

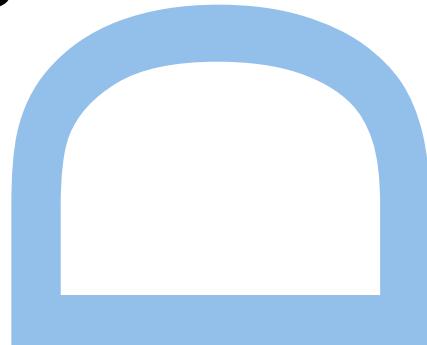
Advanced statistical data analysis methods for the detection of other Earths

João David Ribeiro Camacho

Programa Doutoral em Astronomia

Departamento de Física e Astronomia

2021

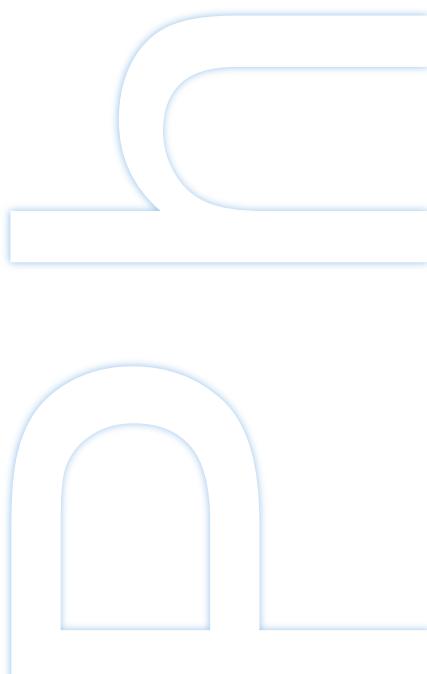


Orientador

Pedro T. P. Viana, Professor Associado, Faculdade de Ciências

Coorientador

João P. S. Faria, Investigador Auxiliar, Instituto de Astrofísica e Ciências do Espaço



Title page illustration:
Radial velocity curve of 51 Peg from Mayor & Queloz (1995)

Acknowledgements

I hate to write the acknowledgements. I never know what to write, and I am always afraid I will forget to mention someone. If I do, I am sorry! Also, I'm writing this right after taking the booster dose for the Janssen vaccine. If I forget someone, I will use that as an excuse.

These last four years were a challenge. One that I would never have completed without the support of my supervisors Pedro and João. The challenges we faced and the patience they had to help me finish my PhD is something I will never forget. Secondly, I need to thank Nuno and the rest of the Planet's team. Without them, I would have never started the PhD in the first place.

Many other people were fundamental as I did my PhD. To start, Bachar, my coffee dealer and office mate. Then there are, of course, the lunch mates that are too many to name them all. The friends I made while on my Master degree that thought I was crazy for doing a PhD. I can not also forget Sandy and Zon, the last two people that have been with me since the start of my academic life back in 2009, and Ângela, for all the messages we exchanged while she was in Colorado. For last but not least is my mother and my grandmother that believed in me since day one.

Also, I managed to finish the PhD and survive a global pandemic! I don't think many people can brag about that.

Abstract

The discovery of the first extra-solar planet was a breakthrough that revolutionized our understanding of the Universe. The increasing number of planets discovered combined with the growing instrumental precision allows us to be closer to the new milestone of finding an Earth-like planet around a Sun-like star. As we approach this discovery, new challenges arise and with it the need for ever more powerful data analysis methods.

For this thesis, we present a new Gaussian process framework to improve radial velocity measurements analysis. This new framework, known as Gaussian process regression network, combines the stellar activity information on auxiliary datasets to improve the identification and removal of these signals from the respective radial velocity time series. The novel concept of a Gaussian process regression network lies in its non-stationary setup, ideal for modelling stellar activity signals.

We describe all the mathematical foundations necessary to implement a Gaussian process regression network and understand its differences from other frameworks in use. To facilitate its use, we created a python package that we employed to analyze radial velocity observations of the Sun. The results allowed us to comprehend the complexity and constraints our work possesses. It showed that the Gaussian process regression network, in its current form, has limitations not exhibited on a traditional Gaussian process regression. Notwithstanding, it allowed to identify several steps required to improve the capabilities of the regression network.

We also applied this new framework to four stars possessing different stellar activity levels. Done for the EXPRES Stellar-Signals Project, it allowed us to determine that the Gaussian process regression network shown to be sensitive to

the stellar activity level presenting better results on more active stars. These results were also crucial to conclude that, even considering its limitations, the regression network obtains results similar to other time series modelling methods.

Resumo

A descoberta do primeiro planeta extra-solar foi um evento que revolucionou a nossa compreensão do Universo. O número crescente de planetas descobertos em conjunto com a crescente precisão instrumental permite-nos aproximar do objectivo de encontrar planetas semelhantes à Terra em órbita de uma estrela semelhante ao Sol. À medida que nos aproximamos desta descoberta, novos desafios surgem e com eles a necessidade usar métodos de análise de dados cada vez mais complexos.

Com esta tese, apresentamos um novo método com processo Gaussianos para melhorar a análise de medições de velocidade radial. Este novo método, conhecido como Rede de Regressão com Processos Gaussianos, combina as informações da atividade estelar existente de dados auxiliares para melhorar a identificação e remoção desses mesmos sinais das respetivas séries temporais de velocidade radial. O novo conceito por trás de uma rede de regressão com processos Gaussianos consiste na sua configuração não estacionária, ideal para modelar sinais de atividade estelar.

Também descrevemos todos os conceitos matemáticos necessários para implementar esta rede de regressão com processos Gaussianos e compreender as diferenças em relação a outros métodos com processos Gaussianos. Para facilitar o seu uso, criamos um repositório em python que criamos para analisar observações da velocidade radial do Sol. Os resultados permitiram-nos compreender a complexidade e limitações que o nosso trabalho possui. Estes mostraram que a rede de regressão de processos Gaussianos, na sua forma atual, tem limitações não existentes quando é feita uma simples regressão com processos Gaussianos. Não obstante, isto permitiu-nos identificar várias etapas necessárias para melhorar as

viii Advanced statistical data analysis methods for the detection of other Earths

capacidades desta nova rede de regressão.

Também aplicamos esta rede de regressão a quatro estrelas com diferentes níveis de atividade estelar. Este trabalho feito para o EXPRES Stellar-Signals Project, permitiu-nos determinar que a rede de regressão com processo Gaussiano é sensível ao nível de atividade estelar, apresentando melhores resultados em estrelas mais ativas. Estes resultados também foram fundamentais para concluir que, mesmo considerando as suas limitações, esta rede de regressão obtém resultados semelhantes a outros métodos de modelação de séries temporais.

Contents

1	Introduction	21
1.1	Exoplanets search overview	21
1.1.1	Detection methods	24
1.1.2	Milestones	31
1.1.3	Statistical analysis	32
1.2	This thesis	36
2	Doppler spectroscopy	37
2.1	Mathematical background	37
2.1.1	Keplerian orbits	38
2.1.2	Orbital elements	40
2.1.3	Calculating the Keplerian function	43
2.2	Spectroscopic observations	47
2.2.1	The past, present and future of spectroscopy	47
2.2.2	Statistical and systematic errors	48
2.2.3	The cross-correlation function	49
2.3	The stellar activity problem	54
2.3.1	Stellar activity features	54
2.3.2	Stellar activity proxies	60
3	Gaussian Process Regression Network	67
3.1	Gaussian processes	67
3.1.1	The Gaussian distribution	68
3.1.2	The Gaussian process	69
3.1.3	The Covariance function	74
3.2	Gaussian processes frameworks	85

3.2.1	Rajpaul et al. (2015) framework	85
3.2.2	Jones et al. (2017) framework	86
3.2.3	artGPN framework	87
3.3	Gaussian processes regression networks	90
3.3.1	Variational inference	91
3.3.2	Simulating a quasi-periodic GP with the GPRN	98
3.3.3	Using a GPRN on simulated data	101
4	Application to radial velocity observations	105
4.1	Priors definition	105
4.1.1	Individual analysis	106
4.1.2	Combined analysis	108
4.2	Analysis of Sun-as-a-star observations	110
4.2.1	Sun observations	110
4.2.2	Results obtained on the solar data	116
4.2.3	Interpretation of the results	142
4.3	EXPRES data	149
4.3.1	HD 10700	149
4.3.2	HD 34411	152
4.3.3	HD 101501	154
4.3.4	HD 26965	156
4.3.5	Final thoughts on the EXPRES results	158
5	Conclusions	159
5.1	Future work	161
Appendices		163
A	Sampled datasets	165
A.1	Priors	166
A.2	Scatterplot matrices	167
B	Datasets summary	171
C	Scatterplot matrices	173
C.1	Parameters	174

C.2 Plots	175
D Other EXPRES results	191

List of Figures

1.1	Mass-period distribution of all detected exoplanets.	23
1.2	Orbital motion of a star and a planet around a common center of mass, and the blue and red shifts on its light caused by orbit of the star.	25
1.3	Transit of a planet in front of a star and respective flux variation ΔF by it. The transit time duration is represented as t_T and the time of totality as t_F	27
1.4	The brown dwarf 2M1207 (centre) and its planetary companion (bottom left). Source: ESO	28
1.5	Gravitational microlensing amplified by an extrasolar planet.	31
1.6	Left: Hot-Neptune desert derived by Mazeh et al. (2016) with the dash-dotted lines delimiting the boundaries of this desert. Right: Radius gap from Fulton et al. (2017) where the gap around 1.8 Earth radii creates a bi-modal distribution of planets around Sun-like stars.	35
2.1	The two-body problem scheme.	39
2.2	Elliptical orbit of a celestial body in three dimensions. Source: Wikimedia Commons.	41
2.3	ESPRESSO's first light spectrum. Source: ESO/ESPRESSO team. .	49
2.4	Simulated structure of a CCF.	51
2.5	Example of the equivalent width of a spectral line.	53
2.6	Solar granulation over an area of 19000 squared kilometers. Source: Big Bear Solar Observatory/New Jersey Institute of Technology. .	56
2.7	A Sunspot capture by the Inouye Solar Telescope on January 28, 2020. Source: NSO/NSF/AURA.	58

2.8	Monthly mean of the number of sunspot observed since 1749. Source: https://wwwbis.sidc.be/silso	60
2.9	Simplified illustration of the impact of a spot on the CCF.	61
2.10	Expected behavior of the FWHM measurements on the example I gave on figure 2.9. On the x-axis are represented the five panes.	62
3.1	Three 1-dimensional Gaussian distributions. At solid blue a dis- tribution with $\mu = 0$ and $\sigma = 1$, at dotted red one with $\mu = 2$ and $\sigma = 0.5$, and at checkered green one with $\mu = -1$ and $\sigma = 2$.	69
3.2	Left: Three samples drawn from a Gaussian process prior. Right: Three samples drawn from the conditional posterior distribution after two function values have been observed.	70
3.3	Samples of a squared-exponential kernel with different length- scales. Solid blue $\eta_2 = 1$, dashed red $\eta_2 = 10$, and dotted green $\eta_2 = 10$.	76
3.4	Samples of a Matérn 5/2 kernel with different length-scales. Solid blue $\eta_2 = 1$, dashed red $\eta_2 = 10$, and dotted green $\eta_2 = 10$.	78
3.5	Samples of a periodic kernel with $\eta_3 = 1$ but varying η_4 . Solid blue $\eta_4 = 0.1$, dashed red $\eta_4 = 1$, and dotted green $\eta_4 = 10$.	81
3.6	Scheme of our framework	88
3.7	Example of a GPRN diagram	92
3.8	Comparison of the probability density function of a Gaussian and a Bessel distribution. Both with a mean of 0 and a standard deviation of 1.	98
3.9	Samples of a GP defined with a quasi-periodic kernel.	99
3.10	Samples of a GPRN defined with node with a periodic GP and a weight with a squared-exponential GP.	100
3.11	Samples of a GPRN defined with node with a quasi-periodic GP and a weight with a squared-exponential GP.	100

3.12 The black markers in each panel represent the data sampled (node, weight, or time series) from the three GPRNs. Each GPRN uses a different colour for the obtained fits. On the top, drawn with blue, are the data points and fits for node, weight, and final time series for the 1st model, on the middle with red, is the data and resulting fits for the 2nd model, and on the bottom plotted with green, is the 3rd model.	103
4.1 RV, BIS, FWHM, and $\log R'_{hk}$ measurements of the Sun obtained after the nightly binning. Below each time series is its respective GLS periodograms after removing a long-term trend from the data. The vertical solid lines indicate a period of 27 days. The horizontal dashed line represents the false alarm probability (FAP) level of 1%. These measurements are available at https://github.com/jdavidrcamacho/datasets	112
4.2 Daily number of sunspot observed from the start of 2015 til the end of Solar cycle 24 on late 2019 (e.g. McIntosh et al., 2014; Nandy et al., 2020). Source: https://wwwbis.sidc.be/silso	114
4.3 Discrete cross-correlation functions between the RV and the BIS (dash-dotted blue), FWHM (dashed red), and $\log R'_{hk}$ (dotted green).	115
4.4 Fits and residuals obtained using the MAP values from the GPRN (blue) and the GP (red) on the RV measurements.	119
4.5 Fits and residuals obtained using the MAP values from the GPRN (blue) and the GP (red) on the BIS measurements.	122
4.6 Fits and residuals obtained using the MAP values from the GPRN (blue) and the GP (red) on the FWHM measurements.	125
4.7 Fits and residuals obtained using the MAP values from the GPRN (blue) and the GP (red) on the $\log R'_{hk}$ measurements.	128
4.8 Fits and respective residuals obtained by the GPRN framework on the RV and the BIS time series.	132
4.9 Fits and respective residuals obtained by the artGPN framework on the RV and the BIS time series.	133
4.10 Fits and respective residuals obtained by the GPRN framework on the RV and the FWHM time series.	136

4.11	Fits and respective residuals obtained by the artGPN framework on the RV and the FWHM time series.	137
4.12	Fits and respective residuals obtained by the GPRN framework on the RV and the $\log R'_{hk}$ time series.	140
4.13	Fits and respective residuals obtained by the artGPN framework on the RV and the $\log R'_{hk}$ time series.	141
4.14	Left: Plots of the weight and node using the respective MAP values from the RV analysis. Middle: Product on the weight and node. Right: GPRN fit on the RV data points with the obtained parameters.	143
4.15	Top: ELBO and $\log \mathcal{Z}$ values given by the variational inference and DYNESTY, respectively. Bottom: Difference between the $\log \mathcal{Z}$ and the ELBO. The average value of the difference, of approximately 1.852, is plotted by the dashed horizontal line.	144
4.16	Left: Structure of each component of the GPRN. Middle: Product of the node by each respective weight. Right: GPRN fit to the RV and FWHM when using the MAP values obtained on the analysis.	146
4.17	Left: Superimposed normalized predictives from GP and the GPRN. Right: Discrete CCF between the predictives of the RV and the activity indicator.	147
4.18	Left: RV observations of HD 10700. Right: Periodogram of the data after subtracting a long term trend. Also on it is marked, as a red vertical line, the expected rotation period of 34.5 days, and the FAP of 1% as a dashed horizontal line.	151
4.19	Left: Components of the GPRN. The weight_{RV} connects the node to the RV, while the weight_{FWHM} connects it to the FWHM. Right: Final fits of the GPRN to the RV and the FWHM.	151
4.20	Left: RV observations of HD 34411. Right: Periodogram of the data after subtracting a long term trend. Also on it is marked the FAP of 1% as a dashed horizontal line. No rotation period is known for this star.	153
4.21	Left: Contribution of the node and weights to the final fit. The weight_{RV} connects the node to the RV, while the weight_{FWHM} connects it to the FWHM. Right: GPRN fit to the measurements of HD 34411.	153

4.22	Left: RV observations of HD 101501. Right: Periodogram of the data after subtracting a long term trend. Also on it is marked, as a red vertical line, the expected rotation period of 17.1 days, and the FAP of 1% as a dashed horizontal line.	155
4.23	Left: Contribution of the node and weights to the final fit. The weight_{RV} connects the node to the RV, while the weight_{FWHM} connects it to the FWHM. Right: GPRN fit to the measurements of HD 101501.	155
4.24	Left: RV observations of HD 26965. Right: Periodogram of the data after subtracting a long term trend. Also on it is marked, as a red vertical line, the expected rotation period of 40 days, and the FAP of 1% as a dashed horizontal line.	157
4.25	Left: Contribution of the node and weights to the final fit. The weight_{RV} connects the node to the RV, while the weight_{FWHM} connects it to the FWHM. Right: GPRN fit to the measurements of HD 26965.	157
A.1	Posterior distributions of the parameter from the 1st GPRN model.	168
A.2	Posterior distributions of the parameter from the 2nd GPRN model.	169
A.3	Posterior distributions of the parameter from the 3rd GPRN model.	170
C.1	Posterior distributions of the parameter of the GP used in the individual analysis of the Sun's RV measurements.	176
C.2	Posterior distributions of the parameter of the GPRN used in the individual analysis of the Sun's RV measurements.	177
C.3	Posterior distributions of the parameter of the GP used in the individual analysis of the Sun's BIS measurements.	178
C.4	Posterior distributions of the parameter of the GPRN used in the individual analysis of the Sun's BIS measurements.	179
C.5	Posterior distributions of the parameter of the GP used in the individual analysis of the Sun's FWHM measurements.	180
C.6	Posterior distributions of the parameter of the GPRN used in the individual analysis of the Sun's FWHM measurements.	181
C.7	Posterior distributions of the parameter of the GP used in the individual analysis of the Sun's $\log R'_{hk}$ measurements.	182

C.8	Posterior distributions of the parameter of the GPRN used in the individual analysis of the Sun's $\log R'_{hk}$ measurements.	183
C.9	Posterior distributions of the parameter of the GP used in the combined analysis of the Sun's RV and BIS measurements.	184
C.10	Posterior distributions of the parameter of the GPRN used in the combined analysis of the Sun's RV and BIS measurements.	185
C.11	Posterior distributions of the parameter of the GP used in the combined analysis of the Sun's RV and FWHM measurements.	186
C.12	Posterior distributions of the parameter of the GPRN used in the combined analysis of the Sun's RV and FWHM measurements.	187
C.13	Posterior distributions of the parameter of the GP used in the combined analysis of the Sun's RV and $\log R'_{hk}$ measurements.	188
C.14	Posterior distributions of the parameter of the GPRN used in the combined analysis of the Sun's RV and $\log R'_{hk}$ measurements.	189
D.1	Fit and respective residuals of the MAP values from the GP on the RV and the FWHM time series of HD 10700.	192
D.2	Fit and respective residuals of the MAP values from the GP on the RV and the FWHM time series of HD 34411.	193
D.3	Fit and respective residuals of the MAP values from the GP on the RV and the FWHM time series of HD 101501.	194
D.4	Fit and respective residuals of the MAP values from the GP on the RV and the FWHM time series of HD 26965.	195

List of Tables

3.1	Median values of parameters obtain for the node and weight of each GPRN model. To calculate the errors we used the 16th and 84th percentiles.	102
4.1	Parameters and prior distribution on the analysis of a single for both GP and GPRN. Here, y represents the observed output and t the times of observation. In this individual analysis, the measurements y could either be the RVs, FWHM, BIS, or $\log R'_{hk}$	107
4.2	Parameters and prior distributions when considering the analysis of the RVs with an activity indicator. y_{RV} represents the RV time series while y_{AI} the activity indicator (FWHM, BIS, or $\log R'_{hk}$) time series, as previously t represents the time.	108
4.3	Parameters obtained by the GPRN and the GP on the analysis of the RV time series. On the GPRN's variables, the exponent represents if the variable is either from a node (n) or a weight (w).	118
4.4	Parameters obtained by the GPRN and the GP on the analysis of the BIS time series. On the GPRN's variables, the exponent represents if the variable is either from a node (n) or a weight (w).	121
4.5	Parameters obtained by the GPRN and the GP on the analysis of the FWHM time series. On the GPRN's variables, the exponent represents if the variable is either from a node (n) or a weight (w).	124
4.6	Parameters obtained by the GPRN and the GP on the analysis of the $\log R'_{hk}$ time series. On the GPRN's variables, the exponent represents if the variable is either from a node (n) or a weight (w).	127

4.7	Parameters obtained by the GPRN and the GP on the analysis of the RV and the BIS time series. On the GPRN's variables, the exponent represents if the variable is either from a node (n) or a weight (w).	131
4.8	Parameters obtained by the GP and the GPRN on the analysis of the RV and the FWHM time series. On the GPRN's variables, the exponent represents if the variable is either from a node (n) or a weight (w).	135
4.9	Parameters obtained by the GP and the GPRN on the analysis of the RV and the $\log R'_{hk}$ time series. On the GPRN's variables, the exponent represents if the variable is either from a node (n) or a weight (w).	139
A.1	Parameters and prior distribution used on the MCMC analysis. $\mathcal{U}(a, b)$ stands for a uniform distribution between a and b . $\mathcal{LU}(a, b)$ stands for log-uniform distribution with shape parameters a and b . \mathcal{MLU} stands for modified log-uniform distribution, a log-uniform distribution that allows the support of the distribution to include zero. On the parameters of the distributions, y stands for the input and t the time. t_{AV} is the average time between consecutive observations. σ is the standard deviation of the vector used, while PTP stands for its peak-to-peak value. We set the amplitude of the node (η_1^n) to $1m/s$ to simplify the models used.	166
B.1	Summary with the number of points, time spans, peak-to-peak amplitude, average error, and initial RMS of the datasets analysed on this thesis.	172
C.1	Parameters and its respective units.	174

Chapter 1

Introduction

How many extra-solar planets, also known as exoplanets, exist? That is a simple yet increasingly tricky question to answer. It is possible to say for a given day, but that same day a planet is found that surpass any expectations astronomers might have had. Twenty-six years after the first exoplanet detection, we have continuous growth in the number of known exoplanets.

With the current and planned instruments, this trend will continue for many years to come. With it, many challenges will arise. New methodologies will be required to solve it. With that in mind, this thesis tries to develop advanced statistical data analysis methods for the detection of other Earths.

To reach this, however, it is fundamental to understand how the field evolved. In this first chapter, I will summarise the main milestones that marked the field, the methods most commonly used, and how the exoplanets changed our perception of the Universe.

1.1 Exoplanets search overview

The answer to what was the first exoplanet discovered might, surprisingly, depend on which side of the Atlantic the person is. One might answer that the discovery of a planetary system around pulsar PSR1257+1 (Wolszczan and Frail, 1992) or the discovery of 51 Pegasi b (Mayor and Queloz, 1995) as the

milestone that began the extra-solar planet search field.

As in any other field, earlier claims will push the boundaries of technology and scientific knowledge to the limit. That means such claims will likely be received with disbelief. Sometimes to independently verify the results, many years are required. Examples of these early claims are, for example, a companion to Barnard's Star announced by van de Kamp (1982) later disproved by Choi et al. (2013), or the Jupiter mass planet Gamma Cephei Ab made by Campbell et al. (1988) that was only confirmed decades later by Hatzes et al. (2003).

As of the 2nd of December 2021, 4878 planets have been discovered¹. These populate a wide variety of orbital and mass ranges as shown by figure 1.1. Visually it is easy to identify three main groups. The first is the group of giant planets with short periods. The majority discovered using transit photometry. In the second group, we have giant planets at long periods found with radial velocities. The last group includes lower-mass planets at short periods. These were discovered with both transit and radial velocity methods and are currently the main focus on exoplanet search. Next, I will focus on describing the main methods used for these discoveries.

¹Source: <http://exoplanet.eu/>

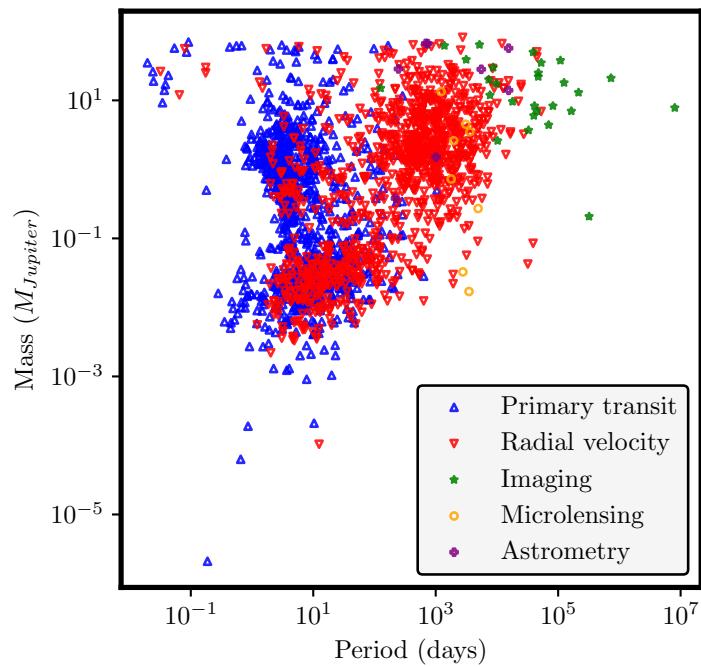


Figure 1.1: Mass-period distribution of all detected exoplanets.

1.1.1 Detection methods

There are several methods to detect exoplanets. Each one has unique strengths and limitations. The two most efficient are the radial velocity and transits methods. These together discovered more than 4000 planets alone. Other successful methods include the use of direct imaging, astrometry, and gravitational microlensing. In this section, I will give a background on these five methods, how we can combine them to characterize exoplanets, and what is the future for these methods.

Radial velocities

The radial velocity method, also known as Doppler spectroscopy, was used by Mayor and Queloz (1995) to discover 51 Pegasi b. This method uses the shifts in the spectral lines of a star caused by the presence of a planet. The gravitational effect of the orbiting planet will cause both star and planet to orbit a common centre of mass. This orbital motion will cause a blue and redshift in the stellar spectrum as it approaches or moves away from our line of sight, respectively. In figure 1.2 we have a representation. Precise spectrographs on Earth can then measure these periodic shifts.

If the orbits were perfect circles, the shift observed would have a sinusoidal shape. That would mean we could use a sine function to interpret the observations. Unfortunately, orbits are ellipses. That implies the orbits are elongated as measured by the orbital eccentricity. To explain this, a function a bit come complex than a sine is required. This function is commonly known as a Keplerian function.

Usually we can translate the radial velocity shifts into physical meaningful information using the equation

$$v_r = K[\cos(\omega + \nu) + e \cos(\omega)] + \gamma. \quad (1.1)$$

On this equation, K represents the radial velocity semi-amplitude. ν represents the true anomaly, while ω is the argument of the pericenter, e is the orbital eccentricity of the planet, and γ is the proper motion of the centre of mass (Murray

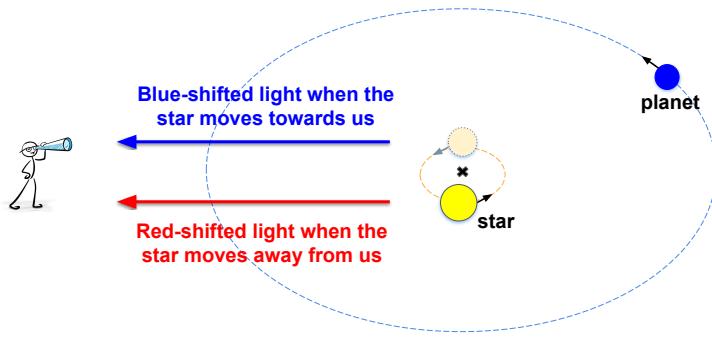


Figure 1.2: Orbital motion of a star and a planet around a common center of mass, and the blue and red shifts on its light caused by orbit of the star.

and Correia, 2010; Perryman, 2011). See chapter 2 for a detailed derivation of this formula.

With some approximations, the radial velocity semi-amplitude is found to be

$$K = 28.4 \text{ m.s}^{-1} \left(\frac{P}{\text{1 year}} \right)^{-1/3} \left(\frac{M_{\text{planet}} \sin i}{M_{\text{Jupiter}}} \right) \left(\frac{M_{\text{star}}}{M_{\text{Sun}}} \right)^{-2/3}, \quad (1.2)$$

where we have the orbital inclination i , the orbital period P in years, and the stellar mass M_{star} in solar masses. $M_{\text{planet}} \sin i$ is called minimum mass (Perryman, 2011). In Doppler spectroscopy, we are unable to obtain the real mass of the planet, but its minimum mass, which is statistically very close to the real planetary mass (Santos and Faria, 2018). I leave more details about this method for chapter 2.

That means, of course, that with this method is not possible to determine many other planetary parameters. For example, we have the mass but not the radius. Meaning a density for the planet cannot be estimated. That is where other methods come into play, as it allows us to combine the different information we obtain with each and better characterize an exoplanet. Next, I present the transits

method, many times used in combination with the radial velocity method with this exact purpose.

Transits

We can define a transit quite easily. It is the passage of a celestial body directly in front of another. Occurring among the line of sight, but not in the correct position to create an eclipse. That, for example, can happen in the Solar system when we observe the transits of Mercury and Venus. It can also occur when a planet is in orbit of another star.

If the planet has an orbital axis that is nearly perpendicular to the line of sight, then there will be a period in time where it will pass between the star and us, shown by figure 1.3. As the planet passes in front of the star, it obscures a small percentage of the star, diminishing the flux we receive from the star by an amount ΔF . Assuming spherical shapes and negligible flux from the planet, this will be related to the ratio of the radius of planet R_p and star R_s by (Deeg and Alonso, 2018)

$$\frac{\Delta F}{F} = \left(\frac{R_p}{R_s} \right)^2 = k^2, \quad (1.3)$$

where F is the total observed flux, and k represents the radius ratio.

A transit will carry valuable information about the exoplanet performing it. If a transit occurs equation 1.3 allows us to, knowing the stellar radius, estimate the radius of the planet. Two consecutive transits will give us the orbital period P of the planet. Using Kepler's laws, it is possible to estimate the semi-major axis of the orbit a .

If the stellar radius is unknown, it is possible to combine the information from the total time of transit t_T , the time for which the planetary disk is in front of the stellar disk t_F , the period P , the radius ratio, and the orbital inclination. This relation was derived by Seager and Mallén-Ornelas (2003). One of the solutions allow us to write the ratio between a and R_s as

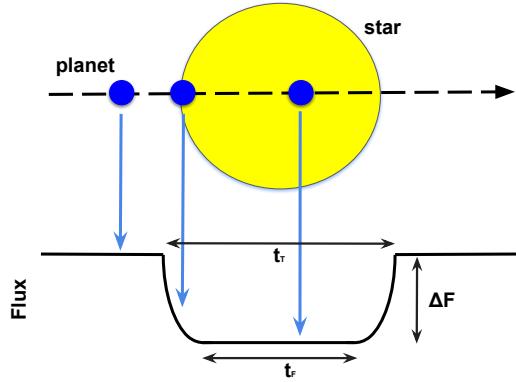


Figure 1.3: Transit of a planet in front of a star and respective flux variation ΔF by it. The transit time duration is represented as t_T and the time of totality as t_F .

$$\frac{a}{R_s} = \frac{2P}{\pi} \frac{\Delta F^{1/4}}{(t_T^2 - t_F^2)^{1/2}} \quad (1.4)$$

and with all the required parameters, estimate the radius of the star. Similarly and if necessary, it is possible to estimate the density of the star with

$$\frac{\rho_s}{\rho_{Sun}} = \frac{32}{G\pi} P \frac{\Delta F^{3/4}}{(t_T^2 - t_F^2)^{3/2}}. \quad (1.5)$$

We can complement the knowledge obtained from the transit method with information gathered using the radial velocity method. Firstly, detecting an exoplanet using RV and transits, two independent methods, confirms the existence of such an exoplanet. Secondly, we can combine the planetary mass obtained using the RV method with the planetary radius obtained by the transit method. That allows us to reach an estimate of the density of the exoplanet.

Additionally, if the planetary system has more than one planet, the gravitational interactions between them can affect the time and duration of the transits. These transit timing variations (TTV) and transit duration variations (TDV) can be used to detect additional planets as small as the Earth (Holman and Murray, 2005).

Imaging

The idea of taking a picture of a planet around another star might sound like a straightforward idea. That, however, does not take into consideration many technical and physical limitations. Already in 1990 Brown and Burrows (1990) were developing the principles to search for exoplanets with this method. However, the first detection was only possible more than a decade later.

The main limiting factor is due to the star being many orders of magnitude brighter than the planet. Due to this, it is of no surprise that, and looking back at figure 1.1, the detections through imaging are of high mass and long orbital period planets. For example, the first exoplanet detected with imaging was 2M1207-39 b, shown in figure 1.4, by Chauvin et al. (2004). This planet, with four times the mass of Jupiter, orbits a brown dwarf from a distance of around 46 AU (Ducourant et al., 2008).

One advantage of direct imaging is that it is a direct detection method. That creates a unique opportunity to study the chemical and atmospheric properties of

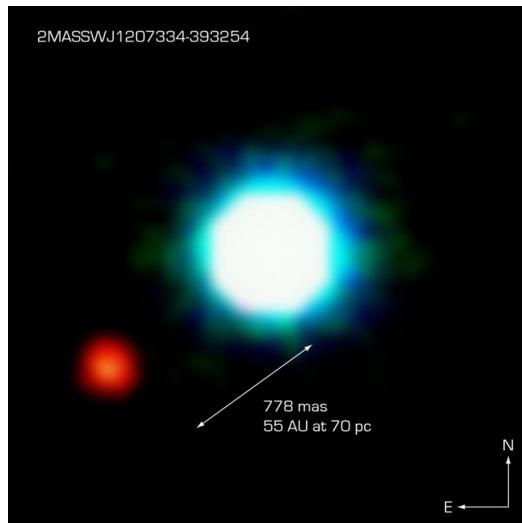


Figure 1.4: The brown dwarf 2M1207 (centre) and its planetary companion (bottom left). Source: ESO

exoplanets. That is not achievable with any other technique. Imaging processing techniques like the angular differential imaging (Marois et al., 2006) and the reference star differential imaging (Ruane et al., 2019), together with instruments like SPHERE (Beuzit et al., 2008) explore this opportunity. With future instruments like the Nancy Grace Roman Space Telescope (Johnson et al., 2020) under development, this method will be fundamental to the study of exoplanets.

Astrometry

Astrometry is probably one of the oldest branches of astronomy and can be traced back to the time of Hipparchus, around 2200 years ago. The principles of using astrometry to detect an exoplanet are similar to the ones used on radial velocities. If a star has a planet, both will orbit a common centre of mass. As such, if we measured the star's position every day, an offset would eventually appear as it crosses the sky.

That astrometric offset, which I will call α , is created due to the planet. Considering that the exoplanet will have a mass of M_p in Earth-masses and it orbits a star with a mass M_s in Solar-masses at a distance of d parsecs from us. Then the offset created will be (Malbet and Sozzetti, 2018)

$$\alpha = 3.3 \frac{a_P}{d} \frac{M_p}{M_s}. \quad (1.6)$$

The parameter a_P is the semi-major axis of the planet's orbit. Also important to mention is that α is in units on micro arc-second (μas).

Unfortunately, the challenges on measuring this offset only now were surpassed with the ESA GAIA mission (de Bruijne, 2012). The main difficulties arose due to the small astrometric offset expected for most cases, of the order of 1 μas (Santos and Faria, 2018). Even before the launch of GAIA, exoplanets detected by other methods were confirmed using astrometry. The first exoplanet discovered using only astrometry only occurred in 2010. This planet, in orbit around the star HD 176051, has a mass of around 1.5 Jupiter masses and a semi-major axis of 1.76 AU (Muterspaugh et al., 2010).

Microlensing

In my opinion, the gravitational microlensing method is one of the most fascinating used in exoplanet detection. This method uses an effect called the gravitational lens effect that occurs because mass bends the space around it (e.g. Sauer, 2008). To understand how the gravitational lens effect occurs, we need to use figure 1.5.

If we have two stars, the closer one to us labelled *lens star*, and the further away labelled *source star*. As the stars move in the galaxy, and under the right conditions, the source star will pass behind the lens star. The light will not be blocked, as this occurs. The mass of the lens star bends the space around it. That means the light emitted from the source star, instead of blocked by the lens star, sees its trajectory warped. As this event occurs, we will detect a change in the light we receive due to the lens effect.

Even more interesting is to use this event to detect exoplanets. If the planet exists, then there will be an extra mass that will bend the light of the source star. This extra bend will occur on only one side of the lens, where the planet is, but not on the opposite side. I represent such an event in figure 1.5. Due to the presence of the exoplanet, the light suffers an extra bend as it travels instead of taking the dashed line trajectory.

The physics behind this method implies that we are observing neither the planet nor the star it is orbiting. We are observing instead the effect they cause on the light of other stars. That makes this method useful to detect low-mass planets on wide orbits. I refer to Gaudi (2010) and Gaudi (2012) for a more in-depth review on the physics behind this method.

The first successful detection with this method occurred in 2003. That year, a 2.6 Jupiter-mass planet was detected, using the Optical Gravitational Lensing Experiment (Udalski et al., 1997), on a 4.3 AU wide orbit (Bond et al., 2004). Unfortunately, this method has one crucial constraint. These events are one-time events and, a second detection of the same planet using microlensing is very unlikely.

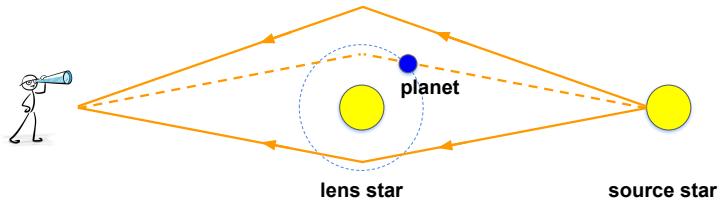


Figure 1.5: Gravitational microlensing amplified by an extrasolar planet.

1.1.2 Milestones

The early 90s marked a change in the previous paradigm when technology finally allowed the first confirmed planetary discoveries. Wolszczan and Frail (1992) presented the discovery of a planetary system around the millisecond pulsar PSR1257+1. This work discovered not only one but two planets with a 66 and 98 days orbital period. Later these discoveries were complemented by the discovery of a third planet in 1994 (Wolszczan, 1994). If today the existence of these three extra-solar planets is not questioned, one might claim these discoveries are not the "true" first exoplanets due to the nature of the system host, a pulsar.

The 6th of October 1995 marked, without a doubt, the beginning of a revolution. Analysing periodic Doppler shifts in the spectra of 51 Pegasi, Mayor and Queloz (1995) confirmed the first planet around another star. This $0.47M_{Jup}$ (Jupiter masses) exoplanet not only impacted the community for being the first exoplanet confirmed around a star. The hot Jupiter discovered, a classification given to gas giants with very short orbital periods, challenged the theories of planet formation.

At the time, and with only our Solar System as a case study, the theories supported that gas giants would only form and orbit beyond the snow-line (e.g. Pollack et al., 1996). Not only did 51 Peg b not follow these characteristics, but

its existence was also later confirmed by Marcy and Butler (1995). In 1996 Lin et al. (1996) showed 51 Peg b most likely had formed by the gradual accretion of solids and capture of gas around 5 AU from the star and gradually migrated inwards.

To the creation and migration of hot Jupiters, it was later given the name of type II disk migration. That occurs when the planet mass is large enough to cause density gaps along its orbit (for more see e.g. Lin and Papaloizou, 1986; D'Angelo et al., 2003). The occurrence of planet migration allowed the development of the Nice Model in 2005, which explains the evolution into their current position of the gas giants of the Solar System (Gomes et al., 2005; Morbidelli et al., 2005; Tsiganis et al., 2005).

The theories against 51 Peg b were slowly proven wrong with the discovery of other hot Jupiters (e.g. Marcy and Butler, 1996; Butler et al., 1997). The definitive proof for this class of planets came in the year 2000. Charbonneau et al. (2000), using the transit method, confirmed the existence of HD 209458 b, also detected by Henry et al. (2000) using radial velocity measurements. The detection of this $0.69 M_{Jup}$ planet on a 3.52 days orbit using two independent methods refuted any argument against hot Jupiters.

The following years saw an explosion in numbers of not only these gas giants but of new classes of planets, see figure 1.1. The increasing number of instruments and methods available to astronomers also led to the discoveries of Neptune and Super-Earth planets (e.g. Rivera et al., 2005; Santos et al., 2004b).

1.1.3 Statistical analysis

With a growing number of planets discovered, it is impossible not to classify and group them according to their characteristics. The previously mentioned *Hot-Jupiters* populate the region with periods inferior to 10 days and masses above $0.1 M_{Jup}$ (Udry and Santos, 2007). The discovery of Gliese 436b, a Neptune-size planet, led to the definition of *Hot-Neptunes* for Neptune-like exoplanet with short periods (Butler et al., 2004; Gillon et al., 2007).

With the same mass ranges but longer periods ($P > 100$ days), we have cold

gas giant planets, often called *Cold-Jupiters*. A famous example of these types of exoplanets is Kepler-16b, a Saturn size and mass planet. It was also the first exoplanet discovered in a circumbinary orbit around a binary star system (Doyle et al., 2011).

Another class of exoplanets we can call as *Mini-Neptunes*, for gas planets smaller than Neptune. Two examples of this type of planets are Gliese 1214b (Charbonneau et al., 2009) and Kepler-138d (Rowe et al., 2014). These planets are often also labelled as *Super-Earths*. This other class is for rocky planets with a mass higher than the Earth, e.g., 55 Cancri e (McArthur et al., 2004), Gliese 876d (Rivera et al., 2005), and Corot-7b (Léger et al., 2009). The broad definition of these two classes, and the scientific and instrumental limitations of the discovery methods employed, can end with a planet being labelled in either class by different studies.

With a large number of exoplanets, one can also make a statistical study of their main features. These studies are possible by the surveys performed over the years, such as the Californian Planet Search survey (Marcy et al., 2005), the Geneva survey of FGK stars (Mayor et al., 2011; Sousa et al., 2011) and the Kepler space telescope survey (Borucki et al., 2009). Some of the main statistical results are listed below.

Ocurrence rate

We now believe that most Sun-like stars have planets. The quantification of this statement is not a trivial task. One could try to calculate the average number of planets per star n . However, the range of properties (size, mass, and orbital parameters) a planets display, added together to the host star properties, such as mass, age, and metallicity, make it impossible to derive a function for it.

The most recent surveys show an important difference in the occurrence between giant and small planets (Winn, 2018). Several studies (e.g. Cumming et al., 2008; Santerne et al., 2016) agree that $\sim 10\%$ of the Sun-like stars will have a giant planet with an orbital period of a few years. Studying the long-period giants this percentage increase to $\sim 52\%$ (e.g Bryan et al., 2016), while the occurrence of Hot-Jupiter decreases to 1% or lower depending on the survey (e.g. Wright et al., 2012; Fressin et al., 2013). It is now believed that around 50% of all Sun-

like stars have a small planet with an orbit shorter than 100 *days*. It decreases to around 27% for periods shorter than 50 *days* (Mayor et al., 2011).

Stellar metallicity and planets

A star forms within a molecular cloud. It is natural to think that its element abundance will reflect itself on the star composition. It is also natural to consider that such element abundance will exist in the protoplanetary disk. That will, of course, influence the planets soon to be born. We find this principle in the correlation between gas-giants and stellar metallicity.

It has been observed that gas-giant planets occur more frequently around metal-rich stars (e.g. Fischer and Valenti, 2005; Santos et al., 2004a). With the increasing number of planets discovered it was also concluded that the correlation is not observed on low-mass planets (e.g. Sousa et al., 2008). Metal-poor stars have been suggested to host longer period planets and less eccentric orbits (e.g. Adibekyan et al., 2013).

These discoveries, correlations, or lack thereof, are crucial for the development of planet formation theories. For example, the relation between gas-giants and the metal-rich star has an origin on the primordial cloud. The cloud, being metal-rich, allows the core to form faster. Due to this, it reaches a sufficient mass to gravitational attract gas more rapidly. I suggest the reader to Udry and Santos (2007) and Adibekyan (2019) for a more in-depth review on this issue.

Hot Neptune desert

One surprising fact about the exoplanets discovered so far is the lack of short-period Neptune-sized planets (e.g. Szabó and Kiss, 2011). This result also called *sub-Jovian desert*, is defined as the lack of exoplanets with masses around $0.1 M_{Jup}$ on short-period orbits of $P < 4$ *days* (West et al., 2019). This anomaly, shown in figure 1.6, is believed to be associated with a combination of photo-evaporation of the atmospheres and tidal disruptions due to gas giants migrations (West et al., 2019). The existence of this desert also seems to indicate that Super-Earths and Hot-Jupiters have distinct formation processes (Owen and Lai, 2018).

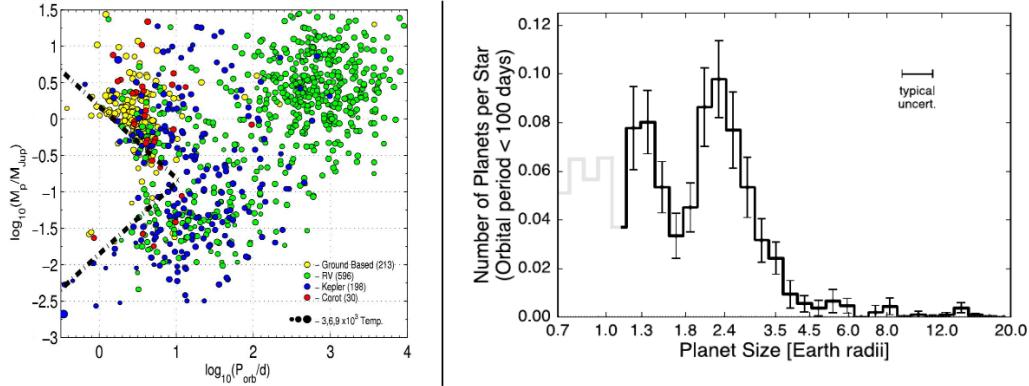


Figure 1.6: Left: Hot-Neptune desert derived by Mazeh et al. (2016) with the dash-dotted lines delimiting the boundaries of this desert. Right: Radius gap from Fulton et al. (2017) where the gap around 1.8 Earth radii creates a bi-modal distribution of planets around Sun-like stars.

Exoplanet radius valley

The exoplanet radius valley is a gap in the radius of small planets observed in the California Kepler Survey (Petigura et al., 2017; Fulton et al., 2017). This gap (see figure 1.6) known for planets with radius around the 1.8 Earth radius had already been predicted by numerical models due to photoevaporation (e.g Lopez and Fortney, 2013; Owen and Wu, 2013). Another explanation is that the luminosity of the cooling rocky core erodes the envelopes, only preserving heavy ones. Such a mechanism would produce a lack of intermediate-sized planets (Ginzburg et al., 2018).

While first identified for FGK stars, more recent works also included M stars. The results indicate that this gap occurs at a lower planets' equilibrium temperature than expected makes M dwarfs stars more effective at reducing the size of their planets (Hirano et al., 2018). That seems to confirm that the driving process creating this gap is photoevaporation.

1.2 This thesis

Now, more than ever, methods to improve our planet detection capabilities are necessary. The most recent instruments are capable of reaching a precision never before seen. Meaning every day, we are closer to detect a second Earth.

My work on this thesis will focus on using radial velocity data. Combining the flexibility of Gaussian processes with the structure of a neural network, I will adopt a framework known as a Gaussian process regression network to improve the detection limits using radial velocity observation.

I organized this thesis into four more chapters. In the next one, I will briefly explore the main aspects of the radial velocity method. I will start with the mathematical background that supports this method. I will follow by seeing how to obtain these observations and the obstacles we need to tackle when using radial velocity data.

In the third chapter, I will present the Gaussian process regression network. I will start with a basic introduction to Gaussian processes, explore the covariance functions of interest when performing a Gaussian process regression. The chapter ends with a presentation of the mathematics behind a Gaussian process regression network and how we intend to use it on radial velocity data.

I dedicate the fourth chapter to exploring a fully functioning Gaussian process regression network in radial velocity data. I will focus on exploring its capabilities on measurements obtained from the Sun. I will also present some of the results obtained on four stars from the EXPRESS Stellar Signal Project. In the fifth and last chapter, I compile the conclusions we achieved with our new framework, its pros and cons, and the future possibilities and improvements still required.

Chapter 2

Doppler spectroscopy

In this chapter, we describe the key aspects behind the measurement of precise radial velocities. We start by presenting the mathematical/physical principles behind Doppler spectroscopy. It is followed with a brief discussion of the radial velocity calculation techniques. The chapter concludes with the challenges stellar activity create on these observations and the mitigation strategies currently employed.

2.1 Mathematical background

As seen in the previous chapter, a planet around a star causes both to orbit a common centre of mass. If we observe the spectrum of said star, we will see that its lines shift with the star radial motion. These shifts in wavelength or relative radial velocity between the star and the observer can be measured using the relativistic Doppler shift z (Sher, 1968)

$$z \equiv \frac{\lambda_{\text{observed}} - \lambda_{\text{rest}}}{\lambda_{\text{rest}}} = \frac{\gamma}{1 + \frac{v_r}{c}} - 1, \quad (2.1)$$

where λ is either the line wavelength observed or the one measured in a rest frame. γ is a relativistic factor, c is the speed of light in the vacuum, and v_r is the radial velocity. Considering that for our problem $\gamma \approx 1$ and $c \gg v_r$ we can simplify

equation 2.1 and obtain

$$v_r = c \frac{\lambda_{\text{observed}} - \lambda_{\text{rest}}}{\lambda_{\text{rest}}}. \quad (2.2)$$

To understand how we use this measure to determine if a set of RV measurements includes one or more planets, we need to understand how we reach the keplerian function from equation 1.1 and understand its main parameters. For it I will use the works of Murray and Correia (2010), Perryman (2011), and Wright and Gaudi (2013) to derive all the necessary components.

2.1.1 Keplerian orbits

The two-body problem allows us to understand the motion of a planet and a star around their barycenter. If we consider Newton's law of universal gravitation, a body attracts another body with a force directly proportional to the product of their masses m_1 and m_2 , and inversely proportional to the square of their distances r (Newton, 1833). Mathematically this force F_g is

$$F_g = G \frac{m_1 m_2}{r^2}, \quad (2.3)$$

m_1 and m_2 , with $m_2 > m_1$, is the mass of the two bodies, r the distance separating them, and G is the universal gravitational constant valued at $6.67430 \times 10^{-11} m^3 kg^{-1} s^{-2}$ (NIST, 2019). To simplify the notation, and unless specified otherwise, I will use m_1 and m_2 to name the respective bodies.

