Planets around young active solar-type stars: assessing detection capabilities from a non-stabilized spectrograph

A. Heitzmann [1],[*] S. C. Marsden [1], P. Petit,[2] M. W. Mengel [1], D. Wright,[1] M. Clerte,[1] I. Millburn,[1] C. P. Folsom [3,4] B. C. Addison [1], R. A. Wittenmyer[1] and I. A. Waite[1]

[1]*Centre for Astrophysics, University of Southern Queensland, West Street, Toowoomba, QLD 4350, Australia*
[2]*Institut de Recherche en Astrophysique et Planétologie, Université de Toulouse, CNRS, CNES, 14 Avenue Edouard Belin, F-31400 Toulouse, France*
[3]*Department of Physics and Space Science, Royal Military College of Canada, PO Box 17000 Station Forces, Kingston, ON K7K 0C6, Canada*
[4]*Tartu Observatory, University of Tartu, Observatooriumi 1, Tõravere, 61602 Tartumaa, Estonia*

ABSTRACT

Short-orbit gas giant planet formation/evolution mechanisms are still not well understood. One promising pathway to discriminate between mechanisms is to constrain the occurrence rate of these peculiar exoplanets at the earliest stage of the system's life. However, a major limitation when studying newly born stars is stellar activity. This cocktail of phenomena triggered by fast rotation, strong magnetic fields, and complex internal dynamics, especially present in very young stars, compromises our ability to detect exoplanets. In this paper, we investigated the limitations of such detections in the context of already acquired data solely using radial velocity data acquired with a non-stabilized spectrograph. We employed two strategies: Doppler Imaging and Gaussian Processes and could confidently detect hot Jupiters with a semi-amplitude of $100 \, \mathrm{m \, s^{-1}}$ buried in the stellar activity. We also showed the advantages of the Gaussian Process approach in this case. This study serves as a proof of concept to identify potential candidates for follow-up observations or even discover such planets in legacy data sets available to the community.

Key words: techniques: radial velocities – planets and satellites: detection – planets and satellites: formation – stars: activity – stars: individual: HD 141943 – stars: pre-main-sequence.

1 INTRODUCTION

In the quest to understand the processes governing planetary system development, the peculiar case of short-orbit gas giants (i.e. Jovian and sub-Jovian exoplanets, orbiting their star with periods of less than a few weeks that we will refer as hot Jupiters or HJs hereafter) is a real challenge as classical theories describing their formation and evolution do not predict their presence in the vicinity of their parent star. Although representing a significant fraction of all exoplanets discovered (between 10 and 15 per cent[1]), their true occurrence rate is estimated to be around 1 per cent for mature solar-type stars (Wright et al. 2012). Even though this discrepancy can be explained through observing biases, their scarcity raises the question of the formation channel generating this population.

In the most accepted explanation, future HJs form in the colder region of the protoplanetary disc (more than a few au) and later experience orbital decay to eventually reach a close-in orbit. Two migration mechanisms are proposed: gas disc migration (see Baruteau et al. 2014, for a review), where the planet migrates inwards as the result of angular momentum exchange between the gas giant and the disc, and high-eccentricity tidal migration. In this last scenario, the planet is sent to a highly eccentric orbit following a strong perturbation (planet–planet scattering, e.g. Chatterjee et al. 2008,

or secular interactions, see Beaugé & Nesvorný 2012; Petrovich 2015; Petrovich & Tremaine 2016; Hamers et al. 2017 for the different proposed mechanisms). Now being close enough to the star at periastron, tidal forces exerted by the star act to circularize the planet's orbit.

Confronting migration theories is *in situ* formation, where the HJ forms in the vicinity of the host star and remains in close orbit. This explanation has been historically rejected as it sets restrictive constraints on the inner stellar disc, i.e. there must be enough available material to form the core of these gas giants and that core forming process needs to be completed before the star depletes all the gas from the area for the future HJ to successfully accrete its gaseous envelope. Due to these constraints, it is unlikely to occur according to the *Solar nebula theory*, assuming a disc composition similar to the one that gave birth to our Solar system (Perryman 2011). Now realizing that our Solar system may be far from being the norm in the great diversity of planetary systems, *in situ* formation has come back under the spotlight (Batygin, Bodenheimer & Laughlin 2015; Boley, Contreras & Gladman 2015). Recent studies, such as Bailey & Batygin (2018) or Dawson & Johnson (2018), suggest that HJs could have a different origin in different systems and/or that a combination of the proposed mechanisms could be at play.

In their review paper, Dawson & Johnson (2018) propose to test the different theories by searching for correlations between properties of HJs and their parent stars. Among the 15 studied properties, two are flagged as requiring further observations: HJ obliquities and host star ages. This paper focuses on the latter.

Studying young stars is a privileged approach as it would help us to discriminate between early-stage mechanisms such as *in situ* formation or gas disc migration versus more prolonged and late-stage mechanisms, like high-eccentricity migration. However, Dawson & Johnson (2018) warn that high-eccentricity migration-driven HJs could arrive in close-in orbit fairly early in the system's formation, showing that the dependence of these mechanisms with stellar age is not yet completely clear. Therefore, young stars (<20–50 Myr) and, even more so, younger (<10 Myr) low-mass (<3 $M_{\odot}$) T Tauri pre-main-sequence (PMS) stars as defined in Appenzeller & Mundt (1989) are probably the best candidates.

Unfortunately, with the exception of direct imaging surveys, the youngest stars are typically systematically avoided when searching for exoplanets, as they exhibit particularly strong intrinsic variability, or stellar activity. For such stars, this activity-induced correlated noise results primarily from surface brightness features, linked to complex internal processes and a strong magnetic field. Surface features yield spurious radial velocity (RV) signatures that generally completely mask exoplanet signatures, hence preventing their discovery. Additionally, Nava et al. (2020) showed that activity can generate unexpected spurious peaks in a periodogram analysis, which could lead to false positives if no adequate treatment of the activity is applied.

Filtering or mitigating this stellar activity becomes crucial if one hopes to find traces of exoplanets orbiting young active stars. It is also important to note that effective activity mitigating strategies are key in the search of Earth-sized planets around less active stars. In those cases, both the activity level and the planetary signature are up to two orders of magnitude smaller, but present a similar situation in relative terms. However, it is still slightly different as additional phenomena are also at play (i.e. granulation and pulsations). The exoplanet community is actively trying to develop and assess these strategies (see Cabot et al. 2021).

Available data on planets with periods less than 15 d orbiting very young stars (<50 Myr) are very scarce. Six planet-hosting stars have been found from transits (David et al. 2016, 2019; Newton et al. 2019; Bouma et al. 2020; Plavchan et al. 2020; Rizzuto et al. 2020) and three from RV searches: CI Tau b (Johns-Krull et al. 2016; Flagg et al. 2019), V830 Tau b (Donati et al. 2015, 2016, 2017), and TAP 26 b (Yu et al. 2017b). Recently, however, the existence of both V830 Tau b and CI Tau b has been challenged by Damasso et al. (2020) and Donati et al. (2020). V830 Tau b and TAP 26 b were found by the MaTYSSE (Magnetic Topologies of Young Stars and the Survival of massive close-in Exoplanets) observation programme in a sample of 33 weak-line T Tauri stars (Yu 2017a). If real, these two planets would indicate a fraction of HJs as high as 6 per cent for newly born stars. In this context, it is crucial to carry on the search for close-in gas giants around young stars to better estimate their occurrence rate at that stage.

In this paper, we investigated the case of searches for short-period gas giants orbiting very young and active stars solely using RV data. More specifically, we injected various RV signatures mimicking single circular planet systems behind real data of the young active G dwarf HD 141943 (not known to host a massive planetary companion) and assessed our detection limits using two distinct strategies: Doppler Imaging (DI) activity filtering (Section 3.1) and Gaussian Process (GP) Regression (Section 3.2).

Although already used in the past (DI + GP in Donati et al. 2016, 2017; Yu et al. 2017b, 2019; Klein et al. 2020 and GP in most exoplanet searches for the past few years), testing the respective performance of these two methods in legacy data sets has not been performed. We note that the underlying data were not optimized to search for exoplanets and were obtained using a non-stabilized spectrograph (e.g. with ≈50–100 m s^{-1} uncertainty on radial velocities). The limitations we describe should therefore be significantly improved with RV stabilized data sets. However, they provide a strong baseline for what is achievable and are useful to investigate other data sets of this nature already available [i.e. in the Bcool (Marsden et al. 2014) or TOUPIES (Folsom et al. 2016, 2018) surveys]. We compared our results to the planet 'hide and seek' study done on the same star with no specific treatment for stellar activity (Jeffers et al. 2014).

This paper is organized as follows: Details on the techniques used to reduce the data, more specifically to get from raw spectra to radial velocities, are given in Section 2. We then cover the methods addressing stellar variability in Section 3. Section 4 of this paper focuses on our reanalysis of HD 141943's raw data set. Section 5 explains how we set up our simulated data sets, and results from the analysis are laid out in Section 6. Finally, we give our conclusions and future prospects in Sections 7 and 8.

2 DATA ANALYSIS

2.1 From spectra to line profiles

Both methods we utilized to disentangle stellar activity from planetary signals take RV time series as input. The extraction of RV values from raw stellar spectra was performed by finding the centroid (described in Section 2.2) of a 'mean line profile' obtained using least-squares deconvolution (LSD; Donati et al. 1997; Kochukhov, Makaganiuk & Piskunov 2010). LSD convolves an observed stellar spectrum with a spectral line mask. Given an appropriate mask, the result is an enhanced peak signal-to-noise ratio (S/N) 'mean line profile' exhibiting stellar activity-induced line features. We chose the stellar mask best matching our star in the list of masks designed in the scope of the Bcool survey (Marsden et al. 2014) using VALD (Kupka 2000) for a star with an effective temperature of $T_{\rm eff}$ = 6000 K, a surface gravity of log g = 4.5 cm s^{-2}, and [Fe/H] = +0.2. Only spectral lines deeper than 20 per cent of the maximum line depth were kept for the LSD computation, yielding a total of 4097 lines. The outcome was an S/N increase from ≈50–100 for the observed spectra (depending on the spectrum and spectral order considered) to ≈1000 for the LSD mean line profiles.

2.2 From line profiles to radial velocities

Classically, each RV is taken to be the mean of a Gaussian profile fitted to the obtained line profile. However, for active stars, the distortion and here the 'flat bottom' (see the centre plot on Fig. 1) of the line show that a Gaussian fit is not suitable. We considered two alternatives.

First, we chose a generalized normal distribution (GND; Nadarajah 2005), as shown in green on the central plot of Fig. 1 and described by the following p.d.f:

$$\text{GND}(x) = \frac{\beta}{2\sigma\,\Gamma\left(\frac{1}{\beta}\right)} \exp\left(-\left|\frac{x-\mu}{\sigma}\right|^{\beta}\right), \tag{1}$$

where Γ denotes the gamma function, μ the position parameter (mean), σ the scale parameter, and β the shape parameter. $\beta <$ 2 results in wings more extended than a normal distribution and a sharper distribution peak. When $\beta = 2$, the GND becomes a Gaussian distribution (where σ is the standard deviation). For $\beta > 2$, the distribution yields wings less extended than a normal distribution
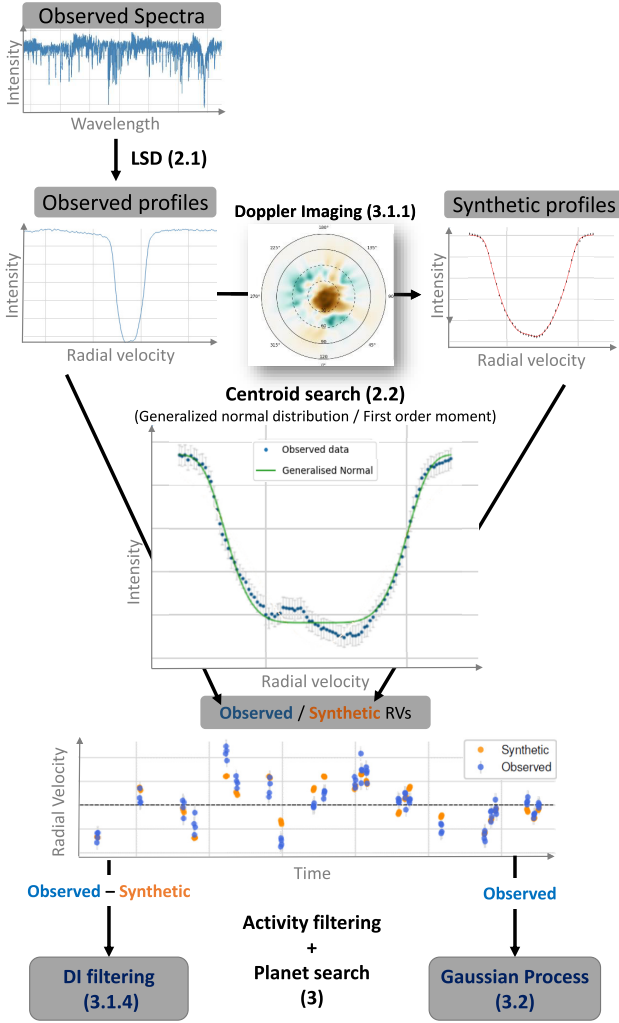
Figure 1. Diagram of the data analysis procedure, from raw spectra to periodic signature identification. Each block is a step of the process. Bold text and the associated numbers in parenthesis, respectively, indicate the method used to progress from one block to the next and the section of this paper detailing the corresponding process. Each point on the bottom plot results from the analysis of a single spectrum using the entire procedure described here.

and tends to a uniform distribution as $\beta \to \infty$. This grants more flexibility to the distribution, resulting in a better fit to broadened profiles. Error bars on the GND parameters are given by the fitting method. The centroid μ and the associated error bars for each LSD profile constituted our RV time series.

Secondly, we derived RV values using the first-order moment (generalized centroid, FOM hereafter) of each LSD profile, computed as

$$\mathrm{RV} = \frac{\int (I_c - I(\nu)) \, \nu \mathrm{d}\nu}{\int (I_c - I(\nu)) \, \mathrm{d}\nu}, \tag{2}$$

with $I(\nu)$ the intensity of the profile at radial velocity ν and I_c the continuum level. Here, error bars were propagated using the LSD-derived uncertainties. We note that results given by FOM are sensitive to integration limits (i.e. the limits on the line profiles used to compute it). This is further described in Section 5.3.

3 STELLAR ACTIVITY FILTERING

Stellar activity distorts line profiles, causing a shift in the line's centroid and therefore in the measured RV. Modelling the activity thus aims to correct for these distortion-induced shifts.

3.1 Method #1: filtering activity using DI

Section 3.1.1 describes DI, representing the core of our filtering process, following Donati et al. (2014). Sections 3.1.2 and 3.1.3 describe magnetic imaging (Zeeman DI, ZDI) and differential rotation, complementary to the DI technique. The actual filtering process is described in Section 3.1.4.

3.1.1 Doppler imaging

DI is a tomographic technique that, for rapidly rotating stars ($v \sin i \gtrsim 10 \, \mathrm{km \, s^{-1}}$), uses spectroscopic observations to infer the brightness features at their surface (Brown et al. 1991; Donati & Brown 1997). Practically, a time series of observed pseudo-line profiles obtained through LSD is iteratively adjusted using a tomographic algorithm. Irregularities in the profiles are interpreted as surface bright/dark spots that enhance/block Doppler-shifted light due to stellar rotation. Then, iteratively, synthetic profiles, derived from the DI surface maps, are fitted to the observed ones. To reach a unique solution to the ill-posed problem of DI inversion (as a single line profile can be generated from different surface map solutions), a maximum entropy selection of the solution is adopted (i.e. minimizing the information content of the brightness map), while ensuring that the χ^2 is kept below a defined threshold. This is done following the routine of Skilling & Bryan (1984) and using the entropy as defined in Hobson & Lasenby (1998). Further details can be found in appendix B of Folsom et al. (2016). The model output is constituted of a synthetic set of LSD profiles, and of the brightness surface map producing this spectral information.

