Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **507,** 1847–1868 (2021) https://doi.org/10.1093/mnras/stab2192
Advance Access publication 2021 July 30

A HARPS-N mass for the elusive Kepler-37d: a case study in disentangling stellar activity and planetary signals

V. M. Rajpaul [1,2★] L. A. Buchhave [3] G. Lacedelli [4,5] K. Rice [6,7] A. Mortier [1,8] L. Malavolta [4,5] S. Aigrain,[2] L. Borsato [5] A. W. Mayo [9] D. Charbonneau,[10] M. Damasso,[11] X. Dumusque,[12] A. Ghedina,[13] D. W. Latham,[10] M. López-Morales,[10] A. Magazzù,[13] G. Micela,[14] E. Molinari,[15] F. Pepe,[12] G. Piotto,[4,5] E. Poretti,[13,16] S. Rowther [17,18] A. Sozzetti [11] S. Udry[12] and C. A. Watson[19]

Affiliations are listed at the end of the paper

Accepted 2021 July 23. Received 2021 July 12; in original form 2021 May 29

ABSTRACT

To date, only 18 exoplanets with radial velocity (RV) semi-amplitude <2 m s^{-1} have had their masses directly constrained. The biggest obstacle to RV detection of such exoplanets is variability intrinsic to stars themselves, e.g. nuisance signals arising from surface magnetic activity such as rotating spots and plages, which can drown out or even mimic planetary RV signals. We use Kepler-37 – known to host three transiting planets, one of which, Kepler-37d, should be on the cusp of RV detectability with modern spectrographs – as a case study in disentangling planetary and stellar activity signals. We show how two different statistical techniques – one seeking to identify activity signals in stellar spectra, and another to model activity signals in extracted RVs and activity indicators – can each enable a detection of the hitherto elusive Kepler-37d. Moreover, we show that these two approaches can be complementary, and in combination, facilitate a definitive detection and precise characterization of Kepler-37d. Its RV semi-amplitude of 1.22 ± 0.31 m s^{-1} (mass 5.4 ± 1.4 M$_\oplus$) is formally consistent with TOI-178b's $1.05^{+0.25}_{-0.30}$ m s^{-1}, the latter being the smallest detected RV signal of any transiting planet to date, though dynamical simulations suggest Kepler-37d's mass may be on the lower end of our 1σ credible interval. Its consequent density is consistent with either a water-world or that of a gaseous envelope (~ 0.4 per cent by mass) surrounding a rocky core. Based on RV modelling and a re-analysis of Kepler-37 TTVs, we also suggest that the putative (non-transiting) planet Kepler-37e should be stripped of its 'confirmed' status.

Key words: methods: data analysis – techniques: radial velocities – techniques: spectroscopic – planetary systems – stars: activity – stars: individual: Kepler-37.

1 INTRODUCTION

Since the discovery of the first exoplanet over 25 yr ago (Mayor & Queloz 1995), Doppler spectroscopy – also known as the radial velocity (RV) method – has been a cornerstone of exoplanetary science. While it is a key tool for planet discovery, it is also indispensable for confirming and characterizing candidates discovered via other techniques, particularly transit photometry (e.g. Konacki et al. 2003). As the RV method constrains planetary masses, it can shed light on planets' likely compositions, formation histories, atmosphere scale heights, and more. As of 2021 May, Doppler spectroscopy has been responsible for the discovery of around one in five known exoplanets, while Doppler spectroscopy and transit photometry have, together, led to the discovery of around 95 per cent of all confirmed exoplanets.[1]

The precision of RV surveys has been steadily improving over the years, thanks to numerous technical advances. While the RV spectrographs of 50 yr ago produced RVs with nominal errors in excess of 1 km s^{-1} per measurement, absorption-cell spectrographs (Campbell & Walker 1979; Marcy & Butler 1992; Butler et al. 1996) have in recent years demonstrated precisions of order 1 m s^{-1} (Butler et al. 2017), whereas the newest generation of ultra-stabilized spectrographs (i.e. using the so-called 'simultaneous reference' technique; Baranne et al. 1996; Probst et al. 2014) today boast sub-m s^{-1} precisions, and aim to achieve 10 cm s^{-1} precisions (Probst et al. 2014; Schwab et al. 2016; Thompson et al. 2016; Petersburg et al. 2020; Pepe et al. 2021). Such precision is sufficient, in principle, to detect the signal of an Earth-analogue exoplanet. Ambitious plans for next-generation ultra-stabilized spectrographs call for stability at the 1 cm s^{-1} level (Pasquini et al. 2008; Fischer et al. 2016).

Despite enormous advances in instrumentation, a few significant obstacles impede the discovery or characterization of planets with RV signatures at the $\lesssim 1$ m s^{-1} level. The semi-amplitude K_p of the RV reflex motion induced by an orbiting planet scales as $K_\mathrm{p} \propto P_\mathrm{p}^{-1/3}$ and $K_\mathrm{p} \propto M_\mathrm{p} \sin i_\mathrm{p} (M_\star + M_\mathrm{p})^{-2/3}$ (Perryman 2011), where P_p is the planet's orbital period, M_p is its mass, i_p is its angle of orbital inclination, and $M_\star$ is the mass of its host star. The paucity of $K_\mathrm{p} \lesssim 1$ m s^{-1} RV detections thus translates into a dearth of low-mass and, to a lesser extent, long-period RV exoplanets.

★ E-mail: vr325@cantab.ac.uk

[1]Based on counts from the NASA Exoplanet Archive, available online at exoplanetarchive.ipac.caltech.edu.

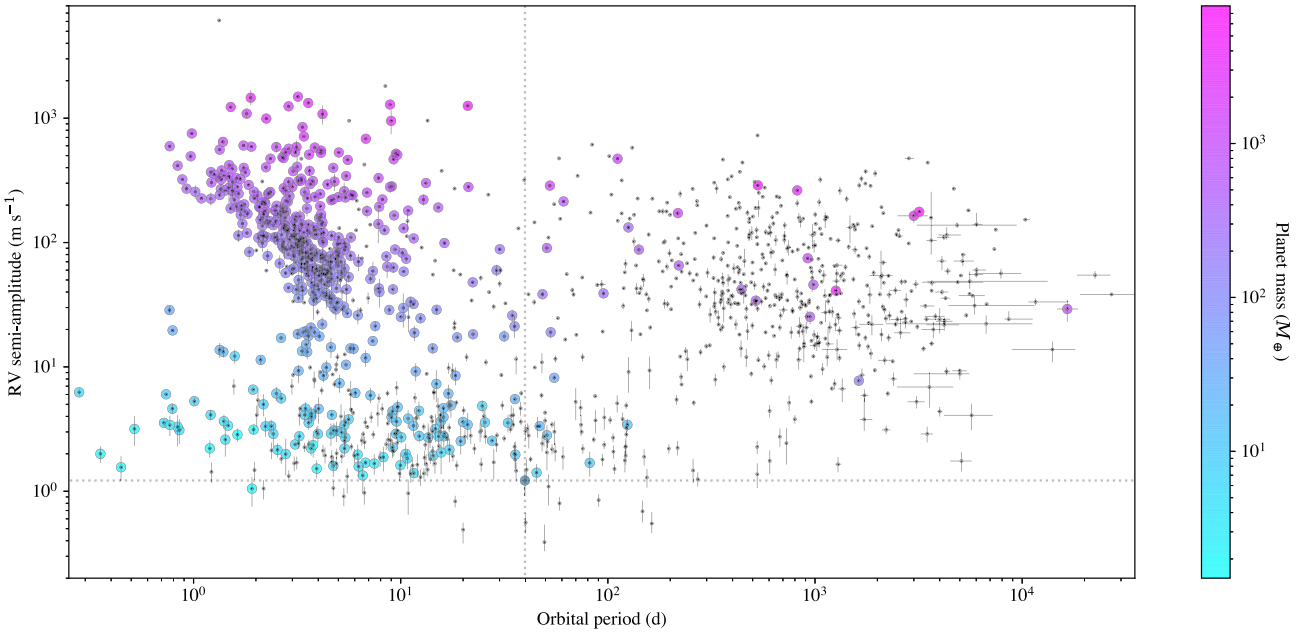

Figure 1. RV semi-amplitudes versus orbital periods of the 1321 (at the time of writing) exoplanets with at least a 3σ detection in RVs. Of these, 556 have known masses (rather than minimum masses), which are indicated via the colour scale. All data retrieved from the NASA Exoplanet Archive on 2021 May 4, except for the point corresponding to Kepler-37d – indicated by the intersection of the grey, dotted lines – which is characterized in this work.

At the time of writing, the NASA Exoplanet Archive contained 1321 confirmed exoplanets with RV signals detected and inconsistent with zero at a 3σ level; of these, only 556 have a true mass (rather than minimum mass $M_{\rm p}\sin i_{\rm p}$) measurement. Moreover, of the latter 556 planets, only 18 have RV semi-amplitudes <2 m s^{-1}, with the smallest being TOI-178b's $1.05^{+0.25}_{-0.30}$ m s^{-1} (Leleu et al. 2021): still an order of magnitude larger than the precision of ESPRESSO, the spectrograph that characterized it (Pepe et al. 2021), or of other cutting-edge instruments such as EXPRES (e.g. Blackman et al. 2020; Petersburg et al. 2020). See Fig. 1.

The most vexatious obstacle to $\lesssim 1$ m s^{-1} detections – and thus the detection of Earth analogues – is variability intrinsic to stars themselves. These stellar nuisance signals, due e.g. to surface magnetic activity such as rotating spots and plages, may be characterized by covariance structures similar to – but amplitudes orders of magnitudes larger than – signals expected from orbiting exoplanets (Dumusque 2012). Thus they can drown out or even mimic exoplanetary signals.

There has therefore been significant effort devoted to developing ways to disentangle stellar activity signals from planetary ones (Boisse et al. 2009; Lanza et al. 2010; Aigrain, Pont & Zucker 2012; Haywood et al. 2014; Robertson & Mahadevan 2014; Tuomi et al. 2014; Rajpaul et al. 2015; Jones et al. 2017; Gilbertson et al. 2020). These approaches have been modestly successful, e.g. enabling the characterization of planets that would otherwise have remained undetected. Up until quite recently, though, most such efforts have been based on *post hoc* attempts to model stellar activity signals already present in RVs. That is, RVs produced by a given pipeline are taken as a starting point, then combined with supplementary information (e.g. knowledge of a stellar rotation period; ancillary photometry; or spectroscopic diagnostics that should be sensitive to activity but not planets) to assess which variability in RVs might be attributable to stellar variability and which to planets.

What about the process of getting actual RV measurements out of stellar spectra, however? For several decades, the standard approach to extracting RVs from spectra taken with stabilized spectrographs was to align or cross-correlate observed spectra with a template (Griffin 1967; Simkin 1974; Baranne, Mayor & Poncet 1979; Tonry & Davis 1979; Bouchy, Pepe & Queloz 2001). The latter is typically either a synthetic template based on model stellar atmospheres, knowledge of atomic line locations, etc., or a high signal-to-noise ratio (SNR) spectrum derived from real observations (e.g. Nordström et al. 1994; Baranne et al. 1996; Balona 2002), in either case with various numerical weights and/or masking applied to different parts of the template (Pepe et al. 2002). The CCF approach retains wide currency and is employed, for example, in the primary data reduction pipelines of HARPS (Rupprecht et al. 2004) and HARPS-N (Cosentino et al. 2012), as well as in the pipelines of newer third-generation instruments such as ESPRESSO (Di Marcantonio et al. 2018).

More recently, though, there has been a rapidly growing number of efforts to develop improved ways of extracting RVs from stellar spectra, particularly with a view to mitigating stellar activity and/or telluric contamination from the final RVs (e.g. Anglada-Escudé & Butler 2012; Dumusque 2018; Zechmeister et al. 2018; Bedell et al. 2019; Collier Cameron et al. 2020; de Beurs et al. 2020; Rajpaul, Aigrain & Buchhave 2020; Zhao & Tinney 2020). These techniques are often data-driven, exploiting the vast quantity of wavelength- and/or time-dependent information in stellar spectra – each typically containing hundreds of thousands of flux/wavelength pairs – to extract 'cleaner' RVs (i.e. less contaminated by activity and/or telluric signals) than may have been possible with traditional approaches. Some techniques also employ forward modelling of observed stellar spectra in the process of deriving RVs (e.g. Zechmeister et al. 2018; Petersburg et al. 2020).

In this paper, we use HARPS-N spectra of Kepler-37 as a case study in both stellar activity modelling (i.e. modelling stellar signals already present in RVs extracted by some pipeline) and spectral-level mitigation (identifying and suppressing the effects of activity at the level of stellar spectra, in order to extract 'cleaner' RVs). For short, throughout this paper we shall refer to these approaches simply as

Table 1. A few key properties of Kepler-37, drawn from the following sources: (1) Stassun, Collins & Gaudi (2017); (2) Berger et al. (2018); (3) Schuler et al. (2015); (4) Morton et al. (2016); (5) Gaia Collaboration (2016); (6) Gaia Collaboration (2018); (7) Andrae et al. (2018); (8) Gaia Collaboration (2021); (9) Walkowicz & Basri (2013); (10) Høg et al. (2000); (11) Mamajek, Meyer & Liebert (2002); (12) Mamajek, Meyer & Liebert (2006); (13) Cutri et al. (2003); (14) Cutri & et al. (2014).

Stellar property	Value	Reference
Mass ($M_\odot$)	0.87 ± 0.15	(1)
Radius ($R_\odot$)	$0.787^{+0.033}_{-0.031}$	(2)
Density ($\rho_\odot$)	1.76 ± 0.23	(1)
$T_{\rm eff}$ (K)	5406 ± 28	(3)
$\log g$ (cm s^{-2})	4.49 ± 0.13	(3)
Age (Gyr)	$3.8^{+3.3}_{-2.0}$	(4)
[Fe/H] (dex)	-0.32 ± 0.07	(2)
Luminosity ($L_\odot$)	0.4791 ± 0.0014	(5), (6), (7)
Distance (pc)	63.999 ± 0.043	(5), (6)
Spectroscopic $P_{\rm rot}$ (d)	22.9 ± 6.9	(9)
Photometric $P_{\rm rot}$ (d)	28.8 ± 3.3	(9)
$v \sin i$ (km s^{-1})	1.70 ± 0.50	(9)
RA (J2000)	$284.0592 \pm 0.0088^\circ$	(5), (8)
Dec (J2000)	$44.518 \pm 0.011^\circ$	(5), (8)
Parallax (mas)	15.625 ± 0.010	(5), (8)
Absolute RV (km s^{-1})	-30.58 ± 0.30	(5), (6)
B-band (mag)	10.446 ± 0.028	(10), (11), (12)
V-band (mag)	9.770 ± 0.028	(10), (11), (12)
J-band (mag)	8.356 ± 0.018	(13)
H-band (mag)	8.000 ± 0.024	(13)
K-band (mag)	7.942 ± 0.013	(13)
W1 (mag)	7.867 ± 0.025	(14)
W2 (mag)	7.933 ± 0.020	(14)
W3 (mag)	7.901 ± 0.019	(14)

'modelling' and 'mitigation'. We shall show how either approach can facilitate the detection and characterization of a hitherto-elusive planet, and moreover, how stellar activity mitigation and modelling can be leveraged in tandem for superior results.

The remainder of the paper is structured as follows. The next section draws on extant literature to summarize salient information about Kepler-37 and its planetary system. Section 3 provides information about the HARPS-N observations that form the basis of this study, and we use these observations to derive updated stellar parameters for Kepler-37. Section 4 describes our methods for extracting RVs and mitigating activity when doing so, while Section 5 discusses our modelling of extracted RVs. We present and discuss our main results in Section 6, then conclude in Section 7. Additionally, Appendix A contains comments on supplementary (though ultimately inconclusive) analyses we performed using HIRES spectra.

2 THE KEPLER-37 SYSTEM

2.1 The star

Kepler-37 – also variously designated TYC 3131-1199-1, KIC8478994, KOI 245, and UGA-1785 – is a main sequence dwarf, slightly cooler and smaller than the Sun, situated 64 pc from Earth in the Lyra constellation. Table 1 summarizes various key properties. Detailed abundances for 19 elements, derived using Keck/HIRES spectra, are provided in Schuler et al. (2015).

Though its spectral type has not been formally determined (Schuler et al. 2015), its temperature, luminosity, mass, radius, etc. are indicative of a late-G or early-K dwarf (G8V/K0V: Pecaut & Mamajek 2013). Despite the low luminosity and low amplitude oscillations associated with cool dwarfs, Barclay et al. (2013) were able to detect asteroseismic oscillations of Kepler-37 in Kepler photometry: at the time of the aforesaid analysis, Kepler-37 was the smallest and densest star in which Solar-like oscillations had been detected (see also Huber et al. 2013), though it no longer holds this record.

Kepler-37 is undetected in the *ROSAT* All Sky Survey (Boller et al. 2016), indicating that it is a weak X-Ray source, which would be consistent with low activity levels. Nevertheless, given Kepler-37's probable spectral type, it should have an outer convective envelope; given also its estimated ∼4 Gyr age and ∼29 d rotation period, one could expect that rotational modulation of possible stellar magnetic features may lead to non-trivial photometric and (apparent) RV variability (Schrijver & Zwaan 2000). This inference will be borne out in Section 4, where a cursory examination of our Kepler-37 RVs will reveal significant correlations with stellar activity indicators.

2.2 Planetary system

2.2.1 *Three transiting planets*

Kepler-37 is orbited by three known transiting planets, the periods of which are within 1 per cent of a 5:8:15 commensurability, hinting at a possible mean-motion resonance chain (Barclay et al. 2013).

With a radius of just $0.30\,R_\oplus$, the inner planet, Kepler-37b, is smaller than Mercury, and only slightly larger than the Moon. The NASA Exoplanet Archive indicates that, as of May 2021, Kepler-37b still holds the distinction of having the smallest measured radius of any exoplanet. The next smallest transiting planet, Kepler-444b, has a radius of $0.40\,R_\oplus$, while only two other confirmed planets, Kepler-102b and Kepler-444c, have radii estimates smaller than $0.5\,R_\oplus$. Kepler-37b orbits Kepler-37 with a period of 13.37 d at a distance of 0.10 AU, and is probably rocky, with no atmosphere or water (Barclay et al. 2013).

Kepler-37c is around three-quarters of the size of Earth – still very much at the lower end of the radius distribution of confirmed exoplanets – and orbits Kepler-37 every 21.30 d at a distance of 0.14 AU. Kepler-37d, on the other hand, is approximately twice the size of Earth, and orbits Kepler-37 every 39.79 d at a distance of 0.21 AU. Given their proximity to Kepler-37, Kepler-37c and Kepler-37d are also not expected to contain surface liquid water.