Now if we consider the two bodies m_1 and m_2 (or a planet and a star), with position vectors \mathbf{r}_1 and \mathbf{r}_2 in relation to an origin O (figure 2.1). The relative motion of m_1 with respect to m_2 is then $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$. This allow us to write the vector representation¹ of the gravitational force each body apply as

$$\begin{aligned} \mathbf{F}_1 &= -G \frac{m_1 m_2}{r^3} \mathbf{r}, \\ \mathbf{F}_2 &= +G \frac{m_2 m_1}{r^3} \mathbf{r}. \end{aligned} \quad (2.4)$$

¹ The vector representation of Newton's law of universal gravitation is $\mathbf{F} = G \frac{m_1 m_2}{r^2} \hat{\mathbf{r}}$, where $\hat{\mathbf{r}}$ is a unit vector with the direction of the force.

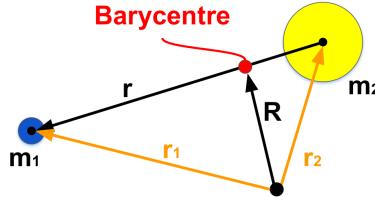


Figure 2.1: The two-body problem scheme.

Now considering Newton's second law of motion² we can re-write it as

$$\begin{aligned} m_1 \ddot{\mathbf{r}}_1 &= -G \frac{m_1 m_2}{r^3} \mathbf{r}, \\ m_2 \ddot{\mathbf{r}}_2 &= +G \frac{m_2 m_1}{r^3} \mathbf{r}. \end{aligned} \quad (2.5)$$

Since we now have $\ddot{\mathbf{r}}_1$ and $\ddot{\mathbf{r}}_2$ lets not forget that $\ddot{\mathbf{r}} = \ddot{\mathbf{r}}_1 - \ddot{\mathbf{r}}_2$. This allows us to combine the terms from 2.5 into

$$\ddot{\mathbf{r}} = -\frac{G}{r^3} (m_1 + m_2) \mathbf{r}. \quad (2.6)$$

This is the equation we need to solve to obtain the formula for a closed elliptical orbit.

The following step requires the use of the cross product³. For an orbiting body, the position and acceleration are collinear vectors. This means

$$\begin{aligned} \dot{\mathbf{r}} \times \mathbf{r} &= 0 \\ \mathbf{r} \times \dot{\mathbf{r}} &= \text{constant} = \mathbf{h}, \end{aligned} \quad (2.7)$$

having \mathbf{h} as a vector perpendicular to \mathbf{r} and $\dot{\mathbf{r}}$.

To solve equation 2.6 we need to change coordinate system into the polar coordinate system (r, θ) . The origin being centered on the body m_2 , and an

² Newton's second law of motion relates the force applied to a body to its change of momentum. It is defined as $\mathbf{F} = \frac{d\mathbf{p}}{dt} = m\mathbf{a}$. Where \mathbf{F} is the force, \mathbf{p} the momentum, m the mass, and \mathbf{a} the acceleration of the body.

³ The cross product is defined $\mathbf{v}_1 \times \mathbf{v}_2 = ||\mathbf{v}_1|| ||\mathbf{v}_2|| \sin \theta \mathbf{n}$, where θ is the angle between the vectors \mathbf{v}_1 and \mathbf{v}_2 , and \mathbf{n} is a unit vector perpendicular to the plane created by \mathbf{v}_1 and \mathbf{v}_2 .

arbitrary reference line defined for $\theta = 0$.

Considering the equations of motion in polar coordinates⁴, and with the appropriate substitutions, we first obtain

$$h = r^2 \dot{\theta}, \quad (2.8)$$

and manage to transform equation 2.6 into a second-order, linear differential equation. This equation has a general solution given by

$$\frac{1}{r} = \frac{G (m_1 + m_2)}{h^2} [1 + e \cos (\theta - \omega)], \quad (2.9)$$

where e and ω are two constants of integration. In reality, this equation is the general equation for a conic section where e is the eccentricity.

Following some more algebraic manipulations, we reach the Keplerian orbit for an ellipse in polar coordinates

$$r = \frac{a (1 - e^2)}{1 + e \cos (\theta - \omega)} = \frac{a (1 - e^2)}{1 + e \cos \nu}, \quad (2.10)$$

where a is the semi-major axis of the ellipse, e its eccentricity, and ν a quantity called a true anomaly. This solution is nothing else but the Newtonian derivation of Kepler's first law of planetary motion.

2.1.2 Orbital elements

In equation 2.10, I defined the quantity $\nu = \theta - \omega$ but did not explain its meaning. That and other quantities of interest are shown in figure 2.2 and arise from the various angles in the orbit plane.

⁴ These are $\mathbf{r} = r \hat{\mathbf{r}}$, $\dot{\mathbf{r}} = \dot{r} \hat{\mathbf{r}} + r \dot{\theta} \hat{\theta}$, and $\ddot{\mathbf{r}} = (\ddot{r} - r \dot{\theta}^2) \hat{\mathbf{r}} + [\frac{1}{r} \frac{d}{dt} (r^2 \dot{\theta})] \hat{\theta}$

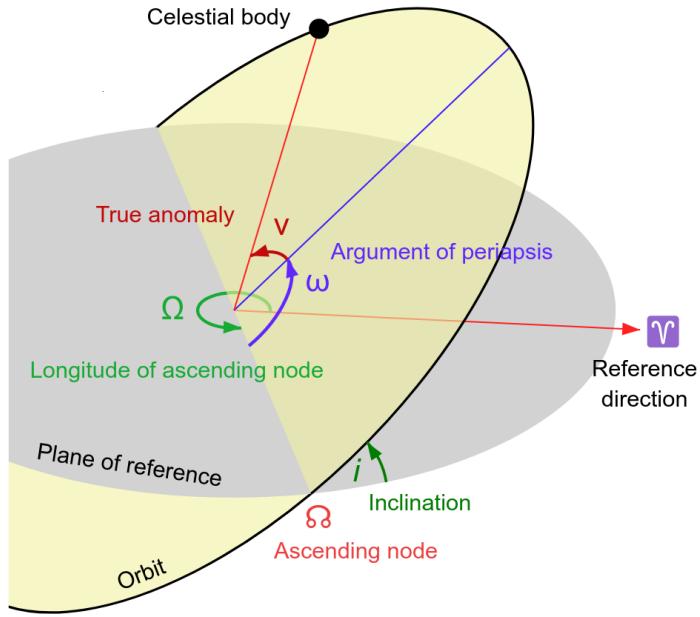


Figure 2.2: Elliptical orbit of a celestial body in three dimensions. Source: Wikimedia Commons.

Inclination

The orbit inclination, denoted as i , is the angle between the orbit plane and the plane of reference. It varies between 0 and π .

Ascending node

The ascending node Ω is the point where the planet traverses the reference plane. It occurs when the planet is moving from below to above the reference plane.

Argument of periapsis

The argument of periapsis is also designated periastron for orbits around stars. Usually represented by an ω , it is the angle between the ascending node and the periapse of the orbit. When calculated, it is done considering the direction of motion and varies between 0 and 2π .

Longitude of ascending node

The longitude of the ascending node Ω , estimated in the reference plane, is the angle where the object moves away from the observer into the direction of the ascending node.

True anomaly

The previously mentioned true anomaly v , or seldom expressed as $v(t)$ as it is dependent on the time, is the angle between the direction of the periapsis and the position of the planet at time t .

Eccentric anomaly

The eccentric anomaly $E(t)$ is an angle obtained from the circumcircle of ellipse⁵ as the planet moves along its elliptic orbit. The true and eccentric anomalies are related by

$$\cos v(t) = \frac{\cos E(t) - e}{1 - e \cos E(t)}, \quad (2.11)$$

that after using some trigonometric properties give us

$$\tan \frac{v(t)}{2} = \left(\frac{1+e}{1-e} \right)^{1/2} \tan \frac{E(t)}{2}. \quad (2.12)$$

⁵ The circumcircle of an ellipse is an auxiliary circle with centre on the centre of the ellipse and radius equal to the semi-major axis.

Mean anomaly

We use the mean anomaly $M(t)$ to determine the true anomaly. Over an orbit, a planet does not move at a constant angular rate. Even so, we can calculate an average. That allows us to calculate the mean anomaly at time $t - t_p$, with t_p being the moment of the passage of the periapsis and define the mean anomaly as

$$M(t) = \frac{2\pi}{P} (t - t_p), \quad (2.13)$$

with P the period of the orbit. Using orbital mechanics we can relate the mean and eccentric anomalies with

$$M(t) = E(t) - e \sin E(t). \quad (2.14)$$

2.1.3 Calculating the Keplerian function

Having defined the main orbital elements, we can finally determine the Keplerian function to model a planetary signal. Let us start by remembering Kepler's third law⁶ writing it as

$$P^2 \propto a^3, \quad (2.15)$$

where the proportionality constant to be inserted will take slightly different values accordingly to the type of orbit. If, for example, we consider the orbit of the body m_2 around the system barycenter, its period and semi-major axis relate as

$$P_{m_2}^2 = \frac{4\pi^2}{GM} a_{m_2}^3 \quad \text{with} \quad M = \frac{m_1^3}{(m_1 + m_2)^2}. \quad (2.16)$$

Omitting some trigonometry manipulations, the position of the body m_2 among the the line-of-sight $z(t)$, perpendicular to the plane of reference of figure 2.2, is

⁶ Kepler's third law states that the squares of the orbital period of a planet are proportional to the cube of its semi-major axis

$$z(t) = r(t) \sin i \sin(\omega + \nu(t)), \quad (2.17)$$

with $r(t)$ being the distance to the barycenter. To obtain the velocity is then necessary to derive the expression concerning t , giving us

$$\dot{z}(t) = \sin i [\dot{r}(t) \sin(\omega + \nu(t)) + r(t) \dot{\nu}(t) \cos(\omega + \nu(t))]. \quad (2.18)$$

This velocity $\dot{z}(t)$ is no more than the projected motion along the line-of-sight, thus being a radial velocity ($v_r \equiv \dot{z}$). Even so, we still need some more steps to obtain a similar expression to equation 1.1. For it, we make some algebraic manipulations with $r(t)$ and $\dot{r}(t)$ to reach

$$v_r(t) = K[\cos(\omega + \nu(t)) + e \cos(\omega)] + \gamma, \quad (2.19)$$

with

$$K = \frac{2\pi}{P_{m_2}} \frac{a_{m_2} \sin i}{\sqrt{1 - e^2}}. \quad (2.20)$$

Using Kepler's third law from equation 2.16 we can re-write the semi-amplitude as

$$K = \left(G \frac{2\pi}{P} \right)^{1/3} \frac{m_1 \sin i}{(m_1 + m_2)^{2/3}} \frac{1}{(1 - e^2)^{1/2}}. \quad (2.21)$$

That allows us to obtain two other equations of interest. First, for the minimum mass where we have

$$m_1 \sin i = 4.919 \times 10^{-3} K \left(1 - e^2 \right)^{1/2} P^{1/3} (m_1 + m_2)^{2/3}. \quad (2.22)$$

On it, $m_1 \sin i$ is given in Jupiter masses, K in meters per second, P in days, and $m_1 + m_2$ in Solar masses.

The second equation is identical to expression 1.2. For it, we need to consider circular orbits and the masses $m_1 \ll m_2$ to reach

$$K = 28.4 \frac{m_1 \sin i}{P^{1/3} m_2^{2/3}}. \quad (2.23)$$

In this equation we have K in meters per second, P in years, $m_1 \sin i$ in Jupiter masses, and m_2 in Solar masses⁷.

We now have all the necessary equations to characterize a set of RV points. That, however, is not the end. Equation 2.19 allows only a single planetary signal. For multiple-planet fitting, we can do a linear sum for an n number of planets

$$v_r(t) = \sum_{i=1}^n K_i [\cos(\omega_i + \nu_i(t)) + e_i \cos(\omega_i)] + \gamma. \quad (2.24)$$

Another common issue with RV observations arises when we have multiple instruments observations. These observations will most likely include an instrument-dependent RV offset. That will require the use of different offsets, one for each different sections of the measurements, accordingly to the instruments.

Also quite common in RV analysis is to include a linear trend parameter d to fit any long term trends that might appear in the data not necessarily related to planetary signals. That means using a function of the type

$$v_r(t) = \sum_{i=1}^n K_i [\cos(\omega_i + \nu_i(t)) + e_i \cos(\omega_i)] + \gamma + d(t - t_0), \quad (2.25)$$

where d is the slope of the long term signal, and t_0 is the average of t .

The last major setback to fully characterize any planetary signal appears on the relation between the true, eccentric, and mean anomalies. To obtain the value of true anomaly is necessary to use its relation to the eccentric anomaly. That then requires the use of the mean anomaly to calculate the eccentric anomaly. The setback on these steps is on equation 2.14 being transcendental⁸. To calculate, we

⁷ Usually equations 2.22 and 2.23 are written with the units in the denominator. That is helpful to avoid using the wrong units but, I did not use it in this chapter.

⁸A transcendental function is a function that can not be express by a finite sequence of the algebraic operations

require the use of iterative methods, like the Newton-Raphson method. We start by calculating an initial value for $M(t)$ and $E(t)$ with

$$\begin{aligned} E_0 &= M + e \sin M + \frac{e^2}{2} \sin 2M, \\ M_0 &= E_0 - e \sin E_0, \end{aligned} \quad (2.26)$$

calculating M using the time of the periapsis passage. We then calculate any new term iteratively with

$$\begin{aligned} E_1 &= E_0 + \frac{M - M_0}{1 - e \cos E_0}, \\ M_1 &= E_1 - e \sin E_1. \end{aligned} \quad (2.27)$$

Convergence, within a few iterations, should be achieved. Following it, we are ready to calculate the true anomaly and, with it, the Keplerian function.

2.2 Spectroscopic observations

In the previous section, we saw the mathematics behind a radial velocity measurement. That, however, is only possible after having observed the spectra of a given star and appropriately processed the data. That implies transforming the spectra observed into a single measurement. Of course, this is not an easy task. It all starts in the spectrograph that, in simple terms, will receive the light from the telescope, disperse it to create a spectrum, and then record this spectrum on a detector.

2.2.1 The past, present and future of spectroscopy

When Mayor and Queloz (1995) discovered 51 Pegasi b, spectroscopic observations were done using the ELODIE spectrograph (Baranne, A. et al., 1996). This instrument entered into operation in 1993 and was capable of an RV precision of 7 m/s initially and later of 3 m/s, after an upgrade on 2014 (Pepe et al., 2018). Over the years, the developed spectrographs pushed the limits of the precision achieved. For example, HARPS (Pepe et al., 2002) and HARPS-N (Cosentino et al., 2012), both spectrographs with an ever-increasing number of exoplanets discovered when began operations managed to achieve a precision of 1 m/s.

That, however, is not enough when searching for a second Earth. We need to break the "barrier" of detecting a signal of 9 cm/s to accomplish it. Spectrographs like ESPRESSO (Pepe et al., 2014) and EXPRES (Jurgenson et al., 2016), with a precision goal of 10 cm/s, are the most recent bet to achieve this objective. Although obtaining a higher precision brings several challenges, both technical and scientific, astronomers and engineers keep pushing the detection limits. For example, a future goal is to detect signals around the two centimetres per second mark with instruments like CODEX (Pasquini et al., 2010).

Developing and building all these instruments, however, is only the first step. With it, we aim at obtaining a spectrum with as many lines as possible. If successful, we get something similar to what appears in figure 2.3. In this figure, we see many dark absorption spectral lines. Later we use it to obtain a radial

velocity measure. A high number of lines, however, is not the only requirement for high precision measurements. Radial velocity measurements are also affected by several potential error sources. Next, I will mention two sources of this noise.

2.2.2 Statistical and systematic errors

Let us consider, for example, an Earth-like planet orbiting a sun-like star at $1AU$. The radial velocity expected is approximately $0.1m/s$. If we look at equation 2.2, and use lines on the visible part of the spectrum, we can calculate the wavelength shift $\Delta\lambda$ required to detect such planet. Under some approximations⁹ this implies a shift of

$$\Delta\lambda = 5 \times 10^{-7} \frac{0.1}{3 \times 10^8} \approx 2 \times 10^{-16} m. \quad (2.28)$$

We enter into a regime where we need to address the statistical and systematic errors when measuring such shifts. For example, the spectrograph needs to be stabilized and precisely calibrated in wavelength to be photon-efficient and precise on a sub-metre-per-second level (Pepe et al., 2014). For more on instrument stability errors and other instrumental factors we needed to take into account, I recommend, for example, Pepe and Lovis (2008).

Another issue arises from the shape of spectral lines. More specifically, we need to address the errors due to photon noise. Bouchy et al. (2001) determined how much this noise contribute to the final RV error when assuming a Gaussian-shaped line. Using the expression from Figueira (2018) we can calculate with

$$\sigma_{RV} = \frac{(\pi \log 2)^{-1/4}}{2} \frac{\sqrt{FWHM}}{SNR} \frac{\sqrt{PXLSC}}{C} F(C_{eff}), \quad (2.29)$$

where C is the contrast of the Gaussian line, SNR the signal-to-noise ratio of the spectrum, $PXLSC$ the dimension of the pixel as measured in velocity, $F(C_{eff})$ is a polynomial function of the effective contrast. On the previously mentioned

⁹ The visible spectrum have a wavelength range from 3.8×10^{-7} to 7.5×10^{-7} meters, to simplify the calculations, I used the value of 5×10^{-7} meters. Another approximation done was for the speed of light in vacuum (c), I consider it as roughly $3 \times 10^8 m/s$.

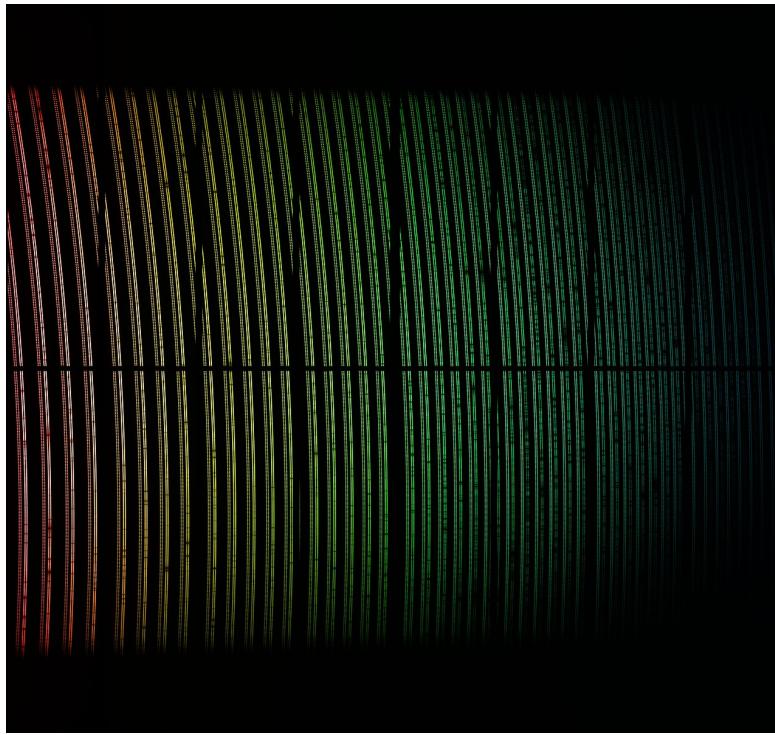


Figure 2.3: ESPRESSO’s first light spectrum. Source: ESO/ESPRESSO team.

HARPS, for example, this noise can have levels of 0.5 m/s with appropriate exposure times (Haywood, 2016b). I will not further explore the photon noise error. Instead I refer to Butler et al. (1996) and Bouchy et al. (2001), for more information on this issue.

2.2.3 The cross-correlation function

As mentioned at the start of this chapter, we use absorption spectral lines to calculate the radial velocity. If we identify the lines and compare them with the theoretical value from the Solar System rest frame, we can measure its radial velocity. This being represented as $\lambda_{\text{observed}} - \lambda_{\text{rest}}$ in equation 2.2.

One of the objectives when gathering RV measurements is to increase the precision as much as possible. Selecting a single line and calculating its RV shift

will not guarantee us that. Instead, we use the mean spectral line obtained from an average of over thousands of spectral lines. To accomplish that, we use the cross-correlation method (Queloz, 1995).

Let us start by a simple definition for cross-correlation from Allende Prieto (2007). Let us consider two arrays of size k . One is from the stellar spectrum (S), and another from a template or mask (M). Cross-correlating them, we obtain the relative velocity between them. That is done using, for example, equation

$$CCF_i = \sum_k M_k S_{k+1}. \quad (2.30)$$

If the spectrum S and M are identical but shifted by an integer amount of p pixels, the maximum value on C will be the element where $i = p$.

In our case, we will use a mean spectral line and cross-correlate it with a line mask to determine the wavelength shift. It will thus be of interest to use a weighted cross-correlation. That is calculated using (Pepe et al., 2002)

$$CCF(v_r) = \int S(\lambda) M(\lambda_{v_r}) d\lambda, \quad (2.31)$$

where $\lambda_{v_r} = \sqrt{\frac{c-v_r}{c+v_r}}$, In this equation $S(\lambda)$ is the measured spectrum, and $M(\lambda_{v_r})$ the Doppler-shifted numerical mask.

The expression becomes simpler when representing M as the sum of masks M_i , for each stellar absorption line i . Considering the different relative depths of each spectral line used, we define a respective weight w . That allows us to reach the expression

$$CCF_{weighted}(v_r) = \sum_i S(\lambda_i) M_i(\lambda_{v_r}) w_i. \quad (2.32)$$

Fitting the CCF with a Gaussian function allows us to determine the radial velocity (see figure 2.4). This measure, obtained considering the location of the peak of the CCF, will be shifted from zero by the star's radial velocity. That, however, is not the only measure of interest we can obtain from the CCF. Other properties are also present.

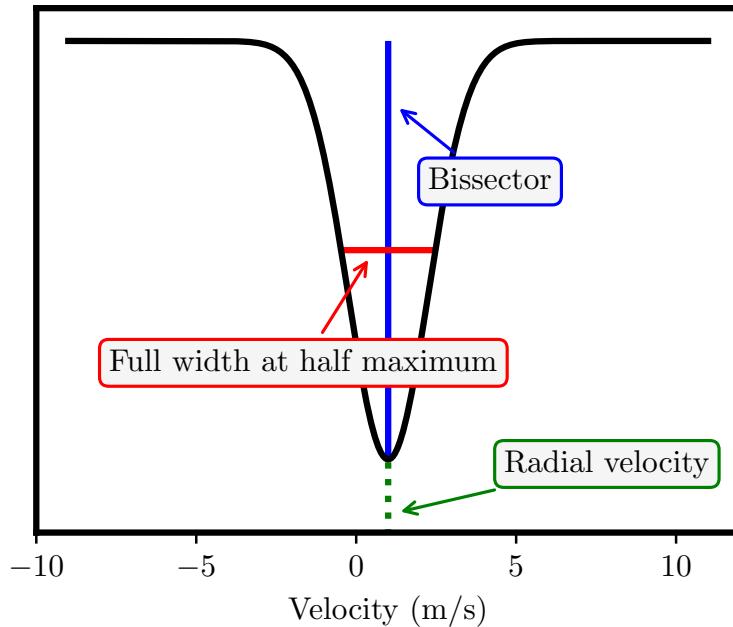


Figure 2.4: Simulated structure of a CCF.

For example, the stellar rotation velocity $v \sin i$, can be inferred from the CCF width (e.g Delfosse et al., 1998; Browning et al., 2010; Reiners et al., 2012). Another useful measure is estimated from the equivalent width. Knowing beforehand the effective temperature of the star and using the CCF equivalent width, one can estimate the star's metalicity (e.g. Mayor, 1980; Santos et al., 2002).

The shape of the CFF is close to that of a Gaussian. That implies we can find in the CCF characteristics usually measured on a Gaussian curve. Properties like the equivalent width and full width at half maximum (FWHM) are relatively easy to obtain using the mathematics behind a Gaussian curve. I will present some of these properties next.

Full width at half maximum

We can define the FWHM as the span within two points on the Gaussian curve, at which its function reaches half of its maximum value. Knowing that the probability density function of a Gaussian distribution is (Patel and Read, 1996)

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right), \quad (2.33)$$

where we have a mean μ , standard deviation σ , and variance σ^2 , we can calculate the FWHM with

$$FWHM = 2\sigma\sqrt{2\log 2}. \quad (2.34)$$

This measure will be sensitive to the variations in the CCF caused by activity on the surface of a star. With it, we can correlate any variation in the FWHM with the RV variations caused by the presence of starspots (Haywood, 2016a). Soon I will explore more about the influence of stellar activity on the FWHM.

Bisection

A property not exclusive to only Gaussian curves is the bisection. A bisection is the division of any geometric figure into two others with the same shape and size. The line dividing the two new figures is called a bisector. Shown by the blue vertical line of figure 2.4, this property will be convenient later for the definition of some activity indicators.

Equivalent width

As mentioned earlier, we can use the equivalent width (EW) to estimate the metalicity. In a spectral line, this measure is obtained from the rectangle with a height equal to that of continuum, and a width that creates a rectangle with area equal to the area in the spectral line (see figure 2.5).

On a CCF we can calculated it with (Borgniet et al., 2019)

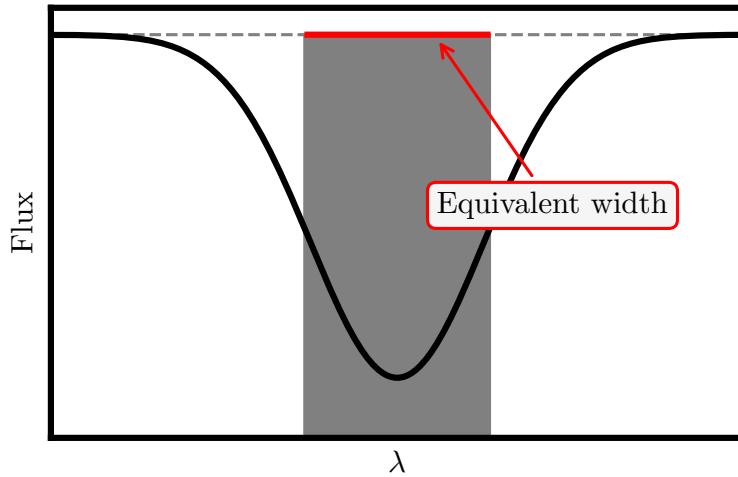


Figure 2.5: Example of the equivalent width of a spectral line.

$$EW = \int \left(1 - \frac{CCF(v_r)}{\max(CCF(\Delta))} \right) dv_r. \quad (2.35)$$

Where ($CCF(\Delta)$) is the area centred on the CCF peak, below the CCF lower wing. On a normalized CCF, the EW will thus be a rectangle with a height equal to 1 and surface identical to the area given by $CCF(\Delta)$.

2.3 The stellar activity problem

We now know how to obtain RV measurements from a star. These measures, however, require careful analysis. One of the main challenges in exoplanets detection arise from the activity from the star contaminating the observation. These stellar signals come from physical processes that induce radial velocity variations on the stars, besides that of a planet, of course. That is known to be able to hide or mimic planetary signals in RV measurements (e.g. Santos et al., 2014; Robertson et al., 2015).

As previously mentioned, a planet induces a Doppler shift on the star spectral lines. That is reflected on the CCF by a change of its peak location that we can measure (Figueira, 2018). That implies that any physical process affecting the shape of the spectral lines will impact any shift measured from the CCF. Already in 1997 Saar and Donahue (1997) showed this would occur when inhomogeneities on a star surface were present. These would affect the shape of the spectral lines. The Doppler shifts it would create could then be mistaken by planetary signals. With spectrographs aiming to achieve a precision of 10 cm/s, the expected radial velocity amplitude of an Earth-like planet, is fundamental to correctly deal with any signal capable of originating a false detection.

2.3.1 Stellar activity features

The stellar activity has a wide range of sources and timescales. These range from a few minutes to dozens of years. That will affect short and long period planets differently. Understanding each feature is fundamental to implementing the necessary strategies required to be deal with each (Dumusque, X. et al., 2011a; Dumusque, X. et al., 2011b).

On timescales of minutes to hours

- *Oscillations*

Pressure or acoustic waves (p-modes for short) propagate at the surface of

stars, causing the exterior envelopes to expand and contract. Usually, on G, K, and M class stars, reach timescales less than 15 minutes. In the case of the Sun, these are known to last from 5 to 15 minutes (Schrijver and Zwaan, 2000; Arentoft et al., 2008; Broomhall et al., 2011).

These oscillations originate when the plasma flows originated from the stellar convective envelope create acoustic waves. These combine, either constructively or destructively, to create stationary waves, whose periodicity of the effect last around five minutes (Cegla, 2019). Individually these waves only create an RV signal on the order of the centimetre per second, but dozens of modes generate a combined signal capable of reaching a few meters per second (Bedding and Kjeldsen, 2003; Bedding and Kjeldsen, 2007).

When we plan spectroscopic observations, we can deal with the induced signals of oscillation. Observations with 10 to 15 minutes exposures are capable of suppressing these signals on solar-like stars, (Dumusque, X. et al., 2011a; Dumusque, X. et al., 2011b). On other stars Chaplin et al. (2019) showed these exposures range from 4 minutes for low mass stars to 100 minutes for hotter higher mass stars.

With the increased precision of spectrographs, another type of oscillation is now of concern. Recently studies showed the existence of Rossby waves, known as r-modes, in the Sun (Löptien et al., 2018). These originated from toroidal modes on a rotating star, can produce RV signals of a few dozens of centimetres per second, and periods between 10 to 20 days (Lanza et al., 2019). As a recent discovery, the r-modes is a future activity source to consider.

- *Flares*

Flares are a sudden increase in the brightness of a star near its surface. Often they are associated, in the case of the Sun, with coronal mass injections. These sudden and random events generate spikes on an RV observation that can range from 0.5 to 10 m/s (Saar, 2009; Reiners, 2009). These, however, are rare and easily identifiable on, for example, the H α emission profile.

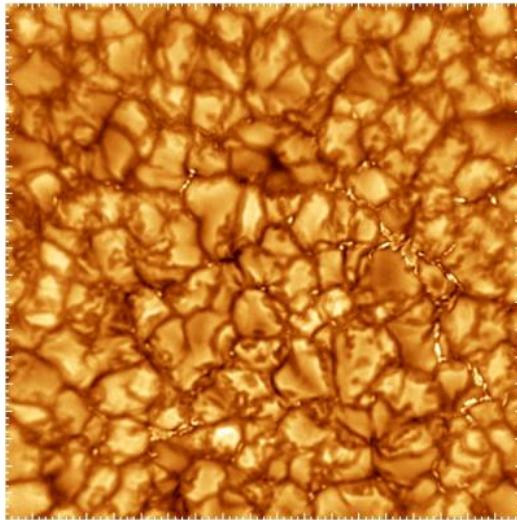


Figure 2.6: Solar granulation over an area of 19000 squared kilometers. Source: Big Bear Solar Observatory/New Jersey Institute of Technology.

- *Granulation*

Another source of noise on a spectroscopic observation is due to a phenomenon known as granulation (see figure 2.6). These phenomena occur when the convection in the external layers of solar-type stars transport bright and hot material from deeper layers, flowing horizontally, cooling down and descending through the darker boundary regions known as intergranular lanes (e.g. Rieutord and Rincon, 2010; Rutten and Severino, 2012). This upward and downward movement of material does not occur equally. With the granules covering more surface area than the intergranular regions, the absorption lines of the spectrum end with an overall net blueshift (Cegla, 2019).

The average lifetime for a granule on the Sun, to reach up to 25 minutes (Bahng and Schwarzschild, 1961). The evolution of these granules and their ratio with the intergranular lines generates a constantly changing RV signal that can reach the meter per second on the Sun (Cegla, 2019). The individual granules also group together, creating supergranules where plasma flows at a much slower rate, generating RV signals capable of reaching lifetimes of 1.5 days (Del Moro, 2004; Rieutord and Rincon, 2010).

Recently, supergranulation phenomena shown to pose a challenge bigger-than-expected on Earth-like planets detection (Meunier, N. and Lagrange, A.-M., 2019; Meunier and Lagrange, 2020).

To eliminate some effects of this phenomenon is recommended to use the appropriate observational strategy. For example, Dumusque, X. et al. (2011a) concluded that three observations per night, separated by 1 to 2 hours, reduce its effects on RV observations. They also determined that while granulation signals are of 1 m/s on the Sun. These can reach values of 30 m/s on other spectral classes.

On timescales of days to months

- *Spots*

Spots are darker regions on the surface of the stars (see figure 2.7). These originate when the magnetic field of the star interacts with the plasma. It inhibits the outgoing convective heat flux of a region and reduces its brightness and temperature (Solanki, 2003). First identified and studied on the Sun, where are known as sunspots, are closely linked to the solar magnetic cycle (Hathaway, 2015).

For the last centuries, we have systematically gathered records of its number and periodic evolution. That allowed astronomers to extrapolate their properties to other stars, where are known as starspots. For a more in-depth review on sunspots and their connection to the solar magnetic cycle, I recommend the reader to both Solanki (2003) and Hathaway (2015).

Since a spot emits less light than the surrounding surface, it will break the balance between the blue-shifted and red-shifted parts. As the star rotates, it originates a Doppler shift on its spectrum that deforms the CCF shape (Saar and Donahue, 1997; Queloz et al., 2001; Dumusque, X. et al., 2011b). That originates signals from 0.2 to 5 m/s on solar age stars with periodicities close to the stellar rotation (Lagrange, A.-M. et al., 2010; Meunier, N. et al., 2010). Younger, more active, and rapidly rotating stars have an activity signal that is mostly dominated by spots (e.g. Reinhold et al., 2019).

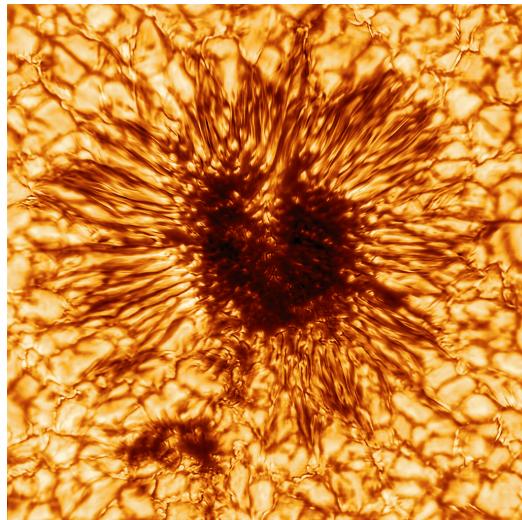


Figure 2.7: A Sunspot capture by the Inouye Solar Telescope on January 28, 2020.
Source: NSO/NSF/AURA.

There are several approaches to deal with these signals. The first uses activity indicators, where the activity but not the planetary signals will be present. I will discuss more of these soon. One can treat the spots signals as correlated noise and fit it with, for example, Gaussian processes (e.g. Haywood et al., 2014; Faria et al., 2016). Another approach consists in using numerical simulations of the active regions to create synthetic RV measurements. These are then compared to the observed RVs to determine their similarities. See Oshagh (2018) for a bit more about this.

- *Faculae and plages*

Faculae are bright structures on the photosphere of a star. In the Sun, these are formed by concentrations of magnetic field lines becoming hotter than the surrounding photosphere (Spruit, 1976). They originate in the space between the granules. Also, they appear close to the solar limb in plages or active regions (Keller et al., 2004). Their contribution to the total solar irradiance varies with the sunspot cycle. It is higher during the sunspots maximum (Berger et al., 2007).

A plage is a bright structure of the chromosphere of a star. It originates in regions of higher temperature and density near the spots. Similarly to spots and faculae, plages are associated with magnetic fields and are born when these fields compress the chromosphere (Solanki, 1993).

Magnetic in origin, both plage and faculae are known to suppress the convective blueshift inside active regions and lead to RV variations up to 10m/s (Méunier, N. et al., 2010). These are the processes that dominate the activity signal on slow rotating and older stars. (e.g. Reinhold et al., 2019).

On timescales of years

- *Magnetic cycles*

Spots, plages and faculae have a close relationship with the solar magnetic cycle. This cycle, on average 11-years long, is produced by dynamo processes in the Sun (Charbonneau, 2020). On a cycle, the number of spots, plages and faculae change as it progress (Hathaway, 2015). For example, since 1749, we have been registering the number of sunspots. On these records, we see a clear increase and decrease as one cycle evolves (see figure 2.8). This magnetic cycle have been observed in other stars (e.g Lovis et al., 2011; Dumusque et al., 2011).

With the search of long orbital period planets, there is now the need to deal with any signal related to this cycle. On other stars, the magnetic cycle can last between 7 up to 30 years, depending on its class and age (Baliunas et al., 1998). On the Sun, studies showed it creates a Doppler signal up to 16.4 m/s and reach values as high as 25m/s on other stars(Lovis et al., 2011).

Luckily these magnetic cycles correlate well with the variation present on activity indicators. That means we can identify the signal using, for example, a linear relationship between the RV and the indicator (e.g. Dumusque et al., 2012). Even so, if the planet's orbital period is very similar to the cycle's period, problems can arise. In the future, we will most likely require better methods to deal with the magnetic cycles signal.

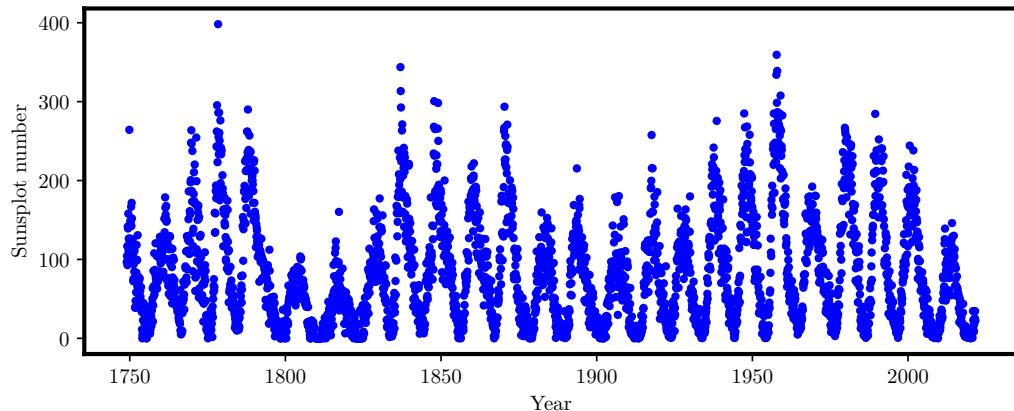


Figure 2.8: Monthly mean of the number of sunspot observed since 1749.
Source: <https://wwwbis.sidc.be/silso>

2.3.2 Stellar activity proxies

Earlier, I mentioned we can use activity indicators to deal with some stellar activity but did not explain it. An activity indicator or proxy is an auxiliary time series, usually derived simultaneously with the RV and carrying identical timestamps. These might be obtained from the same CCF we use to calculate the RV or from the spectroscopic observations.

Let us, for example, consider the case of only one spot on the surface of a star. For it, I use a simple scheme I made in figure 2.9 simulating it. As the spot crosses our field of view, going from 1 to 5, its signal affects the CCF shape. In our case, when the spot enters our field of view (2), it distorts the CCF. This distortion "evolves" until the spot is out of the field of view (5) and the CCF returns to normal.

Let us focus on the impact the crossing of the spot will have on the activity indicators. It is easy to notice in figure 2.9 that any change in the CCF will have an impact on both the FWHM (represented on the red horizontal line) and the bisector (represented by the blue vertical line).

In the case of the FWHM, if from the CCF represented by (1) has the value

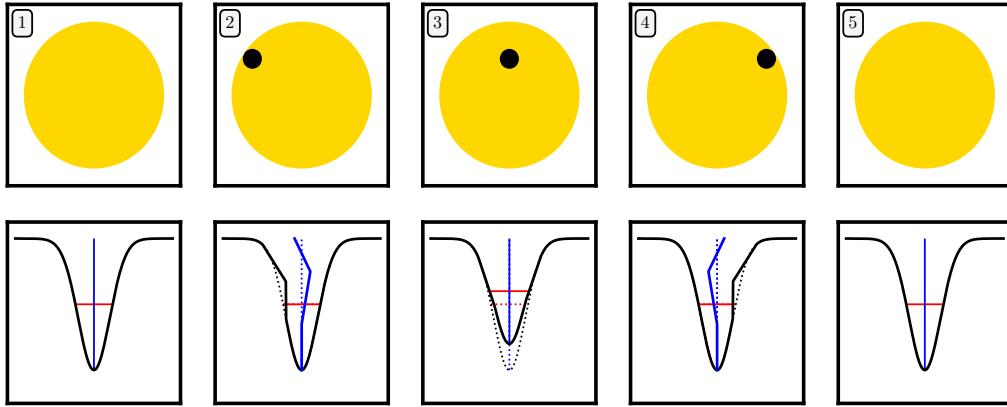


Figure 2.9: Simplified illustration of the impact of a spot on the CCF.

y_1 . The asymmetry on (2) does not make the FWHM change position, but its value will be smaller than y_1 . On pane (3), the spot affects the peak of the CCF and, consequently, the FWHM changes position. Considering this is the only effect on the CCF, the value should be bigger than y_1 . As the CCF is a Gaussian curve, the width where the FWHM is, is bigger than before. On pane (4), the spot makes the FWHM decrease to a value equal to (2). As the spot exits the stellar disk, the FWHM increases to the value y_1 on the pane (5). This is roughly plotted on figure 2.10. This effect occurs due to non-planetary signals as a planet would only shift all of the CCF from the 0 m/s mark.

The previous paragraph was a simplification of the impact of a spot. However, it is paramount to understand that, as the star rotates, a stellar spot enters and leaves our line of sight. That, consequently, deforms the CCF. Due to the rotation of the star, a periodicity will appear. That periodic signal will have a value close to the rotation period of the star. Using that information, we can associate similar periodic signals on the RV and reject them as activity-related. Similarly, we can try to interpret the bisector behaviour of figure 2.9 and associate its changes with the impact of the spot. Is this that allows the creation of time-series like the bisector inverse slope.

It is the information contained on these other time-series that, for example, Rajpaul et al. (2015) and Jones et al. (2017) use to improve planetary detection.

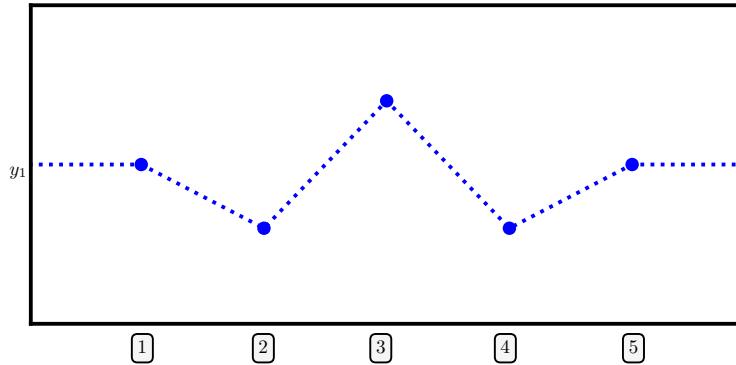


Figure 2.10: Expected behavior of the FWHM measurements on the example I gave on figure 2.9. On the x-axis are represented the five panes.

Similarly, in my thesis, I create a new framework using Gaussian processes that use auxiliary time series to analyse RV measurements. Next, I will define some of the activity proxies time series we can use.

Bisector inverse slope

The bisector inverse slope (BIS) is an asymmetry measure of the CCF. We can find an early use of it on, for example, on Gray (1997). The authors suggested using the overall slope of the bisector, measured between bisector points at 0.50 and 0.90 of the continuum. As these measures will be different between the top and bottom parts, it is sensitive to any signal contaminating the line profile of the CCF. (Gray and Hatzes, 1997).

Improvements to the original definition of the BIS have been developed. Queloz et al. (2001) used the difference between the top and bottom regions of the bisector, \overline{V}_t and \overline{V}_b respectively. That increased the accuracy of the BIS by defining it as

$$BIS = \overline{V}_t - \overline{V}_b. \quad (2.36)$$

One can define different boundaries for the top and bottom regions. That, however, requires some caution when comparing results (Queloz et al., 2001;

Figueira et al., 2013; Santerne et al., 2015). As with the FWHM, it is possible to connect the contamination of stellar activity in our data to the measures obtained for the BIS. Similarly, I will explore this better up ahead.

V_{span}

The bisector velocity span (V_{span}) is an asymmetry measurement of the CCF proposed by Boisse et al. (2010). We can describe the V_{span} as a two-step fit of the CCF, using two Gaussians. First, we fit the top part of the CCF (RV_{high}) and the lower part (RV_{low}). We can then calculate the V_{span} as

$$V_{span} = RV_{high} - RV_{low}. \quad (2.37)$$

This proxy is less sensitive to noise, meaning this measurement can be advantageous over the BIS for observations with low S/N (Figueira et al., 2013).

V_{asy}

This indicator was developed by Figueira et al. (2013) and is known as V_{asy} . It compares the RV information of a spectral line to asymmetries between the red and the blue wings of the spectral line. On it, we compute the spectral information between the red and the blue wings of the line profile

$$V_{asy} = \frac{\sum_{flux} (W_i(red) - W_i(blue)) \times \overline{W}_i}{\sum_{flux} \overline{W}_i}, \quad (2.38)$$

where W_i is the weight for the point calculated at the flux level i. To address some limitations, Figueira et al. (2015) defined W_i as

$$W_i = c^2 \frac{\left(\frac{\partial CCF(i)}{\partial RV(i)} \right)^2}{\sigma_{CCF}(i)^2}. \quad (2.39)$$

$\sigma_{CCF}(i)$ is the error on the RVs induced by the flux measurement in pixel i, and c is the speed of light in the vacuum.

This indicator shows significant correlations with RV for active stars. It is better correlated with the RV for signals caused by line deformations, making it more advantageous than the BIS (Figueira et al., 2013).

S index

The S index is a magnetic activity proxy developed for the HK Project by the Mount Wilson Observatory (Vaughan et al., 1978). Created from the regularly observed CaII H and K emission lines of bright dwarf stars to characterize their magnetic activity (Wilson, 1978). We can derive it as

$$S = \alpha \frac{N_H + N_K}{N_R + N_V}, \quad (2.40)$$

where N_H and N_K are the fluxes in 1.09Å triangular bands centred on the CaII H and K lines, N_R , N_V are 20Å reference bandpasses in the continuum, and α is a calibration constant determined nightly (Vaughan et al., 1978).

Some limitations in this index led to the development of the $\log(R'_{HK})$.

$\log(R'_{HK})$

The R'_{hk} index was developed by Noyes et al. (1984). This index uses the fluxes of the H and K lines cores to calculate a star's bolometric luminosity radiated as chromospheric H and K emission (Hall, 2008). It is defined as

$$R'_{HK} = \frac{F_H + F_K}{\sigma T^4}, \quad (2.41)$$

where F_H and F_K are respectively the fluxes of the H and K lines subtracted by the chromospheric fluxes of a reference star, σ the Stefan-Boltzmann constant and T the star's temperature (Martínez-Arnáiz, R. et al., 2010). Most commonly this index is used in its logarithmic form $\log(R'_{hk})$.

This index is widely used to identify the level of activity of a star. Generally the boundary between active and inactive lies at $\log(R'_{hk}) = -4.75$, a limit known

as Vaughan-Preston gap (Vaughan and Preston, 1980). This implying more active stars will have a $\log(R'_{hk}) > -4.75$, and less active a $\log(R'_{hk}) < -4.75$.

Chapter 3

Gaussian Process Regression Network

In this chapter, we present the main mathematical background for Gaussian processes analysis. We follow it with the presentation of some frameworks currently used on radial velocity analysis. The chapter concludes with a detailed explanation of the framework used on this PhD project. We show the principles behind a Gaussian process regression network and how to use it on exoplanet search.

3.1 Gaussian processes

Currently Gaussian processes is one of the most advanced fields in statistics, used in various fields such as Finance (e.g. Nirwan and Bertschinger, 2020), pharmacology (e.g. Sahli Costabal et al., 2019), and more recently COVID-19 outbreaks prevention (e.g. Ketu and Mishra, 2020). This tool also proved to be successful in astronomy in fields such as cosmology (e.g. Melia and Yennapureddy, 2018), and in exoplanet detection to model the stellar activity (e.g. Faria et al., 2016; Serrano et al., 2018; Barros et al., 2020). A GP is often seen as an esoteric tool that researchers use because it works, while the user ignores the mathematical background. Hoping to change that, I will start with a brief introduction to the Gaussian

distribution before diving into a GP per se. This chapter will only be a brief introduction to GPs and their properties. I recommend the reader to Rasmussen and Williams (2006) for a more detailed approach.

3.1.1 The Gaussian distribution

In nature, many random processes follow a Gaussian probability distribution function¹ This function is known for its "bell" shape and has a probability density function² (PDF) given by

$$p(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \quad (3.1)$$

This 1-dimensional (1-d) Gaussian distribution is thus defined by a mean μ and standard deviation σ (Patel and Read, 1996). That and its "bell shape" means it will have its maximum when $x = \mu$, and width proportional to σ (Sivia and Skilling, 2006). In figure 3.1 it is shown that is a symmetric distribution that shifts on the x-axis accordingly to its mean and increases or decreases in width with the value of its standard deviation. If a random variable x follows a Gaussian distribution with a mean μ and standard deviation σ , we can write it as

$$x \sim \mathcal{N}(\mu, \sigma^2). \quad (3.2)$$

The same way we define a 1-d Gaussian distribution, we can represent a 2-d, 3-d, or n-d with $n \in \mathbb{N}^+$ multivariate Gaussian distribution. This distribution has a PDF defined as

$$p(X|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2} (X - \mu)^T \Sigma^{-1} (X - \mu)\right) \quad (3.3)$$

If $X = [X_1, X_2, \dots, X_k]$ is a n-dimensional random vector, then μ is now defined as a mean vector of size n, and Σ as a $n \times n$ covariance matrix contain-

¹ We can also label a Gaussian distribution as a *Normal distribution*.

² The probability density function of a continuous random variable, is a function that returns us the probability that a random variable X , will have the value x (Grinstead and Snell, 2012).

