Revisiting K2-233 spectroscopic time-series with multidimensional Gaussian processes

Oscar Barragán [1]⋆ Edward Gillen [2] Suzanne Aigrain,[1] Annabella Meech [1] Baptiste Klein [1]
Louise Dyregaard Nielsen [1,3] Haochuan Yu,[1] Niamh K. O'Sullivan,[1] Belinda A. Nicholson[1,4]
and Jorge Lillo-Box[5]

[1]*Sub-department of Astrophysics, Department of Physics, University of Oxford, Oxford OX1 3RH, UK*
[2]*Astronomy Unit, Queen Mary University of London, Mile End Road, London E14NS, UK*
[3]*European Southern Observatory (ESO), Karl-Schwarzschild-Straße 2, D-85748 Garching bei München, Germany*
[4]*Centre for Astrophysics, University of Southern Queensland, West Street, Toowoomba, QLD 4350, Australia*
[5]*Depto. de Astrofísica, Centro de Astrobiología (CAB, CSIC-INTA), ESAC Campus, E-28692 Villanueva de la Cañada (Madrid), Spain*

ABSTRACT

Detecting planetary signatures in radial velocity time-series of young stars is challenging due to their inherently strong stellar activity. However, it is possible to learn information about the properties of the stellar signal by using activity indicators measured from the same stellar spectra used to extract radial velocities. In this manuscript, we present a reanalysis of spectroscopic High Accuracy Radial Velocity Planet Searcher data of the young star K2-233, which hosts three transiting planets. We perform a multidimensional Gaussian process regression on the radial velocity and the activity indicators to characterize the planetary Doppler signals. We demonstrate, for the first time on a real data set, that the use of a multidimensional Gaussian process can boost the precision with which we measure the planetary signals compared to a one-dimensional Gaussian process applied to the radial velocities alone. We measure the semi-amplitudes of K2-233 b, c, and d as $1.31^{+0.81}_{-0.74}$, $1.81^{+0.71}_{-0.67}$, and $2.72^{+0.66}_{-0.70}$ m s^{-1}, which translate into planetary masses of $2.4^{+1.5}_{-1.3}$, $4.6^{+1.8}_{-1.7}$, and $10.3^{+2.4}_{-2.6}$ M$_\oplus$, respectively. These new mass measurements make K2-233 d a valuable target for transmission spectroscopy observations with *JWST*. K2-233 is the only young system with two detected inner planets below the radius valley and a third outer planet above it. This makes it an excellent target to perform comparative studies, to inform our theories of planet evolution, formation, migration, and atmospheric evolution.

Key words: techniques: radial velocities – planets and satellites: individual: K2-233 – stars: activity.

1 INTRODUCTION

Young exoplanets (<1 Gyr) are crucial to understanding planetary evolution. However, their detection is difficult with indirect methods, such as transit and radial velocity (RV), due to strong stellar signals imprinted in photometric and spectroscopic time-series. Significant progress has been made in the detection of transit signals in young stars' light curves. The time-scale difference between the transit and stellar signals in photometric data has enabled detections of tens of young exoplanets by filtering the relatively low-frequency stellar signals. The *K2* (Howell et al. 2014) and *Transiting Exoplanet Survey Satellite* (*TESS*; Ricker et al. 2015) missions have discovered around a dozen transiting exoplanets or exoplanet candidates around young stars (e.g. Mann et al. 2016a, b, 2022; David et al. 2018; Newton et al. 2019, 2021; Bouma et al. 2020; Rizzuto et al. 2020; Hobson et al. 2021; Martioli et al. 2021). A natural next step after finding a young transiting planet is to follow it up with spectroscopic observations in order to detect the RV wobbles induced by the planet on its host star.

However, disentangling planet- and star-induced signals in RV data is still a big challenge in exoplanet science.

Gaussian processes (GPs) have became a popular mathematical framework to model activity-induced RVs given their flexibility to describe stochastic variations (see e.g. Aigrain & Foreman-Mackey 2022). Haywood et al. (2014) were the first to show that activity-induced RV signals can be modelled as a GP with a quasi-periodic (QP) covariance. Since then, multiple authors have used GPs to model stellar and planetary signals of active stars (e.g. Grunblatt, Howard & Haywood 2015; Fulton et al. 2018; Dai et al. 2019; Suárez Mascareño et al. 2021). This approach creates flexible models for the RV time-series, where there are risks that potential planetary signals can be absorbed or modified by the GP activity model (see e.g. discussions in Ahrer et al. 2021; Rajpaul et al. 2021).

Stellar activity also manifests in other observables that come from the same spectra from which the RVs are measured. They can be obtained from deformations of chromospheric lines and intensity of the lines, amongst others (e.g. Wilson 1968; Queloz et al. 2001; Boisse et al. 2009). This generates simultaneous ancillary time-series that contain information about the stellar activity and for this reason they are called *activity indicators*. They usually contain

⋆ E-mail: oscar.barragan@physics.ox.ac.uk

Table 1. Main identifiers and parameters for K2-233.

Parameter	Value	Source
Main identifiers		
Gaia DR3	6253186686054822784	Gaia Collaboration (2020)
TYC	6179−186−1	Høg et al. (2000)
2MASS	J15215519−2013539	Cutri et al. (2003)
Spectral type	K3	David et al. (2018)
Equatorial coordinates, proper motion, and parallax		
α(J2000.0)	$15^{\rm h}21^{\rm m}55\overset{\rm s}{.}1983$	Gaia Collaboration (2020)
δ(J2000.0)	$-20°13'53\overset{''}{.}9909$	Gaia Collaboration (2020)
μ_α (mas yr^{-1})	-20.031 ± 0.052	Gaia Collaboration (2020)
μ_δ (mas yr^{-1})	-30.963 ± 0.020	Gaia Collaboration (2020)
π (mas)	14.7719 ± 0.0188	Gaia Collaboration (2020)
Distance (pc)	67.695 ± 0.086	Gaia Collaboration (2020)
Magnitudes		
B	11.81 ± 0.15	Høg et al. (2000)
V	10.88 ± 0.09	Høg et al. (2000)
Gaia	10.4229 ± 0.0028	Gaia Collaboration (2020)
J	8.968 ± 0.020	Cutri et al. (2003)
H	8.501 ± 0.026	Cutri et al. (2003)
Ks	8.375 ± 0.023	Cutri et al. (2003)
Stellar parameters		
$T_{\rm eff}$ (K)	4796 ± 66	Lillo-Box et al. (2020)
log g (cgs)	4.53 ± 0.22	Lillo-Box et al. (2020)
[*Fe/H*]	-0.082 ± 0.028	Lillo-Box et al. (2020)
Age (Myr)	360^{+490}_{-140}	David et al. (2018)
$P_{\rm rot}$ (d)	9.754 ± 0.038	David et al. (2018)
$v\sin i$	4.5 ± 1.0	David et al. (2018)
(km s^{-1})		
Mass (M$_\odot$)	0.79 ± 0.01	Lillo-Box et al. (2020)
Radius (R$_\odot$)	0.71 ± 0.01	Lillo-Box et al. (2020)
Density	3.17 ± 0.07	Lillo-Box et al. (2020)
(g cm^{-3})		

only information about the stellar signal, and not about the planets (although note that signals associated with planets can occur in activity signals, see Klein et al. 2022, for an example). Therefore, they can be used to set priors on the pattern of the stellar activity in the RV time-series. Rajpaul et al. (2015) created a framework that uses spectroscopic activity indicators together with RVs in order to set better constraints on the activity-induced signal in the RV time-series. This is done by combining all time-series within a multidimensional GP (hereafter multi-GP) framework that exploits the correlations between them. Barragán et al. (2019b) showed that this multi-GP can lead to the detection of planetary signals tens of times smaller than the stellar signal. This approach has now been used to spectroscopically confirm a handful of transiting young planets (e.g. Mayo et al. 2019; Barragán et al. 2022b; Nardiello et al. 2022; Zicher et al. 2022). Furthermore, Barragán et al. (2022a) showed (with simulated data) that the multi-GP approach can lead to more precise planetary signal detection in comparison with RV-only GP regressions (see also Ahrer et al. 2021; Rajpaul et al. 2021).

In this manuscript, we explore this further with real data for the K2-233 system. K2-233 is a young ($\sim$360 Myr) star hosting three transiting planets with orbital periods of $\sim$2.5, 7, and 24 d (David et al. 2018). K2-233's main identifiers and properties appear in Table 1. Lillo-Box et al. (2020) performed an intensive spectroscopic follow-up of the system, confirming the planetary nature of the three transiting signals and putting constraints on the composition of the three planets. In this work, we reanalyse the Lillo-Box et al. (2020) data sets with a multi-GP approach to put new constraints on the

planetary masses. This manuscript is part of a series of papers under the project *GPRV: Overcoming stellar activity in radial velocity planet searches* funded by the European Research Council (ERC; PI S. Aigrain). The outline of this manuscript is as follows: Section 2 describes the K2-233 photometric and spectroscopic data used for the analyses presented in Section 3. Section 4 is devoted to the discussion, followed by Section 5 where we summarize the most important results of this work.

2 K2-233 DATA

2.1 Light curve

K2-233 was observed by *K2* during its Campaign 15 between 2017 August 23 and 2017 November 20. Using these data, David et al. (2018) discovered and validated three transiting planets with periods of 2.5, 7, and 24 d, and radii of 1.3, 1.3, and 2.6 R$_\oplus$, respectively. We downloaded K2-233's light curve as provided by Vanderburg & Johnson (2014).[1] Fig. 1 shows K2-233's *K2* photometric time-series. K2-233's astrophysical signal is a result of activity regions on the stellar surface and transiting signals.

In order to analyse the transiting signals, we proceed to detrend the light curve using the public code CITLALICUE (Barragán et al. 2022a). In short, CITLALICUE uses a GP regression as implemented in GEORGE (Ambikasaran et al. 2015) to model the low-frequency signals caused by stellar activity in the light curves. We input to CITLALICUE the light curves normalized to 1, as well as the ephemeris of the three transiting signals. We proceed by binning the data to 3 h and masking out all the transits from the light curve while performing the GP regression using a QP kernel (see Ambikasaran et al. 2015, for more details). We apply an iterative optimization to find the maximum likelihood together with a 5σ clipping algorithm to find the best model describing the out-of-transit variations. We then divide the *K2* light curve by the inferred model to obtain a flattened light curve with transiting signals. Fig. 1 shows the detrended light curve.

2.2 Spectroscopic observations

Lillo-Box et al. (2020) performed a spectroscopic follow up of K2-233 using the High Accuracy Radial Velocity Planet Searcher (HARPS; Mayor et al. 2003) spectrograph mounted at the 3.6-m European Southern Observatory (ESO) telescope at La Silla Observatory in Chile. They collected 124 HARPS spectra in a window of 105 d (based on observations collected under ESO programmes 198.C-0169 and 095.C-0718).

