 ± 8	+2 ± 9	2.62	1.66
+ linear log $R'_{\text{HK}}$	-71 ± 3	-16 ± 3	0 ± 3	-6 ± 4	-2 ± 4	2.18	1.51
+ $FF'$	-64 ± 1	-14 ± 2	0 ± 1	-3 ± 5	-9 ± 6	2.47	1.53
Gaussian process	-73 ± 1	-19 ± 2	0 ± 1	-4 ± 3	-23 ± 6	2.18	1.45
Circular low activity	-45 ± 3	-10 ± 4	0 ± 4	-4 ± 4	+9 ± 4	3.08	2.08
Kepler low activity	-37 ± 1	-5 ± 1	0 ± 2	+14 ± 1	+7 ± 2	2.17	1.33
+ BIS $P_{\text{rot}}$ 3rd harm.	-38 ± 1	-16 ± 2	0 ± 2	+1 ± 4	+3 ± 2	2.30	1.46
+ $FF'$	-38 ± 4	-7 ± 4	0 ± 2	-5 ± 4	+5 ± 7	4.15	3.08
Gaussian process low activity	-41 ± 1	-7 ± 1	0 ± 2	-2 ± 4	-3 ± 1	1.75	1.09
Circular high activity	-15 ± 1	-4 ± 1	0 ± 2	+3 ± 2	+3 ± 1	3.83	2.57
Kepler high activity	-14 ± 1	-2 ± 1	0 ± 1	+4 ± 4	+4 ± 2	3.35	2.07
+ BIS $P_{\text{rot}}$ 3rd harm.	-7 ± 2	+2 ± 2	0 ± 2	+2 ± 3	+4 ± 2	3.61	2.32
+ $FF'$	-16 ± 4	-2 ± 4	0 ± 2	-1 ± 4	-6 ± 7	3.75	2.42
Gaussian process high activity	-16 ± 1	-4 ± 2	0 ± 3	-1 ± 3	-3 ± 2	2.47	1.57

**Table 9.** HD 13808: Periods for a putative third planet found in the various runs by the different models, their corresponding semiamplitudes and the uncertainty (per Table 8) in the model evidences. Values in bold occur in at least two of the runs.

Model	$N_{\text{runs}}$	$\sigma_{\ln Z}$	Periods (d)	Semi-amplitudes (m s <sup>-1</sup> )
Circular	5	3	<b>11</b> , 32, 63	~0.6, 0.7, 0.7
Kepler	3	4	8, <b>11</b> , 19	~1.1, 0.7, 0.8
+ magn. cycle	3	4	12, <b>19</b>	~0.6, 0.8
+ linear BIS	3	2	10, 12, 77	~0.5, 0.5, 0.9
+ BIS $P_{\text{rot}}$ 1st harm.	3	7	11, 34, 64	~0.7, 0.9, 0.7
+ BIS $P_{\text{rot}}$ 2nd harm.	3	6	10, 12, 34	~0.2, 0.2, 0.2
+ BIS $P_{\text{rot}}$ 3rd harm.	3	4	<b>12</b> , 26	~0.9, 0.9
+ magn. cycle	3	8	<b>18</b> , 63	~1.2, 1.7
+ linear log $R'_{\text{HK}}$	3	4	9, 23, 29	~0.7, 0.5, 0.5
+ $FF'$	3	5	12, 29, 43	~1.1, 0.9, 0.5
Gaussian process	3	3	28, <b>80</b>	~0.3, 0.5
Circular low activity	4	4	12, 22, 29, 74	~0.7, 0.5, 1.1, 0.6
Kepler low activity	3	1	10, 33, 49	~0.8, 0.7, 1.0
+ BIS $P_{\text{rot}}$ 3rd harm.	3	4	8, 17, 23	~0.2, 0.4, 0.1
+ $FF'$	3	3	9, 21, 65	~1.0, 0.7, 0.9
Gaussian process low activity	3	4	9, 26, 34	~0.5, 0.5, 0.2
Circular high activity	3	2	11, 28, 76	~1.2, 1.0, 1.0
Kepler high activity	4	4	12, 19	~1.7, 1.6
+ BIS $P_{\text{rot}}$ 3rd harm.	3	3	9, 11, 34	~0.8, 1.3, 2.3
+ $FF'$	3	4	10, <b>12</b>	~1.6, 1.4
Gaussian process high activity	3	3	<b>5</b> , 19	~0.4, 2.0

**Table 10.** HD 13808: Comparison of the orbital parameters semiamplitude  $K$ , period  $P$ , and eccentricity  $e$  for the two planets HD 13808b and HD 13808c for each model. The parameters for the two-planet models of the run with the highest evidence from all runs are shown here. Note that not all models favour the two-planet case.