Some key properties of the three transiting planets are summarized in Table 2.

2.2.2 *Kepler-37e?*

The paper presenting the discovery of the three transiting planets (Barclay et al. 2013) noted the presence of a fourth planet candidate orbiting Kepler-37, then referred to as KOI-245.04, with a 51.2 d orbital period. However, Barclay et al. (2013) commented that they did 'not trust that KOI-245.04 is a valid planet candidate' as the inclusion of more data since the release of the (then) most recent planet candidate catalogue *decreased* the signal to noise of the putative transit signal, suggesting the transit-like signal in question was likely caused by random noise, or correlated stellar or instrumental noise.

Subsequently, Hadden & Lithwick (2014) analysed transit timing variations (TTVs) obtained over 3 yr (Q1–Q12) of the Kepler mission, to extract densities and eccentricities of 139 sub-Jovian planets, using ephemerides from the TTV catalogue published by Mazeh et al. (2013). In the former authors' analysis, the apparent TTVs of 'Kepler-37e' were considered to derive constraints on the

Table 2. Key properties of the three planets known to transit Kepler-37, and of the putative fourth planet Kepler-37e. Data from the following sources: (1) Gajdoš et al. (2019); (2) Berger et al. (2018); (3) Stassun et al. (2017); (4) Van Eylen & Albrecht (2015); (5) Barclay et al. (2013); (6) Hadden & Lithwick (2014).

Planet	Orbital period (d)	Ref.	Radius ($R_\oplus$)	Ref.	Inclination (deg)	Ref.	Eccentricity	Ref.	Transit mid-point (BJD−2 455 000)	Ref.
Kepler-37b	13.367020 ± 0.000060	(1)	$0.277^{+0.033}_{-0.025}$	(2)	88.63 ± 0.41	(3)	$0.08^{+0.210}_{-0.080}$	(4)	17.0473 ± 0.0037	(1)
Kepler-37c	21.301848 ± 0.000018	(1)	$0.725^{+0.040}_{-0.032}$	(2)	$89.07^{+0.19}_{-0.33}$	(5)	$0.090^{+0.180}_{-0.090}$	(4)	24.83997 ± 0.00087	(1)
Kepler-37d	39.7922622 ± 0.0000065	(1)	$1.917^{+0.085}_{-0.076}$	(2)	$89.335^{+0.043}_{-0.047}$	(5)	$0.15^{+0.07}_{-0.10}$	(4)	8.24982 ± 0.00013	(1)
'Kepler-37e'	∼51.196?	(6)	?	—	?	—	?	—	n/a	—

mass of Kepler-37d; however, this seemed to pre-suppose that KOI-245.04 was a real planet, which (to the best of our knowledge) had not yet been established.

A number of later studies suggested that Kepler-37's transiting planets do *not* exhibit significant TTVs (e.g. Holczer et al. 2016; Gajdoš, Vaňko & Parimucha 2019). Sowing further doubt, two separate planet-searching pipelines (Huang & Bakos 2014; Kunimoto & Matthews 2020) applied to Kepler data explicitly failed to identify the signal associated with KOI-245.04 as a *bona fide* transit. In fact, Kunimoto & Matthews (2020) noted that KOI-245.04 was the only 'planet' completely missed by the pipeline they applied to ∼ 200 000 FGK dwarfs observed by Kepler.

Given the scant evidence in support of a detection, and the lack of a literature consensus, we hereafter tread with caution and treat KOI-245.04 or 'Kepler-37e' as a *putative* but not indisputably confirmed planet, despite several catalogues treating it as such.[2] We shall aim to shed light on this candidate planet's existence through our main RV analysis, complemented by an updated TTV investigation.

2.2.3 *Expected RV detectability*

Using our knowledge of the radii of the transiting planets, we can calculate the semi-amplitude of their associated RV signals, assuming various possible compositions (Zeng et al. 2019).

Even if Kepler-37b had an improbably-high density, it would certainly *not* be detectable by HARPS-N, nor indeed by any extant spectrograph. Assuming a 100 per cent Fe composition, say, its consequent 0.03 $M_\oplus$ mass would induce reflex motion in Kepler-37 with $\lesssim 1$ cm s^{-1} semi-amplitude, which is an order of magnitude smaller than the semi-amplitude of the RV reflex motion induced in the Sun by Earth. Kepler-37c has equally unrealistic prospects for RV characterization: a 100 per cent Fe composition would give rise to a meagre 14 cm s^{-1} RV signal.

Kepler-37d, on the other hand – the focus of this study – has more realistic prospects for RV detection. Considering all planets in the NASA Exoplanet Archive with radii within 0.25 $R_\oplus$ of Kepler-37d's radius (and non-trivial mass measurements), 84 per cent have masses greater than 4.1 $M_\oplus$, and 50 per cent have masses greater than 7.3 $M_\oplus$; assuming the latter masses for Kepler-37d, its consequent RV signal would be 0.9 or 1.6 m s^{-1}, respectively. Earth-like rocky (32.5 per cent Fe and 67.5 per cent $MgSiO_3$) and 100 per cent $MgSiO_3$ compositions for Kepler-37d would be associated with RV signals with semi-amplitudes of 2.7 and 1.9 m s^{-1}, respectively. Even a 50 per cent Earth-like rocky core plus 50 per cent mass in an H_2O layer, assuming 500 K equilibrium temperature and 1 mbar surface pressure, would be associated with a 1.1 m s^{-1} RV signal: difficult, though not unfeasible, to detect with HARPS-N (cf. remarks in Section 1 on the dearth of <2 m s^{-1} detections).

[2]At the time of writing, 'Kepler-37e' is listed as a confirmed planet by sources including the NASA Exoplanet Archive (in which Kepler-37e bears no 'controversial' flag), the Open Exoplanet Catalog (available online at www.openexoplanetcatalogue.com), and SIMBAD (Wenger et al. 2000). The Extrasolar Planets Encyclopaedia (Schneider et al. 2011) dissents, listing only three confirmed planets in the system.

3 HARPS-N OBSERVATIONS

3.1 Details of observations

The basis of the analyses in this paper is a set of 110 high-SNR spectra obtained with HARPS-N (Cosentino et al. 2012), an $R \approx 115\,000$ spectrograph covering a wavelength range from 383 to 690 nm.

A total of 115 HARPS-N spectra were obtained between 2014 April 15 and 2019 September 10, with the majority from 2014 and 2015 (59 and 53 spectra, respectively), and only three spectra taken in 2019. The majority of the observing time originated from two proposals in period AOT29 and AOT31 (PI Buchhave) that were respectively awarded 29.5 and 17.5 h of observing time for Kepler-37, corresponding to roughly 94 spectra. The remaining observations were acquired by the HARPS-N GTO program. 108 of the 115 spectra had exposure times of 1800 s, with the remaining exposure times being between 930 and 1500 s. Across all observations, the median SNR/pixel at 570 nm was 115; the median seeing was 1.27 arcsec, and the median airmass was 1.16.

Following visual inspection of RVs extracted by the HARPS-N Data Reduction Software (DRS; see Section 4.1), we identified five RVs as possible outliers, viz. from spectra taken on the nights of 2014 April 20, 2014 July 11, and 2015 May 11, 14, and 15. The SNRs for all five spectra were several (up to about 10) times lower than for the majority of spectra. Moreover, some of these spectra were associated with activity indicator values that appeared more anomalous than their formal error bars suggested (e.g. $\log R'_{\rm HK}$), DRS drift correction quality control flags marked as 'failed', unusually poor seeing, etc. To err on the side of caution, we excluded these five spectra from further consideration, and worked with the remaining 110 spectra.

3.2 Refined stellar parameters

The high SNR needed to extract precise RVs from spectra means spectra used for this purpose are generally more than adequate for deriving stellar parameters. Accordingly, we took advantage of our high-SNR high-resolution HARPS-N spectra to derive refined stellar parameters for Kepler-37.

We derived stellar atmospheric parameters via three independent methods: ARES+MOOG, `CCFPams`, and the Stellar Parameter

Table 3. Refined properties of Kepler-37 derived in this work using our HARPS-N spectra.

Stellar property	Value	Note
$T_{\rm eff}$ (K)	5357 ± 68	(1)
$\log g$ (cm s^{-2})	4.60 ± 0.12	(1)
[Fe/H] (dex)	-0.36 ± 0.05	(1)
ξ_t (km s^{-1})	0.93 ± 0.08	(2)
$v \sin i$ (km s^{-1})	<2.0	(3)
Mass ($M_\odot$)	$0.790^{+0.033}_{-0.030}$	(4)
Radius ($R_\odot$)	$0.789^{+0.0064}_{-0.0056}$	(4)
Density ($\rho_\odot$)	$1.624^{+0.096}_{-0.093}$	(4)
Age (Gyr)	$7.6^{+3.4}_{-3.1}$	(4)
Distance (pc)	63.999 ± 0.042	(4)

(1) Averaged parameters from ARES+MOOG, `CCFPams`, and SPC analyses; (2) microturbulent velocity – from ARES+MOOG analysis; (3) from SPC analysis; (4) averaged parameters from `isochrones` and MIST analyses.

Classification tool (SPC). ARES+MOOG is a curve-of-growth method based on neutral and ionized iron lines, and is explained in Sousa (2014) and references therein; the `CCFPams` method, described in Malavolta et al. (2017b), uses cross-correlation function (CCF) equivalent widths to obtain effective temperature, surface gravity, and iron abundance via empirical calibration; lastly, SPC is a spectrum synthesis method, described in detail in Buchhave et al. (2012, 2014). The first two methods use a co-added master spectrum, while SPC uses the individual spectra and takes a median of the individual results. Surface gravity estimates from ARES+MOOG and `CCFPams` were corrected for accuracy (Mortier et al. 2014), and systematic errors were added to our precision errors for effective temperature, surface gravity, and iron abundance as derived by ARES+MOOG and `CCFPams` (Sousa et al. 2011). Our final adopted stellar atmospheric parameters, which appear in Table 3, are inverse variance-weighted averages of the results from these three methods, following a methodology first used by Malavolta et al. (2018) that has since been well tested and widely adopted (e.g. Rice et al. 2019; Mortier et al. 2020).

We obtained stellar mass, radius, age, and distance estimates from isochrones and evolutionary tracks. We used the `isochrones` package (Morton 2015) and two separate stellar evolution models, viz. Dartmouth (Dotter et al. 2008) and MIST (`MESA` Isochrones and Stellar Tracks; Dotter 2016) to estimate these parameters, following the methodology described in Mortier et al. (2020). As inputs, we used our spectroscopically determined effective temperature and iron abundance, the Gaia EDR3 parallax, and magnitudes from the visible to mid-infrared (as listed in Table 1). We ran each code three times – varying only between the sets of spectroscopic parameters estimates by the ARES+MOOG, `CCFPams`, and SPC approaches – to produce a total of six estimates for stellar mass, radius, age and distance. Our final estimates for each parameter, in Table 3, were obtained by combining the six posterior distributions for each parameter.

We note that all of our revised stellar parameters in Table 3 are consistent within 1σ with the literature values in Table 1. While the precisions of our surface gravity and metallicity estimates are essentially the same as the literature values, our stellar mass and radius estimates have error bars five times smaller than the literature values (thanks to a much improved parallax value). Furthermore, while the parameters in Table 1 are drawn from many distinct data sets and analyses, our stellar parameters have the advantage of being derived in a uniform way from a single data set.

4 RV EXTRACTION

We consider RVs extracted from HARPS-N spectra using two independent methods: the HARPS/HARPS-N DRS pipeline, and a Gaussian process (GP) method that infers RVs by aligning pairs of spectra.

4.1 DRS RV extraction and activity indicators

Both HARPS and HARPS-N spectra are, by default, processed by the DRS pipeline, which is optimized for exoplanet searches (Lovis et al. 2006; Cosentino et al. 2012). The DRS cross-correlates observed spectra with a stellar mask chosen from a template library, and thus computes a CCF for each spectrum. The mask itself consists of a list of wavelength ranges that identify spectral lines, and weights defining the contribution of each spectral line to the cross-correlation. Masks optimized for G2-, K5-, and M2-type stars are readily available and have been widely used, though more recently, bespoke masks for other stellar types have also been developed (e.g. Rainer, Borsa & Affer 2020). The DRS uses the resulting CCF both for RV extraction and for building stellar activity indicators, since e.g. stellar activity can induce asymmetries and other distortions to the CCF.

As the DRS has been widely used in published RV studies, we do not discuss it in detail here; we simply note that we used a K5 mask and version 3.7 of the DRS – the latest version available in 2020, when most of our analyses were carried out – to extract our RVs.[3] A periodogram and time series representation of these RVs appear in Figs 2 and 3, respectively. In addition to RVs, we also extracted activity indicators including $\log R'_{\rm HK}$, bisector span (BIS), and full width at half-maximum (FWHM) time series; the activity indicators referred to in the rest of this paper are always those extracted by the DRS, since our second method of RV extraction, described below, does not compute these activity indicators.

Summary statistics pertaining to key stellar activity indicators for Kepler-37, derived by the DRS, are given in Table 4. In particular, we note here that the mean $\log R'_{\rm HK}$ value over our 110 spectra, -4.871, would lead Kepler-37 to be classified as (relatively) 'inactive,' following the scheme given by Maldonado et al. (2010), in which a star is classified as 'active' when $\log R'_{\rm HK} > -4.75$, and 'very active' when $\log R'_{\rm HK} > -4.2$. For comparison, the Sun varies between about $\log R'_{\rm HK} \sim -4.85$ and -5.0 as it moves between high- and low-activity phases.

GLS (Zechmeister & Kürster 2009) periodograms for the same stellar activity indicators appear in the middle panel of Fig. 2. Periodograms for the tightly-correlated $\log R'_{\rm HK}$ and FWHM time series show prominent peaks around 14, 29, 35, and 48 d, all of which could be associated with Kepler-37's estimated ∼29-d rotation period (see Table 1) and possible active region evolution. Power around ∼100 and ∼200 d could be associated with active region evolution or changes as a result of a longer-term magnetic cycle (we touch again on the latter possibility in Section 6.2). The BIS periodogram has a prominent peak around 26 d, again close to the estimated rotation period. Note that the stellar rotation period should not be expected

[3]As of 2021, a newer version of the DRS (counter-intuitively numbered 2.2.8), based on the ESPRESSO pipeline, is available. The difference in Kepler-37 RVs extracted by the two DRS versions turns out to be negligible: 95 per cent of RVs are consistent within 2σ, and 100 per cent within 2.3σ. Correlations between RVs and activity indicators (see Table 5) are as strong or marginally stronger in the newer RVs. We also re-ran a representative subset of our modelling (per Section 5) using the newer RVs; our conclusions were identical.

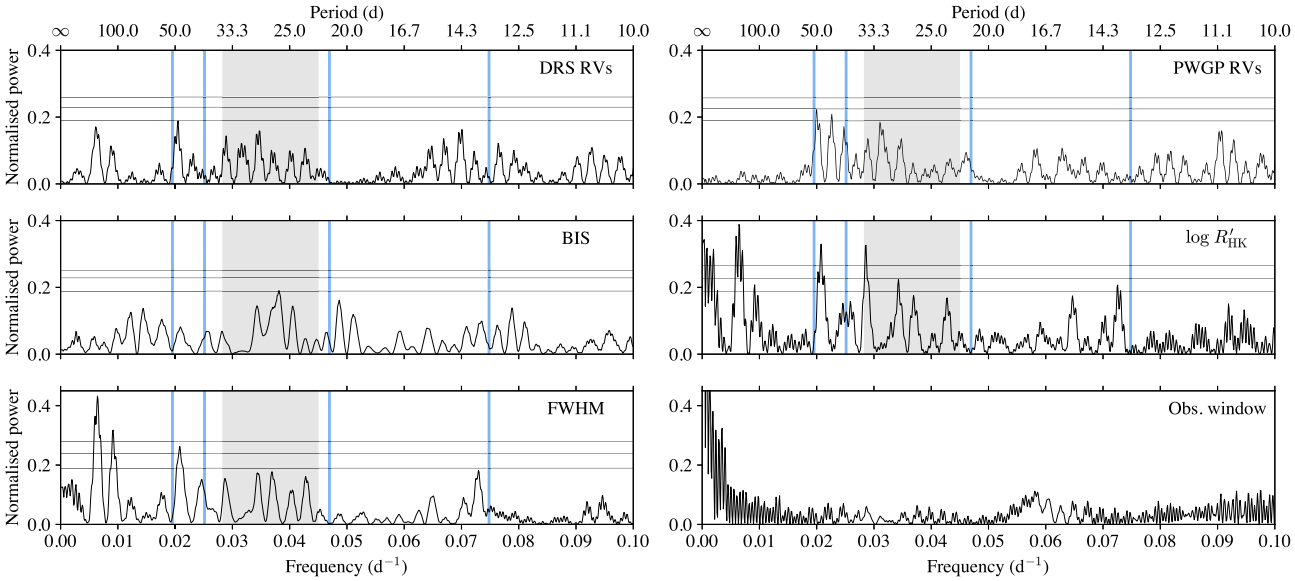

Figure 2. Generalized Lomb-Scargle (GLS) periodograms of Kepler-37 RVs, as extracted by the DRS and pairwise GP (PWGP) approaches (top panels); of three different stellar activity indicators (middle and lower left-hand panels); and of the observing times (lower right-hand panel). Vertical blue lines indicate the orbital periods of Kepler-37b through to Kepler-37e, while the grey shaded box covers a $\pm 2\sigma$ credible interval around Kepler-37's photometric $P_{\rm rot}$, as estimated by Walkowicz & Basri (2013). The horizontal lines indicate, from bottom to top, estimated 10 per cent, 1 per cent, and 0.1 per cent false alarm probability (FAP) thresholds, respectively.