Synthetic line profiles are obtained by integrating the Doppler-shifted flux (due to the rotation of the star) emerging from each point of the visible hemisphere. This flux is scaled according to the local surface cell projected area, brightness, and limb darkening. The local line profiles are calculated using a Voigt profile, a convolution of a Gaussian, and a Lorentzian profile.

Output products of DI include a set of synthetic profiles and a surface brightness map (or a magnetic map for ZDI; see the next section). The use of DI also enables us to constrain the stellar fundamental parameters by selecting the parameter values that optimize the brightness model (i.e. inclination of the stellar rotational axis with respect to the line of sight i, line-of-sight projected equatorial rotation velocity $v \sin i$, stellar equatorial rotation period P_{eq}, stellar mean radial velocity $\overline{RV}$, and differential rotation $\mathrm{d}\Omega$) and line profile parameters (i.e. line depth and Gaussian and Lorentzian equivalent widths). The DI analysis of HD 141943 is described in Section 4.2 and Fig. 3.

3.1.2 Zeeman Doppler imaging

Although ZDI is not part of the filtering process, it is similar to the stellar mapping process and is therefore described here.

Similarly to DI, ZDI (e.g. Semel 1989) is a technique that uses polarimetric information (i.e. Stokes V LSD profiles) to reconstruct the magnetic field structure at the surface of the star. Here, we used a spherical harmonic expansion to describe the large-scale components of the magnetic field (i.e. poloidal and toroidal; Donati et al. 2006).
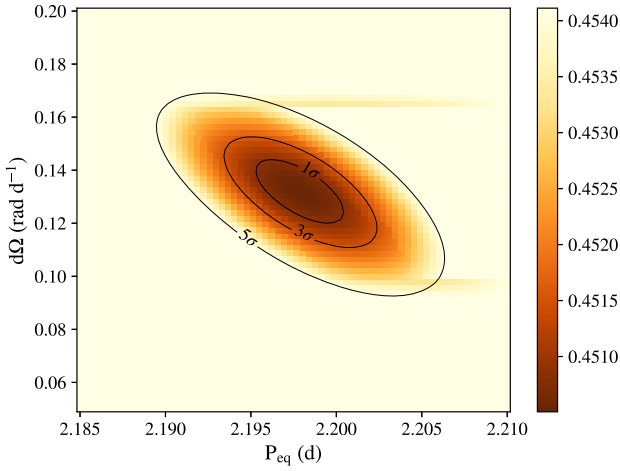
Figure 2. Reduced χ^2 surface of the differential rotation (rad d^{-1}) (*y*-axis) versus equatorial rotation period (*x*-axis), equivalent to the rotation frequency $\Omega_{eq} = 2\pi/P_{eq}$. Contours show confidence levels at 1σ, 3σ, and 5σ. Colour bar shows the reduced χ^2 values.

The Zeeman effect allows one to infer the strength and direction of the surface magnetic field, provided one has high enough S/N line profiles, rendered possible by the LSD technique. Like DI, solving for a magnetic field configuration is an ill-posed problem and ZDI also relies on maximum entropy image reconstruction. The ZDI analysis of HD 141943 is described in Section 4.2 and Fig. 4.

3.1.3 Surface differential rotation

The information used to generate a snapshot of the stellar surface through DI and ZDI often spans multiple stellar rotation cycles. Thus, the effect of differential rotation needs to be accounted for. The code we used models that differential rotation as a simplified solar-like differential rotation law:

$$\Omega(\theta) = \Omega_{eq} - d\Omega \sin^2 \theta, \qquad (3)$$

with $\Omega(\theta)$ the rotation rate at latitude θ, Ω_{eq} ($= \frac{2\pi}{P_{eq}}$) the rotation rate at the equator, and dΩ the difference in rotation rate between the equator and the poles (i.e. the differential rotation). Following Petit, Donati & Collier Cameron (2002) and Donati, Collier Cameron & Petit (2003), we explored the dΩ and Ω_{eq} parameter space, by running DI inversions for various values of the two parameters, looking for the doublet that optimizes the DI model, i.e. the dΩ and Ω_{eq} values that minimize the χ^2 of our model at fixed entropy level. The resulting χ^2 surface is used to derive our uncertainty on these two parameters.

We performed our DI, ZDI, and differential rotation analyses using the PYTHON ZDIPY code (see appendix from Folsom et al. 2018, for a more detailed description of the code). The code has been adapted to run on Fawkes, the High Performance Computing (HPC) facility at the University of Southern Queensland. The HPC allowed us to quickly explore our parameter space. Practically, we varied the stellar parameters (up to 3 at a time) to find the best solution. By best solution we mean the set of parameters that fit our line profiles down to the target χ^2 (<1 due to the LSD process; see Cang et al. 2020, for a similar case and follow-up explanations) and also maximize the entropy value. The differential rotation analysis of HD 141943 is described in Section 4.2 and Fig. 2.

3.1.4 Filtering the activity

Following Donati et al. (2014), we removed the stellar activity contribution by subtracting the RV time series derived from our modelling of the activity alone (i.e. the synthetic profile centroids obtained from DI) from the values obtained from the raw/observed LSD profiles (i.e. the observed/raw LSD profile centroids). We assumed that for very active stars, stellar variability is in a first approximation entirely due to features present on the stellar surface. We then searched for periodicity in the resulting filtered RVs, utilizing a Lomb–Scargle (LS) periodogram (Lomb 1976; Scargle 1982). We point out that residuals exhibit some red noise leftovers, while LS periodograms are designed for uncorrelated/white noise (VanderPlas 2018). We keep this approach here for maximal consistency with previous papers of Donati et al. (2014, 2016) and Yu et al. (2017b, 2019).

The nature of stellar variability (i.e. correlated/red noise), combined with the imperfect filtering (see Fig. 5) of the activity using DI, results in residuals exhibiting some red noise leftovers. As LS periodograms are designed for uncorrelated/white noise (VanderPlas 2018), this approach is limited and should not be used alone to claim a planet detection. To assess significance of a detection, we use the false alarm probability (FAP).[2] To compute the FAP levels, we used the Baluev approximation (see Baluev 2008). We also tried a bootstrap approach, which yielded very similar results.

3.2 Method #2: modelling the activity using a GP regression

Our second approach uses a GP regression to model the stellar activity-induced RV and its temporal evolution as first suggested in Haywood et al. (2014) and Rajpaul et al. (2015). The GP regression treats stellar activity as Gaussian red (correlated) noise. This Bayesian approach is driven by the data points, considered to be random correlated Gaussian variables and the covariance matrix $\mathbf{C}$, specifying the correlation between each pair of data points. Following Haywood et al. (2014), we computed each entry $\mathbf{C}_{ij}$ of this covariance matrix using the following physically driven quasi-periodic kernel made of a sinusoidal component to account for the rotation of the star combined with an exponential component for the surface feature appearance/decay:

$$\mathbf{C}_{ij} = \theta_1^2 \cdot \exp\left[-\frac{(t_i - t_j)^2}{\theta_2^2} - \frac{\sin^2\left(\frac{\pi(t_i - t_j)}{\theta_3}\right)}{\theta_4^2} \right] + (\sigma_i^2 + \sigma_s^2)\delta_{ij}, \qquad (4)$$

where the four hyper-parameters can be interpreted as follows:

(i) θ_1 (km s^{-1}): Semi-amplitude of the activity RV signature.
(ii) θ_2 (d): Decay parameter, or typical surface feature lifetime.
(iii) θ_3 (d): Recurrence time-scale, expected to be very close to P_{eq}.
(iv) θ_4 [0:1]: Smoothing parameter or amount of high frequency in the signal. Smaller and larger values of θ_4, respectively, indicate variations on longer and shorter time-scales. From experience (Jeffers & Keller 2009; Haywood et al. 2018), light curves and RV time series exhibit values of around 0.3 to 0.4 for this parameter. We chose a uniform prior that guarantees to largely encompass these values.

---

[2]The FAP limit indicates the likelihood that a peak caused by random fluctuations in the data would reach a given height/power (see dashed lines in Figs 5, B2, and C1). However, it does not indicate the probability of a data set to have a periodic component given the data.

σ_i is the uncertainty of data point i and σ_s is an extra white, uncorrelated noise parameter accounting for variations due to other sources and not explicitly captured by the model. σ_i and σ_s were added in quadrature and only applied to the diagonal of our matrix (i.e. variance of the data points).

Our global model is the sum of the GP model accounting for the stellar activity (RV_{GP}), a sinusoid for the circular planetary signature (RV_{pla}), and a constant offset (RV_0):

$$RV_{tot} = RV_0 + RV_{GP}(t, \theta_1, \theta_2, \theta_3, \theta_4, \sigma_i, \sigma_s)$$
$$+ RV_{pla}(t, K, P_{orb}, \Phi). \quad (5)$$

We ended up with a parameter space to explore containing 5 (θ_1, θ_2, θ_3, θ_4, and σ_s) + $3 \times n$ parameters + RV_0, for n planets (i.e. 9 parameters for single planet model). Then, two aspects need to be considered in order to confidently claim the presence of a periodic signal in the data. The first part is parameter estimation, where we explore the parameter space yielding posterior distributions from which the most likely set of parameters, as well as their mean and uncertainty values, can be recovered. The second aspect is model selection, where we assess how much more likely a model containing one planet is, compared to one with only stellar activity. Commonly, parameter space exploration is performed using Monte Carlo approaches. Despite the efficiency of some algorithms (e.g. EMCEE from Foreman-Mackey et al. 2013), the bottleneck of planet searches is usually model selection.

Model selection is performed by comparing the *marginal* likelihood (or evidence, $\mathcal{Z}$) of different models (i.e. activity only, activity with one planet, two planets, etc.). A detailed description of the evidence is given in Appendix A. Accurate estimation of this evidence is computationally expensive as it implies multidimensional integration over potentially large parameter spaces. Recently, Nelson et al. (2020) compared different methods for computing the evidence applied by different research groups. Although this was preliminary and would require follow-up studies to completely generalize the results, some approaches proved to be more consistent than others. Following their results, we developed our GP code using PYMULTINEST (Buchner et al. 2014), a PYTHON implementation of MULTINEST (Feroz, Hobson & Bridges 2009). This Importance Nested Sampling algorithm estimates the evidence and provides, as a by-product, the posterior probabilities and can therefore also be used for parameter estimation.

For the rest of this paper, when comparing models, we will refer to the Bayes factor (BF) and/or the associated probability (p) in favour of a single planet model (model $\mathcal{M}_1$) over an activity-only model (model $\mathcal{M}_0$):

$$BF = \frac{\mathcal{Z}_1}{\mathcal{Z}_0}, \quad (6)$$

with $\mathcal{Z}_0$ and $\mathcal{Z}_1$ the marginal likelihood for $\mathcal{M}_0$ and $\mathcal{M}_1$, respectively. We used the metric of Jeffreys (1961) (see Table A1) to assess significance from the BF.

3.2.1 *Likelihood and priors*

Two ingredients are needed to recover the posterior probabilities: likelihoods and prior probabilities.

In our case, the natural logarithm of the likelihood [i.e. probability of the data given the model and its parameters, $p(\boldsymbol{y}|\theta, \mathcal{M}_i)$ or $\mathcal{L}$] is given by

$$2 \ln \mathcal{L} = -n \ln(2\pi) - \ln(|\mathbf{C}|) - \boldsymbol{y}^T (\mathbf{C})^{-1} \boldsymbol{y}, \quad (7)$$

Table 1. Prior distributions of parameters used for the GP regression. The right column gives the prior for each parameter of the model. J(*min*, *max*) stands for Jeffrey's priors, MJ(*max*, *knee*) for Modified Jeffery's priors, $\mathcal{N}$(*mean*, *std*) for Gaussian priors, and $\mathcal{U}$[*min*, *max*] for Uniform priors. $\overline{\sigma}_{RV}$ is the mean of the RV uncertainties. RV_{max} is the largest absolute value in the data set and RV_{std} is the standard deviation of all RV values.

Parameters	Priors
Stellar activity	
θ_1 (km s^{-1})	MJ($1.5 \times RV_{max}$, $\overline{\sigma}_{RV}$)
θ_2 (d)	J(1, 100)
θ_3 (d)	$\mathcal{N}$(2.2, 0.05)
θ_4 [0:1]	$\mathcal{U}$[0:1]
Planet	
K (km s^{-1})	MJ($2 \times RV_{max}$, $\overline{\sigma}_{RV}$)
P_{orb} (d)	J(0.1, 15)
Φ [0:1]	$\mathcal{U}$[0:1]
Telescope and noise	
RV_0 (km s^{-1})	$\mathcal{U}$[$-RV_{max}$:RV_{max}]
σ_s (km s^{-1})	MJ(RV_{std}, $\overline{\sigma}_{RV}$)

with $\boldsymbol{y}$ the vector (of length n) containing the residuals after having removed both RV_{pla} and RV_0 from the original RVs and $\mathbf{C}$ the co-variance matrix computed using our GP kernel from equation (4).

Our priors, physically motivated following Gregory (2007), are listed in Table 1. Because the evidence is dependent on prior probabilities, we emphasize the importance of favouring uninformative priors, such as uniform, Jeffrey's (uniform prior in logarithmic space; see Gregory 2007) or Modified Jeffrey's (Jeffrey's prior, approaching a uniform distribution for values $\ll$ to the *knee* parameter of the modified Jeffrey's prior to handle priors that have 0 as a lower boundary; also see Gregory 2007), or at least priors independent of the studied data when previous and statistically valid knowledge is available in the literature. Using informative priors, without justification, would act to artificially boost the evidence. This is especially true for parameters that are not shared by the compared models (the planetary parameters in our case). The only informative prior we used here is θ_3 as P_{eq} has been constrained from DI.

We ran PYMULTINEST with an efficiency of 0.3 and 2000 live points (see Nelson et al. 2020). For each run, the parameter search drew between $\approx$50 000 samples from the posterior for the model with no activity and $\approx$150 000 for the single-planet model. Details of the results for all data sets are in Table 4.

4 ANALYSIS OF HD 141943

Before attempting to recover injected planets behind HD141943's activity, we analysed the raw observations (data set #5 containing no planet) to recover stellar parameters and ensure the star does not host any planet *that we can detect*.

4.1 Spectropolarimetric data set

Spectroscopic Stokes I (intensity) and V (polarized) observations of HD 141943 used in this study were acquired using the SEMPOL instrument, visitor polarimeter operating together with the University College London Echelle Spectrograph (Donati et al. 2003) and mounted on the 3.9 m Anglo-Australian Telescope (AAT) in Siding

Table 2. Fundamental parameters of HD 141943.

Parameter	HD 141943
Spectral type	G2V
Distance (pc)	60.028 ± 0.083^d
Age (Myr)	$17–32^b$
$M_\star$ ($M_\odot$)	1.3^a
Photospheric temperature T_{eff} (K)	5850 ± 100^a
Spot temperature (K)	$\approx 3950^a$
R_{eq} ($R_\odot$)	$1.5^{+0.06}_{-0.05}$ ^{c}
i (°)	70 ± 10^a
$v \sin i$ (km s^{-1})	35.6 ± 0.7 ^{e}
Equatorial rotation period P_{eq} (d)	2.198 ± 0.002^e
$d\Omega$ (rad d^{-1})	$0.1331^{+0.0095}_{-0.0094}$ ^{e}

^{a}M11A.
^{b}Hillenbrand et al. (2008).
^{c}*Gaia* DR2: Gaia Collaboration (2016, 2018).
^{d}*Gaia* EDR3: Gaia Collaboration (2016, 2021).
^{e}This study.

Spring, Australia. Available data comprise 92 spectra spread over 11 d between 2007 March 30 and April 9, covering 4.68 stellar revolutions, offering a well-sampled rotational phase coverage as required for DI and ZDI (further details on the data can be found in Marsden et al. 2011a) and a suitable time-scale to search for HJs. The 92 spectra were taken in chunks of four 30 min consecutive exposures, each in different polarization states to perform ZDI. As each 2 h observing run represents a very short time frame compared to P_{eq} (stellar equatorial rotation period) and any simulated HJs' orbital period, this data set can be treated as containing 23 epochs rather than 92. A previous DI and ZDI analysis of this data set is available in Marsden et al. (2011a, b) (M11A/B hereafter). The reduction of raw spectra was done using the ESPRIT pipeline (Echelle SPectra Reduction: an Interactive Tool; Donati et al. 1997).