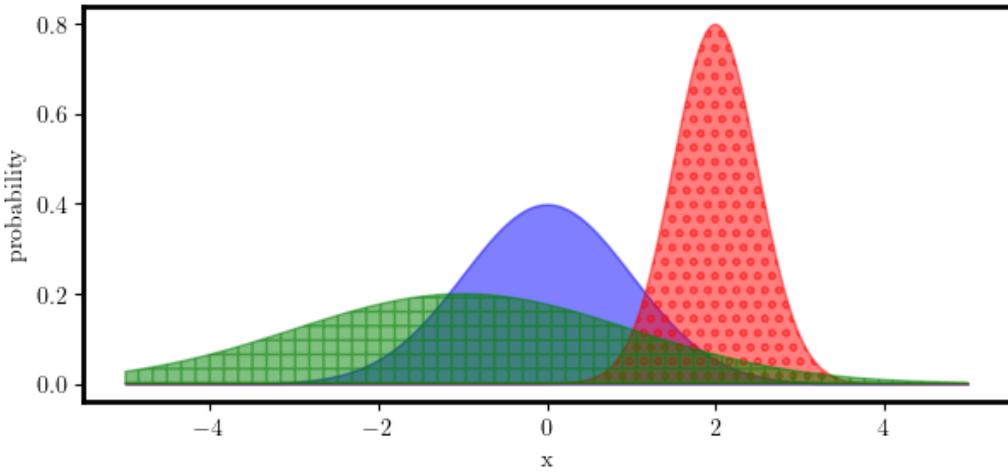


Figure 3.1: Three 1-dimensional Gaussian distributions. At solid blue a distribution with $\mu = 0$ and $\sigma = 1$, at dotted red one with $\mu = 2$ and $\sigma = 0.5$, and at checkered green one with $\mu = -1$ and $\sigma = 2$.

ing the covariance between each pair of elements of the n-dimensional random vector (Gut, 2009). Similarly to the 1-d case, if a random vector X follows a multivariate Gaussian distribution with a mean μ and covariance Σ , we can write it as

$$X \sim \mathcal{N}(\mu, \Sigma). \quad (3.4)$$

3.1.2 The Gaussian process

Now that we defined the multivariate Gaussian distribution, we can look at a GP as its generalization. It is characterized by a mean function $m(t)$ and a covariance function $k(t, t')$ ³.

A Gaussian process provides a useful Bayesian tool for modelling using a

³ Note that, for here on, I will use t as input instead of the usual x . That is due to, in my work, the input usually defines the time of a given observation.

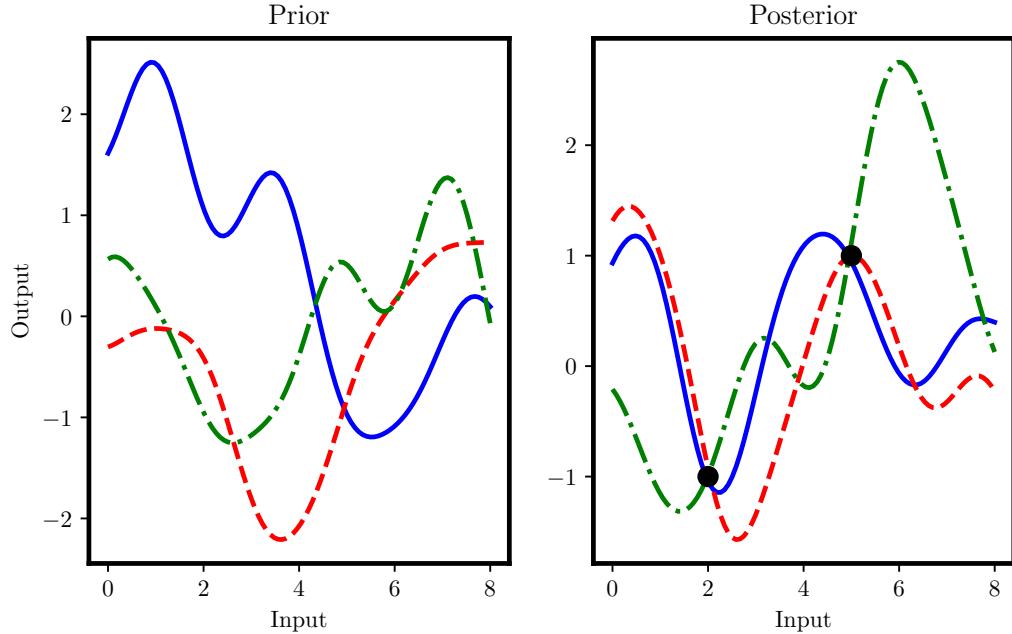


Figure 3.2: Left: Three samples drawn from a Gaussian process prior. Right: Three samples drawn from the conditional posterior distribution after two function values have been observed.

non-parametric model. On a GP, the parameters we are trying to learn are functions of the input, on which we first place a prior distribution and later condition on observed data. The combination of the prior distribution and the data leads, by using the Bayes theorem, to a posterior distribution for the GP (see figure 3.2).

Formally one can define a GP $f(t)$ as a collection of random variables, any finite number of which have a joint Gaussian distribution (Rasmussen and Williams, 2006). We can write the GP as

$$f(t) \sim \mathcal{GP}(m(t), k(t, t')), \quad (3.5)$$

where we have

$$\begin{aligned} m(t) &= E[f(t)], \\ k(t, t') &= E[(f(t) - m(t))(f(t') - m(t'))], \end{aligned} \quad (3.6)$$

where E is the expected value.

As previously mentioned, a GP is a non-parametric model. While in a parametric setting, we know the exact model we will fit the data (e.g. linear regression). In a non-parametric setting, the model structure is determined by the data instead. In exoplanet search, this flexibility allows astronomers to not only constrain a model to the data but to deal with the correlated stellar activity component of the measurements (e.g. Haywood et al., 2014; Faria et al., 2016; Cloutier, R. et al., 2019).

While a GP, for simplicity, is usually defined as having a mean of zero, in the exoplanet context, the mean function is often a parametric model of the observed phenomena. In the case of radial velocities, the most common mean used is a sum of keplerian functions similar to equation 1.1.

With a keplerian function on the mean, the kernel to be the element in a GP that will deal with the stellar activity component present in the data while the mean function model the planetary signals. Since this activity is correlated and quasi-periodic, the most used kernel is known as the quasi-periodic kernel, see section 3.1.3 for a more detailed description of this kernel. This kernel became the go-to of exoplanet search using GPs due to its parameters often associated with physical properties of the stellar activity like the stellar rotation and decaying timespan of the active regions.

After choosing the mean and covariance functions, often parameterized by so-called hyperparameters, the next step is to optimize all the parameters in our model conditioning on observed data. That can be done by either maximizing the GP marginal likelihood or sampling from the posterior distribution for the hyperparameters using MCMC or other methods. Either Minimizing or maximizing objective functions or using Markov chain Monte Carlo (MCMC) methods to obtain samples from the parameter space.

To optimize the hyperparameters, one typically chooses from a set of minimization algorithms such as the Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1963) or the L-BFGS (Liu and Nocedal, 1989) algorithms. A convenient

implementation of several of such algorithms is available on the `Scipy` python library⁴ (Virtanen et al., 2020). Nevertheless, minimizing (maximizing) an objective function is challenging. Many methods are sensitive to the existence of local minima (maxima) or the starting guess. That means that the final result provided by the algorithm might not correspond to the globally optimal values of the hyperparameters.

In terms of sampling methods, we often rely on different MCMC algorithms such as those implemented in the `emcee` (Foreman-Mackey et al., 2019), `dynesty` (Speagle, 2020), and `zeus` (Karamanis et al., 2021) packages. `emcee` uses an affine invariant MCMC ensemble sampler developed by Goodman and Weare (2010) useful to sample badly scaled distributions but not well suited for high dimensional or multi-modal probability surfaces. `dynesty` uses a nested sampling method developed by Higson et al. (2019) capable of calculating Bayesian posteriors and marginal likelihoods (or model evidence). However, like other nested sampling methods, it becomes inefficient when exploring the parameter space in progressively higher likelihoods. Lastly, `zeus` implements the ensemble slice sampling method, an extension of slice sampling developed by Neal (2000) capable of sampling from highly correlated distributions.

Independently of the method used, when we try to optimize the model parameters is common to use the log marginal likelihood ($\log \mathcal{L}$) as an objective function. This quantity takes the form of

$$\log \mathcal{L} = -\frac{1}{2} (y - m)^T K^{-1} (y - m) - \frac{1}{2} \log |K| - \frac{n}{2} \log (2\pi), \quad (3.7)$$

where $y - m$ is the residuals (the subtraction of measurements by the mean function), K the covariance matrix of the GP, and n the number of measurements. The three terms in equation 3.7 have the following interpretation. The first term $\left(-\frac{1}{2} (y - m)^T K^{-1} (y - m)\right)$ is the only one which includes the data y and corresponds to a measure of the goodness-of-fit. The second term $\left(-\frac{1}{2} \log |K|\right)$ will

⁴ Since the computational part of my work uses the Python programming language, I will limit the examples given to this language. Other programming languages exist where similar algorithms are available.

measure and penalize models based on the determinant of the covariance matrix⁵. The remainder is only a normalization constant (Rasmussen and Williams, 2006).

The absolute value of the log marginal likelihood value alone has no interpretation. Its advantage comes when used in a model with two different sets of parameters values. The higher the likelihood, the better the model with those parameters represents the data. For example, a set of parameters whose log marginal likelihood is -100 are better than a set with a log marginal likelihood of -200.

For a complete Bayesian analysis, we must also define prior distributions for the hyperparameters. Then, the fully marginalized likelihood, or evidence, represented as \mathcal{Z} , can be calculated and used to compare different models. This measurement acts as a normalization factor in parameter estimation (Sivia and Skilling, 2006). We define the evidence as

$$\mathcal{Z} \equiv p(D | M_i) = \int p(\theta | M_i) p(D | \theta, M_i) d\theta, \quad (3.8)$$

where M_i is our model, D the observed data, and θ the parameters vector. In practice, this quantity is challenging to compute, but approximations are known, such as the one derived by Perrakis et al. (2014).

From the evidence of two different models M_1 and M_2 , we can then obtain the likelihood ratio known as Bayes factor (K_B)

$$K_B = \frac{\mathcal{Z}_{M_1}}{\mathcal{Z}_{M_2}}. \quad (3.9)$$

We then use the natural logarithm of the Bayes factor as model selection criteria, often assuming equal prior probabilities for the two models (Jeffreys, 1983; Kass and Raftery, 1995). For example, when comparing a 0-planet model against a 1-planet model, a Bayes factor threshold of 150 indicates that there is enough evidence to claim a planetary detection (e.g. Faria et al., 2016; Faria et al., 2020).

⁵ We can call this term as the *Occam's Razor* in GP model selection. Named after William of Occam (1285-1347) famous for the quote "The explanation requiring the fewest assumptions is most likely to be correct".

3.1.3 The Covariance function

The covariance function, most commonly known as the kernel, is the most relevant element in a GP. It is the kernel that controls the characteristics of the functions under the GP prior. It becomes vital to choose a covariance function appropriate to the properties present in the data (e.g. periodicity, smoothness, or stationarity).

If a kernel is a function of only $t - t'$, we say it is *stationary*. Stationarity implies invariance of the kernel to translations in the input, meaning that the kernel value remains the same independently of how the inputs are shifted t (Duvenaud, 2014). If the kernel is a function of $|t - t'|$ we classify it as *isotropic* whose value does not depend on direction (Rasmussen and Williams, 2006).

A kernel $k(t, t')$ has a corresponding Gram matrix K containing the covariance between any two points t and t' . This Gram matrix will be a positive semidefinite matrix, that is, a symmetric matrix with non-negative eigenvalues (Rajpaul, 2017). Taking into consideration the set of input points $\{t_1, t_2, \dots, t_n\}$, K will take the form

$$K = \begin{bmatrix} k(t_1, t_1) & k(t_1, t_2) & \dots & k(t_1, t_n) \\ k(t_2, t_1) & k(t_2, t_2) & \dots & k(t_2, t_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(t_n, t_1) & k(t_n, t_2) & \dots & k(t_n, t_n) \end{bmatrix}. \quad (3.10)$$

Since this Gram matrix contains the covariance of the input points, we call it a covariance matrix.

We can now look at two of the most important properties of a GP kernel, addition and multiplication. You can find more on these in Rasmussen and Williams (2006). The two properties are:

- The sum of two kernels $k_1(t, t')$ and $k_2(t, t')$ is a valid kernel. This means that if we have two independent GPs $f_1(t) \sim \mathcal{GP}(0, k_1(t, t'))$ and $f_2(t) \sim \mathcal{GP}(0, k_2(t, t'))$, we can create a new GP $g(t)$

$$g(t) \sim \mathcal{GP}(0, k_1(t, t') + k_2(t, t')). \quad (3.11)$$

- The product of two kernels $k_1(t, t')$ and $k_2(t, t')$ is also a valid kernel. Again, if we have two independent GPs $f_1(t) \sim \mathcal{GP}(0, k_1(t, t'))$ and $f_2(t) \sim \mathcal{GP}(0, k_2(t, t'))$, we can thus created a GP $h(t)$ defined by

$$h(t) \sim \mathcal{GP}(0, k_1(t, t') \times k_2(t, t')). \quad (3.12)$$

These properties allow the creation of new kernels capable of modelling data with a more complex structure. They are fundamental if we want to determine the best covariance functions to use in RV data analysis. See Duvenaud (2014) for more detailed work on the structures obtained by the sum and multiplication of different kernels.

In the following pages, I present the main kernels of interest in GP analysis. Each covariance function is characterized by a set of parameters, also called hyperparameters in some works. I represent them using the Greek alphabet plus an index. To avoid duplicating parameter definitions, for different kernels, the parameter represented by the same letter and index share the same interpretation unless I mention it otherwise on the text.

Squared-exponential kernel

The first kernel I present is the *squared-exponential kernel*⁶, one of the most widely used kernels in machine learning. It is given by

$$k(t, t') = \eta_1^2 \exp\left(-\frac{(t - t')^2}{2\eta_2^2}\right), \quad (3.13)$$

where η_1 represents the amplitude of the signal and η_2 is a parameter called *length-scale*. One way of interpreting this length-scale is as being the time required for the function to change significantly (Rasmussen and Williams, 2006). The higher this parameter, the smoother the squared-exponential kernel becomes, see figure 3.3. On the limit, if $\eta_2 \rightarrow \infty$ the squared exponential kernel becomes a constant covariance function. Although widely used, this kernel is known to be too smooth

⁶ The squared-exponential kernel is commonly called as the radial basis function or RBF kernel in machine learning.

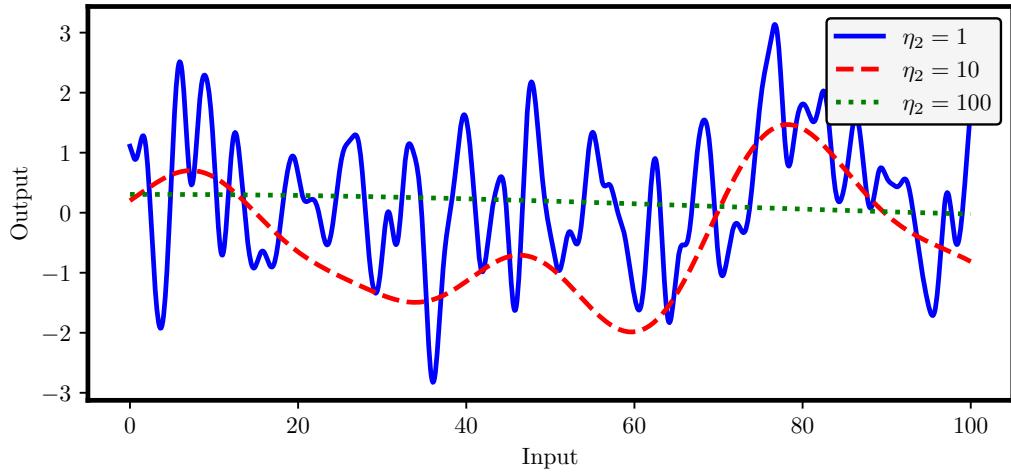


Figure 3.3: Samples of a squared-exponential kernel with different length-scales. Solid blue $\eta_2 = 1$, dashed red $\eta_2 = 10$, and dotted green $\eta_2 = 100$.

to realistically model a considerable number of physical processes (Stein, 1999). As a consequence, I will next present two alternatives to the squared-exponential.

Rational quadratic kernel

A possible replacement of the squared exponential kernel is the *rational quadratic kernel* given by

$$k(t, t') = \eta_1^2 \left(1 + \frac{(t - t')^2}{2\alpha\bar{\eta}_2^2} \right)^{-\alpha}. \quad (3.14)$$

While a squared exponential assumes an underlying function that describes the data evolving at one particular time scale, it might be beneficial to consider a sum of squared exponential kernels, each with a different time scale. The rational quadratic allows this. This kernel is an infinite sum of squared exponential kernels with distinct time scales, controlled by the parameter α . We derive this kernel by expressing the scale mixture of different squared exponential covariance functions through a gamma distribution with shape α and mean $\bar{\eta}_2^{-2}$ (Wilson, 2014).

This derivation leads to $\bar{\eta}_2$ to be the mean (thus the bar over this term) time scale of the infinite sum of different squared exponential kernels, each with a different decaying timescale. When $\alpha \rightarrow \infty$, the rational quadratic kernel converges to the squared exponential kernel, while it converges to a constant when $\alpha \rightarrow 0$ (Rasmussen and Williams, 2006).

Matérn family of kernels

The Matérn family of covariance functions are defined as

$$k(t, t') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2}(t-t')}{\eta_2} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu}(t-t')}{\eta_2} \right), \quad (3.15)$$

where Γ is the gamma function (Artin and Butler, 2015), K_ν is the modified Bessel function of the second kind (Abramowitz and Stegun, 1964), and ν is a positive parameters. If $\nu \rightarrow \infty$ we obtain the squared exponential kernel (Rasmussen and Williams, 2006).

Of this family is of interest to consider ν as a half-integer. The three most common in machine learning are known as the *Matérn 1/2*⁷, *Matérn 3/2* and the *Matérn 5/2* kernels. This last one used in RV analysis by Gilbertson et al. (2020). They are defined as

$$k(t, t') = \eta_1^2 \left(-\frac{(t-t')}{\eta_2} \right), \quad (3.16)$$

$$k(t, t') = \eta_1^2 \left(1 + \frac{\sqrt{3}(t-t')}{\eta_2} \right) \exp \left(-\frac{\sqrt{3}(t-t')}{\eta_2} \right), \quad (3.17)$$

and

$$k(t, t') = \eta_1^2 \left(1 + \frac{\sqrt{5}(t-t')}{\eta_2} + \frac{5(t-t')^2}{3\eta_2^3} \right) \exp \left(-\frac{\sqrt{5}(t-t')}{\eta_2} \right), \quad (3.18)$$

⁷This kernel is also called as exponential kernel in other works.

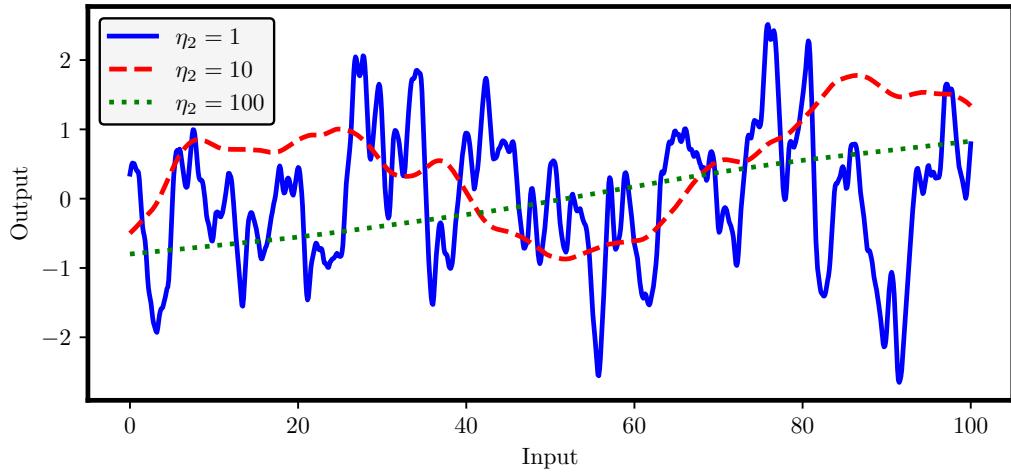


Figure 3.4: Samples of a Matérn 5/2 kernel with different length-scales. Solid blue $\eta_2 = 1$, dashed red $\eta_2 = 10$, and dotted green $\eta_2 = 100$.

respectively.

The η_2 parameters share the same interpretation as the one given for the squared exponential kernel. Some works use the Matérn kernels as a solution to deal with the smoothness of the squared exponential kernel (e.g. Abdessalem et al., 2017). For comparison, and given the same parameters, we see from samples taken with a Matérn 5/2 kernel (figure 3.4), it is far less smooth than the ones drawn for a squared exponential kernel (figure 3.3).

Periodic kernel

Another kernel of interest is the *periodic kernel* used to model data that, as the name implies, has a periodical behaviour. Usually, it is written as

$$k(t, t') = \eta_1^2 \exp \left[-\frac{2}{\eta_4^2} \sin^2 \left(\frac{\pi (t - t')}{\eta_3} \right) \right]. \quad (3.19)$$

This kernel have two new parameters. The period of the repetitions η_3 and a length-scale η_4 that models the structure inside a period. Despite the fact that it is

widely used, its origins are sometimes ignored. Mackay (1998) derived this kernel by mapping (or warping) the squared exponential kernel into the two-dimensional space $u(t) = \left[\cos\left(\frac{2\pi}{\eta_3}t\right), \sin\left(\frac{2\pi}{\eta_3}t\right) \right]$.

The derivation is of interest to show. To differentiate the periodic from the squared exponential kernel, I will name them k_P and k_{SE} , respectively. We start from the squared exponential kernel of equation 3.13 and map it into the two-dimensional space⁸. That implies that the periodic kernel we want is

$$k_P(t, t') = k_{SE}(u(t), u(t')).$$

With the mapping of before we end up with

$$k_P(t, t') = \exp\left(-\frac{1}{2\eta_2^2} z^\top z\right) \quad \text{with } z = \begin{bmatrix} \cos t - \cos t' \\ \sin t - \sin t' \end{bmatrix}.$$

This allow us obtain a kernel given by

$$k_P(t, t') = \exp\left[-\frac{(\cos t - \cos t')^2 + (\sin t - \sin t')^2}{2\eta_2^2}\right].$$

Using the binomial theorem (Abramowitz and Stegun, 1964) we can expand the squared terms into

$$\begin{aligned} (\cos t - \cos t')^2 &= \cos^2 t - 2 \cos t \cos t' + \cos^2 t', \\ (\sin t - \sin t')^2 &= \sin^2 t - 2 \sin t \sin t' + \sin^2 t', \end{aligned}$$

allowing us to use the Pythagorean trigonometric identity $\cos^2 X + \sin^2 X = 1$, to simplify $k(t, t')$ into

$$k_P(t, t') = \exp\left(-\frac{2 - 2 \cos t \cos t' - 2 \sin t \sin t'}{2\eta_2^2}\right).$$

⁸To simplify the notation, I will omit $\frac{2\pi}{\eta_3}$.

We now make use of the prosthaphaeresis formulas (Kung, 1996)

$$2 \cos X \cos Y = \cos(X - Y) + \cos(X + Y), \\ 2 \sin X \sin Y = \cos(X - Y) - \cos(X + Y),$$

and transform the kernel into

$$k_P(t, t') = \exp \left[-\frac{1 - \cos(t - t')}{\eta_2^2} \right].$$

The last steps require the use of the double-angle formula $\cos 2X = \cos^2 X - \sin^2 X$ to give us

$$k_P(t, t') = \exp \left[-\frac{1 - \cos^2 \left(\frac{t-t'}{2} \right) + \sin^2 \left(\frac{t-t'}{2} \right)}{\eta_2^2} \right].$$

Using again the Pythagorean trigonometric identity we obtain

$$k_P(t, t') = \exp \left[-\frac{2}{\eta_2^2} \sin^2 \left(\frac{t - t'}{2} \right) \right].$$

Thus ending the derivation. Plugging back in the $\frac{2\pi}{\eta_3}$ term we obtain the expected periodic kernel of equation 3.19

$$k_P(t, t') = \exp \left[-\frac{2}{\eta_2^2} \sin^2 \left(\frac{\pi}{\eta_3} (t - t') \right) \right].$$

This derivation allows me to mention two details. First, the η_4 of the periodic kernel and the η_2 of the squared exponential kernel are the same or should share the same characteristics. That shows that η_4 is a length-scale that will determine the smoothness, or lack of it, inside a period is given by η_3 (see figure 3.5). The higher the value of η_4 , the smoother is the structure within one period. In other words, as $\eta_4 \rightarrow \infty$, the periodic kernel tends to a smooth sinusoid. The other important

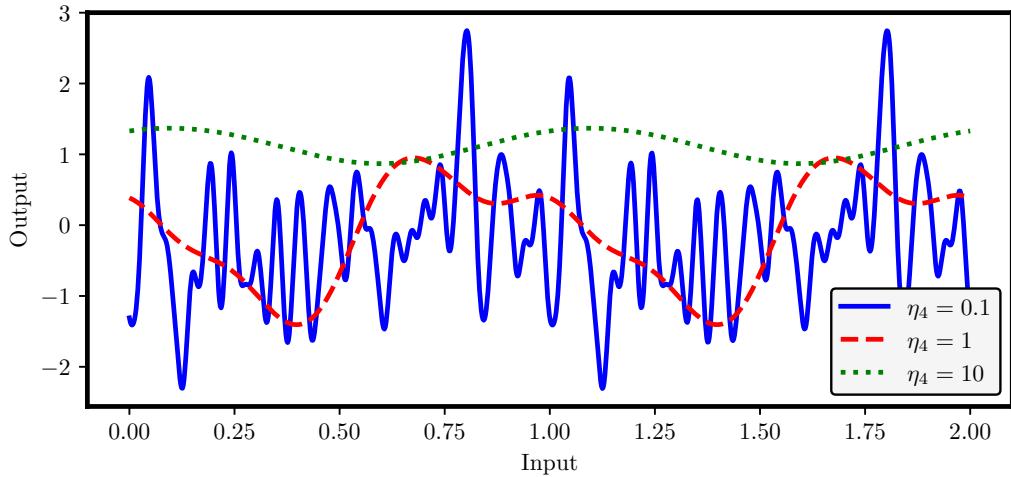


Figure 3.5: Samples of a periodic kernel with $\eta_3 = 1$ but varying η_4 . Solid blue $\eta_4 = 0.1$, dashed red $\eta_4 = 1$, and dotted green $\eta_4 = 10$.

detail the possibility of creating new periodic kernels using the same mapping but starting from different covariance functions. I will explore this further ahead.

Quasi-periodic kernel

The most common kernel in exoplanet search is called *quasi-periodic kernel*. Since the signal induced by the active regions on an RV set is not strictly periodic, its shape evolves. We thus need to use a quasi-periodic function capable of having this into consideration. We attain this by using the property discussed in equation 3.12 and multiplying two "simpler" kernels, the periodic and the squared-exponential kernels. This multiplication creates a model that decays away from the periodicity, allowing a locally periodical structure (Duvenaud et al., 2013). We define the quasi-periodic kernel as

$$k(t, t') = \eta_1^2 \exp \left[-\frac{(t - t')^2}{2\eta_2^2} - \frac{2}{\eta_4^2} \sin^2 \left(\frac{\pi}{\eta_3} (t - t') \right) \right]. \quad (3.20)$$

In this covariance function, we can interpret η_2 as the decaying time scale

of the active regions on the surface of a star. η_3 gives the rotation period of the observed star, while η_4 is a parameter that controls the sinusoidal variations of the periodic signal, that is, the structure within η_3 . When analysing RV data, we also add a white noise term σ to account for any uncorrelated noise present in the data⁹.

Piecewise polynomial kernel

The Wendland compactly supported functions are a family of piecewise polynomial functions with compact support (Wendland, 2004). By definition, a piecewise polynomial kernel with compact support is a kernel that allows the covariance between two points to be zero when its distance exceeds an imposed threshold, creating a sparse covariance matrix in the process (Rasmussen and Williams, 2006). This characteristic implies that by multiplying a *piecewise kernel* by a squared exponential kernel, we obtain a new covariance function suited to model short-scale correlations. In this family, increasing the polynomial order creates increasingly smoother kernels (Davies, 2015). We can, for example, use a third-order piecewise polynomial and define it as

$$k(t, t') = \eta_1^2 \exp\left(-\frac{(t - t')^2}{2\eta_2^2}\right) (3r_p + 1)(1 - r_p)^3, \quad (3.21)$$

where we have $r_p = 0$ if $|t - t'| > \frac{\eta_3}{2}$ and $r_p = (t - t')$ otherwise. I chose to show a third order polynomial due to having a similar behavior to a squared exponential for $0 < \frac{2|r|}{\eta_3} < 1$. It will effectively truncate the squared exponential to small length correlations.

One challenge in the analysis of RV data with a *QP* kernel is the effect that small length correlations might have. These might be responsible for lower η_2 or higher η_4 values than expected, as these parameters try to model correlated noise variations. Having defined this new kernel we can add it to a quasi-periodic kernel. That means having a new covariance function defined as

⁹It is usual to include a white noise term in most GP analysis but, I have omitted this term for simplicity.

$$\begin{aligned} k(t, t') = & \eta_1^2 \exp \left[-\frac{(t-t')^2}{2\eta_2^2} - \frac{2}{\eta_4^2} \sin^2 \left(\frac{\pi}{\eta_3} (t-t') \right) \right] \\ & + \theta_1^2 \exp \left(-\frac{(t-t')^2}{2\theta_2^2} \right) (3r_p + 1)(1-r_p)^3. \end{aligned} \quad (3.22)$$

The idea behind this new kernel is to allow the quasi-periodic term to model other more significant signals while the piecewise part focuses on the 2 to 3 days length variations.

From rational quadratic to periodic

As mentioned earlier, the derivation of the periodic kernel opens the door to the creation of new periodic kernels, and as consequence, new quasi-periodic kernels. Following similar steps we can achieve this using the rational quadratic kernel. In this case we map the rational quadratic kernel into the two-dimensional space $u(t) = [\cos(\frac{2\pi}{\eta_3}t), \sin(\frac{2\pi}{\eta_3}t)]$. Starting from

$$k(t, t') = \left[1 + \frac{(t-t')^2}{2\alpha\bar{\eta}_2^2} \right]^{-\alpha},$$

we obtain

$$k(t, t') = \left[1 + \frac{(\cos t - \cos t')^2 + (\sin t - \sin t')^2}{2\alpha\bar{\eta}_2^2} \right]^{-\alpha}.$$

We need to work a bit on this new kernel. We need to use the expansion of the squared terms, followed by the Pythagorean identity, prosthaphaeresis and double-angle formulas. In the end, we get

$$k(t, t') = \left[1 + \frac{2}{\alpha\bar{\eta}_2^2} \sin^2 \left(\frac{\pi}{\eta_3} (x-x') \right) \right]^{-\alpha}. \quad (3.23)$$

This new kernel has some interesting characteristics. In the rational quadratic kernel, η_2 was the mean time-scale of the infinite sum of squared exponential

kernels controlled by α . Similarly, when α goes to ∞ , this new kernel becomes the periodic kernel from 3.19. There is an advantage of creating a periodic kernel from a rational quadratic. The new covariance function is differentiable in mean square (MS differentiable) for every α . That, for example, does not occur with the Matérn kernels (Rasmussen and Williams, 2006). This implies, for example, $\frac{\partial^2 k(t,t')}{\partial t \partial t'}$ exists (Grigoriu, 2002).

3.2 Gaussian processes frameworks

The stellar activity features mentioned on chapter 2.3.1 give rise to complex activity signals. For example, in the case of the Sun, the magnetic activity dynamics gives rise to non-stationary processes not easily modeled (e.g. Demin et al., 2018). Similar magnetic dynamics is to be expected in other stars, creating similar non-stationary processes on its stellar activity. These processes will reflect on the RV spectroscopic observations. Modelling RV measurements plagued by stellar activity signals non-stationary in nature created even bigger challenges.

The idea of using the information contained in activity indicators to improve the characterization of the signals on RV measurements is not new. Other frameworks already exist that extend the use of GPs, allowing the introduction of the information contained in secondary time series. These proved so far to be an improvement to a standard GP regression framework using only RVs. However, both the traditional GP regression and multivariate Gaussian process models do not tackle the non-stationary behaviour of the stellar activity. That occurs due to the models assuming the parameters stay constant over the whole input space. That implies, for example, to consider a fixed decaying timescale over the entire input space (Plagemann et al., 2008).

Next, I present three GP frameworks that, while not non-stationary, have been developed to improve the characterization of RV data using auxiliary time series. That, however, implies that a non-stationary framework, such as the one I will soon present in this thesis, is an innovation key to explore.

3.2.1 Rajpaul et al. (2015) framework

This first framework I show evolved as an extension of the FF' framework (Aigrain et al., 2012). The framework includes the $\log(R'_{hk})$ index and the BIS to disentangle the activity component from the planetary signals. We then express the RVs and activity indicators as a linear combination of GPs given by

$$\begin{cases} \Delta RV = V_c G(t) + V_r \dot{G}(t), \\ \text{BIS} = B_c G(t) + B_r \dot{G}(t), \\ \log(R'_{HK}) = L_c G(t). \end{cases} \quad (3.24)$$

V_c, V_r, B_c, B_r , and L_c are free parameters that will control the amplitudes of the GPs. $G(t)$ is a GP with kernel $k_G(t, t)$. Each time-series has a respective mean function as well. For the RVs Rajpaul et al. (2015) used a second-order polynomial function while for the other time-series a constant function as an offset was enough. This does not imply that other mean functions can not be implemented. The novel term $\dot{G}(t)$, is the derivative of $G(t)$. Considering that the derivative of a GP with kernel $k_G(t, t)$ is also a GP but with a different kernel. Osborne (2010) showed that for this new GP it is given by

$$k_{\dot{G}}(t, t') = \frac{\partial^2 k_G(t, t')}{\partial t \partial t'}. \quad (3.25)$$

That implies that, for example, if the kernel of $G(t)$ is not MS differentiable, we can not use it in this framework. Some recent results with this framework appear, for example, on Barragán et al. (2021).

3.2.2 Jones et al. (2017) framework

This framework further extends the work of Rajpaul et al. (2015). It considers a class of models where are used multiple stellar activity indicators. Formally it is defined as

$$\begin{cases} u(t) = m_0 + a_{01}G(t) + a_{02}\dot{G}(t) + a_{03}\ddot{G}(t) + a_{04}Z_0(t) + \epsilon_0(t), \\ q_1(t) = m_1 + a_{11}G(t) + a_{12}\dot{G}(t) + a_{13}\ddot{G}(t) + a_{14}Z_1(t) + \epsilon_1(t), \\ \vdots \\ q_p(t) = m_p + a_{p1}G(t) + a_{p2}\dot{G}(t) + a_{p3}\ddot{G}(t) + a_{p4}Z_p(t) + \epsilon_p(t). \end{cases} \quad (3.26)$$

While Rajpaul et al. (2015) included only the first derivative of a GP, Jones et al. (2017) included a additional second derivative to deal with the effects of the

projected areas of the spots. They added zero mean GPs given by $Z_0(t), \dots, Z_p(t)$ to allow the fit of other structures besides those modelled by $G(t)$ and the derivatives $\dot{G}(t)$ and $\ddot{G}(t)$ are capable of. Also included are white noise terms given by ϵ . Recently Gilbertson et al. (2020) used this framework to model stellar activity.

3.2.3 artGPN framework

Early in my PhD, I focused on modelling HARPS-like datasets containing the RV measurements (ΔRV) and three activity indicators (FWHM, BIS, and $\log R'_{HK}$). That contributed to the creation of the GP framework shown in figure 3.6. Each time series or output is an independent GP fitted using a linear combination of kernels. These will either be nodes represented as n_1, n_2, \dots, n_q or weights represented as $w_{11}, w_{22}, \dots, w_{3q}, w_{4q}$. Then, for this framework, each outputs has a covariance function defined by

$$\begin{cases} k_{\Delta\text{RV}} = k_{w_{11}} k_{n_1} + k_{w_{12}} k_{n_2} + \dots + k_{w_{1q}} k_{n_q}, \\ k_{\text{FWHM}} = k_{w_{21}} k_{n_1} + k_{w_{22}} k_{n_2} + \dots + k_{w_{2q}} k_{n_q}, \\ k_{\text{BIS}} = k_{w_{31}} k_{n_1} + k_{w_{32}} k_{n_2} + \dots + k_{w_{3q}} k_{n_q}, \\ k_{\log R'_{HK}} = k_{w_{41}} k_{n_1} + k_{w_{42}} k_{n_2} + \dots + k_{w_{4q}} k_{n_q}. \end{cases} \quad (3.27)$$

That determines that each output to be modelled by GP that is given by

$$\begin{cases} \Delta\text{RV} \sim \mathcal{GP}(m_{\Delta\text{RV}}(t), k_{\Delta\text{RV}}(t, t')), \\ \text{FWHM} \sim \mathcal{GP}(m_{\text{FWHM}}(t), k_{\text{FWHM}}(t, t')), \\ \text{BIS} \sim \mathcal{GP}(m_{\text{BIS}}(t), k_{\text{BIS}}(t, t')), \\ \log R'_{HK} \sim \mathcal{GP}\left(m_{\log R'_{HK}}(t), k_{\log R'_{HK}}(t, t')\right), \end{cases} \quad (3.28)$$

each with a mean $m_i(t)$ of any kind (constant, linear, etc...), and kernel $k_i(t, t')$ defined accordingly to the nodes and weights.

Equation 3.27 implies that this network can have a q number of nodes, each of a different kind. It can be of any kind, from a squared exponential to a periodic kernel. If necessary, the nodes can share parameters among themselves.

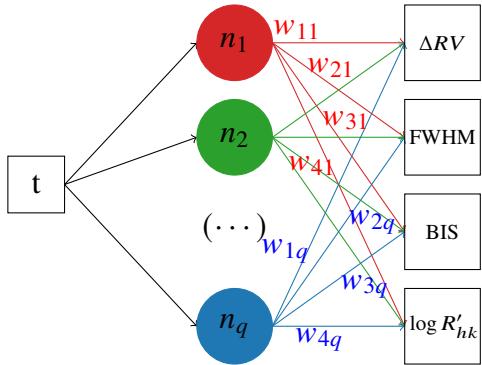


Figure 3.6: Scheme of our framework using HARPS-like datasets. Our regression network will be formed by q nodes (n_1, n_2, \dots, n_q) and $4 \times q$ weights ($w_{11}, w_{21}, w_{31}, \dots, w_{3q}, w_{4q}$) kernels connected similar to a neural network.

For example, using this framework, I could define two nodes. One being periodic, and another quasi-periodic, sharing the same η_3 but with the remaining parameters free. The weights, I usually define as constant kernels. That allows them to act as amplitudes. However, it can take any other form. We can have, for example, squared exponential weights, all with the same η_2 value but different values of η_1 .

This condition allows the fit of the same signals present in different datasets. My model matches the semi-parametric latent factor model of Teh et al. (2005) if the weights are defined to be constant kernels, meaning it only scales the magnitude of the nodes.

Depending on the output, the mean functions can take the form of a Keplerian, linear, quadratic, cubic, or sinusoidal function. The covariance functions are given by equation 3.27.

We can make use of the log marginal likelihood to evaluate this model. Assuming that each GP in the model is independent, then the total log marginal likelihood ($\log p(\mathbf{y}|t)$) is obtained by the sum of each output log marginal likelihood. Meaning we will have a log marginal likelihood given by

$$\log \mathcal{L}_{\text{total}} = \log \mathcal{L}_{\Delta RV} + \log \mathcal{L}_{\text{FWHM}} + \log \mathcal{L}_{\text{BIS}} + \log \mathcal{L}_{\log R'_{HK}}. \quad (3.29)$$

Individually we calculate the log marginal likelihood with equation 3.7. We express, for example, the $\log \mathcal{L}_{\Delta RV}$ as

$$\log \mathcal{L}_{\Delta RV} = -\frac{1}{2} \Delta RV^T K_{\Delta RV} \Delta RV - \frac{1}{2} \log |K_{\Delta RV}| - \frac{n}{2} \log 2\pi. \quad (3.30)$$

Due to its design I named this framework *artificial Gaussian process network* and a python implementation called `artGPN` is available on Github¹⁰.

¹⁰ <https://github.com/jdavidrcamacho/artgpn>

3.3 Gaussian processes regression networks

First developed by Wilson et al. (2012), a Gaussian processes regression network (GPRN) was created, unlike the previously GP frameworks, to be capable of handling non-stationary time series. While previous frameworks only account for fixed correlations among time series, a GPRN has an adaptive structure capable of capturing input dependent correlations across the input space. That occurs by combining the properties of GP regression with the properties of single-layer artificial neural networks (Rumelhart et al., 1986) to create a Bayesian model for multi-output regression.

Wilson et al. (2012) created this framework to model a P -dimensional function $y(t)$ containing signal and noise correlations that vary with t . To do this, they proposed modelling $y(t)$ as

$$y(t) = W(t) [f(t) + \sigma_f \epsilon(t)] + \sigma_y z(t). \quad (3.31)$$

On this network the $Q \times 1$ vector of latent functions (or nodes) $f(t)$ and $P \times Q$ matrix of weight functions (or weights) $W(t)$ are independent GPs defined as

$$\begin{aligned} f_q(t) &\sim \mathcal{GP}(0, k_{f_q}) \text{ for } q = 1, \dots, Q, \\ W_{pq}(t) &\sim \mathcal{GP}(0, k_{w_{pq}}) \text{ for } p = 1, \dots, P \text{ and } q = 1, \dots, Q, \end{aligned} \quad (3.32)$$

and $\epsilon(t)$ and $z(t)$ are white noise processes. This structure creates a linear combination of independent GPs, allowing an adaptive mixture of GPs suitable to deal with nonlinear correlations between the outputs. One advantage of this structure is creating heavy-tail predictive distributions that do not easily over-fit the model (Li et al., 2020).

In a GPRN, the nodes that, for simplification can be re-written as $\hat{f}(t) = f(t) + \sigma_f \epsilon(t)$ ¹¹, are connected together to create the outputs $y(t)$. What this allows is that while each node kernel has a stationary structure, the mixture of kernels $k_{\hat{f}_i}$

¹¹ One change we made to the original GPRN was to the treatment of the jitter term σ_f . We quadratically summed it with σ_y to create $\sigma = \sum_{p=1}^P \sum_{n=1}^N (\sigma_{y_{np}}^2 + \sigma_{f_p}^2)$

will not be stationary. In this structure, the overall amplitudes are defined by the weights $W(t)$. These, being also independent GPs, make the GPRN framework input dependent and non-stationary (Heinonen et al., 2015). The main difference is that, on a GP, a kernel has a constant amplitude η_1 over the input space. The same does not occur on a GPRN. On a GPRN, the kernel of each $y(t)$ is given by

$$k_{y_p}(t_a, t_b) = \sum_{q=1}^Q \left[W_{pq}(t_a) \left(k_{f_q}(t, t') + \sigma_f^2 \delta_{ab} \right) W_{pq}(b) \right] + \sigma_y^2 \delta_{ab}, \quad (3.33)$$

where δ_{ab} is the Kronecker delta. This kernel has an amplitude $W_{pq}(t)$, that being a GP, makes the overall GPRN structure non-stationary. If all the weights use the same kernel, with the same parameters, the connections in the network will vary with t at an equal rate (Wilson et al., 2012).

Another advantage of a GPRN comes from the latent functions being able to take any form. That means we can have q GPs with any kernel (periodic, quasi-periodic, squared exponential). That allows for a mixture of the different input dependent covariance functions and thus non-stationary. That allows for a framework that continuously shifts within regions of completely different covariance structures (Wilson, 2014). The overall flexible structure of the GPRN is, of course, ideal to use on non-stationary signals coming from, for example, stellar activity.

One final detail of the overall structure is that, while both the nodes $f(t)$ and weights $W(t)$ are independent GPs, their product $f(t)W(t)$ will most likely not be a GP (Rasmussen and Williams, 2006). That should not to be confused with the property 3.12 that says that the product of two kernels creates a valid kernel.

3.3.1 Variational inference

The GPRN we worked on makes use of RVs together with a $p - 1$ number of activity indicators (AI), with equal timestamps, in a scheme exemplified in the diagram of figure 3.7. The main objective of this framework is to use the information contained in the activity proxies to better separate activity from planetary

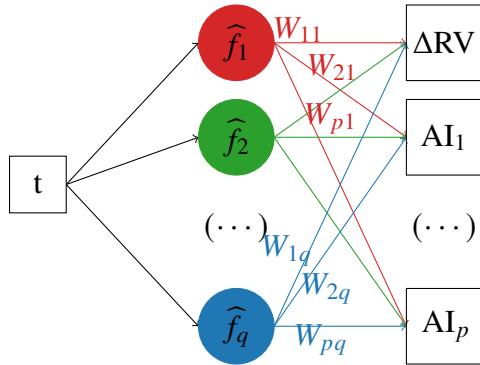


Figure 3.7: Example of a GPRN diagram considering a typical RV time-series with P activity indicator (AI). The regression network will be formed by Q nodes $[\widehat{f}_1(t), \widehat{f}_2(t), \dots, \widehat{f}_q(t)]$ and $P \times Q$ weights $[W_{11}(t), W_{21}(t), \dots, W_{p1}(t), \dots, W_{1q}(t), W_{2q}(t), \dots, W_{pq}(t)]$. For notation simplification we wrote each node as $\widehat{f}_i(t) = f_i(t) + \sigma_{f_i}\epsilon(t)$ and dropped the dependence on t in the diagram.

signals in the RVs. The weight connections in figure 3.7 being dependent on t , will increase or decrease in strength with t . That allows the best nodes to fit the data accordingly, and if necessary, the time-varying weights will make a node have a different impact throughout the data.

Unlike a GP, the non-stationary nature of a GPRN makes inference intractable. What it means is that we can not calculate the posterior on a GPRN analytically. The way to deal with this problem was to use an approximation. The best approach found was to use variational inference (Jordan et al., 1999).

We introduce a family of factorized distributions $q(\mathbf{f}, \mathbf{w})$ to obtain the closest approximating distribution to the posterior $p(\mathbf{f}, \mathbf{w}|D)$ defined by a given set of outputs D . The quality of this approximation is measured with the Kullback-Leiber (KL) divergence (Kullback, 1959),

$$KL(q(\mathbf{f}, \mathbf{w})||p(\mathbf{f}, \mathbf{w}|D)) = E \left[\log \frac{q(\mathbf{f}, \mathbf{w})}{p(\mathbf{f}, \mathbf{w}|D)} \right]. \quad (3.34)$$

To obtain the approximation to the posterior, it is then only required to optimize each member of the family $q(\mathbf{f}, \mathbf{w})$. That allows us to transform this

approximation into an optimization problem (Blei et al., 2017).

To derive the factorized distributions $q(\mathbf{f}, \mathbf{w})$ we use mean-field inference for a GPRN developed by Nguyen and Bonilla (2013). Other variational inference methods exist, e.g. variational message passing (Winn et al., 2005) or nonparametric variational inference (Gershman et al., 2012; Nguyen and Bonilla, 2013), but are more computationally complex to implement.

Under a mean-field approximation we have the family of factorized distribution

$$q(\mathbf{f}, \mathbf{w}) = \prod_{q=1}^Q q(f_q) \prod_{p=1}^P q(w_{pq}) \quad (3.35)$$

where $q(f_q) = \mathcal{N}(\mu_{f_q}, \Sigma_{f_q})$ and $q(w_{pq}) = \mathcal{N}(\mu_{w_{pq}}, \Sigma_{w_{pq}})$.

For a fixed set of nodes and weights, it is required, for each distribution, to determine its parameters μ and Σ . These need to be optimized iteratively, calculating successively better approximations. The starting distributions can be random, but we can use better distributions to speed up the optimization. That can be achieved, for example, by reusing the final distributions from a set of nodes and weights on a new set with similar hyperparameters.

For the nodes, we optimize its μ and Σ values with

$$\begin{aligned} \Sigma_{f_q} &= \left[K_{f_q}^{-1} + \frac{1}{\sigma^2} \sum_{p=1}^P \text{diag}\left(\mu_{w_{pq}} \circ \mu_{w_{pq}} + \text{diag}(\Sigma_{w_{pq}}) \right) \right]^{-1}, \\ \mu_{f_p} &= \frac{1}{\sigma^2} \Sigma_{f_q} \sum_{p=1}^P \sum_{q=1}^Q \left(\mathbf{y}_p - \sum_{k \neq q}^Q \mu_{w_{pk}} \circ \mu_{f_k} \right) \circ \mu_{w_{pq}}, \end{aligned} \quad (3.36)$$

while for the weight they are given by

$$\begin{aligned}\Sigma_{w_{pq}} &= \left[K_{w_{pq}}^{-1} + \frac{1}{\sigma^2} \sum_{p=1}^P \text{diag} \left(\mu_{f_q} \circ \mu_{f_q} + \text{diag}(\Sigma_{f_q}) \right) \right]^{-1}, \\ \mu_{w_{pq}} &= \frac{1}{\sigma^2} \Sigma_{w_{pq}} \sum_{p=1}^P \sum_{q=1}^Q \left(\mathbf{y}_p - \sum_{k \neq q}^Q \mu_{f_k} \circ \mu_{w_{pk}} \right) \circ \mu_{f_q}.\end{aligned}\quad (3.37)$$

On these equations we were forced to adapt σ to our needs. For that we re-wrote it as $\sigma = \sum_{p=1}^P \sum_{n=1}^N (\sigma_{y_{np}}^2 + \sigma_{f_p}^2)$ to include all measurements errors σ_y and jitters σ_f we use. \mathbf{y}_p is the N-dimensional vector with the measurements of dataset p , K is the covariance matrix of the respective node or weight. The $\text{diag}()$ function turns the diagonal elements of a matrix in a vector, or a vector into a diagonal matrix. For last we also have \circ representing the Hadamard product (Styan, 1973).

With all the μ and Σ optimized, we measure the approximation to the true posterior using the KL divergence. That, however, is not analytically possible due to requiring the calculation of the evidence $\log p(D)$. To solve this situation, we algebraically manipulate the KL divergence and re-write it as

$$\begin{aligned}KL(q(\mathbf{f}, \mathbf{w}) || p(\mathbf{f}, \mathbf{w}|D)) &= E[\log q(\mathbf{f}, \mathbf{w})] - E[\log p(\mathbf{f}, \mathbf{w})] \\ &\quad + E[\log p(D)].\end{aligned}\quad (3.38)$$

That allows us to obtain a measurement equivalent to the KL divergence up to a certain constant. This equivalent measurement is known as the evidence lower bound (ELBO) and defined as (Blei et al., 2017)

$$ELBO = E[p(\mathbf{f}, \mathbf{w})] - E[\log q(\mathbf{f}, \mathbf{w})]. \quad (3.39)$$

Having this quantity allow us to use it to calculate the log posterior used on an MCMC analysis. For a the GPRN using mean-field inference, Nguyen and Bonilla (2013) defined the ELBO as

$$ELBO = \mathbf{E}_q[\log p(D|\mathbf{f}, \mathbf{w})] + \mathbf{E}_q[\log p(\mathbf{f}, \mathbf{w})] + \mathcal{H}_q[q(\mathbf{f}, \mathbf{w})], \quad (3.40)$$

where the first term is known as the expected log-likelihood, the second the expected log-prior and the last is known as entropy. To finalize, we need the definition of these three terms. We can see their full derivation in Nguyen (2015).