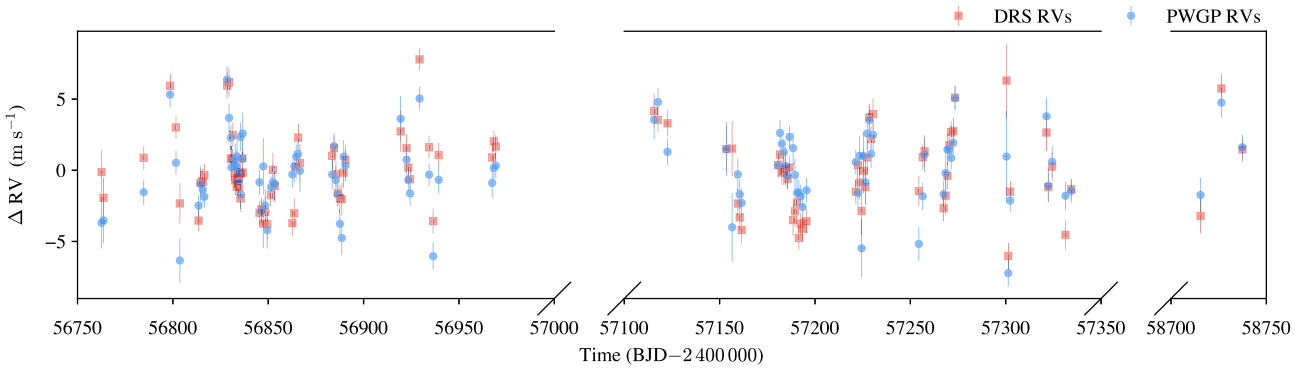

Figure 3. The 110 Kepler-37 HARPS-N RVs used in this paper, as extracted by the DRS pipeline (red squares) and PWGP method (blue circles), along with 1σ error bars from each method. The DRS RVs have their mean value of -30.651 km s^{-1} subtracted, to facilitate comparison with the PWGP RVs, which by construction have zero mean velocity.

to be identifiable via a single periodogram peak (Nava et al. 2020): the GLS periodogram assumes a data model of a *single sine wave plus white Gaussian noise* (VanderPlas 2018). This will always be an imperfect – and sometimes deeply flawed – model for quasi-periodic RVs containing multiple planets and signals arising from dynamic active regions on a diferentially rotating star, which in turn might be subject to long-term magnetic cycles, etc.

4.2 Pairwise GP extraction

The second method we used to extract RVs was the GP-based method presented in Rajpaul et al. (2020). In brief, GPs are used to model and align all pairs of spectra with each other; the pairwise RVs thus obtained are combined to produce differential stellar RVs, without constructing any template. Given the reliance on GP modelling of spectra, and the pairwise nature of the RV extraction, we hereafter refer to this method as the PWGP method.

The rationale for the pairwise comparison of spectra is largely computational: modelling and aligning tens or hundreds of spectra simultaneously would require inversion of enormous covariance matrices (generally the biggest bottleneck to any GP modelling), whilst also sampling parameter spaces of very high dimensionality. By contrast, pairwise RV extraction entails relatively cheap repeated computation that parallelizes trivially, with only a few parameters to be optimized or sampled for any pair of spectra.

The PWGP method can be used to compute differential RVs on a localized basis, e.g. to yield an independent set of RVs for each échelle order, or for much smaller subdivisions of orders. The motivation is that regions of spectra affected by e.g. stellar activity or telluric contribution may be identified and excluded (essentially a data-driven masking of the spectrum, without any knowledge of line locations or properties) from the calculation of the final RVs, which are obtained by an inverse variance-weighted average of the localized RVs.

Table 4. A few statistics summarizing key properties of our HARPS-N spectra of Kepler-37, and measurements extracted from these spectra.

Summary statistic	Value
No. spectra including outliers	115
No. spectra analysed in this work	110
Time span of observations (d)	1974.8
Mean (DRS RV) ($m\,s^{-1}$)	−30 651.51
Mean (PWGP RV) ($m\,s^{-1}$)	0
Mean (DRS RV error) ($m\,s^{-1}$)	1.02
Mean (PWGP RV error) ($m\,s^{-1}$)	1.11
SD (DRS RV) ($m\,s^{-1}$)	2.68
SD (PWGP RV) ($m\,s^{-1}$)	2.50
Mean (BIS) ($m\,s^{-1}$)	4.61
Mean (BIS error) ($m\,s^{-1}$)	2.04
SD (BIS) ($m\,s^{-1}$)	3.47
Mean ($\log R'_{HK}$)	−4.871
Mean ($\log R'_{HK}$ error)	0.0051
SD ($\log R'_{HK}$)	0.023
Mean (FWHM) ($m\,s^{-1}$)	6026.60
Mean (FWHM error) ($m\,s^{-1}$)	2.40
SD (FWHM) ($m\,s^{-1}$)	8.97

To avoid confusion from assigning multiple meanings to the Greek letter sigma, in this table we use 'SD' to denote the standard deviation of a given set of measurements, and 'error' to refer to the estimated 1σ error bars on measurements.

Rajpaul et al. (2020) showed that even a relatively crude implementation of the PWGP method, applied to an inactive star (where only modest if any improvements may have been expected compared to the DRS), resulted in RVs with precisions comparable to and rms scatter about 30 per cent lower than RVs extracted by the DRS and two other commonly-used codes, *viz.* HARPS-TERRA (Anglada-Escudé & Butler 2012) and SERVAL (Zechmeister et al. 2018).

We applied the PWGP method essentially as described in the proof-of-concept paper by Rajpaul et al. (2020). We divided each of the 69 échelle orders per spectrum into 32 'chunks,' for a total of 2208 chunks, each 128 pixels wide (typical width a little under 2 Å). Corresponding chunks across all pairs of spectra were fitted with a Matérn-$\frac{5}{2}$ kernel, and aligned via a maximum likelihood approach. This produced a $110 \times 110 \times 2208$ array of pairwise RV shifts, and an uncertainty array of identical dimension. The arrays are necessarily anti-symmetric and symmetric, respectively, with respect to the first two dimensions.

Up to this point, the PWGP method proceeds autonomously. However, some thought is then required as to how, if at all, to identify and exclude spectral chunks that are likely 'contaminated' – by stellar activity-related variability, as typified e.g. in the Calcium II H & K lines, by telluric variability, typified e.g. in water vapour lines, etc. – before producing a final set of differential RVs. Here, approaches of varying levels of sophistication are possible. For example, a subset of local RVs exhibiting periodicities known to be associated with the stellar rotation period could be excluded, or a clustering analysis might help identify chunks with similar properties and link problematic chunks with one another. We adopted a conservative version of a simple approach that has at least been tested on both synthetic and real spectra (Damasso et al. 2020; Rajpaul et al. 2020); specifically, we excluded all chunks exhibiting the following simple properties:

(i) rms or median error bar $>10\ km\,s^{-1}$; and/or

(ii) significant correlation ($p < .05$ under a non-parametric Spearman rank correlation test) with any activity indicator (BIS, $\log R'_{HK}$, FWHM), barycentric Earth RV, or airmass time series.

Assuming Poisson statistics, one would expect the SNR in RVs extracted from a given wavelength range to scale as $\sqrt{N}$, with N being the photon count. All else being equal, and assuming a precision of $\sim 1\ m\,s^{-1}$ is possible with the full spectrum, a precision of $\sqrt{2208} \sim 50\ m\,s^{-1}$ might be expected from individual chunks. Thus, local RVs satisfying condition (i), i.e. with rms scatter or error bars orders of magnitude larger than could have been expected *a priori*, were likely contaminated by extremely strong stellar, telluric, or instrumental signals, or they may have corresponded to continuum-dominated regions of spectra containing very little Doppler information.[4]

Local RVs satisfying condition (ii) likely suffered mild to strong stellar activity or telluric contamination. Note, however, that an absence of a significant local correlation with e.g. an activity indicator does not prove that a chunk was contamination free – it simply means that any such contamination was not detected via one specific measure of statistical association. In the case of very strong stellar variability, it can well be the case (e.g. Damasso et al. 2020) that even when combining local chunk RVs that all appear uncorrelated with activity indicators – due to low SNR – the resultant RVs nevertheless exhibit overwhelming correlations with activity indicators. On a related note, the traditional activity indicators we used are only imperfect proxies for activity signals in RVs. For example, on Sun-like stars, $\log R'_{HK}$ is not expected to correlate very strongly with activity-driven RV variations over long time-scales (Milbourne et al. 2019). As such, our filtering will only be sensitive to the specific types of activity variation tracked by the indicators used.

Following the above filtering, roughly 39 per cent of chunks remained, and were accordingly combined via an inverse variance-weighted average to produce a final set of 110 differential RVs. While discarding ~ 61 per cent of spectra by wavelength range suggests an aggressive filtering, note for comparison that the combined width of the several thousand lines included in the DRS' G2 mask is only $\sim$170 Å, or ~ 5 per cent of the full 1D spectrum, the latter being dominated by regions containing little Doppler information.

4.3 Comparison of the DRS and PWGP RVs

GLS periodograms of the DRS and PWGP RVs appear in the upper panel of Fig. 2, while the RVs themselves are plotted in Fig. 3. A quick glance at the RVs in Fig. 3 suggests much superficial similarity; indeed, Table 4 confirms that the PWGP RVs have only marginally larger error bars, and a marginally smaller rms scatter, than the DRS RVs. The median and median *absolute* differences between the two sets of RVs are $0.16\ cm\,s^{-1}$ and $1.32\ m\,s^{-1}$, respectively. Their mutual linear (and rank) correlation coefficient is $\rho \sim 0.74$ ($p \ll .001$).

The periodograms in Fig. 2 reveal that both sets of RVs evince periodicities at several periods likely linked to rotationally modulated stellar activity. However, unlike the DRS RVs, the PWGP RVs show no power at periods of $\sim$100 or $\sim$200 d, both of which feature strongly in the periodograms of the $\log R'_{HK}$ and FWHM time series, suggesting some degree of successful activity mitigation by the PWGP approach. The activity mitigation is confirmed in Table 5, which summarizes correlation between the DRS/PWGP RVs and various stellar activity indicators. Whereas the DRS RVs show both strong and significant non-parametric correlations with four different

[4]A single-value cut will inevitably be somewhat arbitrary. We note that we obtained nearly identical results when we used a $1\ km\,s^{-1}$ cut instead.

Table 5. Non-parametric measures of correlation between Kepler-37 RVs extracted using the DRS and the PWGP approach, and various time-varying quantities (columns) that are expected to be independent of dynamically-induced stellar velocity shifts, though possibly sensitive to stellar activity or telluric contamination. The FWHM, contrast, and BIS are all CCF parameters yielded by the DRS; they measure the width, depth, and asymmetry, respectively, of the average spectral line profile (Anglada-Escudé & Butler 2012). BERV is the barycentric Earth radial velocity at the time of observation. For each correlation, we give the Spearman rank correlation coefficient, ρ, and the associated p-value. Correlations significantly nonzero at a $p < .05$ level are typeset in bold.

	FWHM		Contrast		BIS		$\log R'_{\rm HK}$		Airmass		BERV	
	ρ	p	ρ	p	ρ	p	ρ	p	ρ	p	ρ	p
DRS RVs	**+0.50**	**$\ll .001$**	**−0.50**	**$\ll .001$**	**+0.22**	**.02**	**+0.45**	**$\ll .001$**	−0.03	.76	−0.12	.20
PWGP RVs	+0.12	.23	−0.14	.15	−0.04	.68	+0.13	.19	+0.01	.95	−0.02	.85

stellar activity indicators, these correlations are wholly absent from the PWGP RVs.

Intriguingly, there is far more power at Kepler-37d's period in the PWGP RVs than the DRS RVs. Even in the PWGP case, though, this power still falls below a 10 per cent FAP threshold, and is exceeded in power by nearby periodicities around 30, 45, and 51 d. If nothing else, this emphasises how weak Kepler-37d's putative signal is, and how important it will be to model the data as carefully as possible.

5 RV MODELLING

We now describe our modelling of the RVs extracted via the DRS or the PWGP approach. We discuss our modelling of Keplerian signals (Section 5.1); our stellar activity modelling with a GP fitted to multiple time series (Section 5.2); the Bayesian framework we used for all model and parameter inference (Section 5.3); and the practical implementation of this Bayesian inference via the POLYCHORD algorithm (Section 5.4).

5.1 Keplerian signals and RV offsets

We model the contribution of N_p orbiting planets to observed RVs via a standard prescription, ignoring planet–planet interactions (for more details, see e.g. Seager 2010 or Perryman 2011):

$$\mathrm{RV}_{\rm Kepler}(t) = \sum_{p=1}^{N_p} K_p \left[\cos(\nu_p(t) + \omega_p) + e_p \cos(\omega_p)\right], \tag{1}$$

where K_p is the RV semi-amplitude, ν_p the true anomaly, e_p the orbital eccentricity, and ω_p the argument of periastron of the pth planet, respectively. The true anomaly, characterizing the time-dependent angle between a planet and its periapsis position, is computed via

$$\tan\left(\frac{\nu_p(t)}{2}\right) = \sqrt{\frac{1+e_p}{1-e_p}} \tan\left(\frac{E_p(t)}{2}\right). \tag{2}$$

The eccentric anomaly $E_p(t)$, in turn, is found by numerical solution of the following non-linear equation ('Kepler's equation'):

$$E_p(t) - e_p \sin E_p(t) = \frac{2\pi(t - T_{0,p})}{P_p} \equiv M_p(t), \tag{3}$$

where $T_{0,p}$ defines the time at which the pth planet is located at a particular reference point (e.g. periapsis), P_p is that planet's orbital period, and $M_p(t)$ is defined as the mean anomaly. In practice, it is sometimes more convenient to parametrize the orbit using an angular parameter instead of using a reference point in its orbit – e.g. $M_{0,p}$, the mean anomaly at some reference time. In total, then, there are five observables that we can fit for a single orbiting planet, on the basis of RV measurements alone: e_p, P_p, $T_{0,p}$ (or $M_{0,p}$), ω_p, and K_p. The latter is related to the stellar and planetary masses via

$$K_p = \left(\frac{2\pi G}{P_p}\right)^{1/3} \frac{M_p \sin i_p}{(M_\star + M_p)^{2/3}} \frac{1}{(1-e_p^2)^{1/2}}, \tag{4}$$

where i_p denotes the angle of orbital inclination. We solve for M_p from K_p by making the common approximation that $M_p \ll M_\star$.

In addition to Keplerian signals, we also allowed for the possibility that each time series contained long-term trends (due e.g. to a long-term activity cycle, a distant orbiting body, instrumental drifts, etc.), which we modelled with polynomials up to second order. However, all our initial tests showed that (i) polynomial coefficients of first order and above were consistent within 1σ with zero in all time series, and (ii) models containing first- or second-order polynomial trends had less favourable evidence values than those containing offsets only. Therefore, we restrict ourselves here to considering models with only an offset for RV, BIS, and $\log R'_{\rm HK}$ time series – hereafter denoted $\gamma_{\rm RV}$, $\gamma_{\rm BS}$, $\gamma_{\rm HK}$ – without higher-order polynomial components.

5.2 Stellar activity

In cases where we modelled stellar activity explicitly (as opposed to assuming it could be neglected, or at least accounted for via an additive white-noise 'jitter' term), we used the GP framework developed by Rajpaul et al. (2015), hereafter R15, to model RVs simultaneously with $\log R'_{\rm HK}$ and BIS observations – the latter two time series being sensitive to activity-induced variability, but *not* planetary signals. In short, this framework assumes that all observed stellar activity signals are generated by some underlying latent function $G(t)$ and its derivatives; this function, which is not observed directly, is modelled with a GP (Rasmussen & Williams 2006; Roberts et al. 2013).

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

$$\Delta\mathrm{RV} = V_c G(t) + V_r G'(t), \tag{5}$$

$$\mathrm{BIS} = B_c G(t) + B_r G'(t), \text{ and} \tag{6}$$

$$\log R'_{\rm HK} = L_c G(t), \tag{7}$$

respectively, where $G'(t) = \mathrm{d}G/\mathrm{d}t$. The coefficients V_c, B_c, L_c, V_r, and B_r are free parameters relating the individual observations to the unobserved Gaussian process $G(t)$. The first three coefficients pertain to convective blueshift suppression, and the latter two to rotationally modulated signals (hence the subscripts). $G(t)$ itself can be loosely interpreted as representing the projected area of the visible stellar disc covered in active regions at a given time (Aigrain et al. 2012). The GP describing $G(t)$ is assumed to have zero mean and covariance matrix $\mathbf{K}$, where $K_{ij} = \gamma(t_i, t_j)$. As in R15, we adopt the following

quasi-periodic (QP) covariance kernel function,

$$\gamma(t_i, t_j) = \exp\left[-\frac{\sin^2\left[\pi(t_i - t_j)/P_{\rm GP}\right]}{2\lambda_{\rm p}^2} - \frac{(t_i - t_j)^2}{2\lambda_{\rm e^2}}\right], \quad (8)$$

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

By modelling multiple activity-sensitive time series simultaneously, more information can be gleaned on activity signals in RVs,[5] compared to approaches exploiting only simple correlations between RVs and (typically) one activity indicator (Gilbertson et al. 2020). In general, we would not advocate using a GP to model activity in RVs alone. Despite the convenience of such approaches (e.g. Blunt et al. 2018), even if good prior constraints on hyper-parameters are available, there is little safeguard against fitting signals unrelated to stellar activity in order to achieve an optimal data likelihood. The term 'over-fitting' might be subtly misleading in such cases, as model residuals could be perfectly consistent with the formal error bars, even if non-activity signals were inadvertently absorbed.

As our framework uses GP draws and derivatives thereof as basis functions for modelling available time series, it avoids issues inherent in e.g. sinusoidal or other simple parametric models, the inappropriate use of which could easily lead to the introduction of correlated signals into model residuals. The GP basis functions could in principle take any form, although in the GP framework their properties are constrained by the data themselves, and by reasonable prior assumptions about the quasi-periodic nature of stellar activity signals. The GP framework also incorporates the so-called FF' formalism directly as a special case (Aigrain et al. 2012); the former approach may be thought of as a generalization of the latter. For a recent, in-depth study demonstrating advantages of R15's GP framework over a number of other approaches to modelling stellar activity, including the FF' method and a multi-harmonic method, see Ahrer et al. (2021).

In addition to the parameters associated directly with stellar activity in our GP model, we also included white-noise 'jitter' parameters for each time series – which we denote $\sigma^+_{\rm RV}$, $\sigma^+_{\rm BS}$, $\sigma^+_{\rm HK}$ – and which were added in quadrature to the formal error bars for each observation. Under the GP model, these white-noise jitter parameters were intended to encapsulate activity-induced and other signals that are not adequately captured by the GP model, whether because of flaws in the assumed relationship between RVs and indicators, or because certain signals simply do not show up in chosen activity indicators (e.g. stellar pulsation signals). In cases where we did *not* use a GP, the jitter parameters were intended to encapsulate *all* signals (including stellar activity-related ones) that could not be adequately modelled via Keplerian terms.

[5]This would not be true for a hypothetical indicator that was somehow independent of activity-driven variations in RVs; given the correlations in Table 5, such a concern is not warranted here. We also note that it would also not be advantageous to model multiple activity indicators that contained little independent information.