4.2 Stellar parameters and surface mapping

HD 141943 is a young ($\approx 17–32$ Myr, M11A and Hillenbrand et al. 2008), nearby [60 ± 0.08 pc, estimated using Bailer-Jones et al. (2020) with the *Gaia* EDR3 data (Gaia Collaboration 2016, 2021)] active G PMS star. This Sun-like star has a mass of 1.3 M$_\odot$ and a radius of $1.5^{+0.06}_{-0.05}$ R$_\odot$ (*Gaia* DR2; Gaia Collaboration 2018). Soummer et al. (2014) also identified a surrounding near-edge-on debris disc, consistent with a planetesimal belt populated by two dust components at respective grain temperatures of 60 and 202 K. The extended list of stellar parameters can be found in Table 2.

We inferred stellar parameters by analysing the raw HD 141943 data set, containing no injected planet. These are marked with the superscript 'd' in Table 2: $v \sin i = 35.6 \pm 0.7$ km s^{-1}, $i = 43 \pm 10°$, $P_{\text{eq}} = 2.198 \pm 0.002$ d, and $d\Omega = 0.1331^{+0.0095}_{-0.0094}$ rad d^{-1}.

These parameters are close although not exactly matching the previous analysis from M11A/B (see the first two lines from Table 3). This discrepancy could be explained by the fact that the DI/ZDI code used between M11A/B is slightly different than ours. Mainly, ZDIPY lets us map bright and dark surface features (spots) in contrast with only dark spots in M11A/B. The inclination is the parameter with the largest difference (43° versus 70°), and also the hardest to constrain. To further investigate, we derived the best solution when fitting both (i) only for dark spots and (ii) using both dark and bright spots but forcing i to match M11A/B's value (i.e. 70°). Obtained Doppler maps and best parameters for the three cases (dark + bright, only dark and

Table 3. Set of parameters resulting from four different analysis: (i) bright and dark feature mapping, (ii) from M11A, (iii) bright and dark feature mapping with an inclination angle constrained to 70° matching M11A's value, and (iv) only dark features.

Best value	$v \sin i$ (km s^{-1})	P_{eq} (d)	$d\Omega$ (rad d^{-1})	i (°)
This work	35.6	2.198	0.13	43
M11A/B	35	2.182	0.36	70
Fixed i	35.6	2.197	0.12	70 (fixed)
Dark only	35.4	2.214	0.02	42

dark + bright with imposed 70° inclination) are given in Fig. 3 and Table 3, respectively.

These three cases yielded similar results, however, noting the negligible differential rotation when fitting only the dark features. Forcing i to 70° did not change the overall solution, and we found good agreement between the dark + bright non-forced and forced analyses. The contrast difference on the Doppler maps as seen on the bottom-left map of Fig. 3 is due to the effect of projection imposed by i. Spot locations are consistent across all maps and with M11A. The difficulty to constrain the inclination angle prevents a reliable deduction of the stellar radius R_{eq} and we therefore used the *Gaia* DR2's value given in Table 2. Our main objective for this paper was to filter out as efficiently as possible any rotationally modulated signal in RVs. Since setting i to 43° optimizes this task, we adopted this value for the inclination in the rest of this study.

Fig. 4 shows the radial (top), azimuthal (middle), and meridional (bottom) magnetic field distribution, derived with ZDI using Stokes V LSD profiles. We find a magnetic field with 52 and 48 per cent distribution for the poloidal and toroidal components, respectively, well agreeing with the 47 and 52 per cent from M11A. The mean strength is ≈ 52 G, much lower than M11A's value of 91 G. This can be explained again by the difference in inclination angle. Indeed, reapplying ZDI with a forced $i = 70°$ yields a field strength of 85 G, better agreeing with M11A.

4.3 Planet search

Before injecting planets in the HD 141943 data set, we ensured it did not exhibit any sign of hosting a planet.

The top panel of Fig. 5 shows the periodogram of the raw RVs, where we identified P_{eq} and its harmonics, the strongest signature being present at $P_{\text{eq}}/2$. Secondly, third and fourth panels are periodograms of the filtered RVs, respectively, from dark and bright, dark and bright with imposed $i = 70°$ features, and only dark analysis. All show similar features but one peak (around 2.7 d) did show different heights across analysis, and was above the 0.001 FAP threshold for the dark spot-only analysis. However, it did not reach overwhelming significance. This data set did not allow us to assess the impact of the varying DI solutions (dark, dark + bright, and dark + bright with forced inclination) on the planet retrieval as it has no injected planet. To test that, we performed a second analysis using these three configurations for data set 22 (see Section 5 for details on simulated data sets), containing a simulated planet in the 'uncertain' range of detection. We found that the different DI solutions did not change our conclusions regarding the planet search (detailed analysis available in Appendix C).

The GP confirmed that we were not able to detect any significant planet in the raw data set. We found a BF in favour of the single planet model over the activity-only model of only 0.3 ($p \approx 0.23$).
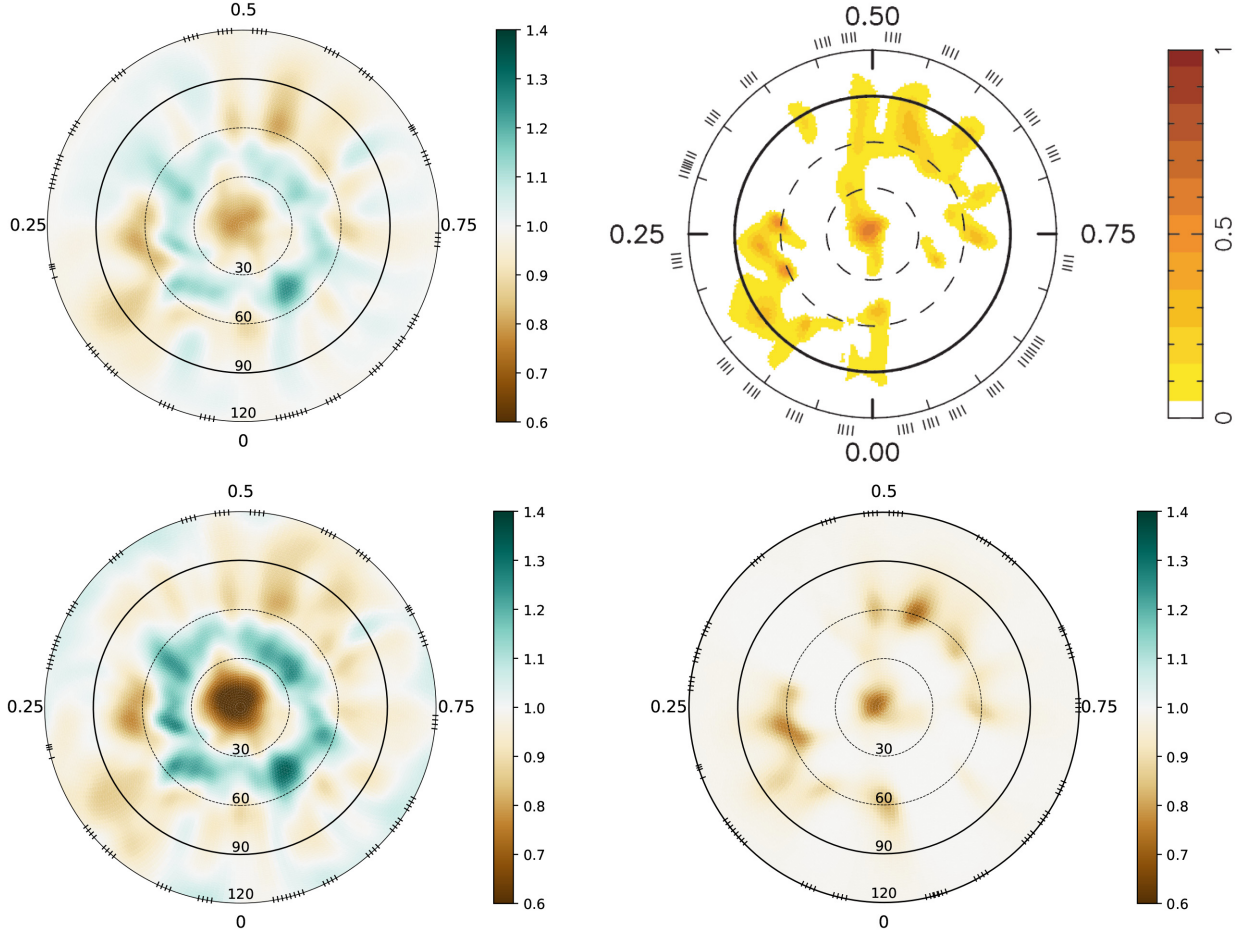
Figure 3. Comparison of the surface brightness maps for HD 141943's original data set (#5). Each map is a maximum entropy reconstructed image of the brightness features at the surface of the star. The blue and brown patches indicate regions that are warmer or colder than the photosphere, respectively. The maps are polar projected, with the centre being the visible rotation pole and the full ring labelled 90 the equator. Values on the outermost ring give the rotational phase and the ticks indicate the phase of each observation. Top left: Map obtained using the parameters from the first line in Table 3 and mapping for bright and dark features. Top right: Map extracted from M11A (second line of Table 3) and mapping only dark features. In their approach, the colour scale expresses the spot filling factor. Bottom left: Map obtained using the parameters from the third line of Table 3, with a forced inclination parameter of 70° and mapping for bright and dark features. Bottom right: Map obtained using the parameters from the fourth line of Table 3 and mapping only dark features.

5 SIMULATED EXOPLANET DATA SETS

We created 37 data sets, each containing a single planet on a circular orbit around HD 141943, following a procedure described in Sections 5.1 and 5.2. Each planet was incorporated into the raw spectra studied in the previous section.

5.1 Injected planets

Injected planets were chosen to be massive short-period exoplanets, with masses ranging from 0.38 up to $5.9 M_J$ and periods shorter than 6 d. We set the orbits to be circular, as it is believed to be the case for most HJs, especially for orbits shorter than 3 d (Dawson & Johnson 2018). We should none the less bear in mind that eccentricity can be a crucial aspect in favour of high-eccentricity migration and should not be overlooked especially when attempting to detect the slightly cooler warm Jupiters ($P_{orb} > 10$ d). The RV shift induced by each planet was defined as

$$RV_{pla}(t) = K \sin \left[2\pi \left(\frac{t}{P_{orb}} - \Phi + 0.5 \right) \right],$$ (8)

with K the semi-amplitude of the signal, P_{orb} the planet's orbital period, and Φ the phase. $\Phi \in [0: 1]$ and was defined such that when $\Phi = 0$, the planet crosses the plane containing the line of sight. We set $\Phi = 0$ to match the mid-point of the observations ($BJD_{\star mid} = 2454195.153776$). For the rest of this paper, we will refer to semi-amplitude values (K) for the planets rather than mass. The equivalence between K and mass is described in Section 6.2.5.

5.2 Complete data sets

To build our data sets, we generated an RV time series using equation (8) at times matching our observing epochs and then shifted each spectrum accordingly in wavelength space. In order to explore our planetary parameter space (made of K, P_{orb}, and Φ) without being overwhelmed with the number of data sets to analyse ($n_K \times n_{P_{orb}} \times n_\Phi$), we used the following strategy:

First, we created seven data sets (#1 to #8, excluding #5, the original one) at a fixed period (3.653 d) with K ranging from 50 to 500 m s^{-1} and random Φ. This initial analysis provided an estimate of the limiting semi-amplitude range for detectability.

Radial Field

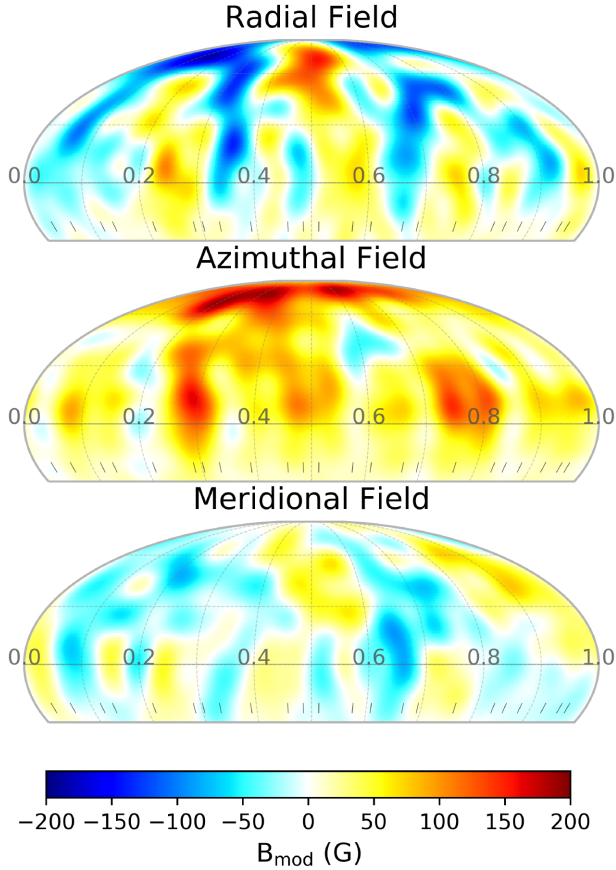## Azimuthal Field

Meridional Field

B_{mod} (G)

Figure 4. Maximum entropy image reconstruction of the radial (top), azimuthal (middle), and meridional (bottom) components of the magnetic field for HD 141943. Positive/negative field modulus values (in Gauss) are displayed in yellow/red and blue, respectively. The horizontal line shows the equator with the number describing the phase. Ticks translate each measurement's epoch.

Then, we generated additional data sets 4–5 at a time, filling areas in the parameter space that seemed relevant, i.e. around the noise limit, around P_{eq} and harmonics and to cover empty areas of the parameter space. After generating each batch of data sets, these were randomly assigned a mock value before analysis to avoid biases.

We ended up with 37 data sets (38 when including the original data set) spanning the following ranges: 42–532 m s^{-1} in K, 0.288–5.69 d for P_{orb}, and 0.02–0.99 in Φ. See Table 4 for specific details.

5.3 RV extraction

We tested the RV extraction from the line profiles using both a first-order moment approach and a GND fit. For the original data set (#5 with no injected planet) and given our limits on the integration for the FOM approach the average difference between FOM- and GND-extracted RVs (from the observed profiles) is 8 m s^{-1} with a maximum difference of 144 m s^{-1} for the most extreme point. The uncertainty on the RVs also differed as GND yielded uncertainties twice as large as those from FOM (148.5 ± 2.0 m s^{-1} versus 70.4 ± 1.8 m s^{-1}).

As previously mentioned, FOM-derived RVs and uncertainties are dependent on the number of points taken into account for its computation (i.e. chosen integration limits for the line profile) and where to cut in the wings of the line profile can be somewhat arbitrary.

On both sides of each profile (and for all of them), we cut at 43 km s^{-1} relative to the line centre. We tested different limits and chose the one that gave the least average difference with the GND approach, which is an analytical function and therefore not prone to this effect.

6 RESULTS

All data sets are extensively described in Table 4 and will be referred to using their number (#1, #2, ..., #38).

6.1 Stellar parameters

Stellar parameters inferred from the data sets containing injected planets are consistent with our refined parameters derived from the raw data set (#5; see Section 4.2). The mean values of the retrieved stellar parameters across all data sets, along with the largest deviation from the mean (given by the ±), are: $v \sin i = 35.6^{+0.27}_{-0.45}$ km s^{-1}, $i = 43^{+5}_{-3}$ °, $P_{eq} = 2.1995^{+0.007}_{-0.006}$ d, and $d\Omega = 0.119 \pm 0.08$ rad d^{-1}. In all cases, spot distributions are similar, with slight differences in terms of contrast. This can be explained by the fact that the fit to the line profiles was sometimes performed to a slightly different χ^2 level. Typically, the presence of a planet with a semi-amplitude significantly larger than the activity level (e.g. for data set #6) slightly impacts the performance of the DI. However, even such a large planet signature did not hamper the capacity of the DI to identify spot locations and recover the planet.