Model	$K_b$ (m s <sup>-1</sup> )	$K_c$ (m s <sup>-1</sup> )	$P_b$ (d)	$P_c$ (d)	$e_b$	$e_c$
Circular	$3.65^{+0.26}_{-0.23}$	$1.95^{+0.24}_{-0.22}$	$14.1801^{+0.0011}_{-0.0014}$	$53.822^{+0.032}_{-0.025}$	–	–
Kepler	$3.62 \pm 0.25$	$2.07 \pm 0.28$	$14.17803^{+0.00098}_{-0.00067}$	$53.794^{+0.054}_{-0.072}$	$0.110^{+0.043}_{-0.054}$	$0.24 \pm 0.13$
+ magn. cycle	$3.60 \pm 0.21$	$2.08 \pm 0.23$	$14.1781^{+0.0013}_{-0.0016}$	$53.801^{+0.050}_{-0.064}$	$0.061^{+0.020}_{-0.056}$	$0.175^{+0.066}_{-0.17}$
+ linear BIS	$3.54 \pm 0.25$	$2.00 \pm 0.24$	$14.1737^{+0.0012}_{-0.00097}$	$53.858 \pm 0.059$	$0.086^{+0.036}_{-0.066}$	$0.140^{+0.046}_{-0.13}$
+ BIS $P_{\text{rot}}$ 1st harm.	$3.55 \pm 0.25$	$1.96 \pm 0.25$	$14.18329 \pm 0.00070$	$53.745 \pm 0.012$	$0.066 \pm 0.017$	$0.212 \pm 0.034$
+ BIS $P_{\text{rot}}$ 2nd harm.	$3.61 \pm 0.25$	$2.14^{+0.25}_{-0.29}$	$14.1780^{+0.0011}_{-0.00086}$	$53.778^{+0.044}_{-0.039}$	$0.086^{+0.036}_{-0.055}$	$0.26^{+0.13}_{-0.15}$
+ BIS $P_{\text{rot}}$ 3rd harm.	$3.61 \pm 0.24$	$2.09 \pm 0.25$	$14.1774^{+0.0013}_{-0.0011}$	$53.804^{+0.033}_{-0.052}$	$0.089^{+0.037}_{-0.065}$	$0.21^{+0.10}_{-0.17}$
+ magn. cycle	$3.56 \pm 0.21$	$2.07 \pm 0.20$	$14.1789^{+0.0014}_{-0.0012}$	$53.803 \pm 0.026$	$0.059^{+0.024}_{-0.052}$	$0.102^{+0.041}_{-0.099}$
+ linear log $R'_{\text{HK}}$	$3.57 \pm 0.21$	$2.15 \pm 0.23$	$14.1782^{+0.0011}_{-0.00065}$	$53.777^{+0.020}_{-0.060}$	$0.073^{+0.033}_{-0.048}$	$0.20^{+0.10}_{-0.14}$
+ $FF'$	$3.57 \pm 0.26$	$1.92^{+0.26}_{-0.34}$	$14.1766 \pm 0.0013$	$53.965^{+0.099}_{-0.20}$	$0.095^{+0.046}_{-0.066}$	$0.237^{+0.099}_{-0.20}$
Gaussian process	$3.67 \pm 0.22$	$2.18^{+0.22}_{-0.20}$	$14.1815 \pm 0.0015$	$53.753^{+0.050}_{-0.082}$	$0.071^{+0.027}_{-0.047}$	$0.156^{+0.050}_{-0.061}$
Circular low activity	$3.71^{+0.26}_{-0.15}$	$2.08 \pm 0.22$	$14.1876^{+0.0049}_{-0.0089}$	$53.71^{+0.15}_{-0.22}$	–	–
Kepler low activity	$3.46^{+0.42}_{-0.36}$	$1.65^{+0.33}_{-0.22}$	$14.191 \pm 0.016$	$53.81^{+0.14}_{-0.26}$	$0.075^{+0.025}_{-0.065}$	$0.145^{+0.044}_{-0.14}$
+ BIS $P_{\text{rot}}$ 3rd harm.	$3.68 \pm 0.25$	$2.06 \pm 0.24$	$14.2019 \pm 0.0017$	$53.365^{+0.024}_{-0.019}$	$0.069^{+0.027}_{-0.066}$	$0.160^{+0.060}_{-0.13}$
+ $FF'$	$3.53 \pm 0.29$	$1.88^{+0.33}_{-0.30}$	$14.193^{+0.013}_{-0.012}$	$53.78^{+0.20}_{-0.28}$	$0.058^{+0.029}_{-0.057}$	$0.122^{+0.037}_{-0.12}$
Gaussian process low activity	$3.43^{+0.53}_{-0.59}$	$1.78^{+0.40}_{-0.34}$	$14.1870 \pm 0.0050$	$53.89^{+0.27}_{-0.087}$	$0.066^{+0.023}_{-0.037}$	$0.098^{+0.025}_{-0.041}$
Circular high activity	$2.81 \pm 0.40$	$1.202^{+0.036}_{-0.058}$	$14.18 \pm 0.30$	$19.1^{+2.0}_{-2.5}$	–	–
Kepler high activity	$2.94 \pm 0.40$	$1.41^{+1.1}_{-0.46}$	$14.132^{+0.046}_{-0.035}$	$19.2^{+2.8}_{-2.2}$	$0.187^{+0.071}_{-0.17}$	$0.45^{+0.23}_{-0.38}$
+ BIS $P_{\text{rot}}$ 3rd harm.	$2.19^{+0.50}_{-0.43}$	$1.20^{+0.46}_{-0.58}$	$14.15^{+0.30}_{-0.39}$	$52 \pm 10$	$0.181^{+0.060}_{-0.16}$	$0.201^{+0.086}_{-0.19}$
+ $FF'$	$3.17 \pm 0.33$	$1.14^{+0.34}_{-0.52}$	$14.155^{+0.035}_{-0.060}$	$53^{+14}_{-2}$	$0.170^{+0.067}_{-0.14}$	$0.28^{+0.33}_{-0.28}$
Gaussian process high activity	$2.84 \pm 0.34$	$1.04^{+0.40}_{-0.57}$	$14.156^{+0.065}_{-0.039}$	$52.7^{+1.3}_{-1.1}$	$0.150 \pm 0.036$	$0.108^{+0.036}_{-0.051}$

the second best model containing three planets with eccentric orbits, albeit with a significantly lower evidence than the two-planet model, and thus rejected. The typical rms of the log evidences was of order  $\sim 3$  for circular orbital fits and of order  $\sim 1$  for eccentric orbital fits.

Over multiple runs with complex models, our tests showed that the default stopping criterion of POLYCHORD was not precise enough<sup>5</sup> for these sort of exoplanet applications. This resulted in inconsistencies where the sampling in some runs missed transitions to areas of higher likelihood (see Fig. 1) and showed very different evidence values and posterior distributions. To ensure consistent and robust runs, the convergence criterion was lowered to values between  $10^{-5}$  and  $10^{-12}$  for runs with a large number of planets and complex stellar activity models. As an historical aside, nested sampling was in fact invented (Skilling 2006) in order to solve precisely these kind of phase transition problems.

## 6 ANALYSIS AND RESULTS

The extent of stellar activity observed from HD 13808 was investigated by looking at the activity indicators BIS, FWHM and  $\log R'_{\text{HK}}$ . None of them showed an obvious linear correlation with the RV measurements. Note that this can occur e.g. when there is a relation but with a simple time lag, as demonstrated with the case of the Sun Collier Cameron et al. (2019). However, the  $\log R'_{\text{HK}}$  measurements do show a notable cycle from a low of  $-5.10$  to a high of  $-4.70$ ,

<sup>5</sup>The convergence criterion in POLYCHORD is defined as the point where the fraction of total evidence contained in the live points drops below the default value of  $10^{-3}$  (Handley et al. 2015).

as displayed in Fig. 2. A cycle with similar shape and period is also seen in the BIS and FWHM time series, although with smaller amplitude. We thus decided to extend our analysis by splitting the full RV data into two subsets based on the median( $\log R'_{\text{HK}}$ ) =  $-4.90$ ; low- and high-activity subsets then corresponded to observations with  $\log R'_{\text{HK}} < -4.90$  and  $\log R'_{\text{HK}} > -4.90$ , respectively. Table 6 summarizes the statistics of the full data set and those two subsets. Note that although we call one subset ‘high activity,’ HD 13 808 is still classified as an inactive star, as an ‘active’ classification would usually require an average of  $\log R'_{\text{HK}} > -4.75$  (Henry et al. 1996). This  $\log R'_{\text{HK}}$  cycle shows great similarities to the one of the Sun with its recent maxima and minima being at  $\log R'_{\text{HK}} \simeq -4.83$  and  $\log R'_{\text{HK}} \simeq -4.96$ ; the extrapolated  $\log R'_{\text{HK}}$  value for the Maunder minimum period is  $\log R'_{\text{HK}} \simeq -5.10$  (Mamajek & Hillenbrand 2008).

### 6.1 Periodogram analysis

To identify strong periodicities in our time series, we computed the Bayesian generalized Lomb–Scargle (BGLS) periodogram (Lomb 1976; Scargle 1982; Zechmeister & Kürster 2009; Mortier et al. 2015) of the two separated data subsets for the RV, the stellar activity indicators BIS, FWHM and  $\log R'_{\text{HK}}$  and a constant function (CF) with identical time stamps.