5.2.1 *Summary of modelling approaches used in this work*

As already noted, we consider in this work RVs extracted via two independent methods, and in each case consider the effect of using a GP to model stellar activity versus using only a white-noise jitter term. Moreover, for every such approach, we consider many different combinations of Keplerian terms being included in the overall model. Therefore, to aid the reader, we summarize here the four main modelling approaches we take in this work, where we use 'approach' as shorthand for one particular combination of RV data set and stellar activity model, regardless of the Keplerian terms included,

(I) DRS RVs – no activity modelling;
(II) DRS RVs – GP activity modelling;
(III) PWGP RVs – no activity modelling; and
(IV) PWGP RVs – GP activity modelling.

We shall often refer to these different approaches using the uppercase numerals (I)–(IV). Adopting the language used earlier in the paper, we may characterize these approaches thus: (I) no activity modelling or mitigation; (II) activity modelling only; (III) activity mitigation only; and (IV) activity mitigation combined with activity modelling. Of course, even in case (I) there is a degree of activity mitigation taking place, since e.g. the masks used by the DRS are designed to avoid lines known to be most susceptible to stellar activity-induced variability. Similarly, a white-noise model that can (in principle) 'absorb' some stellar activity variability is perhaps not ignoring activity completely. Therefore, we use the terms 'modelling' and 'mitigation' in a largely relative sense, where it is understood that case (I) represents our baseline.

5.3 Bayesian model and parameter inference

We use Bayesian inference to evaluate the relative posterior probabilities of (i) different models, and (ii) of different parameter values within given models. We summarize here the relevant formalism.

Bayes' Theorem relates the posterior probability $P(\Theta|\mathcal{D}, \mathcal{M}) \equiv \mathcal{P}(\Theta)$ of some parameters Θ, given data $\mathcal{D}$ and a model $\mathcal{M}$, to

(i) the probability of Θ, $P(\Theta|\mathcal{M}) \equiv \pi(\Theta)$, given $\mathcal{M}$;
(ii) the probability of $\mathcal{D}$, $P(\mathcal{D}|\Theta, \mathcal{M}) \equiv \mathcal{L}(\Theta)$, given Θ, $\mathcal{M}$; and
(iii) the probability of $\mathcal{D}$, $P(\mathcal{D}|\mathcal{M}) \equiv \mathcal{Z}$, given $\mathcal{M}$.

Bayes' theorem may be written as

$$P(\Theta|\mathcal{D}, \mathcal{M}) = \frac{P(\mathcal{D}|\Theta, \mathcal{M})\, P(\Theta|\mathcal{M})}{P(\mathcal{D}|\mathcal{M})}, \quad (9)$$

or, using the notation from Feroz & Hobson (2014), simply as

$$\mathcal{P}(\Theta) = \frac{\mathcal{L}(\Theta)\, \pi(\Theta)}{\mathcal{Z}}. \quad (10)$$

$\mathcal{P}(\Theta)$ is the posterior probability density of the model parameters; $\mathcal{L}(\Theta)$ is the likelihood of the data, and $\pi(\Theta)$ is the parameter prior. The term in the denominator, $\mathcal{Z}$, is usually referred to as the marginal likelihood, model likelihood, or Bayesian evidence – it represents the factor required to normalize the posterior over the entire domain of Θ, i.e. $\mathcal{Z} = \int \mathcal{L}(\Theta)\pi(\Theta)d\Theta$. In general, $\mathcal{Z}$ is notoriously difficult to compute (Nelson et al. 2020). Fortunately, as this term is independent of the parameters Θ, it can be ignored for parameter inference problems, where samples may be drawn from the unnormalized posterior only, as happens e.g. with standard MCMC methods.

For the far more challenging problem of model inference (selection), however, the marginal likelihood or evidence plays a central role. As the evidence may be interpreted as the likelihood averaged over the prior, it is generally larger in a model where more of

Table 6. Jeffreys' scale (Jeffreys 1961) for interpreting Bayesian evidence ratios (Bayes factors). A value $\mathcal{R}_{ij} > 1$ means that model $\mathcal{M}_i$ is favoured more strongly by the data under consideration than model $\mathcal{M}_j$. We give the scale both in terms of Bayes factors and (natural) logarithms of the Bayes factors. We find the former more intuitive to interpret directly, though POLYCHORD outputs the latter, and is more convenient for evidences spanning many orders of magnitudes.

Bayes factor	Log Bayes factor	Strength of evidence
$\mathcal{R}_{ij} < 10^0$	$\ln \mathcal{R}_{ij} < 0$	Negative (supports $\mathcal{M}_j$)
$10^0 < \mathcal{R}_{ij} < 10^{1/2}$	$0 < \ln \mathcal{R}_{ij} < 1.15$	Barely worth mentioning
$10^{1/2} < \mathcal{R}_{ij} < 10^1$	$1.15 < \ln \mathcal{R}_{ij} < 2.30$	Substantial
$10^1 < \mathcal{R}_{ij} < 10^{3/2}$	$2.30 < \ln \mathcal{R}_{ij} < 3.45$	Strong
$10^{3/2} < \mathcal{R}_{ij} < 10^2$	$3.45 < \ln \mathcal{R}_{ij} < 4.61$	Very strong
$\mathcal{R}_{ij} > 10^2$	$\ln \mathcal{R}_{ij} > 4.61$	Decisive

its total parameter space is associated with high likelihoods, and smaller for a model where large areas of parameter space have low likelihood values, even if the likelihood function is sharply peaked. Thus the evidence serves both to penalize 'fine tuning' of a model against observed data, and to automatically and quantitatively implement Occam's Razor or principle of parsimony (e.g. Rasmussen & Ghahramani 2000; Feroz & Hobson 2014).

One can evaluate the relative posterior probabilities of two models $\mathcal{M}_i$ and $\mathcal{M}_j$, given data $\mathcal{D}$, by computing the ratio of their respective posterior probabilities; this ratio is also known as the Bayes factor, which we denote $\mathcal{R}_{ij}$:

$$\mathcal{R}_{ij} = \frac{P(\mathcal{M}_i|D)}{P(\mathcal{M}_j|D)} = \frac{P(\mathcal{D}|\mathcal{M}_i)\,P(\mathcal{M}_i)}{P(\mathcal{D}|\mathcal{M}_j)\,P(\mathcal{M}_j)} = \frac{\mathcal{Z}_i}{\mathcal{Z}_j}\frac{P(\mathcal{M}_i)}{P(\mathcal{M}_j)}. \tag{11}$$

The relative prior probability of the two models, $\frac{P(\mathcal{M}_i)}{P(\mathcal{M}_j)}$, is usually set to unity (unless there happens to be some information available that would suggest favouring one model over the other *a priori*), in which case $\ln \mathcal{R}_{ij} = \ln \mathcal{Z}_i - \ln \mathcal{Z}_j$. This is the approach we take throughout the analysis presented in this paper.

To decide whether the relative posterior probabilities favour one model over the other, we make use of the Jeffreys scale given in Table 6. We emphasize, though, that a Bayes factor $\mathcal{R}_{ij}$ can only be used to *compare* two models: a large value of $\mathcal{R}_{ij}$ may certainly lead to the model or hypothesis $\mathcal{M}_j$ being rejected, but it does not prove that $\mathcal{M}_i$ is 'correct' in an absolute sense, which would in principle require evaluating *all* (infinitely many) alternatives. We recall here the often-quoted words of statistician George Box (Box & Draper 1987): 'All models are approximations. Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind.'

5.3.1 *Likelihood function*

In common with almost all other RV modelling efforts, we make the assumption that our model residuals are drawn from a multivariate Gaussian distribution. This may be justified via the central limit theorem, which establishes that the sum or average of many independent random variables tends towards a Gaussian distribution, even if the original variables are not normally distributed (Fischer 2011). Indeed, measurement errors in astronomical experiments usually contain contributions from many independent sources – photon noise, thermo-mechanical noise, calibration errors, telescope and detector effects, atmospheric effects, etc. – so this assumption is usually well founded, strongly non-Gaussian outliers due e.g. to cosmic ray strikes notwithstanding. From an information theoretic perspective, the maximum entropy principle may also be used to show that in the absence of detailed knowledge of the effective noise distribution (other than assuming it has finite variance), a Gaussian distribution would be the most conservative choice, i.e. maximally non-committal about missing information (Gregory 2005; Sivia & Skilling 2006).

Given our assumption of Gaussianity, the logarithmic likelihood of our data may be computed via the familiar expression

$$\ln \mathcal{L}(\Theta) = -\tfrac{N}{2}\ln 2\pi - \tfrac{1}{2}\ln\det \mathbf{K} - \tfrac{1}{2}\mathbf{r}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{r}, \tag{12}$$

where $\mathbf{K} \in \mathbb{R}^{N\times N}$ defines the covariance between all possible pairs of observations, $\mathbf{r} \in \mathbb{R}^N$ is a vector of residuals, and $N = N_{\mathrm{RV}} + N_{\mathrm{BS}} + N_{\mathrm{HK}}$ is the total number of observations of all types considered.

In cases where we used a GP to model stellar activity variability simultaneously across RV, BIS, and log R'_{HK}measurements, $\mathbf{K}$ was computed via expressions given in section 3.3 and Appendix A2 in R15, thus encoding activity-related covariances between all observations in all three time series, as well as white-noise variance (observational error plus jitter parameter, added in quadrature) associated with individual observations. The vector of residuals, $\mathbf{r}$, was computed by subtracting the appropriate mean function from each type of observation – a constant offset plus a varying number of Keplerian terms for RVs, and constant offsets only for BIS and log R'_{HK} – then concatenating the three sets of residuals into a single vector.

In the simpler cases where we did *not* use a GP to model stellar activity, and instead assumed that all non-planetary variability in RVs could be interpreted as uncorrelated (white) noise, $\mathbf{K}$ was a diagonal matrix, with each diagonal element encoding white-noise variance (observational error plus jitter, as before) associated with particular observations. The vector of residuals, $\mathbf{r}$, was computed as before. We emphasize that under this (non-GP) modelling approach, RVs, BIS, and log R'_{HK} observations were all considered to be independent; the only parameters relating specifically to the BIS and log R'_{HK} observations were constant offsets and white-noise jitter parameters. Nevertheless, we deliberately included all three types of observations in our modelling. This had no effect on inference about RV-specific parameters, e.g. Keplerian parameters, but it *did* have an important effect on the computation of model evidences, since $\mathcal{Z} = P(\mathcal{D}|\mathcal{M})$ always depends on the data being modelled. In other words, in order to compare the evidences of models with and without a GP stellar-activity component, it was imperative that the data in question were identical across all models. Essentially, including the BIS and log R'_{HK} observations in the non-GP modelling ensured the model evidences were correctly normalized to allow comparison with the GP-based models. Thus, we could answer questions such as: 'to what extent do the data support using a more complex GP model to model variability in three (possibly related) time series?'

Table 7. Priors placed over the parameters in the Keplerian components of our RV models. As discussed in Section 5.3.2, we also rejected parameter combinations giving rise to at least one pair of unstable Keplerian orbits. Additional notes: (1) we use the shorthand $s_{\rm RV}$ to denote the standard deviation of the RV time series, as given in Table 4; (2) we also required periods to be sorted, i.e. $P_f < P_g < \ldots$; (3) see Kipping (2013) for more details.

Parameter	Prior	Notes
K_b (m s^{-1})	mod$\mathcal{J}(10^{-4}, 10^{-2})$	Cf. Section 2.2.3
K_c (m s^{-1})	mod$\mathcal{J}(10^{-4}, 0.2)$	Cf. Section 2.2.3
K_i (m s^{-1})	mod$\mathcal{J}\left(10^{-1}, s_{\rm RV}\right)$	$i \in \{d, e, f, \ldots\}$; (1)
P_b (d)	13.367020	Fixed; cf. Table 2
P_c (d)	21.301848	Fixed; cf. Table 2
P_d (d)	39.7922622	Fixed; cf. Table 2
P_e (d)	$\mathcal{N}(51.196, 1)$	cf. Table 2
P_i (d)	$\mathcal{J}(0.5, 4000)$	$i \in \{f, g, \ldots\}$; (2)
e_i	$\beta(0.867, 3.03)$	$i \in \{b, c, \ldots\}$; (3)
$T_{0,\,b}$ (BJD)	2455017.0473	Fixed; cf. Table 2
$T_{0,\,c}$ (BJD)	2455024.83997	Fixed; cf. Table 2
$T_{0,\,d}$ (BJD)	2455008.24982	Fixed; cf. Table 2
M_i (rad)	$\mathcal{U}(0, 2\pi)$	$i \in \{e, f, \ldots\}$
ω_i (rad)	$\mathcal{U}(0, 2\pi)$	$i \in \{b, c, \ldots\}$

5.3.2 *Parameter priors*

We use the symbols $\mathcal{N}$, $\mathcal{U}$, and β to denote normal, uniform, and Beta distributions, respectively, in each case with standard parametrizations. We denote a Jeffreys prior over some parameter $0 < a \leq \theta \leq b$ via $\mathcal{J}(a, b)$, such that

$$p(\theta) = \frac{1}{\theta} \times \frac{1}{\ln(a/b)}. \tag{13}$$

This prior (also known as a log uniform prior) is scale invariant, so is an appropriate choice for parameters whose scale is not known *a priori*; by contrast, a standard uniform prior is inherently biased to larger parameter values (Gregory 2005). The Jeffreys prior is not suitable, however, for parameters whose minimum allowable value is zero, in which case the Jeffreys prior is not normalizable. In such cases, a modified Jeffreys prior over some parameter $0 \leq \theta \leq \theta_{\rm max}$, which we denote mod$\mathcal{J}(\theta_0, \theta_{\rm max})$, is more suitable,

$$p(\theta) = \frac{1}{\theta + \theta_{\rm max}} \times \frac{1}{\ln(1 + \theta_{\rm max}/\theta_0)}; \tag{14}$$

this mirrors the Jeffreys prior closely for large values of θ, though resembles a uniform prior when $\theta < \theta_0$.

The priors we placed on the parameters of the three planets transiting Kepler-37, the putative TTV planet (Kepler-37e), and any other possible planets in our models are given in Table 7. Where informative prior constraints on a (real or putative) planet's expected RV semi-amplitude were not available, we set the maximum allowable RV semi-amplitude to be the standard deviation of the observed RVs. This was essentially a computational convenience, albeit an easily justifiable one: given the known presence of white noise, correlated stellar and other nuisance signals, multiple known planets, etc., it is not plausible that a single planet could account for the entirety of the observed RV variability. We also confirmed, in preliminary runs, that posterior distributions for K_i never peaked near the prior upper limit. We allowed the maximum periods of putative planets to be slightly more than double the temporal baseline of our HARPS-N observations.

Apart from the priors we placed on individual planetary parameters, we also wished to impose the prior restriction that planetary orbits should not be unstable. Accordingly, we checked the mutual orbital stability of all pairs of Keplerian orbits, using a criterion introduced by Gladman (1993), and often used in RV modelling (e.g. Malavolta et al. 2017a): $\Delta > 2\sqrt{3}\, R_H(i, j)$ where $\Delta = a_j - a_i$ is the difference between the semi-major axis of the ith and jth planet, and $R_H(i, j)$ the planets' mutual Hill radius. Any combination of planet parameters giving rise to at least one unstable orbit was rejected by setting the associated data likelihood to zero.

Table 8. Priors placed over the non-planetary – white-noise jitter, offsets, and GP i.e. stellar activity-related – components of our models. Here, we use the shorthand μ_i and s_i to denote the mean and standard deviation of a particular set of measurements, e.g. $\mu_{\rm HK}$ is the mean value of the log $R'_{\rm HK}$time series, and $s_{\rm BS}$ is the standard deviation of the BIS time series.

Parameter	Prior	Notes
$\sigma^+_{\rm RV}$ (m s^{-1})	mod$\mathcal{J}(10^{-1}, s_{\rm RV})$	Cf. Table 4
$\sigma^+_{\rm BIS}$ (m s^{-1})	mod$\mathcal{J}(10^{-1}, s_{\rm BS})$	Cf. Table 4
$\sigma^+_{\rm HK}$	mod$\mathcal{J}(10^{-3}, s_{\rm HK})$	Cf. Table 4
$\gamma_{\rm RV}$ (m s^{-1})	$\mathcal{N}(\mu_{\rm RV}, s_{\rm RV})$	–
$\gamma_{\rm BS}$ (m s^{-1})	$\mathcal{N}(\mu_{\rm BS}, s_{\rm BS})$	–
$\gamma_{\rm HK}$	$\mathcal{N}(\mu_{\rm HK}, s_{\rm HK})$	–
$P_{\rm GP}$ (d)	$\mathcal{J}(10, 100)$	See also equation (15)
λ_e (d)	$\mathcal{J}(10, 1000)$	See also equation (15)
λ_p (m s^{-1})	$\mathcal{J}(10^{-1}, 10)$	See also equation (15)
V_c (m s^{-1})	$\mathcal{N}(0, \sigma_{\rm RV})$	–
V_r (m s^{-1})	$\mathcal{N}(0, \sigma_{\rm RV})$	–
B_c (m s^{-1})	$\mathcal{N}(0, \sigma_{\rm BS})$	–
B_r (m s^{-1})	$\mathcal{N}(0, \sigma_{\rm BS})$	–
L_c	$\mathcal{N}(0, \sigma_{\rm HK})$	–

The priors we placed on all other (i.e. non-planetary) parameters in our models appear in Table 8; these were all, essentially, uninformative priors. For example, we only constrained the additive white-noise jitter parameters to be less than the standard deviation of the time series to which they related (RV, BIS or log $R'_{\rm HK}$), for reasons analogous to our prior constraints on the RV semi-amplitudes K_e, K_f, etc. We constrained our GP hyper-parameters to a broad range likely to cover all physically-plausible possibilities. However, reasoning that we were using the GP to model quasi-periodic stellar activity signals, we did reject parameter combinations that would have resulted in a kernel that was not even quasi-periodic. In particular, we required,

$$\lambda_{\rm e}^2 > 3 P_{\rm GP}^2 \lambda_{\rm p}^2 / 2\pi, \tag{15}$$

which must be satisfied for the QP covariance function to have at least one non-trivial turning point (Rajpaul 2017).[6]

5.4 POLYCHORD

Throughout this study, we use POLYCHORD for all parameter and model inference. POLYCHORD is a state-of-the-art nested sampling algorithm, designed to work well even with parameter spaces with very large dimensionality (Handley, Hobson & Lasenby 2015). A short discussion illuminating its favourable properties compared with MultiNest – its direct predecessor, and a nested sampling tool that has been widely used in exoplanet studies (Feroz & Hobson

[6]This constraint ultimately turned out to be superfluous, as the hyper-parameter posterior probability mass was strongly concentrated far away from regions where this constraint may have been violated (see Table 12) However, we include the constraint here for completeness' sake.