The spectra obtained by Lillo-Box et al. (2020) are high quality with a typical Signal to Noise ratio (S/N) > 50 (at 550 nm) per observation, which leads to a median RV uncertainty of 1.2 m s^{-1}. However, we found that four observations have relatively large error bars, larger than 3 m s^{-1}. They are the observations taken on 2458294.516674118 BJD (2018 June 25 00:24 UT), 2458301.658614401 BJD (2018 July 2 3:48 UT), 2458335.597976999 BJD (2018 August 5 2:21 UT), and 2458354.5192679004 BJD (2018 August 24 0:27 UT). We checked the observational conditions at La Silla when these data were collected using the ESO Astronomical Site Monitor (ASM) online tool.[2] We found that the observations taken on 2018 June 25 00:24 UT

[1] https://lweb.cfa.harvard.edu/~avanderb/k2.html.
[2] https://www.eso.org/asm.
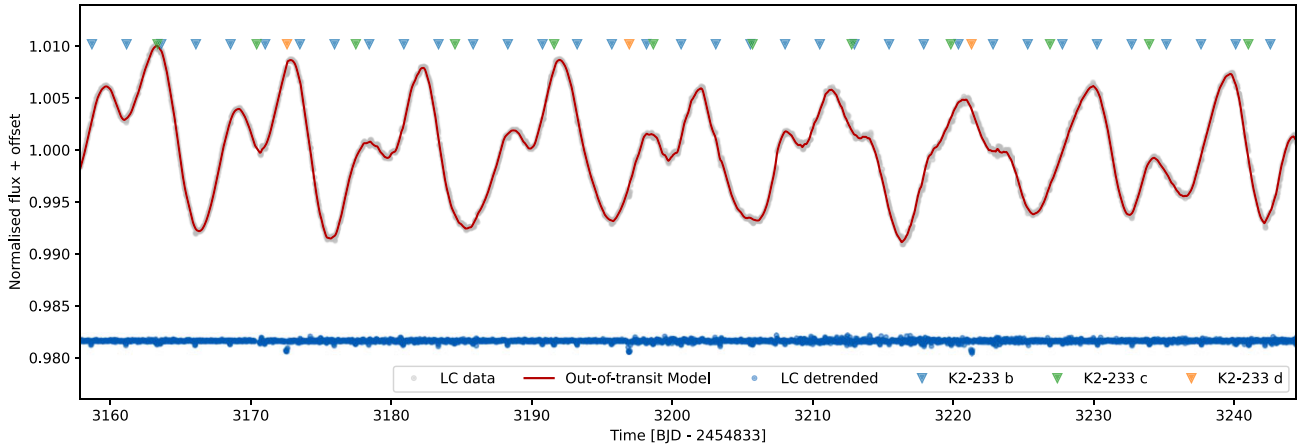
Figure 1. *K2* light curve for K2-233. *K2* data are shown with grey points with the out-of-transit variability model overplotted in red. The resulting flattened light curve is presented with blue points. Transit positions are marked with blue, orange, and green triangles for K2-233 b, c, and d, respectively.

Table 2. HARPS spectroscopic measurements. The full version of this table is available in a machine-readable format as part of the supplementary material.

Time BJD$_{TDB}$−2450000	RV km s^{-1}	σ_{RV} km s^{-1}	FWHM km s^{-1}	BIS km s^{-1}	S_{HK}	$\sigma_{S_{HK}}$	Hα	$\sigma_{H\alpha}$	Na	σ_{Na}
8257.575545	−9.6220	0.0019	7.6508	0.0174	0.6893	0.0068	0.4774	0.0021	0.1713	0.0008
8257.636216	−9.6244	0.0020	7.6542	0.0156	0.6941	0.0073	0.4789	0.0022	0.1682	0.0009
8257.755000	−9.6260	0.0020	7.6635	0.0213	0.7034	0.0079	0.4802	0.0021	0.1693	0.0009
8258.564438	−9.6592	0.0019	7.6835	0.0566	0.7124	0.0065	0.4957	0.0021	0.1695	0.0008
8258.658789	−9.6663	0.0022	7.6591	0.0436	0.6915	0.0083	0.4868	0.0024	0.1682	0.0010
...										

and 2018 August 24 0:27 UT were taken with bad seeing conditions (>2). In particular, the observation taken on 2018 June 25 00:24 UT suffers from a significant change of seeing during the acquisition; this generates an uneven light collection through the fibre that can create unwanted systematics effects in precision RV measurements. We decided to remove the observations taken on 2018 June 25 00:24 UT and 2018 August 24 0:27 UT from our subsequent analyses.

We downloaded the K2-233's data taken from Lillo-Box et al. (2020) from the ESO archive. We reduced the data using the dedicated HARPS data reduction software (DRS) and extracted the RV measurements by cross-correlating the Echelle spectra with a K5 numerical mask (Baranne et al. 1996; Pepe et al. 2002; Lovis & Pepe 2007). We also used the DRS to extract the Ca II H & K lines (S_{HK}), Hα, and Na chromospheric activity indicators, as well as two classic profile diagnostics of the cross-correlation function (CCF), namely, the full width at half maximum (FWHM), and the bisector inverse slope (BIS). These reprocessed HARPS RV measurements have a typical error bar of 1.2 m s^{-1} and an RMS of 9.1 m s^{-1}. We note that these new RVs are equivalent to the values reported in Lillo-Box et al. (2020), but this reprocessing of the spectra allowed us to extract tailored activity indicators. Table 2 lists the HARPS spectroscopic time-series.

We acknowledge that there are eight archival HARPS observations of K2-233 taken in 2014 (under the programmes: 085.C-0019(A) by PI Lo Curto and 072.C-0488(E) by PI Mayor). We note that these observations have suboptimal sampling in order to constrain the shape of the stellar signal in the spectroscopic time-series. Therefore, we do not include them in our analysis.

3 DATA ANALYSIS

We performed a series of joint transit and spectroscopic time-series models. All our models are created using the same configuration for transit modelling (described in Section 3.1). However, we change the way in which we modelled the spectroscopic time-series in order to explore how different methods disentangle the planetary and stellar signals from the RV time-series (Section 3.2).

For all the subsequent analyses, we used the code PYANETI (Barragán, Gandolfi & Antoniciello 2019a; Barragán et al. 2022a). In all our runs, we sample the parameter space with 250 walkers using the Markov chain Monte Carlo (MCMC) ensemble sampler algorithm implemented in PYANETI (Foreman-Mackey et al. 2013; Barragán et al. 2019a). The posterior distributions are created with the last 5000 iterations of converged chains. We thinned our chains by a factor of 10 giving a distribution of 125 000 points for each sampled parameter.

3.1 Transit modelling configuration

In order to perform the transit analysis, we use the flattened *K2* light curve (see Section 2.1). To speed up the transit modelling, we only model data chunks spanning a maximum of 4 h on either side of each transit mid-time. Since *K2* Campaign 15 data were taken with a cadence of 30 min, we resampled the model over 30 steps to account for the data integration (Kipping 2010). We model the transits for the three planets, where for each planet we sample for the time of transit, T_0; orbital period, P_{orb}; orbital eccentricity, e;

angle of periastron, ω; and scaled planetary radius $R_p/R_\star$. We note that we also need to solve for the scaled semimajor axis ($a/R_\star$) for each planet. To do this, we sample for the stellar density, $\rho_\star$, and we recover $a/R_\star$ for each planet using Kepler's third law (see e.g. Winn 2010). We account for the limb darkening using the quadratic limb-darkening approach described in Mandel & Agol (2002) with the q_1 and q_2 parametrization given by Kipping (2013). We also sample for a photometric jitter term (σ_{K2}) to penalize the imperfections of our transit model.

Now, we will describe the priors adopted for the transit parameters for all our runs. We adopt a beta distribution prior for the eccentricity, $\mathcal{B}(1.52, 29)$, as advocated by Van Eylen et al. (2019) for multiplanet systems. For the angle of periastron, we set a uniform prior ranging between 0 and 2π. We set a Gaussian prior on the stellar density based on the stellar parameters given in Table 1. For the rest of the parameters, we set informative uniform priors.

3.2 Detecting the planetary Doppler signals

3.2.1 One-dimensional RV GP regression

We first run a one-dimensional GP modelling of the RVs. We note that this analysis is equivalent to the one presented in Lillo-Box et al. (2020). We model the stellar signal using a GP whose covariance between two times t_i and t_j is given by

$$\gamma_{1D,i,j} = A^2 \gamma_{QP,i,j}, \tag{1}$$

where A is an amplitude term, and $\gamma_{QP,i,j}$ is the QP kernel given by

$$\gamma_{QP,i,j} = \exp\left[-\frac{\sin^2[\pi(t_i - t_j)/P_{GP}]}{2\lambda_P^2} - \frac{(t_i - t_j)^2}{2\lambda_e^2} \right], \tag{2}$$

whose hyperparameters are P_{GP}, the GP characteristic period; λ_p, the inverse of the harmonic complexity; and λ_e, the long-term evolution time-scale.

The planetary signals are included as a parametric mean function of the GP. We use three Keplerian signals, where every Keplerian signal is modelled with a time of minimum conjunction (equivalent to the time of transit for transiting planets), T_0; orbital period, P_{orb}; orbital eccentricity, e; angle of periastron, ω; and Doppler semi-amplitude, K. We also include one offset to account for the systemic velocity of the star, and a jitter term to penalize the imperfections of our model.

We set wide uniform priors for all the GP hyperparameters. For λ_e, we set a uniform prior between 0 and 500 d, for λ_p between 0.1 and 10, and for P_{GP} between 8 and 12 d. We note that the Keplerian models share T_0, P_{orb}, e, and ω with the transit modelling. Therefore, the priors for these parameters are described in Section 3.1. For the rest of the Keplerian parameters and the offset, we set uniform priors. For the jitter term, we set modified Jeffreys priors as defined by Gregory (2005). The posterior creation follows the MCMC guidelines given at the beginning of Section 3.

The recovered QP kernel hyperparameters are $\lambda_e = 17.1 \pm 1.7$ d, $\lambda_p = 0.29 \pm 0.03$, and $P_{GP} = 9.65 \pm 0.09$ d. For a discussion on the interpretation of these hyperparameters, we refer the reader to Appendix A. Fig. 2 shows the recovered posterior distributions for the Doppler semi-amplitudes for K2-233 b, c, and d (black line histograms). The recovered values for the Doppler semi-amplitudes are $1.9^{+1.6}_{-1.2}$, $1.8^{+1.1}_{-1.0}$, and $1.7^{+1.3}_{-1.0}$ m s^{-1} for K2-233 b, c, and d, respectively. These values are consistent with the published values in Lillo-Box et al. (2020). This is expected given that we are reproducing the same analysis in order to compare it with the analysis presented in Section 3.2.2. All these values are constrained to less than a 2σ confidence level. We note that exploring the posteriors for the orbital
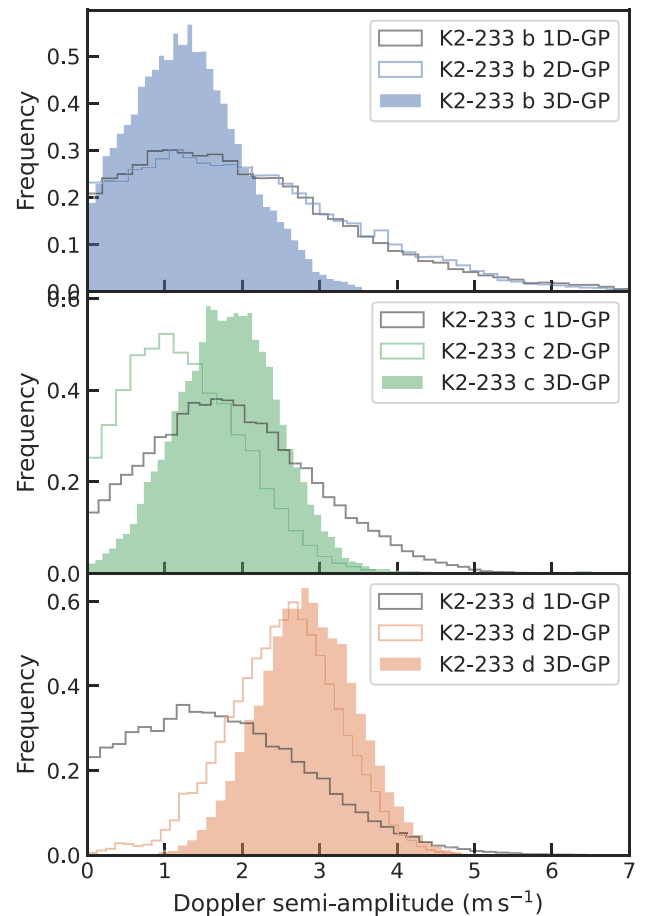
Figure 2. Posterior distributions for the Doppler semi-amplitudes of K2-233 b (top), c (centre), and d (bottom). Shape of the posterior distributions for the one- (solid black line), two- (solid colourful line), and three-dimensional (filled colourful) GP analyses are shown.

eccentricities they all are consistent with a zero eccentricity and the shape of the posterior follows the beta prior.

