Astronomy & Astrophysics manuscript no. aa54336-25
November 10, 2025

EMPEROR

I. Exoplanet MCMC parallel tempering for RV orbit retrieval

Pablo A. Peña R.[1,2] and James S. Jenkins[1,2]

[1] Instituto de Estudios Astrofísicos, Facultad de Ingeniería y Ciencias, Universidad Diego Portales, Av. Ejército 441, Santiago, Chile

[2] Centro de Astrofísica y Tecnologías Afines (CATA), Casilla 36-D, Santiago, Chile

ABSTRACT

Aims. This paper presents `EMPEROR`, an open-source Python-based framework designed for the efficient detection and characterisation of exoplanets by using radial velocity (RV) methods. Its combination of performance, flexibility, and ease of use makes it a robust tool for any exoplanet detection endeavour. `EMPEROR` integrates Dynamic Nested Sampling (DNS) and Adaptive Parallel Tempering (APT) Markov Chain Monte Carlo (MCMC) techniques, supporting multiple noise models such as Gaussian Processes (GPs) and Moving Averages (MA). The framework facilitates systematic model comparison using statistical metrics, including Bayesian evidence (ln $\mathcal{Z}$) and Bayesian Information Criterion (BIC), while providing automated, publish-ready visualisations.
Methods. `EMPEROR` is evaluated across three distinct systems to assess its capabilities in different detection scenarios. Sampling performance, model selection, and the search for Earth-mass planets are evaluated in data for 51 Pegasi, HD 55693 and Barnard's Star (GJ 699).
Results. For 51 Pegasi, the APT achieves an effective sampling increase by a factor of 3.76 over DNS, while retrieving tighter parameter estimates. For HD 55693 the stellar rotation $P_{\rm rot}$=29.72$^{+0.01}_{-0.02}$ and magnetic cycle $P_{\rm mag}$=2557.0$^{+70.1}_{-36.7}$ are recovered, while demonstrating the sensitivity of ln $\mathcal{Z}$ to prior selection. For Barnard's star, several noise models are compared, and the confirmed planet parameters are successfully retrieved with all of them. The best model shows a period of 3.1536±0.0003 d, minimum mass of 0.38±0.03 M$_\oplus$, and semi-major axis of 0.02315±0.00039 AU.
Conclusions. Purely statistical inference might be insufficient on its own for robust exoplanet detection. Effective methodologies must integrate domain knowledge, heuristic criteria, and multi-faceted model comparisons. The versatility of `EMPEROR` in handling diverse noise structures, its systematic model selection, and its improved performance make it a valuable tool for RV exoplanetary studies.

Key words. Methods: numerical – Methods: data analysis – Techniques: radial velocities – Planets and satellites: detection – Planets and satellites: individual: Barnard's star – Planets and satellites: individual: HD 55693

. Introduction

ithin the realm of possible planetary systems orbiting stars other han the Sun, there appears to be an almost unlimited number of ossible configurations. The field of exoplanet research is still elatively young; however, the diversity of planetary systems s notably high (e.g. Mayor & Queloz 1995; Jones et al. 2006; appaport et al. 2012; Anglada-Escudé et al. 2013; Gillon et al. 017, and many others).

Detection is the first step in fully comprehending the nature of lanetary systems observationally. Novel methods are continually developed to uncover systems that are difficult to detect, such as very low-mass planets, multi-planet systems, wide-orbiting companions, or systems with unusual architectures. Numerous new techniques have been proposed that extend beyond the traditional approach of utilising Lomb-Scargle periodograms (LSP, Lomb 1976; Scargle 1982). As such, newer periodogram schemes, as well as new algorithms, have been developed, allowing for more information to be extracted from the data, e.g. Generalised LS (Zechmeister & Kürster 2009), Minimum Mean Squared Error (Jenkins et al. 2014), maximum likelihood periodograms (Baluev 2013), Gaussian Processes (Rajpaul et al. 2015), Deep Learning (Shallue & Vanderburg 2018), amongst many others (see Salman Khan et al. 2016 for a short overview).

The application of Bayesian modelling under a Parallel Tempering Markov Chain Monte Carlo (MCMC) framework (Gregory 2007), particularly when combined with correlated-noise modelling, has recently made strides in the signal detection arena, including the hunt for exoplanets. These methods have led to the discovery of small planet candidates orbiting Sun-like stars (Tuomi & Anglada-Escudé 2013; Jenkins et al. 2013), multi-planet systems orbiting small M dwarf stars (Anglada-Escudé et al. 2016; Gillon et al. 2017), and in known systems they have improved constraints on planet candidates or revealed new ones (Feng et al. 2017). In addition, these methods have helped to disprove or revise previous claims, such as those for α Centauri B b (Dumusque et al. 2012; Hatzes 2013; Rajpaul et al. 2015), GJ 581 d and g (Vogt et al. 2010; Robertson et al. 2015), or Kapteyn b (Anglada-Escude et al. 2014; Robertson et al. 2015).

Despite these advances, the application of MCMC methods in RV inference presents key challenges–the prolonged computing time required to reach an optimal solution; the multi-modal nature of Keplerian signals which complicates ensuring that the global posterior maximum has been identified; and the difficulty of achieving full automation in the analysis.

Several radial velocity (RV) fitting frameworks address different parts of this landscape, for example `RadVel` (Fulton et al.

2018), `Juliet` (Espinoza et al. 2019), `kima` (Faria et al. 2018), and `exostriker` (Trifonov 2019). Whilst these provide flexibility in planet detection and parameter estimation, here we introduce `EMPEROR` (Exoplanet Mcmc Parallel tEmpering for Rv Orbit Retrieval), a Python-based, modular code that offers a single automated workflow that, i) natively employs an affine-invariant Adaptive Parallel Tempering (APT) MCMC engine with built-in evidence estimation; ii) supports modular per-instrument noise and stellar-activity modelling, and optional dynamical-stability priors for multi-planet solutions; and iii) pairs an auto-compiled, parallelised likelihood back-end with publish-ready diagnostics, enabling reliable exploration of broad parameter spaces with minimal manual intervention.

Internally, the posterior is sampled using an APT implementation built on the affine-invariant ensemble sampler, `reddemcee` (Peña R. & Jenkins 2025). Alternative sampling methods, `EMCEE` (Foreman-Mackey et al. 2013), PyMC3 NUTS sampler (Salvatier et al. 2015), and `dynesty` (Speagle 2020), are also available; however, the empirical tests conducted in this study indicate that the APT approach can provide greater confidence in identifying high-probability modes of multi-modal posteriors, subject to the usual caveat that chains substantially exceed the integrated autocorrelation time.

The code supports multi-instrument and multi-planet analysis, as well as various noise models. It conducts automated model comparison to identify and return the optimal model that best describes the data. `EMPEROR` has already been used in published works to detect new planet candidates and confirm existing ones (Wittenmyer et al. 2017; Jenkins et al. 2019a; Barnes et al. 2020; Psaridi et al. 2022; Vines et al. 2023, for example).

This work describes `EMPEROR`'s functionality and the remainder of this manuscript is structured as follows: §2 outlines the Bayesian framework; §3 explains the structure and configuration of `EMPEROR`, including automated pre-processing and post-processing procedures; §4 presents validation tests and benchmarks using three representative cases–1) 51 Pegasi, enabling a direct comparison with published results to evaluate both performance and efficiency of `EMPEROR`; 2) HD 55693, a system where stellar-activity induces RV signals, used to test model comparison under correlated noise; and 3) Barnard's Star, which explores modelling nuances in current ultra-precise data; §5 discusses the complexities of modelling and advantages of using `EMPEROR` in any RV signal detection endeavour; finally, §6 concludes with an overview of the analysis performed in this paper.

2. Bayesian framework

Early RV searches relied on frequentist fitting techniques such as chi-squared minimisation and Lomb-Scargle periodograms (Lomb 1976; Scargle 1982). Coupled with local optimisers such as the Levenberg–Marquardt algorithm (Levenberg 1944; Marquardt 1963), these approaches work well for strong isolated signals, but falter when likelihood surfaces became multi-modal, which is a characteristic of multi-planet systems or low signal-to-noise ratio (SNR) data. Exploring complex Keplerian architectures or alternative noise models quickly became computationally prohibitive, making Earth-mass planets difficult to detect.

Bayesian inference overcomes these hurdles by mapping the full posterior probability of the parameters rather than converging on a single best-fit. MCMC samplers draw from that posterior, naturally quantifying parameter uncertainties (Gregory 2007; Tuomi et al. 2013b). Affine invariant samplers (Goodman & Weare 2010; Hou et al. 2012) are particularly effective in the high-dimensional, highly correlated parameter spaces typical of RV work.

2.1. The Bayesian approach

Bayes' Theorem expresses the posterior $p(\boldsymbol{\theta}|D, M)$ of model parameters $\boldsymbol{\theta}$, given data D and model M, as:

$$p(\boldsymbol{\theta}|D, M) = \frac{p(\boldsymbol{\theta}|M) \cdot p(D|\boldsymbol{\theta}, M)}{p(D|M)} \tag{1}$$

where the numerator combines the prior $p(\boldsymbol{\theta}|M)$ with the likelihood $p(D|\boldsymbol{\theta}, M)$, and the denominator $p(D|M)$ is the Bayesian evidence that normalises the posterior (Robert 2007; Fong & Holmes 2019; Lotfi et al. 2022). Computing this evidence later enables objective model comparison via Bayes factors.

In RV analysis the deterministic component of the signal is a sum of Keplerian curves:

$$\mathcal{K}(t) = \sum_{j=1}^{N_{\rm pl}} K \cdot [\cos(\nu_j(t) + \omega_j) + e_j \cos(\omega_j)] \tag{2}$$

where each planet j is described by the tuple (P_j, K_j, e_j, ω_j, $T_{0,j}$) and ν_j is the true anomaly. Additional terms such as instrumental offsets, long-term trends, or linear correlations with stellar activity indices, expand Eq. 2, without altering its basic form.

The data consist of N RV measurements ($\mathrm{RV_i}$,σ_i,t_i), optionally accompanied by activity indices. Assuming independent Gaussian errors, augmented by an instrument-specific 'jitter' $\sigma_{\rm INS}$, the likelihood is

$$p(D|\boldsymbol{\theta}) = \prod_{\rm INS} \prod_{i}^{N} \sqrt{\frac{1}{2\pi(\sigma_i^2 + \sigma_{\rm INS}^2)}} \exp\left(-\frac{\xi_i^2}{2(\sigma_i^2 + \sigma_{\rm INS}^2)}\right) \tag{3}$$

with residuals $\xi_{i,\rm INS} = \mathrm{RV}_{i,\rm INS} - K(t_i) - \Gamma_{\rm INS}(t_i) - A_{\rm INS}(t_i)$, where Γ captures offsets and trends, and A any linear activity terms. More sophisticated noise models (moving average, red noise, Gaussian Processes, etc.) replace or expand the simple white-noise jitter but fit seamlessly into the same Bayesian machinery.

2.2. Model proposal

A complete RV model combines one or more Keplerian signals with terms that capture instrumental systematics and stellar variability. Each planet contributes the familiar five parameters (P, K, e, ω, ϕ)—period, semi-amplitude, eccentricity, longitude of periastron, and phase of periastron passage—so that the baseline dimensionality is $5N_{\rm pl}$. Nonetheless, Eq. 2 is agnostic to choice of angular variables, and different parameterisations are supported in `EMPEROR`, for example, the textbook (P, K, ϕ, e, ω) (Perryman 2018). The equivalent set (P, K_c, K_s, e_c, e_s), proposed by Hou et al. (2012), where

$$\begin{aligned} K_c &= \sqrt{K}\cos(\phi), \quad & K_s &= \sqrt{K}\sin(\phi) \\ e_c &= \sqrt{e}\cos(\omega), \quad & e_s &= \sqrt{e}\sin(\omega). \end{aligned} \tag{4}$$

This re-parameterisation is used to linearise the circular parameters ω and ϕ, improving sampler performance.

In photometric studies, the time of inferior conjunction T_0 is often preferred over the phase ϕ. The phase-space dependence on T_0 with P makes the circularisation less efficient given the fact that the Jacobian for this change of variable will be different from unity. Consequently, the set (P, K, T_0, e_c, e_s) is recommended for such applications.

Using Mean Anomaly as the angular coordinate is discouraged: it is tied to an arbitrary epoch t_0, making interpretation, and hence uninformative priors, awkward.

Offsets and long-term accelerations are modelled with the polynomial Γ:

$$\Gamma_{i,\mathrm{INS}}(t_i, a) = \sum \frac{\partial^a \gamma}{\partial^a dt} \cdot t^a \tag{5}$$

where a=0 returns a constant offset, a=1, yields a linear term, a=2 a quadratic term, and so on.

First-order linear correlations with activity indices $A_{i,\mathrm{INS}}$, which could be chromospheric stellar activity proxies (e.g. the S-index or $\log R'_{\mathrm{HK}}$, see Jenkins et al. 2008, 2011) or line asymmetry measurements (e.g. full width at half maximum, bisector index slope, see Queloz et al. 2001) are modelled via

$$A_{i,\mathrm{INS}} = \sum_{\mathcal{A}} C_{\mathcal{A},\mathrm{INS}} \cdot \mathcal{A}_{i,\mathrm{INS}} \tag{6}$$

where $\mathcal{A}_{i,\mathrm{INS}}$ denotes each measured activity index, and $C_{\mathcal{A},\mathrm{INS}}$ is the corresponding coefficient.

Short-timescale, colour-correlated noise is captured by a q-order moving average,

$$R_{i,\mathrm{INS}} = \sum_q \Phi_{\mathrm{INS},q} \exp(\frac{-|t_{i-q} - t_i|}{\tau_{\mathrm{INS},q}})\xi_{i-q,\mathrm{INS}} \tag{7}$$

where Φ and τ are the correlation strength and decay timescale, respectively.

With INS counting independent data sets, the number of dimensions of the model is given by $5N_{\mathrm{pl}} + 2(\mathrm{INS} + q) + a + \mathcal{A}$.

2.3. Extended stochastic modelling

`EMPEROR` supports custom modelling options designed to capture low-frequency modulations in stellar signals. Two examples are the single sinusoid (Eq. 8) and the "half-period double sinusoid" (Eq. 9), both of which are used to approximate magnetic cycles or other quasi-periodic behaviours

$$S_i(t) = K_i \cos(\omega_i t + \phi_i) \tag{8}$$

$$M(t) = K_1 \cos(\omega_1 t + \phi_1) + K_2 \cos(2\omega_1 t + \phi_2). \tag{9}$$

For complex or unknown systematics, `EMPEROR` supports Gaussian Processes (Rasmussen & Williams 2006) via the `celerite` library (Foreman-Mackey et al. 2017). GPs are frequently used to model stellar rotation signals, given their quasi-periodic nature. `EMPEROR` provides two distinct pre-configured GP kernels to do so. Both based on mixtures of simple harmonic oscillator (SHO) kernels. The power spectral density of each SHO component is defined by:

$$S(\omega) = \sqrt{\frac{2}{\pi}} \frac{S_0\omega_0^4}{(\omega^2 - \omega_0^2)^2 + \omega^2\omega_0^2/Q^2} \tag{10}$$

where $\rho = 2\pi/\omega_0$ is the undamped period, Q is the quality factor (related to the damping timescale $\tau = 2Q/(\omega_0)$), S_0 is the amplitude of the process, and ω_0 is the undamped angular frequency.