6.2 Planet detection: methods performance

6.2.1 Method 1: DI activity filtering

Results are shown in Fig. 6. Each marker represents a data set with its corresponding number in order to easily refer to Table 4 containing more details. Marker positions indicate the injected planets' K and P_{orb} (as we did not identify a systematic impact of Φ on retrievals, it was omitted for clarity). Each data set is identified by a specific marker/colour combination: A green circle when a periodogram peak was identified above an FAP threshold of 0.001 (0.1 per cent) and with a deviation from the true period of <10 per cent, an orange square when two peaks above an FAP of 0.001 and of similar height were found or when the right peak was identified but with a deviation from the true period of >10 per cent, preventing a safe conclusion, a grey cross when no signature above FAP = 0.001 could be identified, and finally a red cross when a peak was present above the FAP threshold but was not matching the injected planetary period, i.e. a false positive.

This approach yielded 16 positive detections, 6 inconclusive findings, 11 non-detections, and 4 false positives. These four data sets confirm that using FAP as a measure of significance is not the safest approach, as discussed in Section 3.1.4. Rigorous estimation of the significance was performed with the GP analysis.

All six simulated planets with semi-amplitudes larger than 150 m s^{-1} were well retrieved and half (8/18) between 100 and 150 m s^{-1}. This fraction increased to 60 per cent (8/13) when removing all data sets close to P_{eq} and its harmonics. Only 1/13 planets below 100 m s^{-1} were found (noting that #25 was a very weak detection with an FAP of 3.5×10^{-4}).

We note that all analyses that identified the right peak but with deviation of more than 5 per cent (up to 10.6 per cent for #21) from the true period (#19, #21, #25, and #35) had fewer than 2.5 orbital periods within our observation time span. This inaccuracy was due to the width of the peaks in the periodogram arising as the period
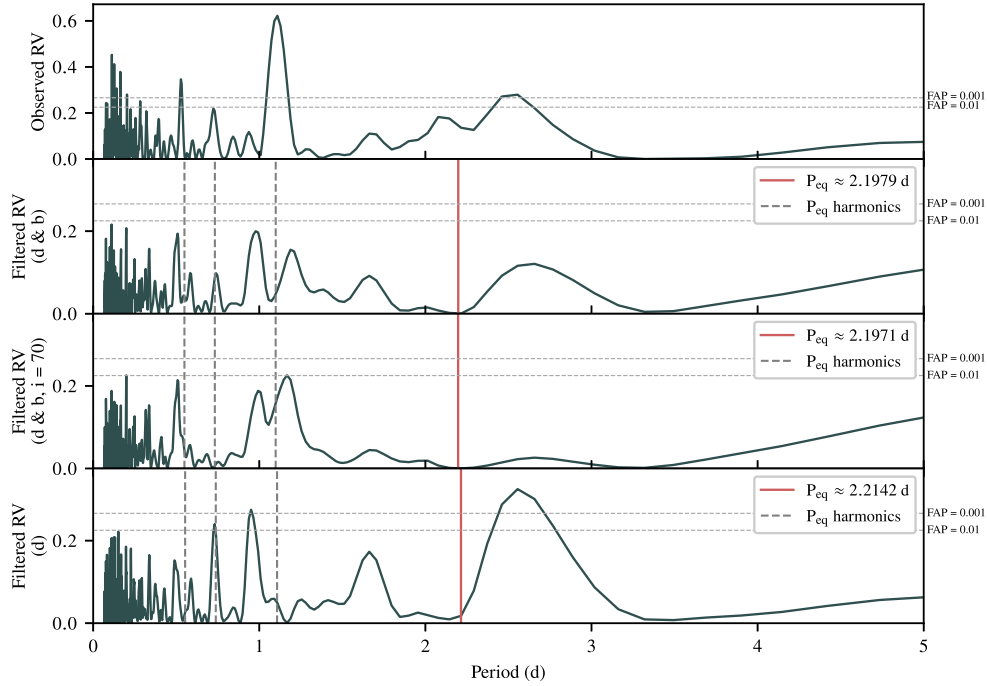
Figure 5. LS periodograms for the original data set (#5, containing no planet). First panel: Observed (raw) RVs. All other panels are for filtered RVs (i.e. raw RVs – synthetic RVs, where synthetic RVs are derived from the DI fitting). Second panel: Filtered RVs using the dark and bright (d and b) features for DI. Third panel: Filtered RVs using the dark and bright features for DI and with the 70° constraint on *i* (d and b, *i* = 70). Fourth panel: Filtered RVs using only dark features (d) for DI. We note that the peak around 2.6 d is likely caused by the rotation period at intermediate latitudes (offering a maximal visibility given the inclination angle of the star). This value is larger than the equatorial period depicted by the red, vertical lines, as expected in the case of a differentially rotating surface (see Section 3.1.3).

represented a significant fraction of the time span. It is safest, given our number of samples, to cover at least 2 to 3 orbital periods to achieve sufficient precision on P_{orb}.

The six 'to be confirmed' (orange squares) data sets #2, #15, #32, #34, and #36 all exhibited two competing peaks above the FAP threshold and at a similar height (FAP of 1.8×10^{-4} and 7.2×10^{-4} for #2, 1.6×10^{-4} and 1.4×10^{-4} for #15, 1.1×10^{-11} and 1.1×10^{-11} for #32, 1.25×10^{-6} and 2.6×10^{-4} for #34, and 1.15×10^{-10} and 2.43×10^{-9} for #36), preventing us from being able to choose the correct period. For #2, #15, and #32, the peaks are just above our detection threshold and it is therefore not surprising to find competing features. For #32 and #36, however, the competing peaks were both very significant. We are not sure as to how to interpret these, which vanished as we filter the signature associated with one of the two peaks. As shown in Fig. B2, these spurious peaks did not seem to correspond to any harmonics of neither the planet nor the star. See periodograms. Although complex interactions between the uneven data sampling and the periodic signatures cannot be ruled out, no significant peak could be identified in the window function (see Fig. B1). The complete analysis of these two data sets can be found in Appendix B. Data set #21 also falls in the 'orange square' category with a very wide identified peak, yielding a 10.6 per cent deviation between the retrieved and injected periods, slightly over our 10 per cent threshold.

False positives arose as the highest peak was not the simulated one, which would lead to false identification (if relying solely on DI) for #1 and #17. Regarding #9 and #24, the peaks were barely above the FAP of 0.001 and would not have led to a significant detection.

Although we could not identify a systematic impact, phase is expected to play a role in the injected planets retrieval and we can

see this occurring in the zoomed box in Fig. 6. The only noticeable difference between #13 and #15 is their phases (respectively, $\Phi_{13} = 0.4769$ and $\Phi_{15} = 0.1093$) and yet planet #13 is recovered but not #15.

Studying periodograms for all data sets indicated that planets with periods close to P_{eq} (#12 and #26), $P_{eq/2}$ (#33 and #34), $P_{eq/3}$ (#9 and #24), and $P_{eq/4}$ (#30 and #32) seem to be affected by the activity filtering. The case of #32 has been discussed above. This effect is to some extent expected as DI has the capacity to distort the line profiles, interpreting the rotationally modulated distortions as produced by spots on the brightness maps at harmonics of the rotation period and therefore is likely to absorb part of a planetary signature close to one of these periods.

For RV searches, the LS periodogram has limitations (choosing an FAP limit, interpreting the significance of a result, limitation to sinusoidal signals; see VanderPlas 2018) and we emphasize that dedicated treatment for stellar activity should be performed. We therefore advocate for incorporating a second, complementary method, presented in the following section, allowing both better quantification of the significance of a retrieved signature and a more comprehensive modelling of the activity.

6.2.2 Method 2: GP regression activity modelling

Again, results are detailed in Table 4 and summarized in Fig. 7. We defined successful retrievals (green circles) where the GP strongly favoured the single planet model over the activity-only model with a probability $p > 0.909$ (computed from the marginal likelihood/BF; see Appendix A and Table A1 for further details). We then have substantial evidence (orange

Table 4. Data sets 1 to 6. Each column represents a simulated data set. The first section (rows 1–4) gives the stellar parameters inferred from the DI analysis. The second section (rows 5–7) gives the values of the three parameters used to simulate the injected planet. The third section (rows 8–13) gives the results of the Method #1 DI filtering. RV_0 is an offset; then, we have the three recovered planet parameters, followed by the rms and χ^2 of the residuals. The fourth section (rows 14–22) gives the result of the GP (Method #2) for the no planet (activity-only) model. For the parameters (θ_1 to σ_s), the values are given as: mean ± std (maximum a posteriori). We then have the rms and χ^2 of the residuals and the resulting natural logarithm of the evidence. The last section (rows 23–36) gives the result of the GP for the single planet model. Again, the parameter (θ_1 to σ_s) values are given as: mean ± std (maximum a posteriori). We then have the rms and χ^2 of the residuals and the resulting natural logarithm of the evidence. Finally, we give the BF, defined as the ratio between $\mathcal{Z}$ from the single planet model (row 34) and the no planet model (row 22). The last row is the probability in favour of the single planet model associated with the BF value. Only the first six rows (data sets) are shown here; the full version is available as online material.

Data set	#1	#2	#3	#4	#5	#6	...
DI inferred stellar parameters							
$i(°)$	43	40.5	41.5	42	42.5	45	...
$v \sin i$ (km s^{-1})	35.47	35.53	35.75	35.789	35.643	35.862	...
P_{eq} (d)	2.206 15	2.203 45	2.1988	2.194 08	2.197 88	2.201 79	...
$d\Omega$ (rad d^{-1})	0.0769	0.097 96	0.133 33	0.146 94	0.133 33	0.123 08	...
Injected planet parameters							
K (m s^{-1})	60.6	82.9	154.9	267.1	–	532.2	...
P_{orb} (d)	3.6531	3.6531	3.6531	3.6531	–	3.6531	...
Φ [0:1]	0.271	0.303	0.445	0.311	–	0.432	...
Method #1 (DI) mean ± std							
RV_0 (m s^{-1})	13.1 ± 9	1.7 ± 9	0.5 ± 9	12.2 ± 10	–	−60.0 ± 10	...
K (m s^{-1})	85.5 ± 12	68.7 ± 11	115.6 ± 12	233.0 ± 13	–	430.2 ± 14	...
P_{orb} (d)	2.549 ± 0.054	1.413 ± 0.022	3.538 ± 0.067	3.649 ± 0.041	–	3.640 ± 0.022	...
Φ [0:1]	0.612 ± 0.026	0.209 ± 0.031	0.435 ± 0.018	0.301 ± 0.009	–	0.427 ± 0.006	...
rms (m s^{-1})	81.9	79.6	79.7	84.4	–	89.9	...
χ^2	1.34	1.28	1.28	1.43	–	1.63	...
Method #2 (GP)/no planet model/mean ± std (MAP)							
θ_1 (m s^{-1})	314.2 ± 113.4(217.4)	276.8 ± 89.9(208.2)	265.4 ± 68.0(217.1)	320.5 ± 74.1(277.5)	357.6 ± 109.1(296.6)	462.9 ± 98.0(394.1)	...
θ_2 (m s^{-1})	10.2 ± 6.3(6.9)	6.9 ± 3.8(6.0)	3.0 ± 1.4(4.2)	1.7 ± 0.5(1.7)	19.2 ± 9.3(15.7)	1.4 ± 0.3(1.1)	...
θ_3 (m s^{-1})	2.190 ± 0.036(2.148)	2.190 ± 0.041(2.141)	2.205 ± 0.045(2.164)	2.206 ± 0.049(2.255)	2.215 ± 0.020(2.216)	2.207 ± 0.049(2.268)	...
θ_4 (m s^{-1})	0.527 ± 0.148(0.399)	0.468 ± 0.136(0.347)	0.415 ± 0.137(0.234)	0.587 ± 0.146(0.511)	0.628 ± 0.083(0.597)	0.723 ± 0.151(0.605)	...
RV_0	12.5 ± 156.8(-20.5)	19.7 ± 127.6(16.6)	35.6 ± 97.7(52.1)	15.8 ± 113.3(-0.7)	25.3 ± 194.0(50.9)	17.4 ± 160.3(49.6)	...
σ_s	12.3 ± 9.5(0.4)	11.7 ± 9.1(2.8)	12.0 ± 9.2(9.8)	12.0 ± 9.2(3.0)	11.5 ± 8.8(0.6)	12.3 ± 9.5(0.7)	...
rms (m s^{-1})	55.9	54.9	54.8	55.8	59.0	56.6	...
χ^2	0.62	0.59	0.59	0.61	0.69	0.63	...
$\ln\mathcal{Z}$	−550.99	−552.15	−560.46	−563.85	−545.03	−571.93	...
Method #2 (GP)/single planet model/mean ± std (MAP)							
θ_1 (m s^{-1})	340.1 ± 117.7(271.1)	360.8 ± 120.8(257.2)	400.9 ± 151.9(301.2)	401.0 ± 138.3(347.7)	357.3 ± 107.9(292.2)	473.5 ± 199.5(338.4)	...
θ_2 (m s^{-1})	14.0 ± 7.8(13.3)	17.2 ± 10.0(17.1)	20.9 ± 13.2(15.3)	21.7 ± 12.3(19.0)	19.9 ± 9.7(15.9)	27.7 ± 16.7(22.3)	...
θ_3 (m s^{-1})	2.204 ± 0.030(2.217)	2.213 ± 0.026(2.233)	2.208 ± 0.024(2.216)	2.215 ± 0.019(2.220)	2.216 ± 0.020(2.228)	2.206 ± 0.020(2.213)	...
θ_4 (m s^{-1})	0.595 ± 0.151(0.602)	0.642 ± 0.154(0.601)	0.641 ± 0.163(0.570)	0.677 ± 0.143(0.645)	0.631 ± 0.081(0.592)	0.705 ± 0.156(0.680)	...
K (m s^{-1})	43.1 ± 23.6(59.8)	59.2 ± 19.7(69.1)	130.8 ± 19.1(138.8)	238.7 ± 14.2(235.6)	29.4 ± 32.1(39.7)	474.7 ± 14.5(473.4)	...
Period (d)	3.008 ± 2.386(3.331)	3.223 ± 1.341(3.420)	3.571 ± 0.438(3.598)	3.580 ± 0.102(3.603)	2.481 ± 2.528(0.172)	3.616 ± 0.035(3.610)	...
Phase [0:1]	0.345 ± 0.210(0.259)	0.305 ± 0.115(0.286)	0.444 ± 0.036(0.448)	0.306 ± 0.013(0.309)	0.571 ± 0.293(0.821)	0.432 ± 0.005(0.430)	...
RV_0 (m s^{-1})	38.6 ± 166.5(-1.6)	52.6 ± 178.1(119.5)	27.8 ± 212.0(81.8)	69.8 ± 211.2(1.2)	28.4 ± 190.2(166.2)	39.2 ± 292.2(121.2)	...
σ_s	11.0 ± 8.3(6.8)	11.1 ± 8.4(3.9)	12.4 ± 9.4(3.2)	11.8 ± 9.1(5.1)	10.7 ± 8.2(6.3)	12.7 ± 9.6(2.8)	...
rms (m s^{-1})	57.3	62.1	61.3	57.9	60.4	64.4	...
χ^2	0.65	0.77	0.75	0.66	0.72	0.82	...
$\ln\mathcal{Z}$	−551.06	−550.91	−556.77	−556.41	−546.25	−560.60	...
Bayes factor	0.9	3.5	40.0	1702.8	0.3	83283.0	...
$p(\mathcal{M}_1)$	0.48	0.78	0.98	1.00	0.23	1.00	...

squares) of a planet (i.e. $0.75 < p < 0.909$) and non-detections (grey crosses, i.e. $p < 0.75$). We note that most of the injected planets (28/37) were correctly identified by the GP, although not always significant enough to lead to a detection claim.