Table 9. Relative Bayesian evidences for models including different numbers of Keplerian terms. For ease of interpretation, the evidence for the GP model including Kepler-37d as the only Keplerian term is here defined to be zero for both the DRS RVs ($\mathcal{Z}_*$) and the PWGP RVs ($\mathcal{Z}_\dagger$); *all* other evidences are given relative to one of these values. Evidences for models for different data (DRS versus PWGP RVs) can not be compared meaningfully. Uncertainties correspond to the standard error on the mean evidence across five POLYCHORD runs, or to the highest individual run error provided by POLYCHORD (the greater of the two).

Keplerians	(I) DRS; no activity model		(II) DRS + GP activity model		(III) PWGP; no activity model		(IV) PWGP + GP activity model	
	$\ln(\mathcal{Z}/\mathcal{Z}_*)$	$\sigma(\ln\mathcal{Z})$	$\ln(\mathcal{Z}/\mathcal{Z}_*)$	$\sigma(\ln\mathcal{Z})$	$\ln(\mathcal{Z}/\mathcal{Z}_\dagger)$	$\sigma(\ln\mathcal{Z})$	$\ln(\mathcal{Z}/\mathcal{Z}_\dagger)$	$\sigma(\ln\mathcal{Z})$
–	−100.524	0.065	−2.834	0.176	−92.371	0.063	−6.138	0.128
b	−101.775	0.078	−4.178	0.160	−94.891	0.082	−6.010	0.156
c	−102.041	0.077	−4.030	0.162	−94.166	0.082	−6.237	0.142
d	−101.377	0.086	≡ 0	0.131	−88.865	0.086	≡ 0	0.124
e	−100.202	0.095	−1.646	0.157	−90.501	0.093	−3.689	0.154
f	−97.563	0.127	−3.159	0.238	−89.786	0.141	−3.127	0.167
b, *d*	−102.248	0.142	−1.829	0.156	−90.669	0.145	−0.362	0.142
c, *d*	−102.556	0.142	−1.773	0.214	−90.823	0.145	−0.465	0.152
d, *e*	−100.939	0.085	−0.507	0.121	−88.394	0.089	+0.008	0.116
d, *f*	−98.671	0.194	−0.673	0.249	−89.116	0.101	−0.914	0.173
d, *e*, *f*	−98.239	0.108	−0.770	0.253	−89.351	0.094	−1.179	0.188
d, *f*, *g*	−97.683	0.176	−0.913	0.284	−89.729	0.103	−1.409	0.228

2008, 2014; Feroz, Hobson & Bridges 2009) – can be found in Hall et al. (2018). A detailed study testing POLYCHORD and then leveraging it specifically for joint modelling of exoplanets and stellar activity in RVs (as in the present study) is given by Ahrer et al. (2021).

POLYCHORD is written in C++ and Fortran, though we called it via the PYPOLYCHORD PYTHON wrapper. We generally left POLYCHORD's sampling parameters at their default values, except for the stopping (precision) criterion, which we changed from the default 10^{-3} to a *much* more stringent 10^{-12}: we found the default value led to evidence values that often differed significantly from run to run (Ahrer et al. 2021). As POLYCHORD natively supports Message Passing Interface parallelization, we were able to run it on a high-performance computing platform, typically using several hundred CPU cores simultaneously. This made it feasible for us to compute accurate Bayesian evidences even for models with high dimensionality and computationally intensive GP components to evaluate.

To obtain robust estimates of the uncertainty in computed model evidences, we considered (i) POLYCHORD's internal estimates provided by individual runs, and (ii) the standard error in the mean of model evidences returned from multiple POLYCHORD runs, and adopted the larger of the two as our most plausible uncertainty estimate (cf. Nelson et al. 2020; Ahrer et al. 2021).

6 RESULTS AND DISCUSSION

6.1 Model selection

Table 9 gives the Bayesian evidences we computed for models featuring various numbers and combinations of Keplerian terms – including ones for all three transiting planets, and the putative planet Kepler-37e – under each of our four modelling approaches. These results are based on multiple, repeated POLYCHORD runs for each model. As this table encapsulates several of our main results, albeit in a very condensed format, we unpack a few of the main findings below. Note that our focus for now is primarily on model selection; in Section 6.2 we shall turn our attention to more fine-grained questions about parameter values inferred under various models.

6.1.1 *Non-detection of Kepler-37b and Kepler-37c*

Regardless of the data set or stellar activity model, models that included either only Kepler-37b or c were *never* favoured over the simpler null (planet-free) models, or models additionally containing Kepler-37d. This was unsurprising, though reassuring, as we did not expect to be able to detect either of the inner two transiting planets.

6.1.2 *Non-detection of new planets*

Under modelling approach (I), representing the case of no activity mitigation or modelling, models with more 'free' Keplerians (in principle, representing hitherto-unknown planets) were favoured over models without these components. However, parameter posteriors revealed that these Keplerians were almost certainly being used to absorb stellar signals. Their periods often corresponded to periodicities in the power spectra of activity indicators; they tended to have large ($e > 0.2$) eccentricities; and their other orbital parameters were weakly constrained. Different runs of identical models led to similar evidence values, but with different periods sometimes being favoured. In short, following the reasoning set out in Ahrer et al. (2021), we concluded that these results were indicative of inadequate stellar activity modelling, *not* of the detection of genuine planets.

Under modelling approaches (II)–(IV), however, models including free Keplerians were never favoured over the model containing Kepler-37d only. In some cases, a model with one free Keplerian was at least favoured over the planet-free model. However, it always turned out that this Keplerian term was used to account for variability at ∼100 or ∼200 d (periodicities that we have already noted likely have a stellar origin), with other orbital parameters poorly constrained.

6.1.3 *Decisive detection of Kepler-37d*

Our most interesting and striking results concern Kepler-37d. Moving from approaches (I)–(IV), the strength of the evidence supporting a model containing Kepler-37d only, versus the null model, increases significantly. In approach (I), the evidence is 'barely worth mentioning', while in the other three approaches, the evidence is 'strong', 'very strong', and 'decisive', respectively, following

Table 10. Bayes factors $\mathcal{R}_{d0} \equiv \mathcal{Z}_d/\mathcal{Z}_0$, quantifying the degree to which a model containing Kepler-37d was favoured versus a planet-free model, under modelling approaches (I)–(IV). While these Bayes factors and uncertainties may be derived from Table 9, we present them here for ease of reference.

Modelling approach	Bayes factor $\mathcal{R}_{d0}$	K-37d detection
(I) DRS; no activity model	$0.426^{+0.048}_{-0.043}$	–
(II) DRS + GP activity model	$17.0^{+4.2}_{-3.3}$	Strong
(III) PWGP; no activity model	$33.3^{+3.6}_{-3.3}$	Very strong
(IV) PWGP + GP activity model	463^{+90}_{-75}	Decisive

Jeffreys' scheme in Table 6. Whether considering the DRS RVs or the PWGP RVs, a model containing Kepler-37d as the only Keplerian contribution to RVs is decisively favoured over all alternatives. These findings are summarized quantitatively in Table 10 which, unlike Table 9, focuses on the strength of detection for Kepler-37d only, rather than numerous alternative models containing various other Keplerians.

As one might have hoped, then, mitigating and modelling stellar activity each enabled the detection of a planetary signal that would otherwise have remained 'buried' under stellar nuisance signals. Moreover, combining these two approaches led to a stronger decisive detection of said planetary signal.

6.1.4 *Non-detection of Kepler-37e*

Models containing only Kepler-37e – or more precisely, a Keplerian with ∼51-d period – were rejected compared to the null model under approaches (I), (II), and (IV); similarly, models containing both Kepler-37d and Kepler-37e were *not* favoured over simpler models containing only Kepler-37d. The only tentative evidence for a Kepler-37e detection came under approach (III), where a model with Kepler-37e only (and $K_e \sim 1.0 \pm 0.4\ \mathrm{m\,s^{-1}}$) was favoured over the null model. However, models including both Kepler-37d and Kepler-37e (in which case $K_e \sim 0.8 \pm 0.4\ \mathrm{m\,s^{-1}}$) were not favoured over the simpler Kepler-37d-only model, with the latter being strongly favoured over the null model. Given also the overwhelmingly lower evidence for all models under approach (III) compared to approach (IV), we conclude that we did not detect Kepler-37e in our RVs.

6.1.5 *Model residual analysis*

In Fig. 4 we show the Lomb-Scargle periodograms of the residuals from our best-fitting Kepler-37d-only model in approaches (I)–(IV); note that, in cases (II)–(IV), the associated model was also favoured over all others considered under the respective approach.

Fig. 4 shows that the residuals from approach (I) are the only ones to contain any significant (FAP < 10 per cent) periodicities. Strikingly, the most significant periodicity occurs around ∼51 d, i.e. the supposed period of Kepler-37e (corresponding to the leftmost blue line in the plot; the associated frequency is $0.02\ \mathrm{d^{-1}}$). In the other three approaches, each of which entails stellar activity modelling and/or mitigation, the 51-d periodicity is absent from the residuals. As 50.9 d is the primary annual alias of 30 d, we suspect the apparent 51-d periodicity was suppressed in tandem with the 30-d activity signal. The presence of the apparent ∼51-d periodicity not only in Kepler-37 RVs and photometry, but also in the log R'_{HK} and FWHM time series (Fig. 2), provides further support for this periodicity/alias being activity related. A planetary origin seems implausible, given that the periodicity is absent even from the residuals in approach (III),

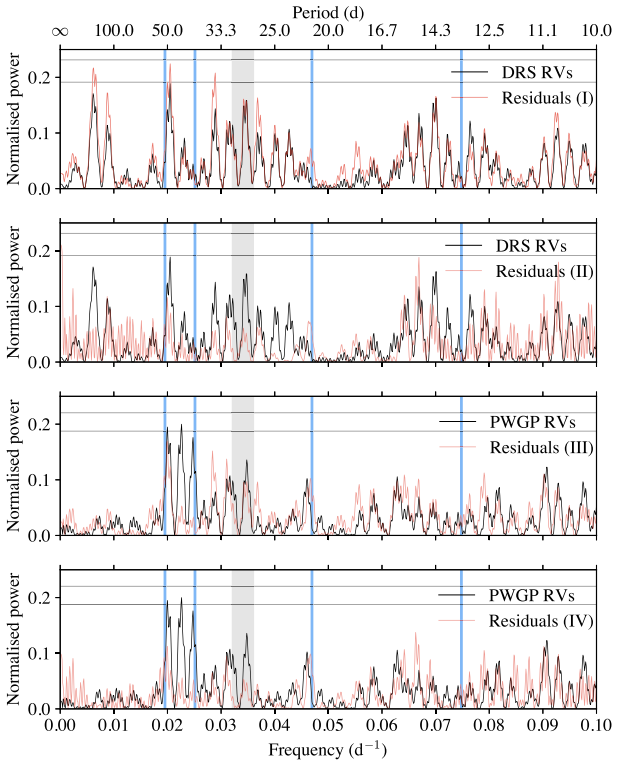

Figure 4. GLS periodogram of Kepler-37 RV residuals, after subtracting our best-fitting Kepler-37d model – plus GP activity model, in cases (II) and (IV) – for the DRS RVs and PWGP RVs. As in Fig. 2, vertical blue lines indicate the orbital periods of Kepler-37b through to Kepler-37e; the grey shaded box covers a $\pm 2\sigma$ credible interval around Kepler-37's estimated rotation period, but now based on our GP modelling (cf. Table 12). The lower and upper horizontal lines indicate estimated 10 per cent and 1 per cent FAP thresholds, respectively.

where one could not contend that a GP model may have inadvertently removed a Keplerian signal with this period, since no GP modelling took place.

The absence of significant periodicities in the residuals from approaches (II)–(IV) does not prove that there are no planetary signals buried in the data. (Indeed, we know Kepler-37b and c are present, and Kepler-37d's signal has FAP > 10 per cent even in the PWGP *RVs*.) On the other hand, this does at least conform with the conclusions from our more robust Bayesian model comparisons, *viz*. that the data do not support the detection of additional Keplerian signals.[7]

Additionally, autocorrelation functions of the residuals from approaches (I)–(IV) showed no significantly non-zero autocorrelations for any lags.

6.1.6 *Quantitative justification for GP modelling*

In both the DRS and PWGP cases, the Bayesian evidence for the best model without a GP was many (tens of) orders of magnitude lower than the worst model including a GP. The interpretation is that the added complexity and larger overall volume of parameter space

[7]In general, we would *not* advocate using periodograms as the basis for conclusions about planet detections or non-detections. We use periodograms here for a quick intuitive way of visualising the end result of fitting models to (quasi-)periodic data, and of diagnosing glaring shortcomings in such fits.

Table 11. Residual rms scatter in RV, BIS, and log $R'_{\rm HK}$ time series, under modelling approaches (I)–(IV), in each case for the model containing Kepler-37d as the only Keplerian. For comparison, the rms of the mean-subtracted DRS and PWGP RVs was 2.68 and 2.50 m s^{-1}, respectively (see Table 4).

Modelling approach	Residual rms		
	RV (m s^{-1})	BIS (m s^{-1})	log $R'_{\rm HK}$ –
(I) DRS; no activity model	2.66	3.47	0.0227
(II) DRS + GP activity model	2.08	2.74	0.0072
(III) PWGP; no activity model	2.35	3.47	0.0227
(IV) PWGP + GP activity model	2.24	2.79	0.0067

introduced by the GP is more than compensated for by the dramatic improvements in the extent to which the GP model can explain non-Keplerian correlated variability in the data (see Table 11). This is despite our fairly stringent requirement that a single GP and its derivative must account for activity-related variability in RVs, BIS, and log $R'_{\rm HK}$ time series *simultaneously* (see Section 5.2). Moreover, the volume of parameter space associated with reasonably good fits seems to be much smaller under the non-GP models (cf. e.g. the narrow error bars on white-noise jitter parameters in Table 12), which is indicative of inappropriate model complexity, and would contribute to the very unfavourable evidence values (Rasmussen & Ghahramani 2000).

While one could make several heuristic arguments in favour of the GP modelling – including the need to account for the manifest though hard-to-parametrize activity contamination noted in Section 4.3, and the fact that in practice, the GP modelling enabled the most secure RV detection of Kepler-37d – it is nevertheless interesting to be able to justify the GP's use via Bayesian model comparison.

6.1.7 *Computational burdens of our analysis*

The most complex models we evaluated, featuring three Keplerian terms and a GP stellar activity model, required $\mathcal{O}(10^8)$ likelihood evaluations before POLYCHORD converged; the simplest models typically required $\mathcal{O}(10^6)$ likelihood evaluations. Since we considered models featuring many different combinations of Keplerian terms (including but not limited to the 12 combinations given in Table 9), across four different approaches, and evaluated every such combination five times, we ultimately evaluated $\mathcal{O}(10^{10})$ likelihood functions. This required tens of thousands of CPU hours on a high-performance computing platform. One might well ask whether such computationally expensive analyses are worth the trouble. We would respond with a resounding 'yes.'

The lion's share of our computational budget was associated with inverting large covariance matrices, as required for GP likelihood evaluations. The overwhelming benefit of using a GP to model stellar activity across multiple time series is borne out by Tables 9 and 10: if one had neglected to model stellar activity on this (nominally) inactive star, one would simply not have detected Kepler-37d. It is beyond the scope of this study to consider activity models less complex or sophisticated than our GP framework; such an in-depth investigation was, however, carried out by Ahrer et al. (2021), who ultimately argued in favour of using a GP to model stellar activity across multiple time series, despite the computational burdens.

Even without the burdens of GP evaluation, computing Bayesian evidences is notoriously difficult. Yet as an approach to evaluating statistically whether one has detected a planet, it is virtually unassailable in its rigour. Compare, for example, the GLS periodograms in Fig. 2 that, by themselves, would have lent little or no credibility to any claim of a Kepler-37d detection. Similarly, some of our models had posterior semi-amplitudes for certain Keplerians – e.g. Kepler-

Table 12. Summary of marginalized 1D posteriors for all parameters in modelling approaches (I)–(IV), in each case where Kepler-37d was the only Keplerian component in the model. For each parameter, the posterior median and $\pm\sigma$ credible interval around the median are given; parameters are separated into categories following the scheme in Tables 7 and 8. The angular parameter ω_d (argument of periastron) has been transformed to the domain $[\pi, 3\pi]$ to suppress boundary effects associated with the original domain $[0, 2\pi]$; the posterior peaks close to the edges of the latter domain, leading to apparent though spurious bimodality.