$\mathcal{GP}_{\mathrm{rot}}(\sigma, \rho, Q_0, dQ, f)$ follows Foreman-Mackey (2018) and comprises two SHO components at periods ρ and $\rho/2$, effectively capturing the fundamental rotation frequency and its first harmonic. Five parameters control their amplitudes and damping:

$$\begin{aligned} \rho_1 &= \rho, & Q_1 &= \tfrac{1}{2} + Q_0 + \delta Q, & S_1 &= \frac{\sigma^2}{(1+f)\omega_1 Q_1}, \\ \rho_2 &= \frac{\rho}{2}, & Q_2 &= \tfrac{1}{2} + Q_0, & S_2 &= \frac{f\sigma^2}{(1+f)\omega_2 Q_2}. \end{aligned} \tag{11}$$

The second kernel, $\mathcal{GP}_{\mathrm{grot}}(\rho, \tau, A_1, A_2)$, uses support variables A_1 and A_2 instead of S_1 and S_2 directly, and shares the same overall structure as the first kernel but enforces a common damping timescale, τ, for both SHO components:

$$\begin{aligned} \rho_1 &= \rho, & Q_1 &= \frac{\pi\tau}{\rho_1}, & S_1 &= \frac{A_1}{2\tau}\left(\frac{\rho_1}{\pi}\right)^2, \\ \rho_2 &= \frac{\rho}{2}, & Q_2 &= \frac{\pi\tau}{\rho_2}, & S_2 &= \frac{A_2}{2\tau}\left(\frac{\rho_2}{\pi}\right)^2. \end{aligned} \tag{12}$$

Both kernels offer a flexible way to model spot-modulated stellar rotation, and the choice between them depends on the specific system's characteristics or on prior constraints for the damping timescale, the rotation period or the amplitude.

3. EMPEROR code

`EMPEROR` is a modular framework that automates RV planet searches with minimal hand-tuning. Three high-level stages, preprocessing, run compilation, and post-processing are executed in a loop that can optionally add model 'blocks' (i.e. complexity) and re-launch itself (Fig. 1). At its core lies `reddemcee`, the Adaptive Parallel Tempering sampler, backed by alternative engines (`EMCEE`, `dynesty`, `PyMC3`) that can be swapped in through a single keyword.

Modelling is done via 'blocks', which are Python objects that bundle 1) a computational model, 2) parameter-holding objects 'specs', and 3) metadata such as dimensionality or LaTeX representation. A typical single-planet fit (1K) would include a Keplerian block plus Offset, and Jitter blocks. Because every block carries its own solver, priors, and script-writing routine, new physics can be added (or removed) via one-line flags.

This section describes the inner workings of the code and provides guidance on its usage. The Adaptive Parallel Tempering sampler, `reddemcee`[1], the default custom Kepler Equation solver written in `cython`[2], which natively admits different parameterisations, and `EMPEROR`[3] are publicly available, open source, and hosted on GitHub; and they can be easily installed via `pip`.

3.1. Stage 1 – Preprocessing

Sampler selection: `reddemcee` is an APT MCMC sampling algorithm, designed to efficiently handle highly multi-modal posteriors, and serves as the core engine of `EMPEROR`. Any native

---

[1] `https://github.com/ReddTea/reddemcee`

[2] `https://pypi.org/project/fast-kepler`

[3] `https://github.com/ReddTea/astroemperor`

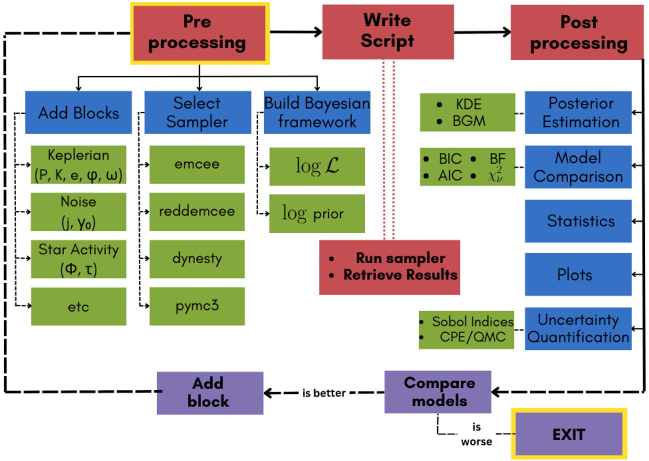

Fig. 1. In red the three distinct steps of the code: pre-processing–where user inputs and data are read; run compilation–where the code is executed; and post-processing–where statistical measures are calculated. The purple blocks with dashed arrows represent the optional step of repeating the loop. Blue highlights the main stages of each step, and available alternatives are displayed in green.

options for the chosen sampler are supplied in a dictionary that `EMPEROR` passes to the secondary script. The available samplers–`reddemcee`, `emcee`, `dynesty`, and `PyMC3`–are briefly explained in Appendix A.

Data ingestion: Each input file is treated as a separate instrument; RVs are mean-subtracted to centre offsets. Columns beyond (time, RV, error) are parsed as activity indices and automatically mapped to the Stellar Activity Block.

Default Bayesian priors: Most parameters employ flat (Uniform) priors, whereas targeted Normal priors are adopted where experience shows it improves convergence (i.e. jitter, eccentricity). In `EMPEROR`, each parameter's prior is attached to its corresponding spec, storing the functional form of the prior and any relevant hyper-parameters. This metadata is then hard-coded into a temporary script generated during runtime, and each of these functions and variables can be adjusted in the code's inputs.

Modelling: Single-line switches activate moving-average noise, Gaussian Process kernels, alternative Keplerian parameterisations (as described in §2.2), and more.

3.1.1. Default Bayesian priors

Period: Uniform prior, where P_{min} is 0.1 days and P_{max} is three times the baseline of the RV data. A common heuristic requires at least one full cycle of data coverage; however, extending the upper bound to three times the observation baseline can reveal long-period signals–detectable with LSPs or Bayesian approaches, although weakly constrained–which could be crucial for disentangling multiple signals.

Semi-amplitude: Uniform prior, with $K_{\mathrm{min}} = 0.1$ (ms^{-1}) and a conservative $K_{\mathrm{max}} = 3 \cdot F$, where F corresponds to the RV value that is furthest from the RV mean, F = max(|RV|). Although some practitioners choose to use 3 times the RMS of the RVs as K_{max}, experience shows that in certain cases (e.g. high amplitude signals with long observational gaps from multiple instruments) large offsets can mimic high amplitudes.

Table 1. Default `EMPEROR` parameter priors.

Parameter	Prior	Parameter	Prior
Keplerian		Others	
P (days)	$\mathcal{U}(0.1, 3 \cdot \max(t))$	γ (ms^{-1})	$\mathcal{U}(-F, F)$
K (ms^{-1})	$\mathcal{U}(0.1, 3 \cdot F)$	θ_{INS} (ms^{-1})	$\mathcal{N}(5, 5^2)$
ϕ (rads)	$\mathcal{U}(0, 2\pi)$	MA Φ	$\mathcal{U}(-1, 1)$
ω (rads)	$\mathcal{U}(0, 2\pi)$	MA τ (days)	$\mathcal{U}(0, 10)$
e	$\mathcal{N}(0, 0.3^2)$	$\dot{\gamma}$ (ms^{-1}/yr)	$\mathcal{U}(-1, 1)$

Notes. $\mathcal{N}$ and $\mathcal{U}$ stand for Normal and Uniform distribution, respectively.

Eccentricity: Truncated Normal prior, $\mathcal{N}(0, \sigma_e^2)$ with σ_e=0.3 as the code's default, where $\Pi(e < 0)$=0. This choice reflects the overall observed distribution of planetary eccentricities, penalising higher eccentricities but allowing them if supported by the data (see Tuomi et al. 2013b). Another common choice for the eccentricity prior is the Beta distribution $\sim \beta(0.867, 3.03)$ (see Kipping 2013).

Offsets and acceleration: Uniform priors, in the intervals $[-F, F]$, and $[-1, 1]$ for γ, and $\dot{\gamma}$, respectively. Where F denotes the full RV coverage.

Jitter: Truncated Normal prior, the default choice is μ_σ=5ms^{-1}, and σ_σ=5ms^{-1}. To avoid the jitter from either blowing up or collapsing to zero, while reflecting typical expectations for the star, it is sensible to use a broad Gaussian prior. These values should be adapted according to the instrument precision or stellar type. For example, when searching for low-mass planets in quiescent stars, a lower mean value is favoured (Wright 2005; Jenkins et al. 2006; Arriagada 2011).

Moving Average: Uniform priors. For the MA Coefficient $\Phi \in [-1, 1]$, to ensure a stationary process that doesn't arbitrarily diverge over time (Tuomi et al. 2013a), and for the MA Timescale $\tau \in [0.1, 10]$ was chosen per Anglada-Escudé et al. (2016). For Φ, another common choice is ± 0.99, avoiding edge behaviour near ± 1, circumventing long auto-correlation tail effects.

3.1.2. Additional priors: dynamical stability

Beyond the standard prior choices (summarised in Table 1), there are additional priors that can be used. A dynamical stability prior can be applied for systems with two or more Keplerian signals, to allow only Hill Stable configurations to be sampled. Following Gladman (1993) first order approximation to the Marchal & Bozis (1982) Hill Criterion:

$$\alpha^{-3}(\mu_1 + \frac{\mu_2}{\delta^2})(\mu_1\gamma_1 - \mu_2\gamma_2\delta)^2 > 1 + \mu_1\mu_2(\frac{3}{\alpha})^{\frac{4}{3}}$$
$$\alpha = \mu_1 + \mu_2; \qquad \gamma_j = \sqrt{1 - e_j^2}; \qquad \delta = \frac{a_2}{a_1} \tag{13}$$

where $\mu_j = \frac{m_j}{M}$, with M the stellar mass, m_j the minimum-mass, a_j the semi-major axis, e_j the eccentricity, and j=1,2 denotes each planet, with subscript 1 standing for the inner orbit.

Alternatively, since the Hill Criterion is less precise for resonant systems, a dynamical stability prior based on the Angular Momentum Deficit (AMD) framework (Petit et al. 2018;

Laskar & Petit 2019) is available. With $\gamma \equiv \frac{m_1}{m_2}$ and $\alpha \equiv \frac{a_1}{a_2}$ the semi-major axis ratio, for a small enough ratio of planetary masses with the host star ε, given that the relative AMD $\mathscr{C}$ verifies the inequality:

$$\mathscr{C} < 1 - (1+\gamma)^{3/2} \sqrt{\frac{\alpha}{\gamma+\alpha}(1 + \frac{3^{4/3}\varepsilon^{2/3}\gamma}{(1+\gamma)^2})} + \gamma\sqrt{\alpha} + O(\varepsilon) \quad (14)$$

3.1.3. Additional priors: parameterisations

Additional priors can become necessary when using changes of variables. For example, in Hou's parameterisation for eccentricity (as in Eqs. 4), the value ranges for e_c, e_s are individually defined as [-1, 1], but the sum of their squares can't go beyond unity, since it would imply eccentricities higher than 1. Another prior enforcing the same eccentricity physical constraint becomes necessary.

3.2. Stage 2 – Compiling the run

Python's process-based parallelism forces every worker to import identical code. To avoid repeated pickling of large data objects, `EMPEROR` auto-generates a secondary script that first loads the data and then defines model functions in situ. This structure removes pickling overhead and delivers near-linear scaling with dataset size. Only the blocks actually requested are written to the script; unused packages and variables are omitted to keep run-time lightweight. Parallelisation back-ends are selected via a single string argument.

3.3. Stage 3 – Post-processing

Posterior estimation: For each parameter, several point estimates are provided: 1) the mean with 1, 2, and 3 standard deviations σ_θ as uncertainties, 2) the median with equivalent $n\sigma_\theta$ percentiles, and 3) the maximum of the posterior with equivalent high-density intervals (HDI).

Optionally, if model-comparison criterion is met, chains can be thinned and fed to a kernel density estimator (KDE) or a Bayesian Gaussian mixture (BGM), to obtain smoother distributions that can be used as priors for subsequent runs. This can be useful for diagnostic runs for signal searching, and should not be used for model comparison.

Statistical diagnostics: `EMPEROR` records acceptance fractions, auto-correlation times, and computes evidence $\ln\mathcal{Z}$ and goodness-of-fit metrics: χ^2 and χ^2_ν–measuring how well the model's predictions align with the observed data. RMSE and RMS$_i$–quantifying the root-mean-square error globally and for individual instruments. AIC, BIC, DIC, HQIC–also for model comparison. `EMPEROR` has the option to increase the number of samples produced, until a pre-determined stopping criterion is met, such as the total length of the chain be equal to n times the auto-correlation time, or having the $\ln\mathcal{Z}$ value stable within a range for n steps. This methodology enables a fully automated algorithm.

Model comparison and outputs: A tolerance-based driver based on the aforementioned goodness-of-fit metrics decides whether to add more signals and repeat the loop (default is $\Delta\text{BIC} \geq 5$). Comprehensive publication-ready figures are generated by an independent suite, which with parallel plotting and data aggregation techniques manages to create them in minimal time.

4. Benchmarks

Three benchmarks assessed `EMPEROR`'s ability to explore the posterior, conduct model comparison, and estimate evidence. The first benchmark uses real data from the well-known star 51 Pegasi, allowing direct comparison with published results to evaluate the sampler's performance. The second benchmark involves model comparison in a system with stellar activity, HD 55693. The final benchmark explores modelling nuances Barnard's Star. These benchmarks will gauge both the performance and efficiency of the `EMPEROR` code.

4.1. 51 Pegasi

51 Pegasi (51 Peg) was the first exoplanet discovered using the RV method (Mayor & Queloz 1995). A quiet Sun-like star with low levels of stellar activity, it presents a single Keplerian signal with high SNR–a hot Jupiter. With enough data and reasonable sampling, the signal is straight-forward to detect, making it an ideal system to validate `EMPEROR`, compare the efficiency of different samplers within the code, as well as comparing against a different exoplanet fitting tool, `Juliet` (Espinoza et al. 2019) in this case.

Stellar parameters and data: 51 Peg is classified as a G2IV star with a visual magnitude of V=5.46, an optical colour of B-V=0.67, and an activity index of $\log R'$HK=–5.054 (Isaacson & Fischer 2010). For this test, 256 LICK RV measurements are utilised, obtained from the Exoplanet Archive [4], based on the work of Butler et al. (2006).

Benchmark discussion: We first test the null hypothesis 0K: a white-noise model comprising offset, jitter, and a linear trend. Then we test a model that adds a single Keplerian component (1K, see Eq. 2).

This benchmark is conducted in three modes, `EMPEROR` coupled with the `reddemcee` sampler (emp-APT), `EMPEROR` with `dynesty` (emp-DNS), and `Juliet` with `dynesty` (jul-DNS). The first comparison, emp-DNS with jul-DNS, seeks to validate `EMPEROR` as a fitting tool, as well as assess its performance. The second one, emp-APT with emp-DNS, validates the APT sampler $\ln\mathcal{Z}$ estimation, as well as its performance compared to DNS. Sampling efficiency is summarised by the effective samples per second, `enits` ≡ ESS/time, which reflects both sampling quality and computational cost. Each configuration is repeated 11 times, we report the mean and standard deviation across repeats.

Since no analytic evidence is available, we assess uncertainty calibration of the Bayesian evidence with a χ^2 consistency test across repeats. For run i with estimate $\ln\hat{\mathcal{Z}}_i$ and reported uncertainty $\hat{\sigma}^2_{\mathcal{Z},i}$, we compute

$$\chi^2_\sigma = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{\ln\hat{\mathcal{Z}}_i - \langle\ln\hat{\mathcal{Z}}\rangle}{\hat{\sigma}_{\mathcal{Z},i}}\right). \quad (15)$$

For χ^2_σ, values ≫1 indicate uncertainty underestimation, ≪1 overestimation, and ≈1 well-calibrated uncertainty estimates.

After performing the emp-DNS runs (1500 live-points, stopping criterion $\Delta\ln\hat{\mathcal{Z}} < 0.01$), `reddemcee` was set to work with (16, 256, 2048) for the number of temperatures, walkers and steps, respectively, to roughly match the total run-time of the K1 of the other method. The Keplerian parameterisation that was available for `Juliet` was used, (P, K, T_0, e, ω), with Period $\sim\mathcal{U}(1, 100)$ and $T_0\sim\mathcal{U}(0, 10)$. Eccentricity and jitter used

---

[4] http://exoplanetarchive.ipac.caltech.edu

Normal priors, $\sim\mathcal{N}(0, 0.3)$, and $\sim\mathcal{N}(0, 15)$, respectively. All other parameters used Uniform priors in accordance to §3.1.1.

Performance differences are apparent: jul-DNS and emp-DNS (see Table 2), using the same sampler, have a similar sampling efficiency (around 3.75% for the $0K$, and around 1.56% for the $1K$). The wall-time, for approximately the same number of evaluations, was considerably lower within `EMPEROR`, at 21% and 10% of the total time, for the $0K$ and $1K$, respectively, leaving the enits in the same proportion.

