

The impact of two non-transiting planets and stellar activity on mass determinations for the super-Earth CoRoT-7b

Ancy Anna John,^{1,2*} Andrew Collier Cameron,^{1,2} Thomas G. Wilson^{1,2}

¹*SUPA, School of Physics & Astronomy, University of St Andrews, North Haugh, St Andrews, KY169SS, UK*

²*Centre for Exoplanet Science, University of St Andrews, North Haugh, St Andrews, KY169SS, UK*

30 June 2022

ABSTRACT

CoRoT-7 is an active star, whose orbiting planets and their masses have been under debate since their initial detection. In the previous studies, CoRoT-7 was found to have two planets, CoRoT-7b and CoRoT-7c with orbital periods 0.85 and 3.69 days, and a potential third planet with a period~9 days. The existence of the third planet has been questioned as potentially being an activity-induced artefact. Mass of the transiting planet CoRoT-7b has been estimated to have widely different values owing to the activity level of the parent star, the consequent RV “jitter”, and the methods used to rectify this ambiguity. Here we present an analysis of the HARPS archival RV (RV) data of CoRoT-7 using a new wavelength-domain technique, SCALPELS, to correct for the stellar activity-induced spectral line-shape changes. Simultaneous modelling of stellar activity and orbital motions, identified using the ℓ_1 -periodogram, shows that SCALPELS effectively reduce the contribution of stellar variability to the RV signal and enhance the detectability of exoplanets around active stars. Using KIMA nested-sampling package, we modelled the system incorporating a Gaussian Process together with SCALPELS. The resultant posterior distributions favoured a three-planet system comprising two non-transiting planets, CoRoT-7c and CoRoT-7d with orbital periods 3.697 ± 0.005 and 8.966 ± 1.546 days, in addition to the known transiting planet. The transiting planet CoRoT-7b is found to be a rocky super-Earth with a mass of $M_b = 6.06 \pm 0.65 M_{\oplus}$. The determined masses of $M_c = 13.29 \pm 0.69 M_{\oplus}$ and $M_d = 17.14 \pm 2.55 M_{\oplus}$ suggest the non-transiting planets CoRoT-7c and CoRoT-7d to be structurally similar to Uranus and Neptune.

Key words: planets and satellites: detection, techniques: radial velocities, stars: activity, line: profiles, individual: CoRoT-7 – planetary system.

1 INTRODUCTION

Orbiting planets exert a gravitational pull on their host star, which results in a periodic variation of the star’s velocity along the line of sight of observation. This wobbling effect can be inferred as radial velocity (RV) information, which is a measure of the associated wavelength shifts. Shortly after the detection of first exoplanet 51 Pegasi-b, Saar & Donahue (1997) recognized that stellar activity severely impacts the measurement of RVs. If the variable stellar surface phenomena influence the RV measurement (by modifying the line shape or its position), then exoplanet detectability and characterization are also affected. These include stellar oscillations, granulation, spots, and faculae/plages, and long-term magnetic activity cycles (Saar & Donahue 1997; Dumusque et al. 2011). The most confusing signals are those caused by the presence of active regions on the stellar surface, which can show periodicities and amplitudes similar to the ones generated by real planetary signals. Studies by Figueira et al. (2010) and Robertson, Roy, & Mahadevan (2015) show that these signals may be harder to disentangle.

Although the stellar signal significantly impacts the detection and

characterization, RVs have been widely used to detect exoplanets, especially for low mass planets (Meunier 2021). These low mass and/or long period planets induce RV semi-amplitudes similar to or smaller than those produced by stellar activity (López-Morales et al. 2016; Dai et al. 2017; Haywood et al. 2018). Therefore, it is highly important to put effort in understanding and correcting for stellar activity, as it is essential for accurate RV detection and hence for characterizing interesting new exoplanets. Instrument stability is no longer the dominant obstacle for the precise and accurate mass determination of low-mass planets orbiting bright stars, considering the advanced latest generation of instruments. However, it is often limited by the intrinsic stellar variability (Saar & Donahue 1997; Queloz et al. 2001; Haywood et al. 2014). So, caution is needed while characterizing planets discovered with the RV method around stars that are active/potentially active, as this method is highly prone to the uncertainty arising from the host star’s activity.

Various analysis techniques have been developed to understand and mitigate the impact of stellar activity effects. The early efforts include decorrelation against proxy indicators of activity, or simply rejecting a planet candidate if the same periodicity was found in the FWHM (Full Width Half Maximum), BIS (Bisector Inverse Slope) or in the chromospheric Ca II H&K emission flux, as in the RV (e.g., Hatzes

* E-mail: aaj1@st-andrews.ac.uk

et al. 2010). Aigrain, Pont, & Zucker (2012) used well-sampled light curves to predict the stellar variability expected in RV measurements. Known as the FF' method, this approach uses the light curve and its first derivative to approximate the effects of spots and faculae, and requires no information about the rotation period or spot modelling. It does, however, require simultaneous high-precision photometry. Recently, Gaussian Process regression has been used to model the correlated noise induced by stellar activity in the RV signal, by fitting the orbital signal simultaneously (Haywood et al. 2014). Although GP regression with a quasi-periodic kernel is successful at fitting rotationally modulated activity, other signals are also present.

As noted by Dumusque et al. (2015) and Cretignier et al. (2021), different systematics on the spectra such as micro-tellurics, tellurics, colour variation due to airmass and detector stitching can also affect the RV, on timescales very different from those modelled with a single GP kernel. YARARA (Cretignier et al. 2021) and WOBBLER (Bedell et al. 2019) were developed on this ground as post-processing pipelines to cope with these features in the spectra. YARARA uses Principal Component Analysis on HARPS spectral time series to correct for the known non-planetary systematics without using models for the different effects. WOBBLER is a data-driven method that can be used to infer the stellar spectra, telluric spectra and RVs simultaneously from the data (Bedell et al. 2019).

A complementary approach is to model the modes of variation present in the spectral line profile independently of their temporal behaviour, as explored by YARARA. While a planet produces a shift of all spectral lines, the stellar activity affects the shape of the spectral lines. The Fourier phase SpecTrum Analysis (FIESTA, a.k.a. Φ ESTA) (e.g., Zhao & Tinney 2020; Zhao & Ford 2022) uses the real and imaginary parts of the Fourier transform of the line profile to disentangle apparent RV shift due to spectral line-profile variability from true Doppler shift. Concurrently, de Beurs et al. (2021) also developed machine learning techniques such as linear regression and neural networks to separate activity signals from true center-of-mass RV shifts using the average spectral line-shape variations, without any time-domain information. Another recent investigation (Klein et al. 2022) employs Doppler imaging to simultaneously model the activity-induced distortions and the planet-induced shifts in the line profiles. It is challenging to differentiate between the effects induced by a planet and stellar activity, as the latter has varied effects on individual spectral lines (Dumusque et al. 2015).

Keeping this in mind, a new algorithm (SCALPELS- *Self-Correlation Analysis of Line Profiles for Extracting Low-amplitude Shifts*) has been developed by Collier Cameron et al. (2021) to decouple the RV information from the time domain aiming to separate the Doppler shifts caused by orbital motion from the apparent shifts caused by spectral line-shape variability triggered by stellar activity. By analysing orthogonal modes of variation in the translationally-invariant autocorrelation function of the line profile, this method of shape-shift separation offers precise Doppler detection and characterization of exoplanets around well-observed, bright main-sequence stars across a wide range of orbital periods, especially for lower-mass planets.

Collier Cameron et al. (2021) demonstrated the algorithm using observations of the solar spectrum from HARPS-N. They successfully verified and validated that the algorithm can accurately detect multiple simulated planets injected into heliocentric solar observations, spanning a wide range of orbital periods. Wilson et al. (2022) also used the SCALPELS to clean up HARPS RVs for the stronger detection of two sub-Neptunes in the TOI-1064 system.

Here we present the first in-depth investigative analysis to study the potential of this approach to “clean-up” the RV of a stellar target other

than the Sun and thus to improve the fidelity of RV measurements. We apply SCALPELS to the archival HARPS data on the CoRoT-7 system, which has been studied extensively with time-domain methods, but for which the precise system architecture remains a subject of debate. In Section 2 we summarize the history of investigations of this system. In Section 3.1 we describe the version of SCALPELS employed here and its joint use with the ℓ_1 periodogram method and KIMA in the following subsections. In subsequent subsections of Section 4 we examine the evidence for a third planet in the system and the impact of the number of non-transiting planets on the planetary mass of transiting planet CoRoT-7b extensively, followed by a discussion and conclusion in Section 5 and 6.

2 TARGET AND OBSERVATIONS

CoRoT-7 is a G9V type main-sequence star of V magnitude = 11.7, slightly cooler ($T_{\text{eff}} = 5250 \pm 60$ K) and younger (1.2–2.3 Gyr) than our Sun (Hatzes et al. 2010). Here, we revisit the published HARPS RV measurements of CoRoT-7. Léger et al. (2009) reported the detection of the first ever known transiting super-Earth, CoRoT-7b with an orbital period of 0.85 days and a measured radius of $1.68 \pm 0.09 R_{\oplus}$, which had the smallest exoplanetary radius at that time. Following this discovery, a 4-month intensive HARPS RV follow-up campaign was launched in order to measure the mass of CoRoT-7b. These observations were carried out under ESO Program IDs 082.C-0120, 082.C-0308(A) and 282.C-5036(A) using the HARPS instrument on ESO 3.6 m telescope at La-Silla. CoRoT-7 was observed for 3 nights in 2008 March and then observed continuously from 2008 November to 2009 February, spanning over 4 months. In immediate course, Queloz et al. (2009) revealed CoRoT-7c, the second planet with an orbital period of 3.69 d.

From the large amplitude of photometric modulation, the authors anticipated the variations in RV to be highly perturbed by the activity of the host star. Aiming to remove the RV variations caused by stellar activity, Queloz et al. (2009) applied a harmonic decomposition to the data, preceded by a pre-whitening procedure. In this ‘cleaning’ procedure, the period of the stellar rotation signal is identified by means of a Fourier analysis, and a sine wave fitted with this period is subtracted from the data. The process of the removal of the next strongest signals (which are the sinusoids at integer multiples of the stellar rotation frequency) from the residuals was repeated until the noise limit was achieved. The 0.85d signal associated with CoRoT-7b was found to agree with the CoRoT transit ephemeris, and thus confirmed its planetary nature. In this significant analysis, the author claims, apart from the 0.85d and 3.69d signals, all other resultant signals to be linked to the harmonics of stellar rotation period. In the study which revealed the discovery of CoRoT-7b, Léger et al. (2009) performed a harmonic decomposition of the rotational period and up to the first three harmonics to filter out the activity signal from RV variations caused by the orbiting planet.

2.1 The second non-transiting planet candidate

Subsequently, the presence of a potential third planet, CoRoT-7d, with an orbital period of 9.02d and a mass of $16.7 \pm 0.42 M_{\oplus}$ was disclosed by Hatzes et al. (2010), making CoRoT-7 a compact system. Hatzes et al. (2010) applied the pre-whitening procedure to the BIS, FWHM and CaII H & K emission lines (spectral quantities used as measures of the host star’s activity) owing to their in-dependency to the planetary dynamics. As no significant signal came up at 0.85

Table 1. A timeline of wide range of reported planet masses, especially for CoRoT-7b. It is evident that the mass of CoRoT-7c is nearly consistent. The measured orbital periods for all (proposed and confirmed) planets are also included.

Authors	Planetary masses			Orbital periods		
	CoRoT-7 b	CoRoT-7 c	CoRoT-7 d	CoRoT-7 b	CoRoT-7 c	CoRoT-7 d
Leger et al.(2009) ^a	$< 21M_{\oplus}$	–	–	0.85 days	–	–
Queloz et al.(2009) ^b	$4.8 \pm 0.8M_{\oplus}$	$8.4 \pm 0.9M_{\oplus}$	–	0.85 days	3.69 days	–
Hatzes et al. (2010) ^b	$6.9 \pm 1.4M_{\oplus}$	$12.4 \pm 0.42M_{\oplus}$	$16.7 \pm 0.42M_{\oplus}$	0.85 days	3.69 days	9.02 days
Ferraz-Mello et al. (2011) ^b	$8.0 \pm 1.2M_{\oplus}$	$13.6 \pm 1.4M_{\oplus}$	–	0.85 days	3.69 days	–
Boisse et al. (2011) ^b	$5.7 \pm 2.5M_{\oplus}$	$13.2 \pm 4.1M_{\oplus}$	–	0.85 days	3.69 days	–
Pont, Aigrain, & Zucker (2011) ^b	$2.3 \pm 1.8M_{\oplus}$	–	–	0.85 days	–	–
Hatzes et al. (2011) ^b	$7.4 \pm 1.2M_{\oplus}$	–	–	0.85 days	3.69 days	9.02 days
Haywood et al. (2014) ^c	$4.7 \pm 0.9M_{\oplus}$	$13.5 \pm 1.08M_{\oplus}$	–	0.85 days	3.69 days	8.58 days (deemed as activity)
Tuomi et al. (2014) ^d	$4.8 \pm 2.3M_{\oplus}$	$11.8 \pm 4.1M_{\oplus}$	$15.4 \pm 6.1 M_{\oplus}$	0.85 days	3.71 days	8.89 days (unsure about origin)
Barros et al.(2016) ^e	$5.5 \pm 0.8M_{\oplus}$	–	–	0.85 days	–	–
Faria et al.(2016) ^c	$5.5 \pm 0.8M_{\oplus}$	$12.6 \pm 0.7M_{\oplus}$	–	0.85 days	3.69 days	8.58 days (deemed as activity)

Data used : ^a CoRoT photometric LRA01& SOPHIE RV data, ^b HARPS RV 2008-9, ^c HARPS RV 2012, ^d HARPS-TERRA RV 2008-9, ^e CoRoT photometric LRA06

and 3.69 days in any of these indicators, the planetary candidacies of CoRoT-7b & CoRoT-7c were strengthened.

However, the calculated masses were not in good agreement with the previous studies. This discrepancy in the mass calculation could possibly be rooted in the difference in the methods adopted to cancel out the effect of stellar activity signals from the RV data. Furthermore, no signal was found in the periodogram of activity indicators at 9.02 days, and thus this RV signal was attributed to a third-planetary companion in the system, CoRoT-7d. This signal was previously detected by (Queloz et al. 2009) as well, but had been attributed to a ‘two frequency beating mode’ arising from an amplitude modulation of a 61 days signal. It was therefore deemed to be associated with stellar activity, as this 61 days signal is quite close to twice the stellar rotation period. Subsequently, Pont, Aigrain, & Zucker (2011) argued that the data cannot be used to search for additional (non-transiting) planets in the 3–10 d period range and that claims of the detection of such planets (‘CoRoT-7c’ and ‘CoRoT-7d’) do not withstand scrutiny.

Aiming to settle these arguments, simultaneous photometric and spectroscopic observations were obtained in 2012 from CoRoT and HARPS (ESO Program ID 088.C-0323) for 26 consecutive clear nights in a row from 2012 January 12 to February 6, with multiple well-separated measurements on each night adding up to 77 observations in total (Haywood et al. 2014). These RV data were reprocessed in the way as the 2008–2009 data (Queloz et al. 2009) using the HARPS data analysis pipeline. Altogether, the system was observed for RV studies using HARPS in 2008, 2009 and 2012 with a total of 177 measurements.

The goal of the 2012 observations was to observe the system with CoRoT and HARPS simultaneously, and to use photometry as a proxy for activity-driven RV variations. Along with updating the planetary parameters for CoRoT-7b, Haywood et al. (2014) analysed this system further for the evidence of the additional planetary companion reported by (Hatzes et al. 2010) at the period of 9.02 days. Although, they found a strong peak in the 6-10 days range in the Lomb-Scargle periodogram, the marginal likelihoods calculated from their GP regression analysis favoured the two-planet solution over a three-planet one including the 9-day signal. They concluded that this signal was more likely associated with the second harmonic of the stellar rotation at ~ 7.9 d. However, they successfully confirmed CoRoT-7c along with CoRoT-7b and improved their planetary parameters. Faria et al. (2016) carried out a model-comparison study employing GP regression to model the activity alongside simultaneous orbit fitting in a nested-sampling scheme using the full CoRoT-7 RV data set. While they found evidence for a weak signal at 9 days, the odds’

ratio again favoured the 2-planet solution. Tuomi et al. (2014) presented an analysis of the HARPS- TERRA (Template-Enhanced RV Re-analysis Application) velocities of CoRoT-7 that suggested a system of two planetary companions (CoRoT-7b & CoRoT-7c), possibly three (CoRoT-7d).