The resulting BGLS periodograms are shown in Fig. 3 where peaks in the RV periodograms are visible at  $\sim 15$  and  $\sim 55$  d. Sets of strong peaks are visible in the BGLS periodograms of the activity indicators  $\log R'_{\text{HK}}$ , BIS and FWHM between 30 and 40 d and a few between 25 and 30 d. Strong peaks are evident at  $\sim 19$  and  $\sim 32$  d in the  $\log R'_{\text{HK}}$  high activity periodogram, with peaks at  $\sim 12$ ,  $\sim 19$ ,

**Table 11.** HD 13808: Rotation periods determined by the different models.

Model	Rotation period (d)
Kepler	
+ BIS $P_{\text{rot}}$ 1st harm.	$36.0414^{+0.0022}_{-0.0056}$
+ BIS $P_{\text{rot}}$ 2nd harm.	$35.2516^{+0.0045}_{+0.00087}$
+ BIS $P_{\text{rot}}$ 3rd harm.	$32.13212^{+0.00096}_{-0.00025}$
+ magn. cycle	$36.0383^{+0.0025}_{-0.00068}$
Gaussian process	$38.99 \pm 0.47$
Kepler low activity	
+ BIS $P_{\text{rot}}$ 3rd harm.	$36.22876^{+0.00094}_{-0.00047}$
Gaussian process low activity	$41.32^{+0.64}_{-0.92}$
Kepler high activity	
+ BIS $P_{\text{rot}}$ 3rd harm.	$37.93 \pm 0.46$
Gaussian process high activity	$40.0 \pm 1.1$

**Table 12.** MAP values and  $\pm 1\sigma$  credible intervals for the parameters of the activity-related parameters of the GP model.

Parameter	Inferred value	Units
$V_c$	$0.48 \pm 0.14$	$\text{m s}^{-1}$
$V_r$	$0.53^{+0.11}_{-0.53}$	$\text{m s}^{-1}$
$L_c$	$0.0515^{+0.0044}_{-0.0054}$	–
$B_c$	$2.67^{+0.23}_{-0.32}$	$\text{m s}^{-1}$
$B_r$	$13.8^{+2.8}_{-3.1}$	$\text{m s}^{-1}$
$P_{\text{GP}}$	$38.99 \pm 0.47$	d
$\lambda_p$	$0.996^{+0.089}_{-0.10}$	–
$\lambda_e$	$91 \pm 10$	d

and  $\sim 34$  d in the BIS periodogram. The FWHM periodograms show peaks in the region of  $\sim 35$  d.

The BGLS periodograms of the high- and low-activity subsets of the data suggest that the RV measurements are affected by stellar activity: note, in particular, the excess power around the putative stellar rotation period and the first two harmonics in the high-activity subsets of the BIS and  $\log R'_{\text{HK}}$  time series. It is also evident that the stellar activity indicators differ quite strongly when the star is slightly more active; the difference between the mean  $\log R'_{\text{HK}}$  value of the two subsets is  $\Delta \langle \log R'_{\text{HK}} \rangle = 0.12$ .

## 6.2 Model comparison

The different models we considered are summarized in Table 7, along with a short description of each. The model including only circular orbital solutions is called ‘Circular’, while the elliptical orbital solutions are referred to as ‘Kepler’. A ‘+’ indicates that the ‘Kepler’ solution was complemented with a specific stellar activity model, such as linear dependence on the BIS (‘Kepler + linear BIS’). This was done equivalently for a subset of models applied to the low and high-activity subsets of the data.

### 6.2.1 Results

The relative Bayesian evidences for each model with different number of Keplerians are shown in Table 8. For ease of comparison, we set the log evidence for models with two Keplerians to zero. Thus, positive values are more favoured and negative ones less favoured than the two-planet models, with significance to be interpreted according to Jeffreys’ scale (Table 5). Table 8 also includes the rms

and median values for the RV residuals computed using the maximum a posteriori (MAP) values of the various two-planet models.

Note that the absolute evidences of different models cannot be meaningfully compared in a global sense, as the evidence values depend, among other things, on the data fitted; hence, an  $n$ -Keplerian model that fits only RVs cannot e.g. be meaningfully compared to another  $n$ -Keplerian model that fits RV simultaneously with a BIS or  $\log R'_{\text{HK}}$  time series. In passing we note, however, that it *was* possible to compare the absolute evidence values of the parametric model ‘Kepler + BIS  $P_{\text{rot}}$  3rd harm. + magn. cycle’ with the GP model, as these fitted the same three time series; the two-planet GP model achieved a log evidence an order of magnitude higher than the former model ( $-584 \pm 3$  versus  $-4810 \pm 6$ ). This overwhelming difference in Bayesian evidences reflects the fact that the GP model achieved a superior fit to the observational data (RVs,  $\log R'_{\text{HK}}$ , and BIS series jointly), and moreover that high-quality fits under the GP model were localized to a comparatively small volume of prior space.

### 6.2.2 Spurious third and fourth planets

Table 8 shows that many of the simple parametric models – though not the GP model – favoured the ‘detection’ of three or even four planets. However, we have strong empirical reasons to reject models with more than two Keplerian components.

First, we found that the periods of the third Keplerian signal (and also the fourth, where applicable), tended to change with every run, while the periods of  $\sim 14$  d and  $\sim 54$  d appeared as MAP values in every run, for all models with two or more Keplerians. This is demonstrated in Table 9, where we list the periods of the additional Keplerian periods found by every model. The most common periods for an ‘extra’ Keplerian were around 12 and 19 d – as it turns out, these periods out are also strongly visible in the high-activity periodogram of the  $\log R'_{\text{HK}}$  and BIS time series (see Fig. 3). As mentioned in the BGLS periodogram analysis, we interpret these periods as harmonics of the stellar rotation period. The fact that these periodicities manifested in Keplerian terms in all but the most complex activity models reflect the inadequacy of the simpler models at capturing all the activity variability – neither the 12 d nor the 19 d Keplerian periodicities showed up under the GP model.

Secondly, it also turned out that these supposed planetary signals described by the Keplerians either did not show up or their semi-amplitudes decrease to below  $1 \text{ m s}^{-1}$  when considering only the low-activity subset of observations. Conversely, the semi-amplitudes of the supposed third planets occurring at  $\sim 12$  d or  $\sim 19$  d increased significantly when only the high-activity data set was modelled compared to when using the full data set (see Table 9).

Perhaps most significantly, our GP model – which we regarded *a priori* as being by far the most realistic of our activity models – decisively favoured a two-planet interpretation of our full set of observations. The GP modelling even favoured a two-planet interpretation when considering only the high-activity subset of observations, and resulted in planet properties consistent with those derived from the low-activity subset or the full data set (see Table 10). This was not true for the simpler parametric models; the  $FF'$  method equivocated between two- and three-planet solutions, and was not reliably able to detect the  $\sim 54$  d periodicity.

Our GP activity model, though relatively complex in its own right, did a comparatively good job of fitting variability in RVs unrelated to planets, when compared to most of the other activity models (see residuals in Table 8). This is despite our requirement that the GP fit activity-related variability in RVs *simultaneously* with BIS

**Table 13.** MAP values and  $\pm 1\sigma$  credible intervals of the planetary parameters of the two planets in the HD 13808 system for the two favoured models: ‘Gaussian process’ and ‘Kepler + BIS  $P_{\text{rot}}$  3rd harm.’