3.2.2 Multidimensional GP regression

We then perform a multi-GP approach to characterize the stellar and planetary signals in our RV time-series (see Rajpaul et al. 2015; Barragán et al. 2022a, for more details). We create N-dimensional GP models, including N time-series $\mathcal{A}_i$, as

$$\begin{aligned} \mathcal{A}_1 &= A_1 G(t) + B_1 \dot{G}(t) \\ &\vdots \\ \mathcal{A}_N &= A_N G(t) + B_N \dot{G}(t), \end{aligned} \tag{3}$$

where the variables $A_1, B_1, \cdots, A_N, B_N$ are free parameters that relate the individual time-series to $G(t)$ and $\dot{G}(t)$. In this approach, $G(t)$ is a latent (unobserved) variable, which can be loosely interpreted as representing the projected area of the visible stellar disc that is covered in active regions as a function of time.

We perform a two-dimensional GP model between the RVs and FWHM, and a three-dimensional run between RVs, FWHM, and BIS. For a further discussion on the use of other activity indicators, we encourage the reader to check Appendix A. The multidimensional covariance matrix was created using the QP kernel given in equation (2) and its derivatives as described in Rajpaul et al. (2015) and Barragán et al. (2022a). We assume that RVs and BIS time-series

can be described as $\mathcal{A}_i = A_i G(t) + B_i \dot{G}(t)$, while the FWHM time-series is photometric-like only, i.e. it is described as $\mathcal{A}_i = A_i G(t)$. For all runs, the mean function corresponding to the RVs was created using three Keplerian curves following the same set-up described in Section 3.2.1, including the offset and jitter terms. We recall the reader that these runs are done together with the transit analysis following the set-up described in Section 3.1. For the activity indicators, the mean function was treated as an offset. We also include a jitter term for each activity indicator. The priors for the mean functions and QP kernel hyperparameters have the same ranges as described in Section 3.2.1. We create posterior distributions using the same MCMC approach as in Section 3.1.

For the two-dimensional GP analysis of the RVs and FWHM, we recover the hyperparameters $\lambda_e = 16.2 \pm 2.0$ d, $\lambda_p = 0.46 \pm 0.05$, and $P_{GP} = 9.73 \pm 0.08$ d. For a more detailed discussion on the interpretation of the hyperparameters, we refer the reader to Section 3.2.1 and to Nicholson & Aigrain (2022). The recovered values for the Doppler semi-amplitudes are $1.9^{+1.6}_{-1.2}$, $1.18^{+0.86}_{-0.69}$, and $2.57^{+0.68}_{-0.74}$ m s^{-1} for K2-233 b, c, and d, respectively. Fig. 2 shows the recovered posterior distributions for the recovered Doppler semi-amplitudes for the three planets (coloured, unshaded histograms). This analysis gives a detection of K2-233 d with a significance of detection of 3.5σ. While for planet b the posterior does not change significantly, and there is an improvement in the detection of planet c, but still below the 2σ significance. This improvement in the detection of the planet signal can be explained by a better constraint of the stellar activity by using the information of the stellar signal in the FWHM time-series. It is worth noting that the posterior distribution for the eccentricity of planet d does not follow the prior. The inferred eccentricity of planet d is $e_d = 0.12^{+0.09}_{-0.07}$, despite the prior pushing for smaller values.

For the three-dimensional GP analysis of the RVs, FWHM, and BIS, we recover the hyperparameters $\lambda_e = 18.0 \pm 2.1$ d, $\lambda_p = 0.50 \pm 0.05$, and $P_{GP} = 9.75 \pm 0.07$ d. For this new analysis, the recovered values for the Doppler semi-amplitudes for K2-233 b, c, and d are $1.26^{+0.79}_{-0.71}$, $1.82^{+0.67}_{-0.69}$, and 2.83 ± 0.65 m s^{-1}, respectively. As in the two-dimensional GP regression, we also have an improvement in the RV Doppler detection of the three planets; in this case, we detect K2-233 b, c, and d with a significance of 1.8σ, 2.7σ, and 4.3σ, respectively. Fig. 2 shows the Doppler semi-amplitude posterior distributions coming from the three-dimensional GP analysis (coloured, shaded histograms). We can see that the inclusion of the BIS time-series does not lead to a significant improvement in the detection of K2-233 d. However, this three-dimensional analysis does boost the detection of K2-233 c. In this particular case, the inclusion of the BIS time-series helps to constrain the activity signal and disentangle it from the Doppler signal of one of the planets. We note too that, although we do not quite achieve a 3σ detection of K2-233 b's Doppler signal, the range of possible masses is better constrained than before. The derived mass for this planet is $2.4^{+1.5}_{-1.3}$ M$_\oplus$ with a maximum mass of 6.12 M$_\oplus$ at the 99 per cent confidence level. Similar to the two-dimensional GP analysis, we found that the posterior distribution for K2-233 d's eccentricity does not follow the prior and in this case, the inferred value is $e_d = 0.19 \pm 0.09$.

3.3 Final model

Based on the analyses presented in this section, our final model for the photometric and spectroscopic data of K2-233 is the transit model described in Section 3.1, together with the three-dimensional GP regression described in Section 3.2.2 (in Section 4.1, we describe the

reasoning to prefer this model over the others). For the RV planetary model, we assume that the two innermost planets have circular orbits while allowing for an eccentricity for K2-233 d. The whole set of sampled parameters and priors are shown in Table 3.

Fig. 3 shows the spectroscopic time-series plots resulting from this joint analysis, while Fig. 4 shows the independent transit and Doppler signals for each of the planets. Fig. 5 shows the posterior distribution for the key parameters of the model. Table 3 shows the inferred sampled parameters, defined as the median and 68.3 per cent credible interval of the posterior distribution. Table 4 shows all the derived planetary and orbital parameters. The detected Doppler semi-amplitudes $1.31^{+0.81}_{-0.74}$, $1.81^{+0.71}_{-0.67}$, and $2.72^{+0.66}_{-0.70}$ m s^{-1} translate to masses of $2.4^{+1.5}_{-1.3}$, $4.6^{+1.8}_{-1.7}$, and $10.3^{+2.4}_{-2.6}$ M$_\oplus$ for planets K2-233 b, c, and d, respectively. This new methodology provides higher precision masses than those reported in the original detection and confirmation work by Lillo-Box et al. (2020). Therefore, we will reassess the implications on planetary composition and perspectives on atmospheric characterization in Section 4.

4 DISCUSSION

4.1 Comparison between models

In Section 3.2, we have shown how the multi-GP analysis provides better constrained posterior distributions for the planetary-induced RV signals in comparison with the RV-only one-dimensional GP regression. From a qualitative point of view, a priori we know that there is an intrinsic correlation between the stellar signal in the RVs and activity indicators time-series (see e.g. Dumusque, Boisse & Santos 2014). In the case where there is no instrumental red noise in our time-series, we can assume that a model that uses activity indicators to constrain the stellar signal in the RV time-series is better than a model of the RVs only. This is especially important while doing GP regression, since an unconstrained GP predictive distribution can absorb or modify the signal from the mean function. On the other hand, the multi-GP approach sets a tighter constraint on the GP predictive distribution for the activity model in the RV time-series, which leads to less biased recovery of the mean model and, consequently, better inference of the mean function parameters (see e.g. Barragán et al. 2022a, for a more extensive discussion on this).

Given the Bayesian approach of the analyses, Bayesian information criterion (BIC) or Bayesian evidence would be the best option to perform a quantitative comparison between our models. However, these methods are useful to compare different models to the same data set. Therefore, we cannot compare the one-dimensional GP model with the two- and three-dimensional GP models given that they all contain different time-series (see also Ahrer et al. 2021; Rajpaul et al. 2021, for a discussion of this problem).

Rajpaul et al. (2021) proposed a solution to perform Bayesian comparisons for multi-GP regressions of different dimensions. Briefly, they modelled the activity indicators simultaneously with the RVs, but each time-series was modelled with different set of parameters that are not connected in any way. The activity indicators (in their case BIS and log R'_{HK}) are modelled with a constant offset and jitter white noise. As it is argued in that manuscript, this approach has no effect on the inference of the RV-related planetary and orbital parameters. This approach also ensures that the model evidences were consistently normalized (having the same number of data), allowing comparisons between all of their models. However, we notice that with this approach, the underlying model that describes the activity indicators is strongly penalized by default given that

Table 3. Model parameters and priors for joint fit.