2.2 The mass of transiting planet

Table 1 lists the reported planet masses for CoRoT-7b spanning a wide range from $2.3M_{\oplus}$ to $8M_{\oplus}$, which is not ideal. This is plausibly due to the activity level of the star that contributes a significant amount of RV ‘‘jitter’’ and how the various methods correct for this ambiguity. It also depends on how many non-transiting Keplerian signals are included in the model. The method of removal of activity signal will also affect the RV amplitude, which in turn can bias the mass determination of companions. All the aforementioned mass determinations used the same HARPS RV data set, which clearly highlights the trouble in planets’ mass determination around an active star like CoRoT-7 (Hatzes et al. 2011).

Collectively, these studies outline the critical role of stellar activity and non-transiting signals in contaminating the RV observations not only in this particular system but also in general. Therefore, it is crucial to understand this role to enhance our ability to detect low-mass planets and thereby measure their masses precisely. The present study aims to redetermine the planet masses using the entire set of HARPS RV measurements of CoRoT-7 obtained so far. We explore techniques for determining the masses and orbital elements of planets discovered around active stars. We expect this study to offer precise RV measurements corrected for stellar activity and hence to resolve the existing debates on the number of companions in CoRoT-7 planetary system and to update their masses with improved precisions.

3 METHODS

3.1 SCALPELS

The analyses of the CoRoT-7 system architecture described above employed time-domain methods such as pre-whitening and GP regression to model the contribution of stellar activity to the RV signal. Collier Cameron et al. (2021) developed the SCALPELS algorithm as an alternative, wavelength-domain method for separating Doppler

shifts of dynamical origin from apparent velocity variations arising from variability-induced changes in the stellar spectrum.

SCALPELS seeks to decouple the effects of genuine dynamical Doppler shifts from spurious variations caused by line shape changes arising from stellar activity and instrumental systematics. SCALPELS uses the translational invariance property of the autocorrelation function (ACF; Adler&Konheim 1962) of the cross-correlation function (CCF) to isolate the effects of shape changes in the CCF from shifts. In their analysis, the ACF, $A(\delta v)$ has been described as the expectation value of the vector cross-product of the CCF with itself at a sequence of lags δv :

$$A(\delta v) = E(\text{CCF}(v) \cdot \text{CCF}(v + \delta v)) \quad (1)$$

Collier Cameron et al. (2021) carried out a Principal-Component Analysis of the CCF and its ACF, using data from the HARPS-N solar telescope feed. The temporal variability of the CCF contains both line shape changes and Doppler shifts. They found that the majority of time variations of the CCF in general appear similar to those of the shift-invariant profile shape changes probed by the ACF. The principal modes of variability in the CCF or ACF can be isolated by calculating the Singular-Value Decomposition (SVD) of the ensemble of CCFs or ACFs respectively, and can be expressed as:

$$\mathbf{C}(v_i, t_j) = \langle \mathbf{C}(v_i) \rangle + \mathbf{U}_C(t_j) \cdot \text{diag}(\mathbf{S}_C) \cdot \mathbf{P}_C(v_i). \quad (2)$$

$$\mathbf{A}(\delta v_i, t_j) = \langle \mathbf{A}(\delta v_i) \rangle + \mathbf{U}_A(t_j) \cdot \text{diag}(\mathbf{S}_A) \cdot \mathbf{P}_A(\delta v_i). \quad (3)$$

The perturbations resulting from solar activity were isolated by projecting the RVs on to the time-domain subspace spanned by the amplitude coefficients \mathbf{U}_A of the ACF's principal components (Collier Cameron et al. 2021). On the contrary, dynamical shifts due to the planets are preserved when projected onto the orthogonal complement of time-domain subspace.

The variation of the shape-driven component of the RV with time can be obtained from the sum of scaled velocity contributions from all principal components of ACF, $\mathbf{v}_{\parallel} = \mathbf{U}_A \cdot \hat{\alpha}$, where $\hat{\alpha}$ is the vector of response factors obtained by taking the inner product \mathbf{U}_A^T with the mean-subtracted time series of observed velocities. The transformation from observed to shape-driven velocities can be considered as a linear projection of the observed velocities into a subspace spanned by the columns of \mathbf{U} :

$$\mathbf{v}_{\parallel} = \mathbf{U}_A \cdot \mathbf{U}_A^T \cdot (\mathbf{v}_{\text{obs}} - \langle v \rangle_{\text{obs}}). \quad (4)$$

These shape-driven perturbations are strongly correlated with the observed RVs, as seen from the first panel of Figure A1, which could reliably reproduce the stellar activity dominated long-term and short-term fluctuations (Collier Cameron et al. 2021). Even though this offers a linear decorrelation, the time-coefficients of the scalpel basis functions exhibit time lags relative to the observed RV signals, which needs to be addressed.

\mathbf{v}_{\perp} is defined as the residual velocity which lies outside the \mathbf{U}_A subspace, and can also be considered as a projection into the orthogonal complement of the ACF subspace, where the planet signals are sought:

$$\mathbf{v}_{\perp} = (\mathbf{v}_{\text{obs}} - \langle v \rangle_{\text{obs}}) - \mathbf{v}_{\parallel} = (\mathbf{I} - \mathbf{U}_A \cdot \mathbf{U}_A^T) \cdot (\mathbf{v}_{\text{obs}} - \langle v \rangle_{\text{obs}}) \quad (5)$$

To avoid overfitting, a reduced-rank version of \mathbf{U} is employed. Collier Cameron et al. (2021) describe the determination of optimal rank using a leave-one-out cross-validation (LOOCV). This method is found to be efficient in identifying the number of leading columns of \mathbf{U} that contains significant profile information.

In this study, we apply this variant of the SCALPELS method to the CoRoT-7 system together with a GP regression. The host star's high level of stellar activity has been a persistent obstacle to previous efforts using other methods to determine the number of non-transiting planets present, and to determine their masses. The system's intrinsic scientific interest complements its suitability as a test target of the SCALPELS method.

3.2 ℓ_1 -periodogram

The ℓ_1 -periodogram (Hara et al. 2017) is primarily designed to identify the sparsest set of orbital signals that fit the observed RVs. This sparsity is achieved by simultaneously fitting Keplerian signals on an optimally sampled grid of frequencies to the data, regularized by minimizing the ℓ_1 norm of their amplitudes. Hara et al. (2017) present the algorithm as a tool to search for a representation of the input signal as a sum of a few sinusoidal components, that is, a representation which is sparse in the frequency domain.

As the primary step, the data is normalized to the mean of the RV data to get an estimate of stellar noise. The selection of frequency grids ω , covariance matrix \mathbf{V} and the width of averaging interval η are done in succession. This algorithm suggests that RV measurements can be well approximated by the linear combination of a few vectors of $e^{-i\omega t}$ and $e^{i\omega t}$, which spans the columns of the \mathbf{A} matrix.

The noise is assumed to be drawn from a Gaussian with the covariance matrix \mathbf{V} . If the stellar rotation information P_{rot} is given, the covariance matrix uses a quasi-periodic kernel. The contribution of activity to the covariance matrix is then modelled as the product of an exponential correlation term and a quasi-periodic term.

$$V_{ij}^{(\text{act})} = \exp\left(-\frac{(t_i - t_j)^2}{2\tau^2}\right) \left(0.5 + \cos\left(\pi \frac{(t_i - t_j)}{P_{\text{rot}}}\right)\right) \quad (6)$$

The full covariance matrix including the noise terms can be then expressed as:

$$\mathbf{V} = \delta_{\mathbf{k}} + \sigma_R^2 V_{ij}^{(\text{act})} + \mathbf{I}\sigma_W^2 + \sigma_C^2 \quad (7)$$

where $\delta_{\mathbf{k}}$ is a diagonal matrix spanned by the RV uncertainties and σ_W is the white noise amplitude, which we considered as 0.1 ms^{-1} . σ_R stands for the red noise, which is 0.5 ms^{-1} here multiplied with an identity matrix \mathbf{I} . In this simple covariance model, a block-diagonal calibration-noise covariance component σ_C is also included to account for the night-to-night calibration errors, which are known to be present in HARPS data at the $\sim 0.5 \text{ ms}^{-1}$ level.

A remarkable advantage of the method is the use of ℓ_1 norm weighting to find periodicity in unevenly sampled signals. Even without any prior knowledge of the number of planets, the intrinsic signal information can be efficiently decoded. The algorithm assigns an amplitude to every frequency on the grid. It then tensions the goodness of fit to the data (using χ^2 , which is the " ℓ_2 norm") against the ℓ_1 norm (median absolute deviation) of all the amplitudes on the frequency grid. The goal is to identify the minimal set of sinusoids that give an optimal fit to the data. The optimal solution is then found iteratively.

Moreover, we also used its capability to fit a set of externally-determined basis functions simultaneously with the orbit fit. If the columns of the reduced-rank matrix \mathbf{U}_A are used for this purpose, this is equivalent to the simultaneous sinusoidal orbit fitting method described in Section 3.3. This inter-operability with the SCALPELS enables us to perform shape-shift signal separation simultaneously with identifying the minimal set of orbit signals present.

3.3 Simultaneous sinusoidal planet orbit fitting

We employed the shape-signal separation simultaneously with fitting the orbits, given prior knowledge of the orbital periods obtained from the ℓ_1 -periodogram analysis. A simultaneous modelling was chosen to be the best way to figure out the orbital solution, as a simple periodogram has the limitation of fitting only a single sinusoid per frequency sample, so that the interaction between various signals may lead to inaccurate amplitude estimation. To determine the impact of the SCALPELS signal separation, (Collier Cameron et al. 2021) used this approach in the solar RV data, with properly injecting a few planet signals that should be orthogonal to all elements of \mathbf{U}_A . This prevented the signals from being partially absorbed during the SCALPELS projection.

Estimation of parameters and signal separation can be achieved in a single linear calculation, once the candidate signal periods have been determined via periodogram search or through prior knowledge of transits. In the case of CoRoT-7, we have an estimate of the periods of the three candidate signals from the blind periodogram search and ℓ_1 -periodogram, while the period and phase of the CoRoT-7b signal were obtained more precisely from transit observations.

As described by Collier Cameron et al. (2021), the net orbital velocity vector \mathbf{v}_{orb} for a set of n planet signals can be modelled as the product of a set of coefficient pairs $\theta_{\text{orb}} = \{A_1, B_1, \dots, A_n, B_n\}$ with an array of time-domain function pairs $\mathbf{F} = \{\cos \omega_1 t_j, \sin \omega_1 t_j, \dots, \cos \omega_n t_j, \sin \omega_n t_j\}$, where ω_k is the orbital frequency of the k^{th} planet:

$$\mathbf{v}_{\text{orb}} = \mathbf{F} \cdot \theta_{\text{orb}} \quad (8)$$

If we consider the planet orbits to be circular, \mathbf{v}_{orb} from Equation 8 could then provide a complete model of the RV data as the sum of shift (planet)-driven velocity variations and the shape (stellar activity)-driven velocity variations (Collier Cameron et al. 2021). Treating this as a least-square problem enabled us to solve for the unknowns, the amplitudes, and phases of \mathbf{F} . The vector θ_{orb} can then be evaluated by minimizing the χ^2 . Similar Implementation has been employed by (Wilson et al. 2022) as well. The detailed algorithm is given in Appendix B2.

3.4 Trans-dimensional Nested Sampling using *kima*

We used the *kima* package of Faria et al. (2018) to sample from the posterior distribution of the orbital model parameters. *kima* employs a Diffusive Nested sampling algorithm of Brewer, Pártay, & Csányi (2010) for calculating the evidence or fully marginalized likelihood not only for a model with a fixed number of Keplerian signals N_p , but also after marginalizing over N_p . The joint posterior distribution was calculated using the prior $p(\Theta | M)$, likelihood $p(D | \Theta, M)$ and evidence information $p(D | M)$ from the Bayes theorem:

$$\mathbf{p}(\Theta | D, M) = \frac{\mathbf{p}(\Theta | M) \mathbf{p}(D | \Theta, M)}{\mathbf{p}(D | M)} \quad (9)$$

where M is the model with Θ as the vector of all considered parameters and D is the RV data series. Undertaking a similar approach to Haywood et al. (2014) and Faria et al. (2016), we also included a quasi-periodic Gaussian process(GP) to account for the correlated noise with the objective of investigating the improvement in the posterior when using the SCALPELS shift RVs instead of the raw RVs uncorrected for the shape variations.

The *kima* covariance matrix also includes a white-noise term. Unlike the ℓ_1 -periodogram, *kima* does not yet model calibration noise by imposing correlations between observations made on the same

night. Nonetheless, this white-noise contribution serves a similar purpose and enables us to optimise the amplitude of the calibration noise used in the ℓ_1 -periodogram. The quasi-periodic component of the GP is defined by the kernel $k(t, t')$ (a.k.a. covariance function):

$$k(t, t') = \eta_1^2 \exp \left(-\frac{(t-t')^2}{2\eta_2^2} - \frac{2}{\eta_4^2} \sin^2 \left(\frac{\pi(t-t')^2}{\eta_3} \right) \right) \quad (10)$$

The hyperparameters of $k(t, t')$ are represented by the variables η_1, η_2, η_3 and η_4 . η_1 and η_2 define the amplitude of correlation between RVs at different time separations and the evolution timescale of active regions. The correlation timescale η_3 reflects the stellar rotation period, η_4 governs the variation timescale of GP latent functions in relation to η_3 . Smaller values of η_4 indicate more short-scale structures within a single stellar rotation period. We used this GP to model the correlated noise occurring on time-scales of the order of stellar rotation period and its harmonics.

We also computed the False Inclusion Probability (FIP) and True Inclusion Probability (TIP) from the joint posterior distribution of orbital parameters and N_p . Hara et al. (2021) introduces the FIP as the probability that there is no planet with period P in a given frequency interval I . On the other hand, the TIP is defined as the counterpart of the FIP, that is, the probability that there is a minimum of one planet with $P \in I$. The TIP is calculated by averaging the planet detection over the possible number of planets, $\frac{N_p(i)}{\sum N_p}$. Hara et al. (2021) presents this detection criterion as an efficient way to evaluate the reliability of significance levels, by effectively accounting for aliases and favouring to discard the presence of planets under a certain confidence level.

3.5 Stacked Bayesian General Lomb-Scargle Periodograms

To probe the statistical implication of the planet candidate signals, a method of stacking the Bayesian General Lomb-Scargle (BGLS) periodograms has been used. This tool was developed for identifying the periodicities caused by stellar activity, and to show how it can be used to track the signal-to-noise ratio (SNR) of the detection over time (Mortier & Collier Cameron 2017). Adding in more observations, should increase the power or SNR measure of the signal, if the signal is real and coherent. On the contrary, if the signal is incoherent or short-lived, adding more data would cause the SNR or Lomb-Scargle power to decrease (e.g., Howard et al. 2011; Hatzes 2013; Mortier & Collier Cameron 2017). Hence, signals arising from stellar activity can be identified by monitoring their variable and incoherent nature. To track the significance of periodic signals, the BGLS periodogram was calculated for a smaller data subset and recalculated repeatedly with subsequent addition of observations.

Mortier & Collier Cameron (2017) recommend that one should not solely rely on this approach to establish the planetary nature of a signal. This can be used as an additional test to reinforce a finding. We are aiming is to devise a workflow by combining the methods mentioned in Section 3.1, 3.2, 3.3, 3.4 and 3.5.