Parameter	HD 13808b		HD 13808c	
	BIS $P_{\text{rot}}$ 3rd harm.	Gaussian process	BIS $P_{\text{rot}}$ 3rd harm.	Gaussian process
$P$ (d)	$14.1774^{+0.0013}_{-0.0011}$	$14.1815 \pm 0.0015$	$53.804^{+0.033}_{-0.052}$	$53.753^{+0.050}_{-0.082}$
$K$ ( $\text{m s}^{-1}$ )	$3.61 \pm 0.24$	$3.67 \pm 0.22$	$2.09 \pm 0.25$	$2.18^{+0.22}_{-0.20}$
$e$	$0.089^{+0.037}_{-0.065}$	$0.071^{+0.027}_{-0.047}$	$0.21^{+0.10}_{-0.17}$	$0.156^{+0.050}_{-0.061}$
$M\sin(i)$ ( $M_{\oplus}$ )	$11.44 \pm 0.80$	$11.2^{+1.2}_{-0.66}$	$10.0 \pm 1.2$	$9.96^{+1.8}_{-0.96}$
$a$ (au)	$0.1051 \pm 0.0010$	$0.10519^{+0.00093}_{-0.0011}$	$0.2558 \pm 0.0024$	$0.2554^{+0.0028}_{-0.0023}$

and  $\log R'_{\text{HK}}$  time series, which ensured that the GP would be very unlikely to try to fit planetary variability. Presumably, then, under the GP modelling, there was little residual activity-induced variability that could be explained using a third or fourth Keplerian term; hence, these more complex models were rejected.

Regarding the parametric activity models (i.e. not the GP model), the addition of extra Keplerian terms beyond the two-planet model in many cases *did* enable a non-trivial fraction of extra activity-related variability to be fitted. This may have been the case, for example, because the amplitudes, phases, and periods of the different terms in the harmonic models were by definition fixed, whereas the quasi-periodic GP model is flexible enough to fit both low- and high-activity variability, possibly with changing phases and arising from a combination of rotation periods (as might be the case due e.g. to differential rotation), rather than a single ‘master’ rotation period (Rajpaul 2017). Thus, the non-GP activity models in many cases favoured solutions with three or four Keplerian terms.

The upshot, then, is that one’s conclusions about the number of planetary signals present in an RV time series is extremely sensitive to whether one has modelled other, non-planetary (e.g. activity-induced) variability present in the RVs.

### 6.2.3 Comments on computational costs

Given the high dimensionality of the joint stellar activity plus planetary models we considered (over 30 parameters for some of the four-planet plus activity models), accurate Bayesian evidence computation required a large number of posterior samples to be drawn – in the case of four-planet models, several hundred million posterior samples were typically required, even with a state-of-the-art sampler such as POLYCHORD. This did not prove too burdensome for the parametric activity models; a typical run for a four-planet model with  $\sim 6$  additional stellar activity modelling parameters took 30–50 h on 10 cores with 2.9 GHz clock speed each.

Evaluating a single GP likelihood, however, required among other things inversion of a  $738 \times 738$ -element covariance matrix; considering that we computed Bayesian evidences for models containing between zero and four planets, and repeated these evidence calculations several times, we ended up implicitly inverting large covariance matrices many billions of times.<sup>6</sup> The upshot was that evaluating our GP models required of order fifty thousand CPU core hours on a high-performance computing platform. (Attempts to speed up computation by reducing POLYCHORD’s precision criterion led to

unacceptably large scatter in computed evidences.) Accurate evaluation of the Bayesian evidences of these GP models would simply not have been feasible on a desktop or even a small computing cluster.

At face value, this might seem to be a serious shortcoming of the GP model. However, we note that Bayesian evidence computation is computationally challenging in general (Nelson et al. 2020), and it would seem that you ‘get what you pay for’: the parametric activity models are certainly far cheaper to evaluate, though as we argued in Section 6.2.2, their oversimplicity inevitably leads to the detection of spurious planets. Such issues are avoided with the GP model.

### 6.3 Planet characteristics

The posterior distributions for the inferred semiamplitude  $K$ , period  $P$  and eccentricity  $e$  of the planets in all of our two-planet models are summarized in Table 10. Despite the wide array of models considered, the planetary parameters show a remarkable degree of consistency, with the credible intervals for their semi-amplitudes and eccentricities agreeing within  $1\sigma$  across almost all models; the same holds true when considering only the low-activity subset of observations. Differences between orbital periods inferred under different models are never greater than a fraction of a per cent. This provides additional evidence that the detected signals are indeed planetary and robust.

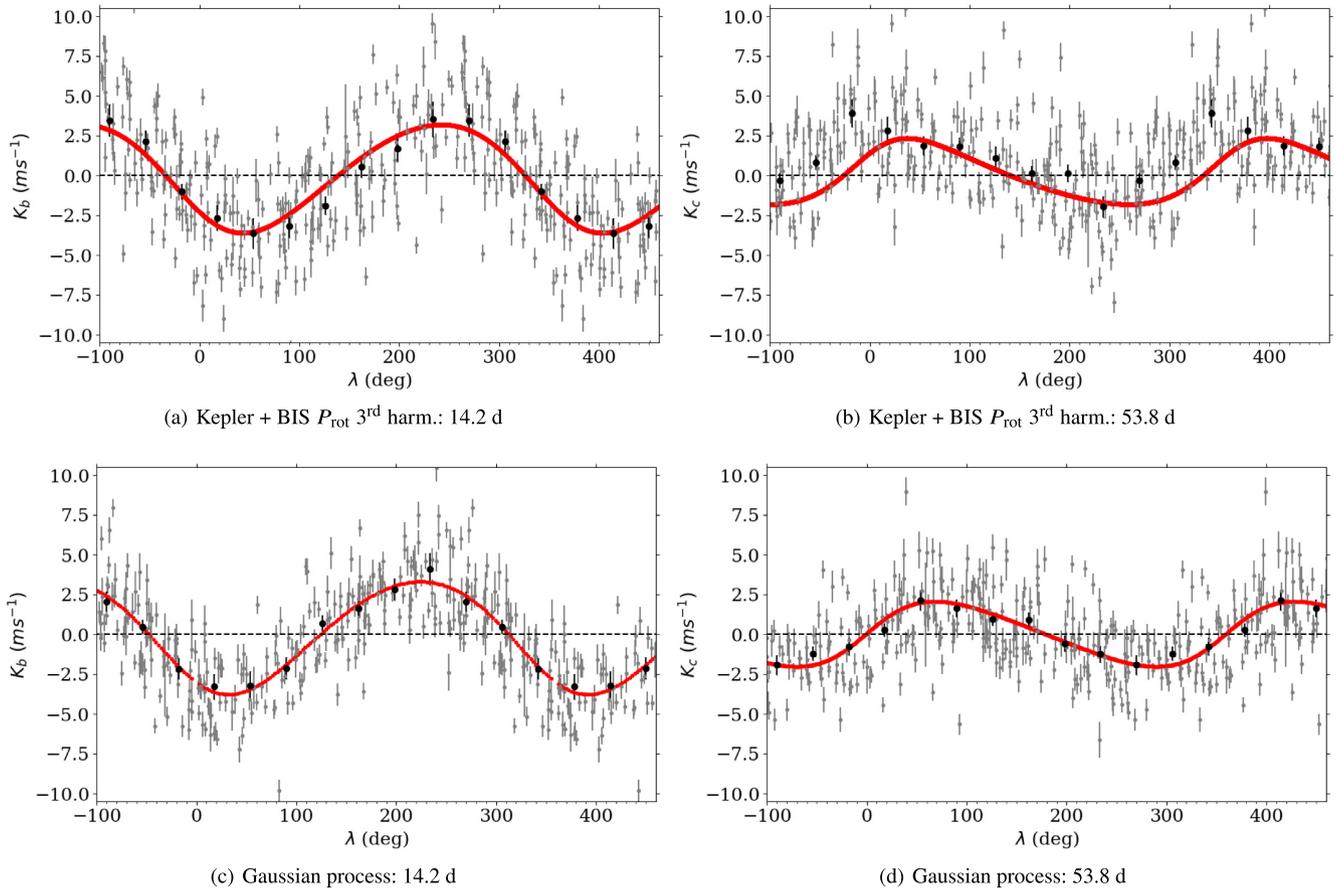
It is worth noting that the GP model favours a marginally larger MAP semiamplitude  $K_b$ , and a marginally larger MAP semiamplitude  $K_c$  than the typical parametric models. However, the values favoured by the GP model are bracketed both above and below by other parametric models, and so are in no sense extreme; concern that a GP might ‘absorb’ some of a planetary signal is clearly unfounded in this case.<sup>7</sup>