The expected log-likelihood is given by

$$\begin{aligned} \mathbf{E}_q [\log p(D|\mathbf{f}, \mathbf{w})] = & -\frac{1}{2} \sum_{n=1}^N \sum_{p=1}^P \left[\log 2\pi \left(\sigma_{yerr_{np}}^2 + \sigma_{y_p}^2 \right) \right] \\ & - \frac{1}{2\sigma^2} \sum_{n=1}^N \left(\mathbf{y}_n^T - \mathcal{W}_{w_n} \mathcal{F}_{f_n} \right)^T \left(\mathbf{y}_n^T - \mathcal{W}_{w_n} \mathcal{F}_{f_n} \right) \\ & - \frac{1}{2\sigma^2} \sum_{p=1}^P \sum_{q=1}^Q \left[\text{diag} \left(\Sigma_{f_q} \right)^T \left(\mu_{w_{pq}} \circ \mu_{w_{pq}} \right) \right. \\ & \left. + \text{diag} \left(\Sigma_{w_{pq}} \right)^T \left(\mu_{f_q} \circ \mu_{f_q} \right) + \text{Tr} \left(\Sigma_{f_q} \Sigma_{w_{pq}} \right), \right] \end{aligned} \quad (3.41)$$

where the $\text{Tr}()$ function returns the trace of the matrix. \mathbf{y}_n^T is a $1 \times P$ vector containing all observations at entry n , \mathcal{W}_{w_n} is a $P \times Q$ matrix containing the variational means μ_w at entry n , and \mathcal{F}_{f_n} is a $P \times 1$ vector containing the variational means μ_f at entry n .

The expected log-prior is given by

$$\begin{aligned} \mathbf{E}_q [\log p(\mathbf{f}, \mathbf{w})] = & -\frac{1}{2} NQ (P+1) \log 2\pi \\ & - \frac{1}{2} \sum_{q=1}^Q \left[\log |K_{f_q}| + \mu_{f_q}^T K_{f_q}^{-1} \mu_{f_q} + \text{Tr} \left(K_{f_q}^{-1} \Sigma_{f_q} \right) \right] \\ & - \frac{1}{2} \sum_{p=1}^P \sum_{q=1}^Q \left[\log |K_{w_{pq}}| + \mu_{w_{pq}}^T K_{w_{pq}}^{-1} \mu_{w_{pq}} + \text{Tr} \left(K_{w_{pq}}^{-1} \Sigma_{w_{pq}} \right) \right], \end{aligned} \quad (3.42)$$

where K is the covariance matrix of the respective node or weight.

For last we have the entropy that is given by

$$\begin{aligned} \mathcal{H}_q [q(\mathbf{f}, \mathbf{w})] &= \frac{1}{2} NQ (P + 1) (1 + \log 2\pi) \\ &\quad + \frac{1}{2} \sum_{q=1}^Q [\log |\Sigma_{f_q}|] + \frac{1}{2} \sum_{p=1}^P \sum_{q=1}^Q [\log |\Sigma_{w_{pq}}|]. \end{aligned} \quad (3.43)$$

Predictive distribution

Having defined all the necessary steps to apply mean-field variational inference, we can now think about the approximated posterior to predict an output y^* on time t^* . These expressions are, again, derived in Nguyen (2015).

Under mean-field inference, the predictive mean of the outputs $y^*(t^*)$ is given by

$$\begin{aligned} \bar{y}_i^*(t^*) &= \sum_{q=1}^Q E[W_{iq}^*] E[q^*] \\ &= \sum_{q=1}^Q K_{w_{iq}}^* K_{w_{iq}}^{-1} \mu_{w_{iq}} K_{f_q}^* K_{f_q}^{-1} \mu_{f_q}. \end{aligned} \quad (3.44)$$

where E is the expected value, K^{-1} is the inverse of the covariance matrix on the respective node or weight, and μ is the variational mean of the respective node or weight. For last, K^* represents the covariance matrices of f or w when evaluated at all pairs of t and t^* .

The predictive covariance is then given by

$$\begin{aligned} cov(y_i^*(t^*)) &= \sum_{q=1}^Q \left[E[W_{iq}^*]^2 var(f_q^*) + var(W_{iq}^*) E[f_q^*]^2 \right] + \delta \sigma_{y_i}^2 \\ &= \sum_{q=1}^Q \left[K_{w_{iq}}^* K_{w_{iq}}^{-1} \mu_{w_{iq}} K_{w_{iq}}^* K_{w_{iq}}^{-1} \mu_{w_{iq}} (K_{f_q}^{**} - K_{f_q}^{*\top} K_{f_q}^{-1} K_{f_q}^*) \right. \\ &\quad \left. + (K_{w_{iq}}^{**} - K_{w_{iq}}^{*\top} K_{w_{iq}}^{-1} K_{w_{iq}}^*) (K_{f_q}^{**} - K_{f_q}^{*\top} K_{f_q}^{-1} K_{f_q}^*) \right. \\ &\quad \left. + K_{w_{iq}}^* K_{w_{iq}}^{-1} \mu_{w_{iq}} K_{f_q}^* K_{f_q}^{-1} \mu_{f_q} \right] + \delta \sigma_{y_i}^2. \end{aligned} \quad (3.45)$$

On the predictive covariance, var gives us the variance of the process, σ_{y_i} the jitter associated to the time series i , and $K^{\star\star}$ the covariance matrices of f or w obtained from t^\star .

There is a final aspect to be addressed on the predictive distribution of a GPRN. In chapter 3 I said that a GPRN will not easily over-fit a model. That being due to, and as stated on Wilson et al. (2012), a GPRN having a heavy-tailed predictive distribution. The reason behind it, although not explained there, is due to how a GPRN was defined.

A GPRN will have at least two latent variables. In the simplest case, we can have one node and one weight. That setting will cause the output (RV, FWHM, etc.) to have a marginal distribution given by two independent Gaussian marginal distributions. Important to remember, each node and weight is an independent GP.

Due to this, each output's marginal distribution will come from the multiplication of two Gaussian distributions. It can be proved, although not shown here, that the product of two Gaussian distributions is not Gaussian (e.g. Cui et al., 2016). This product will have a distribution that follows a re-normalized Bessel function of the second kind of order zero. For more on these functions, I refer to Arfken et al. (2013). This type of Bessel function has indeed heavier tails than a Gaussian function (see figure 3.8). Also important to note is that this distribution occurs independently of the type covariance functions used on the node and the weight GPs.

That will produce a posterior distribution on a GPRN with tails bigger than the predictive of a GP, making it less prone to over-fitting. All the effects this will have on the predictive mean is still not fully understood. With these tails, one could expect the GPRN predictive to show a less "flexible" fit as it would more easily explain data points with the tails. Such would imply that some usual fit evaluation metrics (e.g. root mean squared error) are not good metrics to use on a GPRN, especially when comparing its results to, for example, a GP.

Python implementation

Following the works of Wilson et al. (2012) and Nguyen and Bonilla (2013) we created a python package with the necessary functions to use a GPRN on

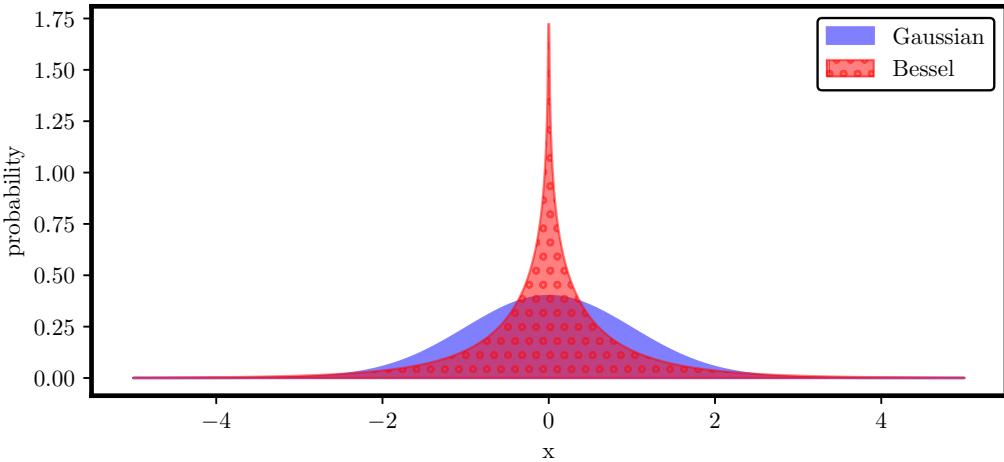


Figure 3.8: Comparison of the probability density function of a Gaussian and a Bessel distribution. Both with a mean of 0 and a standard deviation of 1.

a set of RV and activity indicators measurements. The package called `gpyrn` is publicly available online on Github¹². We wrote this package using python version 3.8, and it should work for all python 3 versions. Since python 2 is no longer supported, we did not perform any tests with that version, and it is unclear if `gpyrn` will work properly with it.

3.3.2 Simulating a quasi-periodic GP with the GPRN

As I mentioned earlier, the GP of choice in exoplanet search uses a quasi-periodic kernel of equation 3.20. One can then question if, under certain conditions, a GPRN can mimic a GP. Or, in the case of RV analysis, if a GPRN can mimic the traditional quasi-periodic GP.

For simplification, let us consider a single time series. For a GPRN, that implies having $P = 1$. On a GP, we only need to define the parameters of the quasi-periodic kernel. I will define the kernel as having $\eta_1 = 1$ m/s, $\eta_2 = 50$ days, $\eta_3 = 20$ days, and $\eta_4 = 0.75$. If we draw three samples of this zero mean GP, we

¹² <https://github.com/jdavidrcamacho/gpyrn>

get figure 3.9. The main characteristic of these samples is an overall periodicity with an evolving structure. That reflects in both a change in the signal amplitude and the structure inside one period.

A single time series on a GPRN implies having $P = 1$. Let us also consider a single node ($Q = 1$) and, as consequence, a single weight ($P \times Q = 1$). We can define one node with a periodic GP and one weight with a squared exponential GP. The parameters of the periodic GP are $\eta_1 = 1$ m/s, $\eta_3 = 20$ days, and $\eta_4 = 0.75$. The decaying time span is non-existent on a periodic kernel. Instead is used on the squared exponential kernel defined with the parameters $\eta_1 = 1$ m/s and $\eta_2 = 50$ days. Drawing three samples from this GPRN gives us figure 3.10.

This first GPRN has a structure whose only change seems to be on the signal amplitude. Unlike the GP, we have no change in the smoothness over time. That indicates a GPRN that will be less flexible than the quasi-periodic GP. That being so, using a GP with a periodic kernel on the node might not be ideal. To correct this result, we can try another configuration on the GPRN.

The second GPRN, again with one node and one weight, will now use a node with a quasi-periodic GP. The kernel will have $\eta_1 = 1$ m/s, $\eta_2 = 50$ days, $\eta_3 = 20$ days, and $\eta_4 = 0.75$. Now the node is exactly equal to the GP used for figure 3.9. That creates the issue of how to define the weight. Considering it using still a squared exponential GP, if its decaying timespan has a large value, it will behave as an almost constant GP. As such I will define the weight kernel as having $\eta_1 = 1$ m/s and $\eta_2 = 10000$ days. If we now draw three samples of the new GPRN, we

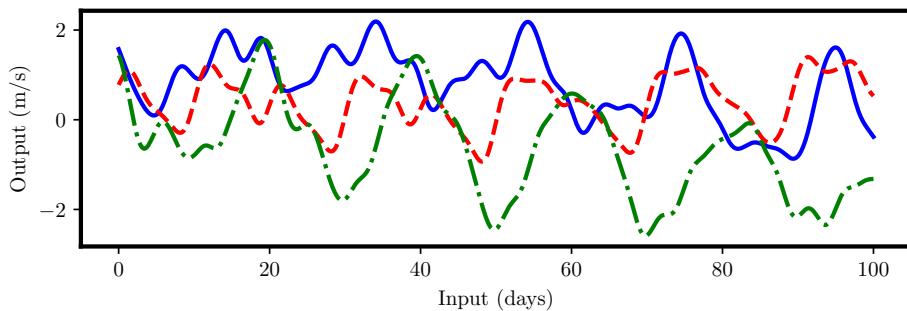


Figure 3.9: Samples of a GP defined with a quasi-periodic kernel.

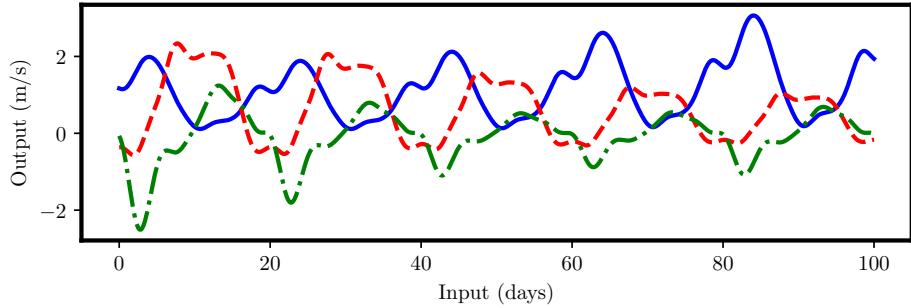


Figure 3.10: Samples of a GPRN defined with node with a periodic GP and a weight with a squared-exponential GP.

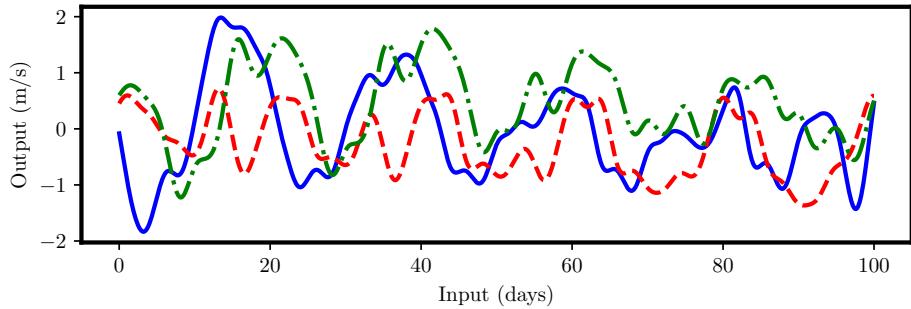


Figure 3.11: Samples of a GPRN defined with node with a quasi-periodic GP and a weight with a squared-exponential GP.

obtain figure 3.11.

This GPRN configuration shows a behaviour far closer to the observed on a quasi-periodic GP. The samples also show an evolving structure from one period to the next, which should be more flexible to fit quasi-periodic signals. How this GPRN performs and compares to a GP will be left for the next chapter.

Other GPRN configurations that simulate a quasi-periodic GP can exist. At first glance, a setup closer to the initial GP would be to use a quasi-periodic kernel on the node with a constant kernel on the weight. Unfortunately, our version of the GPRN makes use of mean-field inference. This inference method uses algorithms that require the inversion of the covariance matrix of the weights. That makes it

impossible to use a constant covariance function.

3.3.3 Using a GPRN on simulated data

One essential question is if a GPRN can determine the expected parameters from a given time series. To determine that, we took samples from three GPRNs that had a quasi-periodic GP on the node and a squared exponential GP on the weight. The sampled node was the same for the three tests, whose parameters were $\eta_1^n = 1 \text{ m/s}$, $\eta_2^n = 17 \text{ days}$, $\eta_3^n = 23 \text{ days}$, and $\eta_4^n = 0.75^{13}$. Each weight, while having the same 7m/s amplitude, had different values for the decaying time span η_2^w . These were 29, 79, and 109 days. To make each GPRN more comprehensible, I will refer to them as 1st, 2nd, and 3rd model, for the GPRN with η_2^w equal to 29, 79, and 109 days, respectively.

The values defined for the parameters have no particular reason. The only curious fact about them, except for η_4^n , is that they are all different prime numbers. In the end, we randomly sampled 50 data points, covering a period of 150 days. The node, weights and final time series are shown on figure 3.12.

In each time series, we then used `emcee` (Foreman-Mackey et al., 2019) to obtain samples from the parameter space. Our models also included a jitter term to account for uncorrelated white noise. We could incorporate an offset, but we set it to zero to simplify the examples. We show the priors used in appendix A.1. Each MCMC used 14 Markov chains (twice the number of parameters) and ran until it achieved the convergence criteria of 25 times the integrated autocorrelation time (τ). This criterion defines τ as the number of steps needed for a chain that forgets where it began or reached its equilibrium (Sokal, 1997). To determine if the criteria had been achieved, at every 5000 iterations, the MCMC calculated the value of τ . We then used 2τ as the number of burn-in iterations of the MCMC.

We show the final posterior distribution of the parameters in appendix A.2, and its inferred values in table 3.1. In figure 3.12 we plot the predictive means using the maximum a posteriori (MAP) solutions. The figure also includes the

¹³ To help differentiate if the parameters are the node or the weight I will use the exponent n when referring to a parameter corresponding to a node, and the exponent w when referring to the parameters of a weight.

Parameter (units)	1st model	2nd model	3rd model
η_2^n (days)	$19.394^{+3.401}_{-2.908}$	$20.294^{+4.528}_{-4.016}$	$23.549^{+3.892}_{-3.636}$
η_3^n (days)	$22.365^{+1.024}_{-0.656}$	$22.523^{+3.973}_{-0.900}$	$22.272^{+0.598}_{-0.692}$
η_4^n	$0.831^{+0.137}_{-0.128}$	$0.756^{+0.147}_{-0.127}$	$0.845^{+0.133}_{-0.113}$
η_1^w (m/s)	$10.437^{+5.967}_{-3.445}$	$14.209^{+10.671}_{-5.272}$	$7.752^{+6.901}_{-2.785}$
η_2^w (days)	$462.987^{+620.391}_{-317.197}$	$144.730^{+90.928}_{-59.462}$	$512.810^{+537.897}_{-293.385}$

Table 3.1: Median values of parameters obtain for the node and weight of each GPRN model. To calculate the errors we used the 16th and 84th percentiles.

individual predictive means for each node and weight.

From these three tests, we observe that the GPRN recovered the values of η_3^n and η_4^n . The decaying time span of the node shows some differences, falling outside the range on our uncertainties on table 3.1. However, the differences between the recovered and initial value show low impact on the predictive means plotted in figure 3.12. The weights have the most significant differences. The values obtained resulted in a far higher value for η_2^w . That indicates that the derived weights have a more constant behaviour than the weight from where we sample our data points. Still, as the GPRN shows a satisfying fit to the data, this had a low impact on the final predictive mean.

We need still to address an intriguing result for the individual predictive means for the nodes and weights. On the 2nd model, the node and weight are well reproduced by the individual predictive. However, when comparing data points and predictive means on the 1st and 3rd models, these are reflected among the x-axes. As this occurs on both node and weight, it has no impact on the final predictive. That seems to indicate that, on a GPRN, the converged result obtained may not be unique, as another that mirrors it might exist with no impact on the final predictive. The effect of these reflections still needs to be further explored to determine if they will pose a problem in the future.

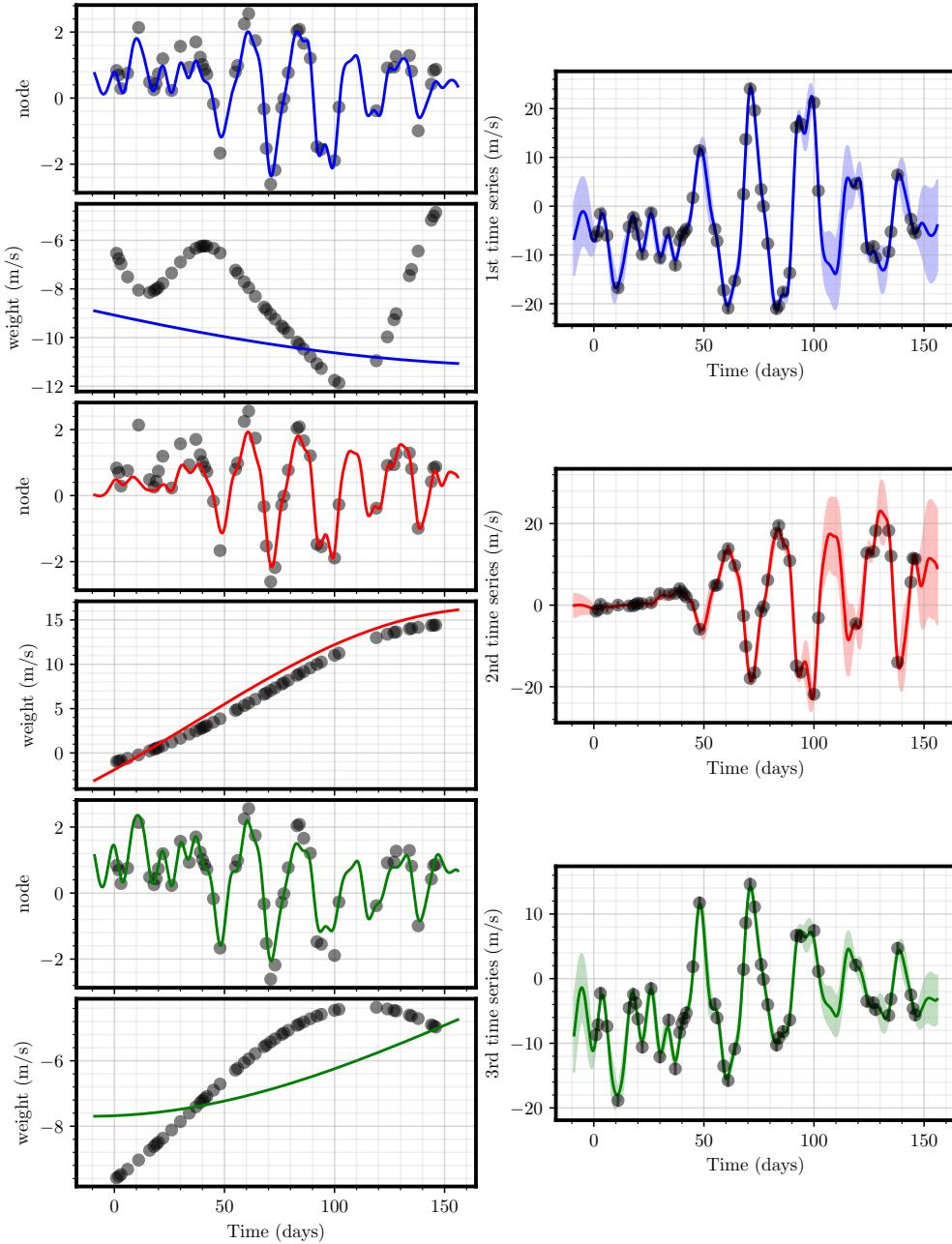


Figure 3.12: The black markers in each panel represent the data sampled (node, weight, or time series) from the three GPRNs. Each GPRN uses a different colour for the obtained fits. On the top, drawn with blue, are the data points and fits for node, weight, and final time series for the 1st model, on the middle with red, is the data are resulting fits for the 2nd model, and on the bottom plotted with green, is the 3rd model.

104 Advanced statistical data analysis methods for the detection of other Earths

Chapter 4

Application to radial velocity observations

In this chapter, we present the results obtained with a GPRN. It starts by showing its performance on radial velocity observations of the Sun. Initially used as a test case for the GPRN development, the results obtained with this dataset also shed a light to future improvements of this framework. In the fourth chapter, we also include the results for four stars on the EXPRES Stellar-Signals Project (ESSP). These stars, with different activity levels, are ideal test cases to the capabilities of the GPRN.

We start by defining the priors used in all our analyses done in this chapter. We follow it with the presentation of the solar data used and all the results obtained with it. In the last section of the chapter, we show the results for the stars of the EXPRES Stellar-Signals Project.

4.1 Priors definition

We start by presenting the priors used on the MCMC to perform all the analysis. For a consistent analysis of all the datasets, the priors were defined to use as much information as possible from the time series, e.g. using its peak-to-peak amplitude (PTP) as a limit. That implied avoiding, whenever possible, undefined

or unexplained values as prior parameters.

The same prior is employed on a given hyperparameter, independently if the analysis is a simple GP regression or inference with the GPRN. That restricts, for example, all the amplitudes η_1 to use the same prior. Considering the GPRN we built, the parameters on it and GP regression should have a similar interpretation. Defining the priors this way will allow us to use the evidence $\log \mathcal{Z}$ as a model comparison more easily. Also relevant to mention, the parameters will carry the same notation used in chapter 3.1.3.

To simplify the presentation, we first describe the priors used to study the individual datasets, either the RVs or the activity indicators. Then, in section 4.1.2 the priors for the combined analysis of RVs and a given activity indicator are presented. Similarly to section 3.3.3 when a parameter comes from a node, we write it with the exponent n when it appears from a weight, we use the exponent w .

4.1.1 Individual analysis

In the individual analysis of the RVs or one of the activity indicators, let y denote the used measurements and t the time of the observations. We list the priors for all the parameters in Table 4.1, where σ stands for standard deviation. y_{PTP} and t_{PTP} are the peak-to-peak amplitude for y or the timespan for t . We represent the minimum and the maximum of the observed measurements by y_{min} and y_{max} , respectively. Finally, t_{AV} is the average time between two consecutive observations.

The parameters that represent amplitudes (or variances) used a modified log-uniform (\mathcal{MLU})¹. That allows the support of the distribution to include zero. On η_2 and η_4 (the time scales) we used a log-uniform distribution (\mathcal{LU}) instead². For the η_3 , we defined a uniform distribution (\mathcal{U}) between 10 and 50, acceptable margins for the rotation period of a Sun-like star. The white noise (or jitter) s used the same prior as η_1 .

¹ This distribution is also known as the *modified Jeffreys prior*.

² This distribution is also known as the *Jeffreys prior*.

Variables	Parameter (units)	Prior
η_1	Amplitude (m/s)	$\mathcal{MLU}(y_\sigma, 2 \times y_{\text{PTP}})$
η_2	Decaying timespan (days)	$\mathcal{LU}(t_{AV}, 10 \times t_{\text{PTP}})$
η_3	Period (days)	$\mathcal{U}(10, 50)$
η_4	Length scale	$\mathcal{LU}(0.1, 5)$
s	White noise amplitude (m/s)	$\mathcal{MLU}(y_\sigma, 2 \times y_{\text{PTP}})$
slope	Slope of the mean function	$\mathcal{N}\left(0, \frac{y_\sigma}{t_\sigma}\right)$
offset	Offset of the mean function (m/s)	$\mathcal{U}(y_{\min}, y_{\max})$

Table 4.1: Parameters and prior distribution on the analysis of a single for both GP and GPRN. Here, y represents the observed output and t the times of observation. In this individual analysis, the measurements y could either be the RVs, FWHM, BIS, or $\log R'_{hk}$.

Important to mention is that the value of η_2 was constrained. This parameter is closely related to the evolution of the active regions on the surface of the stars. To recover more physically and useful results, it has been suggested to constrain it to values higher than the rotation period η_3 (e.g. Kosiarek and Crossfield, 2020).

With these conclusions in mind, we created the following constraints. If the η_2 is on a node, we constrain it to have values higher than $0.5\eta_3$. When the η_2 parameter is on one of the weights, we force it to be higher than the η_2 of the node ($0.5\eta_3 < \eta_2^n < \eta_2^w$). These conditions should, in principle, allow the node to model the evolution of the activity regions and the weight to model longer timescale signals, like those related to the magnetic cycle. For the GP analysis, the framework only has a single η_2 parameter. Due to that, we just constrained it to be higher than $0.5\eta_3$.

For last, it was considered a linear function as the mean function of the GP and the GPRN. This function depends on two parameters, a slope and an offset. The slope parameter used a Gaussian prior (\mathcal{N}) with mean zero and variance defined as the ratio of the standard deviations of the measurements and the time. The offset used a uniform prior between the minimum and maximum values to the

Variables	Parameter (units)	Prior
η_1^n	Amplitude (m/s)	$\mathcal{MLU}([y_{RV}, y_{AI}]_{\sigma}^{\min}, 2 \times [y_{RV}, y_{AI}]_{\text{PTP}}^{\max})$
η_2^n	Decaying timespan (days)	$\mathcal{LU}(t_{AV}, 10 \times t_{\text{PTP}})$
η_3^n	Period (days)	$\mathcal{U}(10, 50)$
η_4^n	Length scale	$\mathcal{LU}(0.1, 5)$
η_1^w	Amplitude (m/s)	$\mathcal{MLU}(y_{\sigma}, 2 \times y_{\text{PTP}})$
η_2^w	Decaying timespan (days)	$\mathcal{LU}(t_{AV}, 10 \times t_{\text{PTP}})$
s	White noise amplitude (m/s)	$\mathcal{MLU}(y_{\sigma}, 2 \times y_{\text{PTP}})$
slope	Slope of the mean function	$\mathcal{N}\left(0, \frac{y_{\sigma}}{t_{\sigma}}\right)$
offset	Offset of the mean function (m/s)	$\mathcal{U}(y_{\min}, y_{\max})$

Table 4.2: Parameters and prior distributions when considering the analysis of the RVs with an activity indicator. y_{RV} represents the RV time series while y_{AI} the activity indicator (FWHM, BIS, or $\log R'_{hk}$) time series, as previously t represents the time.

given input.

4.1.2 Combined analysis

The prior for the node η_1 needed to be adapted to account for two datasets. With both RV and a indicator, the knee or the point where the distribution changes from uniform to log-uniform is defined as the lowest standard deviation of the two time series that denoted by $[y_{RV}, y_{AI}]_{\sigma}^{\min}$. The upper bound of the distribution is specified to be twice the value of the largest of the two peak-to-peak amplitudes obtained from the RVs and the indicator. I represent this value by $[y_{RV}, y_{AI}]_{\text{PTP}}^{\max}$.

For the GP analysis, we use the artGPN framework mentioned in section 3.2.3. The framework combines the kernels on the nodes and the weights to create a single GP per time series. This configuration implies that, for each time series, the final amplitude is the product of the respective node and weight amplitudes. That allows us to simplify the parameter space and fix one amplitude

while the other stays free. In this case, we set the η_1 of the node to 1, while the weights being constant kernels, only had the η_1 parameter. The prior distribution used was equal to the weight η_1^w of table 4.2. The remaining parameters of the quasi-periodic kernel used the corresponding priors defined for the node on table 4.2.

Lastly, as mean functions, we use again linear functions, one per time series. The distributions are identical to the prior of the individual analysis and dependent on the time series.

4.2 Analysis of Sun-as-a-star observations

The Sun is the nearest star to us and consequently has been studied and monitored over millennia. That enabled a determination of its parameters with very high precision. Based on its spectral class, it is a G2V-type main-sequence star. Dominating its interior is a radiative zone that goes up to 0.7 of its radius, covered by a convective shell (Del Zanna and Mason, 2013). It is this convective region that drives the magnetic activity responsible for the stellar signals contaminating RV observations by a physical process known as solar dynamo (e.g. Pagano, 2013; Charbonneau, 2014). This process creates the well known 11-year sunspot cycle, in which, for example, the number of spots, flares and coronal mass ejections varies, and with it, the respective magnetic activity of Sun (for a review on the solar cycle see Hathaway, 2015).

4.2.1 Sun observations

The main results obtained with a GPRN used the three years of solar RV, FWHM, BIS, and $\log R'_{hk}$ measurements taken by Dumusque et al. (2015) and made publicly available at (Dumusque et al., 2021). To obtain these observations they made use of a solar telescope fibre-fed into the HARPS-N spectrometer (Cosentino et al., 2012). The available measurements correspond to a period between 2015 and 2018. This phase corresponds to the end of the solar cycle 24, while the Sun approaches a magnetically less active period. With observations done during this minimum activity period, it is fair to expect less prominent signal contamination due to spots or flares.

Originally the Sun RV observations were released by Collier Cameron et al. (2019). However, to address some issues, a new dataset was created by Dumusque et al. (2021) adapting the ESPRESSO DRS (Pepe et al., 2021) to the HARPS-N spectrograph. They used stricter criteria to evaluate the quality of the spectra over the three years of operation, using only 65% of the available HARPS-N spectra. The result was a substantial decrease in the day-to-day RV scatter and a stronger correlation between the RV and the chromospheric activity indicators (Dumusque et al., 2021).

In total, there were 34550 measurements available with a median cadence of around 5.42 minutes between consecutive observations. We nightly binned to remove oscillation and granulation signals (Dumusque, X. et al., 2011a). To simulate a standard 15-minute exposure RV observation, we nightly binned three consecutive measurements. That immediately excluded days with less than three observations. To bin the data, we randomly picked three sequential measurements in a day, separated by five minutes. We excluded a day when it had no combination of three sequential measurements with a cadence of five minutes. To calculate the final observational error, we quadratically summed the errors of the three used observations.

The final dataset consisted of 497 measurements for each time series, spanning 1094 days (see figure 4.1). We give a small summary of these time series in appendix B. The created observations have no planetary signals. That makes it valuable to understand the effects of stellar activity on RV observations and how to best use the activity indicators to improve planetary signal detection.

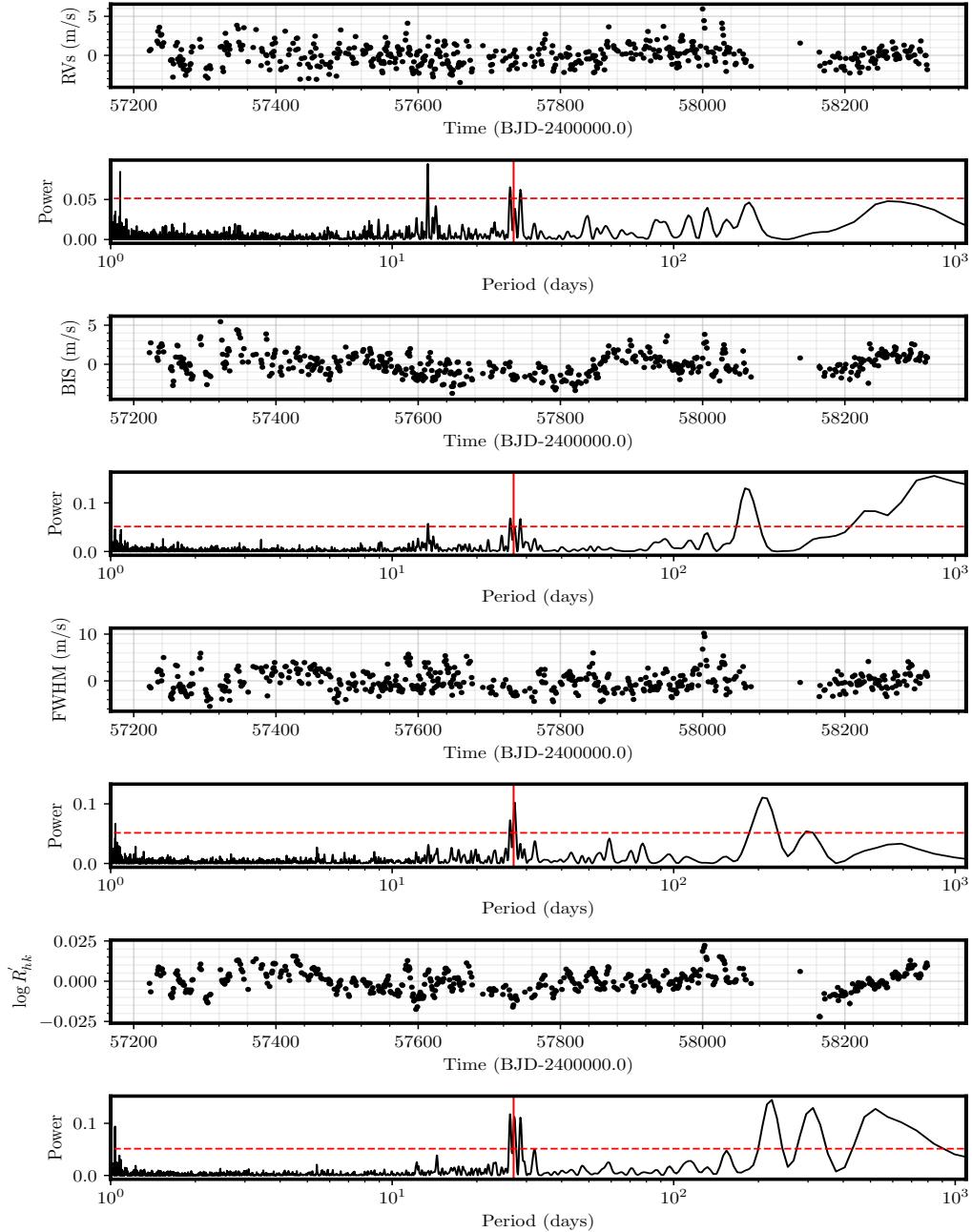


Figure 4.1: RV, BIS, FWHM, and $\log R'_{hk}$ measurements of the Sun obtained after the nightly binning. Below each time series is its respective GLS periodograms after removing a long-term trend from the data. The vertical solid lines indicate a period of 27 days. The horizontal dashed line represents the false alarm probability (FAP) level of 1%. These measurements are available at <https://github.com/jdavidrcamacho/datasets>.

Periodogram analysis

Before applying the GPRN to the HARPS-N solar data, we can analyse the periodogram of the time series. On figure 4.1 4.1 we have the generalized Lomb–Scargle (GLS) periodograms as it is implemented on Astropy (Astropy Collaboration et al., 2018) for the RV, FWHM, BIS, and the $\log R'_{hk}$.

All the time series show peaks in power around the 27 days synodic rotation period of the Sun marked by the red vertical line (Wilcox, 1972). The RV time series has a collection of peaks around 27 days higher than the 1% false alarm probability (FAP). This multiple peak "behaviour" is also present on the activity indicators. That indicates that these peaks are indeed activity-related. They can be the result of, for example, spots on different latitudes or differential rotation. The origin of it, however, is a topic we will not approach in this thesis. The RV data has a more significant peak at the second harmonic of the rotation period, around 13.5 days. It is the only time series with this behaviour.

Of all activity indicators, the BIS has a periodogram structure closest to that on the RV. This indicator also has a peak around 13.5 days higher than the 1% FAP. However, the power is lower than at the rotation period. The second harmonic is either not present or has no significant power on the remaining indicators. The periodogram structure of the FWHM and $\log R'_{hk}$ for the expected rotation period also shows one than one peak with power higher than the 1% FAP level.

Considering the observations occurred with the Sun approaching its less active phase these are interesting results. Although approaching the minima of the magnetic cycle it was not the minima itself. This is seen by the daily sunspots number recorded of figure 4.2. It shows a very considerable number of spots present on the surface of the Sun during the observations. It can also be seen a clear decrease in the number of spots as the Sun approached the cycle's minima. The number of days with spots however is significantly higher than the days without. From the start of 2015 till the end of 2018, there were 1130 days with spots against 331 days without.

Outside the rotation period and harmonics "region", all activity indicator's time series shows a strong signal around the 200 days mark. That signal is probably related to secular trends due to smooth long-term changes in the CCF

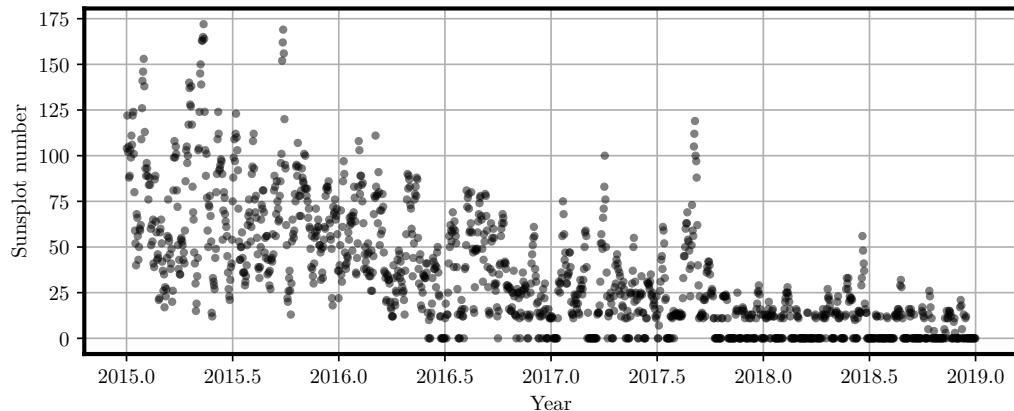


Figure 4.2: Daily number of sunspot observed from the start of 2015 til the end of Solar cycle 24 on late 2019 (e.g. McIntosh et al., 2014; Nandy et al., 2020). Source: <https://wwwbis.sidc.be/silso>

area. Collier Cameron et al. (2019) identified similar periodicities on the BIS time series and fitted it out with a sinusoidal function. In our work, any long term trends should be captured by the mean function and not interfere with the estimation of the rotation period. The RV also shows some structure below the two-days mark that is most likely an artefact of the sampling.

Time lags

The presence of a temporal shift for the Sun measurements between the RV and the activity proxies was reported by Collier Cameron et al. (2019). They determined a lag of 1 and 3 days between the RV and the FWHM and BIS, respectively.

Our measurements have some significant differences from the ones presented there. Collier Cameron et al. (2019) data used the HARPS-N DRS, while Dumusque et al. (2021) adapted the ESPRESSO DRS. That allowed them to obtain more precise RV measurements corrected from instrumental systematics. Besides that, we binned the available measurements to simulate a 15-minutes exposure spectroscopic observation.

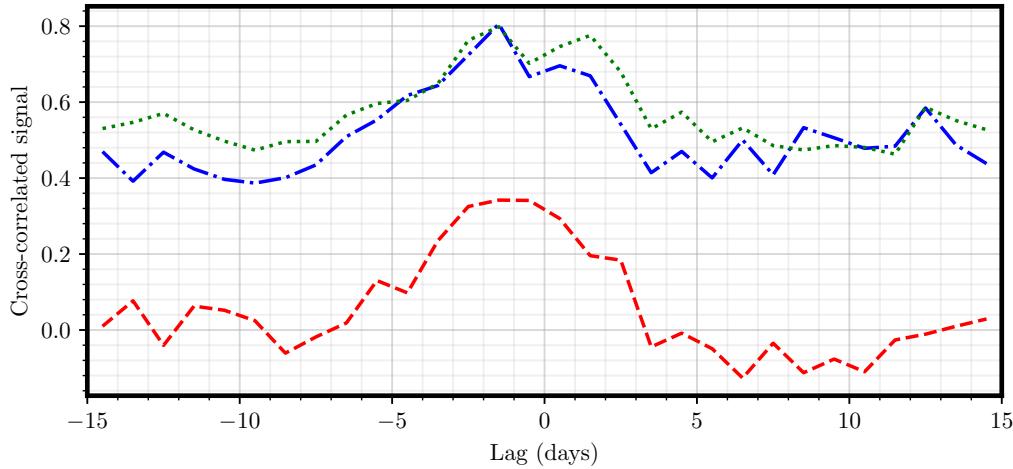


Figure 4.3: Discrete cross-correlation functions between the RV and the BIS (dash-dotted blue), FWHM (dashed red), and $\log R'_{hk}$ (dotted green).

With that in mind, obtaining the same cross-correlation functions is unlikely. However, if they exist, similar time lags should be observable in our time series. Using the same discrete correlation function developed by Edelson and Krolik (1988) we thus calculated the CCFs. The results are shown on figure 4.3. We find that, on our time series, the RV indeed lead the FWHM by 1 to 2 days. This time shift increased to 2 to 3 days when comparing RV and BIS. The chromospheric indicator $\log R'_{hk}$ does not show a clear time lag relative to the RVs.

The correlation of figure 4.3 shows the FWHM and the $\log R'_{hk}$ with significant correlations for all lags. One explanation is the existence of a long-period signal that is not present on the BIS. It seems to be the case as the correlations for longer lags disappear if we subtract a linear trend from the data.

4.2.2 Results obtained on the solar data

In this section, I present the results of applying the GPRN model to the solar observations. A comparison is carried out with a similar GP to understand the benefits of each model. For the GPRN analysis, we used the previously mentioned `gpyrn` package. We consider a GPRN with a single node and one weight per time series. We defined the nodes with a quasi-periodic GP and each of the weights with a squared exponential GP.

On the GP analysis, when using one dataset, we used a package called `tedi`³. With this package, for each time series, we used the quasi-periodic kernel of 3.20 plus a white noise term to then perform a GP regression. The combined GP analysis of RV measurements plus an activity indicator used the artGPN framework mentioned in 3.2.3. Similarly to the GPRN, it only had a single node and one weight per time series. We used a quasi-periodic kernel on the nodes and a constant kernel for each weight.

For these analyses, we used the MCMC algorithm implemented in `emcee` to sample the parameter space and obtain the posterior distributions. The MCMC was left running until the convergence criteria of having a sample size of 25 times the integrated autocorrelation time. Once again, τ was evaluated at every 5000 iterations until it achieved the convergence criteria. To our analysis, we then used 2τ as the number of burn-in iterations of the MCMC. As a metric to determine the best model, we used the Bayesian evidence $\log \mathcal{Z}$ using the approximation of Perrakis et al. (2014). We also considered how much each model diminished the root mean squared (RMS) of the residuals.

In the following pages, we present the GP and GPRN results obtained while analysing the solar data. We describe the obtained results, one by one, accordingly to the time series used. For each model, we show the *maximum a posteriori* (MAP) as well as the 50% quantile (the median⁴) estimates. We use the 16% and 84% quantiles values of the posterior distributions to calculate the uncertainties on the parameters. The corner plots showing the marginal and joint posterior distributions for all parameters are in appendix C. In the next section, we start by

³ <https://github.com/jdavidrcamacho/tedi>

⁴ We used the median value due to being more robust against outliers and skewed distributions.

an interpretation of the results obtained and leave a more careful examination of the behaviour of the GPRN and GP analysis for section 4.2.3.

Individual analysis - RVs

The GP and GPRN were first applied to the RV data alone, converging to similar results. We give the final estimates for the parameters in table 4.3, and the fits obtained with the respective MAP values in figure 4.4.

Table 4.3 shows that, for the two models, the common parameters are consistent within a 1-sigma interval. That implies both the GP framework and the node of the GPRN obtained similar values for η_1 , η_2 , η_3 , and η_4 . Also similar is the jitter value obtained of almost $0.8m/s$. It is also noteworthy to mention that the two models also obtained the correct rotation period η_3 . The values obtained for the parameters of the weight of the GRPN are indicative of the GP characterizing the weight as having an almost constant behaviour.

Reinforcing the similarities of both frameworks are the fits obtained in figure 4.4. Both models show a similar RMS reduction, with a slightly higher decrease on the GP. This model managed to decrease the RMS by 3.075, while the GPRN obtained an RMS reduction of 2.876. It is relevant to remember, however, that the predictive mean of the GP follows a Gaussian distribution, the predictive mean of the GPRN does not. That can result in lower flexibility of the GPRN's predictive distribution, resulting in a higher RMS. I will discuss more this issue in section 4.2.3.

Even so, the evidence gives a slight advantage to the GP. This model obtained a $\log \mathcal{Z}$ of -789.106 ± 0.075 . The GPRN obtained a $\log \mathcal{Z}$ equal to -789.460 ± 0.307 .

A final aspect important to discuss in this analysis is the convergence speed of the models. The GP proved to be considerably faster to achieve the convergence criteria, requiring only 10000 iterations. The GPRN, on the other hand, required 40000 iterations. Considering the mathematical complexity of the GPRN that slows any MCMC used, the need for more iterations to achieve convergence might constrain any future use of a GPRN.

Framework	Variables	MAP value	Median value
GPRN	η_1^n	0.720	$1.412_{-1.097}^{+4.977}$
	η_2^n	19.972	$20.258_{-1.931}^{+2.015}$
	η_3^n	25.921	$26.188_{-0.559}^{+0.657}$
	η_4^n	0.584	$0.610_{-0.063}^{+0.068}$
	η_1^w	1.038	$1.556_{-1.220}^{+5.178}$
	η_2^w	1001.229	$3569.232_{-2123.239}^{+4025.923}$
	s	0.781	$0.793_{-0.040}^{+0.040}$
	slope	-0.005	$-0.005_{-4 \times 10^{-4}}^{+4 \times 10^{-4}}$
	offset	-19.032	$-19.131_{-0.140}^{+0.142}$
GP	η_1	1.145	$1.179_{-0.081}^{+0.088}$
	η_2	20.374	$20.342_{-2.042}^{+2.149}$
	η_3	26.031	$26.239_{-0.562}^{+0.618}$
	η_4	0.625	$0.632_{-0.063}^{+0.073}$
	s	0.797	$0.796_{-0.039}^{+0.040}$
	slope	-0.005	$-0.005_{-4 \times 10^{-4}}^{+4 \times 10^{-4}}$
	offset	-19.124	$-19.117_{-0.148}^{+0.114}$

Table 4.3: Parameters obtained by the GPRN and the GP on the analysis of the RV time series. On the GPRN's variables, the exponent represents if the variable is either from a node (n) or a weight (w).

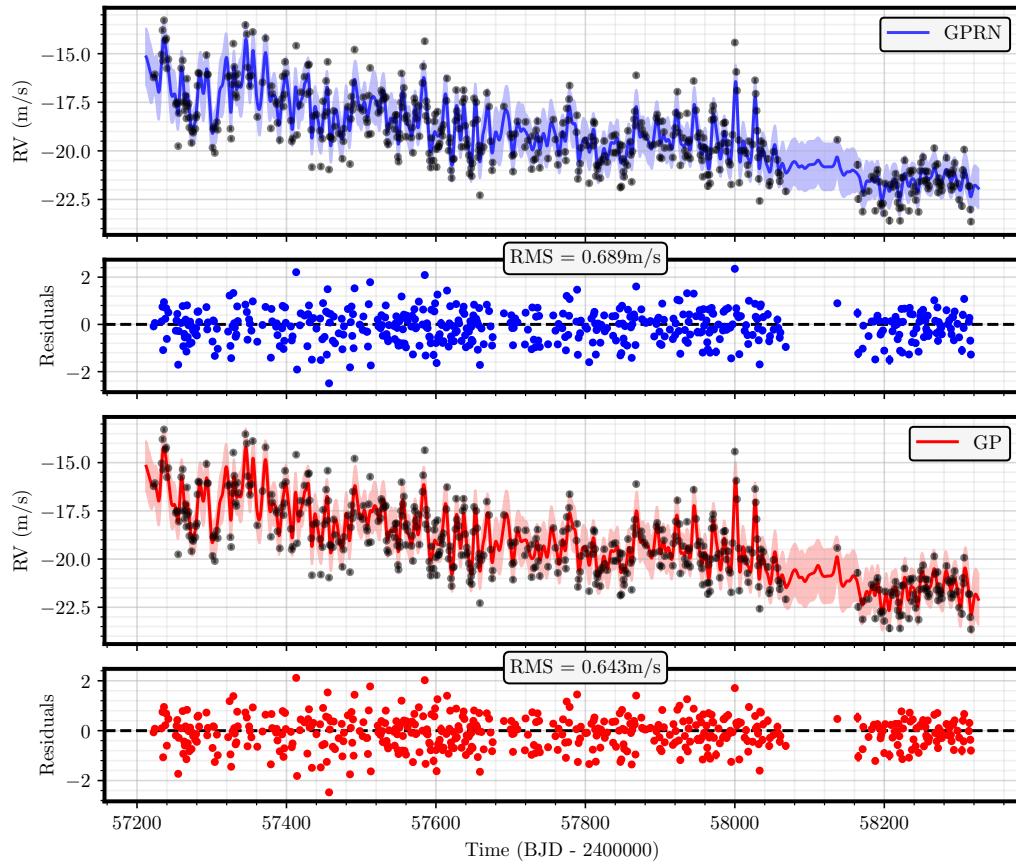


Figure 4.4: Fits and residuals obtained using the MAP values from the GPRN (blue) and the GP (red) on the RV measurements.