4 DATA ANALYSIS AND RESULTS

4.1 Time series analysis

In addition to measuring the RV, the HARPS Data Reduction Software returns several shape diagnostics for the CCF and flux indices for known activity-sensitive spectral lines. The RV data and supplementary information including the S-index, BIS, Area and FWHM

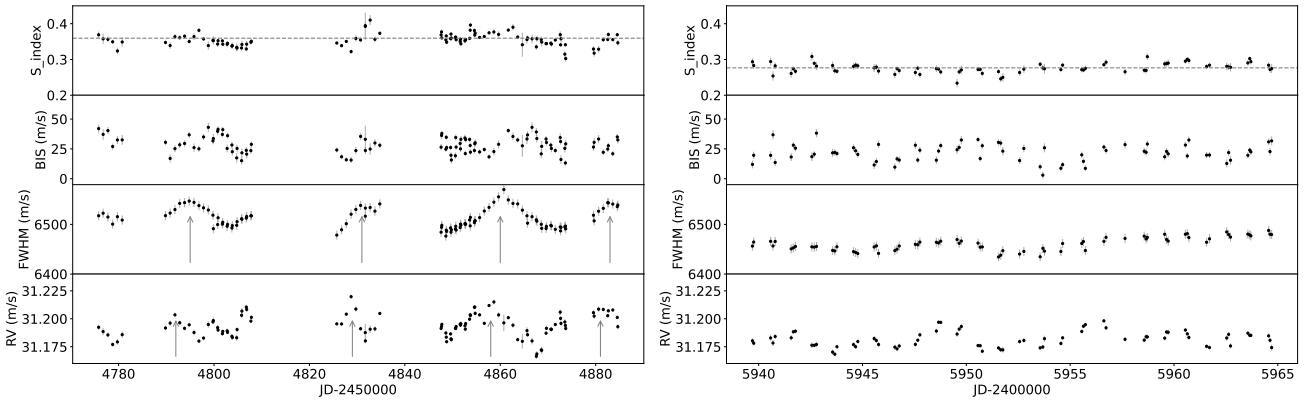


Figure 1. The spectroscopic information measured during the HARPS run spanning over November-2008 and February-2009 is given in the left figure. The RV, FWHM, Area, BIS and S-index data have been plotted against the Barycentric Julian date and shown in each panels from bottom to top. The figure at the right shows the respective representation of spectroscopic data over the span of the 2012 HARPS run. The dashed grey lines running horizontally in the top panels of both figures show the mean values of S-index, indicating the difference in the level of stellar activity during both observing runs. There are arrows given in the RV, FWHM and Area panels of the left figure, pinpointing that the RV peaks occur ~ 2 days before the FWHM and Area peaks. The respective uncertainties are shown using grey error bars.

from 2008-9 and 2012 campaigns are shown in left and right panels of Figure 1. The quantity ‘Area’ is the product of the FWHM of the CCF and its contrast, i.e., the fractional depth of the CCF measured relative to the continuum level. The y-axis values for each parameter in both observing seasons are set to have the same limits for easy comparison.

The supporting measurements of the ‘activity indicators’ such as the FWHM and BIS of CCF and the Ca II H and K values from HARPS allow us to examine the intrinsic stellar variability. [Queloz et al. \(2009\)](#) suggests that one can check two simple characteristics to look for changes in the shape of CCF: the width parameter FWHM and its BIS, computed following [Queloz et al. \(2001\)](#). The chromospheric activity is typically quantified by the S-index, as noted by [Wilson \(1978\)](#). The S-index varies linearly with the chromospheric emission flux in the cores of the Ca II H and K lines in the near ultraviolet ([Salabert et al. 2016](#)). From both panels of the Figure 1, the mean value for S-index can be calculated as 0.3569 in the 2008-9 season and 0.2759 in 2012 and is denoted by the grey horizontal dashed line. The lower mean value in 2012 indicates that the star was less active during 2012.

Similarly, the RV measurements possessed an RMS scatter of 10 ms^{-1} in the 2008-9 run and a lesser scatter value of about 6.8 ms^{-1} in 2012, showing that the amplitudes of RV variability were lower in 2012 than 2008-9. The mean level of RV amplitudes were comparable (31.193 and 31.182) during the two campaigns. However, both sets exhibit obvious multi-periodic variability structures. The FWHM variations exhibit a fairly smooth periodic pattern, with a period of ~ 23 days in both seasons, as seen in [Queloz et al. \(2009\)](#).

A close inspection of Figure 1 shows a temporal offset between FWHM, Area and the RV. The RV maxima occur ~ 2 days prior to those of FWHM and Area. In the Sun, [Collier Cameron et al. \(2021\)](#) observed similar time-lags between the RV, FWHM and BIS. Such temporal shifts between RV and several activity proxies were reported in previous studies as well (e.g., [Santos et al. 2014](#); [Queloz et al. 2009](#)). While [Queloz et al. \(2009\)](#) suggest these time-structured variations between these two parameters to be associated with star-spot related variability in CoRoT-7, [Collier Cameron et al. \(2021\)](#) note that the same phenomenon is present even when the RV signal is dominated by facular suppression of the convective blueshift. [Haywood et al. \(2014\)](#) detail the suppression of convective blueshift

by active regions surrounding star-spots to have a much greater impact on RV than the flux blocked by star-spots, particularly in slowly rotating stars like Sun and CoRoT-7.

The BGLS periodograms for the above-mentioned activity indicators over the entire span of observation are given in Figure D2.

4.2 Blind Periodogram search

We applied SCALPELS to the barycentric RVs of CoRoT-7, and examined the results using a Generalized Lomb-Scargle (GLS) periodogram ([Zechmeister & Kürster 2009](#)) for different subsets of the data (2008-9, 2012 and combined), given the varied level of activity in both seasons as discussed in Section 4.1. The first two curves in Figure 2, shown in blue and orange in all panels, show the observed barycentric RVs from which their own mean is subtracted ($v_{\text{obs}} - \langle v \rangle_{\text{obs}}$), and the shape-driven (v_{\parallel}) RV perturbations derived using the SCALPELS projection respectively. The green curve shows the cleaned shift-driven velocities, obtained when the shape-driven variations are subtracted from the observed RVs. Histograms showing the reduction of RMS scatter after separating the shape-driven signals are presented for all three analyses in Section B1.

Along with the confirmed planet signals with 0.85 day and 3.69 day orbital periods, a third signal with a period of 8.96 days and an amplitude of $\sim 4 \text{ ms}^{-1}$ also appears in all shift-driven periodograms, with no significant counter-part in the shape-driven periodogram as seen from Figure 3. This signal, noted previously by [Hatzes et al. \(2011\)](#) as being attributable to a third planet CoRoT-7d, was suppressed in the GP regression analysis of [Haywood et al. \(2014\)](#), attributing it to be more likely an activity signal. However, we found no significant signal at 8.96 days period in the activity-induced shape-driven periodogram, despite the dataset chosen. This argues against an activity-related origin.

A curious feature of these periodograms is that in the 2012 and combined data, the spurious 1-day periodicity seen in the raw observations is almost entirely accounted for in the shape signal. In the 2012 season, the target was observed 2 or 3 times per night over 26 consecutive clear nights. The target was therefore observed at a range of airmasses and a clear correlation was observed between the airmass and one of the SCALPELS basis vectors (U_2). We consider it possible that chromatic extinction could produce a subtle change

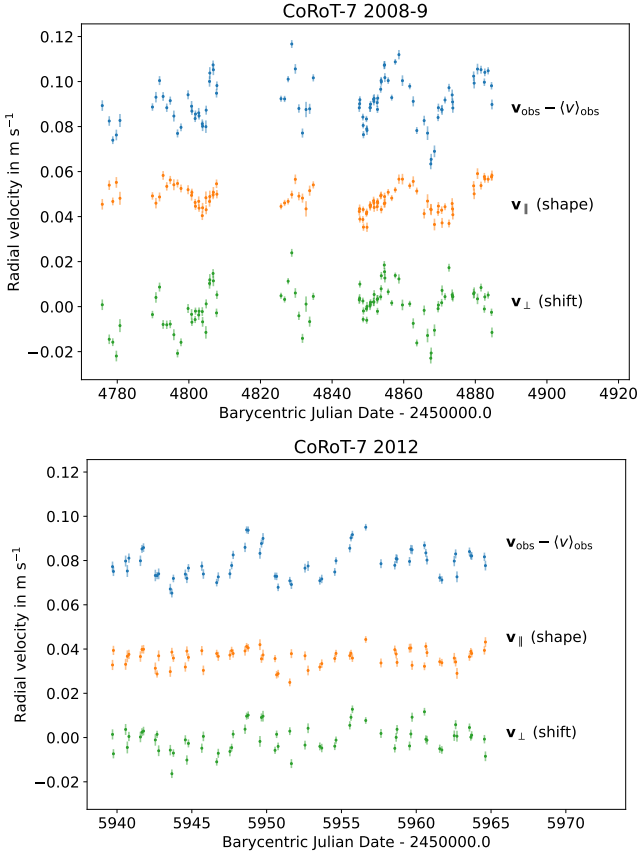


Figure 2. The top panel corresponds to the data from 2008-9 run, while the bottom panel shows 2012 data. In both panels, the blue curve at the top is the barycentric RV subtracted from its own mean. The middle orange curve represents the shape-driven component obtained from the SCALPELS projection, while the green curve manifests the ‘cleaned’ RV, which is the shift-driven component. Please note that an offset of $\pm 60 \text{ m s}^{-1}$ is introduced for clarity. The relatively small uncertainties are shown as semi-transparent error bars to avoid overcrowding.

in CCF shape with airmass, and hence with local sidereal time, despite the care with which the HARPS DRS fits the blaze function and normalizes the continuum prior to cross-correlation. A detailed investigation is, however, beyond the scope of this study.

A strong peak in the combined observed periodogram shows up at around 23 days with an amplitude of 6.42 ms^{-1} , which is the stellar rotation period. Although a shift signal at this period is present, the appearance of a shape-driven signal at the same period casts doubt on a planetary interpretation of this signal. While the presence of a shift signal without a counterpart in the shape signal points strongly to a planetary origin, we cannot completely ignore the possibility of some form of activity in CoRoT-7 producing a shape change in the profile that closely mimics a shift. Therefore, we have to investigate further the nature of the 23 days signal.

The complex frequency structure of the periodogram for the combined 2008-9 and 2012 data (Figure 3) arises from both cycle-count uncertainty and aliasing. In principle, all of this structure should be attributable to a finite number of orbital signals combined with a quasi-periodic activity signal and noise. Given the difficulty in interpreting the system using standard GLS periodograms, specially when the window function is complex and there are multiple, non-harmonic signals present, we recommend performing further analyses to understand similar complex systems to overcome these inherent problems.

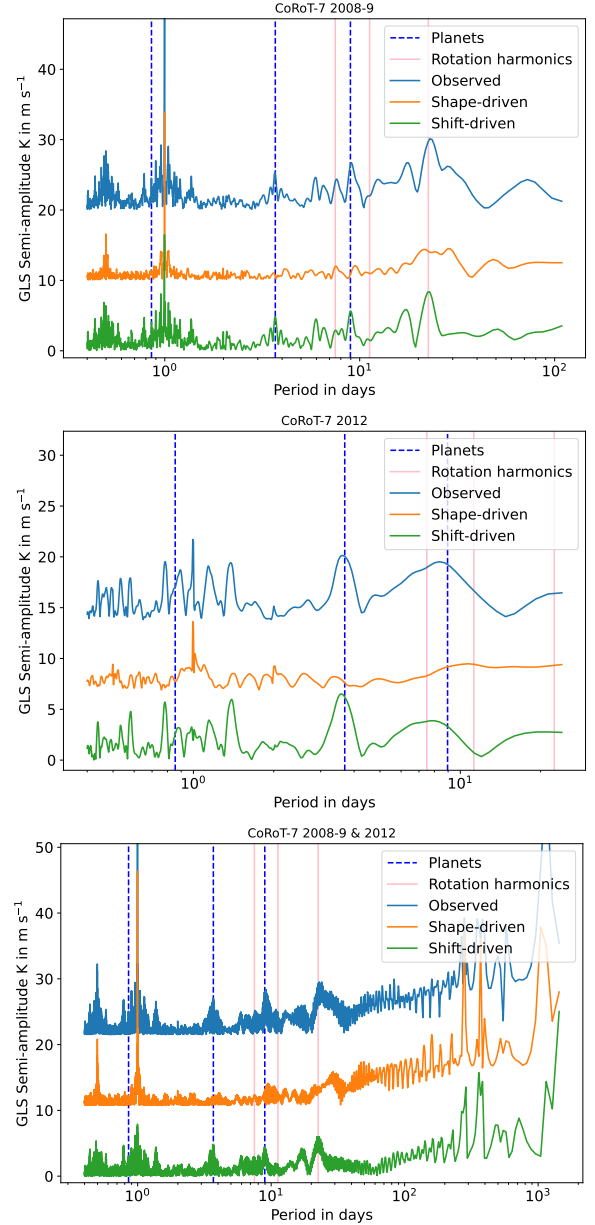


Figure 3. The left, middle and right figures represent 2008-9, 2012 and combined data sets respectively. In each figure, the top periodogram (blue) is for the measured velocities derived from the CCF. The middle trace (orange) is for the shape-driven velocities ($v_{||}$) obtained from the SCALPELS projection. The third periodogram (green) is the shift-driven velocities (v_{\perp}) remaining after the subtraction of shape-driven velocities from the observations. Light blue dotted vertical lines denote the orbital periods of all confirmed and potential planetary companions.

4.3 l_1 -periodogram

Figure 4 shows the l_1 -periodograms, after un-penalizing the 0.85-day signal of the transiting planet. In the l_1 -periodogram terms, this means that we are always including the 0.85 days planet in the fit. The top panel represents the l_1 -periodogram for the raw RVs, dominated by the long-term activity signals.

An l_1 periodogram analysis for the Shift RVs is shown in the bottom panel of Figure 4. After decorrelating the stellar activity-induced shape variations using the SCALPELS U vectors, the two non-transiting

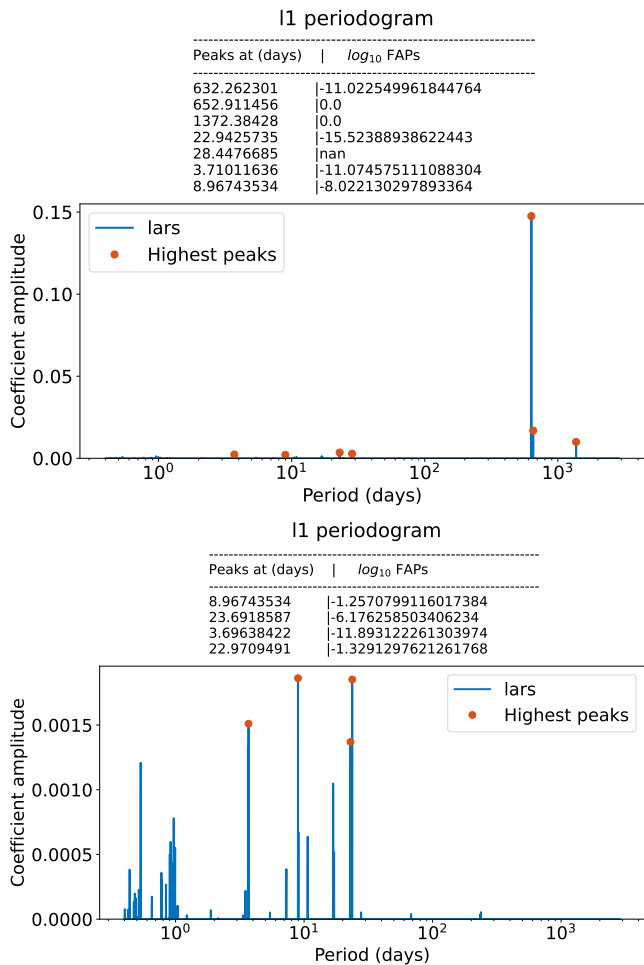


Figure 4. **Top:** The ℓ_1 -periodogram of the raw RVs of CoRoT-7 after unpenalizing the transiting 0.85 days signal of CoRoT-7b, with the strongest spikes marked with red dots at the top. Traces of peaks for the planetary signals of CoRoT-7c (3.7 days) and CoRoT-7d (8.9 days) along with the stellar rotation signal (~ 23 days) are present after the dominant long-term activity signals. **Bottom:** The ℓ_1 -periodogram for the Shift RVs after correcting for the shape variations, showing the strongest peaks at orbital period of planets (8.96 and 3.71 days) and the stellar rotation period with the activity trends suppressed significantly. The corresponding False Alarm Probabilities (FAP) are also listed above each panel.

planet candidates dominate the periodogram. Now the 4 signals with the greater significance and lowest false alarm probabilities are found at periods 3.70 days, 8.96 days and a closely spaced pair (23.69 and 22.94 days) around the stellar rotation period. The long-term signals are weakened considerably. All other strong peaks that showed up in the periodogram are identifiable as one-day aliases ($|\frac{1}{P} \pm 1|^{-1}$) of these four signals.

The presence of two closely-spaced signals near 23d in the both RV sets suggest that, in the ℓ_1 periodogram's sparse representation of the data, the change in the amplitude of this signal between 2008-9 and 2012 is modelled as a 1990 days beat pattern between the two closely-spaced sinusoids.