Under the parametric activity models,  $K_c$  tends to *decrease* in the low-activity data subset compared to the high-activity subset, suggesting that the outer Keplerian (with period broadly similar to the likely stellar rotation period) is actually being used to absorb some RV variability the too-simplistic activity models cannot account for.  $K_b$  appears comparatively insensitive to stellar activity levels, which may reflect the fact that  $P_b$  is about 2.5 times shorter than the star’s putative rotation period, such that little constructive or destructive interference between the Keplerian and activity signal occurs.

Putting aside the fine-grained differences between the GP and parametric activity models, Table 10 shows striking and remarkable consistency between planet parameters across almost all the models we considered, with well-constrained periods, and semiamplitudes

<sup>6</sup>While a number of techniques for speeding up GP regression do exist, e.g. Foreman-Mackey et al. (2017), we are not aware of any that are well-suited to cases where covariances need to be formulated over multiple inputs and outputs simultaneously, e.g. RVs jointly with activity indicators.

<sup>7</sup>After all, the framework from R15 is designed so that the GP component of the model is only able to explain variability simultaneously present in RVs and activity-sensitive time series; therefore, except in extremely pathological cases, there should be little risk of the GP wrongly fitting a planetary signal.



**Figure 4.** Phase folded RVs of HD 13808 for the planets HD 13808b (left) and HD 13808c (right) i.e. the leftover RV when the other planetary signal, the stellar activity model and the polynomial offset is subtracted. The black dots are the averaged RV values computed over bins of 36 deg in mean longitude and the error bars correspond to the rms of the mean, while the grey points are the observed RV values with their error. The red line represents the calculated curve of the planet with the MAP planetary parameter values of each model.

inconsistent with zero at  $>10\sigma$  levels. On this basis, we regard our modelling to represent the secure RV detection of two planets orbiting HD 13808. By way of contrast, it is worth noting that even transiting planets with tightly constrained periods and orbital phases may sometimes prove difficult to characterize, with inferred semi-amplitudes diverging wildly depending on which model is used: Kepler-10c is an excellent example (Fressin et al. 2011; Dumusque et al. 2014; Weiss et al. 2016; Rajpaul, Buchhave & Aigrain 2017). We refer, hereafter, to the planets we have detected as HD 13808b and HD 13808c with periods of  $\sim 14$  and  $\sim 54$  d, respectively.

#### 6.4 Stellar activity characteristics

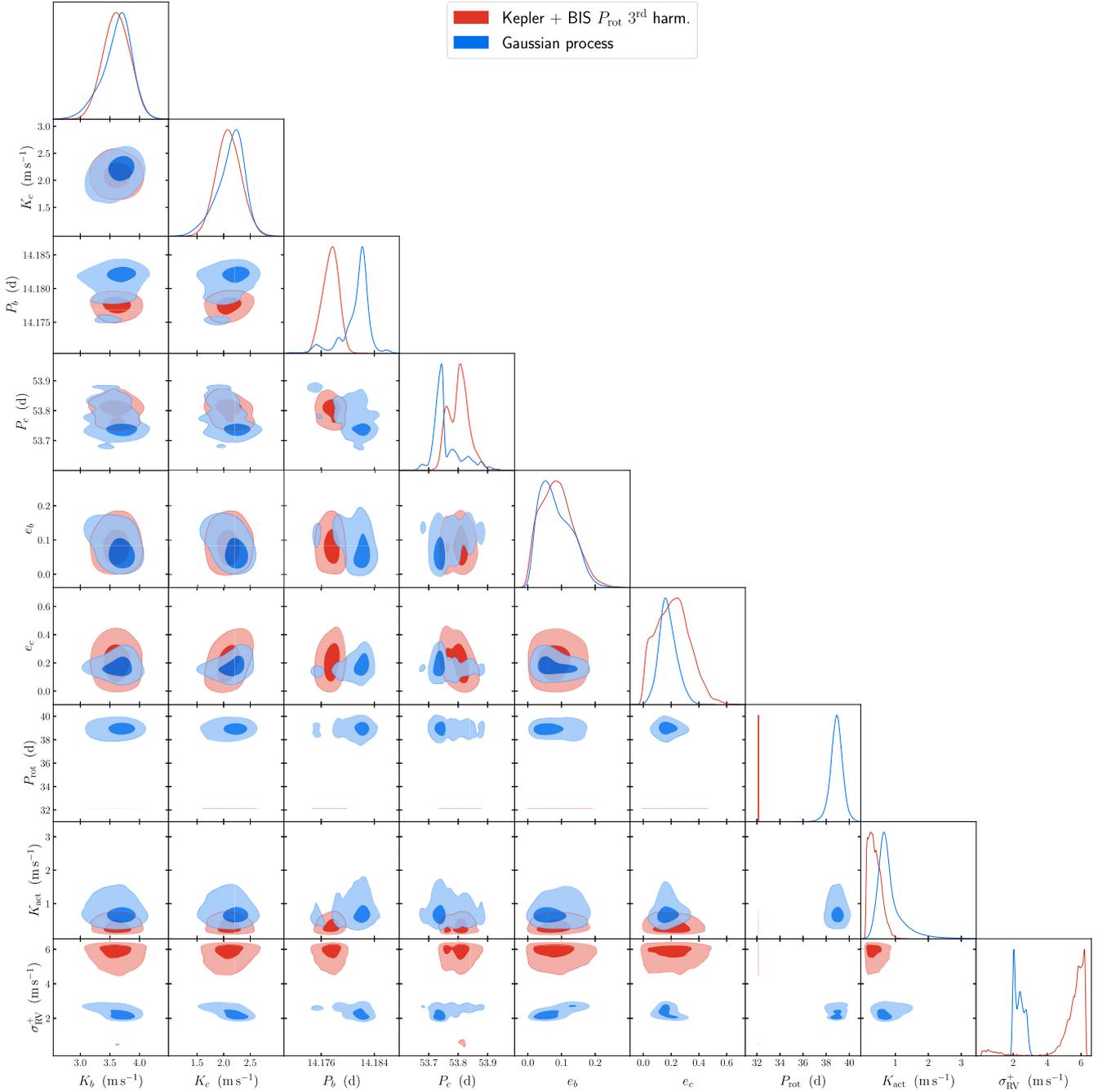
The stellar rotation period of  $38.99 \pm 0.47$  d inferred under the GP model is broadly bracketed by the  $\sim 32$ – $41$  d rotation periods inferred by our various parametric models (Table 11), and agrees well with the  $\sim 40$  d period estimated by Lovis et al. (2011).