The GP yielded 16 positive detections, 4 'to be confirmed' findings (i.e. requiring further observations), 17 non-detections, and more importantly no false positives. Again here, all six simulated planets with semi-amplitudes larger than 150 m s^{-1} were correctly found. This drops to half (9/18) between 100 and 150 m s^{-1} (same ratio as DI although not systematically on the same data sets), and increases to 70 per cent (9/13) when removing all data sets close to P_{eq} and its harmonics. Finally, for planets below 100 m s^{-1}, only 1 out of 13 was found (along with two cases requiring further observations, with #2 a correctly identified planet and #27 a missed identification). The GP, compared to DI, is more conservative yet more reliable (i.e. no false positives) due to its accurate measure of the significance for each finding. Fig. 7 shows that, similar to the DI analysis, it is difficult to identify planetary signatures close to P_{eq} or its harmonics.
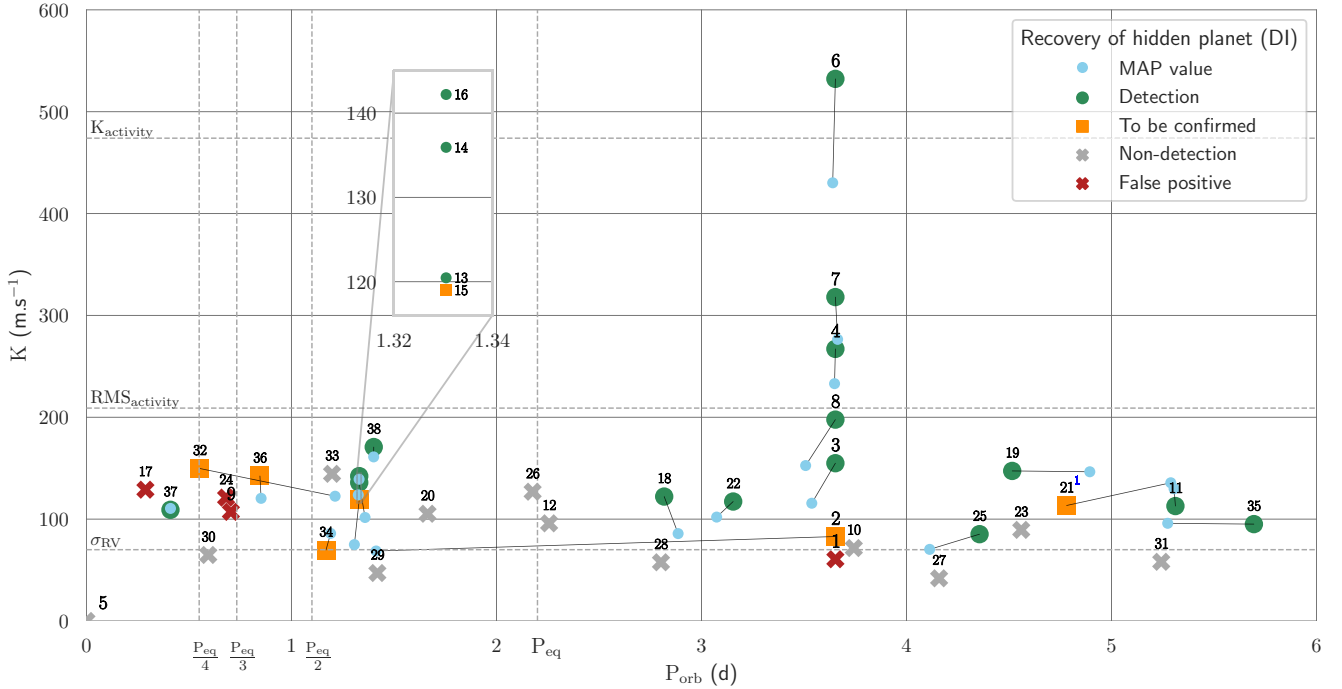
Figure 6. Results using the DI method. Each marker on the plot is a data set containing a simulated injected planet. Orbital period (d) is on the *x*-axis and semi-amplitude (m s^{-1}) is on the *y*-axis. Green circles: identification of the correct planet with a periodogram peak above an FAP of 0.001. Orange square: two peaks of similar height (see Section 6) were found preventing a safe conclusion or when the width of the correctly identified peak yielded a deviation of more than 10 per cent between the retrieved and injected orbital period. Grey crosses: no signature with FAP > 0.001 could be identified. Red crosses: the most significant peak was not matching the injected period peak. Horizontal dashed lines show the stellar activity semi-amplitude ($K_{activity}$) and the error bars on the retrieved RVs (σ_{RV}). Vertical dashed lines show the rotation period of the star and its harmonics. Blue points show, for the 'Detection' and 'To be confirmed' data sets, the P_{eq} and K values corresponding to the highest peak in the periodogram.[1]: For data set #21, the correct peak was found, at an FAP of 1.1×10^{-13}. However, that peak being very broad, it yielded a 10.6 per cent deviation between the retrieved and injected periods, slightly outside our 10 per cent limit.

We finally note that imprecision on the retrieved P_{orb} increases for longer periods (see MAP values indicated on Fig. 7). This is because fewer periods are covered by the data set as we move to the right of Fig. 7.

6.2.3 Consistency between methods

Utilizing two distinct methods serves as cross-validation when a signature is found. However, the GP is the only Bayesian approach, therefore the only one allowing a rigorous quantification of the evidence favouring of a particular model (i.e. presence of a sinusoidal signature in the data or not).

For signatures above 150 m s^{-1} and after removing data sets close to P_{eq} harmonics, both methods yielded systematic detections except for the ambiguity on #36 when using DI. For signatures between 100 and 150 m s^{-1}, the GP showed more consistency than the DI that exhibited three false positives. Out of the 13 data sets below 100 m s^{-1}, we ended up with 1 detection for both GP and DI.

Even though the Bayesian approach using a GP can (i) better handle correlated noise and (ii) more reliably estimate the significance of a detection, the use of the DI filtering method allows an independent validation.

6.2.4 Comparison with previous work

In a study performed in 2014, Jeffers et al. (2014) (J14 hereafter) injected various planets behind simulated activity signatures of

two young G and K stars. Varying parameters were the planet semi-amplitude, orbital period, and $v \sin i$ (shown to be correlated with the activity level). Stellar activity was generated based on DI maps and with different configurations (e.g. adding plages associated with spots, adding extra random spots, etc.; see J14 for more details). The G dwarf was HD 141943, thus making the comparison particularly relevant. Each simulated data set was composed of a single planet in a circular orbit, to which modelled stellar activity and instrumental signatures were added. In that study, the search for injected planets was performed *without* a specific treatment for stellar activity and was considered successful for periodogram peaks below FAP = 0.01 (versus 0.001 in our study).

With 50 observational epochs and for their less complex simulation of activity (only based on the DI maps), J14 were able to retrieve signatures of semi-amplitude $K = 110$ m s^{-1} when $v \sin i = 20$ km s^{-1} and $K = 525$ m s^{-1} when $v \sin i = 50$ km s^{-1}. Regarding their most complex simulation of activity (DI maps + plages + random spots; see J14 for further details), the minimum attainable planetary signature was $K = 525$ m s^{-1} when $v \sin i = 20$ km s^{-1}. In the case of $v \sin i = 50$ km s^{-1}, 200 observational epochs were required to reach the $K = 525$ m s^{-1} detection threshold.

We note that the data sampling is different, which might slightly hinder the comparison.[3] With 23 unevenly spread epochs, the

---

[3]J14 has one datum per night for 50, 100, or 200 consecutive nights, whereas we have 23 epochs over 10 nights.
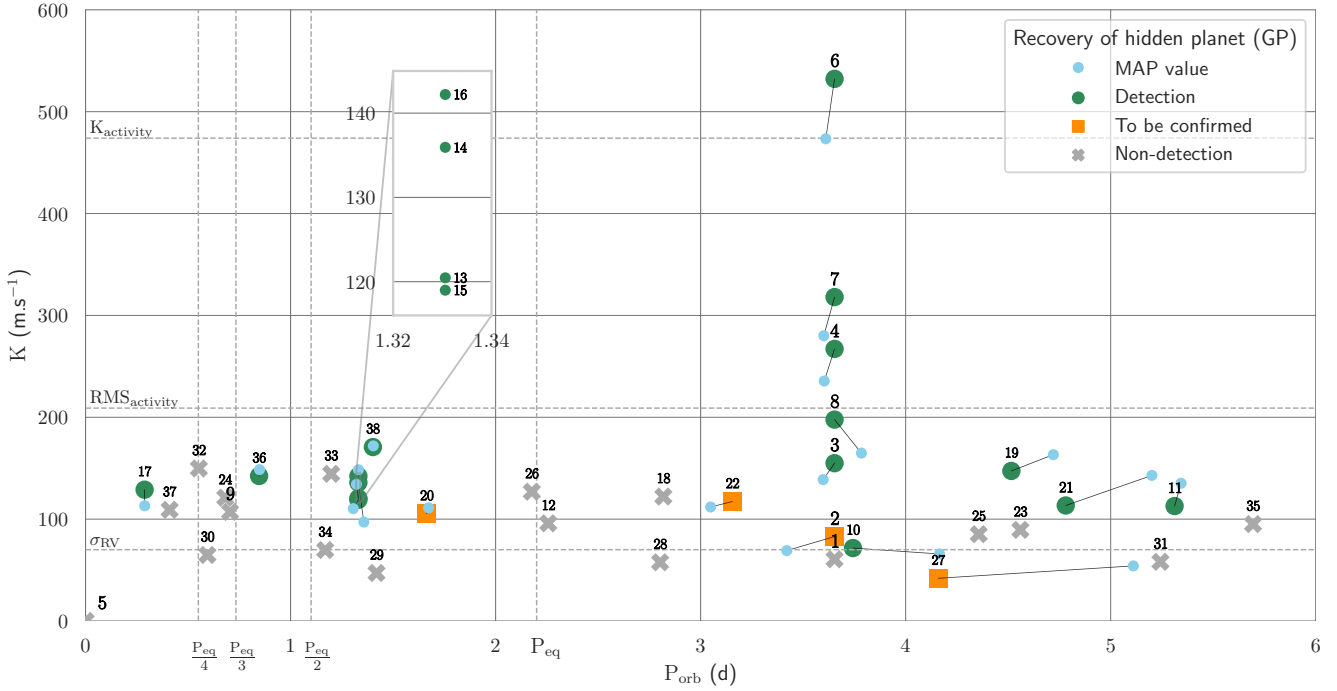
Figure 7. Results using the GP method. Each marker on the plot is a simulated injected planet. Orbital period (d) is on the *x*-axis and semi-amplitude (m s^{-1}) is on the *y*-axis. Green circles: the evidence between the single-planet model and with activity only is $\Delta \ln(\mathcal{Z}) > 10$, associated with at least a strong detection (probability $p > 0.909$). Orange squares: $10 > \Delta \ln(\mathcal{Z}) > 3$ or substantial evidence ($0.95 > p > 0.75$). Grey crosses: $\Delta \ln(\mathcal{Z}) < 3$ ($p < 0.75$). Horizontal dashed lines show the stellar activity semi-amplitude (K_{activity}) and the error bars on the retrieved RVs (σ_{RV}). Vertical dashed lines show the rotation period of the star and its harmonics. Blue points show the Maximum A Posteriori (MAP) values for P_{eq} and K, linked by a line to the corresponding data set.

smallest signature we could reliably detect was $K = 100$ m s^{-1} (down to 70 m s^{-1} for data set #10), emphasizing the benefit granted by our activity filtering approach. Although now systematically applied by the community for planetary searches around active stars, this emphasizes that an independent treatment of stellar activity combined with robust model selection is crucial to improve detection capabilities.

6.2.5 Recovered exoplanets

Here, we translate our results into planetary mass/orbital periods for the case of HD 141943, given $M_\star = 1.3$ M$_\odot$ and $i = 43°$. We consider that the stellar rotation axis is normal to the planet's orbital plan. In this context, our lower detection threshold of 100 m s^{-1} is equivalent to a planet with either

(i) $M_{\text{pla}} = 1$ M_{Jup}, $P_{\text{orb}} = 1.6$ d ($a = 0.03$ au) or
(ii) $M_{\text{pla}} = 2$ M_{Jup}, $P_{\text{orb}} = 12.5$ d ($a = 0.12$ au).

Using M11A's inclination value of $i = 70°$, we get

(i) $M_{\text{pla}} = 1$ M_{Jup}, $P_{\text{orb}} = 5$ d ($a = 0.062$ au);
(ii) $M_{\text{pla}} = 2$ M_{Jup}, $P_{\text{orb}} = 35$ d ($a = 0.23$ au).

In a case of a star of 1 M$_\odot$ with a transiting exoplanet, we can hope to detect a 1 M$_{\text{jup}}$ orbiting at up to 10 d, using typical DI non-stabilized observations. This is, of course, given similar conditions in terms of data quality and quantity, observing constraints, and stellar variability level. As more numerous and precise RVs should be easily obtainable, it is fair to expect better results and identify HJs around very young solar-type stars.

6.2.6 Dependence of planet detection to various parameters

In terms of semi-amplitude, our detection threshold of around 100 m s^{-1} corresponds to half of the stellar activity rms and a quarter of its semi-amplitude ($\approx$400–500 m s^{-1} looking at the maximum of the data points, or 357 ± 100 m s^{-1} according to the GP applied to data set #5). Given the scarcity of planets orbiting very young stars discovered solely using RV, comparisons with the literature are limited. When excluding searches in the small activity regimes (i.e. rms$_{\text{activity}}$ < 50 m s^{-1}) only two planets provide a direct comparison, V830 Tau b (Donati et al. 2017) and TAP 26 b (Yu et al. 2017b).

TAP 26 b is thought to have a semi-amplitude of 160 m s^{-1}, or $\frac{1}{8}$ to $\frac{1}{12}$ of the stellar variability semi-amplitude and V830 Tau b ($K \approx 60$ m s^{-1}) up to $\frac{1}{20}$. They both exhibit activity levels with a semi-amplitude of $\approx$1200 m s^{-1}. We believe that the difference in performance (detection threshold of $\approx\frac{1}{4}$ of the activity level for this work) can be explained by the fact that (i) both Yu et al. (2017b) and Donati et al. (2017) had more data ($\approx$30 and 60 epochs versus 23 for us), (ii) Donati et al. (2017) had slightly better uncertainties on the RVs [$\sigma_{\text{RV}} \approx 50$ m s^{-1} versus 75 for both Yu et al. (2017b) and this study] and most importantly, (iii) both had a longer baseline for the observations: 100 and 35 stellar rotation cycles versus 3 for us. We also used less constraining priors as previous knowledge was not available.

To ensure it was not our method implementation that hindered our capacity to find smaller signatures from our data sets, we ran our code on both Donati et al. (2016)'s and Yu & Donati (2017a)'s data and were able to retrieve the published periodic signatures. We note that with no access to previous knowledge, our prior distributions were less restrictive (i.e. non-Gaussian and/or broader

for the concerned parameters), decreasing the evidence and yielding slightly more conservative results. Our limitations can be seen as an upper boundary and that data of better quality and quantity would be able to detect smaller planets.

We find that detecting planets with orbital periods conflicting (i.e. within 0.1 d of) either P_{eq} or its harmonics was unreliable, as illustrated in Figs 6 and 7. Although longer period planets did not seem to be harder to detect, we noticed a significant loss of precision in the orbital period retrieved once we reach periods larger than 40 to 50 per cent of the observing time frame (see MAP values on Fig. 7). This is expected and good practice to sample a few orbital periods at least to get reasonable constraint. A good example of a similar study can be found in Klein et al. (2020), where the authors required 35/50 data points spread over 3 months ($\approx$15 orbital periods) to reliably detect 5/10 m s^{-1} planets behind stellar activity (about 2/3 times the planetary signature level). Finally, as we saw for data sets #32, #36 (see Appendix B for the complete analysis of these two data sets) and to a lesser extent for #2 and #15, spurious periodicity signatures can appear with no relation to harmonics and no obvious relation with the window function, as suggested by Nava et al. (2020).