The presence of low-amplitude peaks in Figure 4 may also indicate the presence of correlated, non-Gaussian noise, which we expect to be the case here. The ℓ_1 -periodogram is clearly successful at shortlisting the most plausible candidate planet signals in our data

and suppressing the parasitic signals. To validate the physical origin of these signals, we computed additional ℓ_1 periodograms for the shape-driven RVs and the activity-indicators, as shown in Figure C1. None of these show traces of peaks at or around 0.85, 3.69 and 8.96 days, adding to the case for their planetary origins, especially for the third signal. Having identified the most plausible signals, we now turn to fitting them simultaneously with SCALPELS.

4.4 Simultaneous sinusoidal planet orbit fitting

In Section 4.2, we projected the observed RVs, including planet signals, into the space spanned by the SCALPELS basis vectors. Although, we found that there is hardly any correlation between the shape and the shift RVs (See A1), if any of the planet signals are correlated with any of the basis vectors, they may get partially absorbed in the shape signal. This is problematic, as we cannot guarantee orthogonality between any irregularly sampled orbital superposition with the components of the SCALPELS \mathbf{U} vectors. The orbital perturbations of any planet and the scalpels projection process must therefore be modelled self-consistently for the signal separation to recover their semi-amplitudes as reliably as possible. For this reason, Collier Cameron et al. (2021) solved simultaneously for shape-driven variations and a set of circular Keplerian orbits with known periods. Here we use a similar approach, using the set of orbital frequencies obtained with the ℓ_1 -periodogram as inputs. The approach of obtaining a simultaneous solution involved determining the shape-driven variations from the difference between observed RVs and the orbit model velocities (Collier Cameron et al. 2021).

The assumption of circular orbits for candidate signals of the three planets can be justified by considering the tidal circularization timescales. According to Dobbs-Dixon, Lin, & Mardling (2004) we could assume negligible eccentricities for the signals at 0.85 and 3.69 days, having periods less than 6 days and hence accounted for by dissipation of tidal disturbances within their envelopes that are induced by their host stars. For the third signal at a period of 8.96 days, we calculated the circularization timescale as described in Dobbs-Dixon, Lin, & Mardling (2004), and found a value of ~ 610 Gyr for an assumed planetary tidal quality factor $Q'_p = 10^6$. We also relaxed the prior on eccentricity for the analysis in Section 4.5 to allow the data to find the feasible value. The obtained value for eccentricity falls around 0.057 ± 0.062 (see Figure F1), again supporting the assumption of circular orbits.

The left panel of Figure 5 shows the RV periodograms when the signal separation and orbit modelling of all 4 strongest sinusoidal signals having non-zero FAPs from the ℓ_1 -periodogram search along with the transiting planet signal and two pairs of periods at $P_{\text{rot}}/2$ and $P_{\text{rot}}/3$ have been performed simultaneously. The orbital periods of candidate planets are marked by vertical blue lines at 0.85 days, 3.71 days, 8.96 days and beat periods for stellar rotation periods and harmonics by magenta lines at 22.94, 23.75, 10.42, 10.74, 7.30 and 7.35 days.

The top three traces are as defined in the Section 4.2, with the only difference being the shape-driven periodogram here is obtained by subtracting the orbit model from the raw RVs. The fourth trace (red) shows the periodogram of the fitted orbit-model of the candidate signals. The bottom periodogram in purple shows the residuals after subtraction of both the shape-driven and orbital RV models. It is worth noting the level of matching between both periodograms (green & red), even with the finest details, in this case.

A closer look at the orange and green periodograms reveal that the balance of the signal separation has been altered from those in Figure 3, by modelling the orbital motion explicitly at known periods. The

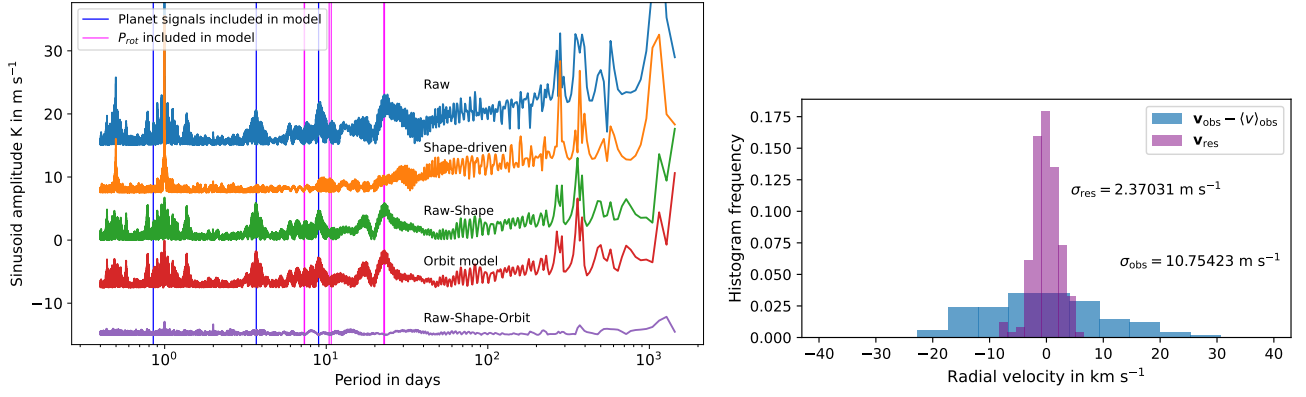


Figure 5. The periodograms of RVs derived from the barycentric CCFs when the signal separation and orbit fitting of the sinusoidal signals of periods derived from the analysis associated with Figure 4 have been performed simultaneously. **Left:** The top three traces are as defined in the caption of Figure 3. The fourth trace (red) shows the periodogram of the fitted model of all the signals, including the 3 candidate planet signals (dark blue vertical bars) and 3 pairs of closely spaced beat periods associated with the rotation period and its harmonics (magenta vertical lines) found by the ℓ_1 -periodogram. The orbital model signal (red) is now much closer in appearance to the shift signal (green), leaving little residuals in the bottom trace (purple), obtained after subtracting both the shape-driven and orbital RV models from the observed RVs. Each trace is offset by $\sim 10 \text{ m s}^{-1}$ for better illustration. **Right:** The resultant histogram of residual RVs shows significantly reduced RMS scatter.

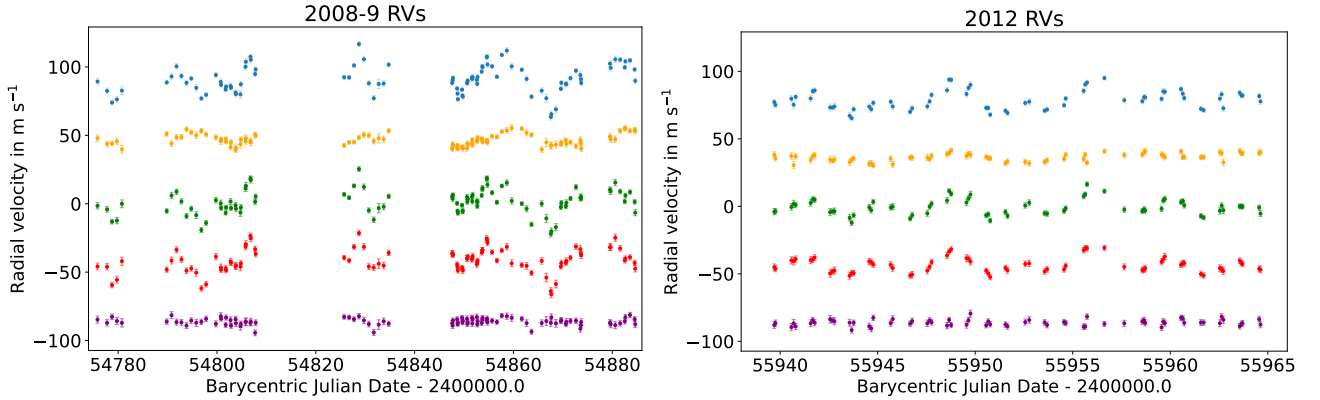


Figure 6. Bottom: The time-domain RV plots obtained from the simultaneous modelling for 2008-9 and 2012 data series are given in the left and right figures, colour-coded corresponding to each periodograms shown in the previous panels. Semi-transparent error bars are also included in the respective curves.

Table 2. The periods and semi-amplitudes of the strongest signals in the periodograms of raw ($k = 0$) and shape-corrected apparent shift-velocities ($k = k_{\text{max}}$) from simultaneous modelling of CCF shape changes and planetary motion, made with prior knowledge of the periods for candidate planets (from the ℓ_1 -periodogram) and pairs of beat period at the stellar rotation period and its harmonics are listed below.

P(days)	$K(\text{ms}^{-1})$ $k = 0$	$\sigma_K(\text{ms}^{-1})$ $k = 0$	$K(\text{ms}^{-1})$ $k = k_{\text{max}}$	$\sigma_K(\text{ms}^{-1})$ $k = k_{\text{max}}$
0.8535	4.185	0.263	3.552	0.306
3.6963	5.484	0.254	6.012	0.312
8.9674	6.877	0.324	5.482	0.380
10.4269	0.472	0.569	1.156	0.606
10.7409	4.215	0.535	2.560	0.597
7.3039	2.173	0.536	2.391	0.620
7.3583	3.004	0.570	3.276	0.655
22.9425	6.915	0.288	4.891	0.353
23.6918	4.077	0.273	3.311	0.348

algorithm efficiently reduces the correlation between the modelled orbit signals and modes of velocity variation driven by line shape changes alone.

To determine the orbital velocity amplitudes that would be recovered without applying profile-shape corrections to the velocities, we reduced the number of principal components k_{max} to zero. We found that, when the sinusoids are fitted to the observed RVs and left uncorrected for line-profile shape variations, the semi-amplitudes of the recovered signals depart from the semi-amplitudes recovered during the simultaneous modelling (see columns 2 and 4 of Table 2).

We also compared the recovered semi-amplitudes of the four strongest signals when fitted with and without the 23.69 days beat signal found in the ℓ_1 -periodogram search, in the appendix (See Table B1). The last two columns of both tables list the semi-amplitudes and associated uncertainties for the shift-driven RVs acquired in both cases. While the differences in the RV amplitudes for planets b, c and d between the 4-signal model and the 5-signal model are small when compared with their uncertainties, the 5-signal model gives a more faithful approximation to the behaviour of the 23-d activity signal between seasons.

The 23-day peak in the shift signal is therefore best modelled as

the beat pattern between the two closely spaced periods, requiring 2 frequencies to be fitted for the stellar rotation period. In the 4-signal case, a single sinusoid is fitted at this frequency, so the 23-day peak shows strong cycle-count interference. We attribute this effect to the 1990 days beat period between the 22.7 days and 22.9 days signals being greater than the entire observation span of ~ 1195 days. When both 23-day periods found by the ℓ_1 - periodogram are fitted, the form of the 23-day peak in the model signal (red) is much closer in appearance to the shift signal (green) (See Figures 5 and B2). However, there is still some power in the residuals at $P_{\text{rot}}/2$ and $P_{\text{rot}}/3$ (marked with pink vertical lines in Figure B2), when only fitting 5 signals. This indicates that a purely sinusoidal model for the residual activity signal, even with two closely-spaced frequencies, might be insufficient.

Therefore, it's better to model the shift-like component of the activity-driven RV signal in such a way as to minimise the residuals and make the residuals as flat as possible. This motivated us to include pairs of sinusoids for the $P_{\text{rot}}/2$ and $P_{\text{rot}}/3$ as shown in Figure 5. The RMS scatter is now significantly reduced to 2.37 ms^{-1} . The resultant orbital periods, semi-amplitudes, and uncertainties of the sinusoidal signals included in the orbit model are listed in Table 2.

Hence, we conclude that the beat model is a better description of the long-term evolution of the 23-day signal, making it unlikely that a fourth planet signal is superposed on the activity signal at this period. The property of the 23-d signal, being a quasi-periodic shift-like change in the CCF shape produced by the activity, complicates the RV analysis.

In brief, all three candidate planet signals reported in the previous sections were recovered successfully with improved precision in their K amplitudes. Particularly for the 8.97d signal, no significant counterpart was detected in the shape RVs obtained from the simultaneous modelling irrespective of the number of planet sinusoids included, strengthening the evidence for the existence of CoRoT-7d, the third planet in the system. Our inability to fit the ~ 23 -day signal satisfactorily with a single sinusoid indicates that it is likely a stellar rotation signal with variable amplitude and phase.

4.5 Nested Sampling using GP +SCALPELS

Using KIMA (Faria et al. 2018), we sampled the posterior distribution of orbital parameters along with a complete Keplerian solver and a Gaussian Process. The posterior distribution constrained by the RV data was then used to probe the credible number of planets given the data, the orbital parameters and the planetary masses. The priors on the GP model parameters hyperpriors were carefully tailored for CoRoT-7, as given in Table.3.

To account for the stellar activity, we decorrelated the SCALPELS U-vectors (when $k = k_{\text{max}}$) that are found to be significant contributors to the shape signal, which are \mathbf{U}_1 , \mathbf{U}_2 & \mathbf{U}_3 . Knowing the parameters of one planet provides a small amount of information about the parameters of another planet (Faria et al. 2016). Keeping this in mind, we treated the transiting planet CoRoT-7b as a known planet, adopting the updated transit parameters in Barros et al. (2014). By putting tight priors on the orbital period P and mid-transit time T_0 we ensure that the phase of this transiting planet signal is consistent with the photometry.

For the known planet, we put a lower limit on K derived from the mass it would have if it were made of compressed water, with a radius of $1.528 R_{\oplus}$ (Barros et al. 2014) that is around $1.56 M_{\oplus}$ (according to the Growth model interpretation of mass distribution proposed by Zeng, Sasselov, & Jacobsen (2016)), giving an RV amplitude of $K_{\text{min}} \sim 1.08 \text{ ms}^{-1}$. The upper limit was set to be about a physically-

Table 3. Prior distribution for model parameters. \mathcal{LU} - Log Uniform ; \mathcal{MLU} - Modified Log Uniform; \mathcal{K} - Kumaraswamy ; \mathcal{U} - Uniform ; \mathcal{G} - Gaussian

Notations	Orbital parameters	Priors
P	Orbital Period	$\mathcal{LU}(0.5, 100)$
K	Semi-amplitude	$\mathcal{MLU}(1, 25)$
e	Eccentricity	$\mathcal{K}(0.867, 3.03)$
ϕ	Orbital phase	$\mathcal{U}(0, 2\pi)$
ω	Longitude of line of sight	$\mathcal{U}(0, 2\pi)$
GP parameters		
η_1	Amplitude of correlations	$\mathcal{LU}(1.6, 2.7)$
η_2	Decay timescale	$\mathcal{LU}(15, 60)$
η_3	Correlation period	$\mathcal{U}(19, 30)$
η_4	Periodic scale	$\mathcal{U}(-1, 0)$
s	Extra white noise	$\mathcal{LU}(0.5, \text{max } v)$
For known planet		
P	Orbital Period	$\mathcal{G}(0.85359, 5.7\text{e-}7)$
K	Semi-amplitude	$\mathcal{MLU}(1.08, 10)$
e	Eccentricity	$\mathcal{U}(0, .1)$
t_0	Time of mid-transit	$\mathcal{G}(54398.0776, 0.0007)$
ω	Longitude of line of sight	$\mathcal{U}(-\pi, \pi)$
β	Activity indicators	$\mathcal{G}(0, 10)$

motivated number of $K_{\text{max}} \sim 10 \text{ m s}^{-1}$ corresponding to the $10 M_{\oplus}$ expected for a pure iron planet of that size.

Taking into account the existing knowledge of the rotation period from the literature (e.g., Queloz et al. 2009; Haywood et al. 2014) and the present work, we restricted the GP hyperparameter η_3 (representing the stellar rotation period) to a sensible range between 19 and 30 days. We adopted a prior period distribution for non-transiting planets with periods between 0.5 days and the duration of the 2008-9 data, i.e., 100 days.

As an initial experiment, the number of planets N_p was set to be a free parameter in the MCMC, with a uniform prior between 0 and 5. Considering N_p as an unknown parameter, KIMA uses a birth-death MCMC proposed by Brewer (2014), which allows estimating the improvements in evidence values when the sampler switches between solutions with different N_p values.

The resultant joint posterior distributions for the semi-amplitudes of the signals are shown in the top panel of Figure 7. We found that two non-transiting planets are favoured by the data, considering the ephemeris of the transiting planet CoRoT-7b as known in the model. Amidst the posterior spread in the parameter space, these are seen as heavily populated regions at $P_c = 3.69$ days and $P_d = 8.96$ days.