The combined RV semi-amplitude ascribed to activity under the GP model (taking into account both the convective and rotational terms in the model, i.e.  $V_c$  and  $V_r$ ) is a little under  $1 \text{ m s}^{-1}$ ; meanwhile, the fitted RV jitter term under the GP model is of order  $2 \text{ m s}^{-1}$ . By contrast, the ‘Kepler + BIS  $P_{\text{rot}}$  3rd harm.’ model ascribes only around  $30 \text{ cm s}^{-1}$  of RV variability to activity, while absorbing nearly  $6 \text{ m s}^{-1}$  of variability as white-noise jitter (see Fig. 5). These differences are not surprising, given that our quasi-periodic

GP function draws do not need to have constant amplitudes and phases, need not be strictly periodic, etc.: consequently, the GP can fit more activity-related variability than the more inflexible parametric model, which must instead absorb the stellar red noise via a large jitter term.

The GP model also allows us to constrain the evolution time-scale of active regions: the inferred value of  $\lambda_e = 91 \pm 10$  d (Table 12) suggests an active region lifetime (or spot evolution time scale) of about two rotation periods, which is significantly longer than is observed for active regions on the Sun (Bradshaw & Hartigan 2014). The hyperparameter  $\lambda_p$  is harder to interpret physically, though the value  $\lambda_p \sim 1$  suggests that the activity signals being modelled by the GP are only moderately more complex than a sinusoidal model, having on average something between three or four inflection points per period, compared with exactly two for a sine wave (Rajpaul 2017).

In addition to the stellar activity linked to the rotation period of the star, we are able to characterize the long-term magnetic cycle of HD 13808. The two models considering this cycle ‘Kepler + magn. cycle’ and ‘Kepler + BIS  $P_{\text{rot}}$  3rd harm. + magn. cycle’ estimate these variations to have a period of  $3682 \pm 14$  d and  $3681 \pm 19$  d with RV semi-amplitudes of  $1.61^{+0.26}_{-0.33}$  and  $1.77^{+0.37}_{-0.28} \text{ m s}^{-1}$ , respectively. The estimate for the RV semi-amplitude of the magnetic cycle in the first model may be affected by the lack of the short-period activity, though the MAP values for both period and semi-amplitude are consistent within their  $1\sigma$  credible intervals.

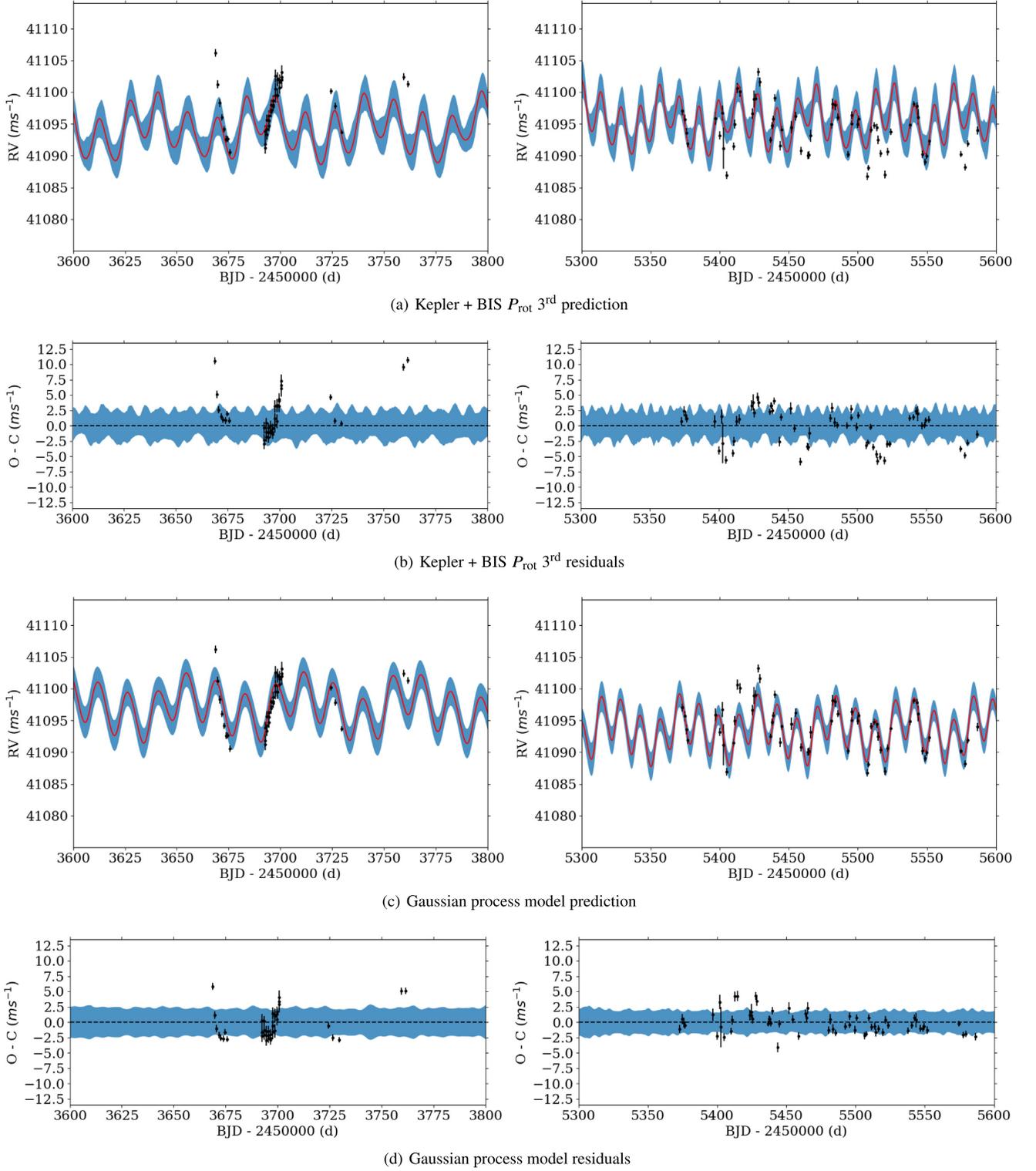


**Figure 5.** Corner plot of the Keplerian semi-amplitudes  $K_b$ ,  $K_c$ , periods  $P_b$ ,  $P_c$ , and eccentricities  $e_b$ ,  $e_c$  of HD 13808b and HD 13808c, respectively, under our two favoured models: the GP model (blue) and the Kepler + BIS  $P_{\text{rot}}$  3rd harm. model (red).  $K_{\text{act}}$  is the semi-amplitude of activity-dependent RV terms combined in quadrature, for the harmonic model or the GP model, and  $\sigma_{\text{RV}}^+$  is the additive noise (‘jitter’) parameter in either model. The dark and light filled regions correspond, respectively, to  $1\sigma$  (39.3 per cent) and  $2\sigma$  (86.5 per cent) joint credible regions.