Within `EMPEROR`, the sampler differences between APT and DNS are evident as well. The wall-time for APT is setup dependant, and since the same setup was used, the $0K$ and $1K$ runs have virtually the same wall-time. Therefore, the enits are used for comparing performance: for the $0K$ the enits are more than tripled with APT (2215.3 vs 666.3), while for the $1K$ they are almost doubled (594.4 vs 319.4). For the APT, the evidence uncertainty χ^2_σ is slightly better calibrated in the $0K$ (1.19 vs 1.28), and much more conservative in the $1K$ (3.13 vs ∼12).

Evidence estimation is consistent between jul-DNS and emp-DNS, for the $0K$ model they arrived at the same value $\ln\hat{\mathcal{Z}} = -1318.7 \pm 0.1$, and for the $1K$ model, emp-DNS and jul-DNS had $\ln\hat{\mathcal{Z}} = -884.12 \pm 2.07$, and -884.53 ± 1.96, respectively.

On the other hand, within `EMPEROR`, the APT and DNS are consistent only for the $0K$ model, with both sharing the estimate $\ln\hat{\mathcal{Z}} = -1318.7 \pm 0.2$. Whereas, for the slightly more complex $1K$ model, APT presents a lower evidence, with a 2-σ difference with respect to the DNS: -880.64 ± 0.52 compared to -884.12 ± 2.07.

To solve this discrepancy, we incrementally increased the 'chain-length' for both methods. In APT, the number of steps were increased up to 10 000, but the estimates remained fairly similar, possibly flagging this $\ln\hat{\mathcal{Z}}$ as a stable estimate.

Similarly, increasing the DNS live-points up to 10 000, showed no changes in the evidence. Nonetheless, by also reducing the stopping criterion to $\Delta\ln\hat{\mathcal{Z}} < 0.001$, the estimate changed to $\ln\hat{\mathcal{Z}} = -880.82 \pm 2.58$ (for 11 runs as well), with a $\chi^2_\sigma = 47.187$. This stricter estimation is consistent with the APT result. For these longer runs, the average run-time was 1473.6 ± 104.1 s, the efficiency was slightly increased to 1.74 ± 0.02, which resulted in 346.3 ± 9.9 enits.

When the phase-space has a complicated shape, there are subtleties that are hard to characterise. For example, as eccentricity tends to zero, both angular parameters become increasingly degenerate. In this system, with eccentricity very close to zero, this degeneracy becomes hard to disentangle. For the longer DNS runs, which had a lower eccentricity estimation $e = 0.006$ (see Table B.1), we can see this phenomena here by looking at the angular parameters, their posteriors flatten, rendering higher uncertainties.

The best solution (see Fig. 2 and Fig. 3) confirms a linear trend in the system, placing tight constraints on the trend, and indicating the presence of a very distant massive companion. The likely companion producing the trend has been imaged by astrometric surveys (Roberts et al. 2011; Kervella et al. 2019) and so the combination of these two detection methods could yield constrained companion properties like mass once at least one RV inflection has been measured.

4.2. HD 55693

Disentangling stellar activity (SA) signals from those of exoplanets presents a significant challenge in RV studies (Dumusque et al. 2011; Fischer et al. 2016). SA can induce RV variations that resemble signals from orbiting exoplanets or interfere with their detection. These variations arise from stellar phenomena such as starspots, plages, and flares, which alter the line profiles of stellar spectra and, consequently, the measured RVs. Additionally, the timescales of stellar activity often overlap with orbital periods of exoplanets (Queloz et al. 2001; Boisse et al. 2011; Díaz et al. 2018), complicating the separation of these signals. Accurately characterising the host star's activity is therefore critical for confirming the presence of an exoplanet and to determine its properties with high precision.

The primary challenge lies in developing robust models capable of effectively distinguishing between stellar activity signatures and the subtle gravitational tug of an exoplanet. Such models often require advanced statistical approaches and long-

Table 2. 51 Peg performance and evidence estimation.

Stat	emp-APT	emp-DNS	jul-DNS
0K			
time (s)	220.5±5.7	49.1±1.4	234.1±1.6
eff (%)	11.37±0.16	3.76±0.01	3.73±0.01
enits	**2215.3±74.5**	666.3±16.6	139.5±0.4
$\ln\hat{\mathcal{Z}}$	-1318.7±0.2	-1318.7±0.1	-1318.7±0.1
$\hat{\sigma}_{\mathcal{Z}}$	0.142±0.005	0.076±0.001	0.076±0.001
χ^2_σ	**1.185**	1.282	1.289
1K			
time (s)	218.8±3.22	250.9±19.2	2630.6±250.3
eff (%)	3.03±0.10	1.65±0.04	1.46±0.16
enits	**594.4 ± 23.3**	319.4±18.3	30.4±2.4
$\ln\hat{\mathcal{Z}}$	-880.64±0.52	-884.12±2.07	-884.53±1.96
$\hat{\sigma}_{\mathcal{Z}}$	0.192±0.027	0.175±0.004	0.175±0.004
χ^2_σ	**3.129**	12.530	11.615

Notes. From left to right, `EMPEROR` with `reddemcee`, `EMPEROR` with `dynesty`, and `juliet` with `dynesty`. From top to bottom, the white noise model ($0K$) and single Keplerian ($1K$), each tab contains descending the total run-time in seconds, eff–the sampling efficiency, enits–effective samples per second, the Bayesian evidence estimate, its estimated uncertainty, and the χ^2 test between internal and empirical evidence uncertainties.

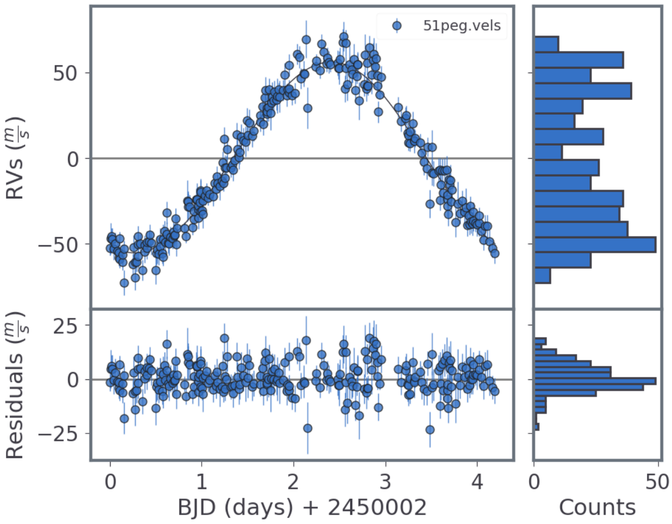

Fig. 2. 51 Peg Keplerian best fit model with `reddemcee`. Top: LICK RVs phase-folded to the period with the best-fit model (black line). Bottom: residuals. Right: histograms of observations and residuals.

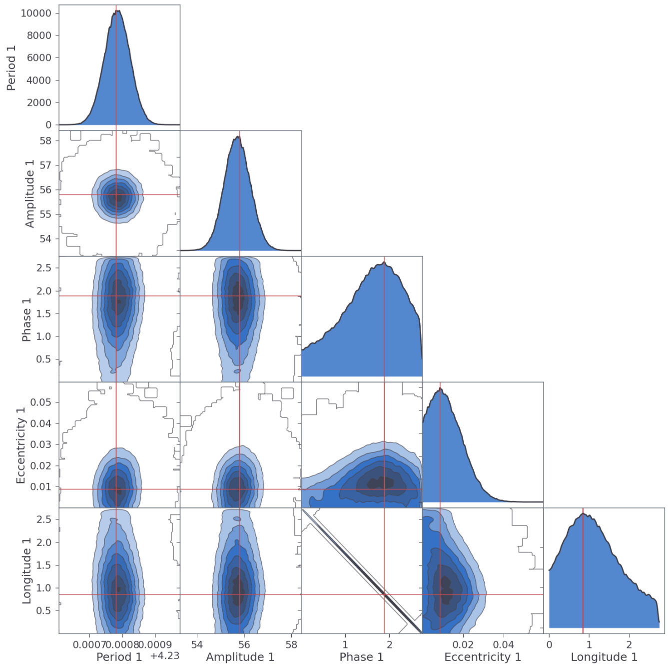

Fig. 3. 51 Peg corner plot of Keplerian parameters. Period, semi-amplitude and eccentricity are well defined narrow Gaussians. The angular parameters, M_0 and ω, look like wide Gaussians, highly correlated with each other, which is the case for $e = 0$.

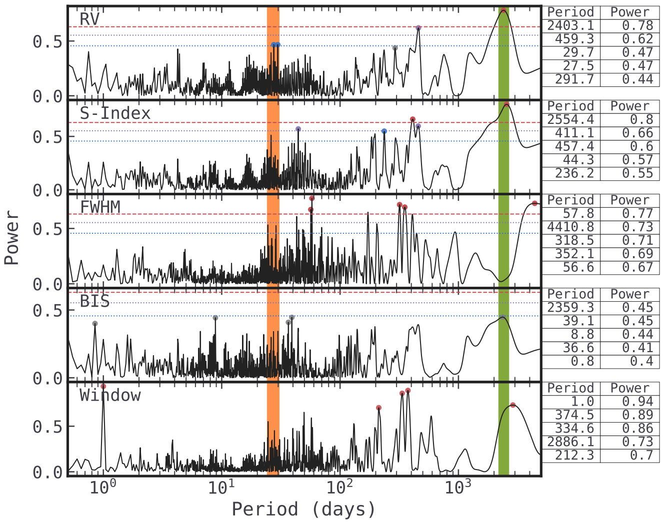

Fig. 4. HD 55693 periodogram for TERRA1 data. Descending, RVs, S-Index, FWHM, BIS, and window function. FAP lines included for 10%, 1% and 0.1%, in dashed red, dotted purple, dotted blue, respectively. Circle markers show the five periods with the greatest power, coloured by FAP region. The orange coloured region corresponds to $P_{\rm rot} = 27.4 \pm 3.2$, while the green one to $P_{\rm mag} = 2403^{266}_{-218}$.

term observational campaigns, including methods like Moving Averages or Gaussian Process regression (Haywood et al. 2014; Rajpaul et al. 2015; Jenkins et al. 2019b; Suárez Mascareño et al. 2020; Vines et al. 2023).

Stellar parameters and data: HD 55693 (HIP 34879) is classified as a G1.5V star (Gray et al. 2006) with a visual magnitude of $V = 7.83$ and an optical colour of $B - V = 0.67$. An activity index of $\log R'_{\rm HK} = -4.963$ was reported by Lovis et al. (2011), indicating a low activity level typical of a G-type star. These authors also identified a significant magnetic cycle with period of $P_{\rm cyc}$=2403^{+266}_{-218} d and a stellar rotation period of $P_{\rm rot}$=27.4±3.2 d.

Two different spectrographs were used for the high-precision RVs–HARPS (Pepe et al. 2000) and PFS (Crane et al. 2010). Since HARPS underwent an optical fibre upgrade in 2015, its RVs were divided into two datasets (before and after the upgrade). The observations were processed with the TERRA pipeline (Anglada-Escudé & Butler 2012), resulting in two datasets designated TERRA1 and TERRA2, containing 29 and 19 RVs, respectively. Each RV is accompanied by stellar activity indices for the S-Index, full width at half maximum (FWHM), and bisector span (BIS). The PFS dataset consists of 36 RV measurements and corresponding S-Index values.

Benchmark discussion: This benchmark will showcase several noise models within the `EMPEROR` framework, as well as the model selection problem: from just the RV data, we will try to recover the star's known rotation period and magnetic cycle, from an uninformed perspective.

Six different noise models are compared: 1) WN–white noise only, 2) SA–linear correlations with stellar activities, 3) MA–an exponentially weighted moving average model, 4) SAMA–a combination of SA and MA, 5) $\mathcal{GP}_{\rm rot}$–Gaussian Processes with a rotation kernel, and 6) $\mathcal{GP}_{\rm rot+SHO}$–Gaussian Processes with rotation and simple harmonic oscillator terms. Each model includes K_n sub-models, where n denotes the number of Keplerian signals. Consistent priors were used across models for equivalent parameters (see Table C.3). A summary of model selection statistics is presented in Table C.1.

An initial inspection of the Lomb-Scargle periodogram (see Fig. 4 for TERRA1) reveals a strong period at 2403 d for RVs (matching the magnetic cycle identified by Lovis et al. (2011)). The S-Index peaks at 2554 d, while the BIS shows a peak in the same range (though below the 10% false-alarm-probability FAP threshold). A signal matching the known rotation period also appears in the RVs (at 29.7 d) and in the S-Indices (26.1 d).

The correlogram for TERRA1 (see Fig. 5) shows a significant correlation between RVs and S-Index (ρ=0.80), as well as between RVs and BIS (ρ=0.56). Notably, the S-Index and BIS are highly correlated at ρ=0.76, suggesting they might be tracing the same physical phenomena. Since only S-Index is available in the PFS dataset, this index alone is employed as part of the SA modelling, for consistency across datasets.

In this preliminary analysis, the high correlation ρ=0.80 implies contamination of the RVs by stellar activity, and the LSP suggests that the ~2500 d peak, shared between RVs and S-index is linked to stellar activity. Other peaks that repeat in both measurements are at ~459 d, ~235 d, ~26 d, and ~193 d.

For a star of this stellar-type and age, we expect a rotation period at ~27-32 days (Mamajek & Hillenbrand 2008), and a magnetic cycle period at ~2200-3300 days (Lovis et al. 2011). This matches the signals appearing in both periodograms at ~2500 d and ~26 d. When using `EMPEROR` for the benchmarked models, each finds first a significant detection at ~2500 d, which we correspond with stellar activity, specifically, the magnetic cycle.

The periodogram of the RV residuals shows peaks under 10% FAP at P=29.7, 56.0, and 19.8 d. And for S-index under 0.1% at P=396.3, 5529.8, and under 1% at P=185.8, 36.7, 27.5, and 19.5 d.

When adding a second signal, the WN(K_2) and MA(K_2) models found P_2=167 d. The $\mathcal{GP}_{\rm rot}(K_1)$ and $\mathcal{GP}_{\rm rot+SHO}(K_0)$ models found P_2=~2.4 d. And the SAMA(K_2) model found P_2=~29.7 d.

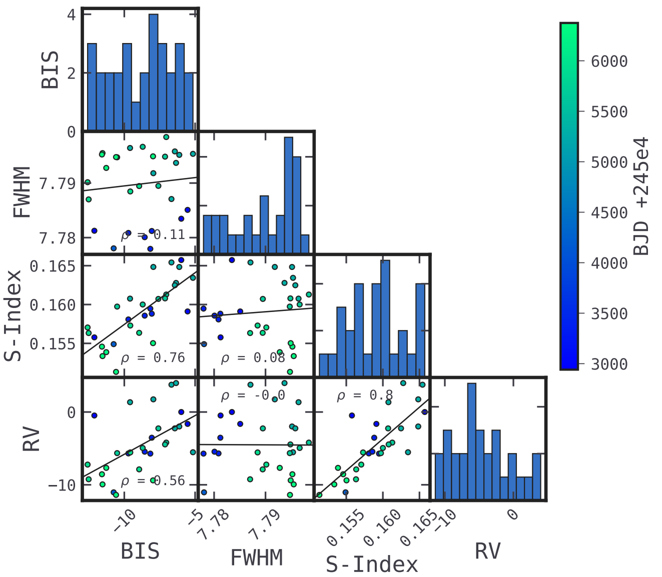

Fig. 5. HD 55693 correlogram for TERRA1 data. Shows the Pearson Correlation coefficient (ρ) between RVs, S-Index, FWHM, and BIS. Diagonal displays the samples distribution. RV presents significant correlation with S-Index ρ=0.8 and BIS ρ=0.56. S-Index with BIS have ρ=0.76, suggesting they describe the same physical phenomena.

A priori, to select the best overall model, consistency across multiple metrics is a good starting point. $\mathcal{GP}_{\mathrm{rot}}(K_1)$ has the highest evidence $\ln\mathcal{Z}$=-211.76 and the lowest RMS=0.81. Nevertheless, its reduced chi-square is well below unity (0.4), flagging over-fitting. For that reason, this model, as well as its cousin $\mathcal{GP}_{\mathrm{rot+SHO}}(K_0)$ are discounted. SAMA(K_2) stands out: It has the second highest evidence $\ln\mathcal{Z}$=-212.03 amongst all models. Its χ^2_ν indicates neither over- nor under-fitting, and its low RMS points to tight residuals. The BIC is just 2.4 higher than the absolute minimum–WN(K_2). This modest penalty is out-weighted by the much larger Bayes factor advantage ($e^{1.26}$≈9.6).