Reviewing this posterior distribution triggered another experiment by fixing the number of planets as $N_p=2$ with the rest of the priors as the same as in Table.3. The resultant posterior distribution of RV semi-amplitudes is represented in the bottom panel of Figure 7, with well-constrained peaks around $P_c = 3.69$ days and $P_d = 8.96$ days, within the orbital period uncertainties reported in the literature. Supporting the presence of CoRoT-7c and CoRoT-7d, the same constrained detail can be seen in the posterior distribution of eccentricities also (See Figure F1). From the 10000 effective samples in the joint posterior distribution, the evidence $\log \mathcal{Z}_2$ for this model was calculated as $\log(p(D | M)) = -475.46$.

If we fix $N_p=1$, we recover the orbital parameters for CoRoT-7c only, similar to those reported by Faria et al. (2016). The evidence is degraded to $\log \mathcal{Z}_1 = -479.20$ for this model. The difference in log evidence between the model with 2 non-transiting planets and the model with 1 non-transiting planet is 3.8. This is in the substantial support range of the Jeffreys scale for Bayesian evidence comparison

Table 4. The mass and mean-density (column 4 & 5) calculated for the transiting planet CoRoT-7b from the posterior distributions of orbital period and RV semi-amplitudes (column 2 & 3) considering different models are listed. Important note: CoRoT-7b is considered as a known planet in all these models with transit parameters from Barros et al. (2014) (See Table 3), who updated the planet radius as $1.528 \pm 0.065 R_{\oplus}$. N_p is the fixed number of planets in each model. The first three rows give results from models with GP + SCALPELS and different number of planets, where the last three rows are the results for different models with GP alone and different N_p .

Model	P_b [days]	K_b [ms^{-1}]	M_b [M_{\oplus}]	ρ_b [$g\ cm^{-3}$]	$\log Z$	$Z_n - Z_{n+1}$
GP + SCALPELS						
$N_p=1$	$0.853592 \pm 6.01e-7$	3.943 ± 0.559	5.595 ± 0.794	8.641 ± 0.269	-479.20	–
$N_p=2$	$0.853592 \pm 5.87e-7$	4.291 ± 0.460	6.056 ± 0.653	9.355 ± 0.235	-475.46	42.31
$N_p=3$	$0.853592 \pm 5.78e-7$	4.278 ± 0.453	6.048 ± 0.643	9.322 ± 0.234	-474.58	2.40
GP only						
$N_p = 1$	$0.853592 \pm 5.48e-7$	4.007 ± 0.550	5.679 ± 0.781	8.786 ± 0.264	-478.58	–
$N_p = 2$	$0.853592 \pm 5.53e-7$	4.293 ± 0.468	6.087 ± 0.664	9.394 ± 0.236	-474.87	40.89
$N_p = 3$	$0.853592 \pm 5.72e-7$	4.331 ± 0.476	6.146 ± 0.676	9.459 ± 0.236	-474.27	1.82

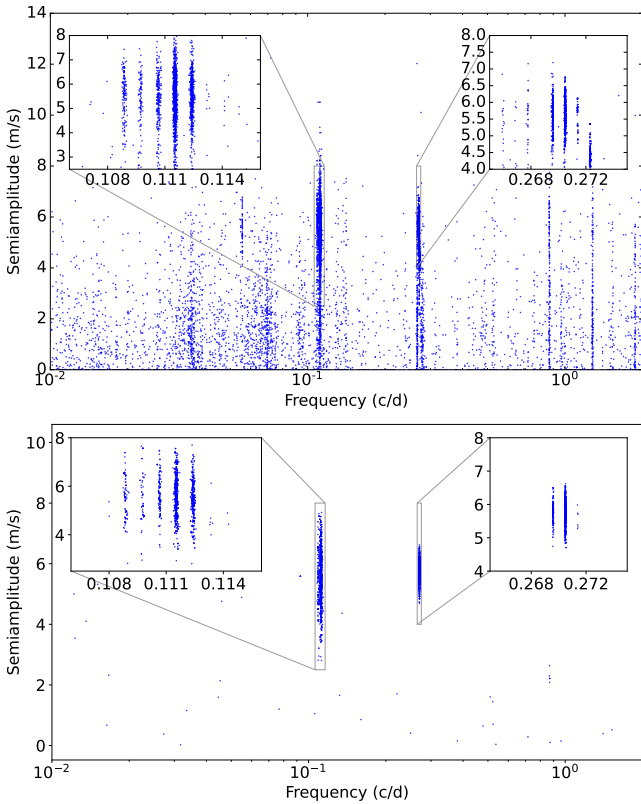


Figure 7. The posterior distribution for semi amplitudes (ms^{-1}) in log frequency space (cd^{-1}) is shown. The transiting 0.85 days planet is considered as ‘known’. The inset plots zoom into the clusterings corresponding to the ~ 9 days planet and ~ 3.7 days planet (right), showing the interference patterns occurred due to the widely separated observing seasons. **Top** : Setting number of planets, N_p as a free parameter with uniform priors between 0 and 5. **Bottom** : N_p is fixed at 2.

(Jeffreys 1961). These results therefore indicate a 2-planet model to be 42.31 times ($Z_1 - Z_2$) more probable than a 1-planet model. We emphasize that, when saying a 2-planet model, we are referring to CoRoT-7c and CoRoT-7d, as the model considers the orbital parameters of transiting planet CoRoT-7b as known.

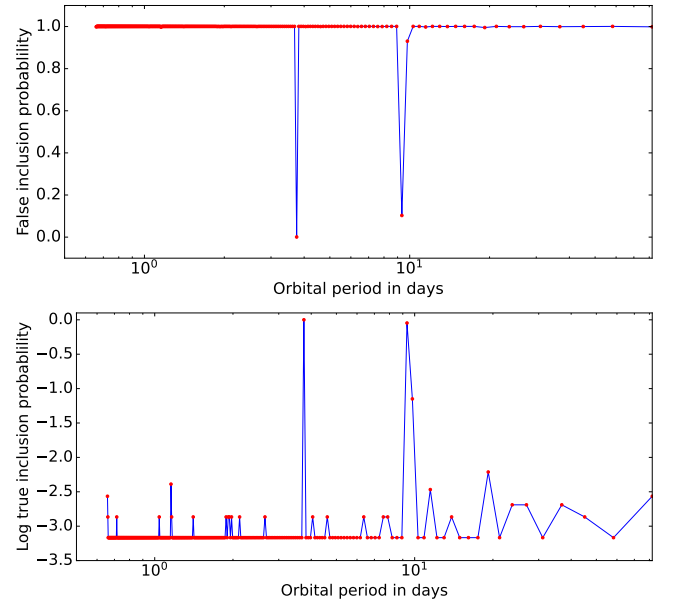


Figure 8. Top: The FIP periodogram of CoRoT-7 HARPS data, computed using the joint posterior distributions obtained from KIMA, modelled with SCALPELS, GP and P_b as known. **Bottom:** TIP periodogram. Both periodograms strongly favour 2 planets at orbital periods 3.69 days and 8.96 days, additional to the known transiting 0.85 day signal.

After ‘confidently detecting’ CoRoT-7b and CoRoT-7c, Faria et al. (2016) pointed out that it is more likely that there are four planets according to the posterior distribution. Their posterior showed smaller peaks around ~ 9 days and ~ 2 days. However, adding the SCALPELS basis vectors to the model for activity decorrelation enabled us to confidently recover the orbital signal of CoRoT-7d at 8.96 days. We were unable to recover any signal around 2 days, as it is removed as being some artefact arising from activity, aliasing or stellar rotation harmonics. In an attempt to validate the improvement in the model with the inclusion of the SCALPELS shape-signal decorrelation vectors along with a GP, we compared the run with the SCALPELS against the one without. The results are given in Table 4. We found that de-trending the RVs for line shape variations using the SCALPELS

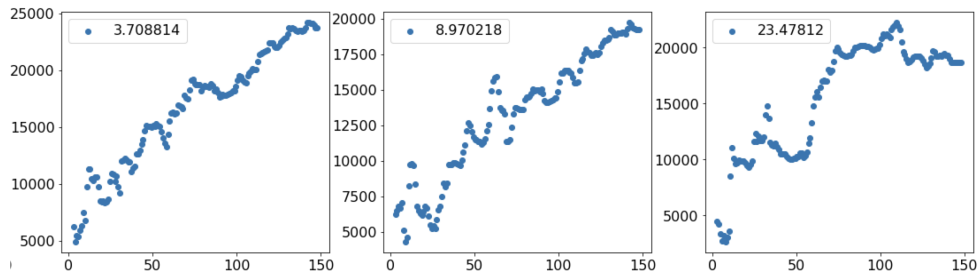


Figure 9. x – axis: Number of observations, y – axis: signal-to-noise ratio. The SNR plotted against the number of observations for the optimal periods of individual candidate signals at periods 3.70 days, 8.97 days and 23.47 days.

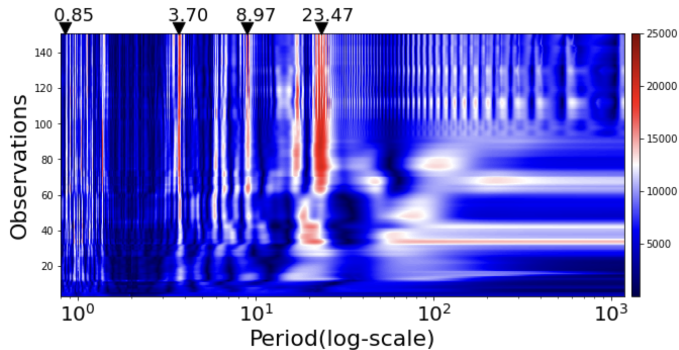


Figure 10. The Stacked Bayesian General Lomb-Scargle periodogram for CoRoT-7 with number of observations in y-axis plotted against period in days displayed in logarithmic scale. The colour gradient designates the logarithm of probability, where redder is more promising.

basis vectors yields a model that is 4.14 times more probable than a model that does not include these activity variations.

4.5.1 FIP and TIP calculation

The zoomed-in view of the over density regions in Figure 7 revealed interference patterns associated with CoRoT-7c and CoRoT-7d at 3.69 days and 8.96 days. It is interesting to consider the two seasons of observations in 2008-9 and 2012 as a ‘double-slit experiment’, in which the path lengths between the light waves from two slits give rise to an interference pattern due to the phase shift. We see an analogous scenario here, with sharp peaks separated by a frequency $\sim \frac{1}{1200} \text{ cd}^{-1}$, accounting for the time separation (1200 days) between two observing campaigns. This clearly indicates that both signals are detected unambiguously in both seasons.

We performed the FIP analysis in the frequency space, setting the bin size equal to the Nyquist frequency resolution of the 2008-9 season rather than the full duration of the data. Since the probability of each planet is spread over several interference fringes, we should marginalize over the uncertainty in the cycle count to establish the existence or otherwise of the planet. A frequency bin size of 0.005 cycles per day captures all probabilities while fully resolving the structure of the 2008-9 periodogram.

Hara et al. (2021) presented FIP and TIP testing as good practice for the analysis of RV data, given that assumptions on the priors and the noise model ($\log \mathcal{L}$) might be incorrect. Following this idea, we searched for various planets simultaneously using a frequency window sliding along the entire frequency space spanned by the posterior distribution of the 2-planet model (with CoRoT-7b being

known). Consequently, two strong peaks were found, at 3.697 days and 8.965 days, both with significantly minimal FIP and maximal TIP, reinforcing their planetary nature (Figure 8). To avoid being misled by the shared power among adjacent interference peaks linked with relevant signals, a basic histogram was used to find these most probable periods, See Figure F2.

Being a trans-dimensional sampling algorithm, KIMA enabled us to directly sample the joint posterior of the number of planets (N_p) and other orbital elements. Several ambiguous detections are prompted at spurious periods due to the cross-talk between different aliases. FIP and/or TIP attenuated the invalid detections associated with these periods. Interesting is the enfeebled TIP and very large FIP (~ 1) at the stellar rotation period of ~ 23 days and its harmonics, providing more evidence against the planetary nature of a body at this period. From Figure 5 alone, one could argue that since the stellar rotation period (~ 23 days) is seen in the shift signal, any other period that also survives when the shape signals are subtracted could also be activity-driven. However, a better validation for the dynamical origin of the candidate signals is achieved here through incorporating a GP to account for any remaining quasi-periodic signals. It is promising to see that both the non-transiting planet signals with orbital periods 3.69 and 8.96 days survives the GP and produces a 2-slit interference pattern as shown in Figure 7.

4.6 Stacked Bayesian General Lomb-Scargle periodogram

To investigate the stability of the periodic signals found in the previous analyses, we used a stacked normalized BGLS periodogram (Mortier & Collier Cameron 2017) of the shift-driven RV data. Following the method of Mortier & Collier Cameron (2017), we derived the stacked periodograms of CoRoT-7 from the cleaned shift-driven RVs obtained after the SCALPELS shape-signal separation. The BGLS periodogram was initially obtained from the first 4 data points and then refined progressively by adding more one by one. This analysis was done for different subsets of data, obtained from the two observing runs (2008-9 & 2012) separately (See Section E1) and also for the entire data set as shown in Figure 10. Figure 10 is colour coded based on the log probability of the periods in the BGLS periodogram. One should expect to see the SNR of a planetary signal increasing asymptotically in proportion to the square root of the number of data points added.

Fringes began to appear in the stacked BGLS periodogram when analysed the both data sets together, as seen from Figure 10. These are caused by the uncertainty in the cycle count elapsed across the three-year gap separating the 2008-9 and 2012 RV campaigns (Mortier & Collier Cameron 2017). However, the combined data set works well for tracking the candidate signals.

Except the 23 days signal, all other candidate signals stand out

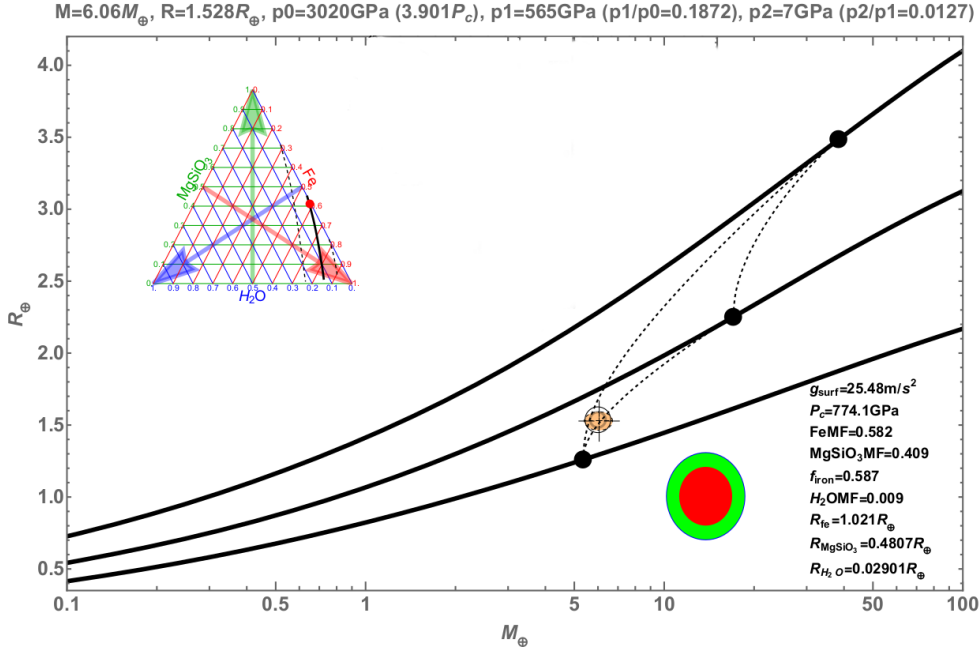


Figure 11. The most plausible composition of the transiting planet CoRoT-7b backed by the derived mass and radius for minimum core pressure is shown as obtained from the *manipulateplanet* tool developed by Zeng, Sasselov, & Jacobsen (2016). The thick black curve in the ternary diagrams shows the degeneracy of tri-layer model due to the trade-off between the three components, but not due to the mass or radius measurement uncertainties.

with clearly growing probability features at periods 0.85, 3.69 and 8.97 days, supporting the coherent nature of these signals. It is worth noting that the 23.47 days signal grows unsteadily, with the strength maximizing around 70 to 100 observations and decreasing later on, see also Figure 9. We looked particularly into the 8.97 days signal to check its behaviour after adding more data points from the 2012 campaign and found that the SNR constantly increases with more observations, as expected for a real and coherent signal.