## 6.5 Conclusion

In summary, our favoured activity model is the GP model; we might consider the ‘Kepler + BIS  $P_{\text{rot}}$  3rd harm.’ (or arguably the  $FF$ ) model to be the best of the parametric activity models, even though it falls short of the GP. Both the GP and the ‘Kepler + BIS  $P_{\text{rot}}$  3rd harm.’ models favour a two-planet solution over a one-planet solution; they both strongly reject solutions containing more than two Keplerians, supporting our non-planetary interpretation for any ‘extra’ Keplerians. In addition, they both show amongst the lowest

residual RV scatters (Table 8). However, the GP model is the only one to reject under all circumstances solutions containing more than two Keplerians, and to reliably detect the second planetary signal even when considering the high-activity subset of observations. Despite its other merits, the ‘Kepler high activity + BIS  $P_{\text{rot}}$  3rd harm.’ model has the unusual disadvantage of having locked on to what appears to be a spurious or at least unusually short stellar rotation period, unlike the GP model and some of the other parametric activity models (see Table 11).



**Figure 6.** Model prediction (red) with posterior predictive uncertainty (shaded blue) and their residuals for ‘Kepler + BIS  $P_{\text{rot}}$  3<sup>rd</sup> harm.’ and ‘Gaussian process’ models; RV measurements and corresponding uncertainties appear in black. Two representative subsets of observations are shown, one from the high-activity (left) and low-activity (right) phases of the star.

The MAP values and their corresponding planetary characteristics for our two favoured models are summarized in Table 13. The inferred minimum masses for HD 13808b and HD 13808c are  $\sim 11 M_{\oplus}$  and  $\sim 10 M_{\oplus}$  (model dependent), and their orbital semimajor axes around the host star are  $\sim 0.11$  AU and 0.26 au, respectively. As Neptune has

a mass of  $\sim 17 M_{\oplus}$ , this leads us to the conclusion that both planets are likely warm Neptunes.

The phase folded RVs for both planets and both models are displayed in Fig. 4, and the corner plot corresponding to the posteriors of both models is presented in Fig. 5. The comparison of the

posteriors of both favoured models shows that – despite taking very different approaches to modelling stellar activity – the planetary parameters do agree within their  $1\sigma$  credible intervals. It is also interesting to note how narrow (and likely wildly overoptimistic) the posterior distribution for the stellar rotation period inferred by the harmonic model is.

To demonstrate the goodness of fit, the RV predictions of our two favoured models in comparison to the data are shown in two parts of the data set in Fig. 6, one being in the high- and one in the low-activity subsets of observations.

## 7 DISCUSSION

We presented a comprehensive study of the HD 13808 RV data provided by the HARPS spectrograph. In particular, we tested and compared multiple approaches to stellar activity modelling.

We found that Bayesian evidence comparison is, by itself, not sufficient for deciding on the number of planets present: when stellar activity is not modelled adequately, extra ‘planets’ might be favoured to account for residual variability. Only our GP model favoured two planets (both of which we believe to be extremely secure detections) over a higher number of planets under all circumstances we considered, including modelling only a high-activity subset of observations; the parameters for the additional ‘planets’ suggested by other models were highly variable and model sensitive, and/or corresponded to likely activity variability (cf. periods in Table 9, particularly for the high-activity subset). However, some of our more complex stellar activity models such as the multiharmonic and  $FF'$  stellar activity models did seem to describe the activity variability better than more simplistic parametric models, based on both the decrease in residual RV rms scatter and the decrease in evidence for a three-planet solution versus the two-planet solution. In short, model comparison using Bayesian evidence is only as good as the models considered.

Reassuringly, when modelling the full set of observations, the RV signals of HD 13808b and HD 13808c were without exception detected at a  $>10\sigma$  level with *every* activity model we considered, with the planets’ characteristics remarkably consistent across all models. Taking all the observations into account, our two favoured models found the two planets HD 13808b and HD 13808c to orbit their host star with periods of  $\sim 14.2$  and  $53.8$  d at distances of  $\sim 0.11$  and  $0.26$  au, with minimum masses of  $11$  and  $10 M_{\oplus}$ , respectively. The planetary parameters inferred from the low-activity observations alone are also mostly within agreement with those derived from the full data set.

The influence of stellar activity on the planets’ characteristics was evident, however, when considering the parameters yielded by the models applied to the high-activity data subset where the  $\log R'_{\text{HK}}$  value of the star was  $> -4.90$ . For example, even under a single model (‘Kepler high activity + BIS  $P_{\text{rot}}$  3rd harm.’), the MAP RV semi-amplitude of HD 13808b decreases from  $3.71$  to  $2.19$   $\text{m s}^{-1}$  when moving from the low- to high-activity subset of observations, although these two semi-amplitudes are inconsistent at only a  $\sim 2\sigma$  level. Using the GP – the only model reliably to detect HD 13808c in the high-activity subset of observations – the MAP RV semi-amplitude of HD 13808c decreases from  $1.78$  to  $1.04$   $\text{m s}^{-1}$ , although in this case at least the semi-amplitudes do remain consistent within  $\sim 1\sigma$ .

To conclude, even though in our case the planetary parameter values were relatively insensitive to the choice of stellar activity model when analysing the full data set (246 measurements), caution would be essential on smaller data sets, or indeed when trying to characterize weaker planetary signals. In particular, the changes

in the inferred planet parameters when considering only the high-activity subset of our data underscores the importance of stellar activity modelling even for a host star nominally classified as ‘inactive’.

## ACKNOWLEDGEMENTS

The authors are grateful to the anonymous reviewer for constructive feedback that helped improve this manuscript, and to Xavier Dumusque for a valuable discussion. VMR acknowledges the Royal Astronomical Society and Emmanuel College for financial support, and the Cambridge Service for Data Driven Discovery (CSD3) – operated by the University of Cambridge Research Computing Service ([www.csd3.cam.ac.uk](http://www.csd3.cam.ac.uk)), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/P020259/1), and DiRAC funding from the Science and Technology Facilities Council ([www.dirac.ac.uk](http://www.dirac.ac.uk)) – for providing computing resources. AM acknowledges support from the senior Kavli Institute Fellowships.

Based on observations made with HARPS spectrograph on the 3.6-m ESO telescope at La Silla Observatory, Chile.

## DATA AVAILABILITY

The data underlying this article will be available via VizieR at CDS.