Individual analysis - BIS

Similar to the RV analysis, the results obtained for the BIS observations show two models with similar performance at determining the best kernel parameters. The parameters obtained are in table 4.4 and the fits of the two models in figure 4.5.

As previously, the parameters in both models are consistent within a 1-sigma. The rotation period was again correctly determined by both models. That results, of course, in very similar fits as shown by figure 4.5. Although similar, the RMS reduction achieved by the GP is again lightly higher. That model decreased the RMS by 5.355 times, while the GPRN decreased it only by 4.866 times. The explanation for a better GP fit lies again on the different distributions each framework have for its respective predictive mean. With a Bessel distribution, the GPRN predictive explains the data points using the longer tails, restricting the reach of the predictive mean on points deviating too much from any trend observed on the other measurements.

The evidence calculated supports the GP as being better at modelling the data. The GP obtained a $\log \mathcal{Z}$ of -639.606 ± 0.064 against a $\log \mathcal{Z}$ of -643.512 ± 0.212 for the GPRN. The speed of the GPRN is again problematic. Similar to the RV analysis performance, the GPRN model showed to be not only slower per iteration but required more iterations to converge. The GP necessitated 10000 to achieve it against the 25000 iterations of the GPRN.

Framework	Variables	MAP value	Median value
GPRN	η_1^n	1.568	$1.557^{+4.511}_{-1.203}$
	η_2^n	23.235	$23.575^{+1.588}_{-1.570}$
	η_3^n	26.853	$26.993^{+0.364}_{-0.381}$
	η_4^n	0.724	$0.735^{+0.056}_{-0.050}$
	η_1^w	0.855	$1.492^{+4.804}_{-1.130}$
	η_2^w	2431.229	$4909.158^{+3588.740}_{-2517.306}$
	s	0.414	$0.418^{+0.030}_{-0.030}$
	slope	-0.005	$-0.005^{+0.001}_{-0.001}$
	offset	-91.689	$-91.669^{+0.177}_{-0.173}$
GP	η_1	1.342	$1.360^{+0.098}_{-0.086}$
	η_2	24.004	$23.739^{+1.584}_{-1.558}$
	η_3	27.064	$27.073^{+0.365}_{-0.360}$
	η_4	0.713	$0.726^{+0.051}_{-0.051}$
	s	0.406	$0.405^{+0.031}_{-0.029}$
	slope	-0.005	$-0.005^{+0.001}_{-0.001}$
	offset	-91.695	$-91.663^{+0.190}_{-0.182}$

Table 4.4: Parameters obtained by the GPRN and the GP on the analysis of the BIS time series. On the GPRN's variables, the exponent represents if the variable is either from a node (n) or a weight (w).

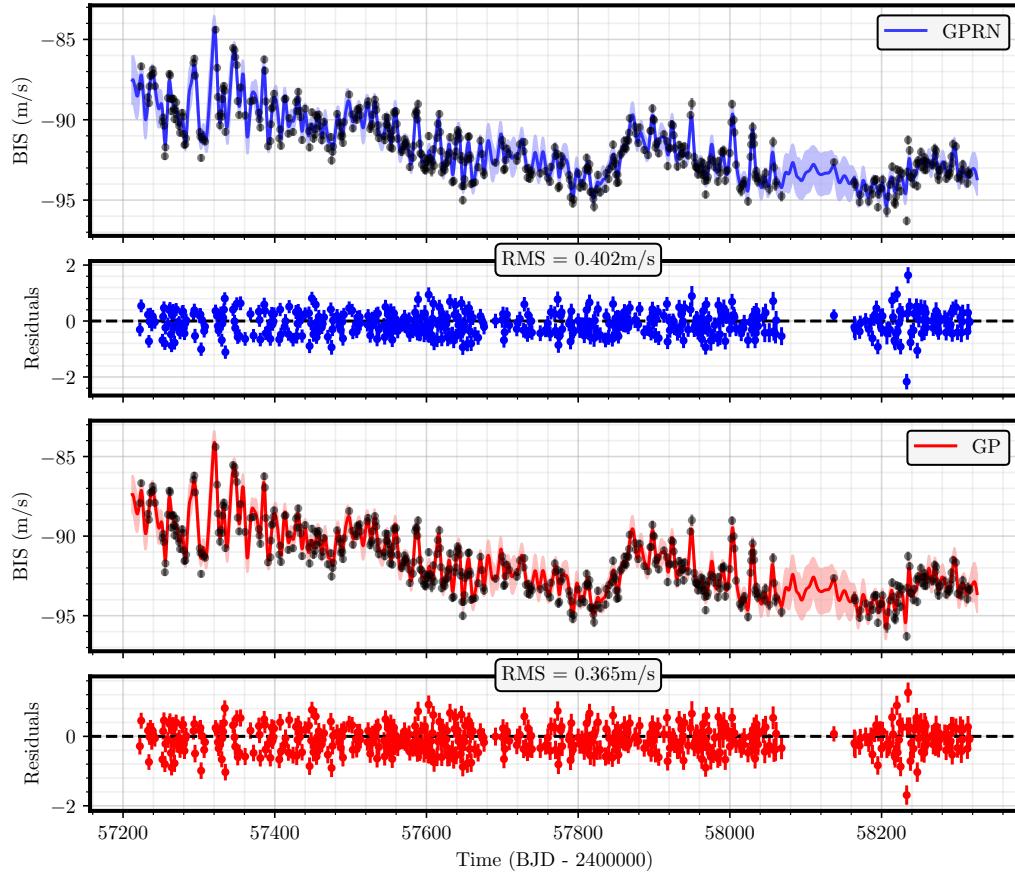


Figure 4.5: Fits and residuals obtained using the MAP values from the GPRN (blue) and the GP (red) on the BIS measurements.

Individual analysis - FWHM

As observed previously, the GP and GPRN obtained similar results, this time on the FWHM analysis. We show the parameters obtained by both models in table 4.5. The fits using each model's MAP values are in figure 4.6.

The GP kernel and the GPRN node parameters are, again, consistent within a 1-sigma interval. That explains both models determining a similar rotation period, around the 27-days mark. The values obtained on the weight, especially for its η_2 , indicate an almost constant weight. That will, of course, not change the quasi-periodic behaviour of the node significantly.

Unlike the previous two analyses, the FWHM time series does not have a long term trend, noticeable by figure 4.6. That, of course, has an impact on the slope of the mean function. In this analysis, the slope converged to values closer to zero.

The previously mentioned flexibility of the models, or lack thereof, is again observed. That means a higher RMS reduction on the GP. This model reduced the RMS by 2.528 times, while the GPRN reduced it only by 2.231 times. The evidence is again is advantageous for the GP. For it, we obtained a $\log \mathcal{Z}$ of -977.182 ± 0.657 . The GPRN, although far behind, achieved a $\log \mathcal{Z}$ of $-979.329.606 \pm 0.496$.

The convergence of the models showed an unexpected result. The GPRN, although again slower per iteration, required only 25000 steps to achieve convergence. The GP required 30000 iterations. This 5000 iterations difference, however, still wasn't enough to obtain results from the GPRN faster than on the GP. If we calculate the integrated autocorrelation time, the lengthy convergence on the GP was due to two parameters, the η_2 and the offset. Their τ values were 1054 and 1050, respectively, twice as long as the remaining parameters. The explanation for it, however, is not easy as their respective posterior distributions as they show no anomalies (figure C.5 of the appendix).

Framework	Variables	MAP value	Median value
GPRN	η_1^n	0.647	$1.777^{+5.770}_{-1.391}$
	η_2^n	19.001	$19.056^{+1.636}_{-1.776}$
	η_3^n	28.244	$27.982^{+0.613}_{-0.553}$
	η_4^n	0.731	$0.723^{+0.091}_{-0.089}$
	η_1^w	1.898	$1.733^{+5.978}_{-1.351}$
	η_2^w	1812.053	$4662.489^{+3829.389}_{-2618.007}$
	s	1.136	$1.129^{+0.060}_{-0.058}$
	slope	-0.001	$-0.001^{+0.001}_{-0.001}$
GP	offset	7785.094	$7785.015^{+0.237}_{-0.226}$
	η_1	1.905	$1.947^{+0.142}_{-0.126}$
	η_2	19.449	$19.123^{+1.732}_{-1.728}$
	η_3	27.903	$28.014^{+0.628}_{-0.582}$
	η_4	0.644	$0.720^{+0.091}_{-0.088}$
	s	1.093	$1.120^{+0.057}_{-0.058}$
	slope	-0.001	$-0.001^{+0.001}_{-0.001}$
	offset	7785.100	$7785.016^{+0.244}_{-0.247}$

Table 4.5: Parameters obtained by the GPRN and the GP on the analysis of the FWHM time series. On the GPRN’s variables, the exponent represents if the variable is either from a node (n) or a weight (w).

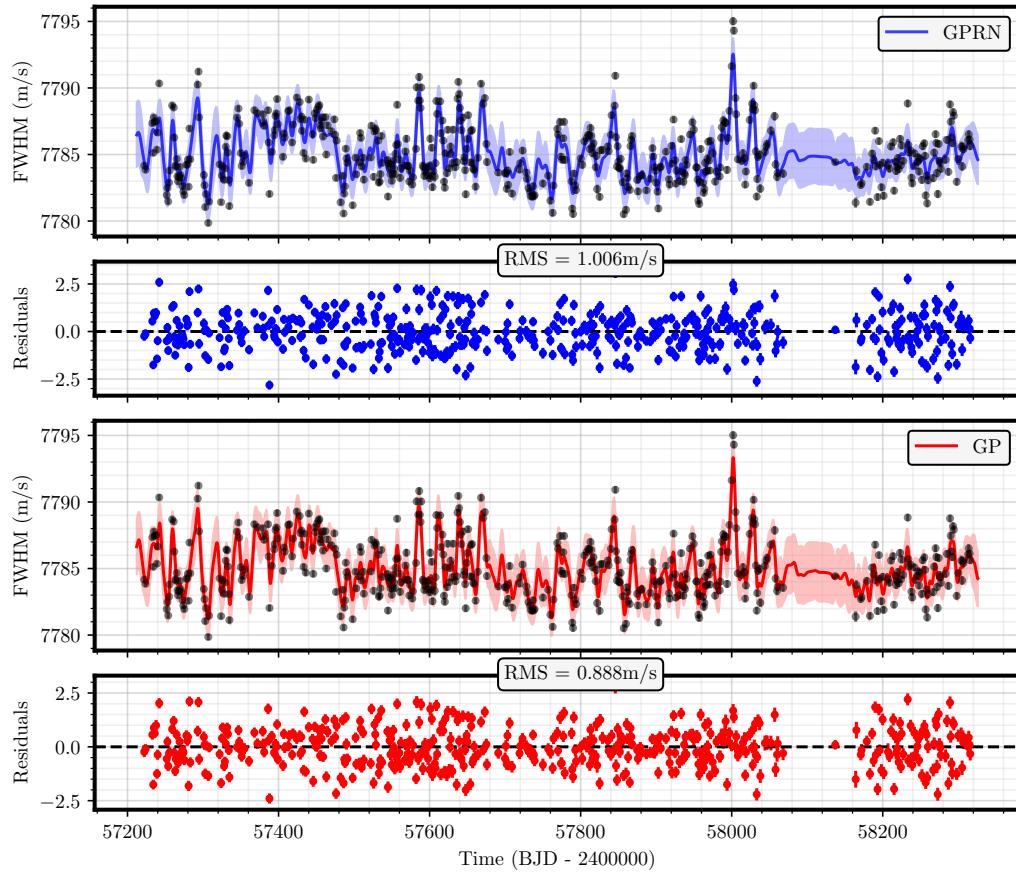


Figure 4.6: Fits and residuals obtained using the MAP values from the GPRN (blue) and the GP (red) on the FWHM measurements.

Individual analysis - $\log R'_{hk}$

The final individual analysis is that of the $\log R'_{hk}$ time series. Once again, both the GP and GPRN obtained consistent results. These are in table 4.6, while the fits using the respective MAP values are in figure 4.7.

Considering the MAP obtained values, both rotation periods are higher than 27 days but not far from the synodic rotation period. Also higher are the values for the η_4 length-scale. On both models, it surpassed the unit. On the GPRN, the weight η_2 timescale obtained the highest values of the four analyses. While the previous analysis had values between 1000 and 2500 days, the MAP value for the $\log R'_{hk}$ was of 4124.169 days, roughly 11.3 years. That gives all the modelling flexibility to the node.

It is necessary to mention that we obtained the highest decrease in RMS on all four time series. The GP had an RMS reduction of 10.101 times, and the GPRN had an RMS reduction of 9.756 times. The evidence shows once more the GP to be the best model obtaining a $\log \mathcal{Z}$ of 2188.970 ± 0.115 and the GPRN getting a $\log \mathcal{Z}$ of 2168.673 ± 0.384 . These values are quite different from those of the other analyses. One hypothesis is due to the overfitting of the models.

Lastly, the GP proved to be more computationally efficient. It required 10000 iterations to achieve our convergence criteria. The GPRN only managed to achieve convergence after 30000 iterations.

Framework	Variables	MAP value	Median value
GPRN	η_1^n	0.010	$0.050^{+0.042}_{-0.024}$
	η_2^n	20.530	$20.468^{+1.351}_{-1.426}$
	η_3^n	28.273	$27.425^{+0.720}_{-0.666}$
	η_4^n	1.074	$1.049^{+0.087}_{-0.079}$
	η_1^w	0.102	$0.050^{+0.042}_{-0.028}$
	η_2^w	4124.169	$5518.676^{+3541.486}_{-2991.382}$
	s	0.001	$0.001^{+9 \times 10^{-5}}_{-8 \times 10^{-5}}$
	slope	-4×10^{-5}	$-4 \times 10^{-5}^{+2 \times 10^{-6}}_{-2 \times 10^{-6}}$
	offset	-4.989	$-4.989^{+0.001}_{-0.001}$
GP	η_1	0.006	$0.006^{+0.001}_{-1 \times 10^{-4}}$
	η_2	21.211	$20.823^{+1.367}_{-1.357}$
	η_3	28.397	$28.414^{+0.661}_{-0.610}$
	η_4	1.052	$1.095^{+0.091}_{-0.079}$
	s	0.001	$0.001^{+8 \times 10^{-5}}_{-8 \times 10^{-5}}$
	slope	-4×10^{-5}	$-4 \times 10^{-5}^{+3 \times 10^{-6}}_{-3 \times 10^{-6}}$
	offset	-4.989	$-4.989^{+0.001}_{-0.001}$

Table 4.6: Parameters obtained by the GPRN and the GP on the analysis of the $\log R'_{hk}$ time series. On the GPRN's variables, the exponent represents if the variable is either from a node (n) or a weight (w).

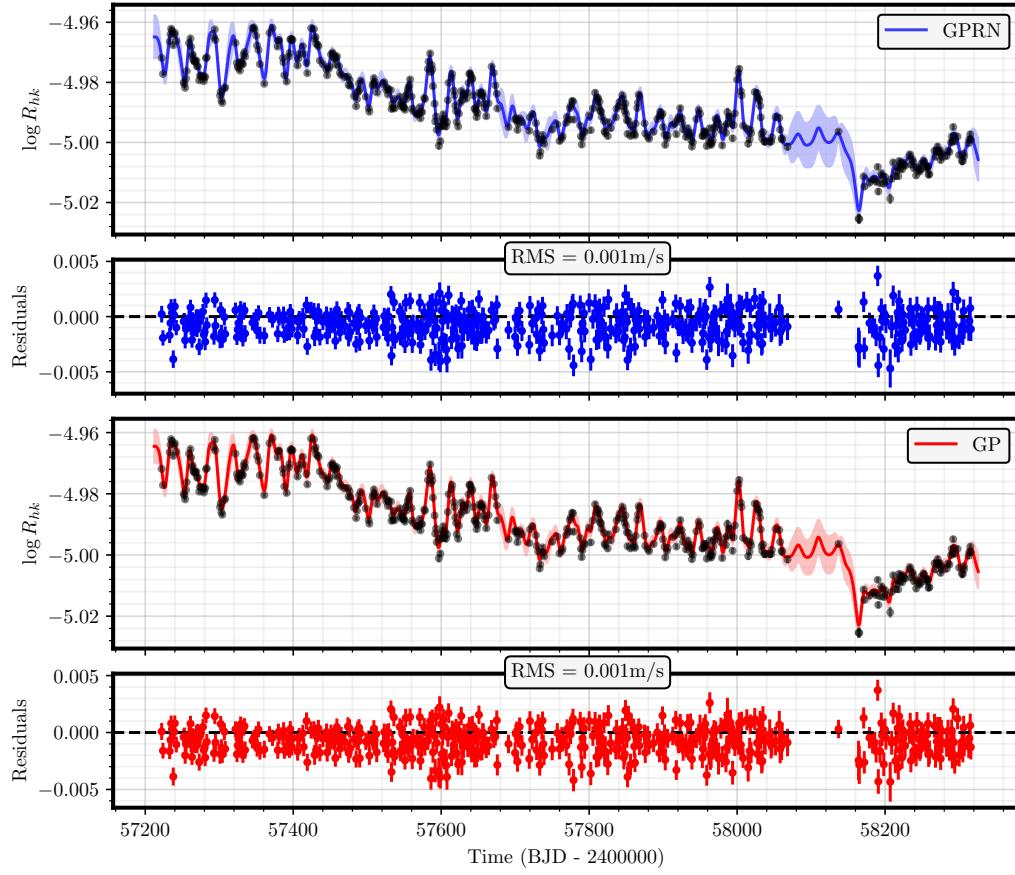


Figure 4.7: Fits and residuals obtained using the MAP values from the GPRN (blue) and the GP (red) on the $\log R'_{hk}$ measurements.

Combined analysis - RVs and BIS

We now present the combined analysis of the RV observations and an activity indicator. There are some significant changes from the previous results. For example, from now on, the GPRN model will have two η_2 time scales. That ends in a GPRN with 14 parameters, plus three than the GP used on the artGPN framework. However, under the right circumstances, these two models should be equivalent.

The first analysis combined the RV and BIS measurements. In table 4.7 we present the parameters obtained by the two models. In figures 4.8 and 4.9 are the fits obtained using the respective MAP values.

Even with this new setting, the equivalent parameters from the two frameworks are consistent. For example, both models manage to obtain the expected rotation period of the Sun given by η_3 . However, the GPRN plot of the posterior distributions in figure C.10 show samples of η_3 for values around 36-37 days and 44-45 days. These correspond to smaller values of η_4 but have a minimal impact on the final solution. One main difference exists on the jitter term of the RVs. The value obtained by the GPRN is higher than both the GP and the individual analysis done by the GPRN. The weights on the GPRN has an interesting behaviour. On the BIS datapoints, the η_2^w value indicates that the weight is becoming a constant function, the η_2^w on the RV is smaller. That might imply that that weight is trying to be more flexible and fit more structure than expected.

The fits of figures 4.8 and 4.9 show a higher RMS on the GPRN. This RMS difference is explicitly visible on the RV data points. One explanation is that the node is only good to fit one time series. If it is true and the node is modelling only the BIS, it leaves weight connected to the RVs to compensate and try to fit the data, thus obtaining a lower value for η_2^w . That would also explain the increase in the jitter value. In section 4.2.3 I will explore more of this issue.

The GP obtained better evidence with a $\log \mathcal{Z}$ of -1422.486 ± 0.195 . The value is comparable to the sum of evidence for the individual GP analysis of the RV and the BIS. The GPRN obtained a $\log \mathcal{Z}$ of -1471.618 ± 0.724 . This value is far lower than the sum of the evidence of its corresponding individual analysis. Again this reveals the GPRN to have a problem with this combined analysis.

130 Advanced statistical data analysis methods for the detection of other Earths

Computationally the artGPN analysis was faster in achieving convergence. The GP converged after 15000 iterations. The GPRN, now an even more complex model, converged after 100000 iterations.

Framework	Variables	MAP value	Median value
GPRN	η_1^n	1.163	$0.547^{+1.450}_{-0.371}$
	η_2^n	24.319	$23.647^{+1.640}_{-1.603}$
	η_3^n	26.745	$27.055^{+0.408}_{-0.390}$
	η_4^n	0.783	$0.744^{+0.058}_{-0.056}$
	RV η_1^w	0.413	$2.209^{+4.748}_{-1.626}$
	RV η_2^w	281.890	$2057.797^{+4117.223}_{-1677.437}$
	BIS η_1^w	0.669	$3.049^{+6.617}_{-2.247}$
	BIS η_2^w	1676.794	$5074.099^{+3661.307}_{-2668.754}$
	RV s	1.187	$1.204^{+0.042}_{-0.039}$
	BIS s	0.423	$0.438^{+0.032}_{-0.030}$
	RV slope	-0.005	$-0.005^{+3 \times 10^{-3}}_{-3 \times 10^{-3}}$
	RV offset	-19.239	$-19.177^{+0.108}_{-0.106}$
GP	BIS slope	-0.005	$-0.005^{+0.001}_{-0.001}$
	BIS offset	-91.780	$-91.780^{+0.172}_{-0.106}$
	η_2	22.478	$22.711^{+1.317}_{-1.241}$
	η_3	26.961	$26.849^{+0.341}_{-0.331}$
	η_4	0.696	$0.705^{+0.043}_{-0.041}$
	RV η_1	1.185	$1.214^{+0.095}_{-0.088}$
	BIS η_1	1.315	$1.336^{+0.086}_{-0.079}$
	RV s	0.801	$0.816^{+0.040}_{-0.037}$
	BIS s	0.374	$0.397^{+0.023}_{-0.028}$
	RV slope	-0.005	$-0.005^{+0.001}_{-5 \times 10^{-3}}$
	RV offset	-19.047	$-19.105^{+0.164}_{-0.160}$
	BIS slope	-0.005	$-0.005^{+0.001}_{-0.001}$
	BIS offset	-91.618	$-91.660^{+0.174}_{-0.175}$

Table 4.7: Parameters obtained by the GPRN and the GP on the analysis of the RV and the BIS time series. On the GPRN’s variables, the exponent represents if the variable is either from a node (n) or a weight (w).

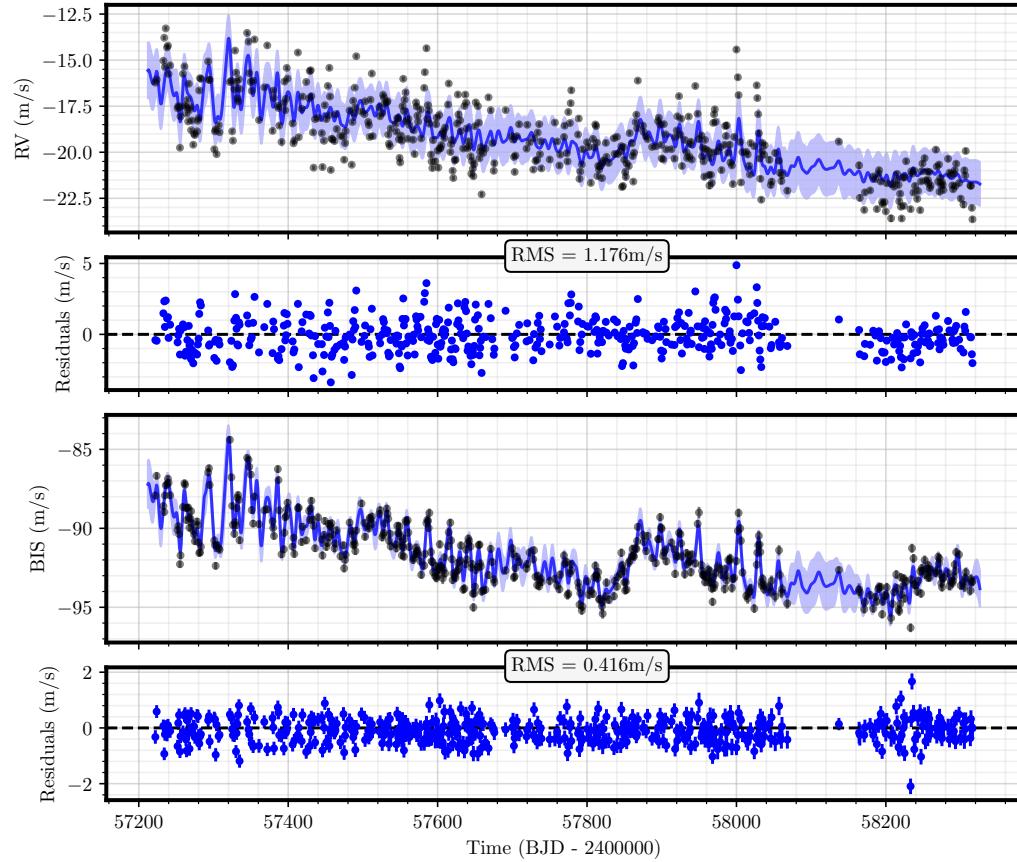


Figure 4.8: Fits and respective residuals obtained by the GPRN framework on the RV and the BIS time series.

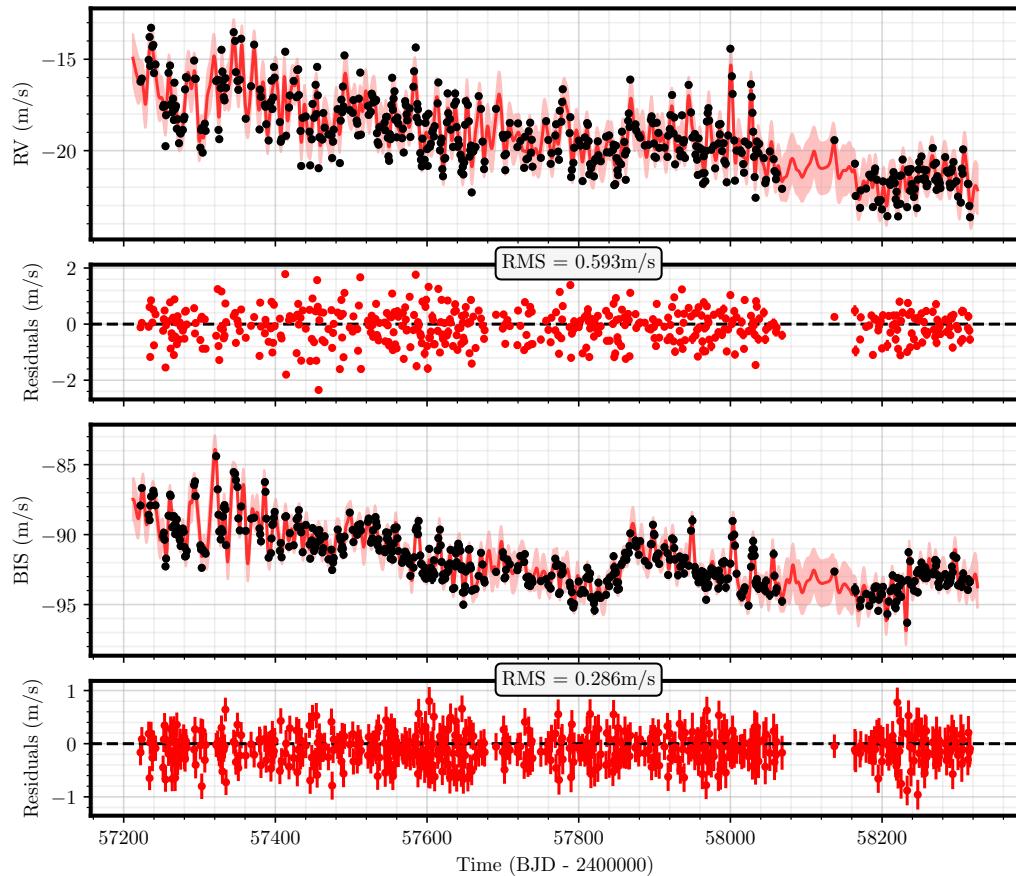


Figure 4.9: Fits and respective residuals obtained by the artGPN framework on the RV and the BIS time series.

Combined analysis - RVs and FWHM

The next analysis makes use of the RV and FWHM measurements. We present the parameters determined in table 4.8. The MAP values fit for the GPRN and the GP are in figures 4.10 and 4.11, respectively.

The two frameworks show a result similar to the previous analysis. Both determined the expected value for the rotation period of the Sun. Although the η_4 on the GPRN is higher than the artGPN value, the remaining parameters on the node are consistent within a 1-sigma interval. Similarly to previously, the weights of the GPRN have opposite behaviour. The weight of the FWHM shows to behave as a constant function. For the RVs, the η_2^w value is consistent with a weight with some structure.

The η_4 of the GPRN has an unusual relationship with offsets on the mean functions. By the posterior distributions of appendix C.12. Both offsets show a bimodal distribution. These correspond to two different regions of the η_4 parameter space. Curiously, the MCMC converged to the region of significantly higher η_4 and rejected the one corresponding to values of $\eta_4 \sim 0.7$.

The RMS obtained shows, again, to be higher when using the GPRN. For the FWHM, the artGPN decreased the RMS by 2.584 times, and the GPRN decreased the RMS by 2.068 times. That difference increases for the RVs modelling. The GPRN decreased the RMS by 1.917 times while the artGPN decreased it by 3.073 times. That again show the GPRN to have difficulties modelling two time series at once. This problem has a reflection on the evidence of the models. The GPRN obtained a $\log \mathcal{Z}$ of -1812.882 ± 3.339 , and the combined GP a $\log \mathcal{Z}$ of -1767.184 ± 0.105 . Computationally, the convergence issues between GP and GPRN continues. The joint GP analysis was faster converging after 15000 iterations. The GPRN took almost ten times more, with 105000 iterations.

Framework	Variables	MAP value	Median value
GPRN	η_1^n	1.951	$0.511^{+1.147}_{-0.340}$
	η_2^n	22.183	$20.779^{+1.934}_{-2.000}$
	η_3^n	27.643	$27.680^{+0.621}_{-0.581}$
	η_4^n	1.008	$0.861^{+0.120}_{-0.109}$
	RV η_1^w	0.440	$1.809^{+3.691}_{-1.257}$
	RV η_2^w	37.319	$66.105^{+56.031}_{-24.361}$
	FWHM η_1^w	0.675	$4.812^{+9.103}_{-3.383}$
	FWHM η_2^w	1660.697	$4673.599^{+3920.324}_{-2729.219}$
	RV s	1.074	$1.129^{+0.056}_{-0.0057}$
	FWHM s	1.229	$1.200^{+0.058}_{-0.056}$
	RV slope	-0.004	$-0.004^{+4 \times 10^{-3}}_{-4 \times 10^{-3}}$
	RV offset	-19.686	$-19.467^{+0.220}_{-0.265}$
GP	FWHM slope	-0.001	$-1 \times 10^{-3}^{+0.001}_{-0.001}$
	FWHM offset	7783.414	$7784.166^{+0.789}_{-0.716}$
	η_2	19.154	$19.674^{+1.302}_{-1.301}$
	η_3	27.336	$27.370^{+0.454}_{-0.425}$
	η_4	0.622	$0.663^{+0.056}_{-0.054}$
GP	RV η_1	1.152	$1.174^{+0.087}_{-0.077}$
	FWHM η_1	1.908	$1.939^{+0.130}_{-0.122}$
	RV s	0.804	$0.805^{+0.038}_{-0.037}$
	FWHM s	1.082	$1.098^{+0.057}_{-0.056}$
	RV slope	-0.005	$-0.005^{+4 \times 10^{-4}}_{-5 \times 10^{-4}}$
	RV offset	-19.080	$-19.116^{+0.148}_{-0.145}$
	FWHM slope	-0.001	$-0.001^{+0.001}_{-0.001}$
	FWHM offset	7785.114	$7785.028^{+0.237}_{-0.234}$

Table 4.8: Parameters obtained by the GP and the GPRN on the analysis of the RV and the FWHM time series. On the GPRN's variables, the exponent represents if the variable is either from a node (n) or a weight (w).

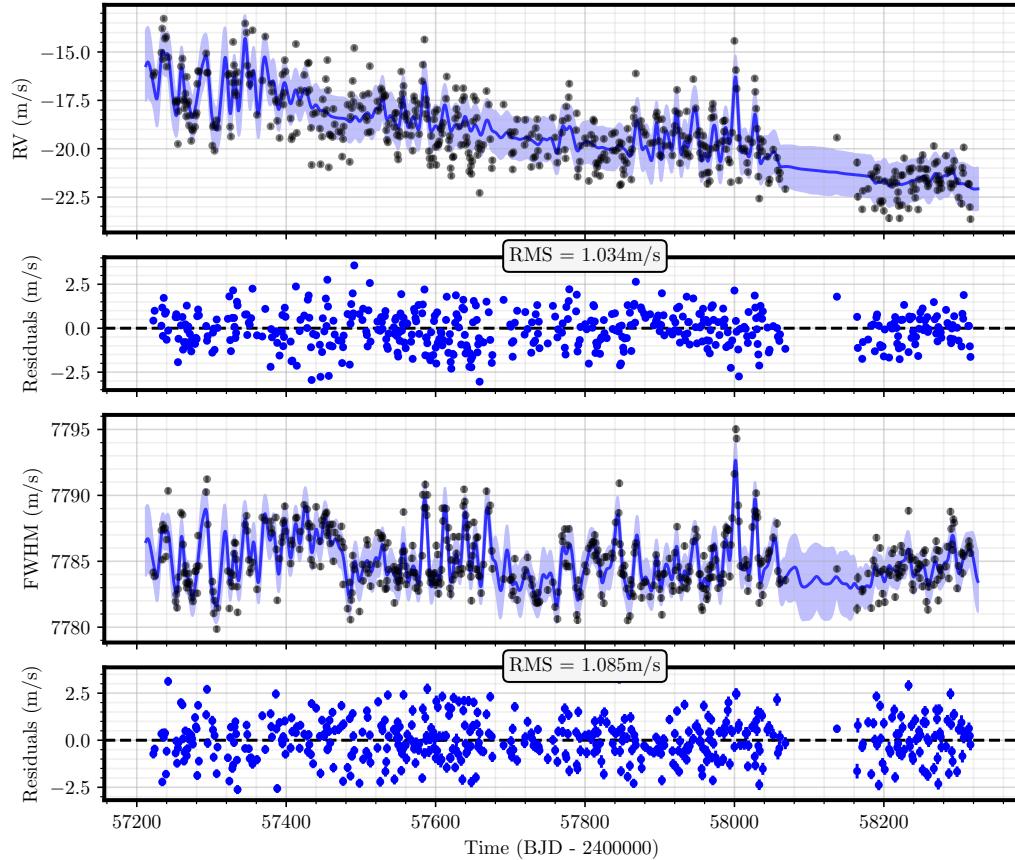


Figure 4.10: Fits and respective residuals obtained by the GPRN framework on the RV and the FWHM time series.

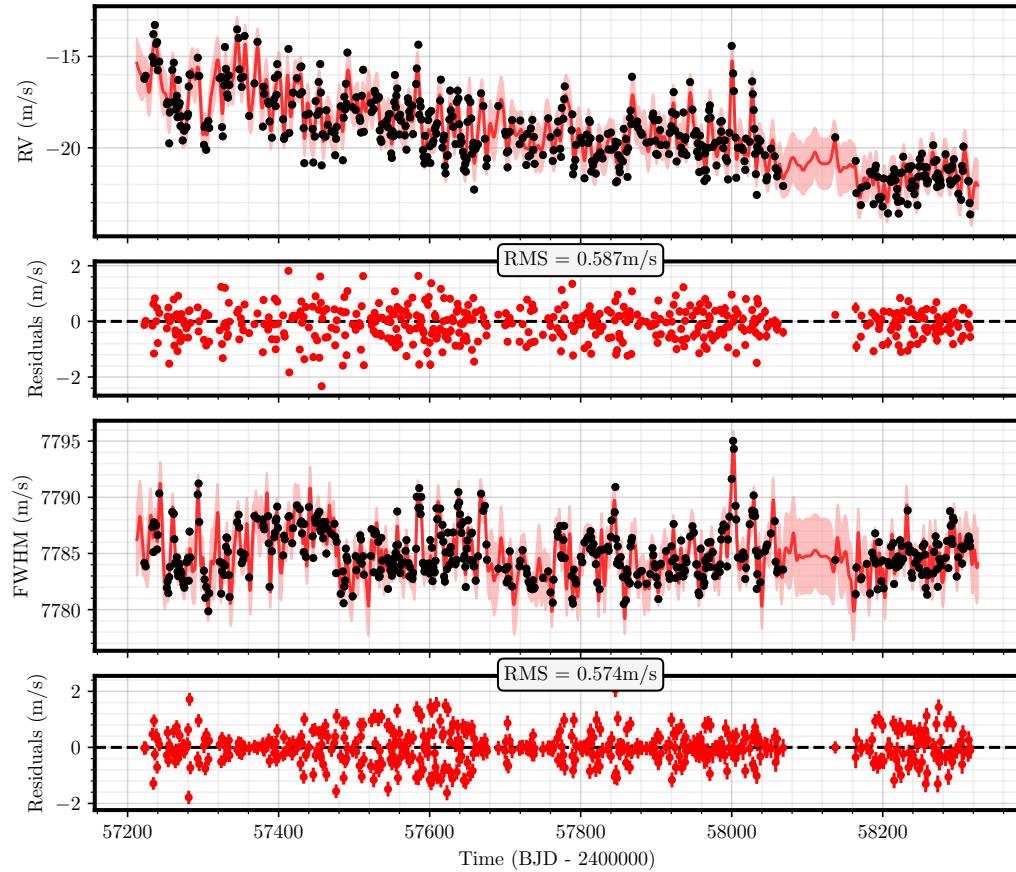


Figure 4.11: Fits and respective residuals obtained by the artGPN framework on the RV and the FWHM time series.

Combined analysis - RVs and $\log R'_{hk}$

The last analysis combines the RV and $\log R'_{hk}$ time series. The derived parameters on the GPRN and GP are in table 4.9. The respective fits are plotted in figure 4.12 and figure 4.13.

Again the related parameters on both frameworks are consistent, with similar values for the rotation period and the activity regions time scales. The weights of the GPRN reveal, again, a behaviour similar to the previous analysis. The weight associated with the $\log R'_{hk}$ behaves as a quasi constant function. The weight connected to the RV measurements has the opposite result. That is its η_2 value that creates a significant structure on the weight.

The figures 4.12 and 4.13 show the GPRN again, more easily fitting the activity indicator with the node. That results in the RV weight compensating the fit on the RV data points. That has an impact on both RMS and evidence of the models. The GPRN has a worse RMS and lower evidence. The evidence obtained on the GP was 1394.419 ± 0.141 , far higher than the one of the GPRN. For the GPRN framework, the $\log Z$ had the value of 1368.207 ± 5.843 . The GPRN decreased the RV and the $\log R'_{hk}$ by 1.797 and 9.628 times, respectively. The values for the GP were 2.745 and 10.757 times, respectively. The decreases in the $\log R'_{hk}$ might be due to the overfitting of the model.

Computationally the efficiency is similar to the previous analysis. The GP managed to converge after 15000 iterations, while the GPRN required 100000 iterations.

Framework	Variables	MAP value	Median value
GPRN	η_1^n	0.001	$0.073^{+0.053}_{-0.027}$
	η_2^n	19.654	$20.498^{+1.362}_{-1.392}$
	η_3^n	28.615	$28.316^{+0.706}_{-0.651}$
	η_4^n	1.057	$1.078^{+0.088}_{-0.082}$
	RV η_1^w	5.939	$10.086^{+5.870}_{-4.432}$
	RV η_2^w	86.682	$88.988^{+21.555}_{-19.312}$
	$\log R'_{hk} \eta_1^w$	0.024	$0.059^{+0.037}_{-0.027}$
	$\log R'_{hk} \eta_2^w$	3041.504	$5898.438^{+3312.965}_{-3012.180}$
	RV s	1.126	$1.123^{+0.039}_{-0.037}$
	$\log R'_{hk}$ s	0.001	$0.001^{+9 \times 10^{-5}}_{-9 \times 10^{-5}}$
GP	RV slope	-0.005	$-0.005^{+3 \times 10^{-4}}_{-3 \times 10^{-4}}$
	RV offset	-19.400	$-19.374^{+0.098}_{-0.099}$
	$\log R'_{hk}$ slope	-4×10^{-5}	$-4 \times 10^{-5}^{+2 \times 10^{-6}}_{-2 \times 10^{-6}}$
	$\log R'_{hk}$ offset	-4.990	$-4.990^{+0.039}_{-0.037}$
	η_2	21.318	$21.135^{+1.059}_{-1.076}$
	η_3	27.559	$27.651^{+0.427}_{-0.419}$
	η_4	0.904	$0.924^{+0.059}_{-0.057}$
	RV η_1	1.257	$1.294^{+0.110}_{-0.105}$
	$\log R'_{hk} \eta_1$	0.006	$0.005^{+4 \times 10^{-4}}_{-4 \times 10^{-4}}$
	RV s	0.864	$0.850^{+0.039}_{-0.037}$
GP	$\log R'_{hk}$ s	0.001	$0.001^{+9 \times 10^{-5}}_{-8 \times 10^{-5}}$
	RV slope	-0.005	$-0.005^{+0.001}_{-0.001}$
	RV offset	-19.167	$-19.102^{+0.193}_{-0.187}$
	$\log R'_{hk}$ slope	-4×10^{-5}	$-4 \times 10^{-5}^{+3 \times 10^{-6}}_{-3 \times 10^{-6}}$
	$\log R'_{hk}$ offset	-4.989	$-4.989^{+0.001}_{-0.001}$

Table 4.9: Parameters obtained by the GP and the GPRN on the analysis of the RV and the $\log R'_{hk}$ time series. On the GPRN's variables, the exponent represents if the variable is either from a node (n) or a weight (w).

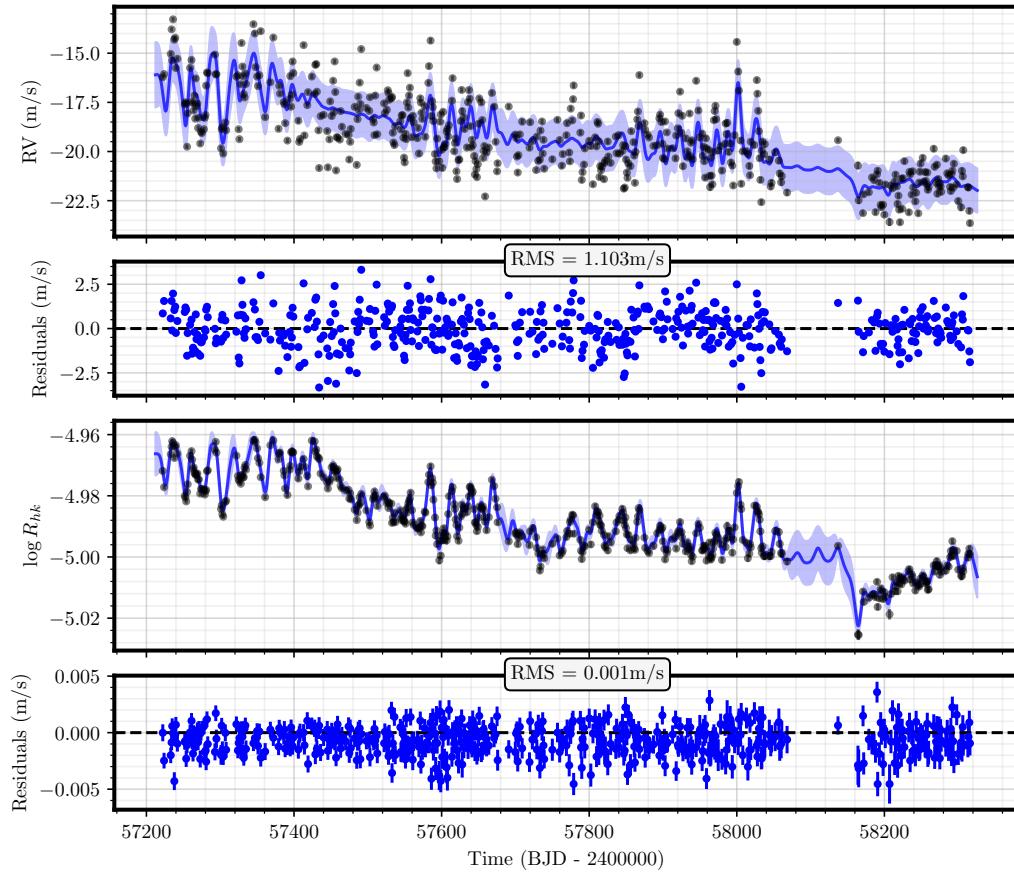


Figure 4.12: Fits and respective residuals obtained by the GPRN framework on the RV and the $\log R'_{hk}$ time series.

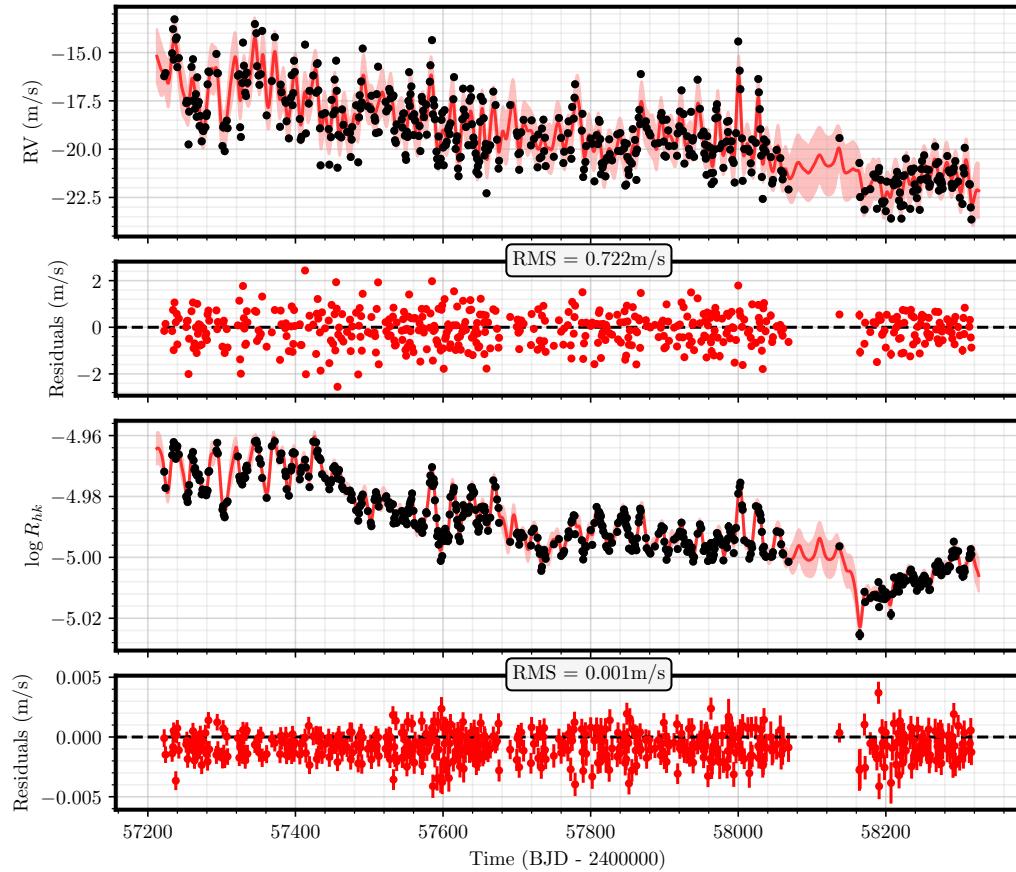


Figure 4.13: Fits and respective residuals obtained by the artGPN framework on the RV and the $\log R'_{hk}$ time series.

4.2.3 Interpretation of the results

The previous pages focused only on presenting the performance of the GPRN and comparing it to a traditional GP. It briefly mentioned the impact of parameters like, for example, the decaying timespan η_2 . That, however, was not backed by any further explanation. This section will thus explore and summarize the results. Since the conclusions of the different analyses can be generalized, we focus on the individual RV analysis and the combined RV and FWHM analysis.

Is vital to understand the behaviour of a GPRN. Knowing the GPRN is constituted by weights and nodes defined as independent GPs, its structure is more complex than a traditional GP. That makes it harder to understand the impact each node and weight have on the final result. To help solve it, plotting each GPRN component separately can help explain the results previously achieved.

The individual analysis of the RV time series from 4.2.2 shows that a GPRN can obtain similar results to a GP. The high value obtained on the η_2^w parameter indicates the weight as becoming a quasi-constant function. That, of course, implies it will impact the amplitude of the node but not change its structure significantly. Plotting each GPRN component alone confirms this. From figure 4.14 we can see a quasi-constant weight and an extremely quasi-periodic node. The only impact of the weight is on changing the amplitude on the node while it models the structure present on the RV data points resulting in the middle plot of figure ???. The final offset and slope are accounted using the mean function resulting on the right-side plot of the figure.

That occurs in all individual analyses as the η_2^w value forces the weight to become constant. However, this relation between the node and the weight does not explain the worse performance the GPRN showed. However, two hypotheses exist that might explain it.

As mentioned in chapter 3.3, inference in a GPRN is intractable. The solution to this issue is to use mean-field inference to find an approximation to the posterior. As a measure of that approximation, mean-field inference uses the ELBO. That is an acronym for evidence lower bound that, as the name indicates, is a lower bound approximation used to evaluate the posterior. It has been proved, albeit not shown here, that the ELBO is the lower bound to the true log-evidence (e.g.