Parameter	Prior[a]	Final value[b]
K2-233 b's parameters		
Orbital period, P_{orb} (d)	$\mathcal{U}[2.4667, 2.4683]$	$2.467\,583^{+0.000\,062}_{-0.000\,067}$
Transit epoch, T_0 (BJD$_{TDB}$−2450000)	$\mathcal{U}[7991.6745, 7991.7077]$	7991.6897 ± 0.0013
Scaled planet radius, $R_p/R_\star$	$\mathcal{U}[0.0, 0.05]$	$0.016\,98^{+0.000\,41}_{-0.000\,38}$
Impact parameter, b	$\mathcal{U}[0, 1]$	$0.20^{+0.11}_{-0.12}$
Orbital eccentricity, e	$\mathcal{F}[0]$	0
Angle of periastron, ω (deg)	$\mathcal{F}[90]$	90
Doppler semi-amplitude variation, K (m s^{-1})	$\mathcal{U}[0, 50]$	$1.31^{+0.81}_{-0.74}$
K2-233 c's parameters		
Orbital period, P_{orb} (d)	$\mathcal{U}[7.0547, 7.0655]$	$7.060\,24^{+0.000\,44}_{-0.000\,43}$
Transit epoch, T_0 (BJD$_{TDB}$−2450000)	$\mathcal{U}[7586.8425, 7586.9105]$	$7586.862^{+0.027}_{-0.028}$
Scaled planet radius, $R_p/R_\star$	$\mathcal{U}[0.0, 0.05]$	$0.016\,42 \pm 0.000\,44$
Impact parameter, b	$\mathcal{U}[0, 1]$	$0.15^{+0.12}_{-0.10}$
Orbital eccentricity, e	$\mathcal{F}[0]$	0
Angle of periastron, ω (deg)	$\mathcal{F}[90]$	90
Doppler semi-amplitude variation, K (m s^{-1})	$\mathcal{U}[0, 50]$	$1.81^{+0.71}_{-0.67}$
K2-233 d's parameters		
Orbital period, P_{orb} (d)	$\mathcal{U}[24.3509, 24.3781]$	$24.3681^{+0.0013}_{-0.0014}$
Transit epoch, T_0 (BJD$_{TDB}$−2450000)	$\mathcal{U}[8005.5640, 8005.6008]$	$8005.5802^{+0.0016}_{-0.0014}$
Scaled planet radius, $R_p/R_\star$	$\mathcal{U}[0.0, 0.05]$	$0.030\,51^{+0.000\,69}_{-0.000\,65}$
Impact parameter, b	$\mathcal{U}[0, 1]$	$0.28^{+0.16}_{-0.18}$
Orbital eccentricity, e	$\mathcal{B}[1.52, 29]^c$	0.18 ± 0.09
Angle of periastron, ω (deg)	$\mathcal{U}[0, 360]$	168^{+20}_{-25}
Doppler semi-amplitude variation, K (m s^{-1})	$\mathcal{U}[0, 50]$	$2.72^{+0.66}_{-0.70}$
GP hyperparameters		
GP period, P_{GP} (d)	$\mathcal{U}[8, 12]$	$9.746^{+0.065}_{-0.070}$
λ_p	$\mathcal{U}[0.1, 5]$	$0.490^{+0.046}_{-0.042}$
λ_e (d)	$\mathcal{U}[1, 500]$	$17.95^{+2.08}_{-1.98}$
A_{RV} (m s^{-1})	$\mathcal{U}[0, 100]$	$6.5^{+2.0}_{-1.5}$
B_{RV} (m s^{-1} d)	$\mathcal{U}[-100, 100]$	$28.2^{+7.0}_{-5.0}$
A_{FWHM} (m s^{-1})	$\mathcal{U}[0, 100]$	$44.8^{+11.3}_{-8.1}$
B_{FWHM} (m s^{-1} d)	$\mathcal{F}[0]$	0
A_{BIS} (m s^{-1})	$\mathcal{U}[-100, 100]$	$-1.9^{+1.4}_{-1.5}$
B_{BIS} (m s^{-1} d)	$\mathcal{U}[-100, 100]$	$-29.1^{+5.2}_{-7.5}$
Other parameters		
Stellar density, $\rho_\star$ (g cm^{-3})	$\mathcal{N}[3.17, 0.14]$	3.08 ± 0.13
K2 parametrized limb-darkening coefficient q_1	$\mathcal{U}[0, 1]$	$0.35^{+0.33}_{-0.20}$
K2 parametrized limb-darkening coefficient q_2	$\mathcal{U}[0, 1]$	$0.32^{+0.34}_{-0.19}$
Offset RV (km s^{-1})	$\mathcal{U}[-10, -9]$	-9.6484 ± 0.0026
Offset FWHM (km s^{-1})	$\mathcal{U}[7, 8]$	7.608 ± 0.018
Offset BIS (km s^{-1})	$\mathcal{U}[-0.5, 0.5]$	0.0386 ± 0.0011
Jitter term σ_{RV} (m s^{-1})	$\mathcal{J}[1, 100]$	$1.10^{+0.71}_{-0.72}$
Jitter term σ_{FWHM} (m s^{-1})	$\mathcal{J}[1, 100]$	$9.4^{+2.0}_{-2.1}$
Jitter term σ_{BIS} (m s^{-1})	$\mathcal{J}[1, 100]$	$6.66^{+0.61}_{-0.57}$
Jitter term σ_{K2} (ppm)	$\mathcal{J}[1, 100]$	81 ± 1.5

[a] $\mathcal{F}[a]$ refers to a fixed value *a*, $\mathcal{U}[a, b]$ to a uniform prior between *a* and *b*, $\mathcal{N}[a, b]$ to a Gaussian prior with mean *a* and standard deviation *b*, $\mathcal{B}[a, b]$ to a beta distribution with shape parameters *a* and *b*, and $\mathcal{J}[a, b]$ to the modified Jeffreys prior as defined by Gregory (2005, equation 16).

[b] Inferred parameters and errors are defined as the median and 68.3 per cent credible interval of the posterior distribution.

[c] Beta distribution to inform eccentricity sampling using the beta distribution for multiplanetary systems as defined by Van Eylen et al. (2019).
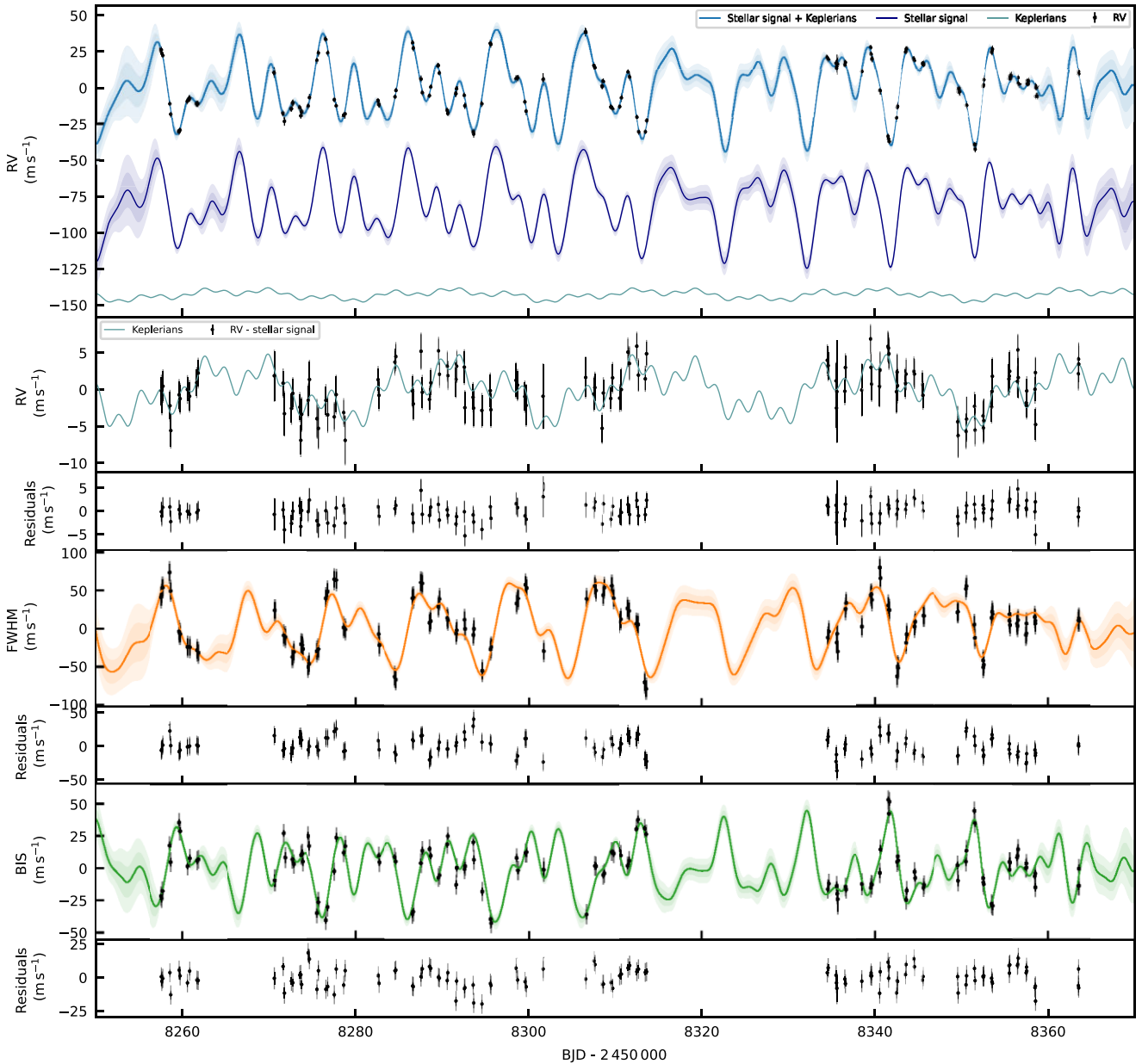
Figure 3. K2-233's RV, FWHM, and BIS time-series after being corrected by inferred offsets. The plot shows (from top to bottom) RV data together with full, stellar and planetary signal inferred models; RV data with stellar signal model subtracted; RV residuals; FWHM data together with inferred stellar model; FWHM residuals; BIS data together with inferred stellar model, and BIS residuals. Measurements are shown with black circles, error bars, and a semitransparent error bar extension accounting for the inferred jitter. The solid lines show the inferred full model coming from our multi-GP, light-shaded areas showing the corresponding GP model's 1σ and 2σ credible intervals. For the RV time-series (top panel), we also show the inferred stellar (dark blue line) and planetary (light green line) recovered signals with an offset for better clarity.

they are modelled only as an offset and a jitter term. Therefore, the difference in Bayesian evidence does not only reflect the model difference given by the planetary and activity model in the RVs, but it also is penalized by the suboptimal modelling of the stellar signal in the activity indicators. We therefore assume that this approach is not optimal to compare models with different time-series.

To show quantitatively the advantages of using the multi-GP, we show the different conclusions that we could draw for the different analyses regarding the RV detection of the planetary signals. The question we want to answer is 'can we claim an RV detection of the planetary signals in each analysis?' Also, using the BIC and Akaike information criterion (AIC), we select the preferred model. We note

that BIC and AIC have different properties (see e.g. Burnham & Anderson 2002; Chakrabarti & Ghosh 2011). BIC has been found more useful in helping to select the true model (assuming that the true model is in the sample of tested models), while the AIC is more appropriate in finding the best model (assuming that we do not know the true model). Since we do not believe that any of our models is the true one describing our data, AIC is the best metric to use. However, we will show BIC and AIC given the popularity of the former in the exoplanet comunity.