Looking at how the evidence of each noise model evolves when adding signals can give some insight: For the WN model, K_1 presents very strong Bayes factor over the K_0, of $\approx e^{16.33}$. The K_2 over the K_1, although strong, $\Delta\ln\mathcal{Z}$=4.66, barely misses the strict $\Delta\ln\mathcal{Z}$=5 usually required in exoplanet detection.

The MA model should model the lower frequency modulation produced by the rotation period, and has similar behaviour to the WN model, but with decreased effects, a much better K_1 over K_0 solution ($\Delta\ln\mathcal{Z}$=14.82), and a worse improvement in the K_2 over the K_1, $\Delta\ln\mathcal{Z}$=0.21. This already gives some insight that this second signal might be a by-product of stellar rotation.

The SA model, finds the same P_1~2500 d, but doesn't find a stable P_2, with different solutions (e.g. 86, 138, 167, 5200 d). This model should de-trend signals produced by stellar activity, and it is noticeable by the fact that compared to the WN model, the evidence improvement by adding this first signal is much lower $\Delta\ln\mathcal{Z}$=4.30 (and therefore, rejected) compared to the WN's 16.33. The correlation coefficients also go down considerably (e.g. from $\mathcal{A}_{\mathrm{T1}}$=0.81 to 0.44).

The SAMA model, including both effects, shows an improvement of $\Delta\ln\mathcal{Z}$=2.36 when adding the first signal, the lowest so far. This is a moderate indicator that these signals are produced by stellar activity.

With all of this in mind, a conservative conclusion to this RV analysis would be that there are no planets in this system, a magnetic cycle with a period of ~2500 d, and a high frequency rotation period not well characterised within the RV data. Furthermore, since the true P_{mag} and P_{rot} are known, models describing these phenomena are expected to best describe the system, either by removing the stellar activity noise–like the SA(K_0), SAMA(K_0), or $\mathcal{GP}_{\mathrm{rot+SHO}}(K_0)$ models–or by modelling the activity as signals–like the WN(K_2), MA(K_2), or $\mathcal{GP}_{\mathrm{rot}}(K_1)$ models.

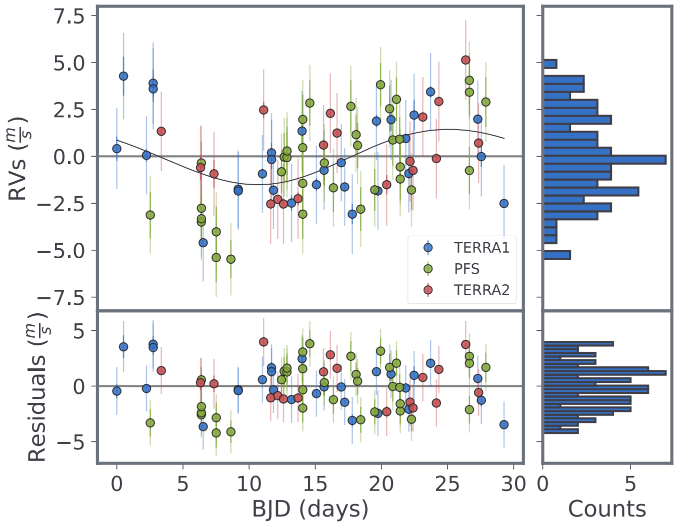

Fig. 6. HD 55693 SAMA(K2) model. Top: RVs phase-folded to P_{rot}=$29.7^{+37.2}_{-15.3}$ d with the best-fit model (black line). Marker colours differentiate datasets, blue (T1), green (PFS), and red (T2). Bottom: RV residuals. Right: histograms of observations and residuals.

The former approach, removing the stellar activity noise, can be seen naturally in this benchmark, by the selection of either the SA(K_0) or SAMA(K_0) as best model. The simplicity of SA(K_0) and its capacity to remove competing signals from the dataset, makes it an attractive plausible physical description of the system.

A simple prior on the SA coefficients support this logic: we compare the SA(K_0) model with different priors for the stellar activity coefficient $\mathcal{A}_{INS}$: 1) default $\sim\mathcal{U}(-1,1)$, 2) constrained around the correlation coefficient ρ and its error (Bonett & Seier 2006) $\frac{(1-\rho^2)}{\sqrt{N-3}}$, and 3) with a Normal prior $\mathcal{N}(\rho,\sigma_\rho)$. The values for ρ and σ_ρ can be found in Table C.2, as well as the metric values. The SA solutions present the same parameter solutions, reflected in the same RMSE=2.73. Notably, these two models, unbounded-normal and bounded-uniform, have similar evidences, but present an improvement of 4.06 and 4.59 over the $\sim\mathcal{U}(-1,1)$ SA(K_0) model. An additional signal under these priors, gives $\ln\mathcal{Z}$=-212.20±0.08 and -212.54±0.13, respectively.

On the other hand, the latter approach, modelling the noise as signals, would require additional information to arrive at, but would give out important details about the system, that is the P_{rot} and P_{mag} of the star. The SAMA(K_2) model features the best overall evidence for non-overfitting models, characterising P_{rot}=$29.7^{+0.01}_{-0.02}$ d (see Fig. 6) and P_{mag}=2557^{+70}_{-38} d (see Fig. 7), in accordance to the presumed true periods for this system. Additional work on the system would be required, like fitting signals on the S-index data alone to impose stronger priors in the RVs, or a simultaneous fit with RVs and S-index, enforcing shared parameters between signals. This is further explored in the following benchmark.

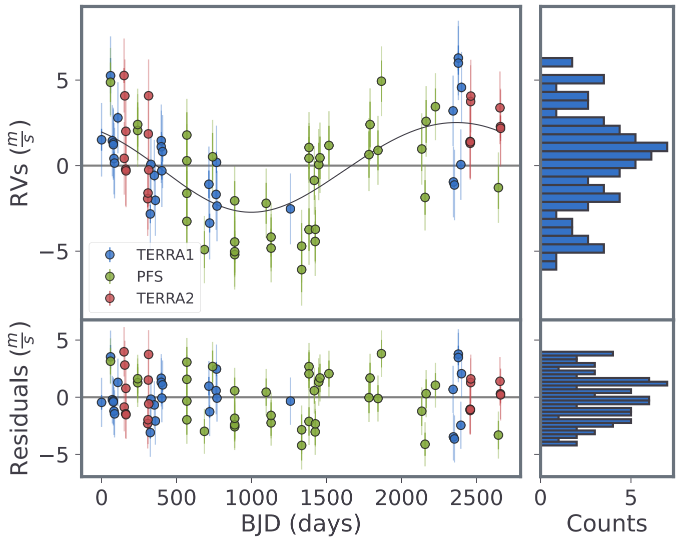

Fig. 7. HD 55693 SAMA(K2) model. Top: RVs phase-folded to $P_{\rm mag}$=2531.6$^{+76.9}_{-8.0}$ d with the best-fit model (black line). Marker colours differentiate datasets, blue (T1), green (PFS), and red (T2). Bottom: RV residuals. Right: histograms of observations and residuals.

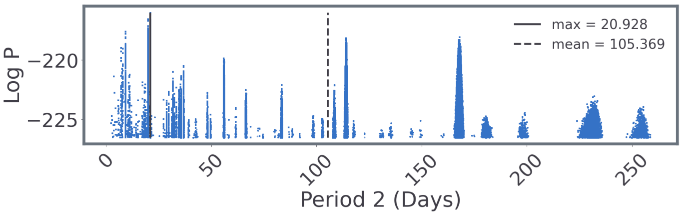

Fig. 8. HD 55693 WN(K_2) model, top of the posterior for $P_{\rm rot}$. There is multi-modality in the solution, with strong peaks at 19, 21, 115 and 170 d. This explains both the high HDI intervals for the parameter (20.93$^{+146.79}_{-19.43}$) and the rejection of this solution.

4.3. Barnard's star

M-dwarf systems, being smaller than the Sun, possess their habitable zone closer to their host. This characteristic makes them prime candidates in the search for Earth-like planets (Scalo et al. 2007; Jenkins et al. 2008; Shields et al. 2016). Barnard's star is not only an M-dwarf but also the closest single-star system to the Sun, at a distance of ∼6 light years. It exhibits the highest known proper motion and a very low level of stellar activity, making it an excellent target in the search for Earth analogues.

Stellar parameters and data: Barnard's star (GJ 699) is classified as an M3.5V-M4V star with a visual magnitude of V=9.51 and an optical colour of B−V=−1.73 (Alonso-Floriano et al. 2015). Toledo-Padrón et al. (2019) reported a low activity level, $\log R'_{\rm HK}$=−5.82, a rotation period of $P_{\rm rot}$=145±15 d and a magnetic cycle of $P_{\rm mag}$=3800±600 d. González Hernández et al. (2024), hereafter GH24, based on 156 ESPRESSO RVs over 4 years found $P_{\rm rot}$=136.2$^{+10.5}_{-9.4}$ d and $P_{\rm mag}$=3325$^{+276}_{-226}$ d, along a Keplerian signal at $P_{\rm Kep}$=3.1533±0.0006 d.

We follow GH24 and analyse 792 RVs obtained with Carmenes (CAR), HARPS (H15), HARPS-N (HAN), and ESPRESSO. Owing to the ESPRESSO fibre-link upgrade, its data is split into pre- and post-upgrade subsets (E18 and E19). After discarding ESPRESSO observations with uncertainties >0.5ms^{-1}, and H15/HAN points with RV uncertainties >0.85ms^{-1} or FWHM uncertainties >2.5ms^{-1}, the final sample comprises 479 CAR, 100 H15, 48 HAN, 9 E18, and 140 E19 measurements (N=776).

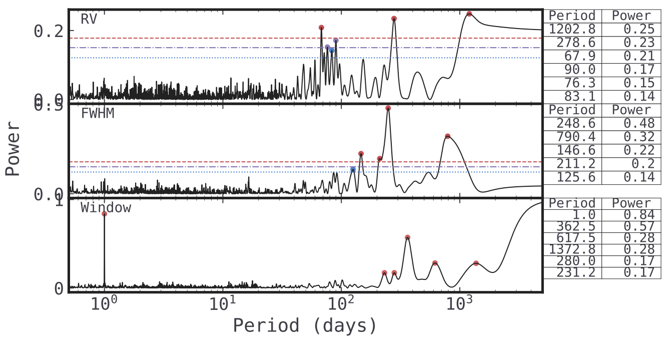

Fig. 9. Barnard's star periodogram for ESPRESSO data. Descending, RVs, FWHM, and window function. FAP lines included for 10%, 1% and 0.1%, in dashed red, dash-dotted purple, dotted blue, respectively. Circle markers show the 5 periods with the greatest power, coloured by FAP region.

Benchmark discussion: The FWHM data was examined first to identify any trends or correlations. Multiple models were tested for the RV data, both with and without incorporating information from the FWHM analysis. This procedure was carried out in parallel for the stand-alone ESPRESSO dataset and for the combined ESPRESSO+HAN+H15 datasets.

The LSP for ESPRESSO FWHM data (see Fig. 9) revealed significant peaks at 249, 790, 147 and 211 d, along with a smaller peak at 126 d. The 249 d peak could correspond to the one-year alias of the 147 d signal, whilst the small bump at 211 d (and the one at 126 d) may be caused by differential rotation. A correlation of ρ=0.43 was found between RV and FWHM (see Fig. 10), suggesting stellar activity contamination in the RVs.

`EMPEROR` was first applied to the ESPRESSO FWHM data with uninformative wide priors (see WN results in Table D.1). The white-noise-only (WN) approach finds first a 245 d signal, followed by 166 d, and 2384 d, each with increasing probability. The first two signals are of special interest, since they also appear in the LSP, and can be recognised as the one-year alias of the rotation period and $P_{\rm rot}$ itself from Toledo-Padrón et al. (2019).

The rotation period was then modelled with GPs. To minimise over-fitting, some sensible priors were applied: The rotation period was constrained to $\sim\mathcal{U}(50,300)$, covering both the presumed rotation period and its one-year alias, and a Normal prior centred on the data's RMS of $\sim\mathcal{N}(3.3,3.3)$ was imposed on its amplitude to prevent the GP from absorbing all observed variability. The most revealing GP models used a rotation kernel ($\mathcal{GP}_{\rm rot}$, defined in Eq. 11), which resulted in a fit $P_{\rm rot}$ of 168 d (see Table D.4 for parameters comparison). Adding a single Keplerian models the 250 d period, whereas adding a magnetic cycle model instead (defined in Eq. 9) models a $P_{\rm mag}$ of 506 d, whilst reducing $P_{\rm rot}$ to 155 d. A tighter prior on the long cycle period, $\sim\mathcal{U}(800,5000)$, produced a $P_{\rm mag}$ of 809 d. Inspecting the posterior shows a boundary solution at a truncated peak, likely the previous 506 d solution. An even tighter prior of $\sim\mathcal{U}(2000,5000)$ resulted in $P_{\rm rot}\sim$ 250 d, and a flat posterior for most of the magnetic cycle parameters, including $P_{\rm mag}$. None of these variants of $P_{\rm rot}$+Mag conveyed a clear peak at ∼ 3200 days.

The $\mathcal{GP}_{\rm rot}(K_0)$ model has an evidence equal to −215.86±0.01, and the inclusion of an additional Keplerian (K_1) provides a marginally worse evidence ($\Delta\ln\mathcal{Z}$=−0.34); so $\mathcal{GP}_{\rm rot}(K_0)$ is selected as the best model. The proposed value of $P_{\rm rot}$ is detected as

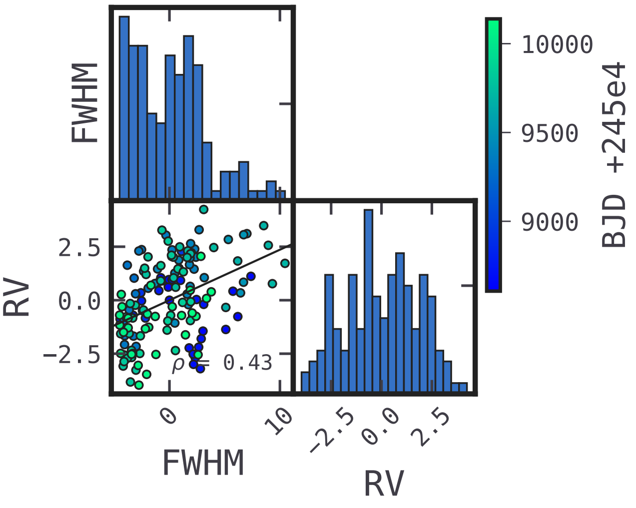

Fig. 10. Barnard's star correlogram for E18 and E19 data. Displays the Pearson Correlation coefficient (ρ) between RVs and FWHM. Diagonal displays the samples distribution. RV presents significant correlation with FWHM ρ = 0.43.

$P_{\rm rot}$=$170.95^{+22.01}_{-1.98}$ d, but the long-period cycle was not detected in any model featuring Uniform priors, likely due the comparatively short baseline of the ESPRESSO data (1533 d).

A GP model variation with $\mathcal{GP}_{\rm grot}$ (see Eq. 12), similar to the approach of GH24, was also tested. Although it found $P_{\rm rot}$=174.2, it was ruled out based on evidence comparison.

The inclusion of HARPS and HAN data presents consistent results with those of ESPRESSO on its own (see Table D.2), presenting the best $\ln\mathcal{Z}$ and BIC values.

Next up, the RV data was analysed, informed by the previous FWHM results. The base model includes two signals, one for stellar rotation and one for the magnetic cycle. For stellar rotation, the period boundaries were mildly constrained to $\sim\mathcal{U}(50,300)$ once again, while for the magnetic cycle, elusive to the previous FWHM analysis, the priors from Toledo-Padrón et al. (2019) were adopted, $\sim\mathcal{N}(3250,300)$, matching the priors used by GH24.