Parameter	(I) DRS; no activity model	(II) DRS + GP activity model	(III) PWGP; no activity model	(IV) PWGP + GP activity model	Notes
K_d (m s^{-1})	$0.340^{+0.039}_{-0.18}$	0.86 ± 0.30	$0.98^{+0.11}_{-0.18}$	1.22 ± 0.31	1σ agreement: (II)–(IV) only
e_d	$0.126^{+0.016}_{-0.050}$	$0.1163^{+0.0079}_{-0.053}$	$0.148^{+0.028}_{-0.044}$	$0.142^{+0.026}_{-0.044}$	1σ agreement for all
ω_d	$6.4^{+1.1}_{-1.0}$	$6.44^{+1.8}_{-0.27}$	$6.37^{+0.27}_{-0.62}$	$6.39^{+0.35}_{-0.64}$	1σ agreement for all
$\sigma^+_{\rm RV}$ (m s^{-1})	$2.344^{+0.080}_{-0.00016}$	$1.989^{+0.078}_{-0.11}$	$2.030^{+0.090}_{-0.035}$	$1.96^{+0.10}_{-0.056}$	1σ agreement: (II)–(IV) only
$\sigma^+_{\rm BS}$ (m s^{-1})	$2.63^{+0.11}_{-0.017}$	1.97 ± 0.31	$2.60^{+0.12}_{-0.0091}$	1.99 ± 0.30	1σ agreement: (I) & (III), (II) & (IV)
$\sigma^+_{\rm HK}$	$0.02105^{+0.00070}_{-0.00010}$	0.0038 ± 0.0010	$0.02100^{+0.00077}_{-0.00015}$	$0.00363^{+0.00052}_{-0.00043}$	1σ agreement: (I) & (III), (II) & (IV)
$\gamma_{\rm RV}$(m s^{-1})	-30651.55 ± 0.16	-30651.50 ± 0.19	-0.14 ± 0.15	-0.11 ± 0.17	Cf. comments in Section 3
$\gamma_{\rm BS}$(m s^{-1})	4.60 ± 0.20	4.48 ± 0.23	4.59 ± 0.21	4.47 ± 0.23	1σ agreement for all
$\gamma_{\rm HK}$	-4.8705 ± 0.0012	-4.8703 ± 0.0019	-4.8706 ± 0.0014	-4.8702 ± 0.0020	1σ agreement for all
$P_{\rm GP}$ (d)	–	$29.32^{+0.54}_{-0.80}$	–	$29.46^{+0.57}_{-0.93}$	1σ agreement
$\lambda_{\rm e}$ (d)	–	$29.0^{+1.1}_{-4.3}$	–	$29.3^{+1.2}_{-4.3}$	1σ agreement
$\lambda_{\rm p}$	–	$0.687^{+0.027}_{-0.087}$	–	$0.722^{+0.032}_{-0.098}$	1σ agreement
V_r (m s^{-1})	–	$3.49^{+1.8}_{-0.059}$	–	$1.340^{+0.053}_{-0.88}$	(II) and (IV) discrepant; cf. Section 6.2
V_c (m s^{-1})	–	$-1.65^{+0.20}_{-0.12}$	–	$-0.78^{+0.16}_{-0.14}$	(II) and (IV) discrepant; cf. Section 6.2
B_r (m s^{-1})	–	-5.8 ± 2.0	–	-5.9 ± 2.2	1σ agreement
B_c (m s^{-1})	–	$-1.76^{+0.28}_{-0.12}$	–	$-1.88^{+0.28}_{-0.17}$	1σ agreement
L_c	–	$-0.0253^{+0.0025}_{-0.00086}$	–	$-0.0271^{+0.0030}_{-0.00081}$	1σ agreement

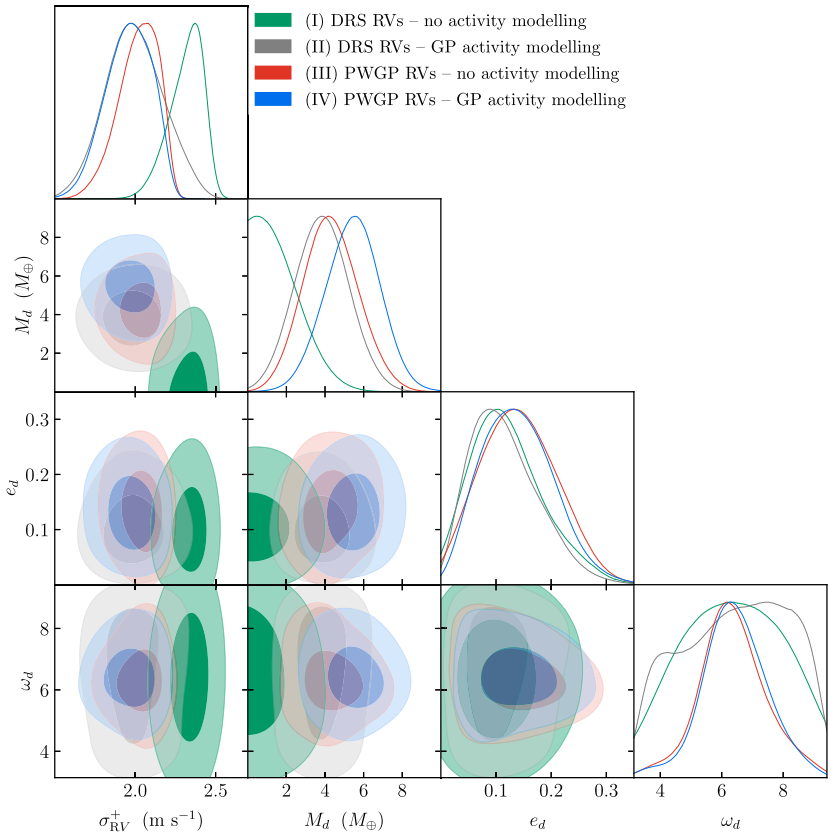

Figure 5. Corner plot showing marginalized 1D and 2D posterior probability densities for four key parameters shared between modelling approaches (I)–(IV), in each case where Kepler-37d was the only Keplerian component in the model. The dark and light filled regions correspond, respectively, to 1σ (39.3 per cent) and 2σ (86.5 per cent) joint credible regions. ω_d has been transformed to the domain $[\pi, 3\pi]$ to suppress boundary effects, per the caption to Table 12.

37e – that were inconsistent with zero at $>2\sigma$ levels, superficially suggesting a detection, even though the model itself was ultimately rejected. A subtle caveat, though, is that Bayesian model selection can never decide whether a model is 'correct' in an absolute sense – hence the need to evaluate many competing, plausible models.

Finally, when computing Bayesian evidences with POLYCHORD, we deliberately used very stringent convergence criteria to try to ensure robust and consistent runs; as shown by Nelson et al. (2020), Ahrer et al. (2021), and others, obtaining very large scatters in model evidences is often a pernicious problem when evaluating multi-Keplerian models, and could easily result in both false-positive or false-negative detections. Reassuringly, we obtained very small scatters in model evidences computed across multiple independent runs.

6.2 Parameter inference

As the model with the highest evidence for both the DRS and PWGP RVs contained Kepler-37d as the sole Keplerian component, we focus here on the stellar and planetary parameters inferred under modelling approaches (I)–(IV), in each case where Kepler-37d was the only Keplerian term in the model.

Table 12 summarizes the marginalized 1D posteriors for all parameters in modelling approaches (I)–(IV). We highlight here a few of the most salient findings.

Crucially, the parameter posteriors for Kepler-37d show strong reassuring consistency across all four modelling approaches – with the exception of the RV semi-amplitude K_d, which is consistent within 1σ between approaches (II)–(IV) only, as Kepler-37d was not detected under approach (I). Far from the GP 'eating up' a planetary signal (a concern often voiced in the RV community), our careful use of GP modelling across multiple time series allowed a subtle planetary signal to be recovered from correlated stellar nuisance signals. Indeed, the RV additive white-noise parameter, σ^+_{RV}, is consistent within 1σ between approaches (II)–(IV), but significantly larger in approach (I), where Kepler-37d's signal evidently remains 'buried' behind stellar nuisance signals that were modelled as white noise. Given the consistency between posteriors from approaches (II)–(IV), we avoid the temptation to over-interpret any supposed differences between the results, e.g. the marginally smaller error bars on K_d in approach (III) versus (IV).

Fig. 5 shows 1D and 2D marginalized posteriors for the aforesaid parameters, albeit with semi-amplitude K_d converted into a more physically interesting mass M_d, across approaches (I)–(IV). Meanwhile Fig. 6 shows Kepler-37 RVs folded to the orbital period of Kepler-37d, after subtracting applicable stellar activity models, again across approaches (I)–(IV).

Turning to the activity modelling, Table 12 shows that, unsurprisingly, σ^+_{BS} and σ^+_{HK} end up significantly larger in approaches (I) and (III) than in (II) and (IV), since in the latter two cases, the GP allows variability in the BIS and $\log R'_{HK}$ time series to be modelled as correlated signals (which they certainly are), rather than white noise.

The quasi-periodic GP hyper-parameters are consistent between approaches (II) and (IV); they suggest a stellar rotation period of $P_{GP} = 29 \pm 1$ d, and an active region evolution time-scale of order one rotation period, i.e. $\lambda_e = 29^{+1}_{-4}$ d. The hyper-parameter $\lambda_p \sim 0.7$ suggests the stellar RV signals modelled were of moderate harmonic complexity, having roughly twice as many turning points per period

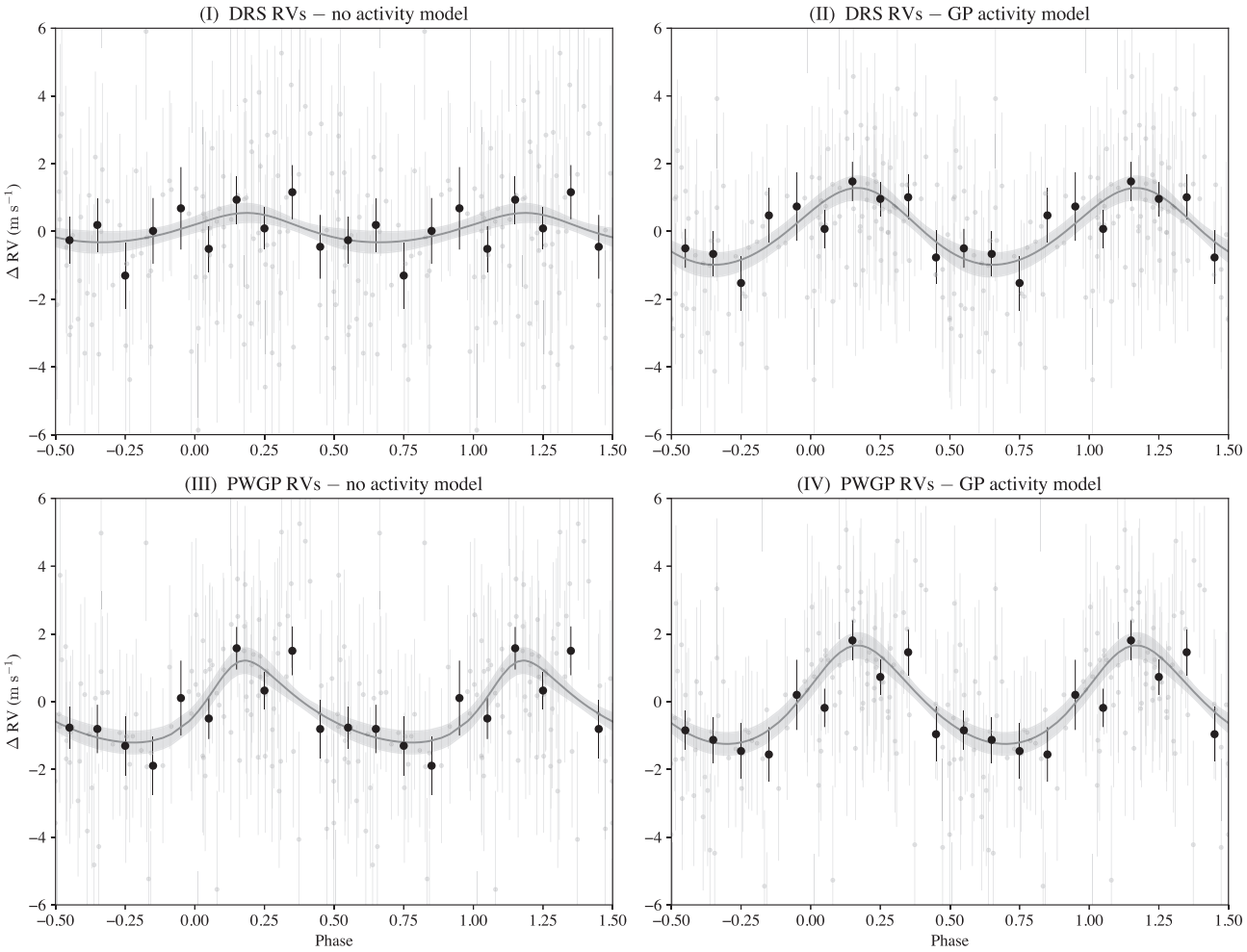

Figure 6. Kepler-37 RVs folded to the orbital period of Kepler-37d. The grey points show the DRS RVs (top panels) or PWGP RVs (bottom panels), either directly from the extraction pipelines (left-hand panels) or after subtracting the best-fitting GP activity model (right-hand panels); the associated error bars are the pipeline 1σ error bars. The solid grey lines indicate maximum *a posteriori* (MAP) models for Kepler-37d in approaches (I)–(IV), with the shaded grey regions indicating $\pm\sigma$ predictive uncertainty, as derived from the posterior uncertainty in the orbital parameters. Finally, the black points represent the RVs averaged into ten equally-spaced phase bins; here, the associated error bars are the standard errors on the mean RV of the points in each bin.

as a sine wave (Rajpaul 2017). The relatively tight constraint on the stellar rotation period under the GP model contrasts strongly with the weak constraints from GLS analyses of activity indicators (see e.g. Fig. 2); these findings are unsurprising, though, given that the activity signal is both non-sinusoidal and fairly rapidly evolving, and therefore inherently unsuitable for characterization via a GLS periodogram.

Importantly, we found that the GP's RV covariance amplitudes (V_r, V_c) ended up about a factor of two smaller in approach (IV) than (II), though the amplitudes for the other time series (B_r, B_c, L_c) were consistent across the two approaches. This is compelling evidence that our efforts to mitigate RV activity contamination via the PWGP approach were successful, albeit imperfect – the latter also evidenced by the improvements seen when moving from approach (III) to (IV). The fact that $|V_r| > |V_c|$ and $|B_r| > |B_c|$ suggests the stellar signals we modelled arose primarily from rotating active regions, rather than their associated suppression of convective blueshift (Aigrain et al. 2012; Rajpaul et al. 2015).

Fig. 7 shows 1D and 2D marginalized posteriors for the aforementioned GP (hyper-)parameters, across approaches (I)–(IV), while our best-fitting model from approach (IV), including the simultaneous GP fit to the RV, BIS, and log $R'_{\rm HK}$ time series, is shown in Fig. 8.

While the log $R'_{\rm HK}$ series is clearly dominated by oscillations with a ∼30 d (quasi-)period, it also shows some evidence for a long-term trend, with the median values in the 2014 season being higher than those in the 2015 season. Though not visible in Fig. 7, this apparent trend continued in 2019, when log $R'_{\rm HK} \leq -4.93$ in all three spectra taken that year, suggesting the star was moving to a lower-activity phase over the years we observed it. A similar trend was evident in the FWHM series. These trends correspond to some of the lower-frequency structures in the log $R'_{\rm HK}$ and FWHM periodograms in Fig. 2, and may be evidence of a longer-term stellar magnetic cycle.

It is also evident, in Fig. 8, that the residual scatter from the GP fit to the log $R'_{\rm HK}$ series is smaller than in the RV and BIS cases. This was to be expected *a priori*, given that the log $R'_{\rm HK}$ error bars are about half the size of the RV or BIS error bars, when considered as a fraction of the rms variation in the respective time series. This difference is further amplified when factoring in the proportionately larger RV and BIS white-noise jitter terms (see Table 12).

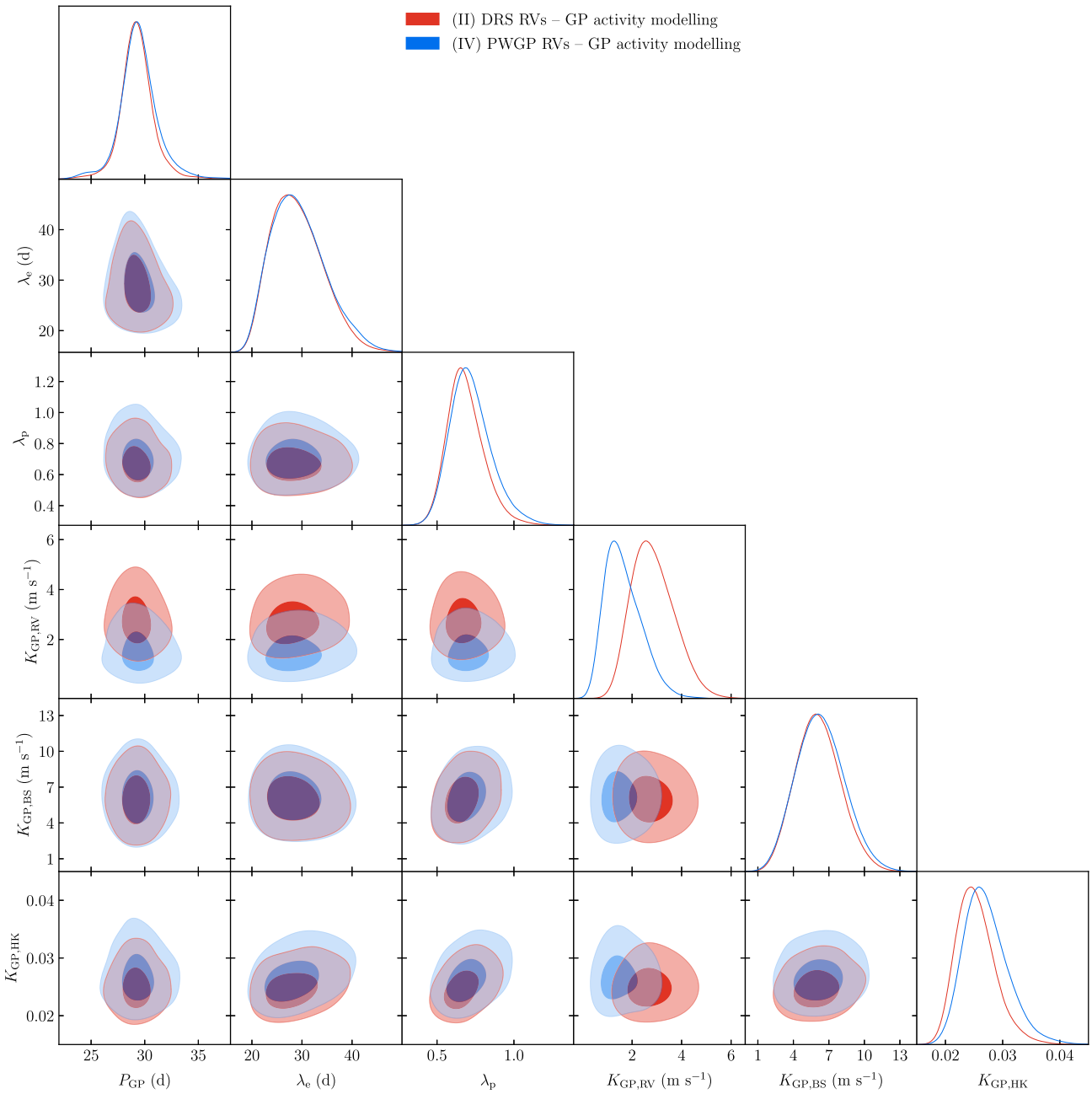

Figure 7. Corner plot as in Fig. 5, but now for GP-related (hyper-)parameters in the two GP-based modelling approaches, in each case where Kepler-37d was the only Keplerian component. To simplify interpretation and suppress sign degeneracy bimodalities inherent in the parameters V_c, V_r, B_c, B_r and L_c, we plot the following (positive semidefinite) semi-amplitudes: $K_{GP,RV} \equiv (V_r^2 + V_c^2)^{1/2}$, $K_{GP,BS} \equiv (B_r^2 + B_c^2)^{1/2}$, $K_{GP,HK} \equiv |L_c|$.

6.3 Dynamical stability analysis

We used the code MERCURY-T (Bolmont et al. 2015) to run full dynamical simulations of the Kepler-37 system, including general relativistic precession and precession due to stellar rotational flattening. Using the parameter values from approach (IV) in Table 12, we found that the system was unstable over 10^6 yr, mostly through the eccentricity excitation and ejection of Kepler-37b; this did not depend on the inclusion of candidate planet e. However, when we lowered Kepler-37d's mass and eccentricity to the lower end of our 1σ posterior credible interval ($M_d = 4\ M_\oplus$, $e_d = 0.075$) and assumed Kepler-37b was iron-rich ($K_b \sim 0.6$ cm s^{-1}, $M_b \sim 0.03\ M_\oplus$), the system generally remained stable over 10^6 yr, again regardless of whether Kepler-37e was included.