4.6.1 Tracking the significance of detection

We tracked the significance of each strong signal in the Figure 10 by plotting the SNR against the number of observations (Figure 9). The optimal frequency for each candidate signal was obtained from the dominant peaks in the top line of the stacked BGLS periodogram.

Each point in Figure 9 represents the maximum SNR for the set of all RVs up to that point in time. We particularly looked into the 3.69 and 8.97 days signals and found that they exhibit a monotonic rise towards asymptotic square-root behaviour, confirming that they are strictly periodic and were present throughout the observations, hence substantiating the existence of CoRoT-7c and CoRoT-7d. As a result of the steady growth, their detections reach $\sim 25\sigma$ and $\sim 20\sigma$ respectively.

On the other hand, the detection significance of the 23.47 days signal first grows rapidly but then levels out and even drops at some point when more observations are added. This is a clear sign of an incoherent signal. We have an advantage of having longer-baseline which make this analysis relevant, which would not be the case for shorter baselines and longer orbital periods. This evidence again strengthens the activity origin of the 23-day signal, due to its unstable nature throughout the period of observation.

5 DISCUSSION

5.1 Mass & Mean density of CoRoT-7b

As discussed in Section 2.1, the mass of the transiting planet CoRoT-7b has been a subject of debate for more than a decade. Table 1 shows the wide range of masses as reported in the literature. This can be attributed to the relatively high level of activity in the host star. Different posterior distributions for semi-amplitudes, considering different models, are listed in Table 3. We consider the model with $GP + SCALPELS + N_{p=2} + \text{CoRoT-7b}_{\text{known}}$ as the optimal one. Several results motivated us to reach this conclusion. The major motivation was the significant improvement of evidence (\mathcal{Z}) when fixing the number of non-transiting planets at 2. This model yielded a RV semi-amplitude of $4.291 \pm 0.46 \text{ ms}^{-1}$ for CoRoT-7b.

The planetary mass was calculated using this semi-amplitude, assuming a circular orbit. The orbital inclination i and stellar mass M_* were obtained from Barros et al. (2014) as 80.98 ± 0.51 degree and $0.915 \pm 0.017 M_{\odot}$ respectively. We determined the mass of CoRoT-7b to be $6.056 \pm 0.653 M_{\oplus}$, making it a super-Earth. This mass measure is within the uncertainty limit of several previous studies (e.g., Faria et al. 2016; Hatzes et al. 2010). Incorporating with the planet radius of $1.528 \pm 0.065 R_{\oplus}$ from (Barros et al. 2014), the planet bulk density was calculated as $9.355 \pm 0.235 \text{ g cm}^{-3}$.

We used the MANIPULATEPLANET tool developed by Zeng, Sasselov, & Jacobsen (2016) to get an estimate on the planetary composition. The derived value of mass and a minimum core pressure (3020 GPa) supports a planet with an iron core (Fe : 58.7% by mass), Silicate mantle (MgSiO_3 : 40.9%) and a thin layer of water on the surface (H_2O : 0.9%), favouring the rocky composition discussed in the literature.

For the given value of the mass and radius of the planet, we varied the core pressure in the model to investigate the degeneracy in the composition of the planet up to a maximum supported core pressure of 3346 GPa. This maximum core pressure proposes a dense iron

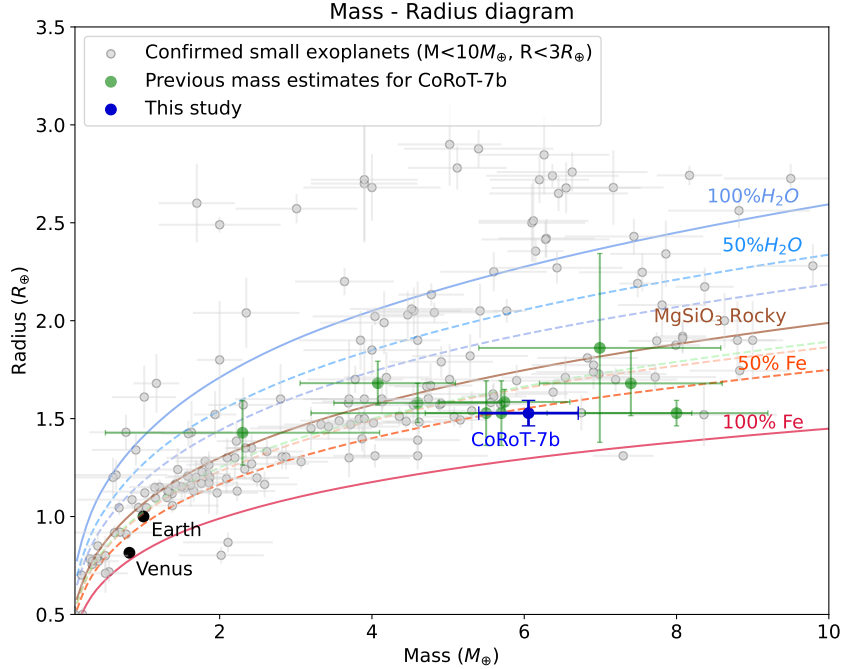


Figure 12. Mass–radius relation for planets with masses $<10 M_{\oplus}$ represented by grey circles. The plot also includes Earth and Venus, for reference. The lines show models of different compositions, with solid lines indicating single composition planets (either H₂O, MgSiO₃, i.e., rock, or Fe). The dashed and dotted lines indicate Mg-silicate planets with different amounts of H₂O and Fe. The green circles show the widely different mass estimates for the transiting planet CoRoT-7b from the literature, where the refined mass estimate from this study is marked in blue, with significant improvement in uncertainty. We want to point out the reason for a few points falling in a horizontal line with our estimate being the similar value of planetary radius considered in the corresponding analyses.

core (Fe : 96.3 %) immersed in deep ocean mantle (H₂O : 12.6 %) with a compromised silicate layer (MgSiO₃ : 3.2 %). Also, the lower value of planet mass ($M-1\sigma$) suggests a composition that can be up to 78.6 % Fe, 6.2% MgSiO₃ and 15.1% H₂O. However, none of the above compositions seem to be feasible, considering the high equilibrium temperature (1756 K) of CoRoT-7b estimated assuming a zero albedo (Barros et al. 2014). Therefore, the most plausible composition occurs to be the one discussed earlier in this section, shown in Figure 11 with little or no surface layer of water. It should be borne in mind that the MANIPULATEPLANET tool assumes a condensed volatile layer typical of a cold planet. More recent studies, (e.g. Zeng et al. 2021) show that a substantial water layer would establish a deep steam atmosphere in a planet as strongly-irradiated as CoRoT-7b, increasing the radius well beyond the observed value.

We also explored the masses of known planet CoRoT-7b for different models, as tabulated in Table 4. As seen from the table, the mass estimation improves and settles for the transiting + 2-planet model incorporated with GP regression and the SCALPELS decorrelation. This approach takes into account the contributions of all additional planets whose existence may not be established conclusively, but whose RV variations influence the mass determination of the known planet(s).

5.2 Masses of non-transiting planets

As listed in Table 5, we also computed the masses of non-transiting planets from the RV semi-amplitudes obtained from the analysis described in Section 4.5. The inner non-transiting planet CoRoT-7c, with an orbital period of 3.69 days, was found to have a mass

Table 5. Planetary parameters for CoRoT-7b, CoRoT-7c and CoRoT-7d obtained from the best model.

Planet	P[days]	K[ms ⁻¹]	M [M_{\oplus}]
CoRoT-7b	$0.8535 \pm 5.87e-7$	4.291 ± 0.46	6.056 ± 0.653
CoRoT-7c	3.697 ± 0.005	5.757 ± 0.298	13.289 ± 0.689
CoRoT-7d	8.966 ± 1.546	5.525 ± 0.792	17.142 ± 2.552

of $13.289 \pm 0.689 M_{\oplus}$. This planet could be structurally similar to Uranus ($\sim 14.5 M_{\oplus}$) in our Solar System. The RV semi-amplitude corresponding to the second non-transiting planet CoRoT-7d at 8.97 days suggests a planetary mass of $17.14 \pm 2.55 M_{\oplus}$. The mass of this long-period planet is comparable to that of Neptune ($\sim 17.1 M_{\oplus}$). However, the bulk density and composition of these non-transiting planets are unknown, due to the lack of radius information from the photometry.

5.3 Areas for further improvement

The quasi-periodic signal with $P \sim 23$ days showed up in the shift-driven periodogram is very close to the stellar rotation period that stands out as a clear peak in CoRoT photometry, activity-indicator periodograms and the SCALPELS basis-vector periodograms (Figure D2 and D1, marked with the rightmost dotted lines). This points out that the ~ 23 days shift signal is more likely to be a stellar rotation artefact than a planetary reflex-motion signal. However, the presence of this signal in both shift periodograms (Figure 3 and Figure 5) and BGLS periodograms of activity indicators strengthens the conclusion

reached in Section 4.2: the property of this signal that produces a shape change in the profile that also closely mimics a shift. A note of caution is due here, since this confusing signal survived the SCALPELS signal separation by imitating the behaviour of a real planet signal. We anticipate this to be the effect of time-lags between the temporal coefficients of \mathbf{U} vectors and the RVs, which makes a linear decorrelation difficult and incomplete.

6 CONCLUSION

In this paper, we present the results of applying the SCALPELS to extract precise RV estimates from the high-resolution spectroscopic planet surveys without using time-domain information. SCALPELS construct an orthogonal basis containing the coefficients of the first few principal components of shape-induced CCF variations, then project the raw RVs onto this basis to obtain a time series of shape-induced RV variations. These are then subtracted from the original RVs to leave shift-only RV variations, where the planet signals are sought.

By separating the activity-induced line-shape changes from the HARPS archival RV information of an active-star CoRoT-7, SCALPELS isolated the planet-induced shift-driven signals. The consequent shift-driven RVs suggest that CoRoT-7 is best modelled as a 3-planet system, with quasi-periodic harmonics of the stellar rotation period also present. An ultra-short period (0.85 days) planet CoRoT-7b; which is transiting, a second planet CoRoT-7c with orbital period of 3.69 days together with a third companion CoRoT-7d at a long-period orbit of 8.97 days, make CoRoT-7 an ultra-compact planetary system. Both the long-period planets (CoRoT-7c & CoRoT-7d) are non-transiting.

We used the ℓ_1 periodogram, a compressed-sensing method developed by Hara et al. (2017), to identify the sparsest set of orbital periods that could simultaneously fit the SCALPELS shift signal. The resulting sparse solution eliminated unwanted alias signals and revealed three dominant candidate signals. This analysis favoured the results from the SCALPELS projection, that CoRoT-7 is better modelled as a 3-planet system than a 2-planet system. Because the ~ 23 -d signal varies slowly in phase and amplitude, we noted the usefulness of a split-period representation for reproducing the detailed frequency structure of the periodogram and for determining the orbital velocity amplitudes of the three dominant planet signals.

A simultaneous modelling of stellar variability and planetary motion was performed to study the impact of the SCALPELS signal separation on the CoRoT-7 system, considering planet signals of known periods. A good match was observed for a 5-signal model comprising three planet signals + two closely-spaced periods near 23 d representing the quasi-periodic stellar rotation signal, as suggested by the ℓ_1 periodogram. This showed the effectiveness of the SCALPELS in reducing the contribution of stellar variability to the RV signal and enabling the detection of exoplanets' signals in data from active stars.

The posterior probability distributions for the orbital parameters were determined along with a complete Keplerian solver and a Gaussian Process using KIMA (Faria et al. 2018). This GP model, incorporating the SCALPELS basis vectors for line-shape decorrelation, is found to be 4 times more probable than a model without SCALPELS. The improvement in the evidence again suggested a system with two non-transiting planets and the known transiting planet.

In addition to improving the reliability of planet detection, our method also allows more rigorous determination of the masses of known, transiting planets whose bulk densities reveal their interior composition. This approach takes into account the contributions of

all additional (non-transiting) planets whose existence may not be established conclusively, but whose RV variations influence the mass determination of the known planet(s). Combined with modelling of any surviving shift-like signal from stellar activity, this approach offers a significant improvement in the precise characterization of exoplanet systems, especially for the low-mass planets whose RV amplitudes are close to the detection threshold.

ACKNOWLEDGEMENTS

AAJ acknowledges the support from World leading St Andrews Doctoral Scholarship. ACC and TGW acknowledge support from STFC consolidated grant numbers ST/R000824/1 and ST/V000861/1, and UKSA grant number ST/R003203/1. AAJ, ACC and TGW thank Nathan C. Hara for the productive discussions on l_1 -periodogram. We are also grateful to Joao P. Faria whose algorithm (KIMA) played a significant role in validating our results.

DATA AVAILABILITY

This paper uses of data from the Data & Analysis Center for Exoplanets (DACE) database dedicated to the visualization, exchange and analysis of extrasolar planets' data. The HARPS data products used for the analyses are publicly available in the DACE platform (<https://dace.unige.ch>). The PYTHON codes and notebooks used to generate the results and diagrams in this paper will be made available through the University of St Andrews Research Portal.

REFERENCES

- Adler, Roy & Konheim, Alan (1962). PROC AMER MATH SOC. 13. 425-425.
- Aigrain S., Pont F., Zucker S., 2012, MNRAS, 419, 3147. doi:10.1111/j.1365-2966.2011.19960.x
- Barros S. C. C., Almenara J. M., Deleuil M., Diaz R. F., Csizmadia S., Cabrera J., Chaintreuil S., et al., 2014, A&A, 569, A74.
- Bedell M., Hogg D. W., Foreman-Mackey D., Montet B. T., Luger R., 2019, AJ, 158, 164.
- Brewer B. J., 2014, arXiv, arXiv:1411.3921
- Brewer B. J., Pártay L. B., Csányi G., 2010, ascl.soft.1010.029
- Chib & Ivan Jeliazkov (2001) Marginal Likelihood From the Metropolis-Hastings Output, Journal of the American Statistical Association, 96:453, 270-281
- Collier Cameron A., Ford E. B., Shahaf S., Aigrain S., Dumusque X., Haywood R. D., Mortier A., et al., 2021, MNRAS, 505, 1699.
- Cretignier M., Dumusque X., Hara N. C., Pepe F., 2021, A&A, 653, A43.
- Dai F., Winn J. N., Gandolfi D., Wang S. X., Teske J. K., Burt J., Albrecht S., et al., 2017, AJ, 154, 226.
- Davis A. B., Cisewski J., Dumusque X., Fischer D. A., Ford E. B., 2017, ApJ, 846, 59.
- de Beurs Z. L., Vanderburg A., Shallue C. J., Harps-N Collaboration, 2021, AAS
- Dobbs-Dixon I., Lin D. N. C., Mardling R. A., 2004, ApJ, 610, 464. doi:10.1086/421510
- Dorn C., Khan A., Heng K., Connolly J. A. D., Alibert Y., Benz W., Tackley P., 2015, A&A, 577, A83.
- Dumusque X., Santos N. C., Udry S., Lovis C., Bonfils X., 2011, IAUS, 276, 527.
- Dumusque X., Boisse I., Santos N. C., 2014, ApJ, 796, 132.
- Dumusque X., Glenday A., Phillips D. F., Buchschacher N., Collier Cameron A., Ceccconi M., Charbonneau D., et al., 2015, ApJL, 814, L21.
- Dumusque X., 2018, A&A, 620, A47.

Faria J. P., Haywood R. D., Brewer B. J., Figueira P., Oshagh M., Santerne A., Santos N. C., 2016, *A&A*, 588, A31.

Faria J. P., Santos N. C., Figueira P., Brewer B. J., 2018, *JOSS*, 3, 487.

Ferraz-Mello S., Tadeu Dos Santos M., Beaugé C., Michtchenko T. A., Rodríguez A., 2011, *A&A*, 531, A161.

Figueira P., Marmier M., Bonfils X., di Folco E., Udry S., Santos N. C., Lovis C., et al., 2010, *A&A*, 513, L8.

Fischer D. A., Anglada-Escude G., Arriagada P., Baluev R. V., Bean J. L., Bouchy F., Buchhave L. A., et al., 2016, *PASP*, 128, 066001.

Gregory P. C., 2016, *MNRAS*, 458, 2604.

Guillot T., Gautier D., 2014, arXiv, arXiv:1405.3752.

Hara N. C., Boué G., Laskar J., Correia A. C. M., 2017, *MNRAS*, 464, 1220.

Hara N. C., Unger N., Delisle J.-B., Díaz R., Ségransan D., 2021, arXiv, arXiv:2105.06995

Hatzes A. P., Dvorak R., Wuchterl G., Guterman P., Hartmann M., Fridlund M., Gandolfi D., et al., 2010, *A&A*, 520, A93.