Amongst the tested models were: $\mathcal{GP}_{\rm 3SHO}$–two SHO kernels for the rotation, and one for the magnetic cycle, $\mathcal{GP}_{\rm grot}$+$\mathcal{S}$–a rotation kernel (see Eq. 11) and a simple sinusoid to describe the magnetic cycle (see Eq. 8), $\mathcal{GP}_{\rm rot+SHO}$–a rotation kernel and a SHO kernel to describe the magnetic cycle), and $\mathcal{GP}_{\rm 2rot}$–two rotation kernels, one for each period, and $\mathcal{GP}_{\rm grot}$+Mag–a variation in the rotation kernel (described in Eq. 12) and a double sinusoid (Eq. 9), akin to the model used in GH24. After the initial evaluation, `EMPEROR` added a Keplerian with period $\sim\mathcal{U}(0.5,50)$.

All models found signals consistent with $P_{\rm rot}$=140 d, $P_{\rm mag}$=3200 d and $P_{\rm Kep}$=3.15 d (see Table D.5). Based on the statistics (see Table D.3), $\mathcal{GP}_{\rm rot}$+$\mathcal{S}$, $\mathcal{GP}_{\rm rot}$+Mag, $\mathcal{GP}_{\rm grot}$+$\mathcal{S}$, and $\mathcal{GP}_{\rm grot}$+Mag had similar statistics and parameters, and $\mathcal{GP}_{\rm grot}$+$\mathcal{S}$ is selected as the best model. This model returned a marginally higher evidence over $\mathcal{GP}_{\rm grot}$+Mag $\Delta\mathcal{Z}$=0.69, and the reduced complexity of using two fewer parameters argues in favour of this model. In addition, the $\Delta(\ln\mathcal{P}\text{-}\ln\mathcal{Z})$ is of 3.15 between the two models, remarking the uniqueness of the $\mathcal{GP}_{\rm grot}$+$\mathcal{S}$ solution. The phase-folded model for the Keplerian is illustrated in Fig. 11.

A re-run of this model was performed, with a $\sim\mathcal{U}(800,5000)$ prior instead of $\sim\mathcal{N}(3250,300)$ for $P_{\rm mag}$, to test its dependence on the Normal prior. The dence on the Normal prior. The resulting parameters were parameters were $P_{\rm rot}$=$140.67^{+7.60}_{-9.61}$ d, $P_{\rm mag}$=2486^{+401}_{-19} d, and $P_{\rm Kep}$=$3.15^{+0.98}_{-1.70}$ d. Although this magnetic cycle is shorter than the ∼3200 d presumed signal, its upper bound is poorly constrained by the limited temporal baseline. Crucially, the Keplerian signal remains at 3.15 d. These results align with GH24 (see Table 3).

Table 3. Planet parameters of GJ 699 b.

Parameter	This work	GH24
$P_{\rm Kep}$ (days)	3.1536± 0.0003	3.1533 ± 0.0006
K_p (ms^{-1})	0.56 ± 0.03	0.55 ± 0.07
m_p sin$I$ (M$_\oplus$)	0.38 ± 0.03	0.37 ± 0.05
a_p (AU)	0.02315 ±0.00039	0.02294 ± 0.00033
e_p	<0.04	<0.16

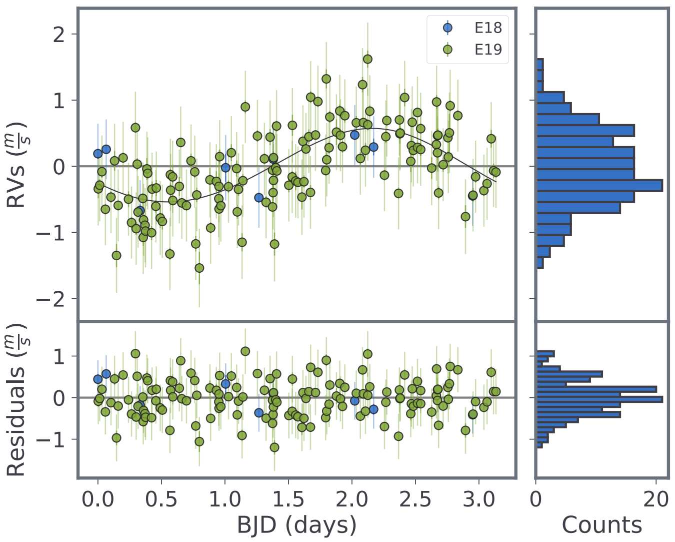

Fig. 11. Barnard's star $\mathcal{GP}_{\rm grot}$+$\mathcal{S}$ model. Top: RVs phase-folded to P = 3.1533 d with the best-fit model (black line). Marker colours differentiate datasets, blue (E18), and green (E19). Bottom: RV residuals. Right: histograms of observations and residuals.

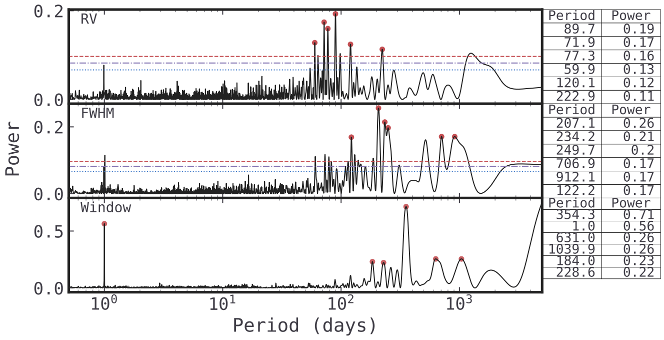

Fig. 12. Barnard's star periodogram for the combined ESPRESSO, HARPS, and HARPS-N data. Descending, RVs, FWHM, and the window function. FAP lines included for 10, 1 and 0.1%, in dashed red, dash-dotted purple, dotted blue, respectively. Circle markers show the 10 periods with the greatest power, coloured by FAP region.

5. Discussion

Bayesian inference has become a fundamental tool in exoplanet detection to push the detectability limits in RV data. Several challenges arise: the highly multi-modal nature of the parameter space, correlations between parameters, data sparsity, low SNR, contamination due to stellar activity, and other complicating factors. `EMPEROR` addresses these issues by employing an original APT MCMC sampler, `reddemcee`, as well as an ample selection of noise models. As shown in the benchmarks with 51 Peg (see §4.1), the increase in effective computation speed due to the sampling algorithm is almost two-fold compared to `dynesty`'s DNS. And the increase due to the `EMPEROR` environment is at least four-fold, compared to a different RV fitter software using the same sampler. This directly translates to a plethora of models getting tested in an equivalent time with other tools. `EMPEROR` also provides the necessary apparatus to make model building (and model comparison) easy and straightforward.

While nested sampling excels in estimating Bayesian evidence (ln $\mathcal{Z}$), the APT approach excels in thoroughly exploring the parameter space, which–in the presence of complex noise structures–allows the algorithm to retrieve precise exoplanetary parameters. This trade-off highlights the importance of selecting the appropriate method based on the specific goals of an analysis; whether to prioritise robust evidence estimation or refining parameter constraints. It is important to bear in mind that dynamic nested sampling tends to underestimate the evidence's error, while thermodynamic integration tends to overestimate it. In the case of 51 Peg, `EMPEROR` successfully retrieved planetary parameters with high precision, APT outperforming NS (see Table B.1).

Statistical methods alone are insufficient for a robust exoplanet detection, as they struggle to distinguish between planetary and stellar activity. Effective model selection requires the integration of domain knowledge, heuristics, and multiple statistical indicators instead of just one. Furthermore, an over-reliance on the Bayesian evidence might hinder the unveiling of the system. As observed in the HD 55693 analysis (see §4.2), a completely uninformed approach to the system could lead to the presence of one or two planets (the WN K_2 model presenting a higher evidence than its predecessors, the WN K_0 and WN K_1). By introducing stellar activity analysis, the first signal (P∼2500 d) becomes suspect of being activity-related, and not Keplerian at all. The same applies for secondary signals at P∼259, 235, 193, and 26 d. By applying some domain knowledge on stars for this spectral type and age (Mamajek & Hillenbrand 2008; Lovis et al. 2011), it becomes reasonable to expect a magnetic cycle period between 2200-3300 d, and a rotation period around ∼30 d.

By trying to model the activity as both rotation and a magnetic cycle, but without normal priors to guide our sampling, with different noise models, several interesting points arise: 1) The MA model effectively removed the rotation signal, fitting exclusively the magnetic cycle, whilst reducing its amplitude compared to other solutions. 2) The SA model managed to remove the rotation as well, but as a single Keplerian was added, the correlation coefficient parameter dropped enough to make room for this signal. Adding a prior on the coefficient severely reduces the degeneracy induced by the Keplerian, effectively removing both the magnetic cycle and the rotation. 3) With a GP for the rotation, a solution with the best evidence is obtained. Nevertheless, the rotation period it fits P_{rot}=2.45 d is much lower than expected. A deeper look at its statistics reveals an overly low χ^2_ν=0.41, hinting over-fitting, and therefore this model should be treated with caution, or discarded.

This benchmark underscores the necessity of cautious prior selection and comparative model evaluation. A rigid reliance on evidence estimation alone may lead to misleading conclusions when prior assumptions are incomplete or introduce biases.

While Gaussian Process regression is a powerful tool for mitigating stellar noise, its application requires careful tuning. As demonstrated in the Barnard's star analysis, five different GP models arrived at the same solution for $P_{\rm rot}$, $P_{\rm mag}$, and $P_{\rm Kep}$, with similar ln $\mathcal{Z}$ values between them. Almost as a cautionary tale, the model with the best ln $\mathcal{Z}$ was the one that had the least parameters ($\mathcal{GP}_{\rm grot}$+S).

6. Conclusions

Three systems were analysed with our new `EMPEROR` code– 51 Peg, HD 55693, and Barnard's star (GJ 699). For 51 Peg §4.1, using 256 LICK RVs we reproduce the results of Butler et al. (2006) with tighter parameters' posteriors for 51 Peg b; P=4.230782$^{+0.000016}_{-0.000013}$ d, K=55.69$^{+0.18}_{-0.23}$ m/s, and e<0.01. The `EMPEROR` code outperforms another RV fitting tool (using the same sampler) by a factor of at least 4, and the `reddemcee` APT sampler is shown to outperform by almost a factor of 2 over `dynesty`'s DNS sampler.

For HD 55693, three datasets were used, TERRA1, TERRA2, and PFS, with 29, 19, and 36 observations, respectively. The data was subjected to preliminary LSP and correlogram analysis, which revealed a strong correlation between S-index and RVs (ρ = 0.80), as well as matching periods for a presumed $P_{\rm mag}$ and $P_{\rm rot}$. With uninformed priors on the periods, the best model, a Gaussian Process with a rotation kernel proves to be false, leading to an unphysically short rotation period. The SA model, composed of stellar activity linear correlations, works best with its prior informed (even modestly) by the correlations, leaving no Keplerian signals in the fit. The SAMA model, a mixture of SA with an exponentially weighted moving average found the periods $P_{\rm mag}$=2557$^{+70}_{-38}$, and $P_{\rm rot}$=29.72$^{+0.01}_{-0.02}$, consistent with the presumed true values. Although with just RV data and uninformative priors, this last model doesn't have enough evidence to supersede the non-Keplerian model. The importance of applying domain knowledge and heuristics to the model's priors, as well as the dangers of over-relying on a single metric for model comparison are also demonstrated, by performing small modifications on the priors' boundaries.

For Barnard's star, the work by (González Hernández et al. 2024) was followed, to test `EMPEROR` on recent discoveries that make use of advanced noise modelling as well as high precision RVs on low SNR data. The presumed $P_{\rm rot}$ was found on both the LSP and the `reddemcee` run of the FWHM data, but $P_{\rm mag}$ did not appear on any of them without applying a strong prior based on a CaHK analysis done previously by Toledo-Padrón et al. (2019). In the analysis of the RV data, the planet signal at $P_{\rm Kep}$ = 3.15 d was consistently recovered across models. Moreover, removing the prior for this long-term cycle found a different value for $P_{\rm mag}$=2486$^{+401}_{-19}$ d, but the Keplerian signal was still successfully retrieved $P_{\rm Kep}$=3.15$^{+0.98}_{-1.70}$ d. In this case, other peaks in the posterior were boosted enough to maintain them in the 95% HDI range used for uncertainties, but still far away from the 3.15 d peak (e.g. compared to the second-highest peak, at 4.12 d, the posterior difference $\Delta p(\boldsymbol{\theta}|D, M) \sim 7$).

The results presented show that the `EMPEROR` code is an effective and flexible tool for planet detection and characterisation using RV data. Its key advantages are its ability to explore a broad multi-modal parameter space efficiently, its modularity which allows various models to be seamlessly integrated, and its computing performance, which democratises the process of exoplanet detection. Its implementation of the APT framework significantly improves convergence rates and ensures that the global posterior maximum is consistently identified. Including Gaussian Processes, Moving Average methods, as well as different depictions of low-frequency modulations, is crucial for mitigating stellar activity contamination. The evidence-based model selection approach employed by `EMPEROR` provides a systematic way to determine the most probable model, in the least amount of time.

Future developments of the `EMPEROR` code will focus on integrating additional data types, notably photometry and astrometry (with the optimised Gaia+Hipparcos-based method developed by Feng et al. 2023), to facilitate a unified and simultaneous analysis of multiple observational channels. Further enhancements to stellar activity modelling are also anticipated, ensuring more robust discrimination between planetary signals and stellar noise. In addition, by applying `EMPEROR` to extensive RV catalogues, it will be possible to conduct population-level studies that unveil broader trends in exoplanet demographics. Through these advancements, `EMPEROR` is poised to become a more comprehensive and adaptive platform for exoplanet detection and characterisation, thereby conquering new realms in the uncharted frontiers of planetary systems.

Acknowledgements. All benchmarks were performed on a computer with an AMD Ryzen Threadripper 3990X 64-core processor with 128 GB of DDR4-3200Mhz RAM, limiting the use to 24 physical cores and 24 threads. This manuscript was written with Overleaf. We are grateful to Fabo Feng and Mikko Tuomi for stimulating early discussions on planet detection, and to María Maass and Raúl Melo for feedback on the initial draft. PAPR and JSJ gratefully acknowledge support by FONDECYT grant 1240738 and from the ANID BASAL project FB210003. We also greatly appreciate the support from the CASSACA China-Chile Joint Research Fund through grant CCJRF2205.