## REFERENCES

- Aigrain S., Pont F., Zucker S., 2012, *MNRAS*, 419, 3147  
 Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Mantelet G., Andrae R., 2018, *AJ*, 156, 58  
 Balan S. T., Lahav O., 2009, *MNRAS*, 395, 1936  
 Baranne A. et al., 1996, *A&AS*, 119, 373  
 Boisse I. et al., 2009, *A&A*, 495, 959  
 Boisse I., Bonfils X., Santos N. C., 2012, *A&A*, 545, A109  
 Bonfils X. et al., 2018, *A&A*, 613, A25  
 Bradshaw S. J., Hartigan P., 2014, *ApJ*, 795, 79  
 Chaplin W. J., Cegla H. M., Watson C. A., Davies G. R., Ball W. H., 2019, *AJ*, 157, 163  
 Collier Cameron A. et al., 2019, *MNRAS*, 487, 1082  
 Delgado Mena E. et al., 2019, *A&A*, 624, A78  
 Dumusque X. et al., 2011, *A&A*, 535, A55  
 Dumusque X., Udry S., Lovis C., Santos N. C., Monteiro M. J. P. F. G., 2011, *A&A*, 525, A140  
 Dumusque X. et al., 2012, *Nature*, 491, 207  
 Dumusque X. et al., 2014, *ApJ*, 789, 154  
 Faria J. et al., 2016, *A&A*, 589, A35  
 Faria J. P. et al., 2020, *A&A*, 635, A13  
 Feroz F., Hobson M. P., 2013, *MNRAS*, 437, 3540  
 Feroz F., Hobson M., Bridges M., 2009, *MNRAS*, 398, 1601  
 Feroz F., Balan S. T., Hobson M. P., 2011, *MNRAS*, 415, 3462  
 Fischer H., 2011, *A History of the Central Limit Theorem, Sources and Studies in the History of Mathematics and Physical Sciences*. Springer, Berlin  
 Foreman-Mackey D., Agol E., Angus R., Ambikasaran S., 2017, *AJ*, 154, 220  
 Fressin F. et al., 2011, *ApJS*, 197, 5  
 Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A., Rubin D. B., 2013, *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton, FL  
 Gillon M. et al., 2017, *A&A*, 601, A117  
 Gladman B., 1993, *Icarus*, 106, 247  
 Goodman S. N., 1999, *Ann. Int. Med.*, 130, 1005  
 Gray R. O., Corbally C. J., Garrison R. F., McFadden M. T., Bubar E. J., McGahee C. E., O’Donoghue A. A., Knox E. R., 2006, *AJ*, 132, 161  
 Grunblatt S. K., Howard A. W., Haywood R. D., 2015, *ApJ*, 808, 127

- Hall R., Thompson S., Handley W., Queloz D., 2018, *MNRAS*, 479, 2968
- Handley W., Hobson M., Lasenby A., 2015, *MNRAS*, 453, 4384
- Hatzes A. et al., 2010, *A&A*, 520, A93
- Haywood R. et al., 2014a, *Int. J. Astrobiol.*, 13, 155
- Haywood R. D. et al., 2014b, *MNRAS*, 443, 2517
- Henry T. J., Soderblom D. R., Donahue R. A., Baliunas S. L., 1996, *AJ*, 111, 439
- Higson E., Handley W., Hobson M., Lasenby A., 2018, *Bayesian Anal.*, 13, 873
- Jeffreys H., 1983, *Theory of Probability*, International Series of Monographs on Physics. Clarendon Press, Oxford
- Jurgenson C., Fischer D., McCracken T., Sawyer D., Szymkowiak A., Davis A., Muller G., Santoro F., 2016, *Proc. SPIE*, 9908, 99086
- Kharchenko N. V., 2001, *Kinematika i Fizika Nebesnykh Tel*, 17, 409
- Lomb N., 1976, *Ap&SS*, 39, 447
- Lovis C., Pepe F., 2007, *A&A*, 468, 1115
- Lovis C., Mayor M., Pepe F., Queloz D., Udry S., 2008, in Fischer D., Rasio F. A., Thorsett S. E., Wolszczan A., eds, *ASP Conf. Ser. Vol. 398, Extreme Solar Systems*. Astron. Soc. Pac., San Francisco, p. 455
- Lovis C. et al., 2011, preprint ([arXiv:1107.5325](https://arxiv.org/abs/1107.5325))
- Malavolta L. et al., 2017, *AJ*, 153, 224
- Mamajek E. E., Hillenbrand L. A., 2008, *ApJ*, 687, 1264
- Mayor M., Queloz D., 1995, *Nature*, 378, 355
- Mayor M. et al., 2003, *Messenger*, 114, 20
- Mayor M. et al., 2009a, *A&A*, 493, 639
- Mayor M. et al., 2009b, *A&A*, 507, 487
- Mayor M. et al., 2011, preprint ([arXiv:1109.2497](https://arxiv.org/abs/1109.2497))
- Mortier A., Faria J. P., Correia C. M., Santerne A., Santos N. C., 2015, *A&A*, 573, A101
- Nelson B. E. et al., 2020, *AJ*, 159, 73
- Ochsenbein F., Bauer P., Marcout J., 2000, *A&AS*, 143, 23
- Pasquini L. et al., 2008, *Proc. SPIE*, 7014, 70141I
- Pepe F., Mayor M., Galland F., Naef D., Queloz D., Santos N. C., Udry S., Burnet M., 2002, *A&A*, 388, 632
- Pepe F. et al., 2014, *Astron. Nachr.*, 335, 8
- Queloz D. et al., 2001, *A&A*, 379, 279
- Queloz D. et al., 2009, *A&A*, 506, 303
- Rajpaul V. M., 2017, PhD thesis, Univ. Oxford
- Rajpaul V., Aigrain S., Osborne M., Reece S., Roberts S., 2015, *MNRAS*, 452, 2269 (R15)
- Rajpaul V., Aigrain S., Roberts S., 2016, *MNRAS*, 456, L6
- Rajpaul V., Buchhave L. A., Aigrain S., 2017, *MNRAS*, 471, L125
- Rasmussen C. E., Williams C. K. I., 2006, *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA
- Roberts S., Osborne M., Ebdon M., Reece S., Gibson N., Aigrain S., 2013, *Phil. Trans. R. Soc. A: Math. Phys. Eng. Sci.*, 371, 20110550
- Santos N. C. et al., 2002, *A&A*, 392, 215
- Santos N. C. et al., 2013, *A&A*, 556, A150
- Santos N. C. et al., 2014, *A&A*, 566, A35
- Scargle J., 1982, *ApJ*, 236, 835
- Schwab C. et al., 2016, *Proc. SPIE*, 9908, 99087H
- Skilling J., 2006, *Bayesian Anal.*, 1, 833
- Skrutskie M. F. et al., 2006, *AJ*, 131, 1163
- Suárez Mascareño A. et al., 2020, *A&A*, 639, A77
- Thompson S. J. et al., 2016, *Proc. SPIE*, 9908, 99086
- Weiss L. M. et al., 2016, *ApJ*, 819, 83
- Zechmeister M., 2018, *A&A*, 619, A218
- Zechmeister M., Kürster M., 2009, *A&A*, 496, 577

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.