For the one-dimensional GP framework, we create two models. One model in which we fix the Keplerian signals to have an amplitude of zero, and another model where we sample for them. Both models
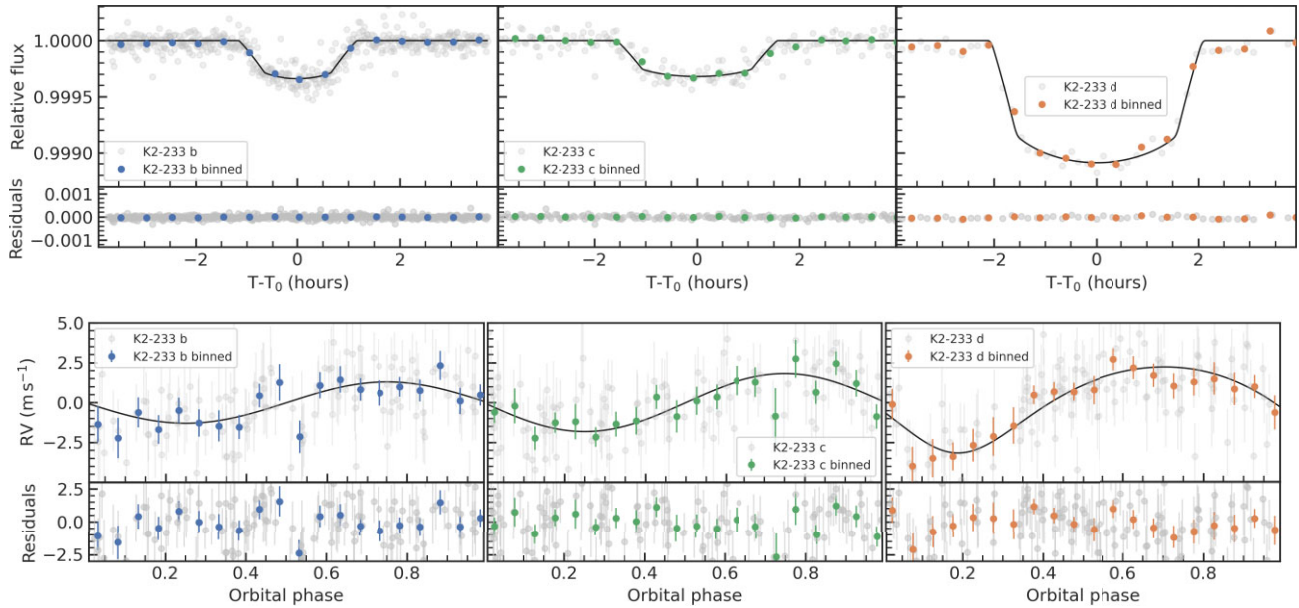
Figure 4. *Top panel*: Phase-folded light curves of K2-233 b (left), K2-233 c (centre), and K2-233 d (right). Nominal *K2* observations are shown in light grey. Solid colour circles represent 30-min binned data. Transit models are shown with a solid black line. The *x*- and *y*-axis in each panel are shown with the same range to facilitate the comparison of the scales of the transit signals. *Bottom panel*: Phase-folded RV signals for K2-233 b (left), K2-233 c (centre), and K2-233 d (right) following the subtraction of the systemic velocities, stellar signal, and other planets. Nominal RV observations are shown as light grey points. Solid colourful points show binned data to 1/20 of the orbital phase.
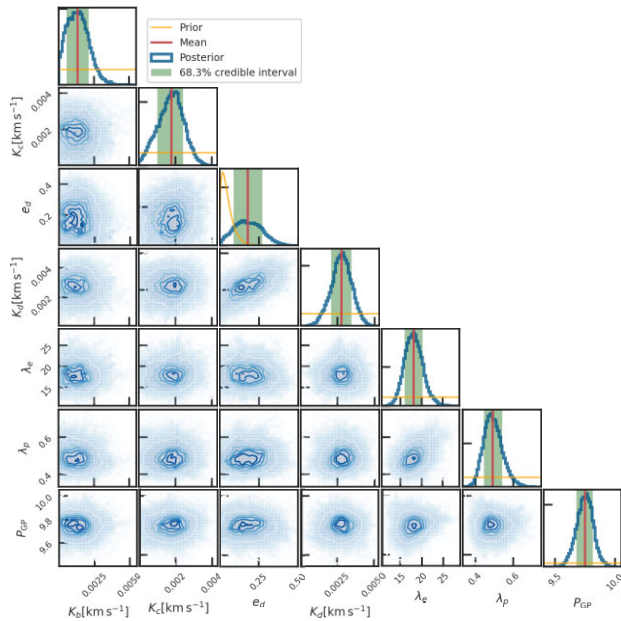
Figure 5. Posterior and correlation plots for some of the key sampled parameters of the joint analysis described in Section 3.3.

give similar likelihoods, but the former model is preferred with a ΔBIC = 19 and ΔAIC = 4. We therefore conclude that from the one-dimensional GP regression we cannot conclude that we detect the RV planetary signals (even if a priori we know they should be there).

We repeat the same experiment but now within the three-dimensional GP regression framework between RVs, FWHM, and BIS. In this case, the model including the RV planetary signals is

preferred over the model where they are zero with a ΔBIC = 3 and ΔAIC = 19. We thus conclude that with the multi-GP analysis we can claim, with a statistical significance, that the model including the planetary Doppler signals is preferred.

These analyses show how for the current data set, the only method that allows to significantly claim the detection of the planetary RV signals is the multi-GP regression. This shows how the multi-GP analysis offers a more robust way of modelling spectroscopic time-series of young active stars and detecting the RV planetary signals in them.

4.2 Dynamical analysis

The architecture of K2-233 is shaped for the orbits of K2-233 b and c being consistent with circular, and K2-233 d with a slight eccentricity of ∼0.18. This relatively small eccentricity of planet d can be reached by planet–disc interaction (e.g. Ragusa et al. 2018). We then performed an orbital stability analysis of the K2-233 system to test if this configuration of the system is stable. We did this using the software MERCURY6 (Chambers 1999). We assume that all planets have coplanar orbits and we use the median of the derived parameters in Tables 3 and 4 to create our MERCURY6 set-up. We evolved the system for 1 Gyr with steps of 0.1 d per integration.

We found negligible changes in the orbital parameters of the three planets, except for the eccentricities that had fluctuations. K2-233 b and c presented eccentricity fluctuations of a maximum of 0.12, while K2-233 d's eccentricity fluctuations were in the 0.01 level. Therefore, we conclude that our orbital solution for the K2-233 system is consistent with a dynamically stable configuration.

4.3 Composition of the planets

Fig. 6 shows a mass–radius diagram for small exoplanets ($1 < R_{\rm p} < 4.5\,R_{\oplus}$ and $1 < M_{\rm p} < 32\,M_{\oplus}$). The plot also shows two-layer

Table 4. Derived parameters for the K2-233 planets.

Parameter	K2-233 b's values	K2-233 c's values	K2-233 d's values
Planet mass M_p ($M_\oplus$)	$2.4^{+1.5}_{-1.3}$	$4.6^{+1.8}_{-1.7}$	$10.3^{+2.4}_{-2.6}$
Planet radius R_p ($R_\oplus$)	1.315 ± 0.036	1.272 ± 0.038	2.363 ± 0.062
Planet density ρ_p (g cm^{-3})	$5.7^{+3.7}_{-3.2}$	$12.4^{+5.1}_{-4.6}$	4.3 ± 1.1
Scaled semimajor axis $a/R_\star$	9.97 ± 0.14	20.10 ± 0.28	45.91 ± 0.65
Semimajor axis a (au)	0.03293 ± 0.00066	0.0664 ± 0.0013	0.1516 ± 0.0030
Time of periastron passage T_p (BJD–2450000)	7991.6897 ± 0.0013	$7586.862^{+0.027}_{-0.028}$	$8009.27^{+1.43}_{-0.86}$
Orbit inclination i_p (°)	$88.86^{+0.72}_{-0.66}$	$89.57^{+0.29}_{-0.35}$	$89.62^{+0.24}_{-0.18}$
Transit duration τ_{14} (h)	$1.882^{+0.037}_{-0.047}$	$2.687^{+0.051}_{-0.064}$	$3.791^{+0.102}_{-0.085}$
Planet surface gravity g_p (cm s^{-2})[a]	1334^{+832}_{-757}	2792^{+1148}_{-1025}	1794^{+437}_{-443}
Planet surface gravity g_p (cm s^{-2})[b]	1343^{+845}_{-762}	2824^{+1143}_{-1044}	1811^{+443}_{-451}
Equilibrium temperature T_{eq} (K)[c]	1127 ± 14	794 ± 10	525 ± 6
Received irradiance ($F_\oplus$)	269 ± 13	66.2 ± 3.3	12.7 ± 0.6
TSM[d]	$6.6^{+8.7}_{-2.6}$	$2.1^{+1.3}_{-0.7}$	$27.1^{+9.0}_{-5.4}$

[a]Derived using $g_p = G M_p R_p^{-2}$.

[b]Derived using sampled parameters following Southworth, Wheatley & Sams (2007).

[c]Assuming a zero albedo.

[d]Transmission spectroscopy metric (TSM) by Kempton et al. (2018).
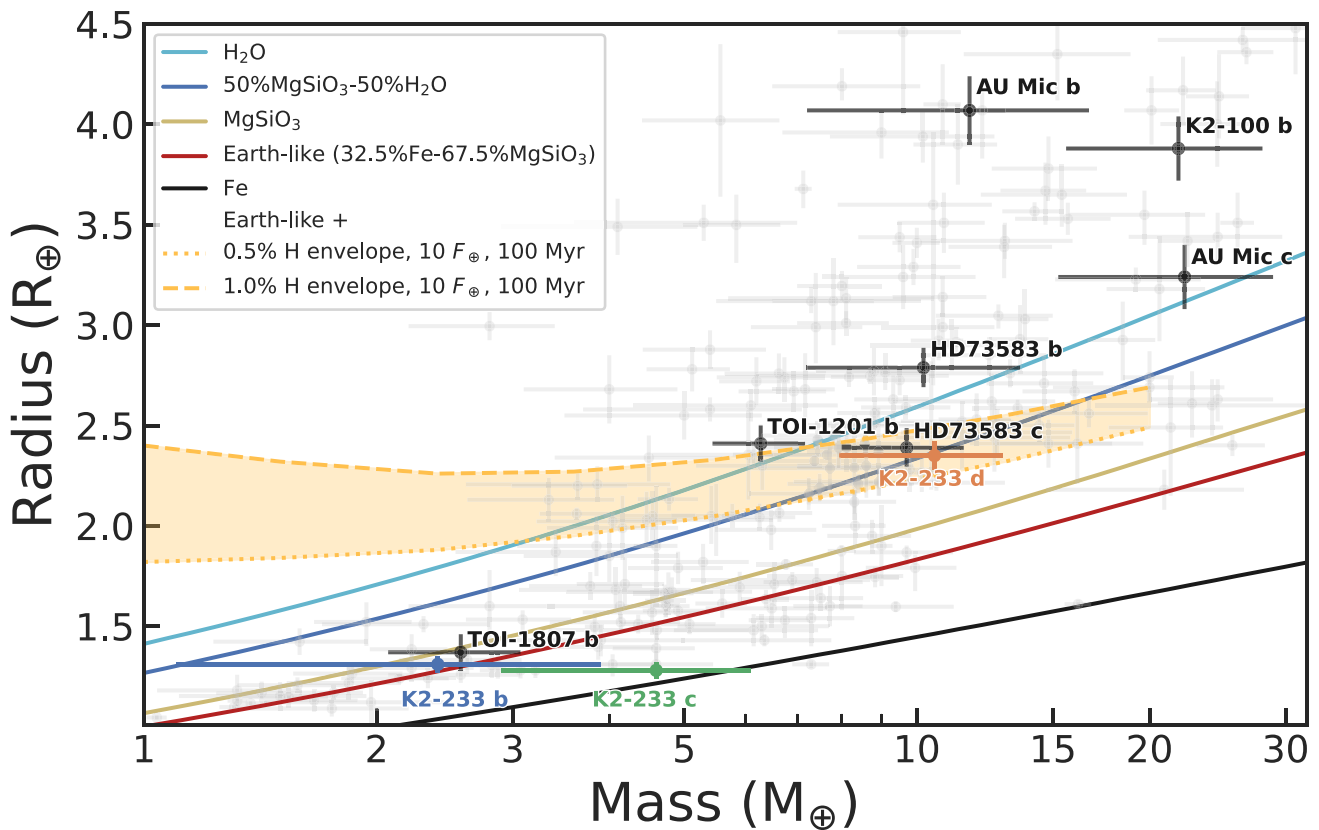
Figure 6. Mass versus radius diagram for small exoplanets ($1 < R_p < 4\,R_\oplus$ and $1 < M_p < 32\,M_\oplus$). Grey points with error bars show planets with mass and radius measurements better than 30 per cent (as in the NASA Exoplanet Archive on 2023 February 2, https://exoplanetarchive.ipac.caltech.edu/; Akeson et al. 2013). Black points and labels refer to young exoplanets (<1 Gyr) with mass and radius measurements. K2-233 b, c, and d are shown with blue, green, and orange circles with corresponding labels identifying each planet. Solid lines represent two-layer models as given by Zeng et al. (2016) with a different colour corresponding to a different mixture of elements. Non-solid lines correspond to rocky cores surrounded by a Hydrogen envelope with 0.5 per cent (dotted line) and 1 per cent (dashed line) on mass for 100-Myr planets with an insolation of $10\,F_\oplus$ (Lopez & Fortney 2014). This plot was created using the same code used to create the mass–radius diagram in Barragán et al. (2018).

exoplanet models by Zeng, Sasselov & Jacobsen (2016), together with the Earth-like interior plus Hydrogen envelope models given by Lopez & Fortney (2014) for 100-Myr old planets receiving an insolation of 10 $F_\oplus$.