Regarding orbital phase, the significant difference in peak height between data sets #13 (periodogram peak power = 0.3, $\Phi = 0.4769$) and #15 (periodogram peak power = 0.5, $\Phi = 0.9183$) given their identical period and comparable semi-amplitude suggests that phase impacts the detection capabilities. It is not surprising that particular phases would have an impact on the periodogram as the irregular sampling can yield different phase coverage. That being said, we did not observe a general trend with phase across all data sets.

Finally, data obviously play a huge role in the detection capabilities with crucial aspects being quality, quantity, and sampling. To better characterize the activity, i.e. improve the hyperparameters of the GP, it is important to optimize the sampling (spanning multiple stellar orbits with as dense and as regular sampling as possible). Another successful strategy is to apply a GP to simultaneous photometric data, or at least not too far apart so that there is not too much evolution of the stellar surface features. We also tested (see Appendix D) the improvement brought by the knowledge of the period of the orbiting planet (i.e. characterizing a known transiting planet).

7 CONCLUSIONS

In this paper, we assessed our capabilities to detect exoplanets behind real stellar activity signatures. We used a previously published set of observations gathered with a non-stabilized spectrograph of the young, active G star HD 141943 in which we injected simulated planets. We then utilized two distinct strategies, DI and a GP regression to filter out the stellar activity variability, aiming to recover these injected planets.

Our dedicated treatment of stellar activity allowed significant improvement in the detection capabilities compared to J14, a planet search study done on the same star with no dedicated activity mitigation. As previously shown by Yu & Donati (2017a), these strategies are among the best tools we have to deal with large activity signatures. Although now widely accepted in the community, we further confirmed that dedicated treatment of the activity is crucial, and showed that we can detect short-orbit gas giants even in non-optimally sampled data sets exhibiting a 50–100 m s^{-1} RV precision.

We tested two alternatives to recover the RVs from the LSD line profiles (GND fit and first-order moment), which yielded slightly different RV time series but more importantly different uncertainties. We favoured the FOM approach but other methods such as broadened profiles could be explored.

With a low number of epochs acquired with a non-stabilized spectrograph, the combined use of both GP and ZDI methods enabled us to set a planet detection threshold of around 100 m s^{-1} or $\approx \frac{1}{4}$ of the activity level. Injected planets under this threshold were either non-detections or would require extra observations to confirm. The limitations we faced give a good idea of the upper limit we can hope to achieve for such systems in similar conditions.

Although DI shows less reliability than the GP, it allows us to strengthen the confidence of a finding. This lack of reliability could be explained by the fact that DI does not take into account surface variability due to the active regions' appearance/disappearance. These can evolve quickly, as shown for another G-type star in Petit et al. (2004). We suggest that claiming planets around active star should be done with a dedicated treatment for stellar variability, and preferentially using a Bayesian framework for robustness and to allow a quantification of the evidence of the presence of an orbiting planet.

We attempted to identify some factors that could improve the likelihood to find exoplanets orbiting young stars. Larger and more precise data set is an obvious one. Efficient sampling is also crucial, where dense sampling of the stellar rotation to better constrain the activity should be combined with coverage of multiple planetary orbits.

Orbital periods close to P_{eq} and its harmonics pose serious difficulties, and often lead to non-detections. In our case, it also appears to be a good rule of thumb to sample at least 2, or even 3, orbital periods to constrain P_{orb} with sufficient precision.

Some data sets (#2, #15, #32, and #36) exhibited significant spurious peaks of mysterious origin that compete with the true planetary period, emphasizing the difficulty of RV-only searches.

Detecting young exoplanets that do not transit is difficult but essential if we want to expand the sample of massive short-period exoplanets orbiting very young stars and progress towards settling the long-lasting debate over their origin. This work demonstrates that we can realistically identify potential candidates for follow-up observations and even detect short-orbit gas giant planets in non-optimized data sets exhibiting large activity variability.

8 FUTURE WORK

As follow-up of this work, and to improve precision, producing mean line profiles with either classical approaches (i.e. CCF, shift, and fit) or more recent proposals (Rajpaul, Aigrain & Buchhave 2020 or Collier Cameron et al. 2021), rather than with LSD for the GP analysis, could be explored. Indeed, the strength of LSD is the increase in S/N it provides, at the cost of a poorer estimation of the uncertainties (usually overestimated). It is more relevant to have a better estimation of the uncertainties when it comes to the RV used for the GP rather than a boosted S/N (required for DI).

Now having a better grasp on the capabilities of these activity mitigation strategies, it is possible to study real data of young solar-type stars. Many projects such as the BCool (Marsden et al. 2014) and TOUPIES[4] (Folsom et al. 2016, 2018) surveys, aimed at characterizing star using DI and ZDI, would be good starting points.

Among the overwhelming number of targets observed by the *TESS* mission (Ricker et al. 2015), many are young Solar analogues. Careful planning of follow-up and the availability of photometry for transiting planets would drastically increase the characterization capabilities; see Appendix D. In general, using complementary tools

---

[4]https://ipag.osug.fr/Anr_Toupies//

to diagnose the activity, such as activity indicators and photometry, is strongly recommended (e.g. Rajpaul et al. 2015; Jones et al. 2017; Oshagh et al. 2017; Kosiarek & Crossfield 2020).

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author. The full content of Table 4 is available as online material.

REFERENCES

Appenzeller I., Mundt R., 1989, A&AR, 1, 291
Astropy Collaboration, 2013, A&A, 558, A33
Bailer-Jones C., Rybizki J., Fouesneau M., Demleitner M., Andrae R., 2020, Gaia eDR3 Lite Distances Subset Cone Search. VO Resource Provided by the GAVO Data Center. http://dc.zah.uni-heidelberg.de/gedr3dist/q/cone/info, Accessed date: May 15th, 2021
Bailey E., Batygin K., 2018, ApJ, 866, L2
Baluev R. V., 2008, MNRAS, 385, 1279
Baruteau C. et al., 2014, in Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds, Protostars and Planets VI. Univ. Arizona Press, Tucson, AZ, p. 667
Batygin K., Bodenheimer P. H., Laughlin G. P., 2015, ApJ, 829, 1
Beaugé C., Nesvorný D., 2012, ApJ, 751, 119
Boley A. C., Contreras A. P. G., Gladman B., 2015, ApJ, 817, L17
Bouma L. G. et al., 2020, AJ, 160, 239
Brown S. F., Donati J. F., Rees D., Semel M., 1991, A&A, 250, 463
Buchner J. et al., 2014, A&A, 564, A125
Cabot S. H. C. et al., 2021, AJ, 161, 26
Cang T. Q. et al., 2020, A&A, 643, A39
Chatterjee S., Ford E. B., Matsumura S., Rasio F. A., 2008, ApJ, 686, 580
Collier Cameron A., 2021, MNRAS, 505, 1699
Damasso M. et al., 2020, A&A, 642, A133
David T. J. et al., 2016, Nature, 534, 658
David T. J., Petigura E. A., Luger R., Foreman-Mackey D., Livingston J. H., Mamajek E. E., Hillenbrand L. A., 2019, ApJ, 885, L12
Dawson R. I., Johnson J. A., 2018, ARA&A, 56, 175
Donati J. F. et al., 2006, MNRAS, 370, 629
Donati J. F. et al., 2014, MNRAS, 444, 3220
Donati J. F. et al., 2015, MNRAS, 453, 3706
Donati J. F. et al., 2016, Nature, 534, 662
Donati J. F. et al., 2017, MNRAS, 465, 3343
Donati J. F., Brown S. F., 1997, A&A, 326, 1135
Donati J. F., Semel M., Carter B. D., Rees D. E., Cameron A. C., 1997, MNRAS, 291, 658
Donati J.-F. et al., 2020, MNRAS, 491, 5660
Donati J.-F., Collier Cameron A., Petit P., 2003, MNRAS, 345, 1187
Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601
Flagg L., Johns-Krull C. M., Nofi L., Llama J., Prato L., Sullivan K., Jaffe D. T., Mace G., 2019, ApJ, 878, L37
Folsom C. P. et al., 2016, MNRAS, 457, 580
Folsom C. P. et al., 2018, MNRAS, 474, 4956
Foreman-Mackey D., 2016, J. Open Source Softw., 1, 24
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Gaia Collaboration, 2021, A&A, 649, A1
Gaia Collaboration, 2016, A&A, 595, A1
Gaia Collaboration, 2018, A&A, 616, A1
Gregory P. C., 2007, MNRAS, 374, 1321
Hamers A. S., Antonini F., Lithwick Y., Perets H. B., Portegies Zwart S. F., 2017, MNRAS, 464, 688
Harris C. R. et al., 2020, Nature, 585, 357
Haywood R. D. et al., 2018, AJ, 155, 203
Haywood R. D., Cameron A. C., Queloz D., Others, 2014, MNRAS, 443, 2517
Hillenbrand L. A. et al., 2008, ApJ, 677, 630
Hobson M. P., Lasenby A. N., 1998, MNRAS, 298, 905
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Jeffers S. V., Barnes J. R., Jones H. R., Reiners A., Pinfield D. J., Marsden S. C., 2014, MNRAS, 438, 2717
Jeffers S. V., Keller C. U., 2009, in Stempels E., ed., AIP Conf. Proc. Vol. 1094, 15th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun. Am. Inst. Phys., New York, p. 664
Jeffreys H., 1961, Theory of Probability/by Harold Jeffreys, 3rd edn. Clarendon Press, Oxford
Johns-Krull C. M. et al., 2016, ApJ, 826, 206
Jones D. E., Stenning D. C., Ford E. B., Wolpert R. L., Loredo T. J., Gilbertson C., Dumusque X., 2017, preprint (arXiv:1711.01318)
Klein B. et al., 2020, MNRAS, 493, L92
Kochukhov O., Makaganiuk V., Piskunov N., 2010, A&A, 524, 1
Kosiarek M. R., Crossfield I. J. M., 2020, AJ, 159, 271
Kupka F. G., 2000, Baltic Astronomy, 9, 590
Lomb N. R., 1976, Ap&SS, 39, 447
Marsden S. C. et al., 2011a, MNRAS, 413, 1922
Marsden S. C. et al., 2011b, MNRAS, 413, 1939
Marsden S. C. et al., 2014, MNRAS, 444, 3517
Nadarajah S., 2005, J. Appl. Stat., 32, 685
Nava C., López-Morales M., Haywood R. D., Giles H. A. C., 2020, AJ, 159, 23
Nelson B. E. et al., 2020, AJ, 159, 73
Newton E. R. et al., 2019, ApJ, 880, L17
Oshagh M. et al., 2017, A&A, 606, A107
Perryman M., 2011, The Exoplanet Handbook. Cambridge University Press, Cambridge
Petit P. et al., 2004, MNRAS, 351, 826
Petit P., Donati J. F., Collier Cameron A., 2002, MNRAS, 334, 374
Petrovich C., 2015, ApJ, 805, 75
Petrovich C., Tremaine S., 2016, ApJ, 829, 132
Plavchan P. et al., 2020, Nature, 582, 497
Rajpaul V. M., Aigrain S., Buchhave L. A., 2020, MNRAS, 27, 3960
Rajpaul V., Aigrain S., Osborne M. A., Reece S., Roberts S., 2015, MNRAS, 452, 2269
Rajpaul V., Aigrain S., Roberts S., 2016, MNRAS, 456, L6
Ricker G. R. et al., 2015, J. Astron. Telesc. Instrum. Syst., 1, 014003
Rizzuto A. C. et al., 2020, AJ, 160, 33
Robert C. P., Chopin N., Rousseau J., 2009, Stat. Sci., 24, 141

Scargle J. D., 1982, ApJ, 263, 835
Semel M., 1989, A&A, 225, 456
Skilling J., Bryan R. K., 1984, MNRAS, 211, 111
Soummer R. et al., 2014, ApJ, 786, 2
VanderPlas J. T., 2018, ApJS, 236, 16
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Wright J. T., Marcy G. W., Howard A. W., Johnson J. A., Morton T. D., Fischer D. A., 2012, AJ, 753, 2
Yu L. et al., 2017b, MNRAS, 467, 1342
Yu L. et al., 2019, MNRAS, 489, 5556
Yu L. F., 2017a, in Reylé C., Di Matteo P., Herpin F., Lagadec E., Lançon A., Meliani Z., Royer F., eds, SF2A-2017: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, held 4-7 July, 2017 in Paris, p. 69

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Table 4. Data sets 1 to 6. Each column represents a simulated data set.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: BAYESIAN MODEL SELECTION

To assess the significance of the presence of periodicity in our data sets, we had to compare different models: activity only versus activity + planet(s). From parameter estimation and for a given model, an MCMC (nested sampling in our case) approach gives access to the most likely set of parameters for the model along with the associated maximum likelihood value. This non-normalized likelihood is, however, not comparable across models. A model with more parameters will have more flexibility and therefore the capacity to better fit the data, yielding a greater likelihood value. To circumvent this, one tries to penalize models with more parameters, following Ockham's razor (law of parsimony). The notoriously hard to compute marginal likelihood or evidence (or posterior distribution normalization constant) naturally applies Ockham's razor and acts to penalize models with higher number of parameters.

We encourage the reader to refer to Robert, Chopin & Rousseau (2009) for a thorough description of this Bayesian framework, based on the work of Jeffreys (1961). We will here give a quick overview.

First let's define the quantities required for this Bayesian framework. With **y** an array containing the data points, θ the set of parameters composing the model, and $\mathcal{M}_i$ the ith model, we define

(i) p($\theta|\mathbf{y}, \mathcal{M}_i$) or $\mathcal{P}(\theta)$: Posterior distribution of the parameters.
(ii) p($\mathbf{y}|\theta, \mathcal{M}_i$) or $\mathcal{L}(\theta)$: Likelihood.
(iii) p($\theta|\mathcal{M}_i$) or $\pi(\theta)$: Prior probability of the parameters.
(iv) p($\mathbf{y}|\mathcal{M}_i$) or $\mathcal{Z}$: Evidence/marginal likelihood/probability of data given the model/posterior distribution normalization constant.
(v) p($\mathcal{M}_i$) or $\pi(\mathcal{M}_i$): Model's prior.

The first four are linked by Bayes' theorem:

$$\mathcal{P}(\theta) = \frac{\pi(\theta) \times \mathcal{L}(\theta)}{\mathcal{Z}}. \tag{A1}$$

To compare models, we look at the BF, defined for two models $\mathcal{M}_0$ and $\mathcal{M}_1$ as

$$\mathrm{BF} = \frac{\mathcal{Z}_1}{\mathcal{Z}_0} \frac{\pi(\mathcal{M}_1)}{\pi(\mathcal{M}_0)}. \tag{A2}$$

Table A1. Values of the BF, i.e. the ratio of marginal likelihood between the single planet model and the 0-planet (stellar activity only) model. Middle and left columns are the associated probability and level of confidence.

Bayes factor	Probability	Level of confidence
<1	<0.5	None
<3	<0.75	Not worth more than a bare mention
<10	<0.909	Substantial
<30	<0.967	Strong
<100	<0.99	Very strong
>100	>0.99	Decisive

In cases where nothing a priori favours one model over the other (i.e. $\frac{\pi(\mathcal{M}_1)}{\pi(\mathcal{M}_0)} = 1$), we are left with the ratio of the marginal likelihoods. To obtain the marginal likelihoods, we need to marginalize (i.e. integrate) over all parameters:

$$\mathcal{Z} = \int \mathcal{L}(\theta)\pi(\theta)\mathrm{d}^N\theta, \tag{A3}$$

with N the number of parameters. For models with a large number of parameters, accurate computation of $\mathcal{Z}$ requires integrating over many dimensions (N) and is therefore often intractable. It quickly becomes too computationally expensive and needs to be approximated. Various approaches are used in the literature; see Nelson et al. (2020) for the most extensive attempt to compare these methods in the context of exoplanet searches.