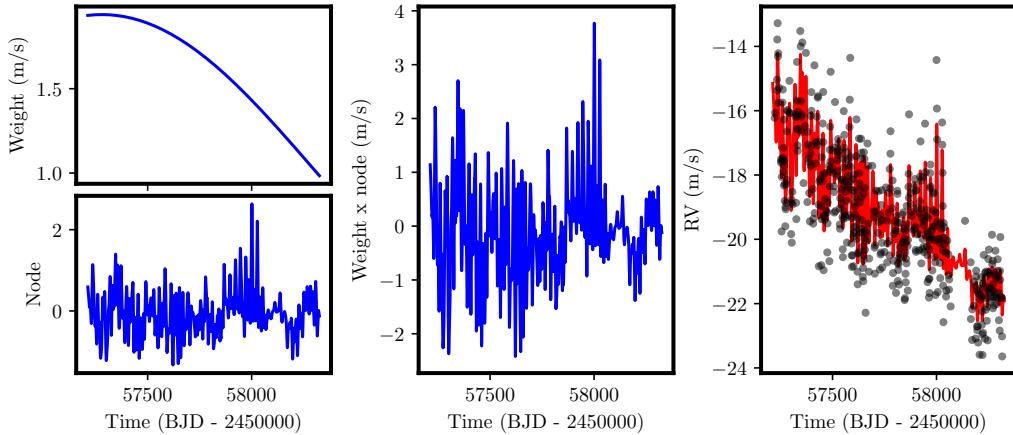


Figure 4.14: Left: Plots of the weight and node using the respective MAP values from the RV analysis. Middle: Product on the weight and node. Right: GPRN fit on the RV data points with the obtained parameters.

Bishop, 2006).

Knowing this implies that the evidence calculated for a GPRN using the Perrakis method is also a lower bound. That ends in that the true evidence value on a GPRN is likely to be greater than the value calculated. That creates two issues that require a solution. First, it is necessary to determine how far this value might be from the correct evidence. Also important to know is if the difference is constant or random for different tests.

Another approximation, closer to the true value, can be calculated for the evidence. Although not feasible to efficiently implement on the GPRN, for a given set of data points and a given set of GPRN parameters, we can calculate it using nested sampling (Skilling, 2006).

That can be tested using, for example, the MAP parameter from table ??, and define a set of nodes and weights. It is then possible to calculate the ELBO and the nested sampling evidence for any number of data points. For the nested sampling calculations, it is easy to use the functions implemented on DYNESTY (Speagle, 2020). The plot with the results is in figure 4.15.

As expected, figure 4.15 show an ELBO constantly lower than the value

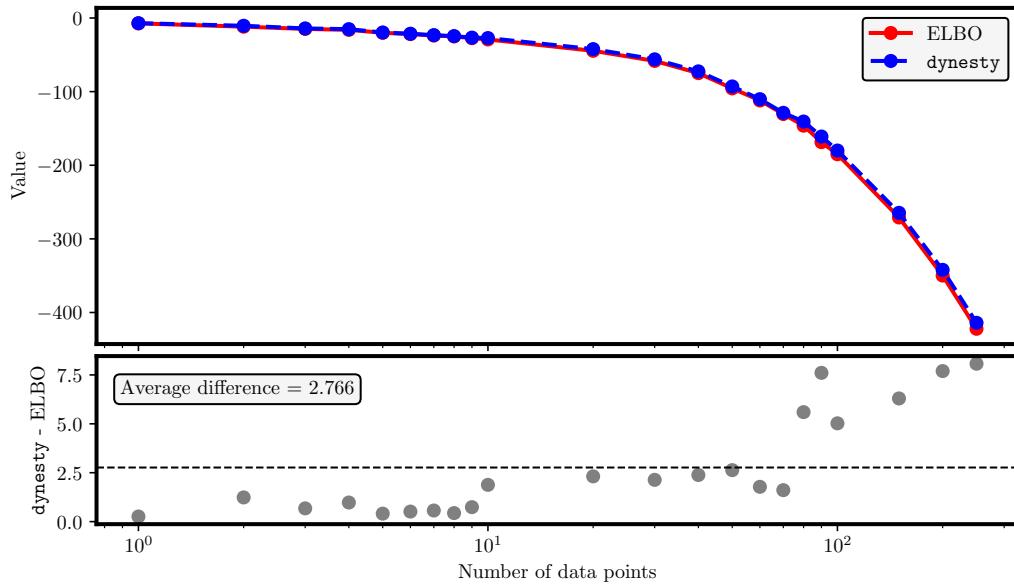


Figure 4.15: Top: ELBO and $\log \mathcal{Z}$ values given by the variational inference and DYNESTY, respectively. Bottom: Difference between the $\log \mathcal{Z}$ and the ELBO. The average value of the difference, of approximately 1.852, is plotted by the dashed horizontal line.

obtained from DYNESTY. On average, the ELBO from mean-field inference and the value calculated by `dynesty` is 2.766. That difference also looks to increase with the number of data points. If this behaviour is usual, the evidence calculated with the GPRN will likely never surpass the values obtained with a GP of equal characteristics.

Another issue common in our analysis is that the results constantly better residuals on the GP models. That, however, can be explained by the way the GPRN is defined. We stated in chapter 3 that a GPRN will not easily over-fit a model. That being due to, and as explained on Wilson et al. (2012), a GPRN has a heavy-tailed predictive distribution. Knowing the prior distribution of a GPRN has a Bessel distribution, it is fair to assume the posterior distribution will also follow a Bessel distribution.

That being true, it creates a heavy-tailed posterior distribution on the GPRN.

That will make its predictive not require to approximate itself to the data as much as the predictive distribution of a GP. The tails from the Bessel function is enough to reach it. That implies the RMS will, most likely, be worse of a GPRN, even when the evidence supports this model.

These two hypotheses are valid for all the analyses made. However, the tests using the RVs and activity indicators measurements, the GPRN behaviour became more difficult to understand. To explain it, we will use the results we obtained on 4.2.2. On that analysis, we used the RV and FWHM measurements. As mentioned, the parameters obtained on the GPRN indicated a quasi-constant weight connected to the FWHM, while weight connected to the RV presented some structure. That is true and observable in figure 4.16. The figure shows each component of the GPRN, in that case, the node and the two weights. The difference in the weights seems related to the structure of the node. In the FWHM case, the node is relatively well behaved to fit the data. That leaves the weight tp only model the overall amplitude. However, for the RV data points, the node is no longer good to represent the data. The weight connected to the RVs has no other choice but to compensate for the node bad fit. That causes then the MCMC to converge to smaller values of η_2^w . Only lower values of the decaying timespan allow the weight to change significantly, becoming more "wiggly".

However, this extra effort from the weight is not enough to overcome the worse fit of the node on the RVs. The first logical explanation is that using a squared-exponential kernel is not enough. Other kernels like the rational quadratic or Matérn kernel might indeed be better. Unfortunately, due to time constraints, there was no possibility to test this.

In all our analyses, the node always chose to fit the activity indicator. Even with a sufficiently malleable weight, that does not explain why the node adapts itself more easily to the activity indicator and not the RV. One explanation is that there might be more information on the activity indicator than on the RV. That begin, a clearer quasi-periodic signal on the indicators. Only further tests can confirm this.

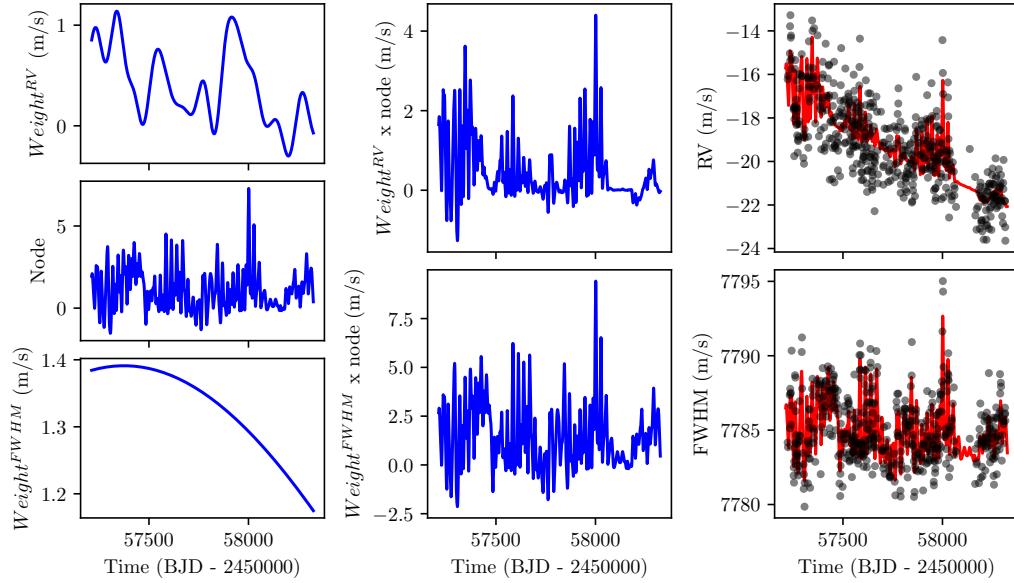


Figure 4.16: Left: Structure of each component of the GPRN. Middle: Product of the node by each respective weight. Right: GPRN fit to the RV and FWHM when using the MAP values obtained on the analysis.

Time lags

As mentioned in section 4.2.1, the activity indicators show a delay with the RV observations. One aspect of interest is how significant these time lags were to our analyses.

To determine the lags of the two frameworks used on the combined analysis of section 4.2.2. The results obtained allow the use of the predictive means to lag capture by the frameworks. The objective is to compare the prediction of the RV points with the prediction for an activity indicator and calculate if any time shift exists. Figure 4.17 shows the results.

On the right are plotted the predictive means of the RV and the indicator accordingly the framework used. To calculate the CCF uses the return values of each predictive. The plots on the right, however, are adjusted for better visualization. The mean function is removed, and each predictive is divided by the

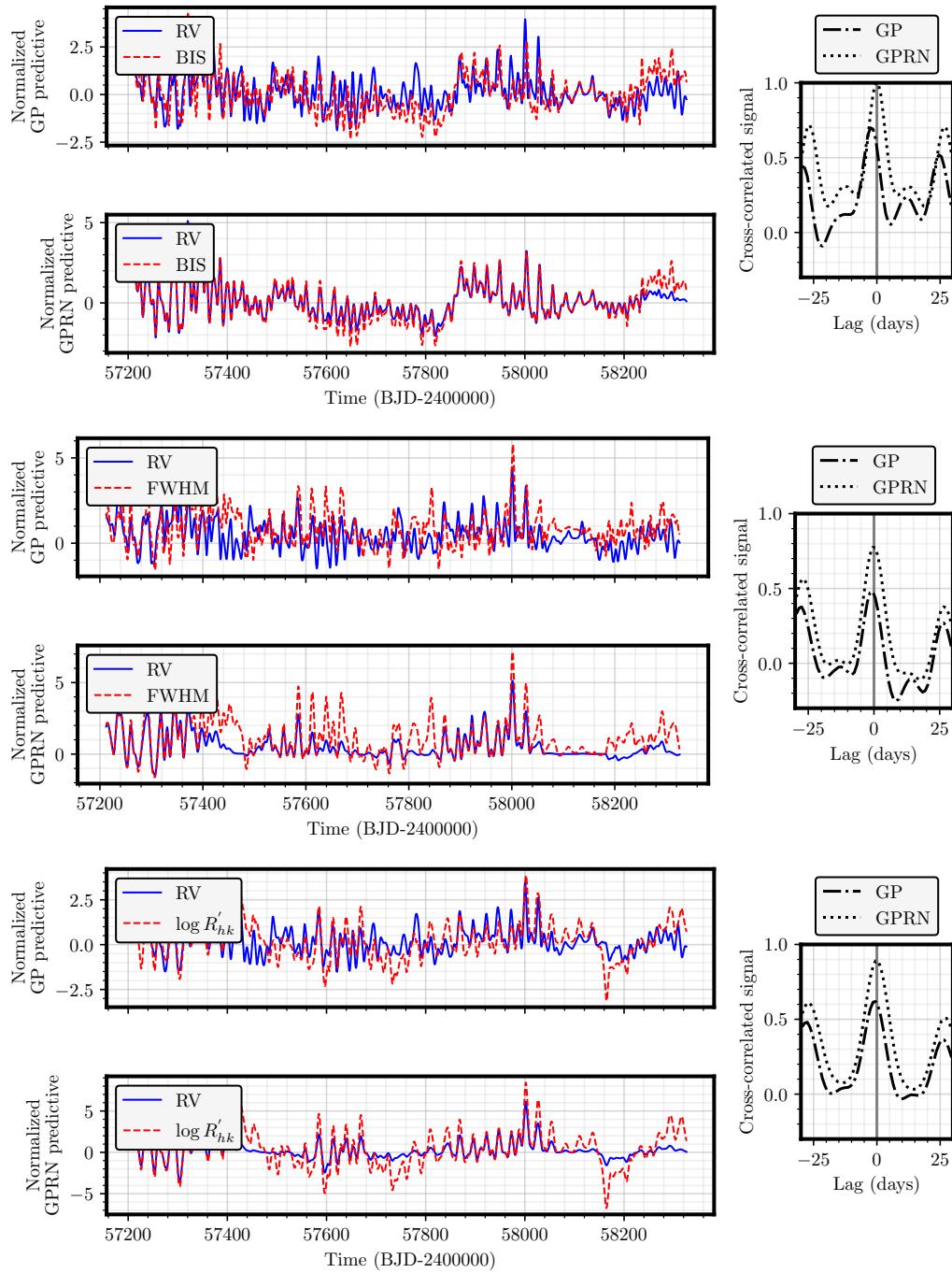


Figure 4.17: Left: Superimposed normalized predictives from GP and the GPRN. Right: Discrete CCF between the predictives of the RV and the activity indicator.

respective η_1 value. That allows to normalize each predictive and more easily compare them visually. That makes it possible, for example, to observe a small shift in the GP predictions of the RV and BIS. The value of these shifts, however, need to be calculated. For it, we used the discrete correlation function of Edelson and Krolik (1988).

On the right of figure 4.17 are the cross-correlation function plots. Each of these plots is between a GP and a GPRN predictive that they represent. These show that the GPRN is incapable of modelling any time lag between the time series. On the contrary, the GP framework shows shifts between the RVs and the activity indicators. For the BIS, the lag is of 2 to 3 days. The GP framework also models a 1-day lag between the RV and FWHM. For last, the GP results show no significant shift between the $\log R'_{hk}$ and RV.

The lack of flexibility on the GPRN to model these shifts helps explain the worse evidence and RMS results. The framework is not capable of shifting one of the time series concerning another originating a worse fit. That added to the GPRN fitting only one of the time series with the node when it tries to fit the remaining time series with the weight compromises all the framework.

These problems observed on the GPRN need to be solved. That might require either to use a GP on the weight with a parameter capable of detecting lags between time series or to adapt the variational inference algorithms to detect the time lags and shift the inputs accordingly. Since the combined GP analysis with the artGPN framework uses an independent GP per time series that share the parameters, it might explain the capability of dealing with phase shifts. Unfortunately, it does not help to find a solution for the lack of flexibility on a GPRN.

4.3 EXPRES data

In this section, we present the analysis of four stars from the EXPRES Stellar-Signals Project (ESSP) (Zhao et al., 2020). The ESSP aimed at comparing several spectroscopic data analysis methods. Thanks to it, we had access to spectroscopic observations carried by the Extreme Precision Spectrograph (EXPRES) (Jurgenson et al., 2016). In total, we had four targets to apply our GPRN framework to it, HD 10700, HD 26965, HD 101501, and HD 34411.

The GPRN used both RV and FWHM to analyse the observations and determine the scatter return from the RV fit. The RMS reduction of the GPRN would then be compared with the other methods to determine the level of success in mitigating the stellar variability. All the methods and results obtained will be published in Zhao et. al. (submitted).

This challenge proved a valuable opportunity for us. It allows us to test the GPRN on stars with different levels of stellar activity and test its detection limits. It also grants us the opportunity to compare our framework with other time-domain methods. The GPRN created used a quasi-periodic GP on the node and a squared exponential GP on each weight. The priors are defined similarly to the priors used in the Sun analysis. These are shown on section 4.1.2.

4.3.1 HD 10700

HD 10700, or commonly known as τ Ceti, is a G8V type star. The $\log R'_{hk}$ value of -4.976 shows it to be chromospherically inactive, with a rotation period of 34.5 days (Keenan and McNeil, 1989; Pizzolato et al., 2003; Isaacson and Fischer, 2010). Four planets dynamically packed have been detected to orbit this star (Feng et al., 2017).

The 174 RV measurements available are shown on figure 4.18. The observations have an RMS of 1.863 m/s and a timespan of 470 days. Performing a periodogram analysis shows that most of the peaks are below the 1% false alarm probability level (figure 4.18). Its low activity level might imply there will not be significant activity contaminating the RV measurements with spurious periodic

signals. That being true, it might prove difficult for the GPRN of obtaining any meaningful result.

After running an MCMC for 225000 iterations, the GPRN achieved convergence. However, it was not capable of, for example, finding the expected rotation period. It instead obtained a period of $5.809^{+1.545}_{-0.562}$ days, with a MAP value of 5.350 days. The lack of stellar activity might have contributed to these results. If spots or other stellar phenomena are not present on the surface, the rotation period signal on the observation would be minimum.

The fit on figure 4.19 also indicated an incapacity of the GPRN to model the RV data. The node has an extremely low decaying timespan (~ 2.715 days). It adapts itself exclusively to the activity indicator, leaving no flexibility to fit the RV. That resulted in an insignificant reduction in the RMS. That meant a reduction of only 2.652% concerning the initial RV measurements. On the FWHM, we had a far more significant reduction on the RMS. It was reduced by almost four times, from 11.761 m/s to 2.944 m/s.

These results might prove a bad performance of the GPRN. However, these might be the expected behaviour for observations with no significant activity signals. The main idea behind a GPRN is to have a non-stationary model capable of modelling the non-stationary signals from stellar activity. Considering the low activity level of HD 10700, there will be a limited activity signal contribution on the RV. That implies no quasi-periodic signals due to stellar activity on the measurements.

Considering the previous performance of the GPRN on the Sun observations, we wanted to confirm the results on HD 10700. To determine if the lack of activity was the reason behind the results of the GPRN, we analysed the data with the artGPN framework from 3.2.3.

The new analysis achieved similar performance. It failed to obtain a rotation period as well. The MAP value of it was of 5.624 days, similar to the GPRN. The fits obtained with these GP are on appendix D.1. These are very similar to the GPRN, with the RV again badly modelled. That indicates all the results indeed due to the lack of quasi-periodic signals on the data.

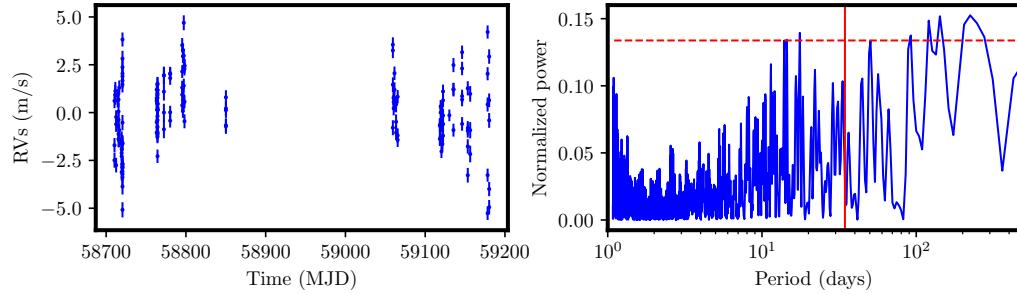


Figure 4.18: Left: RV observations of HD 10700. Right: Periodogram of the data after subtracting a long term trend. Also on it is marked, as a red vertical line, the expected rotation period of 34.5 days, and the FAP of 1% as a dashed horizontal line.

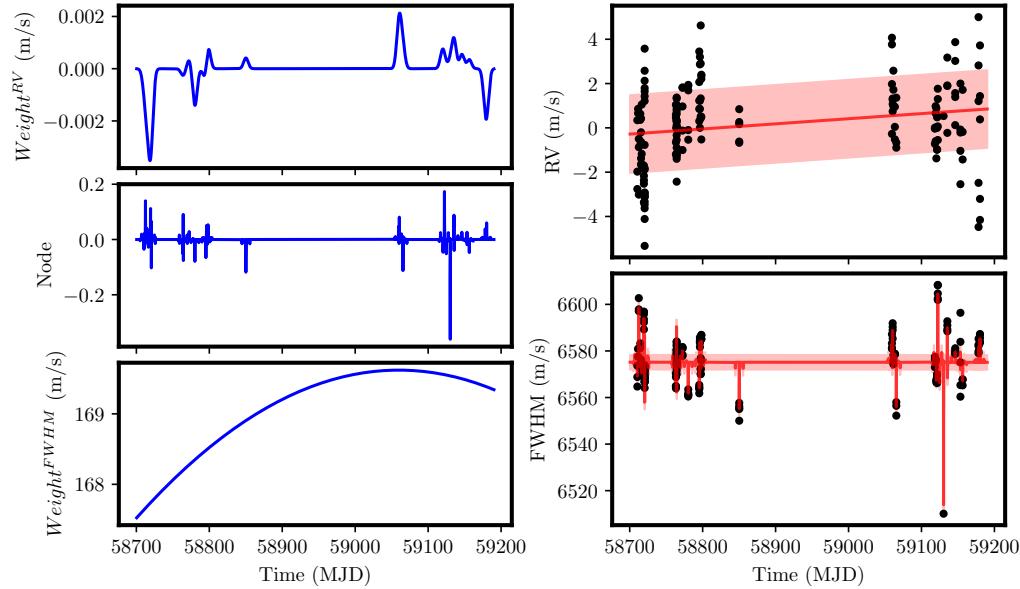


Figure 4.19: Left: Components of the GPRN. The $weight_{RV}$ connects the node to the RV, while the $weight_{FWHM}$ connects it to the FWHM. Right: Final fits of the GPRN to the RV and the FWHM.

4.3.2 HD 34411

HD 34411, also known as λ Aurigae, is a solar-type star classified as a G1V (Boyajian et al., 2012). This star is the least active of the four in analysis. It has a $\log R'_{hk}$ of -5.085 dex, and no known rotation period (Isaacson and Fischer, 2010). Such low value for the $\log R'_{hk}$, indicative of low magnetic activity, makes it a candidate star for a Maunder minimum event on its activity cycle (Lubin et al., 2012).

The low stellar activity can be observed on the periodogram of figure 4.20. On it, there is no peak below 100 days above the 1% FAP threshold. Similar to HD 10700, the lack of activity had an impact on the parameters determined by the GPRN. It was unable to converge to a well-defined period. From its marginal posterior probability distribution, the η_3 remained unconstrained with a final value of $29.425^{+13.165}_{-17.776}$ days. Its MAP value was as of 44.104 days, likely not related to the true rotation period. However, the GPRN still managed to model some structure on the RV. This resulted on a slight decrease of the RMS from 1.768 m/s to 1.421 m/s . On the FWHM, the fit was more successful with a decrease on the RMS from 5.244 m/s to 1.538 m/s .

We show the GPRN fit of the RV and the FWHM in figure 4.21. The node modelling indicator, leaving the weight to fit the RV as best as possible. We also observed problems modelling the RV with a combined GP. See appendix D.2 for the obtained fits with this framework. The GP managed a slightly better RV RMS decrease to 1.181 m/s and an converged value of η_3 of $10.543^{+13.711}_{-1.146}$ days. If this value is related to the rotation period is unknown. The results with a GP again show that the GPRN is not wrong but unable to obtain meaningful results due to lack of activity signals on the measurements.

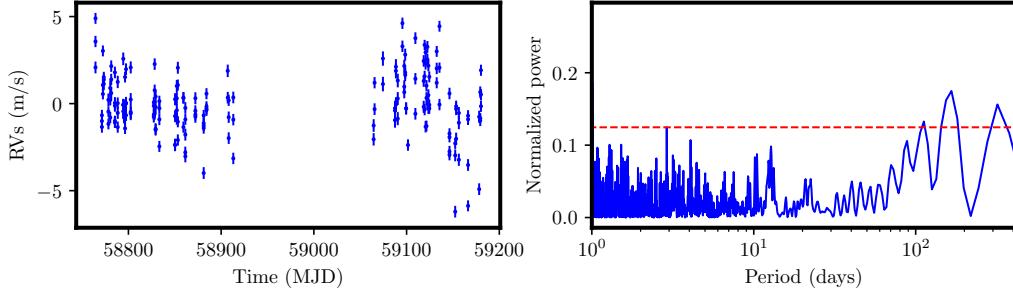


Figure 4.20: Left: RV observations of HD 34411. Right: Periodogram of the data after subtracting a long term trend. Also on it is marked the FAP of 1% as a dashed horizontal line. No rotation period is known for this star.

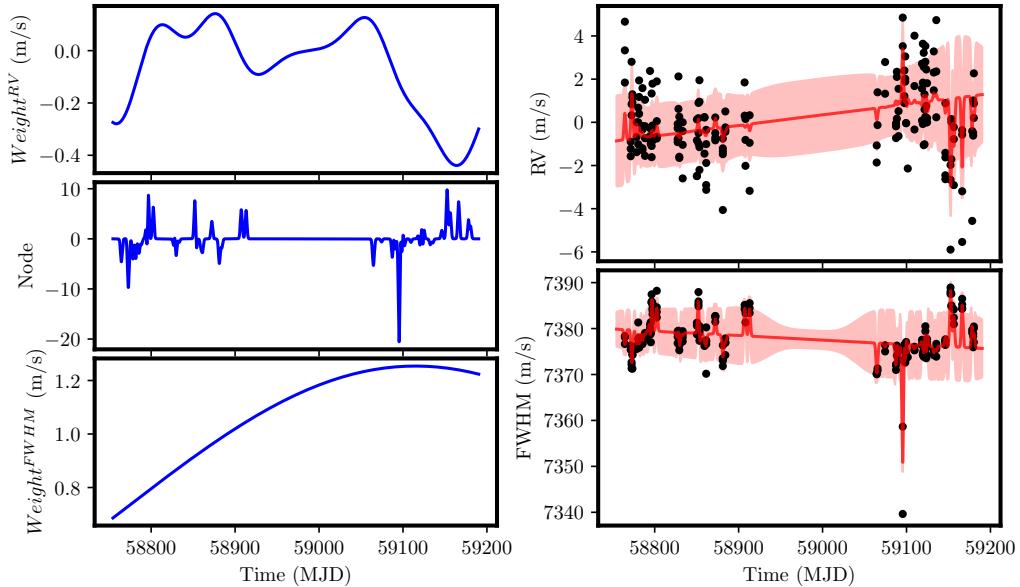


Figure 4.21: Left: Contribution of the node and weights to the final fit. The weight_{RV} connects the node to the RV, while the weight_{FWHM} connects it to the FWHM. Right: GPRN fit to the measurements of HD 34411.

4.3.3 HD 101501

HD 101501, also known as 61 Ursae Majoris, is a late main-sequence star classified as a G8V type star (Wilson, 1962). Of the four stars in this analysis is the most active one. It has a $\log R'_{hk} = -4.483$ and a rotation period of 17.1 *days* (Isaacson and Fischer, 2010; Maldonado et al., 2010). No known planet exist on this star.

The 45 available observations are scattered through a timespan of 655 days. These are plotted on figure 4.22. On the periodogram of the RV observations, the peak due to the rotation period is noticeable as having the highest power. However, another significant peak exists at 12 days of unknown origin. Although not shown here, these peaks are also present on the FWHM periodogram. That seems to indicate that the 12 days peak is related to the activity.

The RV data had an RMS scatter of 4.887 *m/s*, the highest value of the four stars. That supports the high activity level of HD 101501. The GPRN-framework converged at the end of 500000 iterations. The η_3 determine was of $18.557^{+21.571}_{-4.835}$ *days*, consistent with the rotation period. A long tail for a higher period was present, related to higher values of the decaying timespan of the weight connected to the FWHM. For higher values of η_2 , the η_3 becomes unconstrained, forming a tail on the final posterior distribution. This tail, however, also represented regions of lower log-posterior values. The MAP value of the η_3 was 14.981 *days*. The MAP solution of the η_2 on the node was similar, with a value of 14.888 *days*. That is possibly related to the growth and decay of the active regions.

The RV RMS saw a decrease from 4.920 *m/s* to just 2.156 *m/s*. That represents a decrease of 2.279 times, showing the GPRN to be performing better on this star. The fits of both RV and FWHM are in figure 4.23. On it is seen the usual GPRN choice of fitting better the activity indicator with the node. For the activity indicator, the RMS decreased from 24.780 *m/s* to 1.753 *m/s* almost 14 times its initial value. That might be the first proof that the GPRN performs better in more active stars.

We also compared the GPRN with a combined GP analysis. The GP reached a rotational period smaller than expected. On it, the median value was

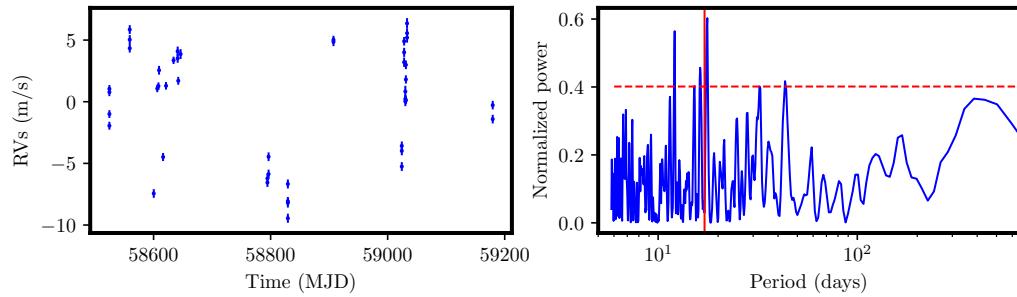


Figure 4.22: Left: RV observations of HD 101501. Right: Periodogram of the data after subtracting a long term trend. Also on it is marked, as a red vertical line, the expected rotation period of 17.1 days, and the FAP of 1% as a dashed horizontal line.

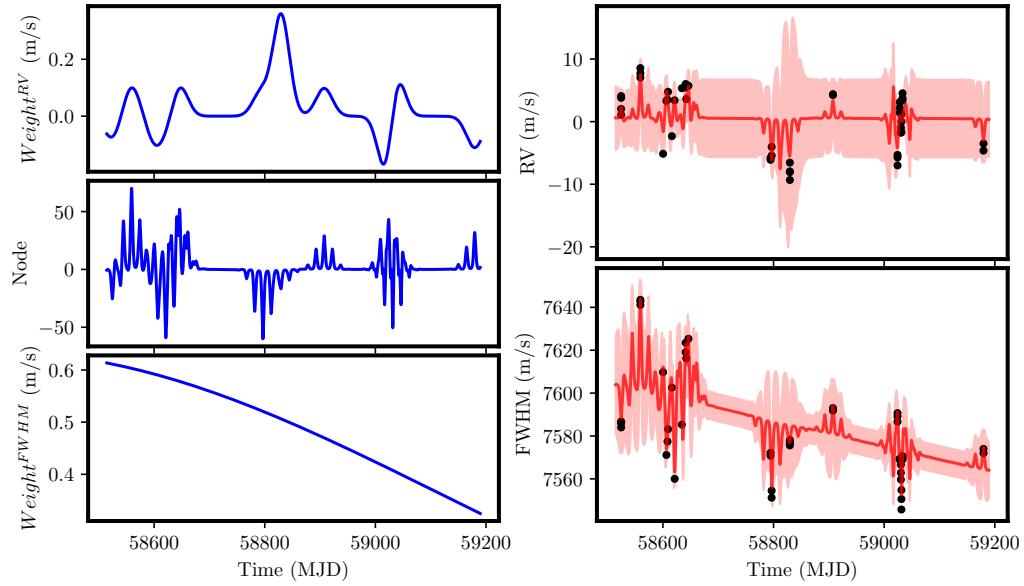


Figure 4.23: Left: Contribution of the node and weights to the final fit. The $weight_{RV}$ connects the node to the RV, while the $weight_{FWHM}$ connects it to the FWHM. Right: GPRN fit to the measurements of HD 101501.

$14.986^{+11.853}_{-0.884}$ days, and the MAP value of 14.374 days. The tail present in this solution is related to the η_4 value. Higher values of η_3 correlate with smaller values of η_4 . These, however, represent solutions of lower log-posterior values. The GP managed to decrease the RMS on the RVs to the value of 0.692 m/s.

4.3.4 HD 26965

HD 26965, also known as 40 Eridani, is a K1V star with a rotation period between 39 and 44.5 days (Ma et al., 2018). It has a $\log R'_{hk}$ of -4.928 dex, making it the second most active star of the four (Pizzolato et al., 2003; Isaacson and Fischer, 2010).

Still debatable is the existence of an exoplanet. A planetary signal with a period of approximately 42 days was detected (Díaz et al., 2018; Ma et al., 2018). However, an orbital period so close to the rotation period of the star makes its existence suspicious.

We had 114 observations with a timespan of 465 days. The RV measurements showed an RMS scatter of around 3.234 m/s. The high RMS and increasing RV amplitude over time (see figure 4.24) show an increase in stellar activity.

The periodogram shows the region of the expected rotation period populated by a group of peaks with significant power. That might explain the difficulties found in determining an accurate rotation period and the still-open debate of if the planetary signal discovered of 42 days is due to stellar activity or not. With a slightly higher power on the periodogram are a set of peaks around 17 days. That is likely related to harmonics due to the periodic signals around the rotation period. We see the same behaviour on the periodogram of the FWHM measurements.

The GPRN ran for 120000 iterations until the MCMC achieved convergence. In the end, we obtained a rotation period of $31.146^{+5.818}_{-13.171}$ days. The corresponding MAP value was 32.785 days. That although lower than expected, is still inside the region dominated by significant peaks on the periodogram. The node η_2 obtained a MAP value of 24.112 days, likely related to the evolution of the active regions. As such, this η_3 value is likely related to the rotation period.

The fit, shown on figure 4.25, resulted on a RMS decrease from 3.234 m/s

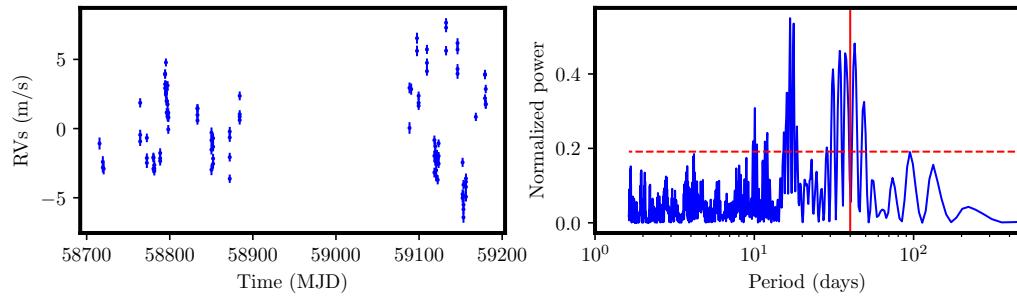


Figure 4.24: Left: RV observations of HD 26965. Right: Periodogram of the data after subtracting a long term trend. Also on it is marked, as a red vertical line, the expected rotation period of 40 days, and the FAP of 1% as a dashed horizontal line.

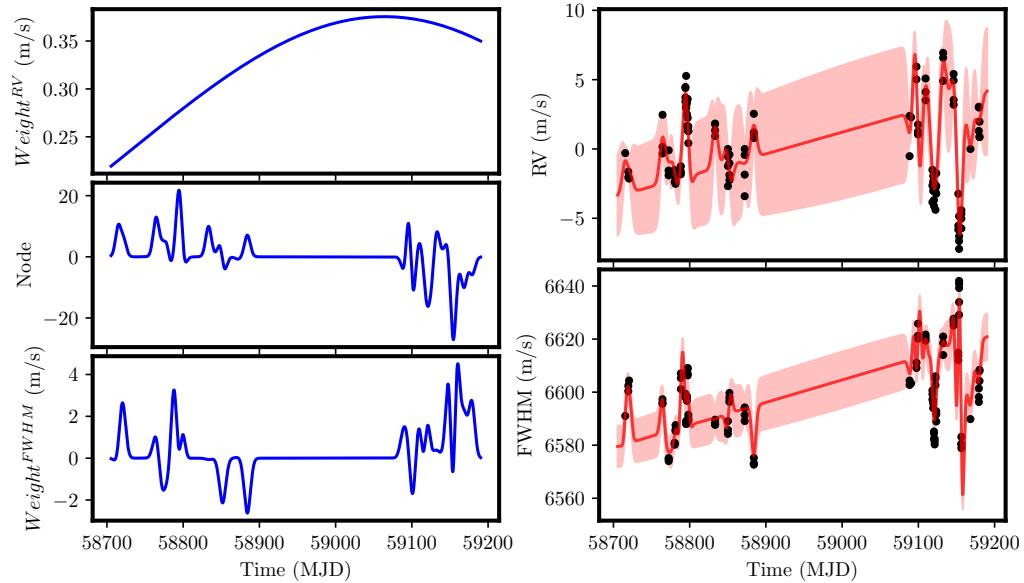


Figure 4.25: Left: Contribution of the node and weights to the final fit. The weight_{RV} connects the node to the RV, while the weight_{FWHM} connects it to the FWHM. Right: GPRN fit to the measurements of HD 26965.

to just 0.815 m/s . That meant an RMS reduction of almost four times its original value. The fit of the GPRN also showed an unexpected result. The behaviour of the node and weight connected to the RV data seems to indicate it is choosing to fit the RV. To compensate for this, the weight connected to the FWHM show a rougher structure to adapt itself to the activity indicator. If the GPRN is absorbing the periodic signals, including the claimed planetary signals, is debatable.

However, if the GPRN is modelling any planetary signals, the same occurs with a combined GP analysis. That analysis obtained a rotation period of $36.983^{+8.402}_{-9.701} \text{ days}$. It also decreased considerable the RMS of the RV, achieving the value of 0.750 m/s . If this signifies the absorption of planetary signals is to be done on future work.

4.3.5 Final thoughts on the EXPRES results

The four stars analysed for the ESSP proved a unique opportunity to test the GPRN on stars of different activity levels. These results indeed show the GPRN to have a better or worse performance accordingly to the activity level of the star. With its non-stationary structure, it is expected to the GPRN to perform better for higher activity levels. That is indeed the case. However, an even better performance was when using a combined GP analysis with the artGPN framework.

Similarly to the case study of the Sun, the GPRN shows an unexpected lack of flexibility that needs to be solved. Considering the limitations seen on the GPRN, addressing them will only improve its performance. Thus, we can see the results of the GPRN as a lower bound of its capabilities.

For the stars with known exoplanets, the next step for a GPRN analysis would be to include one or more keplerian functions. For the current python implementation of the GPRN, it is easy to define them on the mean function. Unfortunately, time constraints didn't allow more careful analysis of it.

Chapter 5

Conclusions

This thesis objective was to use advanced statistical data analysis methods to detect and characterize exoplanets using RV observations. It aimed at improving the detection limits of exoplanets with mass and orbital period similar to the Earth on solar-type stars. To achieve it, we adapted the works of Wilson et al. (2012) and Nguyen and Bonilla (2013) to use it on radial velocity and respective activity indicators data.

A GPRN, being a non-stationary framework, is an improvement when compared with frameworks like the ones of Rajpaul et al. (2015) or Jones et al. (2017). Its application on radial velocity data, however, proved to be more challenging than anticipated. The framework defined by Wilson et al. (2012) did not meet all the requirements, and changes were needed. In the end, we used the mean-field inference for GPRN developed by Nguyen and Bonilla (2013). That, however, didn't come without its constraints. The main one is the variational inference used and the approximations used on its algorithms. Even so, we end this work secure the results showed in this thesis are robust and in agreement with other methods.

In chapter 3, and before introducing the GPRN, we explain the kernels usually used on GP regression. There, we introduced an idea still unexplored when using GP in exoplanet detection. Astronomers can create new covariance functions with interesting new properties. Experimenting with newly derived periodic or quasi-periodic kernels might lead to results not achieved before compared to the "traditional" quasi-periodic covariance function. In the third chapter, we also

present all the mathematical properties needed to create a GPRN. It was tried not to exaggerate the number of derivations required to reach the necessary final solutions. However, some expressions would be easier to understand knowing all the steps used to obtain them. The appendix of Nguyen (2015) is probably the best source to better understand some of them.

The main results achieved in chapter 4 using the GPRN are slightly frustrating. The vast majority of the time employed on this project was to adapt the GPRN and write all the necessary algorithms into a python package. Only at the start of last year, there was a fully functional GPRN. We immediately used it on the Sun observations from Dumusque et al. (2021). However, the sheer number of measurements and the computational complexity of the GPRN meant we only had one shot at obtaining results for the time frame we had. In the end, some tests are future work due to this. We opted by applying a GPRN on its simplest form of one node and one weight per dataset. That does not take full advantage of the capabilities of a GPRN. Every time a GPRN was compared with a GP, the GP came out on top. Considering it occurred when using a limited GPRN considering the number of nodes and weights, the future of the GPRN is still promising.

The GP models consistently achieved better evidence and a lower RMS. Still, the GPRN results agree with those of the GP. For example, the GPRN managed to recover the Sun's rotation period the same way the GP did. Considering all the limitations, maybe on a different combination of nodes and weights, the GPRN might be better than the traditional GP.

5.1 Future work

The results show that the GPRN implemented, although functional, needs to be improved. Using a GPRN with just one node and one weight is not enough to compete against a GP. In this form, the GPRN cannot, for example, successfully model or compensate for time lags between the time series. If the issues found should be addressed by using weights with a different kernel and/or different priors or increasing the number of nodes is up for debate.

Knowing these time lags exist is necessary to create a GPRN capable of identifying them. Considering the analysis with two time series, using two nodes could be sufficient for the GPRN to achieve this. If the GPRN used or focused each node on different time series, the individual predictive means should, in principle, be shifted. However, only future tests will allow us to determine this.

When this project started, one of the main objectives was to use it on a set of observations and determine how well we would capture a planetary signal. The use of a Keplerian on the mean function of the GPRN remains untested. If a GPRN improves the detectability of planetary signals compared to a standard GP is a crucial step to do in the future. Ideally, combining the GPRN with, for example, the diffusive nested sampling algorithm (Brewer et al., 2011) would create a framework capable of inferring the GPRN hyperparameters and the number of planets.

Another issue while using a GPRN is its implementation and complexity. Some computational improvements are required. Due to its intractability, a GPRN approximates the posterior distributions via mean-field inference. This inference method is not unique, and we cannot exclude the possibility of improving our results using other variation inference methods. That could improve the parameters obtained with our model (one node and one weight per dataset) and help us achieve more precise model evidence. The current python package is also highly inefficient. The mathematical complexity and amount of matrices inversions required is a setback that currently has no solution. One way of speeding the calculation would be writing the GPRN with another programming language, e.g. C++.

Appendices

Appendix A

Sampled datasets

A.1 Priors

Variables	Parameter (units)	Prior
η_2^n	Decaying timespan (days)	$\mathcal{LU}(t_{AV}, 10 \times t_{PTP})$
η_3^n	Period (days)	$\mathcal{U}(10, 50)$
η_4^n	Length scale	$\mathcal{LU}(0.1, 5)$
η_1^w	Amplitude (m/s)	$\mathcal{MLU}(y_\sigma, 2 \times y_{PTP})$
η_2^w	Decaying timespan (days)	$\mathcal{LU}(t_{AV}, 10 \times t_{PTP})$
s	White noise amplitude (m/s)	$\mathcal{MLU}(y_\sigma, 2 \times y_{PTP})$
offset	Offset of the mean function (m/s)	$\mathcal{U}(y_{min}, y_{max})$

Table A.1: Parameters and prior distribution used on the MCMC analysis. $\mathcal{U}(a, b)$ stands for a uniform distribution between a and b . $\mathcal{LU}(a, b)$ stands for log-uniform distribution with shape parameters a and b . \mathcal{MLU} stands for modified log-uniform distribution, a log-uniform distribution that allows the support of the distribution to include zero. On the parameters of the distributions, y stands for the input and t the time. t_{AV} is the average time between consecutive observations. σ is the standard deviation of the vector used, while PTP stands for it peak-to-peak value. We set the amplitude of the node (η_1^n) to 1m/s to simplify the models used.

A.2 Scatterplot matrices

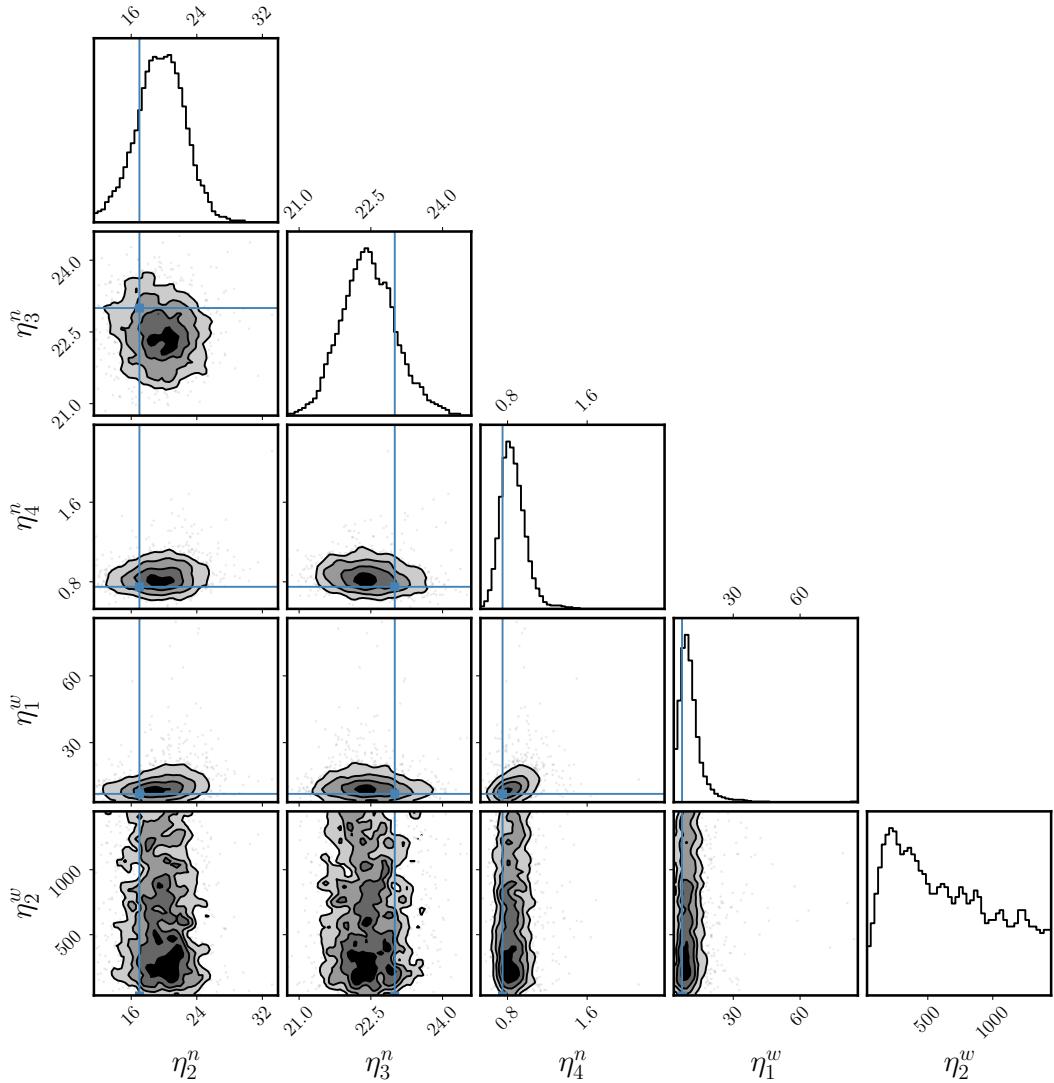


Figure A.1: Posterior distributions of the parameter from the 1st GPRN model.

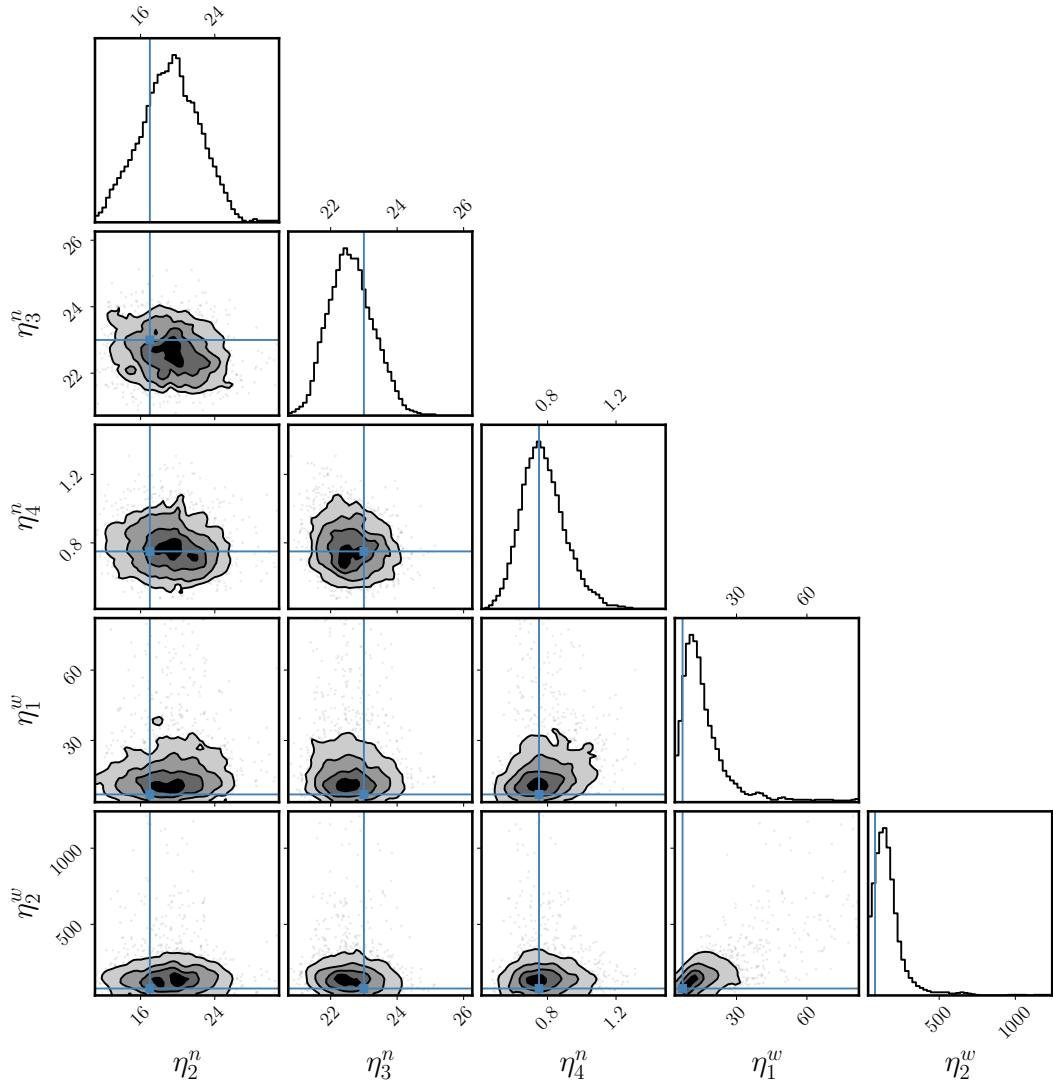


Figure A.2: Posterior distributions of the parameter from the 2nd GPRN model.