With a mass of $10.3^{+2.4}_{-2.6}$ $M_\oplus$ and radius of 2.363 ± 0.062 $R_\oplus$, K2-233 d has a density of 4.3 ± 1.1 g cm^{-3}, placing it in the water world composition line regime (see Fig. 6). Therefore, we can say that K2-233 d could be consistent with a water-rich world that consists of 50 per cent water ice with a small 50 per cent of silicates. However, K2-233 d is also consistent with a planet made of an Earth-like interior, surrounded by a Hydrogen envelope that could account for 0.5–1 per cent of its mass (taking into account its age and insolation). It is important to note that K2-233 d lies well above the small planet radius valley (a lack of planets between 1.5 and 2 $R_\oplus$; Fulton et al. 2017). We therefore expect that the planet has a volatile envelope rather than being a solid water world (as suggested by previous works e.g. Fulton et al. 2017; Van Eylen et al. 2018). Despite the degeneracy in composition for K2-233 d inferred from the mass–radius diagram, for the remainder of the discussion in this manuscript, we will assume that K2-233 d is a world with a Hydrogen-rich volatile envelope.

The other two planets, K2-233 b and c, lie below the radius valley. Therefore, we would expect them to have a rocky composition (as discussed by Lillo-Box et al. 2020). Owen & Wu (2017) posit that the position of the radius valley points to a universal Earth-like rock and iron core composition. Beyond that, it is interesting to compare the expected composition between these two innermost planets given their similar size. The measured mass and radius of K2-233 c present a relatively high density of $12.4^{+5.1}_{-4.6}$ g cm^{-3}. We can see in Fig. 6 that, at face value, K2-233 c is consistent with a planet made principally of iron with a small fraction of silicates. This result is not unexpected, and it is consistent with the dispersion of exoplanets from that part of the mass–radius diagram. We must be cautious when interpreting our results for K2-233 b, given the marginal detection of the signal in the RV time-series. Its location in the mass–radius diagram suggests a lower density (hence more silicates and less iron) than K2-233 c, but the two planets' densities are actually consistent at the 1σ level. Therefore, we cannot claim to have detected a significant difference in composition between the two small planets in the system.

Fig. 6 also shows the position of other young small planets with mass and radius measurements. It is worth noting that K2-233 is the only young system with characterized planets on both sides of the radius valley. Therefore, this system is a valuable laboratory for comparative studies of planetary evolution and formation, and to test theories such as photoevaporation (e.g. Adams & Laughlin 2006; Raymond et al. 2009; Owen & Wu 2013; Lopez & Fortney 2014) and core-powered mass-loss (e.g. Ginzburg, Schlichting & Sari 2016; Gupta & Schlichting 2019, 2021). For the rest of this discussion section, we will assume that K2-233 b and c are solid rocky worlds, while K2-233 d is a planet with a significant volatile envelope accounting for most of its radius. None the less, we note that further observations are needed in order to reduce the mass uncertainties and pinpoint the true composition of these worlds more precisely.

4.4 Photoevaporation and the radius valley

The true origin of the small planet radius valley is still disputed. Gas-poor formation and a subsequent distinct, inherently rocky super-Earth population has mostly been ruled out, considering the slope of the radius valley (Van Eylen et al. 2018; Petigura et al. 2022). However, atmospheric erosion due to (i) photoevaporation (Owen & Wu 2017; Van Eylen et al. 2018) and (ii) core-powered mass-loss

(Ginzburg et al. 2016; Gupta & Schlichting 2019; Berger et al. 2023) are both theories that could explain the observed underabundance of planets sized between ∼1.5 and 2 $R_\oplus$.

We note that K2-233 has three, mutually aligned, close-in planets. Therefore, we expect that they migrated through the disc, and hence have been on close orbits since ≲5 Myr and have been suffering from high levels of stellar radiation since then (see e.g. Kley & Nelson 2012). Since arriving at their present-day orbits, the planets will have been subjected to high levels of stellar irradiation. Theory predicts photoevaporation to be more effective at driving hydrodynamic escape during the first ∼100 Myr after formation (e.g. Owen & Wu 2017). At an age of 360 Myr (David et al. 2018), the K2-233 system is thus ideal to test the role of photoevaporation in sculpting the radius valley. On the other hand, we would not expect the planets in the K2-233 system to have suffered a significant atmospheric loss due to core-powered mass-loss, since this mechanism operates over Gyr time-scales (Ginzburg et al. 2016). The characteristics of the inner K2-233 system are suggestive of a history of significant atmospheric mass-loss, which would support the photoevaporation evolution scenario. A more conclusive test, however, would be to search for signatures of ongoing evaporation, for example by searching for escaping Helium in transmission (e.g. Zhang et al. 2022).

4.5 Atmospheric characterization perspectives

Well-determined masses and radii are crucial in order to perform further follow-up analysis of exoplanets, such as atmospheric characterization. In particular, Batalha et al. (2019) show that a precision of at least 50 per cent in the mass determination is required to enable even a basic retrieval of atmospheric properties from transmission spectroscopic observations with *JWST* or *Hubble Space Telescope* (*HST*), while a precision of 20 per cent or better is recommended to enable a more detailed atmospheric characterization. Fig. 7 shows a mass versus mass uncertainty plot for small planets. Thanks to the updated mass measurements obtained in this work, both K2-233 c and d now have sub-50 per cent uncertainties on their masses, which meet the basic criterion of Batalha et al. (2019). We also computed the transmission spectroscopy metric (TSM; Kempton et al. 2018) of each planet, and report it in Table 4. Due to their small radii, transmission spectroscopy observations of planets b and c would be challenging (for example, multiple transits of each planet would be required with *JWST*). Sub-Neptune K2-233 d is likely to host a H/He-dominated atmosphere. Based on this assumption, the TSM of planet d, while modest, is on par with that of K2-18b, which was recently characterized with *HST* (Benneke et al. 2019). We note that the recovered mild eccentricity of K2-233 d is not expected to be sufficient to induce observable time-variable atmospheric effects. Studies to date have shown that only high eccentricities have the potential to impact the atmospheric temperature profile and chemistry (e.g. Langton & Laughlin 2008; Tsai et al. 2023).

Theory predicts a wide diversity for sub-Neptune atmospheres (Moses et al. 2013; Guzmán-Mesa et al. 2022). Observations to date have presented confounding results, some displaying clear water absorption features and others a featureless spectrum (Knutson et al. 2014; Kreidberg et al. 2014; Benneke et al. 2019). Thick cloud and haze layers are thought to be more dominant for these smaller planets (than the Jupiter population), and this is the accepted explanation for dampened spectral features. At $T_{eq,d} = 525 \pm 6$ K, we would expect clouds to play a role in the atmosphere of K2-233 d, and this would have to be considered in the planning of atmospheric observations. That said, adding to the sample of observed sub-Neptunes is vital to
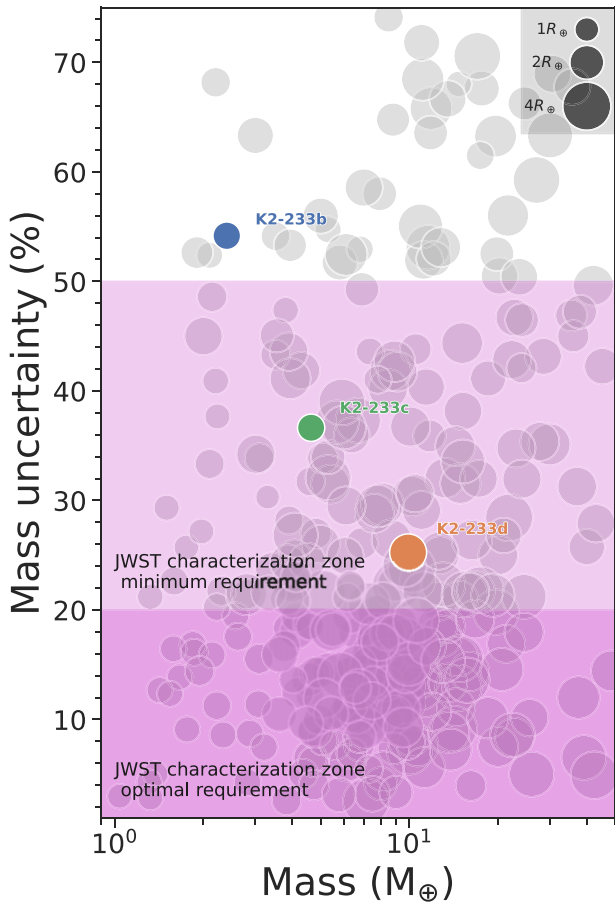
Figure 7. Mass uncertainty versus mass plot for small exoplanets ($1 < R_{\rm p} < 4\,{\rm R}_\oplus$ and $1 < M_{\rm p} < 50\,{\rm M}_\oplus$). Grey circles correspond to planets with mass and radius measurements (as in the NASA Exoplanet Archive on 2023 February 2; Akeson et al. 2013). K2-233 b, c, and d are shown with blue, green, and orange circles with corresponding labels identifying each planet. Circle sizes correspond to planet radii, with the scale displayed in the upper right corner. The two shaded regions denote the minimum (lighter) and optimal (darker) regions for atmospheric characterization with *JWST*.

understand this mysterious population, of which there is no analogue in the Solar system. Furthermore, constraining the volatile content will inform formation process for close-in small planets (Bean, Raymond & Owen 2021).

4.6 On follow-up observations

K2-233 is a keystone target that can help to understand the evolution of multiplanetary systems, but additional follow-up observations are needed to improve the precision of the planetary and orbital parameters. In particular, additional RV monitoring to refine the planet masses would be useful, as the results presented in this paper suggest that K2-233 b and c have significantly different masses, although their radii are nearly identical. This could arise from a discrepancy in composition between the two planets, which would challenge standard models of small-planet formation and evolution, but with the current data we cannot rule out that this is merely a statistical fluctuation.

How many more RV measurements would we need to improve the mass measurements of K2-233b and c to the point where we can confirm or rule out this tentative mass discrepancy? To answer this

question, we used CITLALATONAC (Barragán et al. 2022a) to simulate multiple realizations of additional HARPS-like observations and analysed them in the same way as we analysed the existing data in this paper. We used the GP hyperparameters obtained in Table 3 to create the stellar signal as samples of a multi-GP. We simulated synthetic time-series of activity-induced RV, FWHM, and BIS and added the expected Keplerian signals for the three planets and sampled the data assuming that K2-233 is observed from La Silla with a maximum airmass of 1.5 in an intensive 3-months campaign, and added realistic white and instrumental red noise (based on the existing HARPS data). We then attempted to recover the stellar and planetary signals, using the same multi-GP framework and combining the existing and simulated data, as we have done with the available real data. By varying the number of synthetic observations, we established that we need around 50 epochs spread over 3 months to model the stellar signal robustly and to detect the signal of planet b to 3σ confidence or better, and improve the detection of planets c and d.