Hatzes A. P., Fridlund M., Nachmani G., Mazeh T., Valencia D., Hébrard G., Carone L., et al., 2011, *ApJ*, 743, 75.

Hatzes A. P., 2013, *AN*, 334, 616.

Haywood R. D., Collier Cameron A., Queloz D., Barros S. C. C., Deleuil M., Fares R., Gillon M., et al., 2014, *MNRAS*, 443, 2517.

Haywood R. D., Vanderburg A., Mortier A., Giles H. A. C., López-Morales M., Lopez E. D., Malavolta L., et al., 2018, *AJ*, 155, 203.

Howard A. W., Johnson J. A., Marcy G. W., Fischer D. A., Wright J. T., Henry G. W., Isaacson H., et al., 2011, *ApJ*, 726, 73.

effreys S. H., 1961, *The Theory of Probability*. Oxford Univ. Press, Oxford

Jurgenson C., et al., 2016, *SPIE*, 99086T, *SPIE*.9908

Klein B., Zicher N., Kavanagh R. D., Nielsen L. D., Aigrain S., Vidotto A. A., Barragán O., et al., 2022, *MNRAS*, 512, 5067. doi:10.1093/mnras/stac761

Ksanfomaliti L. V., 1999, *SoSyR*, 33, 482

Léger A., Rouan D., Schneider J., Barge P., Fridlund M., Samuel B., Ollivier M., et al., 2009, *A&A*, 506, 287.

Lomb N. R., 1976, *Ap&SS*, 39, 447

López-Morales M., Haywood R. D., Coughlin J. L., Zeng L., Buchhave L. A., Giles H. A. C., Affer L., et al., 2016, *AJ*, 152, 204.

Martins J. H. C., Figueira P., Santos N. C., Melo C., Muñoz A. G., Faria J., Pepe F., et al., 2018, *MNRAS*, 478, 5240

Meunier N., 2021, arXiv, arXiv:2104.06072

Mayor M., Queloz D., Marcy G., Butler P., Noyes R., Korzennik S., Krockenberger M., et al., 1995, *IAUC*, 6251

Mégevand D., et al., 2014, *SPIE*, 91471H, *SPIE*.9147

Mortier A., Collier Cameron A., 2017, *A&A*, 601, A110.

Motalebi F., Udry S., Gillon M., Lovis C., Ségransan D., Buchhave L. A., Demory B. O., et al., 2015, *A&A*, 584, A72.

Nava C., López-Morales M., Haywood R. D., Giles H. A. C., 2020, *AJ*, 159, 23.

Pepe F., Mayor M., Delabre B., Kohler D., Lacroix D., Queloz D., Udry S., et al., 2000, *SPIE*, 4008, 582.

Pepe F. A., Cristiani S., Rebolo Lopez R., Santos N. C., Amorim A., Avila G., Benz W., et al., 2010, *SPIE*, 7735, 77350F

Pont F., Aigrain S., Zucker S., 2011, *MNRAS*, 411, 1953.

Queloz D., Henry G. W., Sivan J. P., Baliunas S. L., Beuzit J. L., Donahue R. A., Mayor M., et al., 2001, *A&A*, 379, 279.

Queloz D., Bouchy F., Moutou C., Hatzes A., Hébrard G., Alonso R., Auvergne M., et al., 2009, *A&A*, 506, 303.

Robertson P., Roy A., Mahadevan S., 2015, *ApJL*, 805, L22.

Saar S. H., Donahue R. A., 1997, *ApJ*, 485, 319

Salabert D., García R. A., Beck P. G., Egeland R., Pallé P. L., Mathur S., Metcalfe T. S., et al., 2016, *A&A*, 596, A31.

Santos N. C., Mortier A., Faria J. P., Dumusque X., Adibekyan V. Z., Delgado-Mena E., Figueira P., et al., 2014, *A&A*, 566, A35.

Santos N. C., Buchhave L. A., 2018, *haex.book*, 181.

Scargle J. D., 1982, *ApJ*, 263, 835

Seager S., Kuchner M., Hier-Majumder C. A., Militzer B., 2007, *ApJ*, 669, 1279.

Schwab C., et al., 2016, *SPIE*, 99087H, *SPIE*.9908

Suárez Mascareño A., Faria J. P., Figueira P., Lovis C., Damasso M., González Hernández J. I., Rebolo R., et al., 2020, *A&A*, 639, A77

Thompson S. J., Queloz D., Baraffe I., Brake M., Dolgoplov A., Fisher M., Fleury M., et al., 2016, *SPIE*, 9908, 99086F.

Tuomi M., Anglada-Escude G., Jenkins J. S., Jones H. R. A., 2014, arXiv, arXiv:1405.2016

Udry S., Mayor M., 2001, *ESASP*, 496, 65

Vogt S. S., Burt J., Meschiari S., Butler R. P., Henry G. W., Wang S., Holden B., et al., 2015, *ApJ*, 814, 12

Wilson O. C., 1978, *ApJ*, 226, 379. doi:10.1086/156618

Wilson T. G., Goffo E., Alibert Y., Gandolfi D., Bonfanti A., Persson C. M., Collier Cameron A., et al., 2022, *MNRAS*, 511, 1043.

Wright J. T., Robertson P., 2017, *RNAAS*, 1, 51

Wu Y., Chen S., Wang P., Zhou S., Feng Y., Zhang W., Wei R., 2021, *MNRAS*, 503, 3032

Zechmeister M., Kürster M., 2009, *A&A*, 496, 577

Zeng L., Sasselov D. D., Jacobsen S. B., 2016, *ApJ*, 819, 127.

Zeng L., Jacobsen S. B., Hyung E., Levi A., Nava C., Kirk J., Piaulet C., et al., 2021, *ApJ*, 923, 247. doi:10.3847/1538-4357/ac3137

Zhao J., Ford E. B., 2022, arXiv, arXiv:2201.03780

Zhao J., Tinney C. G., 2020, *MNRAS*, 491, 4131. doi:10.1093/mnras/stz3254

APPENDIX A:

A1 SCALPELS

The SCALPELS identified shape-driven perturbations are strongly correlated with the observed RVs, as seen from the first panel of Figure A1, which could reliably reproduce the stellar activity dominated long-term and short-term fluctuations.

APPENDIX B:

B1 Blind search periodograms

We applied SCALPELS to the barycentric RVs of CoRoT-7, and examined the results using a Generalized Lomb-Scargle periodogram (Zechmeister & Kürster 2009) for different subsets of the data (2008-9, 2012 and combined). Histograms showing the reduction of RMS scatter after separating the shape-driven signals are presented for all three analyses in Figure B1.

B2 Simultaneous sinusoidal fit : Steps

After computing the Singular Value Decomposition(SVD) of the ACF rows;

- (i) Compute $\mathbf{F} = \{\cos \omega_1 t_j, \sin \omega_1 t_j, \dots, \cos \omega_n t_j, \sin \omega_n t_j\}$.
- (ii) Concatenate $\mathbf{A}_{(m \times (2n+l))} = [\mathbf{F}_{m \times 2n} \quad \mathbf{U}_{A, (m \times l)}]$ using reduced rank l .
- (iii) Compute reduced-rank covariance matrix using:

$$\frac{1}{m} \mathbf{R}_{k_{\max}}^T \cdot \mathbf{R}_{k_{\max}} = \frac{1}{m} (\mathbf{P}_{C, k_{\max}} \cdot \text{diag}(\mathbf{S}_{C, k_{\max}}^2) \cdot \mathbf{P}_{C, k_{\max}}^T). \quad (\text{B1})$$

- (iv) Compute row variances σ_j^2 of $\mathbf{R} - \mathbf{R}_{k_{\max}}$
- (v) Compute model of full covariance matrix as:

$$\Sigma_{\xi, \eta}(t_j) \simeq \sigma_j^2 \times \max\left(1 - \frac{|v_{\xi} - v_{\eta}|}{\delta v}, 0\right) + \frac{1}{m} \mathbf{R}_{k_{\max}}^T \cdot \mathbf{R}_{k_{\max}}. \quad (\text{B2})$$

- (vi) Compute \mathbf{C}' and $\text{Var}(\mathbf{v}(t_j))$ using:

$$\text{Var}(\mathbf{v}(t_j)) = \frac{1}{\mathbf{C}'(v, t_j)^T \cdot \Sigma^{-1}(t_j) \cdot \mathbf{C}'(v, t_j)}. \quad (\text{B3})$$

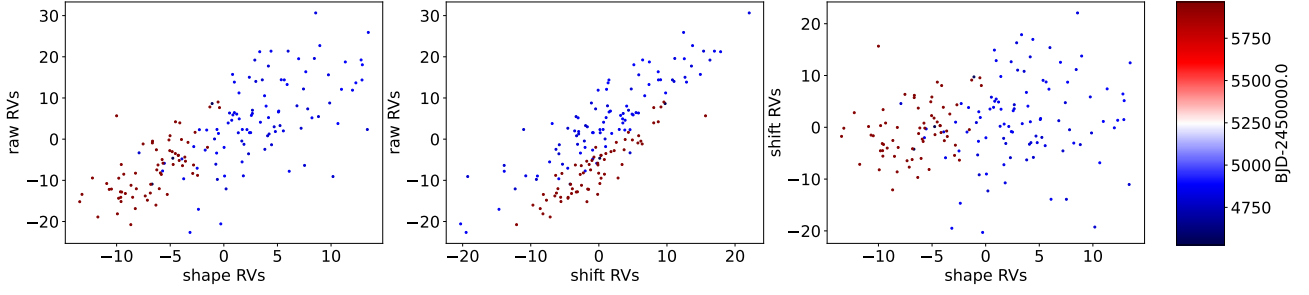


Figure A1. The correlation plots for the observed velocities (v_{obs}), shape-driven component ($v_{||}$) and the shift component (v_{\perp})

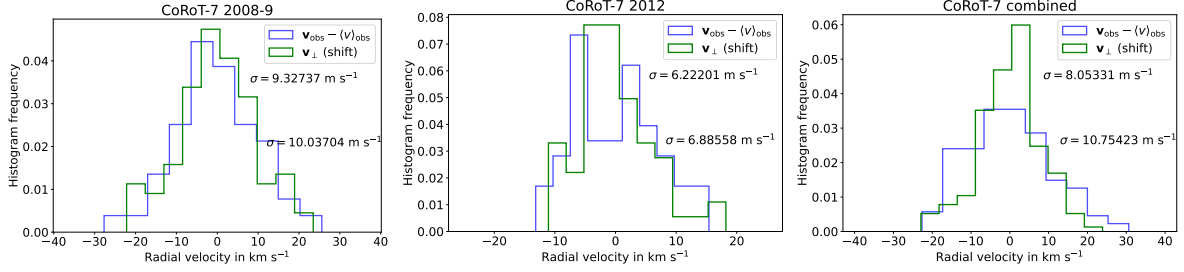


Figure B1. Histograms showing the decrease in RMS scatter in the shift-driven velocities (shown in green) during the blind-period search for 2008-9, 2012 and combined data respectively.

Table B1. Top: The periods and semi-amplitudes of the strongest signals in the periodograms from simultaneous modelling of CCF shape changes and planetary motion, made with prior knowledge of the periods (from I1) of 3 planet candidates on a single period representing the stellar rotation are listed. **Bottom** A pair of beat period corresponding to the stellar rotation is included instead of the single period, considering the varying phase and amplitude.

P(days)	K(ms ⁻¹)	σ_K (ms ⁻¹)
0.8535	2.915	0.280
3.6963	5.112	0.302
8.9674	3.223	0.292
22.9425	5.437	0.321
0.8535	3.099	0.284
3.6963	5.343	0.303
8.9674	3.320	0.293
22.9425	5.142	0.326
23.6968	2.923	0.337

(vii) Construct $\Sigma = \text{Diag}(\text{Var}(\mathbf{v}(t_j)))$.

(viii) Solve

$$(\mathbf{A}^T \cdot \Sigma^{-1} \cdot \mathbf{A}) \cdot \theta = \mathbf{A}^T \cdot \Sigma^{-1} \cdot \mathbf{v}_{obs} \quad (\text{B4})$$

to obtain θ and $\text{Var}(\theta) = 1/\text{Diag}(\mathbf{A}^T \cdot \Sigma^{-1} \cdot \mathbf{A})$.

(ix) Partition θ into coefficients of columns of \mathbf{F} and $\hat{\alpha}$.

(x) Compute $\mathbf{v}_{||} = \hat{\alpha} \cdot \mathbf{U}_A$ in reduced-rank basis.

(xi) Compute $\mathbf{v}_{\perp} = \mathbf{v}_{obs} - \mathbf{v}_{||}$.

Return: RV amplitudes and variances, $\mathbf{v}_{||}$, \mathbf{v}_{\perp} .

APPENDIX C:

C1 l_1 -periodograms for shape RVs & activity indicators

The physical origin of the planet candidate signals, were validated by computing additional l_1 periodograms for the shape-driven RVs and the activity-indicators, as shown in Figure C1.

APPENDIX D:

D1 BGLS periodogram for activity indicators

We studied the shape changes of CCFs comprehensively to investigate intrinsic stellar variability. The spectral parameters such as FWHM, BIS, and Area are used for this diagnostic, as these offer different measures of shape changes in the CCF profile. The Ca II H&K S-index is also used as a proxy for active regions on the stellar surface. While the FWHM carries information about change in the width of the line profile, the BIS records the profile asymmetry. The Area is obtained by taking the product of FWHM and the central line depth. The time series information of these all these indicators for different observing runs are detailed in Section 4.1 along with Figure 1.

Here, we analyse the Bayesian Generalized Lomb-Scargle periodograms of the above-mentioned activity indicators (Figure D2). The periods corresponding to the stellar rotation period and its harmonics are represented by red dotted vertical lines, while the candidate signal periods are marked in green. The primary intention is to examine whether any of these periodograms has peaks at or near the candidate signals of interest. We also examined the BGLS periodograms of the first 4 leading principal components that contributed significantly in modelling the stellar activity (of Figure D1), to understand how efficiently the Principal Component Analysis of the ACF correlate with the known aforesaid activity proxies.

We find that the second principal component (\mathbf{U}_1) exhibits a strong resemblance to the variability of CCF area. Another correlation is found between the third principal component (\mathbf{U}_2) and the FWHM, indicating that the changes in the width of the profile are also reliably considered while modelling the shape-driven component.