Once in possession of the evidence for each model, we can compute the BF and assess whether the data favour model 1 over model 0. It is common practice to work with the *natural* logarithm of the evidence; we then have

$$\mathrm{BF} = \exp(\ln \mathcal{Z}_1 - \ln \mathcal{Z}_0) = \exp(\Delta \ln \mathcal{Z}). \tag{A4}$$

The different level of confidence is then interpreted from the BF, following Jeffreys (1961), as summarized in Table A1. We emphasize that this level of evidence is assessing the significance of the model given the data, and does not take into account how accurately the data reflect what we wish to observe nor the plausibility of the model on its own [although that can be added as $\pi(\mathcal{M}_i)$ in equation A2].

A1 Notes on prior probabilities

Prior probabilities are at the core of Bayesian inference, and express knowledge previously acquired on a particular aspect of the problem, i.e. a parameter of the model, or on the model itself. One must be mindful of the choice for these priors. Some are 'uninformative' (Uniform, Jeffrey's or Modified Jeffrey's priors; see Gregory 2007, for more details on the last two), meaning that they do not contain a priori information (they do to some extent as a uniform prior has limits, but these are rather physical than inferred from previous analyses). On the other hand, so-called 'informative' priors, typically a Gaussian prior, can strongly constrain the parameter space to be explored. This results in a boosted marginal likelihood compared with the use of an 'uninformative prior'.

Therefore, one has to be extremely cautious when using 'informative' priors. The previous knowledge yielding that prior has to be statistically robust, to not mislead the analysis by boosting the evidence.

A2 Note on evidence versus planet detection

Accurate estimation of the evidence grants a statistically robust framework to measure the significance of *one model relative to others*. We should keep in mind that models could be wrong and

that drawn conclusions about physical phenomena are only justified as long as these models and their underlying assumptions are reasonable.

Here, our model assesses the likelihood of the presence of a periodic signature in a data set. It does not, however, inform us about the origin of such signatures. Assumptions on the nature of the underlying physical phenomenon or the accuracy of the data collection and reduction are required to go from 'a periodic signal in the data' to 'a planet orbiting the observed star'. For example, a previous exoplanet detection claim was later attributed to the window function by Rajpaul, Aigrain & Roberts (2016). It has been the case again very recently for V830 Tau b (Damasso et al. 2020), where the signature found by Donati et al. (2016, 2017) could not be found in a new data set. Whether this means the planet does or does not exist is a rather challenging question, but it once more highlights the difficulty of RV-only searches.

APPENDIX B: SPURIOUS PERIODOGRAM PEAKS

For the two following data sets:

(i) #32: $K = 149.72$ m s^{-1}, $P_{orb} = 0.5526$ d, and $\Phi = 0.5701$
(ii) #36: $K = 142.36$ m s^{-1}, $P_{orb} = 0.8463$ d, and $\Phi = 0.6225$

the periodogram resulting from the DI analysis showed, in each case and after the activity filtering, the apparition of a spurious peak

at periods seemingly unrelated to the stellar rotation harmonics, the orbital period harmonics, or a peak in the window function (see Fig. B1). As shown in Fig. B2, $P_{spurious32} \approx 1.22$ d and $P_{spurious36} \approx 5.7$ d. After removing the identified highest peak from the filtered data (matching with the true period in the case of #36 but not for #32), the competing peak was also filtered out, suggesting an effect of the uneven sampling of the observations.

Figs B3, B4, and B5 show the corner plot of the posterior distribution and the resulting fit to the data for these two data sets following the GP analysis.

For data set #32 (Fig. B3 and top plot in Fig. B5), the GP does not seem to be fooled and clearly identifies the right period. However, the evidence is surprisingly low for such a large planet (BF = 1.67 or $p = 0.62$ in favour of the single model), yielding a non-detection. We strongly suspect that the ambiguity in both methods is due to the orbital period of the injected planet being very close to the third harmonic of the stellar period ($\frac{P_{eq}}{3} = 0.547$ d and $P_{orb} = 0.553$ d).

Regarding data set #36 (Fig. B4 and bottom plot in Fig. B5), once again the correct period is identified by the GP. We also note a slight appearance of the conflicting period around 6 d on the posterior distribution (Fig. B4). This time, however, the evidence is strongly in favour of a detection with a BF of 175.9 ($p = 0.994$).

Although it is not clear where the spurious peak arises from, one of these two cases could be settled by the GP.
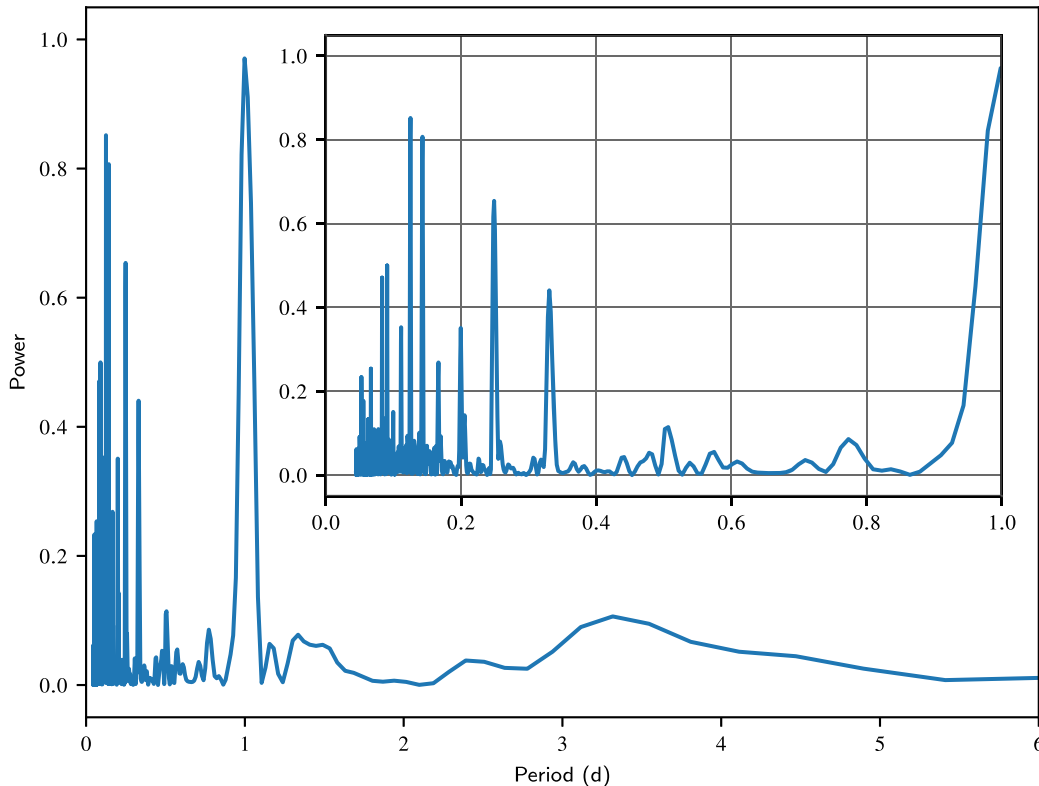

Figure B1. Window function of the observed RVs. The horizontal axis shows the period in days and the vertical axis the periodogram power. The upper right zoomed window shows extra details for periods <1 d.
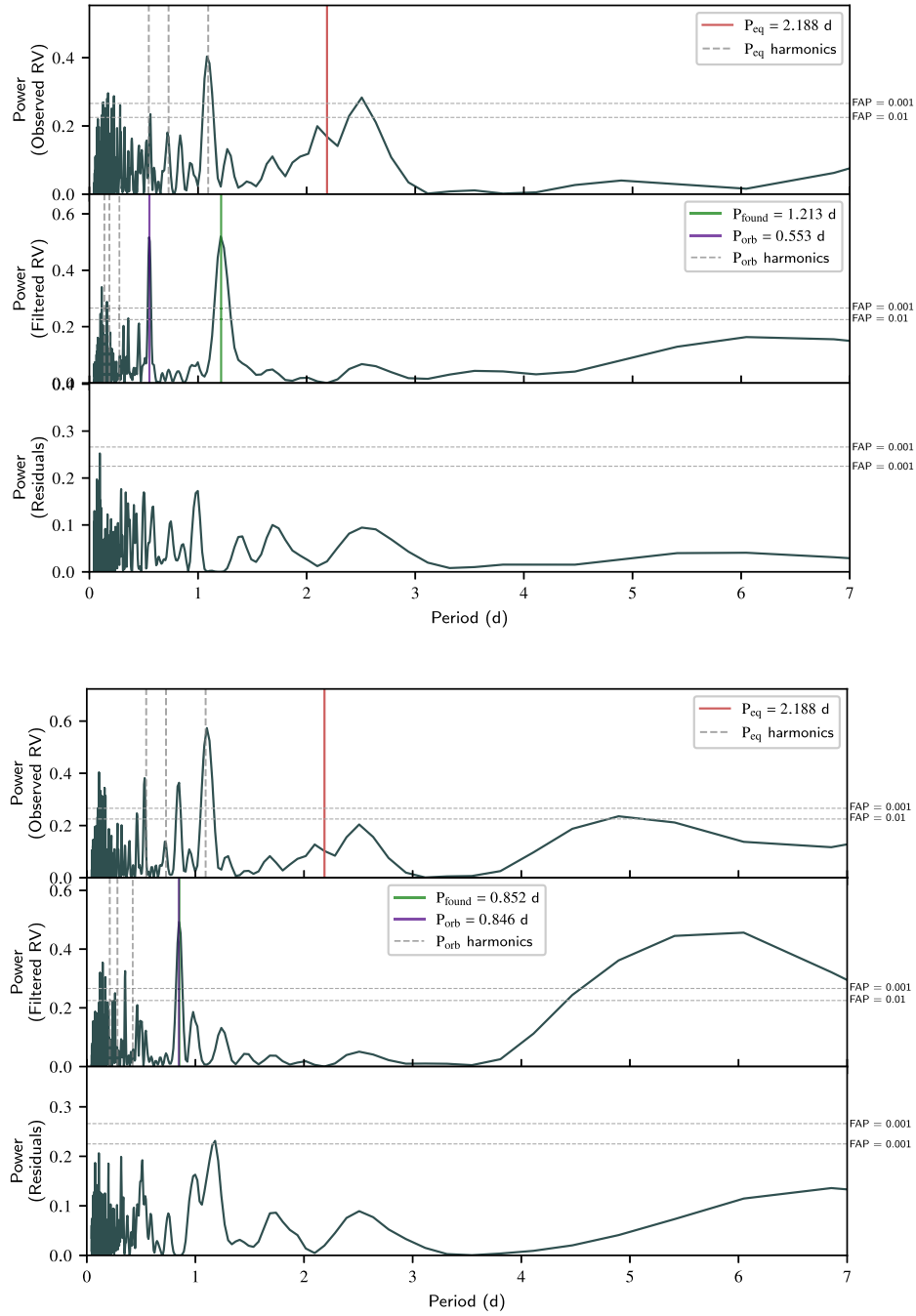
Figure B2. Periodograms of data sets #32 (top) and #36 (bottom). Each plot has three subplots showing the periodograms of the raw observed data (top), DI filtered (middle), i.e. observed – synthetic RVs and residuals (bottom), i.e. filtered – identified periodic signature. Vertical lines on the top subplots mark the stellar rotation period and its harmonics. In the middle subplots, we showed the recovered periodicity (green vertical line) and the true injected period (purple vertical line), along with its harmonics (grey dashed vertical lines).
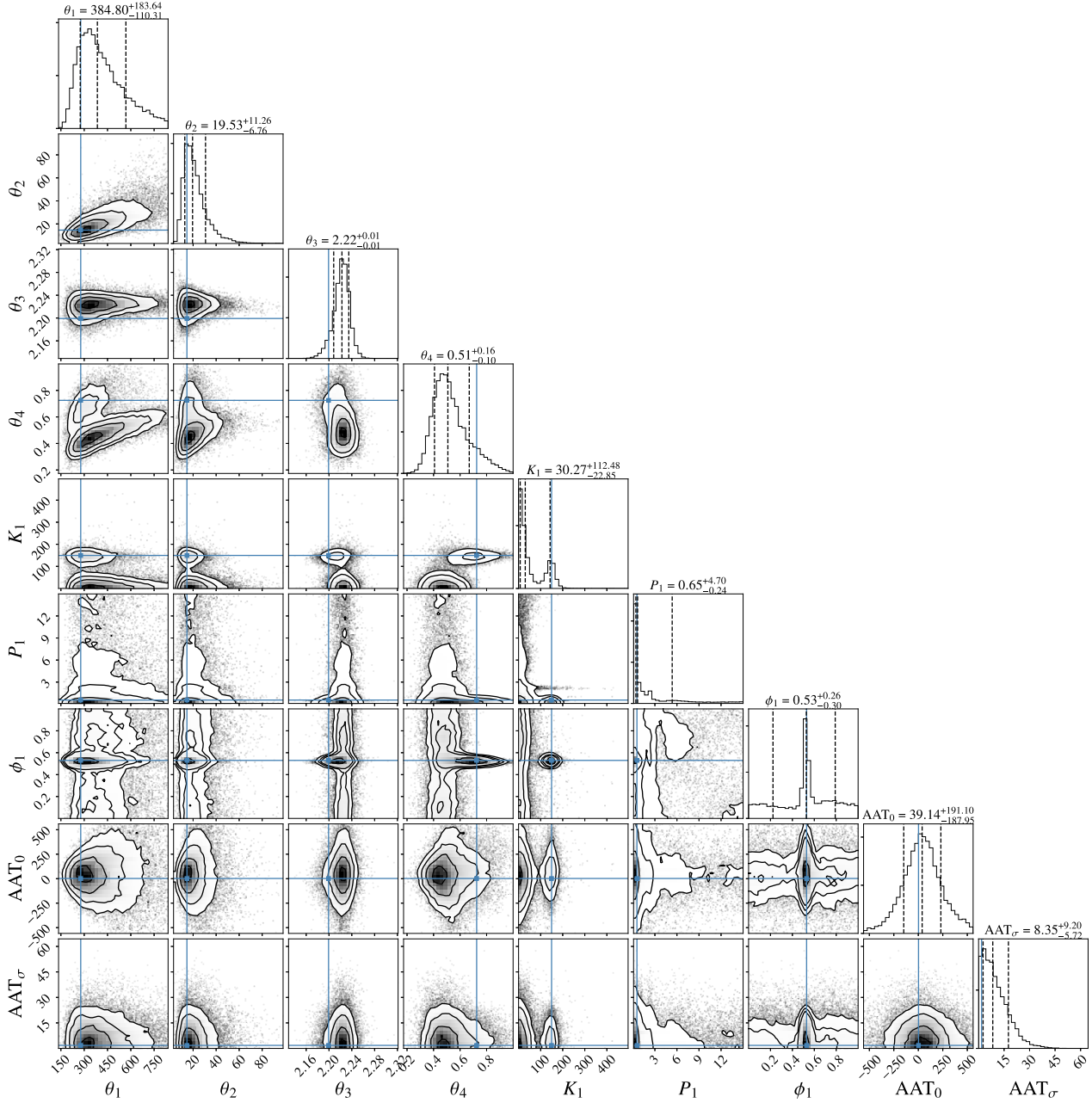
Figure B3. Posterior distribution of the parameters from the GP analysis of data set #32. Prior weighted posterior samples drawn from the PYMULTINEST analysis. Blue lines show the most likely parameters. Dashed vertical lines are 0.16, 0.5, and 0.84 quantiles. Contours are 1σ, 2σ, and 3σ levels (representing, respectively, 39, 86, and 99 per cent for a 2D distribution).
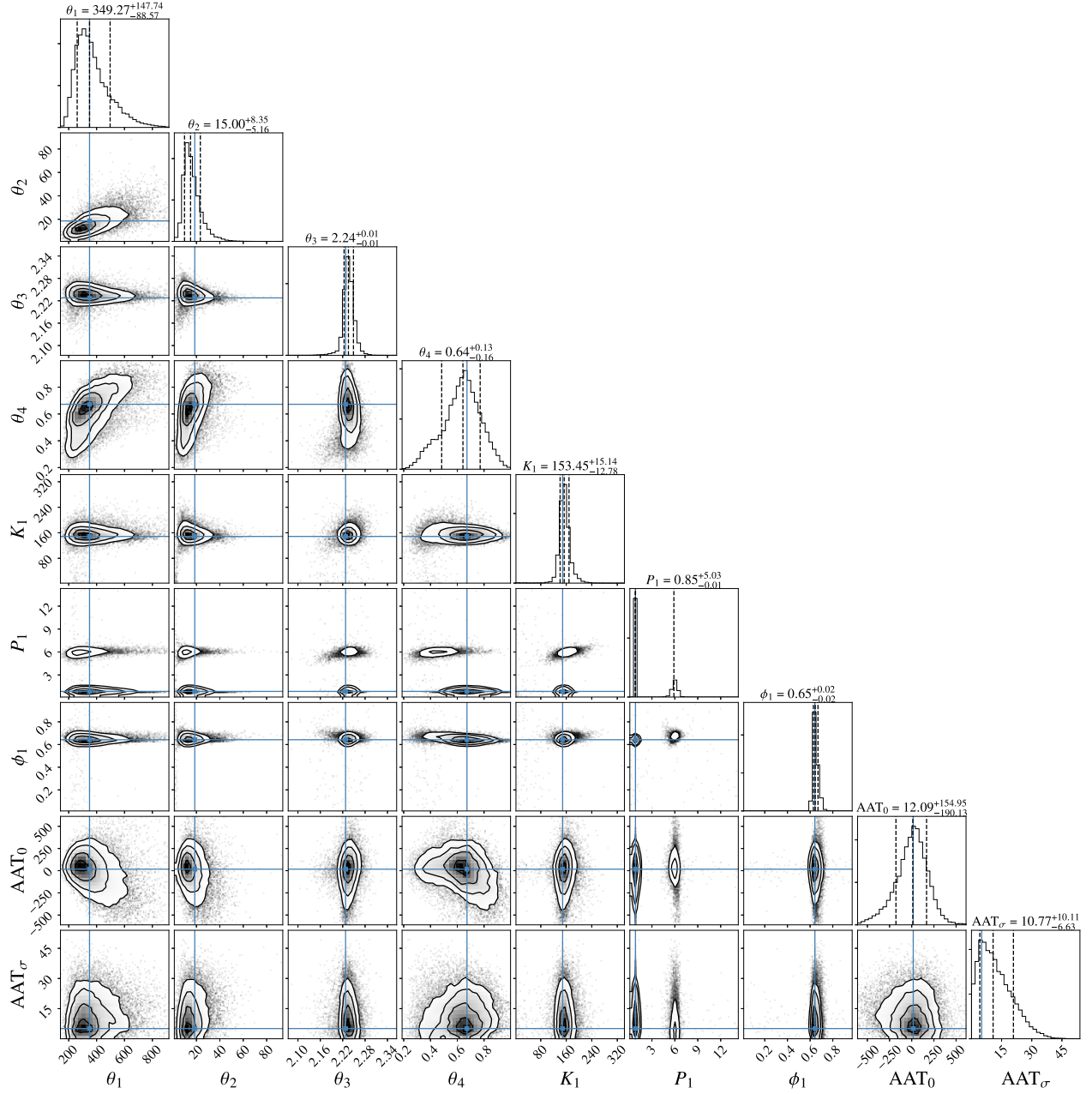
Figure B4. Posterior distribution of the GP analysis of data set #36. Prior weighted posterior samples drawn from the PYMULTINEST analysis. Blue lines show the most likely parameters. Dashed vertical lines are 0.16, 0.5, and 0.84 quantiles. Contours are 1σ, 2σ, and 3σ levels (representing, respectively, 39, 86, and 99 per cent for a 2D distribution).
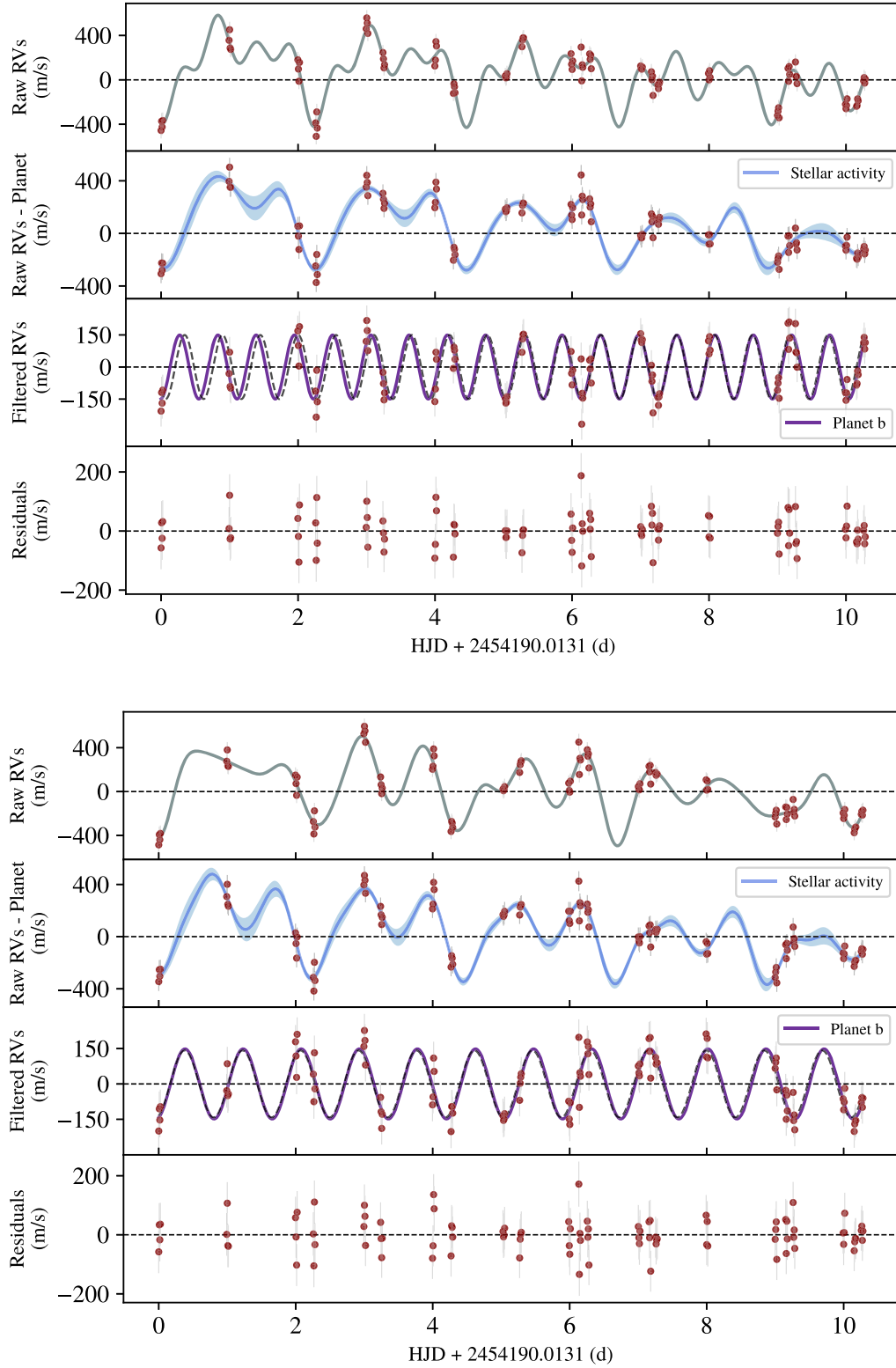
Figure B5. Resulting fits for the GP analysis of data sets #32 (top) and #36 (bottom). Top panel is the total model (grey line) and the observed raw RVs (red points). The second panel is the GP fit with the uncertainty shown by the shadowed area (analytically computed predictive standard deviation of the GP); red points are (raw RVs) – (recovered planet). The third panel is the recovered planet (purple line) on top of the true injected planet (dashed curve). Red points are (raw RVs) – (GP fit). The bottom panel is the residuals, i.e. (raw RVs) – (GP model) – (recovered planet).