As well as spectroscopic data, more photometric data are also needed to refine the radii of the three planets and to keep track of the planet ephemerides. The latter is crucial in order to perform any transmission spectroscopy effort. Unfortunately, K2-233 has not been observed by *TESS* and according to the Web *TESS* viewing tool,[3] the star does not fall in any of the planned sectors until sector 69. This is due to the preferred orientation of *TESS* towards the poles and K2-233 being an ecliptic target. We also checked whether K2-233 could be observable with the *CHaracterising ExOPlanet Satellite* (*CHEOPS*; Benz et al. 2021). According to the *CHEOPS* Observers Manual,[4] K2-233 is observable each year for $\sim$70 d (between April and May) with an observing efficiency of 50 per cent.[5] Thanks to its exquisite photometric precision, *CHEOPS* could be used to observe the transits of all K2-233 planets, including the two small ones. It is also tempting to think about ground-based photometric observations of K2-233. However, the shallow transits of K2-233 b and c are not observable from the ground. K2-233 d would be challenging but observable with state-of-the-art facilities, such as the Next Generation Transit Survey (Wheatley et al. 2018).

5 CONCLUSIONS

We reanalysed the spectroscopic observations of K2-233 published originally by Lillo-Box et al. (2020). We used the activity indicators to constrain the stellar signal in the RV time-series using a multi-GP approach. This led to an improvement in the detection of the Doppler signals with respect to the values reported in Lillo-Box et al. (2020). These improvements in the planetary mass measurements allow us to put better constraints on these planets' compositions.

The outermost planet, K2-233 d, is consistent with a volatile-rich world, while K2-233 b and c are likely rocky, potentially hosting a volatile envelope of an insignificant mass fraction. Despite their similar sizes, K2-233 b and c seem to have different masses, suggesting a surprising discrepancy in composition between the two planets. However, further RV monitoring is needed to confirm this tentative discrepancy. Having improved the precision of the planetary masses, we add the K2-233 planets to the small sample of young planets amenable to space-based transmission observations.

[3] https://heasarc.gsfc.nasa.gov/cgi-bin/tess/webtess/wtv.py.
[4] https://www.cosmos.esa.int/web/cheops-guest-observers-programme/cheops-observers-manual.
[5] Each *CHEOPS* observation has an observing efficiency, which is the fraction of time on target that is not interrupted due to its particular orbit.

ACKNOWLEDGEMENTS

DATA AVAILABILITY

The codes used in this manuscript are freely available at https://github.com/oscaribv. The spectroscopic measurements that appear in Table 2 are available as supplementary material in the online version of this manuscript.

REFERENCES

Adams F. C., Laughlin G., 2006, ApJ, 649, 1004
Ahrer E. et al., 2021, MNRAS, 503, 1248
Aigrain S., Foreman-Mackey D., 2022, preprint (arXiv:2209.08940)
Aigrain S., Pont F., Zucker S., 2012, MNRAS, 419, 3147
Akeson R. L. et al., 2013, PASP, 125, 989
Ambikasaran S., Foreman-Mackey D., Greengard L., Hogg D. W., O'Neil M., 2015, IEEE Trans. Pattern Anal. Mach. Intell., 38, 252
Astropy Collaboration, 2013, A&A, 558, A33
Astropy Collaboration, 2018, AJ, 156, 123
Astropy Collaboration, 2022, ApJ, 935, 167
Baranne A. et al., 1996, A&AS, 119, 373
Barragán O. et al., 2018, A&A, 612, A95
Barragán O., Gandolfi D., Antoniciello G., 2019a, MNRAS, 482, 1017
Barragán O. et al., 2019b, MNRAS, 490, 698
Barragán O., Aigrain S., Rajpaul V. M., Zicher N., 2022a, MNRAS, 509, 866
Barragán O. et al., 2022b, MNRAS, 514, 1606
Batalha N. E., Lewis T., Fortney J. J., Batalha N. M., Kempton E., Lewis N. K., Line M. R., 2019, ApJ, 885, L25
Bean J. L., Raymond S. N., Owen J. E., 2021, J. Geophys. Res. (Planets), 126, e06639
Benneke B. et al., 2019, ApJ, 887, L14
Benz W. et al., 2021, Exp. Astron., 51, 109
Berger T. A., Schlieder J. E., Huber D., Barclay T., 2023, preprint (arXiv:2302.00009)
Boisse I. et al., 2009, A&A, 495, 959
Bouma L. G. et al., 2020, AJ, 160, 239
Burnham K., Anderson D., 2002, Model Selection and Multimodel Inference: A Practical Information-theoretic Approach. Springer-Verlag, Berlin
Chakrabarti A., Ghosh J. K., 2011, in Bandyopadhyay P. S., Forster M. R., eds, Handbook of the Philosophy of Science, Vol. 7, Philosophy of Statistics. Elsevier, Amsterdam, p. 583
Chambers J. E., 1999, MNRAS, 304, 793
Cutri R. M. et al., 2003, VizieR Online Data Catalog, p. II/246
Dai F., Masuda K., Winn J. N., Zeng L., 2019, ApJ, 883, 79
David T. J. et al., 2018, AJ, 155, 222
Díaz R. F., Cincunegui C., Mauas P. J. D., 2007, MNRAS, 378, 1007

Dumusque X., Boisse I., Santos N. C., 2014, ApJ, 796, 132
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Fulton B. J. et al., 2017, AJ, 154, 109
Fulton B. J., Petigura E. A., Blunt S., Sinukoff E., 2018, PASP, 130, 044504
Gaia Collaboration, 2020, VizieR Online Data Catalog, p. I/350
Ginzburg S., Schlichting H. E., Sari R., 2016, ApJ, 825, 29
Gomes da Silva J., Santos N. C., Bonfils X., Delfosse X., Forveille T., Udry S., 2011, A&A, 534, A30
Gregory P. C., 2005, ApJ, 631, 1198
Grunblatt S. K., Howard A. W., Haywood R. D., 2015, ApJ, 808, 127
Gupta A., Schlichting H. E., 2019, MNRAS, 487, 24
Gupta A., Schlichting H. E., 2021, MNRAS, 504, 4634
Guzmán-Mesa A., Kitzmann D., Mordasini C., Heng K., 2022, MNRAS, 513, 4015
Harris C. R. et al., 2020, Nature, 585, 357
Haywood R. D. et al., 2014, MNRAS, 443, 2517
Hobson M. J. et al., 2021, AJ, 161, 235
Høg E. et al., 2000, A&A, 355, L27
Howell S. B. et al., 2014, PASP, 126, 398
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Kempton E. M. R. et al., 2018, PASP, 130, 114401
Kipping D. M., 2010, MNRAS, 408, 1758
Kipping D. M., 2013, MNRAS, 435, 2152
Klein B. et al., 2022, MNRAS, 512, 5067
Kley W., Nelson R. P., 2012, ARA&A, 50, 211
Knutson H. A., Benneke B., Deming D., Homeier D., 2014, Nature, 505, 66
Kreidberg L. et al., 2014, Nature, 505, 69
Langton J., Laughlin G., 2008, ApJ, 674, 1106
Lillo-Box J. et al., 2020, A&A, 640, A48
Lopez E. D., Fortney J. J., 2014, ApJ, 792, 1
Lovis C., Pepe F., 2007, A&A, 468, 1115
Mandel K., Agol E., 2002, ApJ, 580, L171
Mann A. W. et al., 2016a, AJ, 152, 61
Mann A. W. et al., 2016b, ApJ, 818, 46
Mann A. W. et al., 2022, AJ, 163, 156
Martioli E., Hébrard G., Correia A. C. M., Laskar J., Lecavelier des Etangs A., 2021, A&A, 649, A177
Mayo A. W. et al., 2019, AJ, 158, 165
Mayor M. et al., 2003, Messenger, 114, 20
Moses J. I. et al., 2013, ApJ, 777, 34
Nardiello D. et al., 2022, A&A, 664, A163
Newton E. R. et al., 2019, ApJ, 880, L17
Newton E. R. et al., 2021, AJ, 161, 65
Nicholson B. A., Aigrain S., 2022, MNRAS, 515, 5251
Oshagh M. et al., 2017, A&A, 606, A107
Owen J. E., Wu Y., 2013, ApJ, 775, 105
Owen J. E., Wu Y., 2017, ApJ, 847, 29
Pepe F., Mayor M., Galland F., Naef D., Queloz D., Santos N. C., Udry S., Burnet M., 2002, A&A, 388, 632
Petigura E. A. et al., 2022, AJ, 163, 179
Queloz D. et al., 2001, A&A, 379, 279
Ragusa E., Rosotti G., Teyssandier J., Booth R., Clarke C. J., Lodato G., 2018, MNRAS, 474, 4460
Rajpaul V., Aigrain S., Osborne M. A., Reece S., Roberts S., 2015, MNRAS, 452, 2269
Rajpaul V. M. et al., 2021, MNRAS, 507, 1847
Raymond S. N., Barnes R., Veras D., Armitage P. J., Gorelick N., Greenberg R., 2009, ApJ, 696, L98
Ricker G. R. et al., 2015, J. Astron. Telesc. Instrum. Syst., 1, 014003
Rizzuto A. C. et al., 2020, AJ, 160, 33
Southworth J., Wheatley P. J., Sams G., 2007, MNRAS, 379, L11
Suárez Mascareño A. et al., 2021, Nat. Astron., 6, 232
The pandas development team, 2020, pandas-dev/pandas: Pandas, available at https://doi.org/10.5281/zenodo.3509134
Tsai S.-M., Steinrueck M., Parmentier V., Lewis N., Pierrehumbert R., 2023, MNRAS, 520, 3867
Vanderburg A., Johnson J. A., 2014, PASP, 126, 948

Van Eylen V., Agentoft C., Lundkvist M. S., Kjeldsen H., Owen J. E., Fulton B. J., Petigura E., Snellen I., 2018, MNRAS, 479, 4786
Van Eylen V. et al., 2019, AJ, 157, 61
Wheatley P. J. et al., 2018, MNRAS, 475, 4476
Wilson O. C., 1968, ApJ, 153, 221
Winn J. N., 2010, preprint (arXiv:1001.2010)
Zeng L., Sasselov D. D., Jacobsen S. B., 2016, ApJ, 819, 127
Zhang M., Knutson H. A., Wang L., Dai F., Barragán O., 2022, AJ, 163, 67
Zicher N. et al., 2022, MNRAS, 512, 3060

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

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: CHARACTERIZING THE STELLAR SIGNAL

We perform a one-dimensional GP regression of different activity indicators to study the shape of the stellar signal in different time-series $\mathcal{A}_k$. We intend to analyse which time-series behave photometric- (that can be described only with G) and RV-like (that also need to be described with $\dot{G}$).