We are obliged to conclude that Kepler-37d's RV semi-amplitude and/or eccentricity are on the lower end of our inferred values, or that Kepler-37b's RV semi-amplitude is higher than our MAP value ($K_b \sim 0.3$ cm s^{-1}). However, our RVs essentially do not constrain the orbital parameters of Kepler-37b – their posteriors being overwhelmingly shaped by our broad priors – yet *do* place strong constraints on the orbit of Kepler-37d. Under all modelling approaches, we ended up with a moderate eccentricity ($e_d \sim 0.12$ or $e_d \sim 0.14$) inconsistent with zero at a 2–3σ level. As such, we are inclined to favour Kepler-37b being iron-rich (perhaps having ejected much of a former

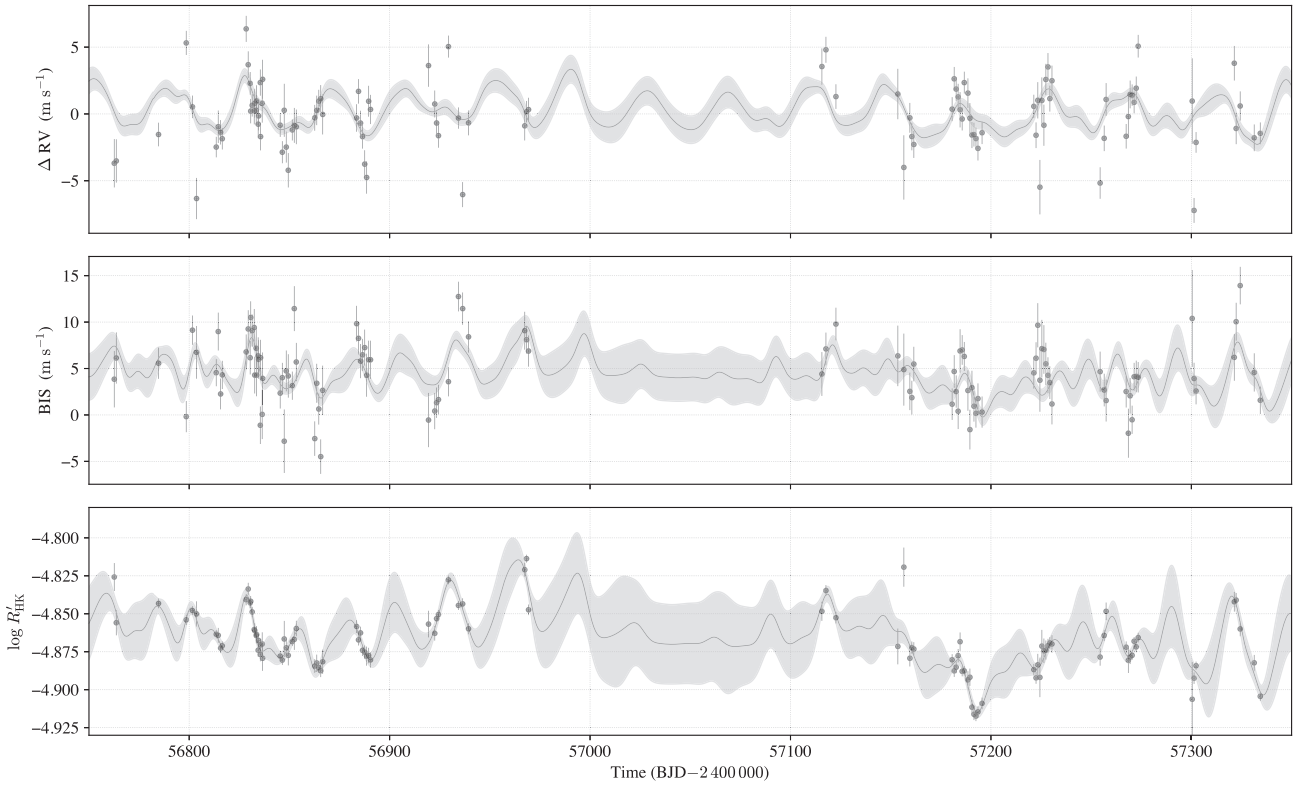

Figure 8. Joint fits to RV, BIS, and log $R'_{\rm HK}$ time series under approach (IV), for the model containing Kepler-37d but no other planets – i.e. the model with the highest marginal likelihood. The dark grey points indicate observed values (with error bars); the solid grey lines indicate GP posterior predictive means (plus best-fitting Keplerian for the RVs), and the shaded regions denote $\pm\sigma$ posterior predictive uncertainties. Three observations from 2019 are not visible here.

silicate mantle in a collision), rather than Kepler-37d's orbit having negligible eccentricity. In any case, whether $e_{\rm d}$ is zero or moderately large has negligible impact on the mass $M_{\rm d}$ inferred from $K_{\rm d}$.

There may *also* be an orbital configuration that could stabilize the whole system, even with Kepler-37d being relatively massive and on a significantly eccentric orbit. However, exploring a much wider range of parameter space is beyond the scope of this paper (cf. comments about computational burdens in Section 6.1.7). In the future, joint GP activity modelling and dynamical simulations would likely help refine the orbital parameters for Kepler-37d, if not the other transiting planets.

6.4 Kepler-37e: re-examining Kepler-37d's TTVs

We carried out a separate dynamical analysis and re-examination of the Kepler-37 TTVs, hoping to shed light on the nature of the ∼51-d period signal currently ascribed to Kepler-37e.

The periods of Kepler-37d and e are close to a 4:3 commensurability, hinting at a first-order mean motion resonance, which may suggest the presence of a detectable TTV signal induced on planet d due to strong dynamical interactions between the two planets. Consequently, by studying the TTV signal of Kepler-37d, we can infer some information on the hypothetical perturbing planet, i.e. Kepler-37e. According to Mazeh et al. (2013), whose transit times were used in the analysis by Hadden & Lithwick (2014), Kepler-37d has a shallow TTV signal with an amplitude of order 1 min, as reported in the observed minus calculated diagram (O−C; Agol & Fabrycky 2018) in Fig. 9.

We numerically integrated the orbits of Kepler-37b, c, d, and e using the *N*-body integrator `ias15` within the `rebound` package

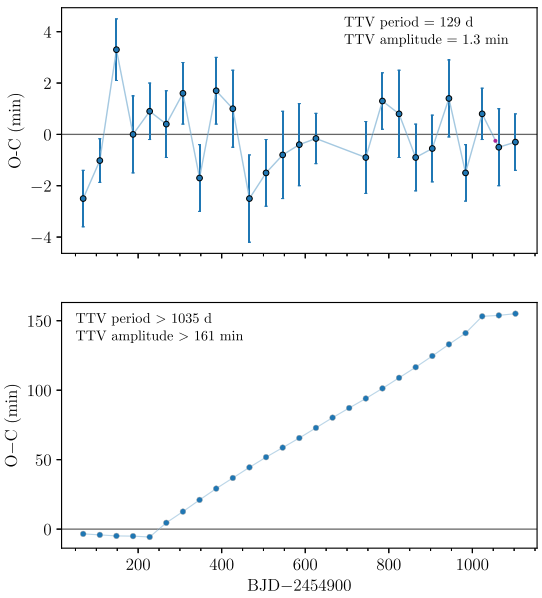

Figure 9. Top: Measured TTV signal of Kepler-37d according to the transit times reported in Mazeh et al. (2013). The TTV period and amplitude are computed via the GLS periodogram (Zechmeister & Kürster 2009). Bottom: Predicted TTV signal of Kepler-37d according to our numerical simulation, assuming the presence of a planet with ∼51-d period. The simulation suggests a TTV period longer than the observed baseline, and amplitude $\gtrsim$ 160 min.

Table 13. Selected fitted and derived parameters for the three planets known to transit Kepler-37. Orbital periods, radii, and orbital inclinations in Table 2 were used to derive planetary masses and densities. The parameter posteriors for Kepler-37b were essentially the same as our priors.

Parameter	Posterior	Notes
K_b (m s^{-1})	<0.007	95 per cent upper limit
M_b ($M_\oplus$)	<0.02	95 per cent upper limit
ρ_b (g cm^{-3})	<6	95 per cent upper limit
K_c (m s^{-1})	<0.1	95 per cent upper limit
M_c ($M_\oplus$)	<0.6	95 per cent upper limit
ρ_c (g cm^{-3})	<5	95 per cent upper limit
K_d (m s^{-1})	1.22 ± 0.31	3.9σ detection
e_d	$0.142^{+0.026}_{-0.044}$	–
a_d (AU)	0.2109 ± 0.0029	Orbital semi-major axis
M_d ($M_\oplus$)	5.4 ± 1.4	3.9σ detection
ρ_d (g cm^{-3})	$4.29^{+0.52}_{-0.74}$	–

(Rein & Liu 2012), assuming as reference time the transit mid-point time ($T_{0,\,d}$) of Kepler-37d. As initial configuration, we assumed the planetary parameters in Table 2, except for the mass and eccentricity of Kepler-37d, for which we used our inferred values (Table 13). For planet e, since Hadden & Lithwick (2014) only report the planetary period, we used the mid-transit time as originally reported by Batalha et al. (2013), $T_0 = 2455028.727 \pm 0.0096$ BJD$_{\rm TDB}$, and we derived the radius from the planet-to-star radius ratio given by Batalha et al. (2013): $R_e/R_\star = 0.0054 \pm 0.0002$, $R_e = 0.43 \pm 0.03$ $R_\oplus$. We estimated the masses of planets b, c, and e using the Wolfgang, Rogers & Ford (2016)[8] probabilistic mass−radius relation: $M_b = 0.01$ $M_\oplus$, $M_c = 0.6$ $M_\oplus$, $M_e = 0.06$ $M_\oplus$. During the integration, we computed the synthetic transit times (O) of each planet following the procedure described in Borsato et al. (2019), and compared the inferred transit times with the linear ephemeris (C) of Mazeh et al. (2013), obtaining the synthetic O−C diagram to identify a possible TTV signal.

Despite the low expected mass of the putative planet e, the predicted TTV signal induced on Kepler-37d has a high amplitude due to the suggested resonant configuration of the planets and to the eccentricity of planet d. Fig. 9 shows that neither the amplitude nor the period of such TTV signal corresponds to the observed TTVs.

It is worth noting that the TTV prediction is highly dependent on the planetary parameters, and in particular, even small changes in the eccentricity of Kepler-37d imply a variation of TTV amplitude from 0 to more than 160 min. However, our simulations predict that in order to obtain a TTV amplitude of order of the observed one, the eccentricity of Kepler-37d would need to be negligible ($e_d \leq 0.01$), whereas our RV modelling points towards a non-trivial eccentricity.

We also computed the same forward modelling assuming a three-planet system, i.e. without the presence of Kepler-37e. In this case, the amplitude of the predicted TTV signal of Kepler-37d was of order 0.2 min, that is, lower than the average error on the transit times of Mazeh et al. (2013), indicating that the two inner planets do not significantly perturb the orbit of planet d.

In summary, our TTV analysis disfavours the presence of the putative planet Kepler-37e. Considering the dubious way in which Kepler-37e became a 'confirmed' planet in the first place (see Section 2.2.2), the fact that our RV analysis did not lead to the detection of a signal with ∼51-d period, the prominence of the ∼51-d periodicity in two activity indicators, and the additional doubts introduced by our TTV forward modelling, we suggest that Kepler-37e be stripped of its status of a 'confirmed' planet.

We do not exclude the possibility of a non-transiting planet inducing the small observed TTV signal of Kepler-37d, which cannot be totally accounted for by a three-planet system, but the properties of such a non-transiting planet would not seem to correspond to those of Kepler-37e as reported in the literature.

[8] https://github.com/dawolfgang/MRrelation.

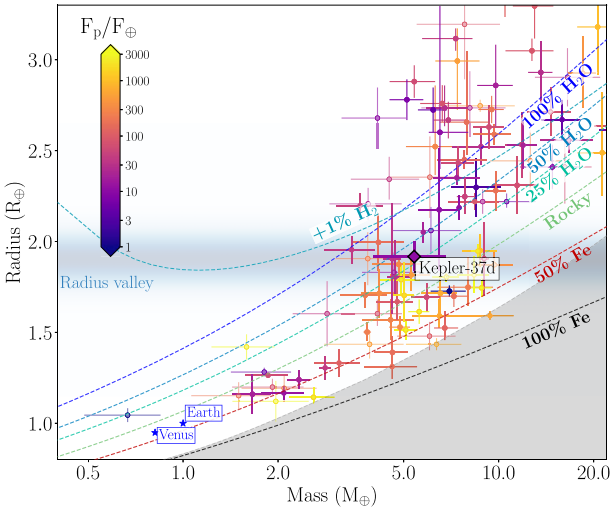

Figure 10. Mass−radius diagram for exoplanets with masses and radii measured with better than 30 per cent precision, colour coded according to their incidental flux (Earth units). Kepler-37d is indicated with a large diamond. The dashed coloured lines represent theoretical mass−radius curves for various chemical compositions according to Zeng et al. (2019). The shaded grey region denotes the maximum Fe content predicted by collisional stripping (Marcus et al. 2010). Data taken from the The Extrasolar Planets Encyclopaedia (http://exoplanet.eu/catalog/) on 2021 March 26.

6.5 Final characterization of transiting planets

In Table 13 we summarize our final fitted and derived parameters for Kepler-37's transiting planets – for Kepler-37d, drawn from our highest-evidence model from approach (IV), and for planets b and c, drawn from the best-fitting model including these Keplerians under approach (IV). We assume Kepler-37e is not present, and disregard any ambiguities raised by our dynamical modelling (Section 6.3).

In Fig. 10 we show the position of Kepler-37d in the mass−radius diagram of exoplanets with masses and radii both measured with better than 30 per cent precision. According to our derived density ($\rho_d = 4.29^{+0.52}_{-0.74}$ g cm^{-3}), the interior composition of Kepler-37d differs from that of an Earth-like planet (around 30 per cent Fe and 70 per cent silicates) at a $>2\sigma$-level in both radius and mass. Instead, its density is consistent with a ∼25 per cent H_2O composition, making it a compatible with a water-world scenario, where 'water worlds' are planets with massive water envelopes, in the form of high pressure H_2O ice, comprising >5 per cent of the total mass. However, its density could also be explained via a gaseous envelope surrounding a rocky core. Following Lopez & Fortney (2014), assuming a rocky Earth-like core, a Solar composition H-He envelope, and using our derived planetary and stellar parameters, we estimate that an H-He envelope comprising ∼0.4 per cent of the planet mass could also explain the observed properties of Kepler-37d.

On the other hand, if Kepler-37d actually has a mass closer to $M_d \sim 4$ $M_\oplus$ (as our dynamical modelling tentatively suggests could be the case), its consequent density $\rho_d \sim 3.2$ g cm^{-3} would be more consistent with a ~ 50 per cent water composition.

We note that Kepler-37d lies within the so-called small planet radius gap or 'Fulton gap' (Fulton et al. 2017). According to recent XUV irradiation modelling by Modirrousta-Galian, Locci & Micela (2020), which reproduces the bimodal exoplanet radius distribution observed by Fulton, roughly equal proportions of planets with Kepler-37d's 1.9 $R_\oplus$ radius are expected to have atmospheres versus no atmospheres.

7 CONCLUSIONS

With an RV semi-amplitude of 1.22 ± 0.31 m s^{-1}, Kepler-37d is one of only a handful of transiting planets securely detected below a 2 m s^{-1} RV threshold. Its RV semi-amplitude is only slightly larger than (though formally consistent with) TOI-178b's $1.05^{+0.25}_{-0.30}$ m s^{-1}, the latter RV signal detected with ESPRESSO, and currently the smallest of any transiting planet detected to date. By extension, Kepler-37d has (to our knowledge) the smallest RV semi-amplitude of any transiting planet characterized with HARPS-N alone, or indeed with any single second-generation spectrograph. Based on our RV modelling alone, our best estimate of Kepler-37d's mass is 5.4 ± 1.4 $M_\oplus$, although full dynamical simulations suggest its mass might be on the lower end of that 1σ credible interval.

As expected, we did not detect either Kepler-37b or c, whose RV semi-amplitudes lie far below the threshold of current detectability. Our RV modelling and re-examination of Kepler-37's TTVs suggested that the putative planet Kepler-37e should probably be stripped of its 'confirmed planet' status.

Stepping back from the specifics of the Kepler-37 system, our analysis served as a case study in stellar activity modelling and mitigation. We showed how (i) careful modelling of stellar activity in HARPS-N pipeline RVs and activity indicators, or (ii) mitigating stellar activity *before* RV extraction, in either case using state-of-the-art tools, enabled a RV detection of a planet that had eluded secure RV characterization for many years. The PWGP RV extraction technique we leveraged in the latter approach appears to have some advantages over the DRS; a systematic comparison of the DRS with many different extraction techniques – in the vein of the EXPRES Stellar-Signals Project (Zhao et al. 2020), ideally including spectra with injected planetary signals – could be useful to illuminate their relative merits.

Importantly, we demonstrated that the aforesaid approaches of activity mitigation and modelling not only yielded consistent results, but can be complementary: when combined, they led to a decisive detection and more precise characterization of Kepler-37d. Such a two-pronged approach might be crucial for enabling the detection of Earth-twin exoplanets with next-generation surveys such as the Terra Hunting Experiment (Hall et al. 2018).

ACKNOWLEDGEMENTS

We are grateful to the anonymous reviewer whose insights helped improve this work. We also wish to thank Howard Isaacson and Courtney Dressing for their assistance interpreting the published HIRES RVs, and for providing newer RVs extracted using the latest HIRES pipeline.

VMR acknowledges the Royal Astronomical Society and Emmanuel College for financial support, and the Cambridge Service for Data Driven Discovery (CSD3) – operated by the University of Cambridge Research Computing Service (www.csd3.cam.ac.uk), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/P020259/1), and DiRAC funding from the Science and Technology Facilities Council (www.dirac.ac.uk) – for providing computing resources. GL and LBo acknowledge funding from the Italian Space Agency (ASI) via 'Accordo ASI-INAF n. 2013-016-R.0 del 9 luglio 2013 e integrazione del 9 luglio 2015 CHEOPS Fasi A/B/C'; GL also acknowledges support from CARIPARO Foundation, via agreement CARIPARO-Università degli Studi di Padova (Pratica n. 2018/0098). SA acknowledges funding from the UK Science and Technology Facilities (STFC) research council via consolidated grant ST/S000488/1, and from the European Research Council (ERC) via grant agreement n. 865624. AMo acknowledges support from the senior Kavli Institute Fellowships. FP acknowledges the Swiss National Science Foundation (SNSF) for supporting research with HARPS-N through SNSF grants n. 140649, 152721, 166227, and 184618.