Except the first principal component and the S-index, all other principal components and activity indicators trace down the stellar rotation period efficiently and hence show significant power at ~ 23 d. Unsurprisingly, no traces of power are found in any

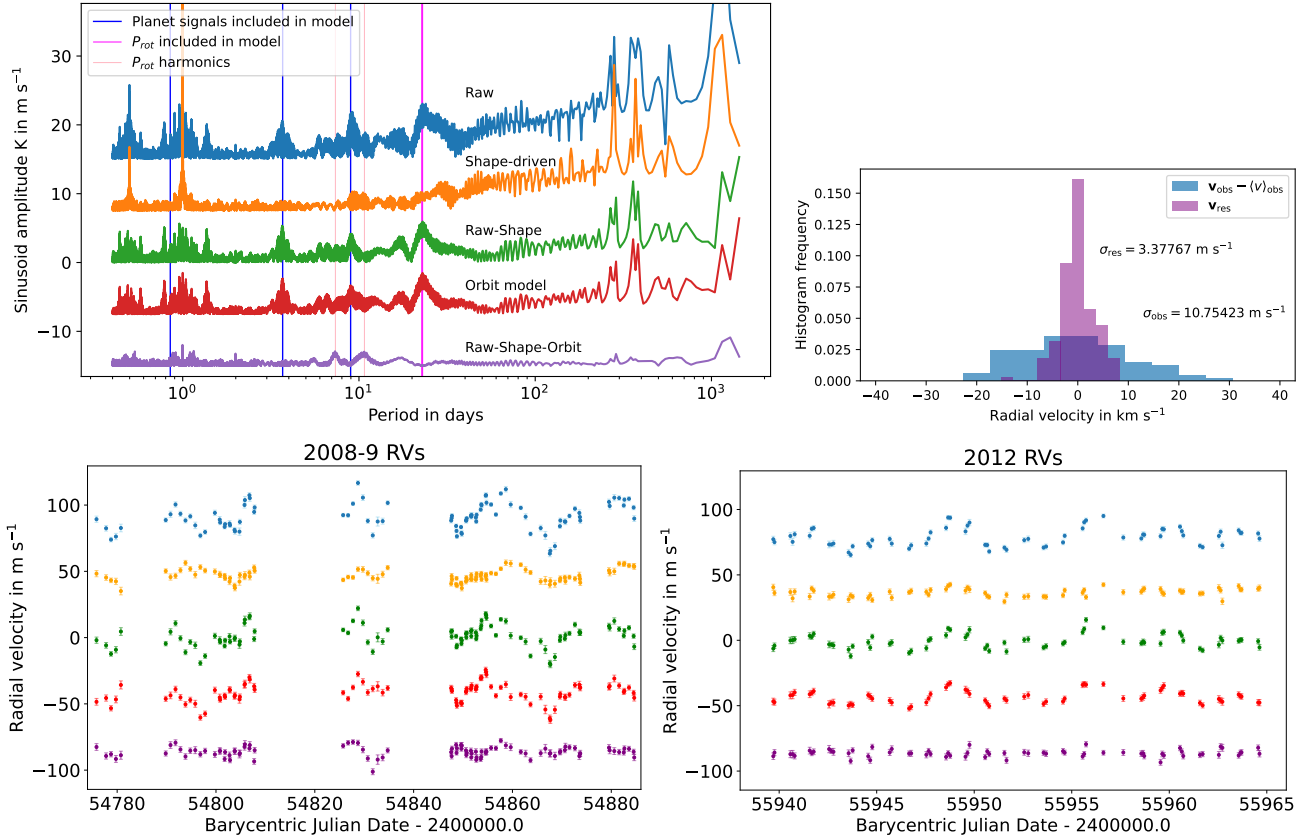


Figure B2. **Left:** All the signals found by the ℓ_1 -periodogram are fitted including the 3 planet candidate signals (blue vertical lines) and 1 pairs of closely spaced beat period associated with the rotation period (magenta vertical lines). The orbital model signal (red) matches with the shift signal (green), except for the harmonics (light pink vertical bars) of the stellar rotation period. **Right:** The resultant histogram of residuals with significantly reduced RMS scatter is shown.

of the periodograms at 0.85 days, 3.70 days and 8.97 days, thus ruling out activity as a plausible origin for these signals. We also noticed that, while \mathbf{U}_3 traces some stellar rotation harmonics very efficiently ($P_{\text{rot}}, \frac{P_{\text{rot}}}{2}, \frac{P_{\text{rot}}}{3}, \frac{P_{\text{rot}}}{6}$ and $\frac{P_{\text{rot}}}{7}$), the Bisector Inverse Slope (BIS) has peaks at almost every harmonics ($P_{\text{rot}}, \frac{P_{\text{rot}}}{2}, \frac{P_{\text{rot}}}{3}, \frac{P_{\text{rot}}}{4}, \frac{P_{\text{rot}}}{5}, \frac{P_{\text{rot}}}{6}$ and $\frac{P_{\text{rot}}}{7}$). In the BIS periodogram, the second harmonic ($\frac{P_{\text{rot}}}{3} = 7.8\text{d}$) has a significantly greater power than the original stellar rotation period ($P_{\text{rot}} \sim 23\text{d}$).

APPENDIX E:

E1 Stacked BGLS periodogram- significance check

The top panel shows the Stacked BGLS periodogram for the 2008-9 observing season (Queloz et al. 2009) with 106 observations, where the periods of four candidate signals are marked using arrows. As anticipated, the region around period ~ 1 day appears to be crowded due to the presence of aliases of several signals, which makes it difficult to spot the comparatively weaker signal at 0.8549 days corresponding to the transiting planet. However, the zoomed-in plot clearly shows the stable nature of this signal. The 1-day alias of 0.854 days signal is clearly seen at 5.92 d, appearing to share power with the real signal.

Another strong signal is present at 3.689 days which becomes apparent after about 30 observations, growing steadily with more observations (Figure E1&9). The 8.97 days signal appears slightly different, with the probability passing through a short-lived maximum at $\sim 60^{\text{th}}$ observation and then decaying briefly before resum-

ing a monotonic increase thereafter, as expected for a real planetary signal. We expect to get a clear picture of this signal by analysing the entire data set. There is a broad strong signal around 23.47 days with the SNR showing strong fluctuations over the span of observations, which is not expected for a coherent signal. Its initial structure is also confusing with a single sparse signal bifurcating into two strong signals with periods 17.35 days and 23.47 d, where the 17.35 day signal can be explained as the beat period occurring from the 8.97 days signal and the 5.92 days alias of the transiting signal.

The observations from 2012 (Haywood et al. 2014) returned a stacked BGLS periodogram, as shown in the bottom panel of Figure E1. We noticed that the transit signal does not appear to have the same period and stability as seen from the previous figure. Instead, two other strong signals are visible at 0.78d and 0.88 days probably sharing the power from the real planet signal. We can explain the former and the latter ones as the $|\frac{1}{p} \pm 1|^{-1}$ aliasing of the 0.85d signal and the 8.97 days signal, respectively. Unsurprisingly, the 3.68 days signal again presents itself, with the strength steadily increasing towards the maximum number of observations. The broadly spread signal spanning from 6 to 13 days complicates the identification of the 8.97 days signal. The shorter observation baseline during the 2012 observing run probably is insufficient to capture the 23.47 days signal. Mortier & Collier Cameron (2017) also analysed the stacked BGLS periodograms for different subsets and for the entire CoRoT-7 data from both RV campaigns, with particular interest in investigating the planetary significance of the ~ 9 days signal. They reported the origin of the ~ 9 days signal to be more likely from stellar activity than a third

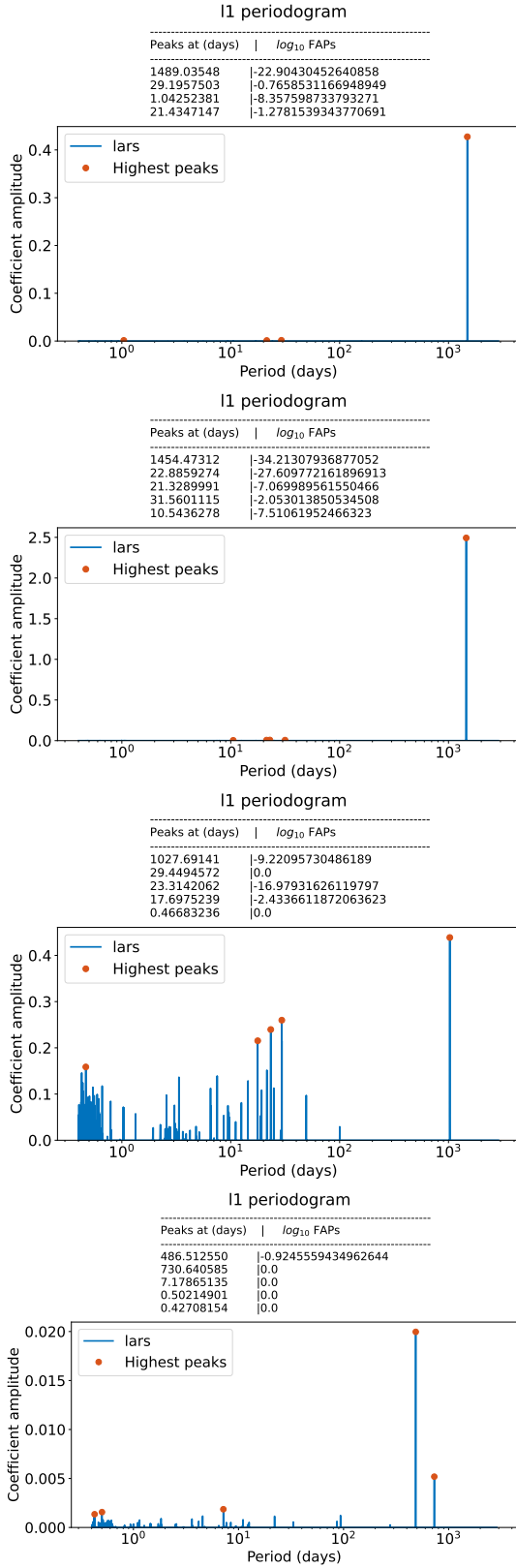


Figure C1. Top: The ℓ_1 -periodogram for the shape-driven RVs and traditional activity indicators like FWHM, Area and the BIS are shown from top to the bottom panels. None of these show significant signals at 0.85 days, 3.69 days and 8.96 days, adding to the case for a dynamical origin. The corresponding False Alarm Probabilities (FAP) are also listed above each panel.

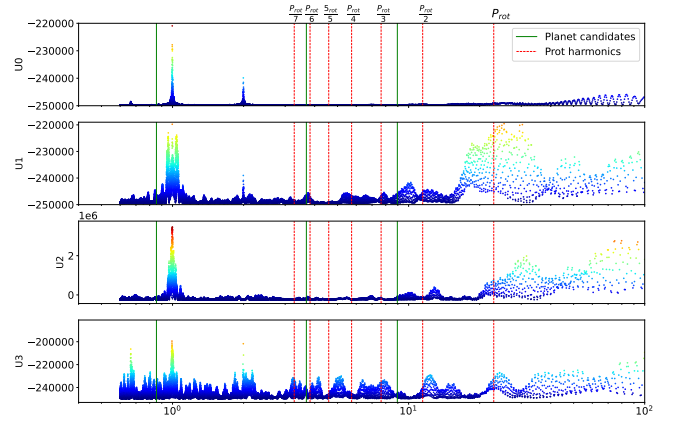


Figure D1. Bayesian Generalized Lomb-Scargle periodograms of the 4 leading principal components of the residual CCF of RV time series. Except the first, all other principal components exhibit power around the rotation period and the first and second harmonics, denoted by red-dotted vertical bars at ~ 23 d, ~ 11.5 days and ~ 7.6 d. Similar vertical bars in green marks the periods of candidate planet signals.

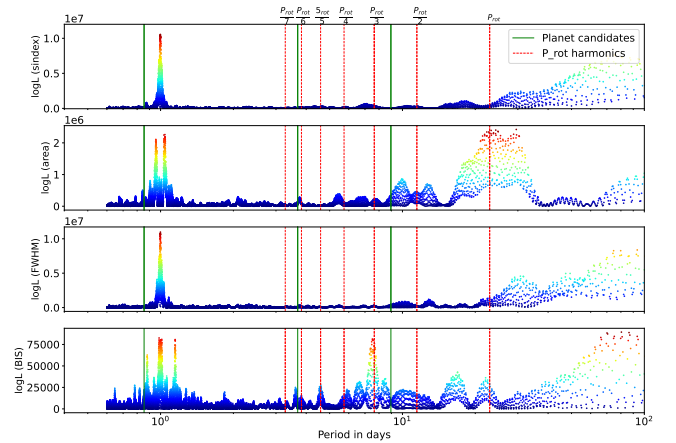


Figure D2. Bayesian Generalized Lomb-Scargle periodograms of activity proxies such as S-index, Area, Full Width Half Maximum (FWHM) and Bisector span (BIS). The vertical lines have the same meaning as in the panel above.

planet based on the unstable nature recorded in the original dataset (Queloz et al. 2009), the follow-up campaign (Haywood et al. 2014) and in the combined data set. In the present analysis, we have the advantage of handling the SCALPELS-treated data, providing us with less complicated, activity corrected, RV information spanning over a longer baseline. Having effectively eliminated the short-term and long-term activity and trends, we are left with well-defined signals whose nature is less ambiguous.

It is also worth noting that most of the aliasing and beat period signals that appeared in the individual data set analyses (Section E1) are no more obvious in the stacked BGLS periodogram for the entire data set.

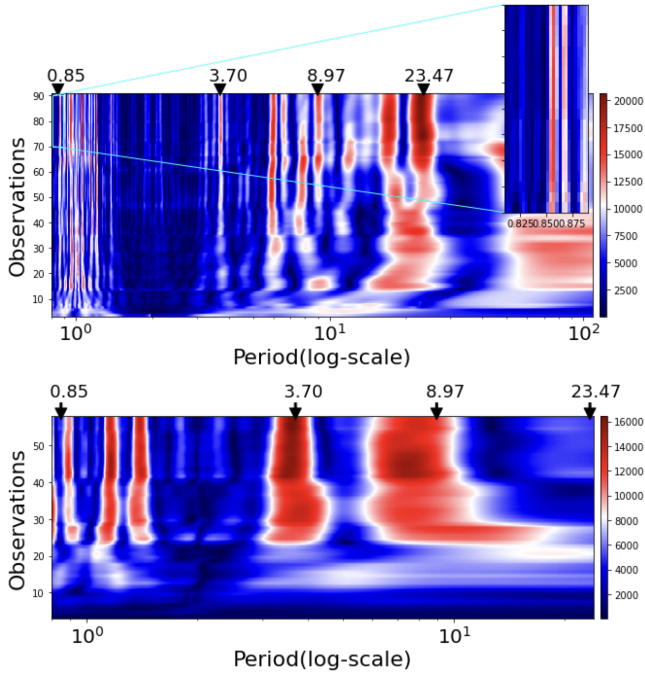


Figure E1. SBGLS periodograms for the 2008-9 and 2012 data subsets are given in the left and right panels respectively.

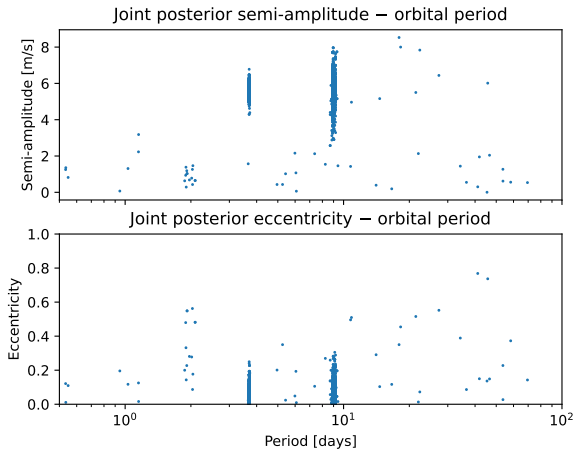


Figure F1. Joint posterior distribution for the semi-amplitudes and eccentricities along with the orbital periods in the x-axes

APPENDIX F:

F1 KIMA posteriors

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

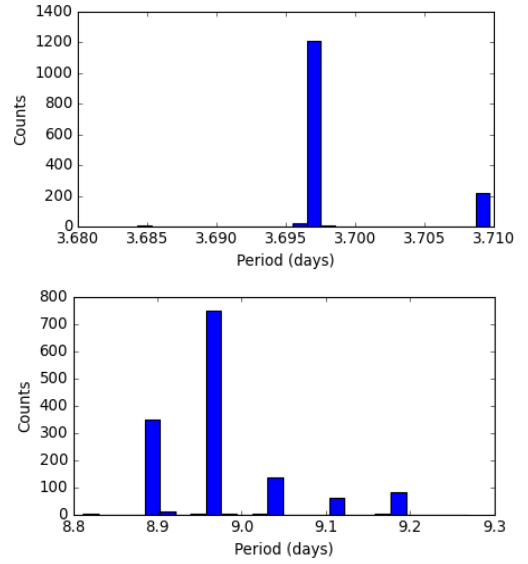


Figure F2. The FIP histograms showing the most probable periods for CoRoT-7c and CoRoT-7d from the corresponding interference patterns.

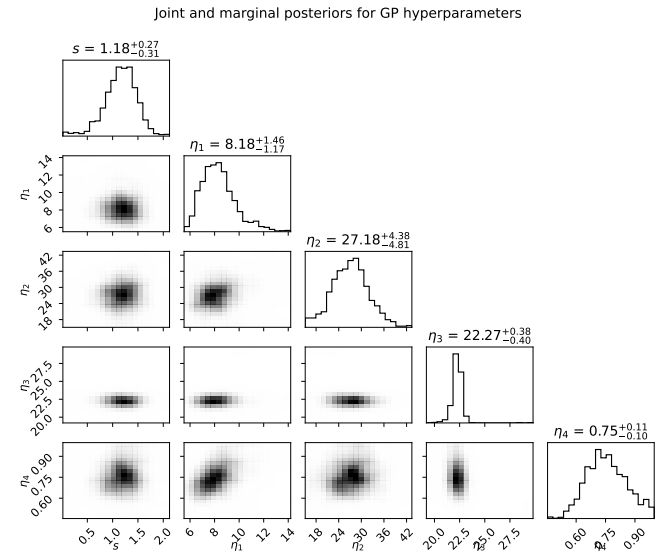


Figure F3. Posterior distributions for the GP parameters and the extra white noise. The samples for all values of N_p were combined. The median of the posterior and the uncertainties calculated from the 16% and 84% quantiles are given in respective titles.