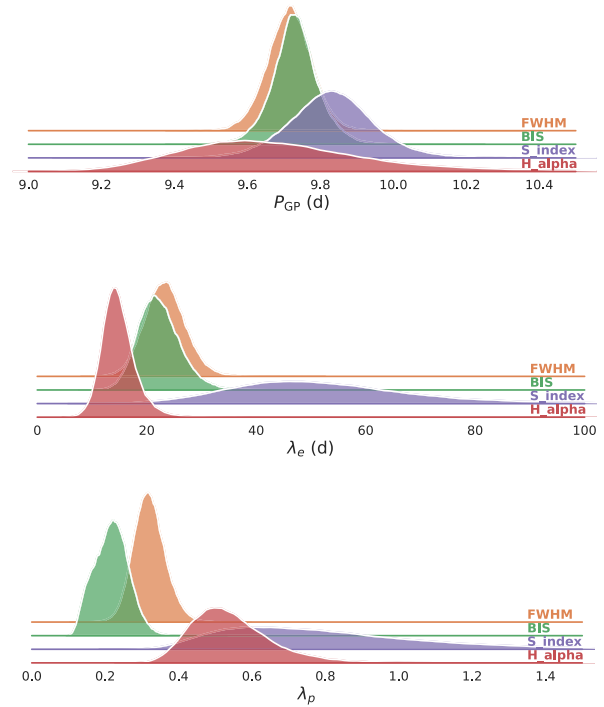We model the covariance between two times t_i and t_j for each time-series $\mathcal{A}_k$ with the function

$$\gamma_{\mathrm{1D}} = A^2 \gamma_{i,j}, \tag{A1}$$

where A is an amplitude term, and $\gamma_{i,j}$ is the QP kernel given in equation (2). We use the code PYANETI (Barragán et al. 2019a, 2022a) to create posterior distributions of the QP kernel hyperparameters.

We sample for six parameters in each run: four GP hyperparameters (A, P_{GP}, λ_{e}, and λ_{p}), one offset, and a jitter term. We set wide uniform priors for all the parameters. In particular, for λ_{e} we set a uniform prior between 0 and 500 d, for λ_{p} between 0.1 and 10, and for P_{GP} between 8 and 12 d. We sample the parameter space using the built-in MCMC sampler in PYANETI. The posterior creation follows the same guidelines as in Section 3.1.

Table A1 shows the inferred P_{GP}, λ_{e}, and λ_{p} hyperparameters for all time-series. Fig. A1 shows the inferred posterior distributions for the P_{GP}, λ_{e}, and λ_{p} parameters for all the modelled time-series, except for Na. Fig. A2 shows the data and inferred models for the different time-series.

The first thing to note is that the Na does not behave as a QP signal in Fig. A2. Díaz, Cincunegui & Mauas (2007) described how the Na D1 and D2 lines could be used as chromospheric activity indicator for active late-type stars ($B - V > 1.1$), typically M dwarfs (Gomes da Silva et al. 2011), but not for earlier stars as K2-233 (B

Table A1. Recovered hyperparameters for one-dimensional GP regressions.

Time-series	P_{GP} [d]	λ_{e} [d]	λ_{p}
FWHM	9.72 ± 0.06	$23.3^{+3.8}_{-3.6}$	0.32 ± 0.04
BIS span	9.73 ± 0.06	$22.0^{+4.1}_{-3.5}$	0.22 ± 0.05
S_{HK}	9.83 ± 0.11	$50.5^{+18.4}_{-14.1}$	$0.75^{+0.33}_{-0.22}$
Hα	$9.65^{+0.28}_{-0.22}$	$14.5^{+3.1}_{-2.5}$	$0.53^{+0.12}_{-0.09}$
Na	$9.54^{+0.98}_{-0.81}$	$10.9^{+2.1}_{-2.0}$	$2.23^{+1.17}_{-0.70}$

Figure A1. Posterior distributions for P_{GP}(top), λ_{e} (middle), and λ_{p} (bottom). Results for FWHM (orange), BIS (green), S_{HK} (purple), and Hα (red) shown for each subpanel.
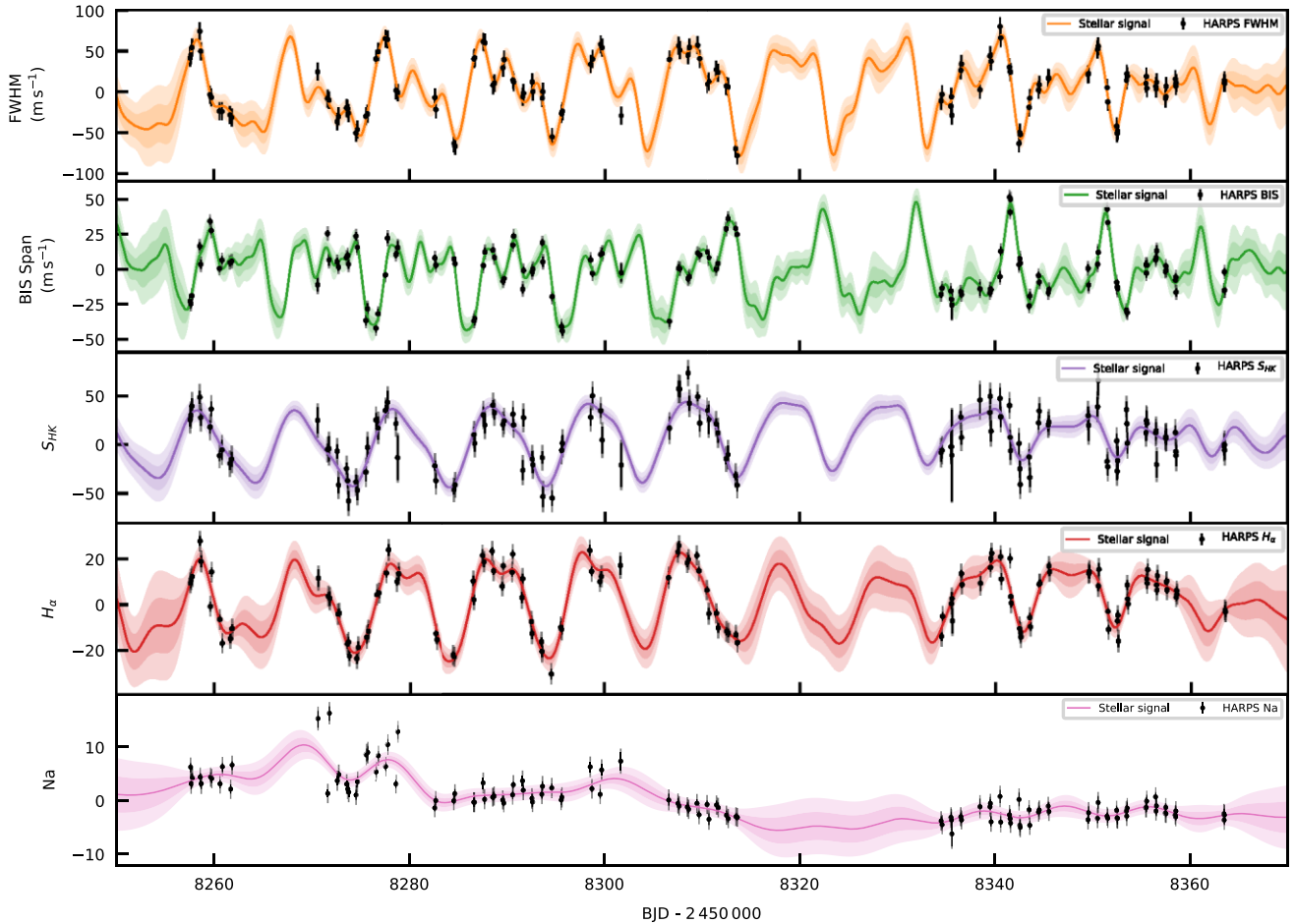
$- V \approx 0.9$). This suggests that the Na is not a good activity indicator to constrain K2-233's stellar activity. This also shows how not all activity indicators are good tracers of the stellar signal that is also contained in the RV data. From now on, we will exclude the Na time-series from our analysis.

From Fig. A1 (and Table A1), we can see that all time-series provide values for P_{GP} that agree well with each other. These values are also in agreement with the stellar rotation period of $9.754 \pm 0.038\,\mathrm{d}$ reported by David et al. (2018) from the *K2* photometric analysis. This suggests that all these activity indicators are good tracers of the stellar rotation period.

When we look at the recovered values for the inverse of the harmonic complexity, λ_{p}, we can see that we have different values for different activity indicators (see Fig. A1). First, we note that the recovered λ_{p} for the BIS time-series is consistently smaller (i.e. higher harmonic complexity) than the other activity indicators. As discussed in Aigrain, Pont & Zucker (2012) and Barragán et al. (2022a, b), it is expected that the BIS time-series is *RV-like*, meaning that it depends on the change of location of the active regions from the red- to the blue-shifted stellar hemisphere, and vice versa. We then see that the FWHM peaks around a value of 0.3, implying a lower harmonic complexity than the BIS time-series (see Table A1). This behaviour is also expected, given that previous analyses suggest that these time-series are good tracers of the areas covered by active regions on the stellar surface (e.g. Oshagh et al. 2017). For this reason, we call these kinds of activity indicators *photometry-like*. It is worth noting that these photometric and RV-like behaviours are visible in Fig. A2. The BIS time-series behaves visually as the derivative of the FWHM time-series. This is expected and it is the base of FF' models as the one used in the multi-GP approach (see Aigrain et al. 2012; Rajpaul et al. 2015; Barragán et al. 2022a). Finally, we note that the

Figure A2. K2-233 time-series (from top to bottom) for FWHM, BIS span, S_{HK}, Hα, and Na. The corresponding measurements are shown with black circles with error bars with a semitransparent error bar extension accounting for the inferred jitter. Solid coloured lines show the corresponding inferred signal coming from our GP regression, while light coloured shaded areas show the 1σ and 2σ credible intervals of the corresponding GP model.

recovered processes for the chromospheric activity indicators, Hα and S_{HK}, have a relatively low harmonic complexity (i.e. values of λ_p that are relatively large). This is unexpected given that both Hα and S_{HK} are expected to be photometry-like activity indicators. Therefore, we would have expected that they would behave similar to FWHM. Moreover, from Fig. A2, we can see that data in these two time-series have a relatively high white noise in comparison with the amplitude of the recovered signal. This significant white noise can be explained because the star is relatively faint ($V \sim 10.88$ mag). We speculate that the relatively high white noise in these time-series degrades our ability to recover complex patterns (i.e. higher harmonic complexity) of the stellar signal. This means that even if the underlying process has a significant harmonic complexity, we cannot recover it due to the suboptimal data.

We then analyse the long-term evolution time-scale, λ_e (see Table A1). Both the FWHM and the BIS time-series give a λ_e that peaks around $\sim$22 d, that is twice the P_{GP}. This implies that the local coherence of the periodic signal lasts two rotation periods. When analysing the recovered λ_e for the chromospheric activity indicators, we notice that they also produce different values when compared with the CCF ones (see Table A1).

A priori, we expect that FWHM, S_{HK}, and Hα behave as photometric-like activity indicators and that BIS behaves as RV-

like activity indicators. So, we would have expected that all activity indicators trace similar P_{GP} and λ_e, while we would suppose a higher harmonic complexity for RV-like activity indicators. This is the general tendency that we observe, but with significant annotations. First, it seems that all activity indicators are good tracers of the rotational period of the star. This is expected given that the stellar rotation is the dominant imprint of the stellar signal in the spectroscopic time-series. Secondly, more subtle characteristics of the stellar signal, which can be traced with λ_e and λ_p, seem to vary between different time-series. We hypothesize that this difference is caused by a degradation of the recovered stellar signal due to difference in data quality for each time-series. But we note that this difference can also be due to the fact that activity indicators are mapping different areas of the stellar surface, as well as we are reaching the limits of the model based on an FF'-like relation between the RVs and activity indicators.

From these analysis and discussion, we will assume that, for this specific system, the CCF diagnostics are better activity indicators than the chromospheric data to characterize the stellar signal in the spectroscopic time-series. We therefore use the FWHM and BIS time-series for the multi-GP analyses presented in Section 3.2.2.

This paper has been typeset from a TeX/LaTeX file prepared by the author.