This work is based on observations made with the Italian Telescopio Nazionale Galileo (TNG) operated on the island of La Palma by the Fundación Galileo Galilei of the INAF (Istituto Nazionale di Astrofisica) at the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofisica de Canarias. The HARPS-N project was funded by the ESA-PRODEX Program of the Swiss Space Office (SSO), the Harvard University Origin of Life Initiative (HUOLI), the Scottish Universities Physics Alliance (SUPA), the University of Geneva, the Smithsonian Astrophysical Observatory (SAO), the Italian National Astrophysical Institute (INAF), University of St. Andrews, Queen's University Belfast, and University of Edinburgh.

DATA AVAILABILITY

All RVs and activity indicators used in our analyses will be available via VizieR at CDS.

REFERENCES

Agol E., Fabrycky D. C., 2018, Handbook of Exoplanets. Springer, Cham, p. 797
Ahrer E. et al., 2021, MNRAS, 503, 1248
Aigrain S., Pont F., Zucker S., 2012, MNRAS, 419, 3147
Andrae R. et al., 2018, A&A, 616, A8
Anglada-Escudé G., Butler R. P., 2012, ApJS, 200, 15
Balona L. A., 2002, MNRAS, 337, 1059
Baranne A. et al., 1996, A&AS, 119, 373
Baranne A., Mayor M., Poncet J. L., 1979, Vistas Astron., 23, 279
Barclay T. et al., 2013, Nature, 494, 452
Batalha N. M. et al., 2013, ApJS, 204, 24
Bedell M., Hogg D. W., Foreman-Mackey D., Montet B. T., Luger R., 2019, AJ, 158, 164
Berger T. A., Huber D., Gaidos E., van Saders J. L., 2018, ApJ, 866, 99
Blackman R. T. et al., 2020, AJ, 159, 238
Blunt S., Fulton B., Petigura E., Howard A., Sinukoff E., 2018, American Astronomical Society Meeting Abstracts #231. p. 439.22
Boisse I. et al., 2009, A&A, 495, 959
Boller T., Freyberg M. J., Trümper J., Haberl F., Voges W., Nandra K., 2016, A&A, 588, A103
Bolmont E., Raymond S. N., Leconte J., Hersant F., Correia A. C. M., 2015, A&A, 583, A116
Bonfils X. et al., 2018, A&A, 613, A25
Borsato L. et al., 2019, MNRAS, 484, 3233
Bouchy F., Pepe F., Queloz D., 2001, A&A, 374, 733
Box G. E., Draper N. R., 1987, Empirical Model-Building and Response Surfaces. Wiley, New York

Buchhave L. A. et al., 2012, Nature, 486, 375
Buchhave L. A. et al., 2014, Nature, 509, 593
Butler R. P., Marcy G. W., Williams E., McCarthy C., Dosanjh P., Vogt S. S., 1996, PASP, 108, 500
Butler R. P. et al., 2017, AJ, 153, 208
Campbell B., Walker G. A. H., 1979, PASP, 91, 540
Collier Cameron A. et al., 2020, MNRAS, 505, 1699
Cosentino R. et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham, WA, p. 84461V
Cutri R. M. et al., 2003, VizieR Online Data Catalog. p. II/246
Cutri R. M., et al., 2014, VizieR Online Data Catalog. p. II/328
Damasso M. et al., 2020, A&A, 642, A133
de Beurs Z. L. et al., 2020, preprint (arXiv:2011.00003)
Di Marcantonio P. et al., 2018, in Alison B. P., Robert L. S., Chris R. B., eds, Observatory Operations: Strategies, Processes, and Systems VII. SPIE, Bellingham, WA, p. 107040F
Dotter A., 2016, ApJS, 222, 8
Dotter A., Chaboyer B., Jevremović D., Kostov V., Baron E., Ferguson J. W., 2008, ApJS, 178, 89
Dumusque X., 2012, PhD thesis, University of Geneva
Dumusque X., 2018, A&A, 620, A47
Feroz F., Hobson M. P., 2008, MNRAS, 384, 449
Feroz F., Hobson M. P., 2014, MNRAS, 437, 3540
Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601
Fischer H., 2011, A History of the Central Limit Theorem. Springer, New York
Fischer D. A. et al., 2016, PASP, 128, 066001
Fulton B. J. et al., 2017, AJ, 154, 109
Gaia Collaboration, 2016, A&A, 595, A1
Gaia Collaboration, 2018, A&A, 616, A1
Gaia Collaboration, 2021, A&A, 649, A1
Gajdoš P., Vaňko M., Parimucha Š., 2019, Res. Astron. Astrophys., 19, 041
Gilbertson C., Ford E. B., Jones D. E., Stenning D. C., 2020, ApJ, 905, 155
Gladman B., 1993, Icarus, 106, 247
Gregory P., 2005, Bayesian Logical Data Analysis for the Physical Sciences. Cambridge University Press, Cambridge
Griffin R. F., 1967, ApJ, 148, 465
Grunblatt S. K., Howard A. W., Haywood R. D., 2015, ApJ, 808, 10
Hadden S., Lithwick Y., 2014, ApJ, 787, 80
Hall R. D., Thompson S. J., Handley W., Queloz D., 2018, MNRAS, 479, 2968
Handley W. J., Hobson M. P., Lasenby A. N., 2015, MNRAS, 453, 4384
Haywood R. et al., 2014, Int. J. Astrobiol., 13, 155
Holczer T. et al., 2016, ApJS, 225, 9
Huang C. X., Bakos G. Á., 2014, MNRAS, 442, 674
Huber D. et al., 2013, ApJ, 767, 127
Høg E. et al., 2000, A&A, 355, L27
Jeffreys H., 1961, The Theory of Probability. Oxford University Press, Oxford
Jones D. E., Stenning D. C., Ford E. B., Wolpert R. L., Loredo T. J., Dumusque X., 2017, preprint (arXiv:1711.01318)
Kipping D. M., 2013, MNRAS, 434, L51
Konacki M., Torres G., Sasselov D. D., Jha S., 2003, ApJ, 597, 1076
Kunimoto M., Matthews J. M., 2020, AJ, 159, 248
Lanza A. F. et al., 2010, A&A, 520, A53
Leleu A. et al., 2021, A&A, 649, 29
Lopez E. D., Fortney J. J., 2014, ApJ, 792, 1
Lovis C. et al., 2006, in McLean I. S., Iye M., eds, Proc. SPIE Conf. Ser. Vol. 6269, Ground-based and Airborne Instrumentation for Astronomy, SPIE, Bellingham, p. 62690P
Malavolta L. et al., 2017a, AJ, 153, 224
Malavolta L., Lovis C., Pepe F., Sneden C., Udry S., 2017b, MNRAS, 469, 3965
Malavolta L. et al., 2018, AJ, 155, 107
Maldonado J., Martínez-Arnáiz R. M., Eiroa C., Montes D., Montesinos B., 2010, A&A, 521, A12
Mamajek E. E., Meyer M. R., Liebert J., 2002, AJ, 124, 1670
Mamajek E. E., Meyer M. R., Liebert J., 2006, AJ, 131, 2360
Marcus R. A., Sasselov D., Stewart S. T., Hernquist L., 2010, ApJ, 719, L45
Marcy G. W., Butler R. P., 1992, PASP, 104, 270
Marcy G. W. et al., 2014, ApJS, 210, 20
Mayor M., Queloz D., 1995, Nature, 378, 355
Mazeh T. et al., 2013, ApJS, 208, 16
Milbourne T. W. et al., 2019, ApJ, 874, 107
Modirrousta-Galian D., Locci D., Micela G., 2020, ApJ, 891, 158
Mortier A., Sousa S. G., Adibekyan V. Z., Brandão I. M., Santos N. C., 2014, A&A, 572, A95
Mortier A. et al., 2020, MNRAS, 499, 5004
Morton T. D., 2015, Isochrones: Stellar Model Grid Package. preprint (ascl:1503.010)
Morton T. D., Bryson S. T., Coughlin J. L., Rowe J. F., Ravichandran G., Petigura E. A., Haas M. R., Batalha N. M., 2016, ApJ, 822, 86
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
Nordström B., Latham D. W., Morse J. A., Milone A. A. E., Kurucz R. L., Andersen J., Stefanik R. P., 1994, A&A, 287, 338
Pasquini L. et al., 2008, in Ian S. M., Mark M. C., Proc. SPIE Conf. Ser., Ground-based and Airborne Instrumentation for Astronomy II, SPIE, Bellingham. p. 70141I
Pecaut M. J., Mamajek E. E., 2013, ApJS, 208, 9
Pepe F., Mayor M., Galland F., Naef D., Queloz D., Santos N. C., Udry S., Burnet M., 2002, A&A, 388, 632
Pepe F. et al., 2021, A&A, 645, A96
Perryman M., 2011, The Exoplanet Handbook. Cambridge University Press, Cambridge
Petersburg R. R. et al., 2020, AJ, 159, 187
Probst R. A. et al., 2014, in Ground-based and Airborne Instrumentation for Astronomy V. SPIE, Bellingham, WA, p. 91471C
Rainer M., Borsa F., Affer L., 2020, Exp. Astron., 49, 73
Rajpaul V. M., 2017, PhD thesis, University of Oxford
Rajpaul V., Aigrain S., Osborne M. A., Reece S., Roberts S., 2015, MNRAS, 452, 2269 (R15)
Rajpaul V., Buchhave L. A., Aigrain S., 2017, MNRAS, 471, L125
Rajpaul V. M., Aigrain S., Buchhave L. A., 2020, MNRAS, 492, 3960
Rasmussen C. E., Ghahramani Z., 2000, in Proceedings of the 13th International Conference on Neural Information Processing Systems. NIPS'00. MIT Press, Cambridge, MA, USA, p. 276
Rasmussen C. E., Williams C. K. I., 2006, Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA
Rein H., Liu S. F., 2012, A&A, 537, A128
Rice K. et al., 2019, MNRAS, 484, 3731
Roberts S., Osborne M., Ebden M., Reece S., Gibson N., Aigrain S., 2013, Phil. Trans. R. Soc., 371, 20110550
Robertson P., Mahadevan S., 2014, ApJ, 793, L24
Rupprecht G. et al., 2004, The Exoplanet Hunter HARPS: performance and First Results. SPIE, Bellingham, WA, p. 148
Schneider J., Dedieu C., Le Sidaner P., Savalle R., Zolotukhin I., 2011, A&A, 532, A79
Schrijver C. J., Zwaan C., 2000, Solar and Stellar Magnetic Activity. Cambridge University Press, Cambridge
Schuler S. C. et al., 2015, ApJ, 815, 5
Schwab C. et al., 2016, in Ground-based and Airborne Instrumentation for Astronomy VI. SPIE, Bellingham, WA, p. 99087H
Seager S., ed., 2010, Exoplanets (Space Science Series). University of Arizona Press
Simkin S. M., 1974, A&A, 31, 129
Sivia D., Skilling J., 2006, Data Analysis: A Bayesian Tutorial, 2 edn, Oxford University Press
Sousa S. G., 2014, in Ewa N., Barry S., Wojtek P., eds, ARES + MOOG: A Practical Overview of an Equivalent Width (EW) Method to Derive Stellar Parameters. Determination of Atmospheric Parameters of B-, A-, F- and G-Type Stars, Springer, Cham, p. 297
Sousa S. G., Santos N. C., Israelian G., Lovis C., Mayor M., Silva P. B., Udry S., 2011, A&A, 526, A99
Stassun K. G., Collins K. A., Gaudi B. S., 2017, AJ, 153, 136

Thompson S. J. et al., 2016, in Christopher J. E., Luc S., Hideki T., eds, Ground-based and Airborne Instrumentation for Astronomy VI. SPIE, Bellingham, WA, p. 99086F
Tonry J., Davis M., 1979, AJ, 84, 1511
Tuomi M., Anglada-Escude G., Jenkins J. S., Jones H. R. A., 2014, preprint (arXiv:1405.2016)
Van Eylen V., Albrecht S., 2015, ApJ, 808, 126
VanderPlas J. T., 2018, ApJS, 236, 16
Walkowicz L. M., Basri G. S., 2013, MNRAS, 436, 1883
Wenger M. et al., 2000, A&AS, 143, 9
Wolfgang A., Rogers L. A., Ford E. B., 2016, ApJ, 825, 19
Zechmeister M., Kürster M., 2009, A&A, 496, 577
Zechmeister M. et al., 2018, A&A, 609, A12
Zeng L. et al., 2019, PNAS, 116, 9723
Zhao J., Tinney C. G., 2020, MNRAS, 491, 4131
Zhao L., Fischer D. A., Ford E. B., Henry G. W., Roettenbacher R. M., Brewer J. M., 2020, Res. Notes Am. Astron. Soc., 4, 156

APPENDIX A: KECK-HIRES RADIAL VELOCITIES

There exist 33 publicly-available Keck-HIRES RVs for Kepler-37 (Barclay et al. 2013), extracted from spectra taken between 2010 and 2012, with 18 of them taken in 2011 August. Their RMS is 3.05 $\mathrm{m\,s^{-1}}$, and mean error bar 1.44 $\mathrm{m\,s^{-1}}$. Using these HIRES RVs, Marcy et al. (2014) derived the following masses for Kepler-37's transiting planets (95 per cent upper limits in parenthesis): $M_\mathrm{b} = 2.8 \pm 3.7$ (10.0) $\mathrm{M_\oplus}$, $M_\mathrm{c} = 3.4 \pm 4.0$ (12.0) $\mathrm{M_\oplus}$, $M_\mathrm{d} = 1.9 \pm 9.1$ (12.2) $\mathrm{M_\oplus}$. At face value, these HIRES masses are all consistent with zero.

We ran our models from approaches (I) and (II) using the HIRES RVs alone, as well as HARPS-N plus HIRES RVs. However, we ultimately omitted these thornier analyses from the body of our paper for several reasons. First, the PWGP RV extraction method is not designed for iodine-cell spectra; even if we had the original HIRES spectra at our disposal, we could not have investigated activity mitigation (a primary focus of this study) with those spectra. Secondly, the HIRES RVs were accompanied by Mount Wilson *S*-index measurements only, but no CCF-derived indicators such as the BIS. Thirdly, joint modelling of RVs and activity indicators from both instruments required significantly more free parameters, to accommodate instrumental offsets, independent additive white-noise terms, etc. The HIRES *S*-index series also evinced quadratic-like variability over its 3-yr span, whereas the HARPS-N activity indicators did not.

The upshot from approaches (I) and (II) was that we did not detect Kepler-37d in the HIRES RVs, and made only a doubtful detection when combining them with the HARPS-N RVs under approach (II). In the latter case, we inferred $K_d = 0.6 \pm 0.28$ $\mathrm{m\,s^{-1}}$, albeit with $\mathcal{R}_{d0} \gtrsim 1$. Re-running all our models with HIRES RVs extracted with the newest pipeline available in 2020 July did not change our conclusions. This seems to parallel the confounding situation in which HIRES RVs, HARPS-N RVs, and their combination each implies discrepant masses for Kepler-10c (e.g. Rajpaul, Buchhave & Aigrain 2017).

We may speculate about why the HIRES RVs hinder our characterization of Kepler-37d (which, unlike Kepler-10c, our HARPS-N RVs indicate to have unexceptional properties). The HIRES spectra were taken years before the first HARPS-N spectra, when Kepler-37 exhibited significantly more activity-related variability (based on *S*-index measurements); stronger stellar signals in the HIRES RVs may exert disproportionate influence on model fits. The HIRES and HARPS-N spectra cover slightly different wavelength ranges (360–800 nm versus 383–690 nm), so stellar signals measured by each instrument could have different properties – something our modelling did not account for – regardless of when the star is observed.

Instrumental systematics peculiar to HIRES are another possibility. We note that 8 of the 33 HIRES observations were made using the B5 decker/slit combination, with the remainder made using the C2 decker/slit combination. While the modes share a $R = 48000$ resolution and 0.861 arcsec slit width, their slit lengths differ, the former being 3.5 arcsec versus 14 arcsec for the latter. Ongoing investigations suggest HIRES RVs taken with the B5 aperture may be associated with significantly higher RV scatters, at least in the case of Kepler-10. Thus, a relatively small number of outlying HIRES RVs, with under-estimated error bars, might be dominating model fits.

Whatever the case, considering that many hundreds of RVs and ongoing analyses have still not reconciled HIRES and HARPS-N RVs for Kepler-10, attempting to do so with Kepler-37 was well beyond the scope of this study. Fortunately, our detection of Kepler-37d was straightforward and secure using HARPS-N observations alone.

[1]*Astrophysics Group, Cavendish Laboratory, JJ Thomson Avenue, Cambridge CB3 0HE, UK*
[2]*Sub-department of Astrophysics, Department of Physics, University of Oxford, Oxford OX1 3RH, UK*
[3]*DTU Space, National Space Institute, Technical University of Denmark, Elektrovej 328, DK-2800 Kgs. Lyngby, Denmark*
[4]*Department of Physics and Astronomy, Università degli Studi di Padova, Vicolo dell'Osservatorio 3, IT-35122 Padova, Italy*
[5]*INAF – Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, IT-35122 Padova, Italy*
[6]*SUPA, Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, Scotland, UK*
[7]*Centre for Exoplanet Science, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK*
[8]*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[9]*Department of Astronomy, University of California Berkeley, Berkeley, CA 94720-3411, USA*
[10]*Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*
[11]*INAF – Osservatorio Astrofisico di Torino, via Osservatorio 20, I-10025 Pino Torinese, Italy*
[12]*Observatoire Astronomique de l'Université de Genève, Chemin Pegasi 51b, CH-1290 Versoix, Switzerland*
[13]*INAF – Fundación Galileo Galilei, Rambla José Ana Fernandez Pérez 7, E-38712 Breña Baja, TF, Spain*
[14]*INAF – Osservatorio Astronomico di Palermo, P.za Parlamento 1, I-90134 Palermo, Italy*
[15]*INAF – Osservatorio Astronomico di Cagliari, via della Scienza 5, I-09047 Selargius, Italy*
[16]*INAF – Osservatorio Astronomico di Brera, via E. Bianchi 46, I-23807 Merate (LC), Italy*
[17]*Centre for Exoplanets and Habitability, University of Warwick, Coventry CV4 7AL, UK*
[18]*Department of Physics, University of Warwick, Coventry CV4 7AL, UK*
[19]*Astrophysics Research Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, UK*

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