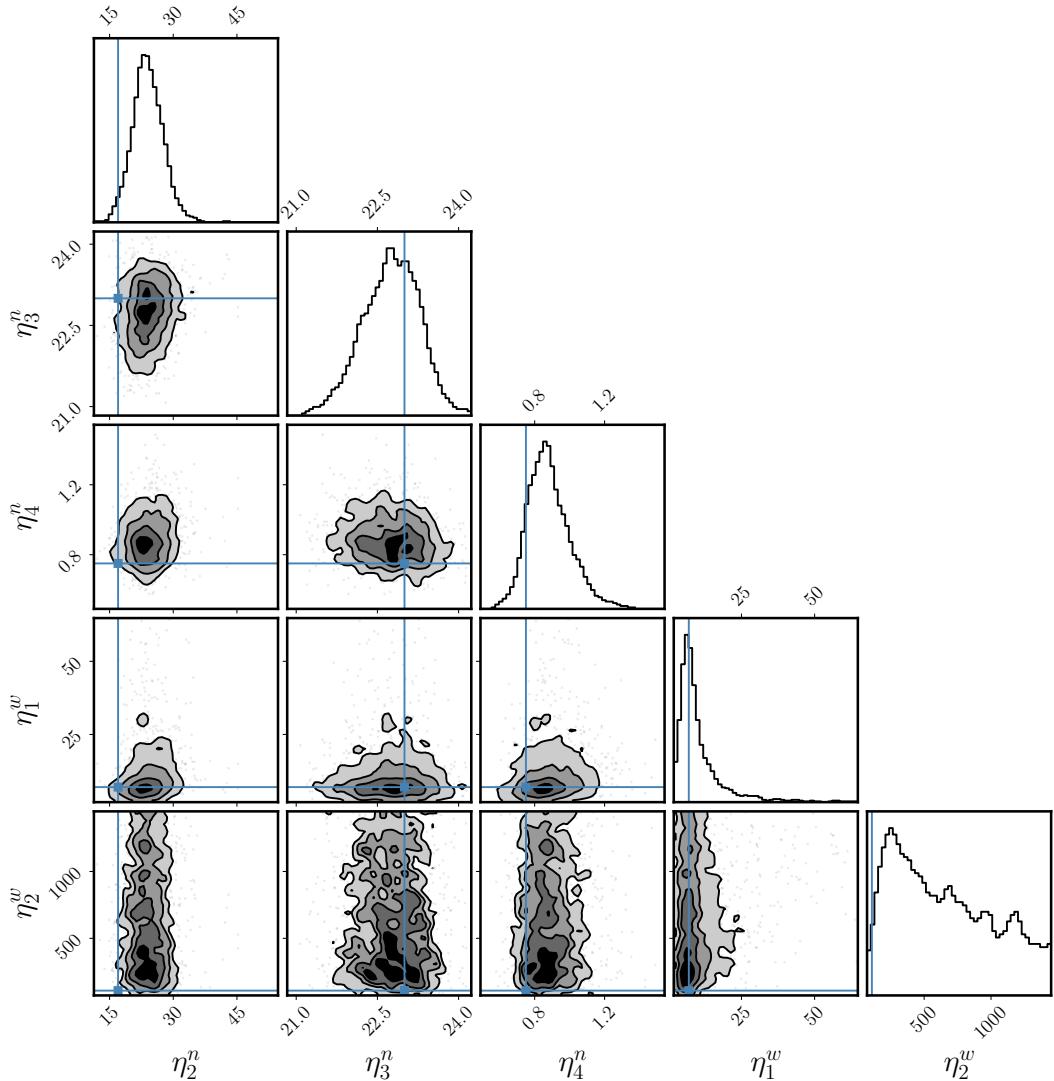


Figure A.3: Posterior distributions of the parameter from the 3rd GPRN model.

Appendix B

Datasets summary

Star	Dataset	N	Time span (days)	PTP amplitude (m/s)	Average error (m/s)	RMS (m/s)
Sun	RV	497	1094	10.364	0.133	1.982
Sun	BIS	497	1094	11.900	0.265	1.954
Sun	FWHM	497	1094	15.160	0.265	2.244
Sun	$\log R'_{hk}$	497	1094	0.064	0.001	0.0127
HD 10700	RV	174	470	10.328	0.370	1.884
HD 10700	FWHM	174	470	98.257	0.739	11.758
HD 34411	RV	188	416	10.744	0.337	1.767
HD 34411	FWHM	188	416	49.294	0.674	5.244
HD 101501	RV	45	655	17.881	0.353	4.920
HD 101501	FWHM	45	655	97.777	0.707	24.480
HD 26965	RV	114	465	14.144	0.334	3.234
HD 26965	FWHM	114	465	69.244	0.667	15.792

Table B.1: Summary with the number of points, time spans, peak-to-peak amplitude, average error, and initial RMS of the datasets analysed on this thesis.

Appendix C

Scatterplot matrices

C.1 Parameters

Following an MCMC analysis is common to plot the obtained posterior distribution for each parameter. To detect existing correlations with the remaining parameters, one generally uses a scatterplot matrix. In the following pages, I present the scatterplot matrices obtained on the analysis of the Sun dataset. For I made use of Python module know as CORNER (Foreman-Mackey, 2016).

Considering the number of parameters some of the plots carry, I tried to simplify them as much as possible to facilitate the visualization. In table C.1 a summary of the parameters and their respective units is shown. I represent the GP parameters by the greek letter η , following the notation I used in chapter 3. The interpretation of each parameter is the same as the one given in that chapter. On the GPRN, the parameters related to the GP of the node will carry the exponent n . The GP of the weights will carry the exponent w followed by a subscript abbreviation of each related time series. For example, a node will carry the notation η_1^n and a weight $\eta_1^{w_{RV}}$.

Variables	Parameter	Unit
η_1	Amplitude	Metre per second
η_2	Decaying time-span	Days
η_3	Period	Days
η_4	Length scale	-
s	White noise amplitude	Metre per second
slope	Slope of the mean function	-
offset	Offset of the mean function	Metre per second

Table C.1: Parameters and its respective units.

C.2 Plots

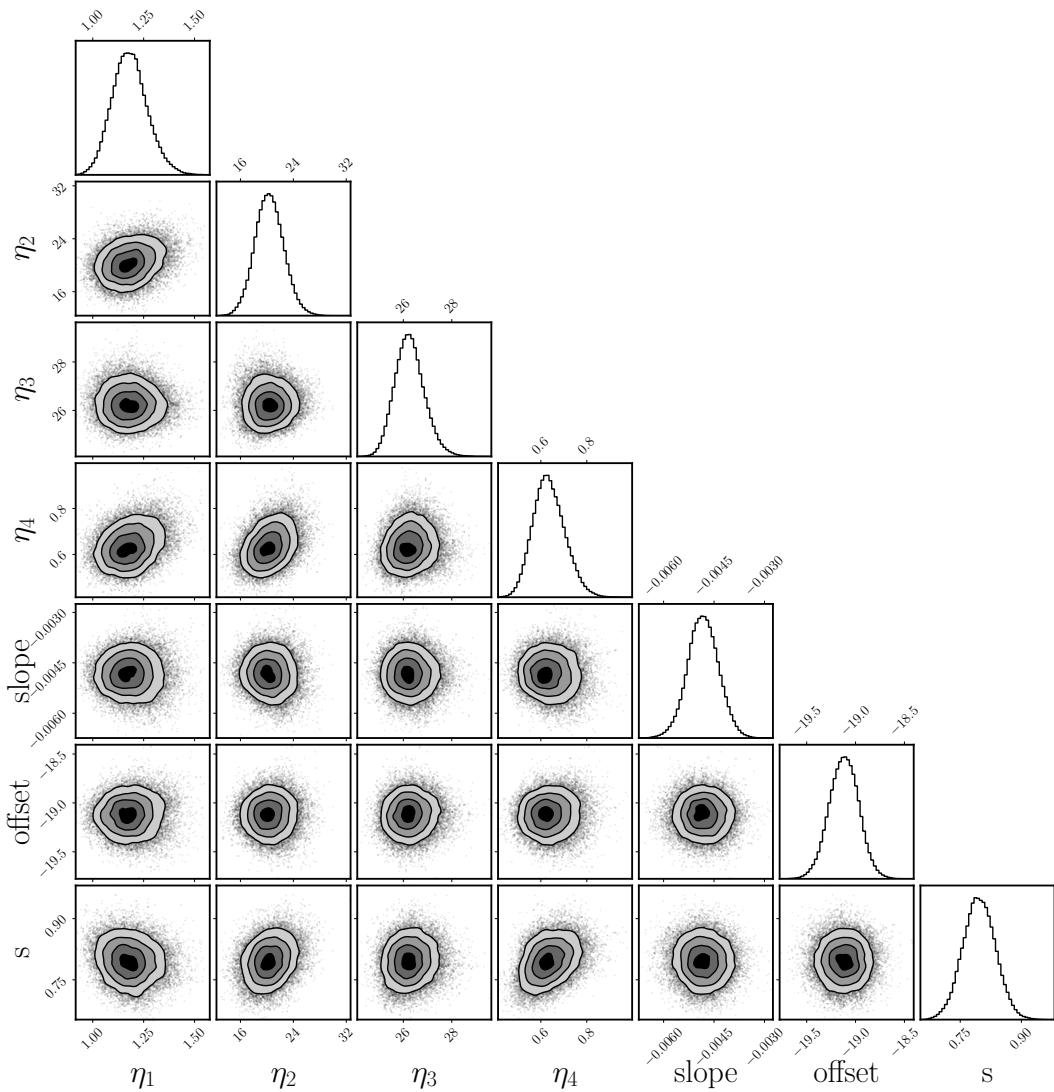


Figure C.1: Posterior distributions of the parameter of the GP used in the individual analysis of the Sun's RV measurements.

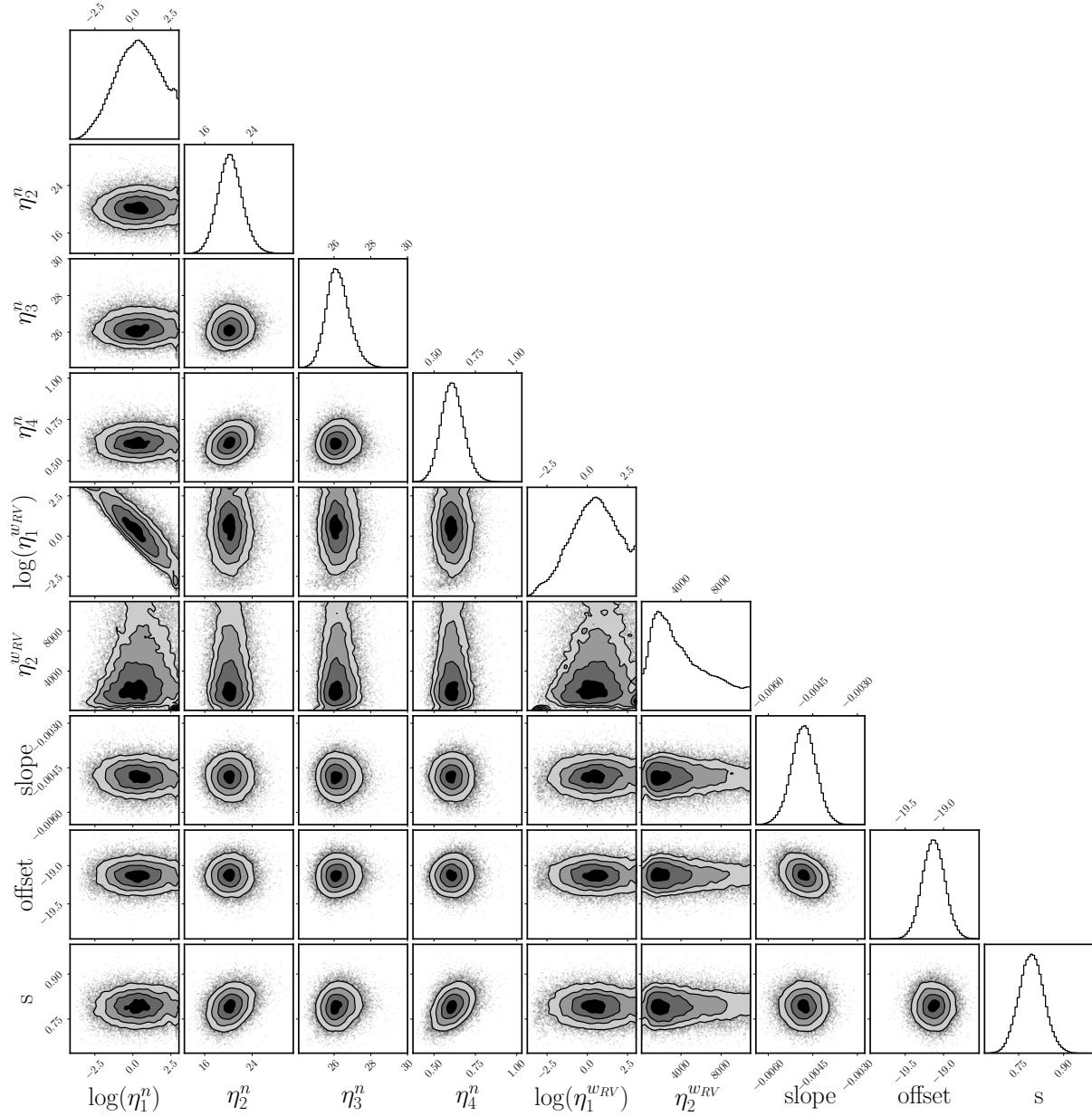


Figure C.2: Posterior distributions of the parameter of the GPRN used in the individual analysis of the Sun's RV measurements.

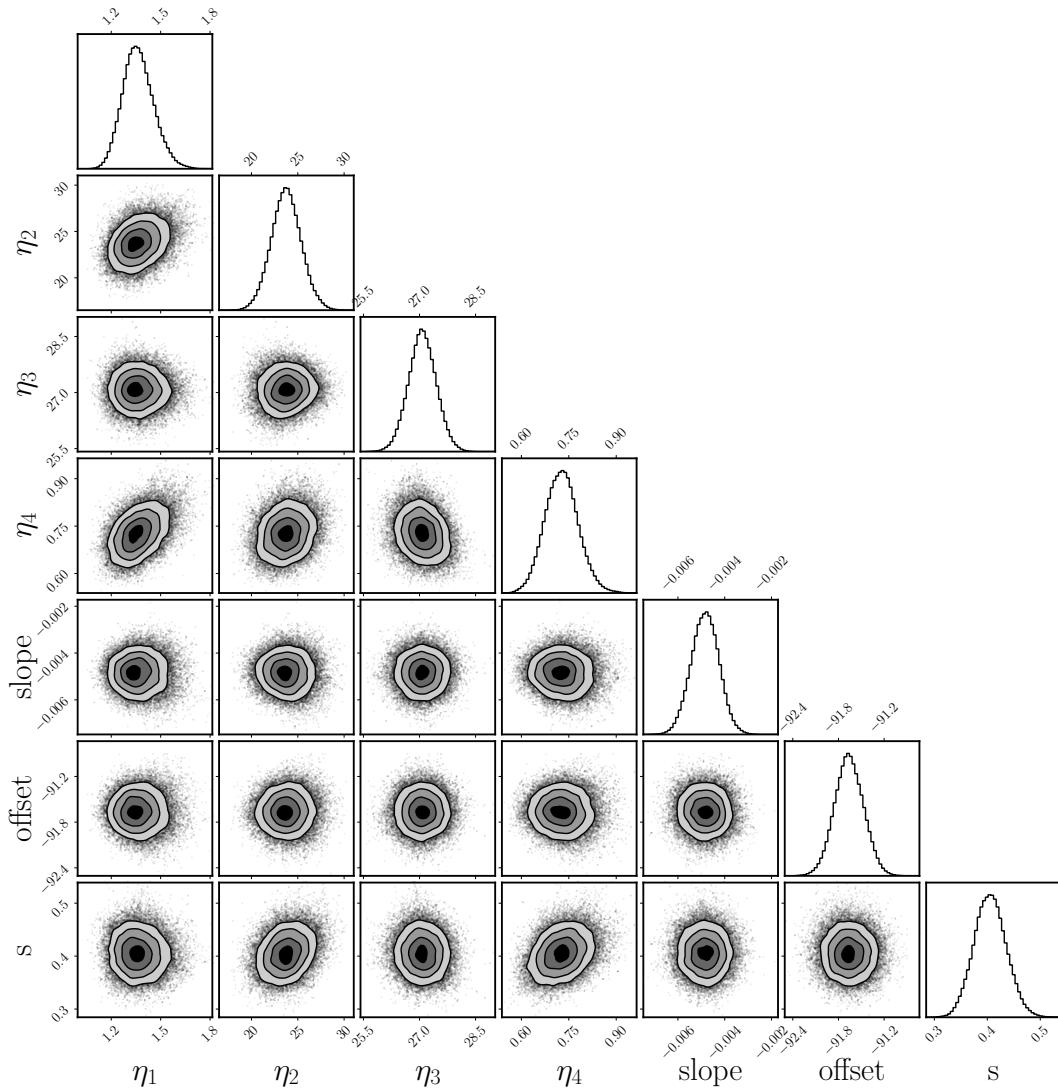


Figure C.3: Posterior distributions of the parameter of the GP used in the individual analysis of the Sun's BIS measurements.

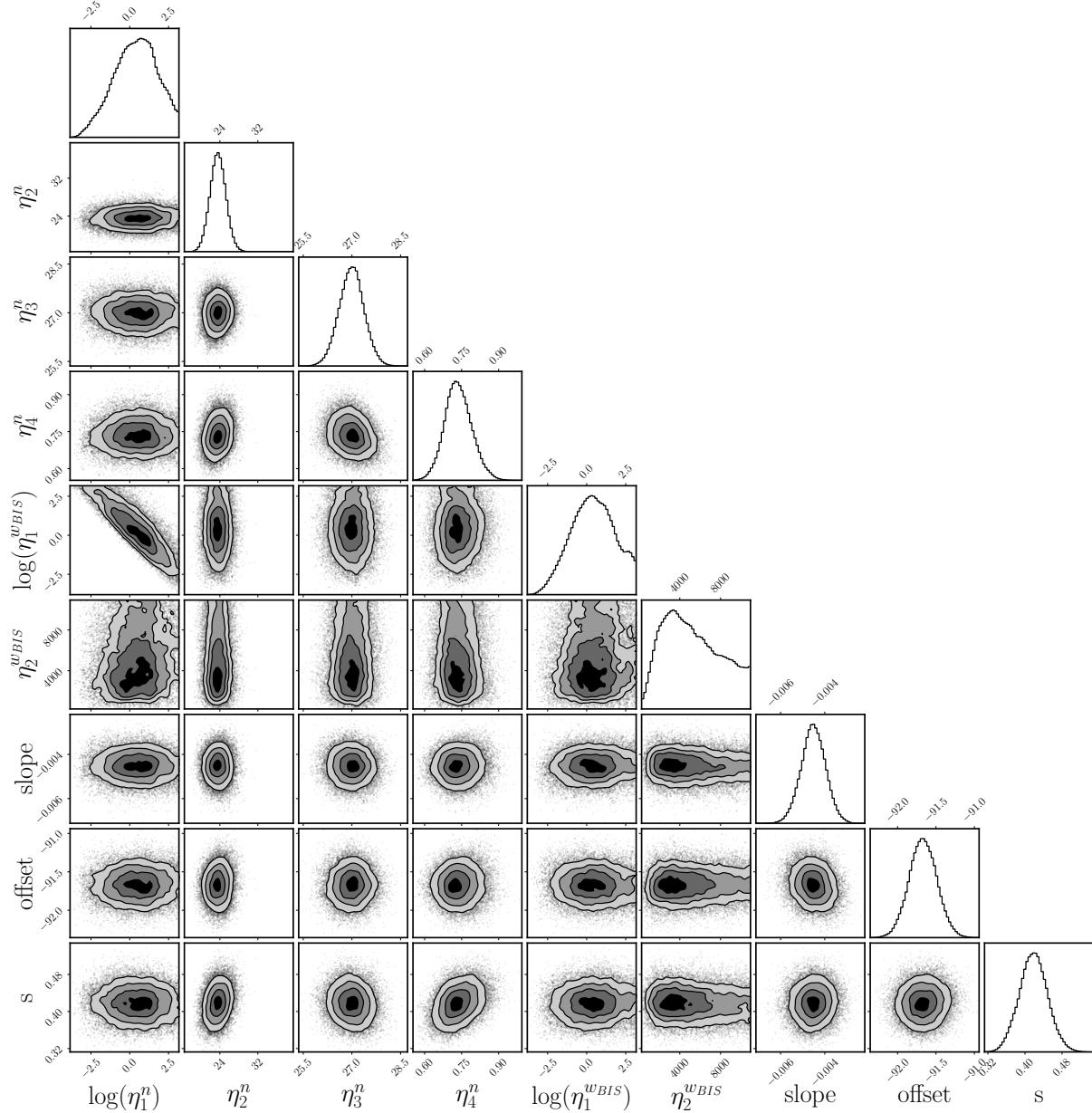


Figure C.4: Posterior distributions of the parameter of the GPRN used in the individual analysis of the Sun's BIS measurements.

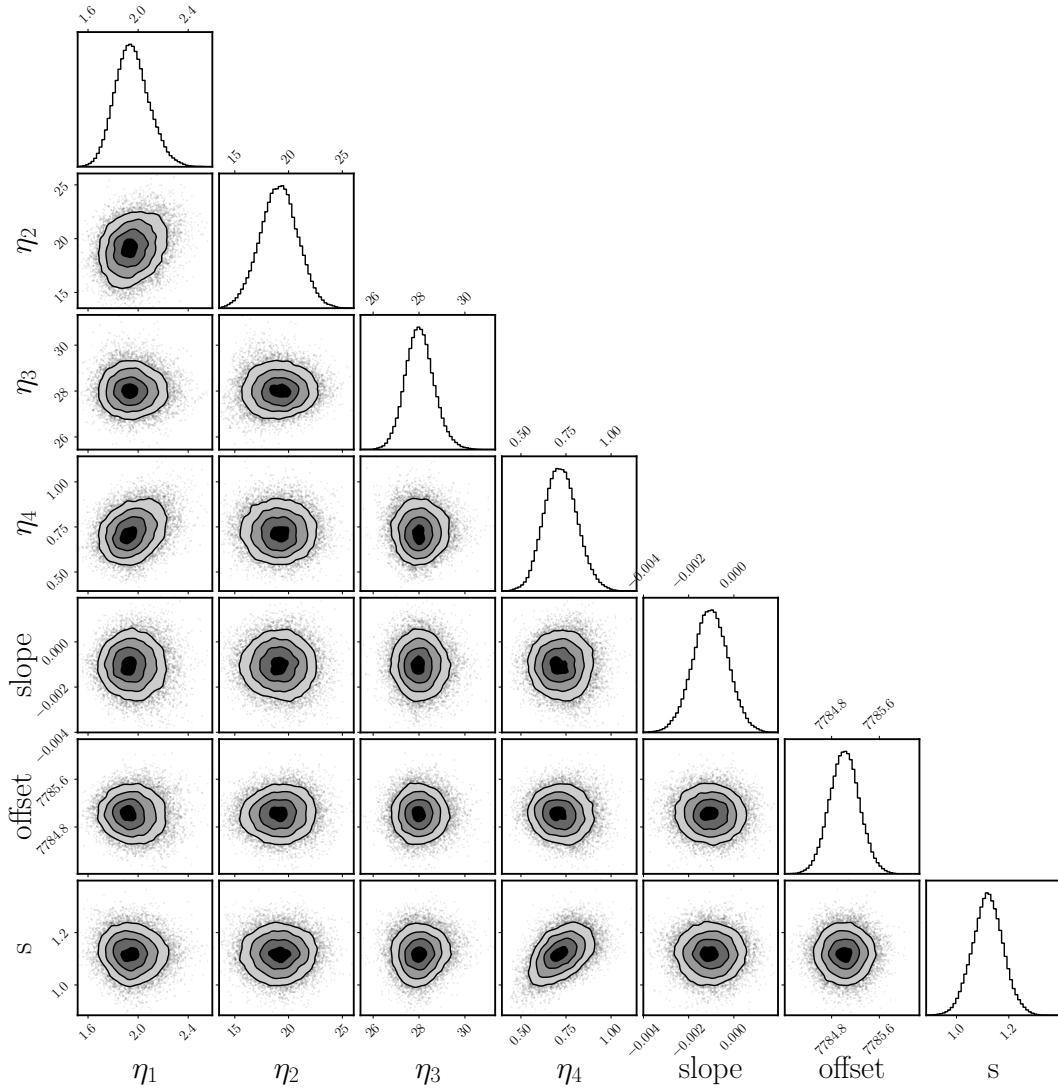


Figure C.5: Posterior distributions of the parameter of the GP used in the individual analysis of the Sun's FWHM measurements.

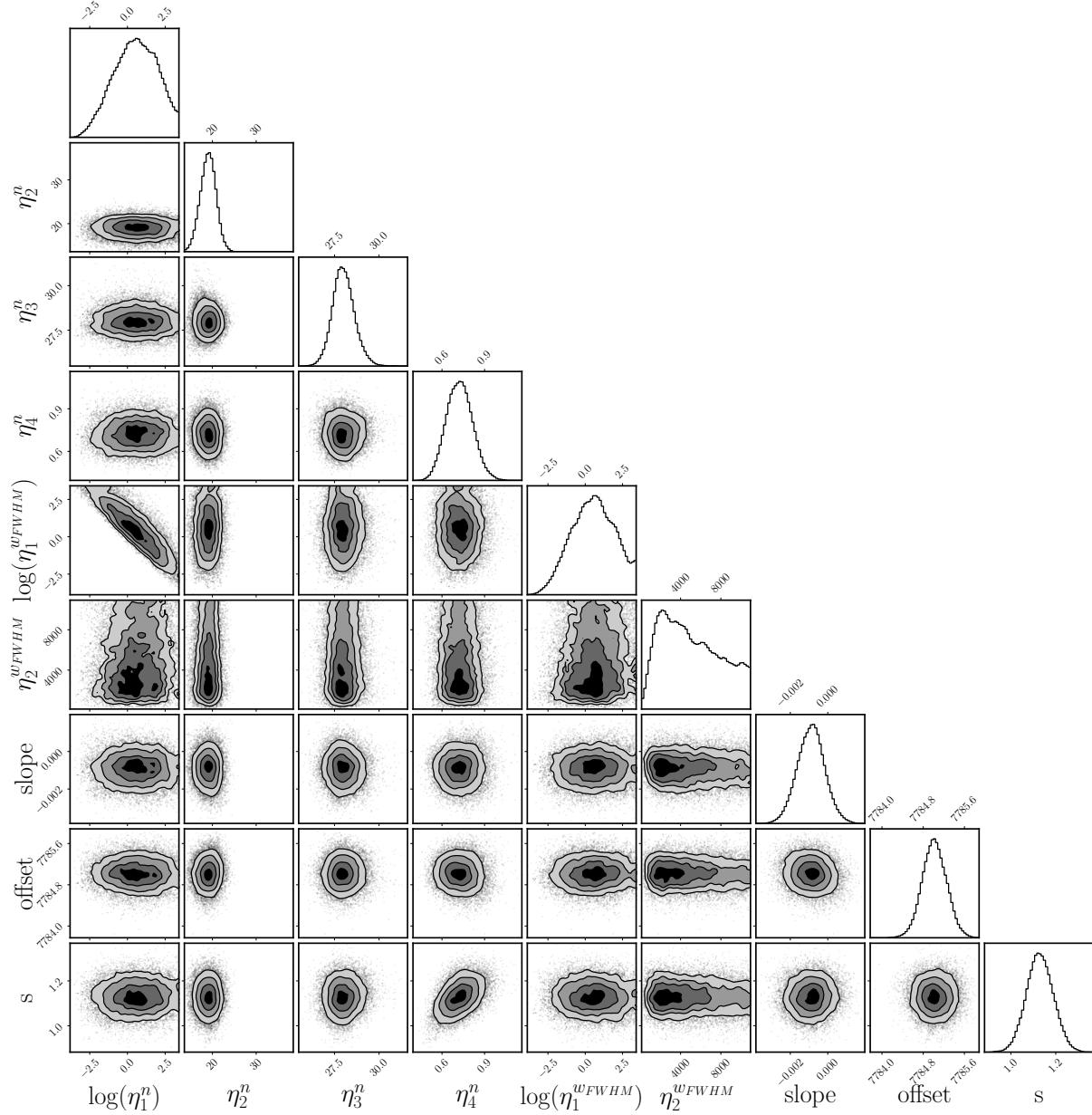


Figure C.6: Posterior distributions of the parameter of the GPRN used in the individual analysis of the Sun's FWHM measurements.

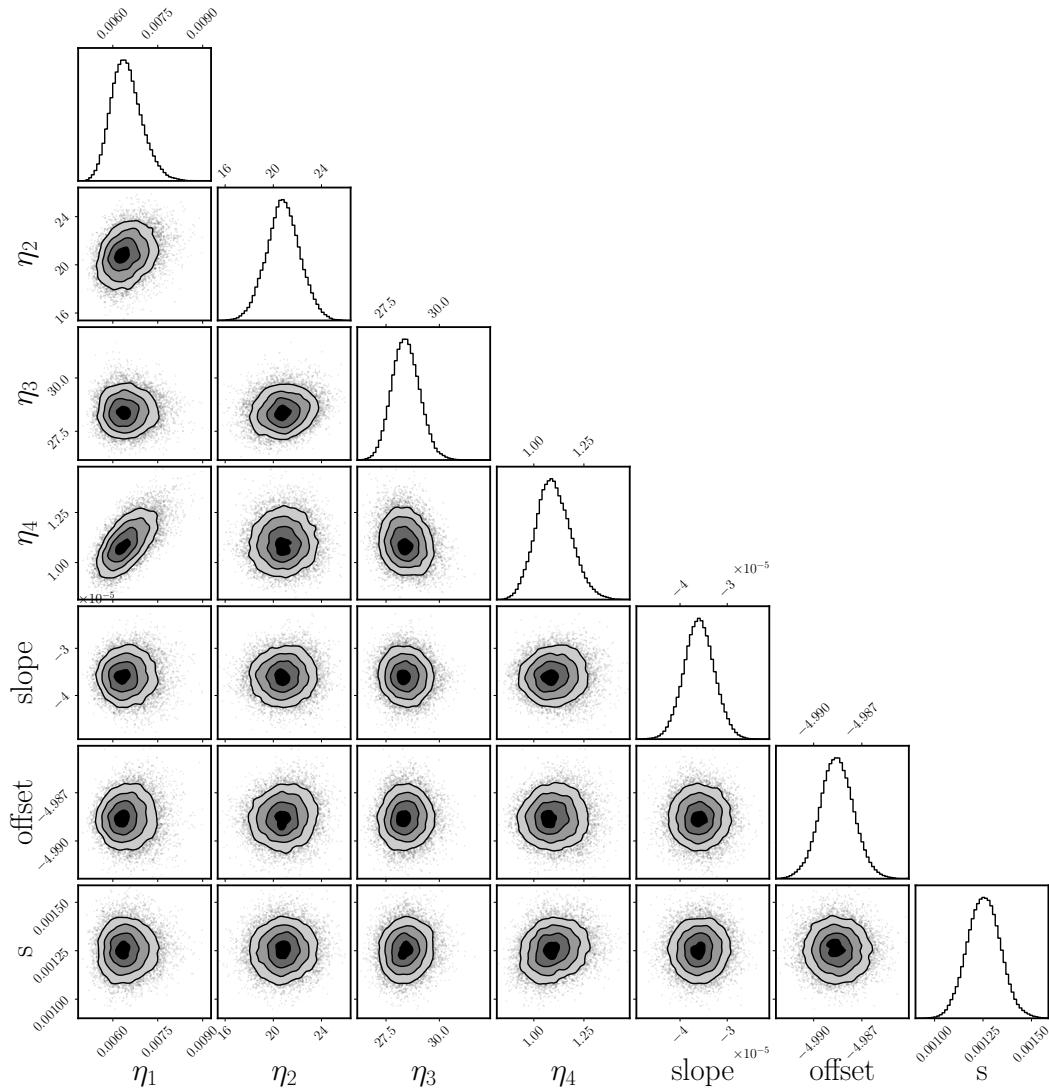


Figure C.7: Posterior distributions of the parameter of the GP used in the individual analysis of the Sun's $\log R'_{hk}$ measurements.

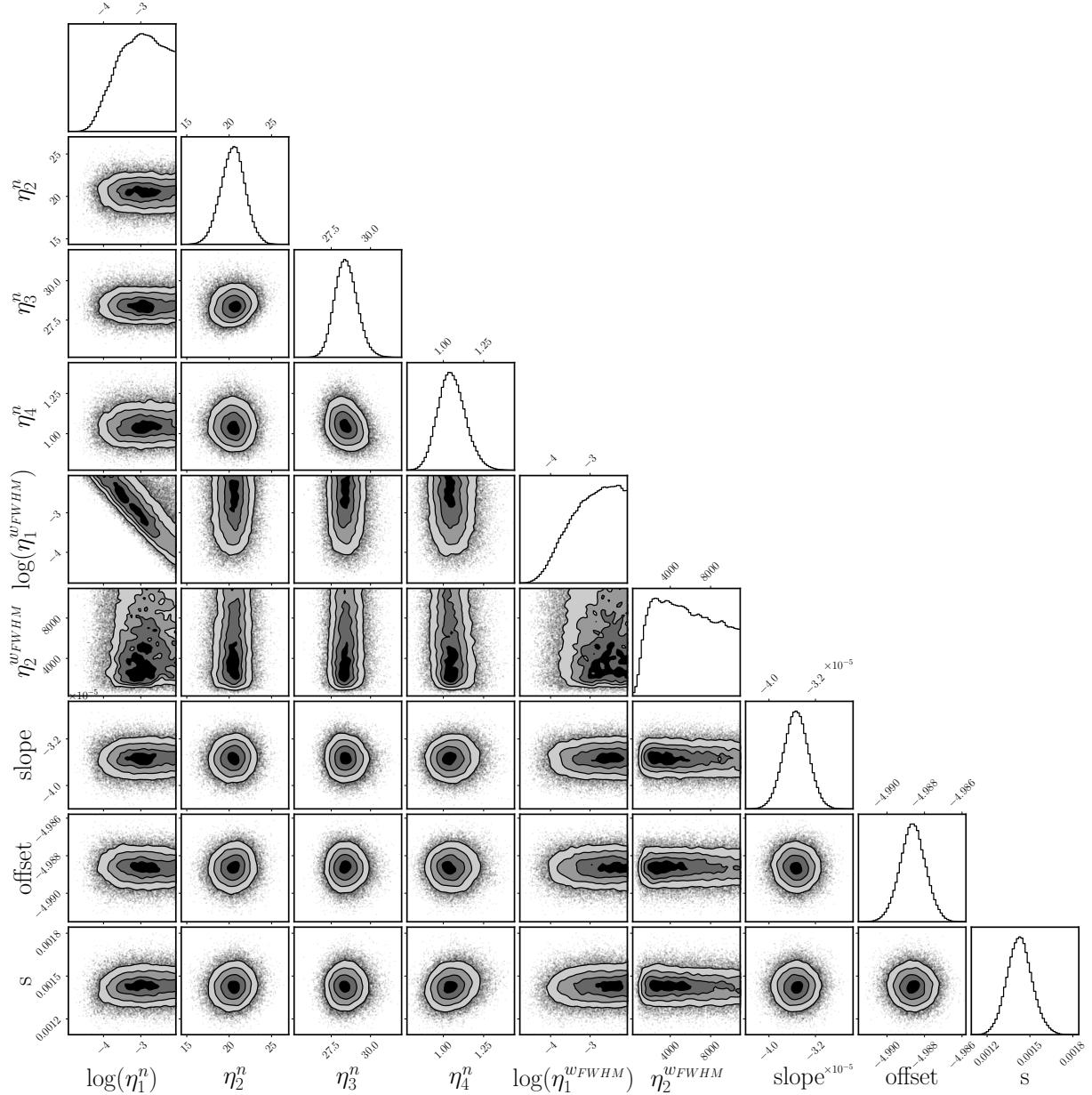


Figure C.8: Posterior distributions of the parameter of the GPRN used in the individual analysis of the Sun's $\log R'_{hk}$ measurements.

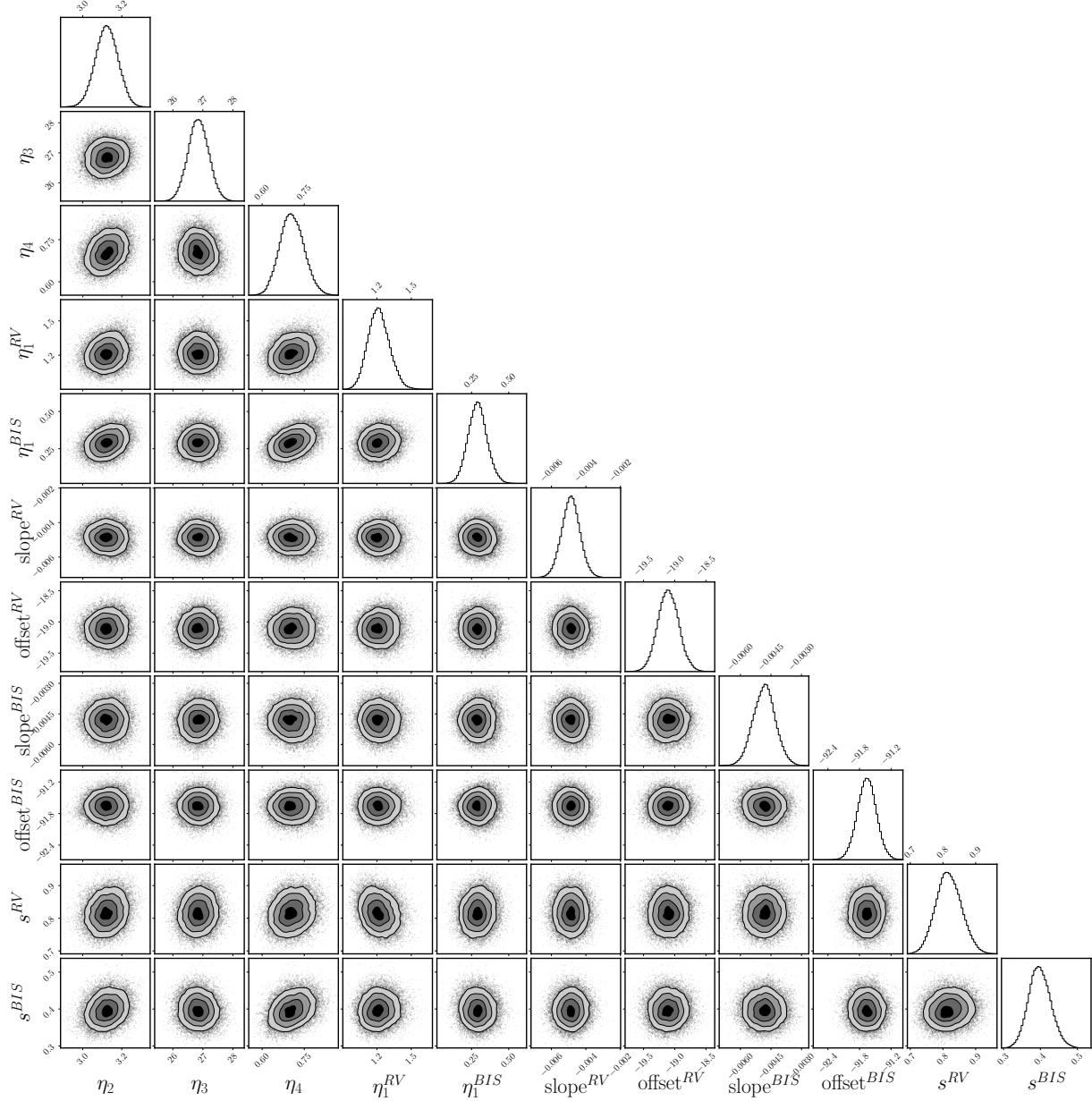


Figure C.9: Posterior distributions of the parameter of the GP used in the combined analysis of the Sun's RV and BIS measurements.

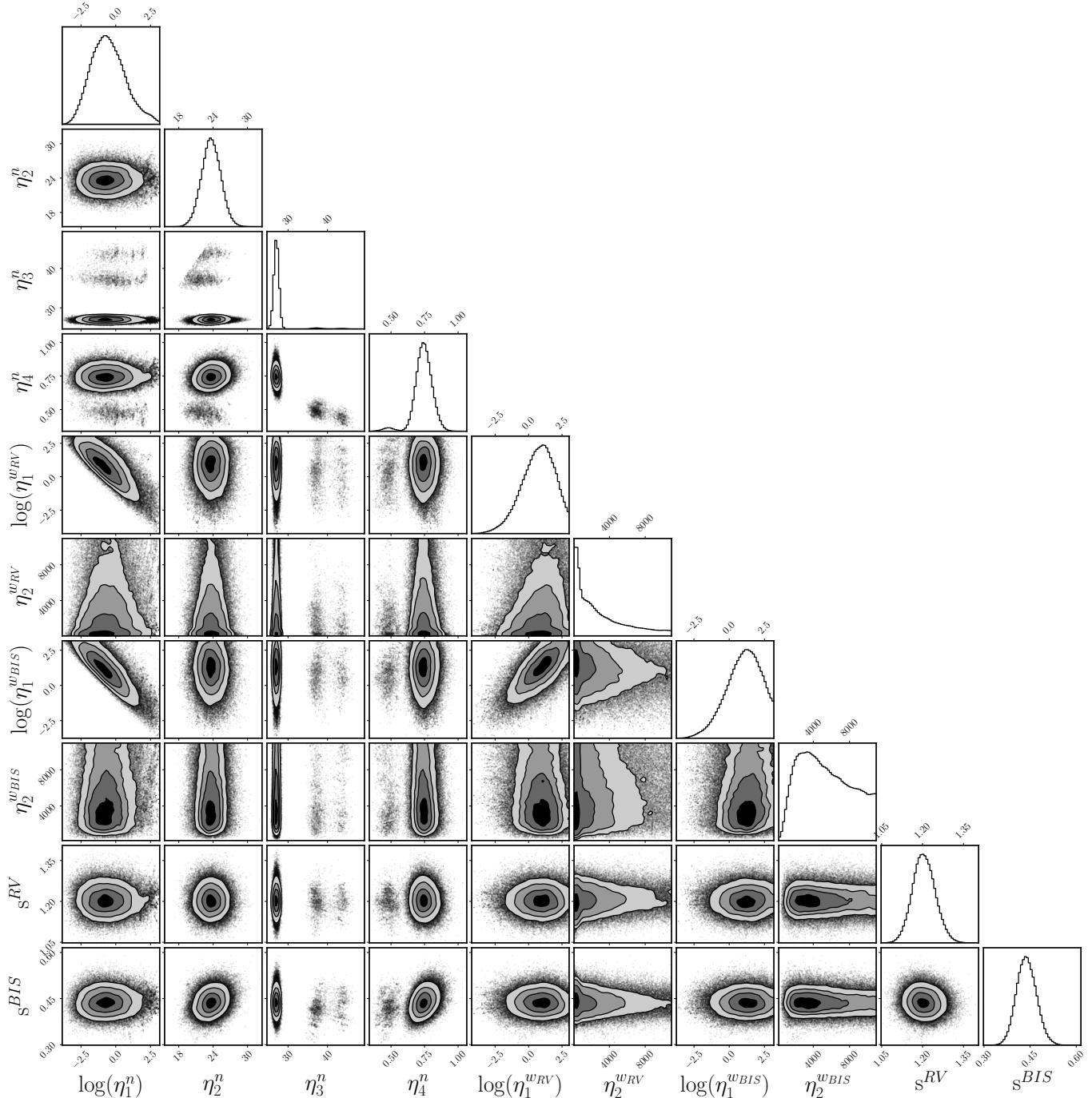


Figure C.10: Posterior distributions of the parameter of the GPRN used in the combined analysis of the Sun's RV and BIS measurements.

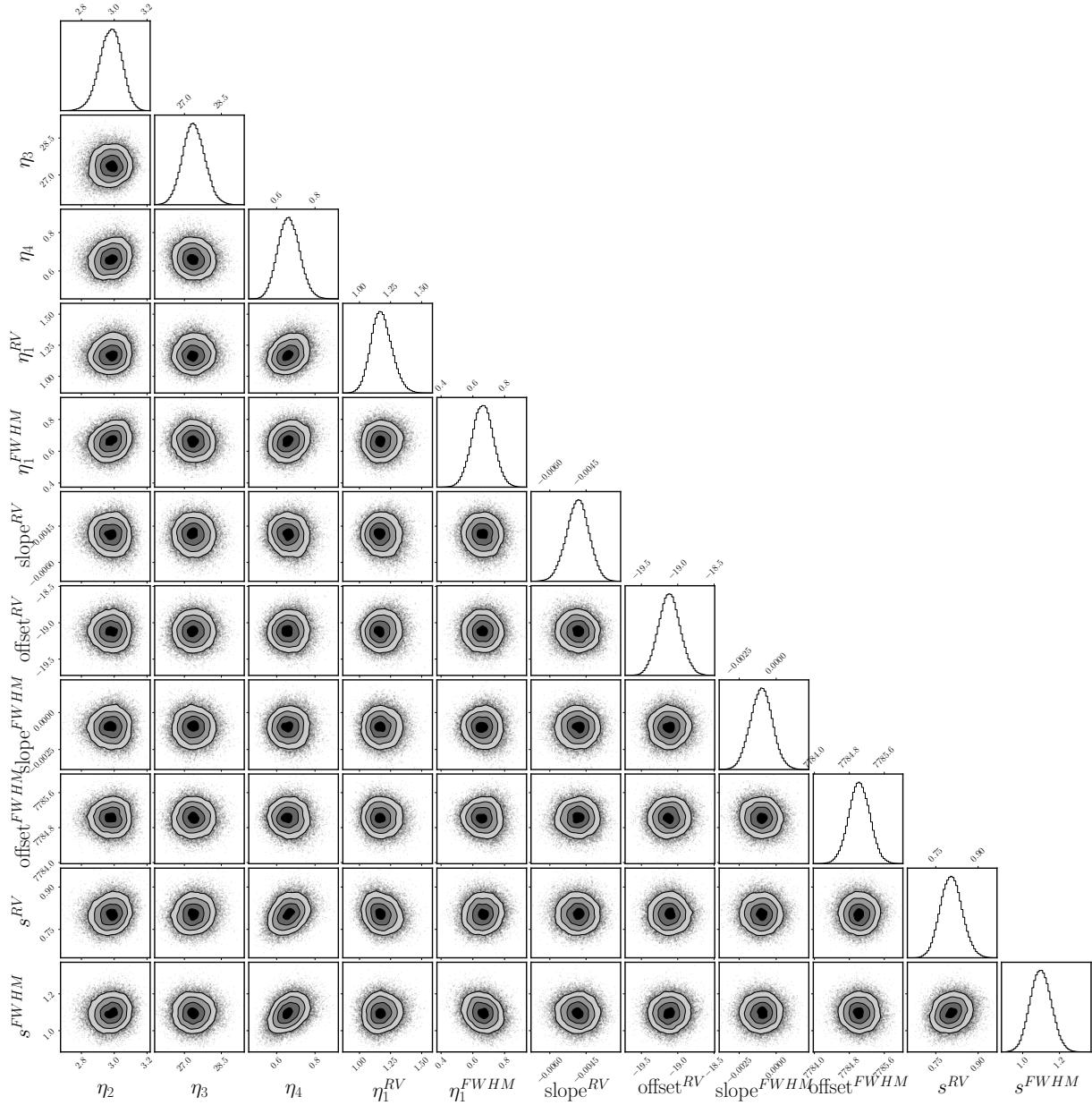


Figure C.11: Posterior distributions of the parameter of the GP used in the combined analysis of the Sun's RV and FWHM measurements.

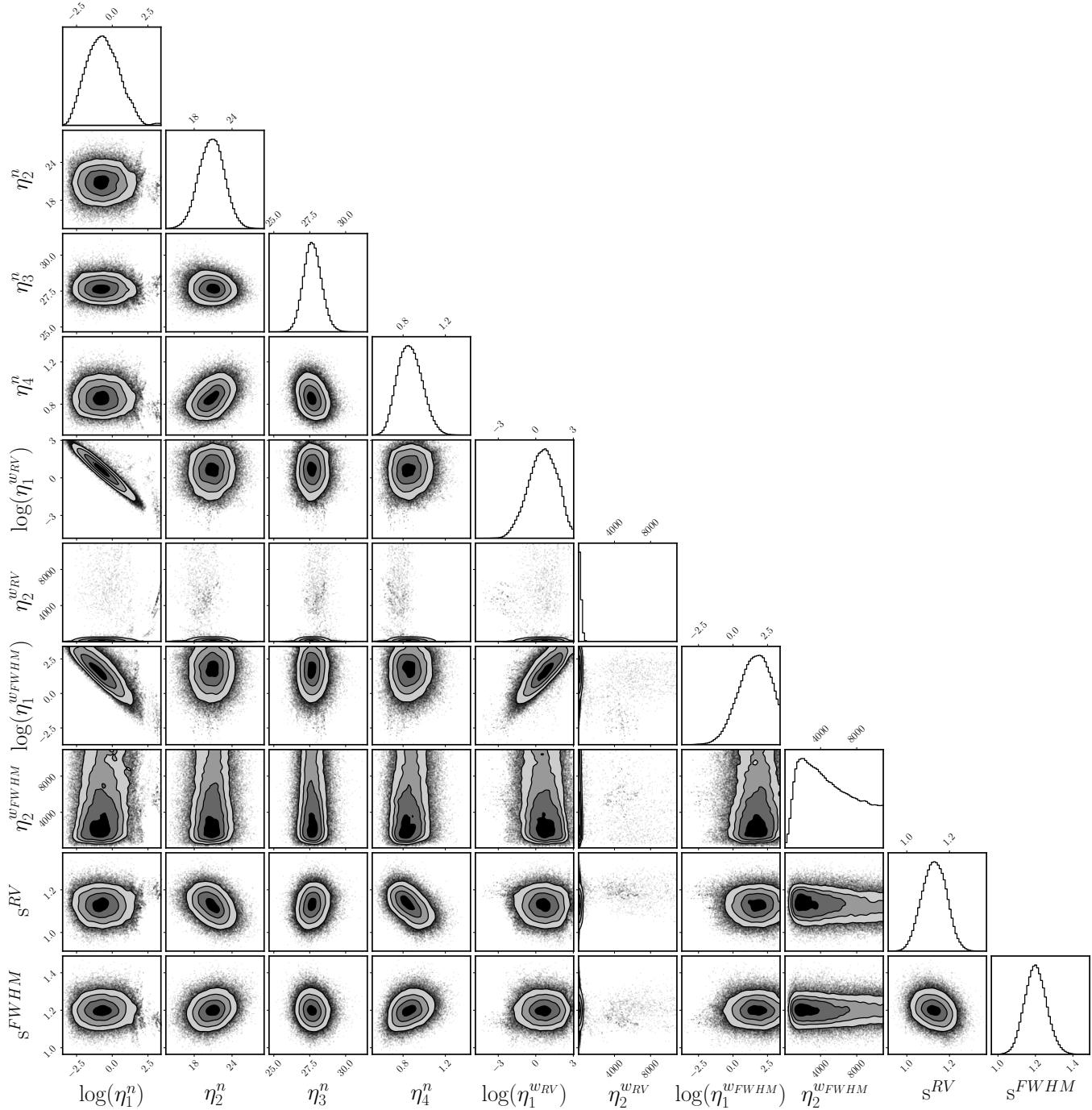


Figure C.12: Posterior distributions of the parameter of the GPRN used in the combined analysis of the Sun's RV and FWHM measurements.