APPENDIX C: IMPACT OF DIFFERENT DI SOLUTIONS ON DATA SET #22

As we saw in Section 4.2, the stellar parameters found via DI are slightly different from M11A/B. Because the DI filtering uses the synthetic generated line profiles that depend on the stellar parameters, we tested the influence of three different DI solutions on our planet detection capabilities. We re-analysed data set 22, containing an injected planet with $K = 117.28$ m s^{-1}, $P_{orb} = 3.1546$ d, and $\Phi = 0.7077$.

As described in Section 4.2, we derived the alternate DI solutions by fitting (i) only using dark features and (ii) using both dark and bright features but forcing the inclination parameter to 70° (matching M11A/B's value). The optimum parameters are displayed in Table C1.

Regarding the maps, results are similar to data set #5 (no planet), with mainly a difference in contrast. The periodograms of the filtered RVs (synthetic RVs – observed RVs) for each of the three cases are displayed in Fig. C1. In all cases, the features in the periodogram are quite similar. The analysis with dark spots only is the most different. It appears, in that case, that the original peak due to activity was not completely filtered and blends with the peak of the injected period after the filtering process. This acts to slightly hinder the accuracy of the period retrieved.

The fixed inclination case performs slightly better than the dark and bright case, as the retrieved period is 0.03 d away from the injected period for the former versus 0.08 d for the latter. However, the shape of the peak is very similar and we cannot conclude whether the better result is in fact due to a better solution for the DI or not.

Table C1. Test of different DI solutions fitting for dark and bright spots (top row), dark and bright features with the inclination parameter forced to 70° (middle row), and only dark features (bottom row).

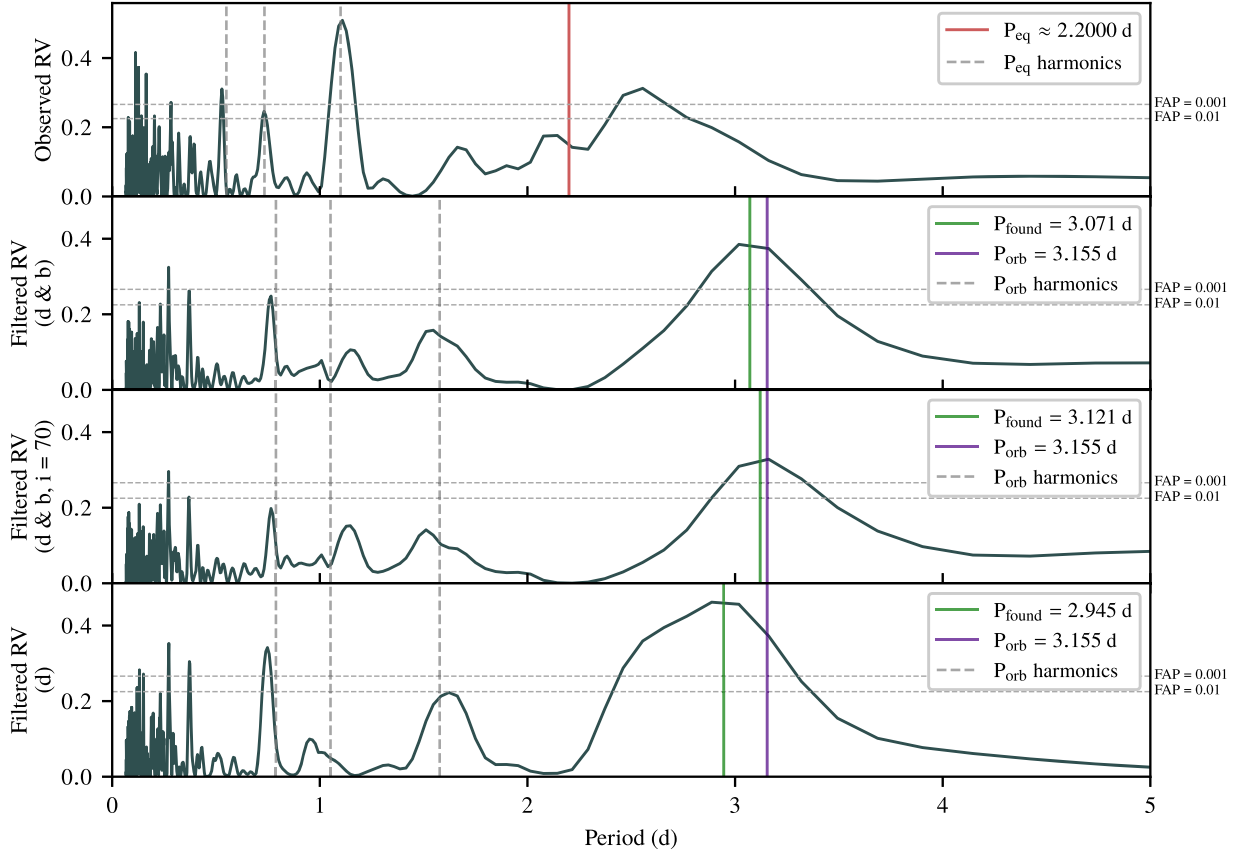| Fitting for	$v \sin i$	P_{eq}	dΩ	rad d^{-1}
Dark and bright	35.54	2.2003	0.1231	47
Dark and bright (fixed i)	35.58	2.1982	0.1138	70 (fixed)
Dark	35.38	2.2115	0.0441	44

Figure C1. Top: periodogram of the raw data from data set #22. Second to fourth plot: comparison of the periodograms for data set #22 obtained from dark and bright-spot DI analysis (2nd, d and b), from dark and bright-spot DI analysis with imposed inclination (3rd, d and b, $i = 70$), and from dark spots-only DI analysis (4th, d).

APPENDIX D: SIMULATING A TRANSITING PLANET

In order to showcase what access to a photometrically detected planet (i.e. the case of a transiting planet) would add to the retrieval capabilities, we investigated a data set allowing for additional prior constraints. For this, we chose data set #29, which exhibits the smallest semi-amplitude among the data sets that are not close to either P_{eq} or its harmonics. It has a semi-amplitude of 47 m s^{-1}, well below our detection threshold. In Table D1, we compared the pieces of evidence (derived from the GP analysis) for three different cases.

The first case (first line of Table D1) uses the same priors as we did throughout this article, simulating access solely to RV data. The second case (second line of Table D1) simulates the availability of transiting data on the planet. We set Gaussian priors for P_{orb} and

Φ, centred on the true injected value and with a standard deviation of, respectively, 0.0001 d and 0.01. For the third case (third line of Table D1), we fixed the value for P_{orb} and Φ to their true values (simulating the best transit value).

For each case, the BF and probability favouring the single planet model over the activity-only model can be found in the last two columns of Table D1. For the first case, the BF is extremely low (0.4, $p = 0.27$) and leads to a non-detection (see grey cross labelled '29' on Fig. 7). For cases 2 and 3, however (lines 2 and 3 of Table D1), their BFs are comparable and around 9, meaning a 0.9 probability in favour of the one-planet model and therefore a strong evidence for the presence of a planet.

This is expected as our constrained priors act to boost the evidence. It also shows how an inappropriate choice of priors can influence the evidence and bias the claim of a finding, following our discussion in A1. We see once more the difficulty of RV-only searches.

Table D1. GP analysis of data set #29 with three different sets of priors. The first column shows the real case configuration leading to the according choice of priors. The second column is the list of parameters with a Gaussian prior. The third column is the list of fixed parameters, i.e. not taking part in the parameter space search process. The last two columns are the BF and the probability favouring the single planet model over the activity-only model resulting from the corresponding analysis. To constrain or fix P_{orb} and Φ, we used the true injected values as they would be available if transits would have been identified.

Analysis	Constrained (i.e. Gaussian prior)	Fixed (not fitted)	BF	Probability
RV only	θ_3	None	0.4	0.27
Transit	θ_3, P_{orb}, Φ	None	9.9	0.91
Transit (fixed)	θ_3	P_{orb}, Φ	8.2	0.89

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