References

Alonso-Floriano, F. J., Morales, J. C., Caballero, J. A., et al. 2015, A&A, 577, A128
Anglada-Escudé, G., Amado, P. J., Barnes, J., et al. 2016, Natur, 536, 437
Anglada-Escude, G., Arriagada, P., Tuomi, M., et al. 2014, MNRAS, 443, L89
Anglada-Escudé, G. & Butler, R. P. 2012, ApJS, 200, 15
Anglada-Escudé, G., Tuomi, M., Gerlach, E., et al. 2013, A&A, 556, A126
Arriagada, P. 2011, ApJ, 734, 70
Baluev, R. V. 2013, MNRAS, 429, 2052
Barnes, J. R., Haswell, C. A., Staab, D., et al. 2020, NatAs, 4, 419
Boisse, I., Bouchy, F., Hébrard, G., et al. 2011, in IAU Symposium, Vol. 273, Physics of Sun and Star Spots, ed. D. Prasad Choudhary & K. G. Strassmeier, 281–285
Bonett, D. & Seier, E. 2006, Biometrical journal. Biometrische Zeitschrift, 48, 144
Butler, R. P., Wright, J. T., Marcy, G. W., et al. 2006, ApJ, 646, 505
Crane, J. D., Shectman, S. A., Butler, R. P., et al. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III, ed. I. S. McLean, S. K. Ramsay, & H. Takami, 773553
Díaz, M. R., Jenkins, J. S., Tuomi, M., et al. 2018, AJ, 155, 126
Dumusque, X., Lovis, C., Udry, S., & Santos, N. C. 2011, in IAU Symposium, Vol. 276, The Astrophysics of Planetary Systems: Formation, Structure, and Dynamical Evolution, ed. A. Sozzetti, M. G. Lattanzi, & A. P. Boss, 530–532
Dumusque, X., Pepe, F., Lovis, C., et al. 2012, Natur, 491, 207
Earl, D. J. & Deem, M. W. 2005, PCCP, 7, 3910
Espinoza, N., Kossakowski, D., & Brahm, R. 2019, MNRAS, 490, 2262
Faria, J. P., Santos, N. C., Figueira, P., & Brewer, B. J. 2018, JOSS, 3, 487
Feng, F., Butler, R. P., Vogt, S. S., Holden, B., & Rui, Y. 2023, MNRAS, 525, 607
Feng, F., Tuomi, M., & Jones, H. R. A. 2017, MNRAS, 470, 4794
Fischer, D. A., Anglada-Escude, G., Arriagada, P., et al. 2016, PASP, 128, 066001
Fong, E. & Holmes, C. 2019, arXiv, arXiv:1905.08737
Foreman-Mackey, D. 2018, RNAAS, 2, 31
Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, AJ, 154, 220
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Fulton, B. J., Petigura, E. A., Blunt, S., & Sinukoff, E. 2018, PASP, 130, 044504
Gillon, M., Triaud, A. H. M. J., Demory, B.-O., et al. 2017, Natur, 542, 456
Gladman, B. 1993, Icar, 106, 247
Goggans, P. M. & Chi, Y. 2004, in American Institute of Physics Conference Series, Vol. 707, Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. G. J. Erickson & Y. Zhai (AIP), 59–66
González Hernández, J. I., Suárez Mascareño, A., Silva, A. M., et al. 2024, A&A, 690, A79
Goodman, J. & Weare, J. 2010, CAMCS, 5, 65
Gray, R. O., Corbally, C. J., Garrison, R. F., et al. 2006, AJ, 132, 161
Gregory, P. C. 2007, MNRAS, 381, 1607
Hatzes, A. P. 2013, ApJ, 770, 133
Haywood, R. D., Collier Cameron, A., Queloz, D., et al. 2014, MNRAS, 443, 2517
Higson, E., Handley, W., Hobson, M., & Lasenby, A. 2019, S&C, 29, 891
Hou, F., Goodman, J., Hogg, D. W., Weare, J., & Schwab, C. 2012, ApJ, 745, 198
Isaacson, H. & Fischer, D. 2010, ApJ, 725, 875
Jenkins, J. S., Harrington, J., Challener, R. C., et al. 2019a, MNRAS, 487, 268
Jenkins, J. S., Jones, H. R. A., Pavlenko, Y., et al. 2008, A&A, 485, 571
Jenkins, J. S., Jones, H. R. A., Tinney, C. G., et al. 2006, MNRAS, 372, 163
Jenkins, J. S., Jones, H. R. A., Tuomi, M., et al. 2013, ApJ, 766, 67
Jenkins, J. S., Murgas, F., Rojo, P., et al. 2011, A&A, 531, A8
Jenkins, J. S., Pozuelos, F. J., Tuomi, M., et al. 2019b, MNRAS, 490, 5585
Jenkins, J. S., Yoma, N. B., Rojo, P., Mahu, R., & Wuth, J. 2014, MNRAS, 441, 2253
Jones, H. R. A., Butler, R. P., Tinney, C. G., et al. 2006, MNRAS, 369, 249
Kervella, P., Arenou, F., Mignard, F., & Thévenin, F. 2019, A&A, 623, A72
Kipping, D. M. 2013, MNRAS, 434, L51
Laskar, J. & Petit, A. 2019, in AAS/Division for Extreme Solar Systems Abstracts, Vol. 51, AAS/Division for Extreme Solar Systems Abstracts, 200.06
Levenberg, K. 1944, Quaterly Journal on Applied Mathematics, 164
Lomb, N. R. 1976, Ap&SS, 39, 447
Lotfi, S., Izmailov, P., Benton, G., Goldblum, M., & Wilson, A. G. 2022, arXiv, arXiv:2202.11678
Lovis, C., Ségransan, D., Mayor, M., et al. 2011, A&A, 528, A112
Mamajek, E. E. & Hillenbrand, L. A. 2008, ApJ, 687, 1264
Marchal, C. & Bozis, G. 1982, CeMec, 26, 311
Marquardt, D. W. 1963, Journal of the Society for Industrial and Applied Mathematics, 11, 431
Mayor, M. & Queloz, D. 1995, Natur, 378, 355
Peña R., P. A. & Jenkins, J. S. 2025, arXiv, arXiv:2509.24870, submitted to A&A
Pepe, F., Mayor, M., Delabre, B., et al. 2000, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4008, Optical and IR Telescope Instrumentation and Detectors, ed. M. Iye & A. F. Moorwood, 582–592
Perryman, M. 2018, The Exoplanet Handbook (Cambridge University Press)
Petit, A. C., Laskar, J., & Boué, G. 2018, A&A, 617, A93
Psaridi, A., Bouchy, F., Lendl, M., et al. 2022, A&A, 664, A94
Queloz, D., Henry, G. W., Sivan, J. P., et al. 2001, A&A, 379, 279
Rajpaul, V., Aigrain, S., Osborne, M. A., Reece, S., & Roberts, S. 2015, MNRAS, 452, 2269
Rappaport, S., Levine, A., Chiang, E., et al. 2012, ApJ, 752, 1
Rasmussen, C. E. & Williams, C. K. I. 2006, Gaussian Processes for Machine Learning (The MIT Press)
Robert, C. 2007, The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation, Springer Texts in Statistics (Springer New York)
Roberts, Jr., L. C., Turner, N. H., ten Brummelaar, T. A., Mason, B. D., & Hartkopf, W. I. 2011, AJ, 142, 175
Robertson, P., Roy, A., & Mahadevan, S. 2015, ApJL, 805, L22
Salman Khan, M., Jenkins, J. S., & Becerra Yoma, N. 2016, arXiv, arXiv:1611.00766
Salvatier, J., Wiecki, T., & Fonnesbeck, C. 2015, arXiv, arXiv:1507.08050
Scalo, J., Kaltenegger, L., Segura, A., et al. 2007, AsBio, 7, 85
Scargle, J. D. 1982, ApJ, 263, 835
Shallue, C. J. & Vanderburg, A. 2018, AJ, 155, 94
Shields, A. L., Ballard, S., & Johnson, J. A. 2016, PhR, 663, 1
Skilling, J. 2004, in American Institute of Physics Conference Series, Vol. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. R. Fischer, R. Preuss, & U. V. Toussaint (AIP), 395–405
Speagle, J. S. 2020, MNRAS, 493, 3132
Suárez Mascareño, A., Faria, J. P., Figueira, P., et al. 2020, A&A, 639, A77
The Theano Development Team, Al-Rfou, R., Alain, G., et al. 2016, arXiv, arXiv:1605.02688
Toledo-Padrón, B., González Hernández, J. I., Rodríguez-López, C., et al. 2019, MNRAS, 488, 5145
Trifonov, T. 2019, The Exo-Striker: Transit and radial velocity interactive fitting tool for orbital analysis and N-body simulations, Astrophysics Source Code Library, record ascl:1906.004
Tuomi, M. & Anglada-Escudé, G. 2013, A&A, 556, A111
Tuomi, M., Anglada-Escudé, G., Gerlach, E., et al. 2013a, A&A, 549, A48
Tuomi, M., Jones, H. R. A., Jenkins, J. S., et al. 2013b, A&A, 551, A79
Vines, J. I., Jenkins, J. S., Berdiñas, Z., et al. 2023, MNRAS, 518, 2627
Vogt, S. S., Butler, R. P., Rivera, E. J., et al. 2010, ApJ, 723, 954
Vousden, W. D., Farr, W. M., & Mandel, I. 2016, MNRAS, 455, 1919
Wittenmyer, R. A., Jones, M. I., Zhao, J., et al. 2017, AJ, 153, 51
Wright, J. T. 2005, PASP, 117, 657
Xie, W., Lewis, P., Fan, Y., Kuo, L., & Chen, M.-H. 2011, Systematic biology, 60, 150
Zechmeister, M. & Kürster, M. 2009, A&A, 496, 577

Appendix A: Available samplers

A.1. `reddemcee`

An Adaptive Parallel Tempering MCMC, based on the excellent `EMCEE` and `ptemcee` (Vousden et al. 2016). Chain tempering has been shown to be necessary to efficiently sample highly multi-modal posteriors (Gregory 2007; Tuomi et al. 2013b), where instead of sampling the posterior of the distribution, a modified posterior is sampled. The modification of the posterior is made to artificially dampen the strength of the posterior maxima, bringing it closer to the posterior noise floor. Normally the inverse temperature $\beta \in [0, 1]$ is used, such that the likelihood will look like $p(\boldsymbol{\theta}|M)^{\beta}$.

The benefit from this implementation is that now the samplers at different temperatures can build proposal densities that are based on chains with other temperatures, and since the walkers in the hotter chains are less constrained, they are less likely to get stuck in regions of the posterior that are much higher than others, bringing confidence to the fact that cold chain (β=1) members have sampled the actual maximum of the posterior and have not gotten trapped in a region of high probability that is not the global maximum.

Another benefit from this method is that with multiple chains at different temperatures, one is able to approximate the Bayesian evidence, through thermodynamic integration (Earl & Deem 2005; Goggans & Chi 2004) or Stepping Stones (Xie et al. 2011).

Given that `EMPEROR` is developed primarily to do a thorough search in a wide parameter space, the use of parallel tempering is preferred and implemented as the native format, in the form of `reddemcee` (Peña R. & Jenkins 2025).

A.2. `emcee`

The MCMC Affine-Invariant Ensemble sampler `EMCEE`[5] (Foreman-Mackey et al. 2013; Hou et al. 2012).

`EMCEE` makes use of a ‘stretch move’ to allow the algorithm’s many w walkers to sample the posterior independently, yet with the collective nature of the $w - 1$ ensemble. This means that the proposal density of each walker is based on the current positions of the full $w - 1$ walker set, and not just the position of the single previous walker, common to other sampling methods like Metropolis-Hastings or Gibbs samplers. Therefore, the algorithm can produce independent samples in much shorter autocorrelation times when compared to these other samplers.

However, when testing `EMCEE` on various data sets, two big issues were found when sampling. In a lot of cases extremely long chains were necessary in order for the samplers to converge to the posterior maximum, due to the high multi-modality. Also, walkers had a tendency to get stuck in local minima, depending on the shape of the posterior. Both of these issues are addressable by constraining the phase-space around the target posterior peak, which requires, of course, knowledge of the solution before-hand.

A.3. `dynesty`

Although the parallel tempering method is highly recommended for broad searches in multi-modal phase-spaces, `dynesty` is an alternative Bayesian posterior sampling engine which uses DNS, Dynamic Nested Sampling (Higson et al. 2019), a generalisation of the Standard Nested Sampling (SNS) algorithm where the live-points (akin to MCMC walkers) vary in number to improve sampling efficiency. In SNS, there is a fixed amount of live-points and most of the computational effort is spent iterating towards posterior peaks. Dynamically varying the points used per posterior peak allows the algorithm to shrink its prior faster and more thoroughly, and to arrive at more precise estimations (more details on SNS and DNS can be found in the aforementioned papers as well as Skilling 2004).

The tests performed have shown that when using Uniform sampling the run-time is exceedingly high for wide priors, taking an exorbitant amount of time to converge. Constraining the phase-space is needed using this scheme. With that said, `dynesty` is excellent as either an alternative or complementary sampler to `reddemcee` for multi-modal posterior distributions, given that the prior volume is not unconstrained.

A.4. `PyMC3`

PyMC3 is a well known and widely used MCMC Python package. It offers many sampling algorithms, but the crown jewel is the No U-Turn Sampler NUTS (Salvatier et al. 2015). It is particularly useful for models with many continuous parameters, taking into account the posterior density gradient for steps, allowing it to meet convergence criteria extremely fast. PyMC3 in particular has several self-tuning strategies for adaptively exploring the posterior distribution, whilst being powered by `Theano` (The Theano Development Team et al. 2016) to transcode to C for improved performance.

NUTS uses a scaling matrix method, which gives a rough shape of the distribution, meaning it does not take vastly different sized steps across dimensions. This matrix is based on the sample variance obtained during the tuning phase (analogous to the burn-in phase for MCMC). As such, efficiency is lost on too differently scaled parameter searches, as well as multi-modal or non-Gaussian distributions, and based on the empirical tests done, this method is not recommended for wide searches.

---

[5] http://dan.iel.fm/emcee/current/

Appendix B: 51 Peg

Table B.1. 51 Peg b model parameters.

Parameter	`reddemcee`	dyn-long	dyn-short	Butler et al. 2006
P (d)	$4.230782^{+0.000016}_{-0.000013}$	$4.230781^{+0.000038}_{-0.000038}$	$4.230784^{+0.000014}_{-0.000014}$	$4.230785^{+0.000036}_{-0.000036}$
K (ms^{-1})	$55.69^{+0.18}_{-0.23}$	$55.66^{+0.54}_{-0.54}$	$55.71^{+0.20}_{-0.19}$	$55.94^{+0.69}_{-0.69}$
T_0 (d)	$1.55^{+0.16}_{-0.38}$	$0.24^{+1.27}_{-1.35}$	$1.56^{+0.23}_{-0.17}$	$1.51^{+0.61}_{-0.61}$
e	$0.0085^{+0.0108}_{-0.0067}$	$0.0059^{+0.0076}_{-0.0043}$	$0.0102^{+0.0032}_{-0.0029}$	$0.0130^{+0.0120}_{-0.0120}$
ω (rad)	$1.07^{+0.24}_{-0.55}$	$4.51^{+1.46}_{-1.73}$	$1.08^{+0.34}_{-0.26}$	1.01
$\dot{\gamma}$ ($\mathrm{ms}^{-1}\mathrm{yr}^{-1}$)	$-1.62^{+0.09}_{-0.05}$	$-1.60^{+0.19}_{-0.19}$	$-1.60^{+0.07}_{-0.07}$	$-1.64^{+0.16}_{-0.16}$
γ (ms^{-1})	$5.41^{+0.08}_{-0.26}$	$5.32^{+0.44}_{-0.44}$	$5.31^{+0.16}_{-0.16}$	-
σ (ms^{-1})	$0.64^{+0.01}_{-0.59}$	$1.034^{+0.841}_{-0.709}$	$1.05^{+0.27}_{-0.27}$	-

Notes. From left to right, parameter estimates for `EMPEROR` with `reddemcee` , `dynesty` random-slice (dyn-rs), and the Butler et al. (2006) solution. T_0 was substracted 2450000d for readability. In Butler 2006, ω had ill defined uncertainties, due the proximity of e to 0. The parameters displayed here are from the run with the median maximum-likelihood for each method.

Appendix C: HD 55693

Table C.1. HD 55693 model statistics.

	$\ln P$-$\ln \mathcal{Z}$	$\ln \mathcal{Z}$	$\hat{\sigma}_{\mathcal{Z}}$	BIC	χ^2_ν	RMS
WN (K_0)	11.48±0.01	-235.28±0.01	0.25	474.19±0.01	1.09±0.01	3.54±0.01
WN (K_1)	23.68±0.03	-218.95±0.01	0.43	439.29±0.06	1.16±0.01	2.54±0.01
WN (K_2)	33.97±0.30	-214.29±0.10	0.09	431.53±0.45	1.22±0.16	2.11±0.01
MA (K_0)	13.84±0.10	-227.94±0.11	0.07	463.64±0.01	1.11±0.01	3.15±0.01
MA (K_1)	22.13±0.09	-213.12±0.05	0.09	439.57±0.16	1.18±0.07	2.41±0.02
MA (K_2)	33.70±0.86	-212.91±0.18	0.10	438.18±1.67	1.36±0.01	2.06±0.02
SA (K_0)	3.48±0.06	-217.79±0.06	0.09	438.89±0.07	1.16±0.04	2.73±0.01
SA (K_1)	25.99±0.12	-213.49±0.08	0.08	437.04±0.17	1.23±0.05	2.27±0.01
SA (K_2)	35.54±1.51	-213.29±0.04	0.09	439.69±3.01	1.05±0.22	1.99±0.08
SAMA (K_0)	-1.23±0.22	-215.74±0.20	0.10	440.54±0.28	1.22±0.06	2.60±0.01
SAMA (K_1)	32.22±0.05	-213.38±0.01	0.09	440.59±1.66	1.27±0.02	2.12±0.15
SAMA (K_2)	41.57±0.28	-212.03±0.07	0.11	433.97±0.54	1.27±0.04	1.85±0.01
$\mathcal{GP}_{\rm rot}(K_0)$	28.01±0.01	-235.17±0.01	0.22	463.06±0.02	1.09±0.011	2.68±0.01
$\mathcal{GP}_{\rm rot}(K_1)$	31.10±0.59	-211.76±0.12	0.10	432.21±1.16	0.40±0.074	0.81±0.05
$\mathcal{GP}_{\rm rot+SHO}(K_0)$	27.55±0.13	-214.70±0.12	0.18	436.34±0.13	0.41±0.010	0.81±0.04

Notes. Runs from different models, (K_n) denotes the number of Keplerian signals. In descending order: white noise only (WN), exponentially weighted moving average (MA), stellar-activity linearly correlated (SA), stellar-activity plus moving average (SAMA), Gaussian Process with a rotation kernel ($\mathcal{GP}_{\rm rot}$), and GP with a rotation kernel plus a simple harmonic oscillator ($\mathcal{GP}_{\rm rot+SHO}$).

Table C.2. HD 55693 WN vs SA model statistics.

	WN		SA	
Statistic	K_1	$\mathcal{U}(-1,-1)$	$\mathcal{U}(\rho\text{-}\sigma_\rho,\, \rho\text{+}\sigma_\rho)$	$\mathcal{N}(\rho, \sigma_\rho)$
$\ln P$-$\ln \mathcal{Z}$	23.68±0.03	3.48±0.06	4.75±0.03	4.52±0.04
$\ln \mathcal{Z}$	-218.95±0.01	-217.79±0.06	-213.20±0.01	-213.73±0.01
BIC	439.29±0.06	438.89±0.07	438.76±0.04	438.83±0.05
χ^2_ν	1.16±0.01	1.16±0.04	1.12±0.02	1.14±0.03
RMSE	2.54±0.01	2.73±0.01	2.73±0.01	2.73±0.01

Notes. From left to right, WN(K_1) model, SA(K_0) model with different ρ priors, $\sim\mathcal{U}(-1, 1)$, $\sim\mathcal{U}(\rho - \sigma_\rho, \rho + \sigma_\rho)$, and $\sim\mathcal{N}(\rho, \sigma_\rho)$. Coefficient values are: ρ_1=0.80, $\sigma_{\rho 1}$=0.07; ρ_2=−0.01, $\sigma_{\rho 2}$=0.20; ρ_3=0.56, $\sigma_{\rho 3}$=0.13.

Table C.3. HD 55693 model parameters estimation.

Parameter	Prior	**WN** (K_1)	WN (K_2)	**MA** (K_1)	MA (K_2)	SA (K_0)	SAMA (K_2)	$\mathcal{GP}_{\mathrm{rot}}(K_1)$
Acceleration, Offsets and Jitter								
γ_{T1} (ms^{-1})	$\mathcal{U}$(-8.5, 8.5)	$-1.17^{+0.5}_{-0.15}$	$-1.95^{+1.36}_{-0.68}$	$-1.46^{+0.79}_{-0.7}$	$-1.32^{+0.47}_{-0.22}$	$-1.2^{+0.29}_{-0.16}$	$-1.38^{+0.19}_{-0.42}$	$-1.55^{+0.63}_{-0.64}$
γ_{PFS} (ms^{-1})	$\mathcal{U}$(-6.7, 6.7)	$0.71^{+0.37}_{-0.14}$	$1.21^{+0.56}_{-0.66}$	$0.4^{+0.79}_{-0.61}$	$1.39^{+0.59}_{-0.77}$	$-0.12^{+0.28}_{-0.27}$	$1.26^{+0.58}_{-0.72}$	$0.75^{+0.31}_{-0.3}$
γ_{T2} (ms^{-1})	$\mathcal{U}$(-5.7, 5.7)	$0.11^{+1.11}_{-1.44}$	$0.14^{+1.17}_{-1.64}$	$0.64^{+1.12}_{-2.06}$	$0.98^{+1.31}_{-2.52}$	$0.3^{+0.16}_{-0.27}$	$-0.55^{+0.22}_{-0.62}$	$0.23^{+1.19}_{-1.34}$
σ_{T1} (ms^{-1})	$\mathcal{N}$(1, 5)	$2.01^{+0.37}_{-0.41}$	$2.42^{+0.42}_{-0.38}$	$1.79^{+0.72}_{-0.41}$	$2.15^{+0.35}_{-0.02}$	$2.37^{+0.23}_{-0.1}$	$1.63^{+0.52}_{-0.42}$	$0.16^{+0.04}_{-0.16}$
σ_{PFS} (ms^{-1})	$\mathcal{N}$(1, 5)	$2.67^{+0.08}_{-0.34}$	$1.56^{+0.96}_{-0.49}$	$2.51^{+0.09}_{-0.27}$	$1.27^{+1.06}_{-0.48}$	$3.05^{+0.22}_{-0.18}$	$1.53^{+0.79}_{-0.47}$	$1.02^{+0.14}_{-0.43}$
σ_{T2} (ms^{-1})	$\mathcal{N}$(1, 5)	$2.32^{+0.52}_{-0.52}$	$2.41^{+0.37}_{-0.1}$	$2.02^{+0.52}_{-0.47}$	$1.92^{+0.62}_{-0.5}$	$1.77^{+0.33}_{-0.39}$	$1.99^{+0.27}_{-0.11}$	$1.04^{+0.73}_{-0.02}$
Magnetic Cycle								
P_1 (days)	$\mathcal{U}$(1, 6712)	$2463.4^{+143.8}_{-96.3}$	$2523.6^{+86.7}_{-81.3}$	$2395.5^{+188.5}_{-93.5}$	$2540.1^{+43.6}_{-44.9}$	-	$2557.0^{+70.1}_{-36.7}$	$2457.7^{+95.3}_{-87.0}$
A_1 (ms^{-1})	$\mathcal{U}$(1e-6, 7.1)	$4.4^{+0.15}_{-0.38}$	$4.43^{+0.52}_{-0.56}$	$3.97^{+0.03}_{-0.19}$	$3.98^{+0.02}_{-0.24}$	-	$2.82^{+0.16}_{-0.15}$	$4.33^{+0.49}_{-0.55}$
M_{01}(rad)	$\mathcal{U}(0, 2\pi)$	$0.79^{+0.39}_{-0.79}$	$2.85^{+1.76}_{-2.04}$	$0.4^{+1.4}_{-0.4}$	$3.93^{+2.36}_{-1.84}$	-	$2.05^{+1.54}_{-0.09}$	$0.47^{+1.48}_{-0.1}$
e_1	$\mathcal{N}$(0, 0.1)	$0.2^{+0.14}_{-0.2}$	$0.14^{+0.06}_{-0.14}$	$0.17^{+0.03}_{-0.17}$	$0.11^{+0.06}_{-0.11}$	-	$0.17^{+0.03}_{-0.12}$	$0.2^{+0.0}_{-0.2}$
ω_1 (rad)	$\mathcal{U}(0, 2\pi)$	$5.73^{+0.54}_{-0.53}$	$3.77^{+2.46}_{-1.71}$	$5.9^{+0.38}_{-1.41}$	$2.86^{+3.38}_{-1.79}$	-	$4.75^{+0.11}_{-1.35}$	$6.05^{+0.23}_{-1.1}$
Stellar Rotation								
P_2 (days)	$\mathcal{U}$(1, 300)	-	$167.23^{+83.81}_{-162.33}$	-	$167.13^{+13.64}_{-59.25}$	-	$29.72^{+0.01}_{-0.02}$	-
A_2 (ms^{-1})	$\mathcal{U}$(0, 7.1)	-	$2.95^{+1.07}_{-1.63}$	-	$4.31^{+1.37}_{-3.0}$	-	$3.03^{+1.52}_{-1.57}$	-
M_{02}(rad)	$\mathcal{U}(0, 2\pi)$	-	$0.99^{+5.01}_{-0.99}$	-	$0.76^{+1.41}_{-0.66}$	-	$5.16^{+0.37}_{-0.39}$	-
e_2	$\mathcal{N}$(0, 0.1)	-	$0.74^{+0.15}_{-0.13}$	-	$0.83^{+0.16}_{-0.07}$	-	$0.67^{+0.29}_{-0.26}$	-
ω_2 (rad)	$\mathcal{U}(0, 2\pi)$	-	$3.29^{+1.38}_{-0.32}$	-	$3.74^{+0.74}_{-0.97}$	-	$2.52^{+0.2}_{-0.81}$	-
$\mathcal{GP}_{\mathrm{rot}}\rho$ (days)	$\mathcal{U}$(1, 300)	-	-	-	-	-	-	$2.45^{+0.04}_{-0.26}$
$\mathcal{GP}_{\mathrm{rot}}\sigma$ (ms^{-1})	$\mathcal{U}$(0, 7.1)	-	-	-	-	-	-	$1.16^{+1.83}_{-0.25}$
$\mathcal{GP}_{\mathrm{rot}}f$	$\mathcal{U}$(0, 10)	-	-	-	-	-	-	$0.01^{+0.99}_{-0.01}$
Noise								
MA ϕ	$\mathcal{U}$(-1, 1)	-	-	$0.58^{+0.16}_{-0.23}$	$0.41^{+0.05}_{-0.1}$	-	$0.1^{+0.2}_{-0.17}$	-
MA τ (days)	$\mathcal{U}$(1, 300)	-	-	$7.68^{+16.18}_{-4.91}$	$27.88^{+7.65}_{-25.81}$	-	$56.33^{+2.06}_{-54.83}$	-
$\mathcal{A}_{\mathrm{T1}}$	$\mathcal{U}$(-1, 1)	-	-	-	-	$0.81^{+0.07}_{-0.04}$	$0.41^{+0.05}_{-0.06}$	-
$\mathcal{A}_{\mathrm{PFS}}$	$\mathcal{U}$(-1, 1)	-	-	-	-	$-0.09^{+0.06}_{-0.06}$	$0.11^{+0.11}_{-0.18}$	-
$\mathcal{A}_{\mathrm{T2}}$	$\mathcal{U}$(-1, 1)	-	-	-	-	$0.58^{+0.04}_{-0.07}$	$0.48^{+0.16}_{-0.15}$	-

Notes. Models with K_n Keplerians in parenthesis, from left to right: White noise (WN), exponentially weighted moving average (MA), linearly correlated stellar-activity (SA), moving average with stellar-activity (SAMA), and GP with rotation kernel.

Appendix D: GJ 699

Table D.1. GJ 699 FWHM ESPRESSO model statistics.

Model	$\ln P - \ln \mathcal{Z}$	$\ln \mathcal{Z}$	$\hat{\sigma}_{\mathcal{Z}}$	BIC	χ^2_ν	RMS
WN (K_1)	3.36±0.11	-354.58±0.11	0.11	705.65±0.23	1.07±0.01	2.23±0.02
WN (K_2)	10.05±0.13	-326.64±0.12	0.12	640.48±1.03	1.11±0.02	1.66±0.01
WN (K_3)	7.13±0.46	-316.01±0.44	0.13	627.35±0.34	1.16±0.01	1.45±0.01
$\mathcal{GP}_{\rm grot}(K_0)$	-10.36±0.62	-217.46±0.06	0.09	445.72±0.31	0.62±0.01	0.36±0.01
$\mathcal{GP}_{\rm rot}(K_0)$	-4.17±0.02	-215.86±0.01	0.08	446.75±0.07	0.67±0.01	0.38±0.01
$\mathcal{GP}_{\rm rot}(K_1)$	-6.61±0.09	-216.20±0.08	0.08	453.24±0.16	0.76±0.01	0.41±0.01
$\mathcal{GP}_{\rm rot}$+Mag(a)	-11.12±0.99	-219.33±0.06	0.10	449.53±1.98	0.69±0.04	0.41±0.01
$\mathcal{GP}_{\rm rot}$+Mag(b)	-18.92±0.22	-219.49±0.08	0.09	465.48±0.38	0.68±0.04	0.37±0.01

Notes. Runs from different models, (K_n) denotes the number of Keplerian signals. In descending order: white noise only (WN), Gaussian process with different parameterisations of the rotation kernel $\mathcal{GP}_{\rm rot}$ and $\mathcal{GP}_{\rm grot}$, and the inclusion of a magnetic cycle (Mag). This last model was tested with two different boundaries for the magnetic cycle: (a)~$\mathcal{U}(0.5, 5000)$, (b)~$\mathcal{U}(800, 5000)$.

Table D.2. GJ 699 FWHM ESPRESSO+HARPS model statistics.

Model	$\ln P - \ln \mathcal{Z}$	$\ln \mathcal{Z}$	$\hat{\sigma}_{\mathcal{Z}}$	BIC	χ^2_ν	RMS
$\mathcal{GP}_{\rm rot}(K_0)$	-3.95±0.11	-558.34±0.11	0.11	1142.38±0.05	0.84±0.02	1.27±0.01
$\mathcal{GP}_{\rm rot}$+Mag(a)	-12.27±0.12	-559.35±0.07	0.11	1158.34±0.37	0.83±0.02	1.26±0.01
$\mathcal{GP}_{\rm rot}$+Mag(b)	-15.51±0.43	-559.21±0.04	0.11	1164.46±0.03	0.86±0.01	1.27±0.01

Notes. Gaussian Process with just a rotation kernel $\mathcal{GP}_{\rm rot}$, and GP with rotation kernel plus a magnetic cycle with different period priors, (a)~$\mathcal{N}(3250, 300)$, and (b)~$\mathcal{U}(800, 5000)$.

Table D.3. GJ 699 RV ESPRESSO model statistics.

Model	$\ln P - \ln \mathcal{Z}$	$\ln \mathcal{Z}$	$\hat{\sigma}_{\mathcal{Z}}$	BIC	χ^2_ν	RMS
$\mathcal{GP}_{\rm 3SHO}(K_0)$	-28.29±0.01	-230.28±0.01	0.11	499.70±0.11	0.75±0.02	0.58±0.01
$\mathcal{GP}_{\rm 3SHO}(K_1)$	-13.66±0.30	-230.84±0.25	0.26	482.01±0.09	0.71±0.02	0.44±0.01
$\mathcal{GP}_{\rm rot+SHO}(K_0)$	-15.70±0.05	-233.14±0.04	0.07	495.65±0.22	0.77±0.01	0.61±0.01
$\mathcal{GP}_{\rm rot+SHO}(K_1)$	-2.38±0.54	-232.56±0.33	0.09	478.10±0.64	0.73±0.07	0.43±0.01
$\mathcal{GP}_{\rm 2rot}(K_0)$	-19.81±0.05	-232.25±0.05	0.07	505.65±0.21	0.78±0.01	0.61±0.01
$\mathcal{GP}_{\rm 2rot}(K_1)$	-7.54±0.30	-231.24±0.30	0.09	489.76±0.12	0.80±0.03	0.46±0.01
$\mathcal{GP}_{\rm rot}+S$ (K_1)	7.55±0.20	-231.21±0.18	0.11	470.94±0.25	0.72±0.03	0.46±0.01
$\mathcal{GP}_{\rm rot}$+Mag (K_1)	3.67±0.42	-230.58±0.42	0.12	480.55±0.62	0.78±0.01	0.45±0.01
$\mathcal{GP}_{\rm grot}+S$ (K_1)	4.34±0.11	-229.86±0.07	0.10	461.31±0.27	0.74±0.03	0.44±0.01
$\mathcal{GP}_{\rm grot}$+Mag (K_1)	1.19±0.32	-230.55±0.28	0.11	472.50±0.40	0.70±0.03	0.44±0.01

Notes. Runs from different models, (K_n) denotes the number of Keplerian signals. In descending order: Gaussian process with 3 SHO kernels, rotation plus SHO kernel, 2 rotation kernels, rotation kernel plus a sinusoid, rotation kernel plus magnetic cycle, and the modified rotation kernel.

Table D.4. GJ 699 FWHM ESPRESSO parameter comparison.

Parameter	Prior	WN(K_3)	$\mathcal{GP}_{\rm grot}(K_0)$	$\mathcal{GP}_{\rm rot}(K_0)$	$\mathcal{GP}_{\rm rot}(K_1)$	$\mathcal{GP}_{\rm rot}$+Mag(a)	$\mathcal{GP}_{\rm rot}$+Mag(b)
				Offsets and Jitter			
$\gamma_{\rm E18}$	$\sim\mathcal{U}(-10.5, 10.5)$	$5.25^{+0.96}_{-1.24}$	$1.17^{+0.46}_{-1.29}$	$2.33^{+0.34}_{-1.59}$	$2.75^{+0.72}_{-1.32}$	$0.14^{+1.74}_{-0.36}$	$0.62^{+2.39}_{-2.33}$
$\gamma_{\rm E19}$	$\sim\mathcal{U}(-10.5, 10.5)$	$0.15^{+0.26}_{-0.13}$	$0.91^{+0.31}_{-0.25}$	$0.89^{+0.38}_{-0.17}$	$0.94^{+0.27}_{-0.13}$	$1.19^{+0.17}_{-0.19}$	$0.17^{+1.41}_{-1.72}$
$\sigma_{\rm E18}$	$\sim\mathcal{N}(0.5, 1)$	$0.98^{+0.34}_{-0.37}$	$0.54^{+0.15}_{-0.07}$	$0.53^{+0.13}_{-0.08}$	$0.50^{+0.18}_{-0.05}$	$0.57^{+0.16}_{-0.08}$	$0.55^{+0.14}_{-0.08}$
$\sigma_{\rm E19}$	$\sim\mathcal{N}(0.5, 1)$	$1.43^{+0.10}_{-0.09}$	$0.32^{+0.02}_{-0.03}$	$0.34^{+0.01}_{-0.03}$	$0.36^{+0.02}_{-0.03}$	$0.38^{+0.01}_{-0.04}$	$0.37^{+0.04}_{-0.06}$
P_1	$\sim\mathcal{U}(50, 300)$	$165.38^{+0.55}_{-0.11}$	-	-	$249.73^{+4.40}_{-3.36}$	-	-
K_1	$\sim\mathcal{U}(0, 6.60)$	$2.75^{+0.22}_{-0.30}$	-	-	$3.57^{+0.52}_{-0.32}$	-	-
ϕ_1	$\sim\mathcal{U}(0, 2\pi)$	$4.41^{+0.15}_{-0.26}$	-	-	$2.54^{+0.63}_{-1.12}$	-	-
e_1	$\sim\mathcal{N}(0, 0.1)$	$0.23^{+0.07}_{-0.09}$	-	-	$0.02^{+0.03}_{-0.02}$	-	-
ω_1	$\sim\mathcal{U}(0, 2\pi)$	$5.24^{+0.39}_{-0.06}$	-	-	$1.28^{+1.29}_{-0.42}$	-	-
P_2	$\sim\mathcal{U}(50, 300)$	$244.41^{+1.50}_{-1.31}$	-	-	-	-	-
K_2	$\sim\mathcal{U}(0, 6.60)$	$4.39^{+0.19}_{-0.01}$	-	-	-	-	-
ϕ_2	$\sim\mathcal{U}(0, 2\pi)$	$1.02^{+0.40}_{-0.33}$	-	-	-	-	-
e_2	$\sim\mathcal{N}(0, 0.1)$	$0.15^{+0.06}_{-0.04}$	-	-	-	-	-
ω_2	$\sim\mathcal{U}(0, 2\pi)$	$2.47^{+0.23}_{-0.29}$	-	-	-	-	-
P_3	$\sim\mathcal{U}(1, 5000)$	2384^{+2616}_{-888}	-	-	-	-	-
K_3	$\sim\mathcal{U}(0, 6.60)$	$1.58^{+0.21}_{-0.15}$	-	-	-	-	-
ϕ_3	$\sim\mathcal{U}(0, 2\pi)$	$5.13^{+1.15}_{-0.15}$	-	-	-	-	-
e_3	$\sim\mathcal{N}(0, 0.1)$	$0.05^{+0.06}_{-0.05}$	-	-	-	-	-
ω_3	$\sim\mathcal{U}(0, 2\pi)$	$4.01^{+0.64}_{-1.00}$	-	-	-	-	-
$\mathcal{GP}_{\rm grot}\rho$	$\sim\mathcal{U}(50, 300)$	-	$221.98^{+8.55}_{-10.41}$	-	-	-	-
$\mathcal{GP}_{\rm grot}\tau$	$\sim\mathcal{U}(50, 600)$	-	$28.65^{+1.28}_{-3.11}$	-	-	-	-
$\mathcal{GP}_{\rm grot}A_1$	$\sim\mathcal{N}(3.3, 3.3)$	-	$2.60^{+0.22}_{-0.61}$	-	-	-	-
$\mathcal{GP}_{\rm grot}A_2$	$\sim\mathcal{N}(3.3, 6.6)$	-	$3.70^{+0.21}_{-0.05}$	-	-	-	-
$\mathcal{GP}_{\rm rot}\rho$	$\sim\mathcal{U}(50, 300)$	-	-	$170.95^{+22.01}_{-1.98}$	$151.46^{+10.27}_{-0.81}$	$155.3^{+2.44}_{-6.04}$	$156.97^{+100.99}_{-34.51}$
$\mathcal{GP}_{\rm rot}\sigma$	$\sim\mathcal{N}(3.3, 3.3)$	-	-	$3.66^{+0.28}_{-0.14}$	$2.80^{+0.63}_{-0.52}$	$2.47^{+0.63}_{-0.51}$	$3.44^{+0.99}_{-0.59}$
$\mathcal{GP}_{\rm rot}Q0$	$\sim\mathcal{U}(0, 100)$	-	-	$1.37^{+0.62}_{-0.97}$	$4.13^{+2.01}_{-3.60}$	$2.81^{+2.74}_{-1.72}$	$1.93^{+0.56}_{-1.56}$
$\mathcal{GP}_{\rm rot}\delta Q$	$\sim\mathcal{U}(0, 100)$	-	-	$0.20^{+0.13}_{-0.20}$	$0.74^{+0.59}_{-0.74}$	$3.90^{+1.06}_{-0.63}$	$0.62^{+1.34}_{-0.62}$
$\mathcal{GP}_{\rm rot}f$	$\sim\mathcal{U}(0, 1)$	-	-	$0.23^{+0.18}_{-0.08}$	$0.51^{+0.24}_{-0.28}$	$0.43^{+0.07}_{-0.18}$	$0.23^{+0.77}_{-0.23}$
$P_{\rm mag}$	$\sim\mathcal{U}(1, 5000)$	-	-	-	-	$506.21^{+4.71}_{-8.03}$	$809.26^{+4190.46}_{-9.26}$
$A1_{\rm mag}$	$\sim\mathcal{U}(0, 6.6)$	-	-	-	-	$1.07^{+0.53}_{-0.29}$	$1.47^{+1.36}_{-1.47}$
$A2_{\rm mag}$	$\sim\mathcal{U}(0, 6.6)$	-	-	-	-	$0.10^{+0.54}_{-0.10}$	$1.64^{+1.03}_{-0.57}$
$\phi1_{\rm mag}$	$\sim\mathcal{U}(0, 2\pi)$	-	-	-	-	$3.75^{+1.04}_{-0.75}$	$2.80^{+1.27}_{-2.80}$
$\phi2_{\rm mag}$	$\sim\mathcal{U}(0, 2\pi)$	-	-	-	-	$4.13^{+0.28}_{-0.27}$	$5.30^{+0.49}_{-1.50}$

Notes. Runs from different models, (K_n) denotes the number of Keplerian signals. In descending order: white noise only (WN), Gaussian process with a rotation kernel ($\mathcal{GP}_{\rm rot}$), and GP with a rotation kernel and a magnetic cycle ($\mathcal{GP}_{\rm rot}$+Mag). This last model was tested with two different priors for the magnetic cycle: (a) $\mathcal{U}(0.5, 5000)$, and (b) $\mathcal{U}(800, 5000)$.

Table D.5. GJ 699 RV parameter comparison table.

Parameter	Prior	$\mathcal{GP}_{\rm 3SHO}$	$\mathcal{GP}_{\rm rot}$S (K_1)	$\mathcal{GP}_{\rm grot}$S	$\mathcal{GP}_{\rm 2rot}$	$\mathcal{GP}_{\rm grot}$+Mag
Offsets and Jitter						
$\gamma_{\rm E18}$	$\sim\mathcal{N}(0,3)$	$-0.03^{+0.32}_{-0.45}$	$-0.39^{+0.28}_{-0.45}$	$-0.52^{+0.72}_{-0.03}$	$-0.20^{+0.55}_{-0.24}$	$-0.40^{+0.61}_{-0.05}$
$\gamma_{\rm E19}$	$\sim\mathcal{N}(0,3)$	$-1.12^{+0.58}_{-0.19}$	$-1.3^{+0.22}_{-0.29}$	$-1.53^{+0.55}_{-0.59}$	$-0.08^{+0.2}_{-0.40}$	$-1.50^{+0.27}_{-0.20}$
$\sigma_{\rm E18}$	$\sim\mathcal{LN}(0.5,1)$	$0.39^{+0.10}_{-0.06}$	$0.55^{+0.2}_{-0.18}$	$0.45^{+0.05}_{-0.11}$	$0.45^{+0.10}_{-0.05}$	$0.56^{+0.20}_{-0.20}$
$\sigma_{\rm E19}$	$\sim\mathcal{LN}(0.5,1)$	$0.53^{+0.05}_{-0.05}$	$0.55^{+0.04}_{-0.01}$	$0.54^{+0.03}_{-0.01}$	$0.55^{+0.04}_{-0.01}$	$0.56^{+0.04}_{-0.01}$
Rotation						
$\mathcal{GP}_{\rm SHO}\rho_1$	$\sim\mathcal{U}(50,300)$	$0.37^{+0.02}_{-0.35}$	-	-	-	-
$\mathcal{GP}_{\rm SHO}\sigma_1$	$\sim\mathcal{LN}(0.5,2)$	$145.63^{+29.93}_{-59.63}$	-	-	-	-
$\mathcal{GP}_{\rm SHO}\tau_1$	$\sim\mathcal{LN}(3,2)$	$19.3^{+4.15}_{-2.46}$	-	-	-	-
$\mathcal{GP}_{\rm SHO}\rho_2$	$\sim\mathcal{U}(25,150)$	$1.50^{+0.01}_{-0.18}$	-	-	-	-
$\mathcal{GP}_{\rm SHO}\sigma_2$	$\sim\mathcal{LN}(0.5,2)$	$64.74^{+5.03}_{-1.43}$	-	-	-	-
$\mathcal{GP}_{\rm SHO}\tau_2$	$\sim\mathcal{LN}(3,2)$	$27.93^{+5.33}_{-5.81}$	-	-	-	-
$\mathcal{GP}_{\rm rot}\rho$	$\sim\mathcal{U}(50,300)$	-	$124.68^{+12.14}_{-14.67}$	-	$148.27^{+2.69}_{-6.15}$	-
$\mathcal{GP}_{\rm rot}\sigma$	$\sim\mathcal{U}(0,7.1)$	-	$1.51^{+0.24}_{-0.22}$	-	$2.04^{+0.06}_{-0.21}$	-
$\mathcal{GP}_{\rm rot}Q0$	$\sim\mathcal{U}(0,50)$	-	$1.74^{+1.05}_{-0.89}$	-	$1.35^{+0.17}_{-0.55}$	-
$\mathcal{GP}_{\rm rot}\delta Q$	$\sim\mathcal{U}(0,50)$	-	$0.28^{+0.05}_{-0.28}$	-	$28.31^{+21.69}_{-14.6}$	-
$\mathcal{GP}_{\rm rot}f$	$\sim\mathcal{U}(0,1)$	-	$0.88^{+0.12}_{-0.01}$	-	$0.82^{+0.18}_{-0.18}$	-
$\mathcal{GP}_{\rm grot}\rho$	$\sim\mathcal{U}(50,300)$	-	-	$124.7^{+14.04}_{-18.01}$	-	$128.18^{+2.49}_{-4.19}$
$\mathcal{GP}_{\rm grot}\tau$	$\sim\mathcal{LN}(3,2)$	-	-	$27.82^{+5.15}_{-5.64}$	-	$50.62^{+0.11}_{-0.62}$
$\mathcal{GP}_{\rm grot}$A1	$\sim\mathcal{LN}(0.5,0.5)$	-	-	$1.30^{+0.72}_{-0.64}$	-	$2.8^{+0.99}_{-0.74}$
$\mathcal{GP}_{\rm grot}$A2	$\sim\mathcal{LN}(0.5,0.5)$	-	-	$3.46^{+1.17}_{-1.24}$	-	$6.9^{+0.17}_{-0.41}$
Magnetic Cycle						
$\mathcal{GP}_{\rm SHO}\rho_3$	$\sim\mathcal{N}(3250,300)$	$1.41^{+0.81}_{-0.02}$	-	-	-	-
$\mathcal{GP}_{\rm SHO}\sigma_3$	$\sim\mathcal{LN}(0.5,0.5)$	$3033.9^{+316.9}_{-275.8}$	-	-	-	-
$\mathcal{GP}_{\rm SHO}\tau_3$	$\sim\mathcal{U}(0,3e4)$	17369^{+12430}_{-8566}	-	-	-	-
P_m	$\sim\mathcal{N}(3250,300)$	-	$2984.0^{+291.1}_{-250.7}$	$3225.31^{+68.66}_{-176.8}$	-	$3227.5^{+35.4}_{-196.7}$
K_m	$\sim\mathcal{LN}(0.5,2)$	-	$2.7^{+0.33}_{-0.29}$	$3.04^{+0.75}_{-0.83}$	-	$2.97^{+0.17}_{-0.38}$
ϕ_m	$\sim\mathcal{U}(0,2\pi)$	-	$4.73^{+0.18}_{-0.18}$	$4.88^{+0.06}_{-0.11}$	-	$4.87^{+0.02}_{-0.11}$
K_2	$\sim\mathcal{LN}(0.5,2)$	-	-	-	-	
ϕ_2	$\sim\mathcal{U}(0,2\pi)$	-	-	-	-	
$\mathcal{GP}_{\rm rot}\rho$	$\sim\mathcal{U}(800,5000)$	-	-	-	$3186.0^{+159.9}_{-87.4}$	-
$\mathcal{GP}_{\rm rot}\sigma$	$\sim\mathcal{U}(0,7.1)$	-	-	-	$1.28^{+0.34}_{-0.23}$	-
$\mathcal{GP}_{\rm rot}Q0$	$\sim\mathcal{U}(0,50)$	-	-	-	$28.24^{+19.78}_{-14.27}$	-
$\mathcal{GP}_{\rm rot}\delta Q$	$\sim\mathcal{U}(0,50)$	-	-	-	$8.16^{+9.13}_{-8.1}$	-
$\mathcal{GP}_{\rm rot}f$	$\sim\mathcal{U}(0,1)$	-	-	-	$0.65^{+0.33}_{-0.28}$	-
Keplerian Orbit						
P_1	$\sim\mathcal{U}(0.5,50)$	3.1536^{+5e-5}_{-5e-4}	3.1529^{+7e-4}_{-5e-4}	3.15^{+6e-5}_{-5e-4}	3.1534^{+2e-4}_{-2e-4}	3.1535^{+3e-4}_{-4e-4}
K_1	$\sim\mathcal{U}(0,6.5)$	$0.59^{+0.08}_{-0.09}$	$0.56^{+0.01}_{-0.08}$	$0.56^{+0.03}_{-0.04}$	$0.55^{+0.04}_{-0.03}$	$0.57^{+0.08}_{-0.08}$
ϕ_1	$\sim\mathcal{U}(0,2\pi)$	$3.95^{+1.68}_{-2.3}$	$4.21^{+1.68}_{-2.56}$	$1.53^{+1.54}_{-0.15}$	$2.44^{+1.07}_{-0.63}$	$1.27^{+1.68}_{-0.05}$
e_1	$\sim\mathcal{N}(0,0.1)$	$0.03^{+0.01}_{-0.03}$	$0.04^{+0.05}_{-0.04}$	$0.03^{+0.01}_{-0.03}$	$0.01^{+0.02}_{-0.01}$	$0.06^{+0.05}_{-0.06}$
ω_1	$\sim\mathcal{U}(0,2\pi)$	$4.22^{+2.06}_{-0.03}$	$3.7^{+1.76}_{-2.26}$	$0.48^{+2.63}_{-0.48}$	$5.58^{+0.7}_{-0.94}$	$0.67^{+1.82}_{-0.67}$