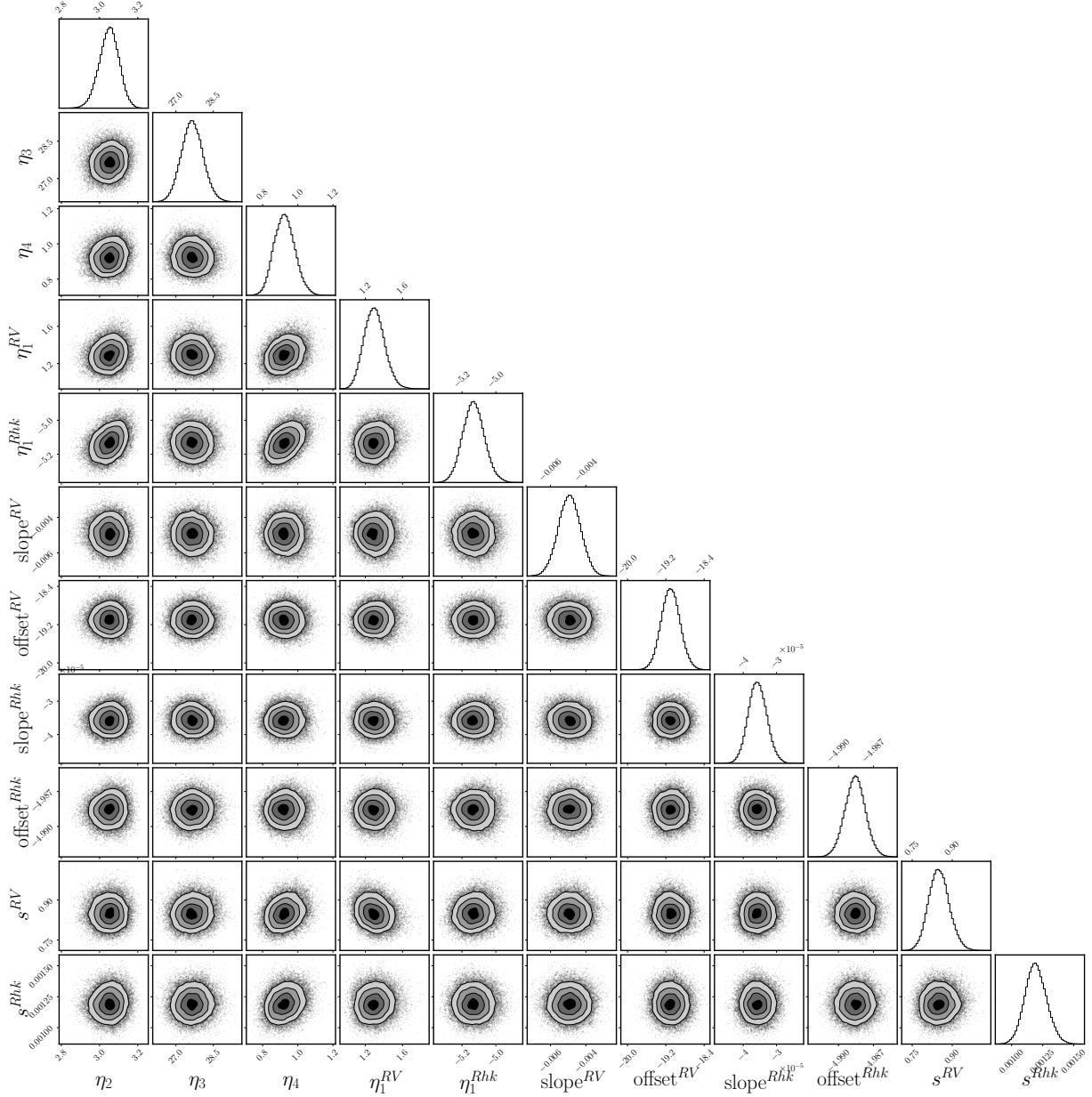


Figure C.13: Posterior distributions of the parameter of the GP used in the combined analysis of the Sun's RV and $\log R'_{hk}$ measurements.

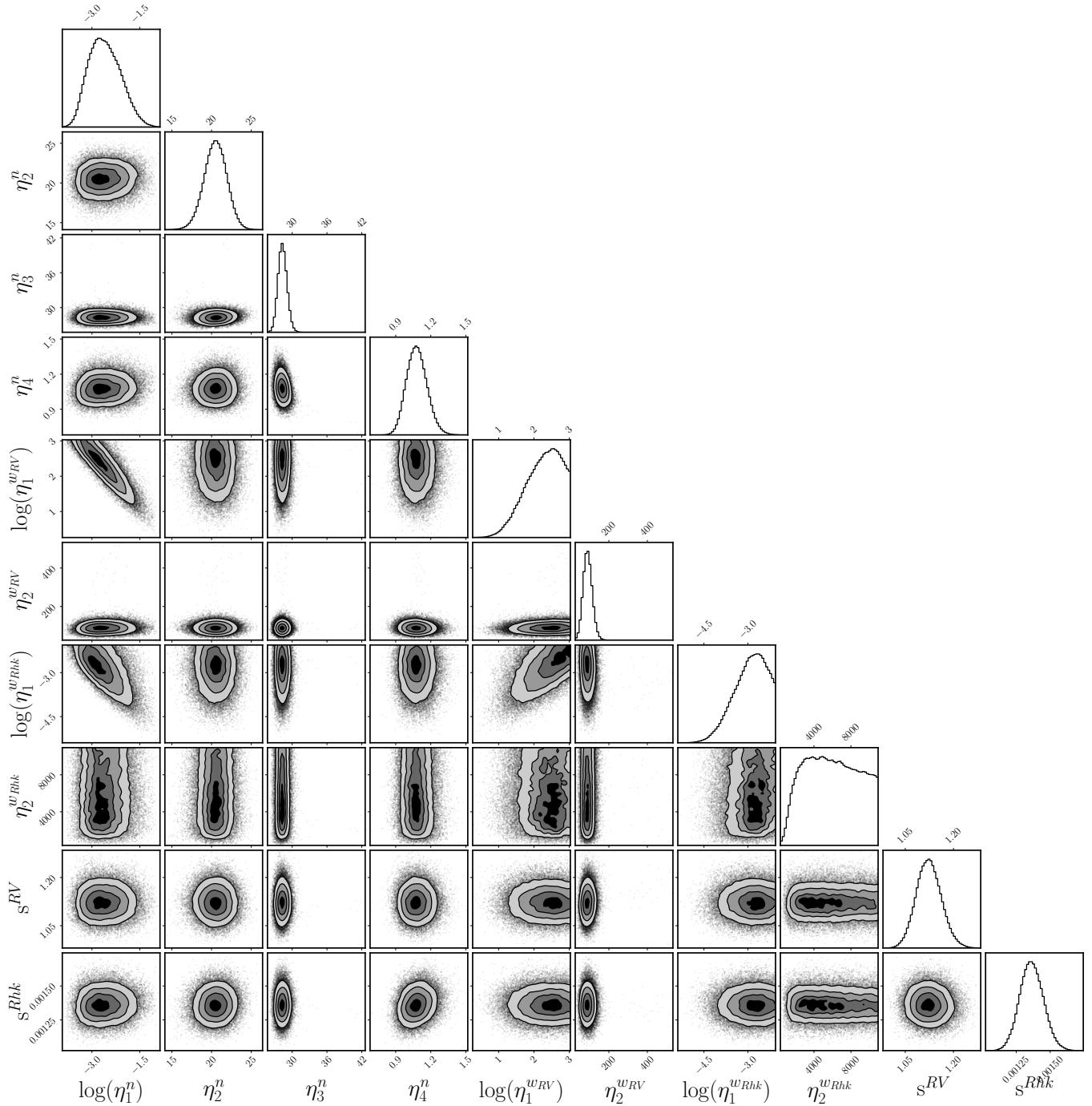


Figure C.14: Posterior distributions of the parameter of the GPRN used in the combined analysis of the Sun's RV and $\log R'_{hk}$ measurements.

Appendix D

Other EXPRES results

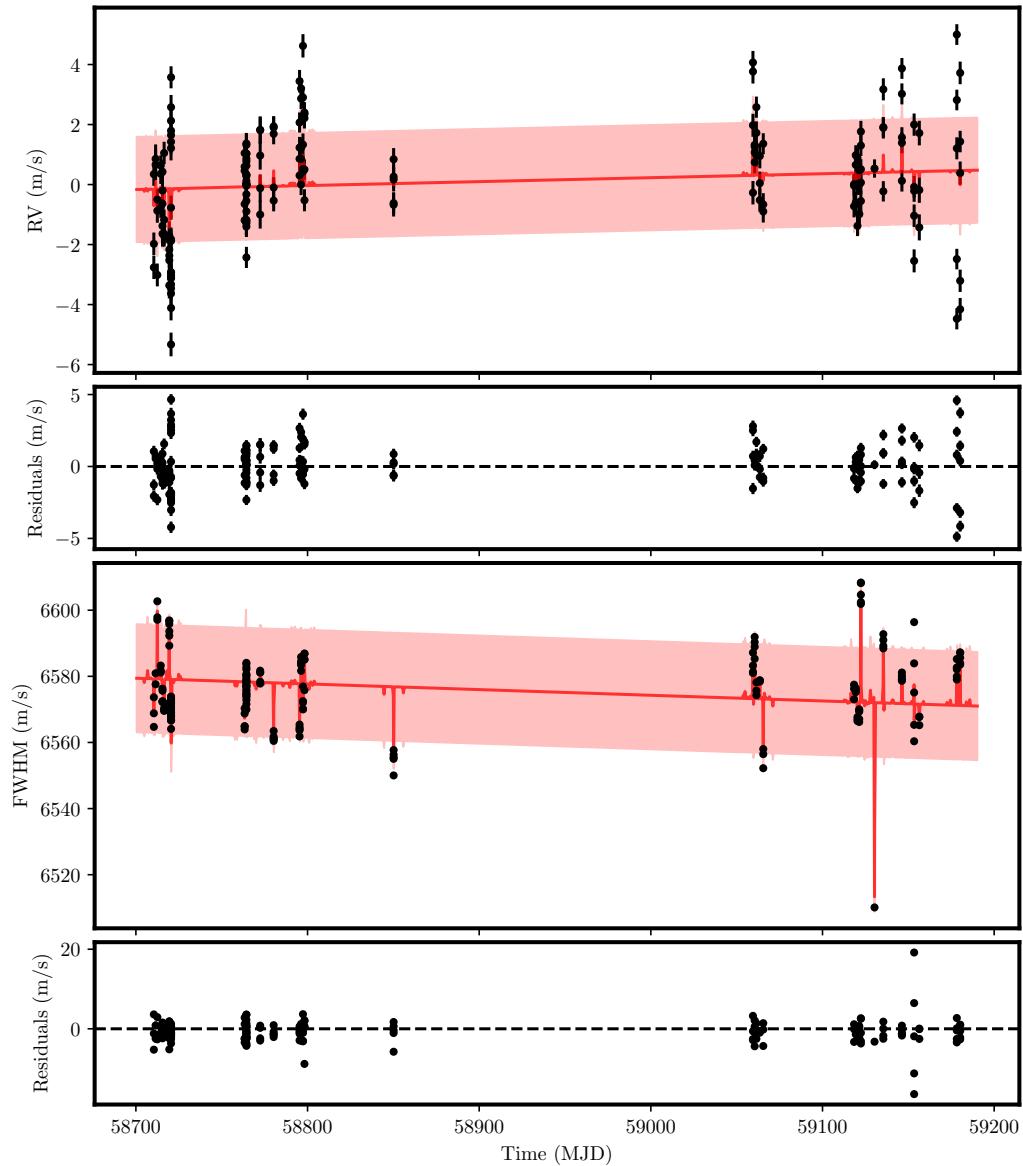


Figure D.1: Fit and respective residuals of the MAP values from the GP on the RV and the FWHM time series of HD 10700.

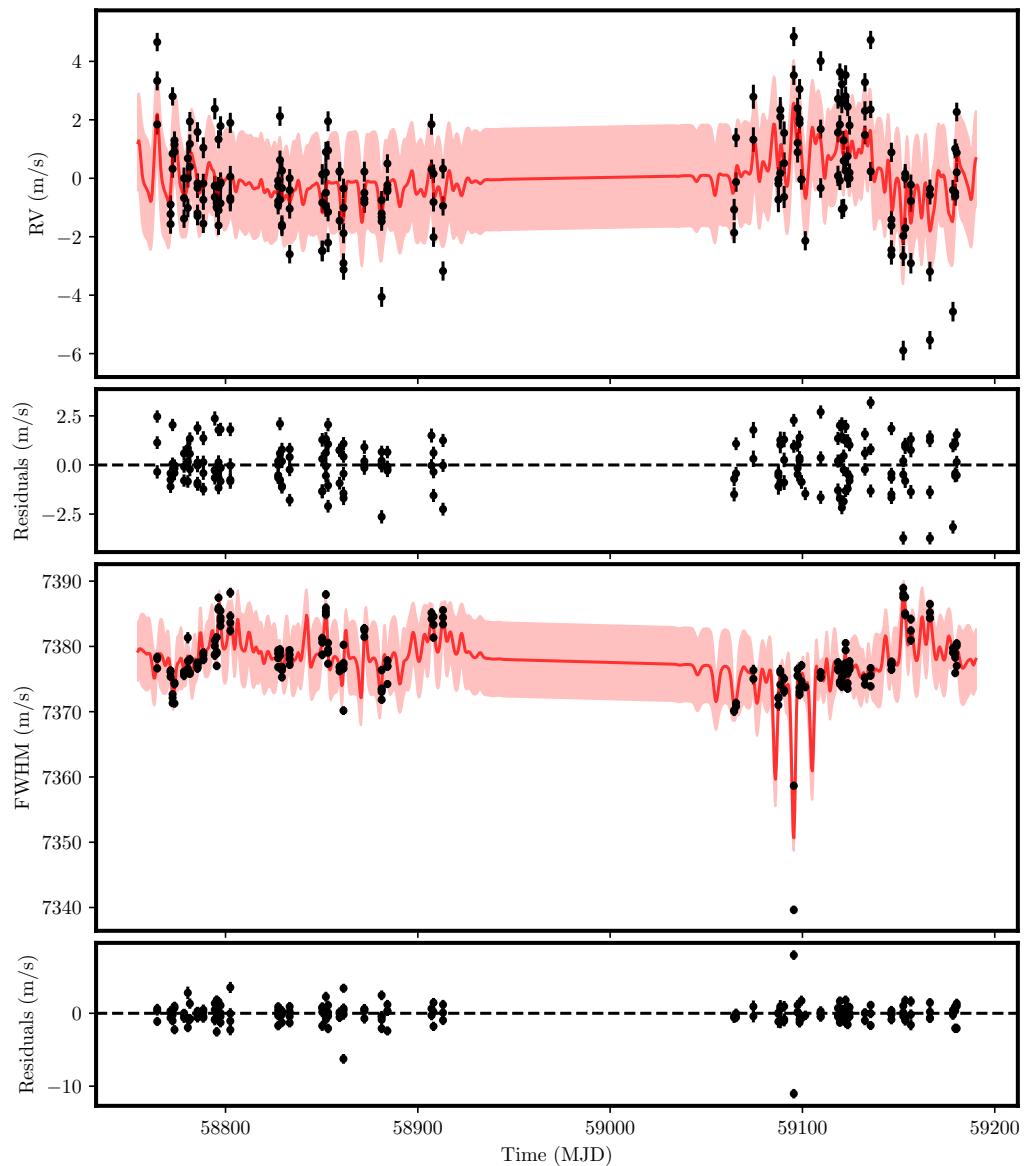


Figure D.2: Fit and respective residuals of the MAP values from the GP on the RV and the FWHM time series of HD 34411.

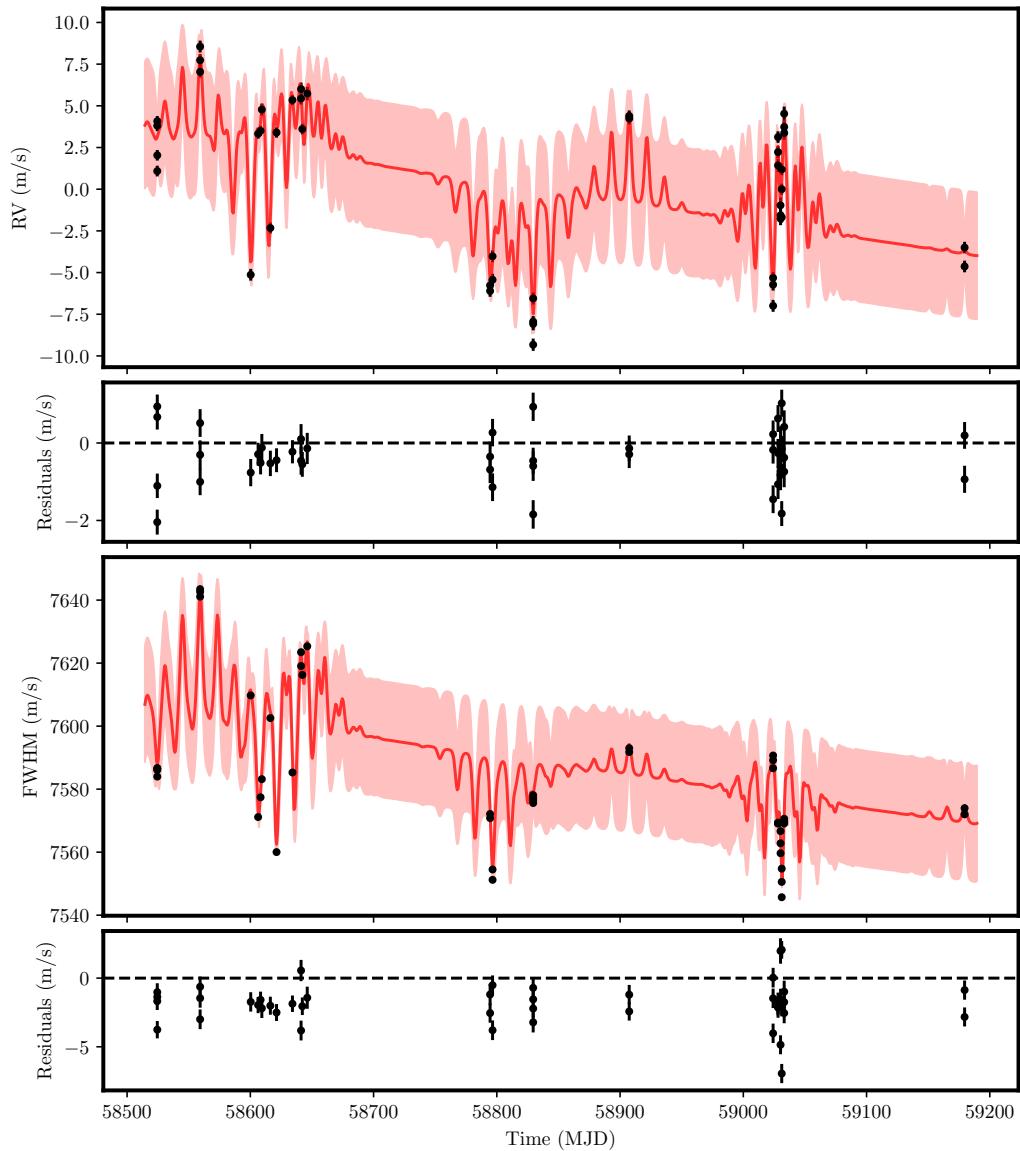


Figure D.3: Fit and respective residuals of the MAP values from the GP on the RV and the FWHM time series of HD 101501.

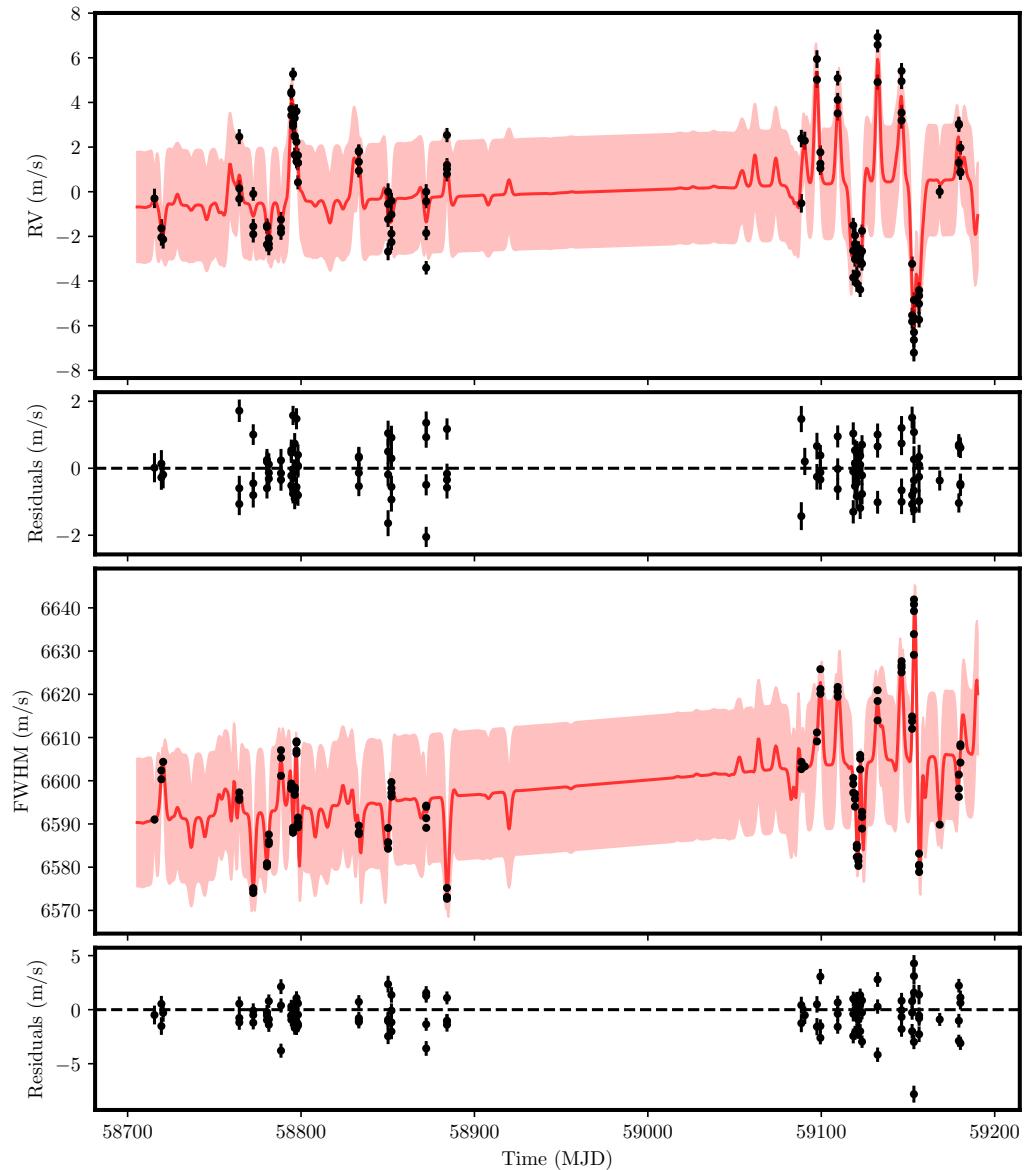


Figure D.4: Fit and respective residuals of the MAP values from the GP on the RV and the FWHM time series of HD 26965.

Bibliography

- Abdessalem, A. B. et al. (2017). “Automatic Kernel Selection for Gaussian Processes Regression with Approximate Bayesian Computation and Sequential Monte Carlo”. In: *Frontiers in Built Environment* 3. doi: [10.3389/fbuil.2017.00052](https://doi.org/10.3389/fbuil.2017.00052).
- Abramowitz, M. and I. Stegun (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Applied mathematics series. U.S. Government Printing Office.
- Adibekyan, V. Z. et al. (2013). “Orbital and physical properties of planets and their hosts: new insights on planet formation and evolution”. In: *A&A* 560, A51. doi: [10.1051/0004-6361/201322551](https://doi.org/10.1051/0004-6361/201322551).
- Adibekyan, V. (2019). “Heavy Metal Rules. I. Exoplanet Incidence and Metallicity”. In: *Geosciences* 9(3). doi: [10.3390/geosciences9030105](https://doi.org/10.3390/geosciences9030105).
- Aigrain, S. et al. (2012). “A simple method to estimate radial velocity variations due to stellar activity using photometry*”. In: *Monthly Notices of the Royal Astronomical Society* 419(4). doi: [10.1111/j.1365-2966.2011.19960.x](https://doi.org/10.1111/j.1365-2966.2011.19960.x).
- Allende Prieto, C. (2007). “Velocities from Cross-Correlation: A Guide for Self-Improvement”. In: *??jnlAJ* 134(5). doi: [10.1086/522051](https://doi.org/10.1086/522051).
- Arentoft, T. et al. (2008). “A Multisite Campaign to Measure Solar-like Oscillations in Procyon. I. Observations, Data Reduction, and Slow Variations”. In: *ApJ* 687(2). doi: [10.1086/592040](https://doi.org/10.1086/592040).
- Arfken, G. et al. (2013). *Mathematical Methods for Physicists: A Comprehensive Guide*. Elsevier Science. ISBN: 9780123846549.
- Artin, E. and M. Butler (2015). *The Gamma Function*. Dover Books on Mathematics. Dover Publications. ISBN: 9780486803005.

- Astropy Collaboration et al. (2018). “The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package”. In: *??jnlAJ* 156(3), 123. doi: [10.3847/1538-3881/aabc4f](https://doi.org/10.3847/1538-3881/aabc4f).
- Bahng, J. and M. Schwarzschild (1961). “Lifetime of Solar Granules.” In: *ApJ* 134. doi: [10.1086/147160](https://doi.org/10.1086/147160).
- Baliunas, S. L. et al. (1998). “Activity Cycles in Lower Main Sequence and POST Main Sequence Stars: The HK Project”. In: *Cool Stars, Stellar Systems, and the Sun*. Ed. by R. A. Donahue and J. A. Bookbinder. Vol. 154. Astronomical Society of the Pacific Conference Series.
- Baranne, A. et al. (1996). “ELODIE: A spectrograph for accurate radial velocity measurements”. In: *Astron. Astrophys. Suppl. Ser.* 119(2). doi: [10.1051/aas:1996251](https://doi.org/10.1051/aas:1996251).
- Barragán, O. et al. (2021). “pyaneti II: A multi-dimensional Gaussian process approach to analysing spectroscopic time-series”. In: *MNRAS*. doi: [10.1093/mnras/stab2889](https://doi.org/10.1093/mnras/stab2889).
- Barros, S. C. C. et al. (2020). “Improving transit characterisation with Gaussian process modelling of stellar variability”. In: *A&A* 634, A75. doi: [10.1051/0004-6361/201936086](https://doi.org/10.1051/0004-6361/201936086).
- Bedding, T. R. and H. Kjeldsen (2003). “Solar-like Oscillations”. In: *Publications of the Astronomical Society of Australia* 20(2). doi: [10.1071/AS03025](https://doi.org/10.1071/AS03025).
- Bedding, T. R. and H. Kjeldsen (2007). “Observations of solar-like oscillations”. In: *Communications in Asteroseismology* 150. doi: [10.1553/cia150s106](https://doi.org/10.1553/cia150s106).
- Berger, T. E. et al. (2007). “What are ‘Faculae’?” In: *New Solar Physics with Solar-B Mission*. Ed. by K. Shibata et al. Vol. 369. Astronomical Society of the Pacific Conference Series.
- Beuzit, J.-L. et al. (2008). “SPHERE: a ‘Planet Finder’ instrument for the VLT”. In: *Ground-based and Airborne Instrumentation for Astronomy II*. Ed. by I. S. McLean and M. M. Casali. Vol. 7014. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 701418. doi: [10.1117/12.790120](https://doi.org/10.1117/12.790120).
- Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer. ISBN: 9780387310732.
- Blei, D. M. et al. (2017). “Variational Inference: A Review for Statisticians”. In: *Journal of the American Statistical Association* 112(518). doi: [10.1080/01621459.2017.1285773](https://doi.org/10.1080/01621459.2017.1285773).

- Boisse, I. et al. (2010). “Disentangling between stellar activity and planetary signals”. In: *Proceedings of the International Astronomical Union* 6(S273). doi: [10.1017/S1743921311015389](https://doi.org/10.1017/S1743921311015389).
- Bond, I. A. et al. (2004). “OGLE 2003-BLG-235/MOA 2003-BLG-53: A Planetary Microlensing Event”. In: *ApJ* 606(2). doi: [10.1086/420928](https://doi.org/10.1086/420928).
- Borgniet, S. et al. (2019). “Consistent radial velocities of classical Cepheids from the cross-correlation technique”. In: *A&A* 631, A37. doi: [10.1051/0004-6361/201935622](https://doi.org/10.1051/0004-6361/201935622).
- Borucki, W. et al. (2009). “KEPLER: Search for Earth-Size Planets in the Habitable Zone”. In: *Transiting Planets*. Ed. by F. Pont et al. Vol. 253. doi: [10.1017/S1743921308026513](https://doi.org/10.1017/S1743921308026513).
- Bouchy, F. et al. (2001). “Fundamental photon noise limit to radial velocity measurements”. In: *A&A* 374. doi: [10.1051/0004-6361:20010730](https://doi.org/10.1051/0004-6361:20010730).
- Boyajian, T. S. et al. (2012). “Stellar Diameters and Temperatures. I. Main-sequence A, F, and G Stars”. In: *ApJ* 746(1), 101. doi: [10.1088/0004-637X/746/1/101](https://doi.org/10.1088/0004-637X/746/1/101).
- Brewer, B. J. et al. (2011). “Diffusive nested sampling”. In: *Statistics and Computing* 21(4). doi: [10.1007/s11222-010-9198-8](https://doi.org/10.1007/s11222-010-9198-8).
- Broomhall, A. M. et al. (2011). “Solar-cycle variations of large frequency separations of acoustic modes: implications for asteroseismology”. In: *MNRAS* 413(4). doi: [10.1111/j.1365-2966.2011.18375.x](https://doi.org/10.1111/j.1365-2966.2011.18375.x).
- Brown, R. A. and C. J. Burrows (1990). “On the feasibility of detecting extrasolar planets by reflected starlight using the Hubble Space Telescope”. In: 87(2). doi: [10.1016/0019-1035\(90\)90150-8](https://doi.org/10.1016/0019-1035(90)90150-8).
- Browning, M. K. et al. (2010). “Rotation and Magnetic Activity in a Sample of M-Dwarfs”. In: ??*jnlAJ* 139(2). doi: [10.1088/0004-6256/139/2/504](https://doi.org/10.1088/0004-6256/139/2/504).
- Bryan, M. L. et al. (2016). “Statistics of Long Period Gas Giant Planets in Known Planetary Systems”. In: *ApJ* 821(2), 89. doi: [10.3847/0004-637X/821/2/89](https://doi.org/10.3847/0004-637X/821/2/89).
- Butler, R. P. et al. (1996). “Attaining Doppler Precision of 3 M s-1”. In: *PASP* 108. doi: [10.1086/133755](https://doi.org/10.1086/133755).
- Butler, R. P. et al. (1997). “Three New “51 Pegasi-Type” Planets”. In: *ApJ* 474(2). doi: [10.1086/310444](https://doi.org/10.1086/310444).

- Butler, R. P. et al. (2004). “A Neptune-Mass Planet Orbiting the Nearby M Dwarf GJ 436”. In: *ApJ* 617(1). doi: [10.1086/425173](https://doi.org/10.1086/425173).
- Campbell, B. et al. (1988). “A Search for Substellar Companions to Solar-type Stars”. In: *ApJ* 331. doi: [10.1086/166608](https://doi.org/10.1086/166608).
- Cegla, H. (2019). “The Impact of Stellar Surface Magnetoconvection and Oscillations on the Detection of Temperate, Earth-Mass Planets Around Sun-Like Stars”. In: *Geosciences* 9(3). doi: [10.3390/geosciences9030114](https://doi.org/10.3390/geosciences9030114).
- Chaplin, W. J. et al. (2019). “Filtering Solar-Like Oscillations for Exoplanet Detection in Radial Velocity Observations”. In: *??jnlAJ* 157(4), 163. doi: [10.3847/1538-3881/ab0c01](https://doi.org/10.3847/1538-3881/ab0c01).
- Charbonneau, D. et al. (2000). “Detection of Planetary Transits Across a Sun-like Star”. In: *ApJ* 529(1). doi: [10.1086/312457](https://doi.org/10.1086/312457).
- Charbonneau, D. et al. (2009). “A super-Earth transiting a nearby low-mass star”. In: *Nature* 462(7275). doi: [10.1038/nature08679](https://doi.org/10.1038/nature08679).
- Charbonneau, P. (2014). “Solar Dynamo Theory”. In: *Annual Review of Astronomy and Astrophysics* 52(1). doi: [10.1146/annurev-astro-081913-040012](https://doi.org/10.1146/annurev-astro-081913-040012).
- Charbonneau, P. (2020). “Dynamo models of the solar cycle”. In: *Living Reviews in Solar Physics* 17(1), 4. doi: [10.1007/s41116-020-00025-6](https://doi.org/10.1007/s41116-020-00025-6).
- Chauvin, G. et al. (2004). “A giant planet candidate near a young brown dwarf. Direct VLT/NACO observations using IR wavefront sensing”. In: *A&A* 425. doi: [10.1051/0004-6361:200400056](https://doi.org/10.1051/0004-6361:200400056).
- Choi, J. et al. (2013). “Precise Doppler Monitoring of Barnard’s Star”. In: *ApJ* 764(2), 131. doi: [10.1088/0004-637X/764/2/131](https://doi.org/10.1088/0004-637X/764/2/131).
- Cloutier, R. et al. (2019). “Confirmation of the radial velocity super-Earth K2-18c with HARPS and CARMENES”. In: *A&A* 621. doi: [10.1051/0004-6361/201833995](https://doi.org/10.1051/0004-6361/201833995).
- Collier Cameron, A et al. (2019). “Three years of Sun-as-a-star radial-velocity observations on the approach to solar minimum”. In: *Monthly Notices of the Royal Astronomical Society* 487(1). doi: [10.1093/mnras/stz1215](https://doi.org/10.1093/mnras/stz1215).
- Cosentino, R. et al. (2012). “Harps-N: the new planet hunter at TNG”. In: *Ground-based and Airborne Instrumentation for Astronomy IV*. Ed. by I. S. McLean et al. Vol. 8446. International Society for Optics and Photonics. SPIE.

- Cui, G. et al. (2016). “Exact Distribution for the Product of Two Correlated Gaussian Random Variables”. In: *IEEE Signal Processing Letters* 23(11). doi: [10.1109/LSP.2016.2614539](https://doi.org/10.1109/LSP.2016.2614539).
- Cumming, A. et al. (2008). “The Keck Planet Search: Detectability and the Minimum Mass and Orbital Period Distribution of Extrasolar Planets”. In: *PASP* 120(867). doi: [10.1086/588487](https://doi.org/10.1086/588487).
- D’Angelo, G. et al. (2003). “Orbital Migration and Mass Accretion of Protoplanets in Three-dimensional Global Computations with Nested Grids”. In: *ApJ* 586(1). doi: [10.1086/367555](https://doi.org/10.1086/367555).
- Davies, A. (2015). “Effective Implementation of Gaussian Process Regression for Machine Learning”. PhD thesis. University of Cambridge.
- de Bruijne, J. H. J. (2012). “Science performance of Gaia, ESA’s space-astrometry mission”. In: 341(1). doi: [10.1007/s10509-012-1019-4](https://doi.org/10.1007/s10509-012-1019-4).
- Deeg, H. J. and R. Alonso (2018). “Transit Photometry as an Exoplanet Discovery Method”. In: *Handbook of Exoplanets*. Ed. by H. J. Deeg and J. A. Belmonte, 117. doi: [10.1007/978-3-319-55333-7_117](https://doi.org/10.1007/978-3-319-55333-7_117).
- Del Moro, D. (2004). “Solar granulation properties derived from three different time series”. In: *A&A* 428. doi: [10.1051/0004-6361:20040466](https://doi.org/10.1051/0004-6361:20040466).
- Del Zanna, G. and H. Mason (2013). “The Sun as a Star”. In: *Planets, Stars and Stellar Systems: Volume 4: Stellar Structure and Evolution*. Ed. by T. D. Oswalt and M. A. Barstow. Springer Netherlands: Dordrecht. ISBN: 978-94-007-5615-1. doi: [10.1007/978-94-007-5615-1_3](https://doi.org/10.1007/978-94-007-5615-1_3).
- Delfosse, X. et al. (1998). “Rotation and chromospheric activity in field M dwarfs”. In: *A&A* 331.
- Demin, S. A. et al. (2018). “Non-stationarity and cross-correlation effects in the MHD solar activity”. In: *Advances in Space Research* 61(2). doi: [10.1016/j.asr.2017.06.055](https://doi.org/10.1016/j.asr.2017.06.055).
- Díaz, M. R. et al. (2018). “The Test Case of HD 26965: Difficulties Disentangling Weak Doppler Signals from Stellar Activity”. In: ??*jnlAJ* 155(3), 126. doi: [10.3847/1538-3881/aaa896](https://doi.org/10.3847/1538-3881/aaa896).
- Doyle, L. R. et al. (2011). “Kepler-16: A Transiting Circumbinary Planet”. In: *Science* 333(6049). doi: [10.1126/science.1210923](https://doi.org/10.1126/science.1210923).
- Ducourant, C. et al. (2008). “An accurate distance to 2M1207Ab”. In: *A&A* 477(1). doi: [10.1051/0004-6361:20078886](https://doi.org/10.1051/0004-6361:20078886).

- Dumusque, X. et al. (2011). “The HARPS search for southern extra-solar planets. XXX. Planetary systems around stars with solar-like magnetic cycles and short-term activity variation”. In: *A&A* 535, A55. doi: [10.1051/0004-6361/201117148](https://doi.org/10.1051/0004-6361/201117148).
- Dumusque, X. et al. (2021). “Three years of HARPS-N high-resolution spectroscopy and precise radial velocity data for the Sun”. In: *A&A* 648, A103. doi: [10.1051/0004-6361/202039350](https://doi.org/10.1051/0004-6361/202039350).
- Dumusque, X. et al. (2012). “An Earth-mass planet orbiting α Centauri B”. In: *Nature* 491(7423). doi: [10.1038/nature11572](https://doi.org/10.1038/nature11572).
- Dumusque, X. et al. (2015). “HARPS-N OBSERVES THE SUN AS A STAR”. In: *The Astrophysical Journal* 814(2). doi: [10.1088/2041-8205/814/2/l21](https://doi.org/10.1088/2041-8205/814/2/l21).
- Dumusque, X. et al. (2011a). “Planetary detection limits taking into account stellar noise - I. Observational strategies to reduce stellar oscillation and granulation effects”. In: *A&A* 525. doi: [10.1051/0004-6361/201014097](https://doi.org/10.1051/0004-6361/201014097).
- Dumusque, X. et al. (2011b). “Planetary detection limits taking into account stellar noise - II. Effect of stellar spot groups on radial-velocities”. In: *A&A* 527. doi: [10.1051/0004-6361/201015877](https://doi.org/10.1051/0004-6361/201015877).
- Duvenaud, D. (2014). “Automatic model construction with Gaussian processes”. PhD thesis. University of Cambridge.
- Duvenaud, D. et al. (2013). “Structure Discovery in Nonparametric Regression through Compositional Kernel Search”. In: ed. by S. Dasgupta and D. McAllester. Vol. 28. Proceedings of Machine Learning Research 3. PMLR: Atlanta, Georgia, USA.
- Edelson, R. A. and J. H. Krolik (1988). “The Discrete Correlation Function: A New Method for Analyzing Unevenly Sampled Variability Data”. In: *ApJ* 333. doi: [10.1086/166773](https://doi.org/10.1086/166773).
- Faria, J. P. et al. (2016). “Uncovering the planets and stellar activity of CoRoT-7 using only radial velocities”. In: *A&A* 588, A31. doi: [10.1051/0004-6361/201527899](https://doi.org/10.1051/0004-6361/201527899).
- Faria, J. P. et al. (2020). “Decoding the radial velocity variations of HD 41248 with ESPRESSO”. In: *A&A* 635, A13. doi: [10.1051/0004-6361/201936389](https://doi.org/10.1051/0004-6361/201936389).
- Feng, F. et al. (2017). “Color Difference Makes a Difference: Four Planet Candidates around τ Ceti”. In: *??jnlAJ* 154(4), 135. doi: [10.3847/1538-3881/aa83b4](https://doi.org/10.3847/1538-3881/aa83b4).

- Figueira, P. et al. (2013). “Line-profile variations in radial-velocity measurements - Two alternative indicators for planetary searches”. In: *A&A* 557. doi: [10.1051/0004-6361/201220779](https://doi.org/10.1051/0004-6361/201220779).
- Figueira, P. et al. (2015). “Line-profile variations in radial-velocity measurements (Corrigendum) - Two alternative indicators for planetary searches”. In: *A&A* 582. doi: [10.1051/0004-6361/201220779e](https://doi.org/10.1051/0004-6361/201220779e).
- Figueira, P. (2018). “Deriving High-Precision Radial Velocities”. In: *Asteroseismology and Exoplanets: Listening to the Stars and Searching for New Worlds*. Ed. by T. L. Campante et al. Vol. 49. doi: [10.1007/978-3-319-59315-9_10](https://doi.org/10.1007/978-3-319-59315-9_10).
- Fischer, D. A. and J. Valenti (2005). “The Planet-Metallicity Correlation”. In: *ApJ* 622(2). doi: [10.1086/428383](https://doi.org/10.1086/428383).
- Foreman-Mackey, D. (2016). “corner.py: Scatterplot matrices in Python”. In: *The Journal of Open Source Software* 1(2). doi: [10.21105/joss.00024](https://doi.org/10.21105/joss.00024).
- Foreman-Mackey, D. et al. (2019). “emcee v3: A Python ensemble sampling toolkit for affine-invariant MCMC”. In: *The Journal of Open Source Software* 4(43), 1864. doi: [10.21105/joss.01864](https://doi.org/10.21105/joss.01864).
- Fressin, F. et al. (2013). “The False Positive Rate of Kepler and the Occurrence of Planets”. In: *ApJ* 766(2), 81. doi: [10.1088/0004-637X/766/2/81](https://doi.org/10.1088/0004-637X/766/2/81).
- Fulton, B. J. et al. (2017). “The California-Kepler Survey. III. A Gap in the Radius Distribution of Small Planets”. In: *??jnlAJ* 154(3), 109. doi: [10.3847/1538-3881/aa80eb](https://doi.org/10.3847/1538-3881/aa80eb).
- Gaudi, B. S. (2010). “Exoplanetary Microlensing”. In: *arXiv e-prints*, arXiv:1002.0332.
- Gaudi, B. S. (2012). “Microlensing Surveys for Exoplanets”. In: 50. doi: [10.1146/annurev-astro-081811-125518](https://doi.org/10.1146/annurev-astro-081811-125518).
- Gershman, S. et al. (2012). “Nonparametric variational inference”. In: *arXiv e-prints*, arXiv:1206.4665.
- Gilbertson, C. et al. (2020). “Towards Extremely Precise Radial Velocities: II. A Tool For Using Multivariate Gaussian Processes to Model Stellar Activity”. In: *arXiv e-prints*, arXiv:2009.01085.
- Gillon, M. et al. (2007). “Detection of transits of the nearby hot Neptune GJ 436 b”. In: *A&A* 472(2). doi: [10.1051/0004-6361:20077799](https://doi.org/10.1051/0004-6361:20077799).
- Ginzburg, S. et al. (2018). “Core-powered mass-loss and the radius distribution of small exoplanets”. In: *MNRAS* 476(1). doi: [10.1093/mnras/sty290](https://doi.org/10.1093/mnras/sty290).

- Gomes, R. et al. (2005). “Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets”. In: *Nature* 435(7041). doi: [10.1038/nature03676](https://doi.org/10.1038/nature03676).
- Goodman, J. and J. Weare (2010). “Ensemble samplers with affine invariance”. In: *Communications in Applied Mathematics and Computational Science* 5(1). doi: [10.2140/camcos.2010.5.65](https://doi.org/10.2140/camcos.2010.5.65).
- Gray, D. F. (1997). “Absence of a planetary signature in the spectra of the star 51 Pegasi”. In: *Nature* 385(6619).
- Gray, D. F. and A. P. Hatzes (1997). “Non-Radial Oscillation in the Solar-Temperature Star 51 Pegasi”. In: *The Astrophysical Journal* 490(1). doi: [10.1086/304857](https://doi.org/10.1086/304857).
- Grigoriu, M. (2002). *Stochastic Calculus: Applications in Science and Engineering*. Stochastic Calculus: Applications in Science and Engineering. Birkhäuser Boston. ISBN: 9780817642426.
- Grinstead, C. and J. Snell (2012). *Introduction to Probability*. American Mathematical Society. ISBN: 9780821894149.
- Gut, A. (2009). “The Multivariate Normal Distribution”. In: *An Intermediate Course in Probability*. Springer New York: New York, NY. ISBN: 978-1-4419-0162-0. doi: [10.1007/978-1-4419-0162-0_5](https://doi.org/10.1007/978-1-4419-0162-0_5).
- Hall, J. C. (2008). “Stellar Chromospheric Activity”. In: *Living Reviews in Solar Physics* 5(1). doi: [10.12942/lrsp-2008-2](https://doi.org/10.12942/lrsp-2008-2).
- Hathaway, D. H. (2015). “The Solar Cycle”. In: *Living Reviews in Solar Physics* 12(1), 4. doi: [10.1007/lrsp-2015-4](https://doi.org/10.1007/lrsp-2015-4).
- Hatzes, A. P. et al. (2003). “A Planetary Companion to γ Cephei A”. In: *ApJ* 599(2). doi: [10.1086/379281](https://doi.org/10.1086/379281).
- Haywood, R. D. et al. (2014). “Planets and stellar activity: hide and seek in the CoRoT-7 system”. In: *Monthly Notices of the Royal Astronomical Society* 443(3). doi: [10.1093/mnras/stu1320](https://doi.org/10.1093/mnras/stu1320).
- Haywood, R. D. (2016a). “Stellar Activity as a Source of Radial-Velocity Variability”. In: *Radial-velocity Searches for Planets Around Active Stars*. Springer International Publishing: Cham. ISBN: 978-3-319-41273-3. doi: [10.1007/978-3-319-41273-3_2](https://doi.org/10.1007/978-3-319-41273-3_2).
- Haywood, R. (2016b). *Radial-velocity Searches for Planets Around Active Stars*. Springer Theses. Springer International Publishing. ISBN: 9783319412733.

- Heinonen, M. et al. (2015). “Non-Stationary Gaussian Process Regression with Hamiltonian Monte Carlo”. In: *arXiv e-prints*, arXiv:1508.04319.
- Henry, G. W. et al. (2000). “A Transiting “51 Peg-like” Planet”. In: *ApJ* 529(1). doi: [10.1086/312458](https://doi.org/10.1086/312458).
- Higson, E. et al. (2019). “Dynamic nested sampling: an improved algorithm for parameter estimation and evidence calculation”. In: *Statistics and Computing* 29(5). doi: [10.1007/s11222-018-9844-0](https://doi.org/10.1007/s11222-018-9844-0).
- Hirano, T. et al. (2018). “Exoplanets around Low-mass Stars Unveiled by K2”. In: ??*jnlAJ* 155(3), 127. doi: [10.3847/1538-3881/aaa9c1](https://doi.org/10.3847/1538-3881/aaa9c1).
- Holman, M. J. and N. W. Murray (2005). “The Use of Transit Timing to Detect Terrestrial-Mass Extrasolar Planets”. In: *Science* 307(5713). doi: [10.1126/science.1107822](https://doi.org/10.1126/science.1107822).
- Isaacson, H. and D. Fischer (2010). “Chromospheric Activity and Jitter Measurements for 2630 Stars on the California Planet Search”. In: *ApJ* 725(1). doi: [10.1088/0004-637X/725/1/875](https://doi.org/10.1088/0004-637X/725/1/875).
- Jeffreys, H. (1983). *Theory of Probability*. International series of monographs on physics. Clarendon Press. ISBN: 9780198531937.
- Johnson, S. A. et al. (2020). “Predictions of the Nancy Grace Roman Space Telescope Galactic Exoplanet Survey. II. Free-floating Planet Detection Rates”. In: ??*jnlAJ* 160(3), 123. doi: [10.3847/1538-3881/aba75b](https://doi.org/10.3847/1538-3881/aba75b).
- Jones, D. E. et al. (2017). “Improving Exoplanet Detection Power: Multivariate Gaussian Process Models for Stellar Activity”. In: *arXiv e-prints*, arXiv:1711.01318.
- Jordan, M. I. et al. (1999). “An introduction to variational methods for graphical models”. In: *Machine learning* 37(2).
- Jurgenson, C. et al. (2016). “EXPRES: a next generation RV spectrograph in the search for earth-like worlds”. In: *Ground-based and Airborne Instrumentation for Astronomy VI*. Vol. 9908. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 99086T. doi: [10.1117/12.2233002](https://doi.org/10.1117/12.2233002).
- Karamanis, M. et al. (2021). “zeus: A Python implementation of Ensemble Slice Sampling for efficient Bayesian parameter inference”. In: *arXiv e-prints*, arXiv:2105.03468.
- Kass, R. E. and A. E. Raftery (1995). “Bayes Factors”. In: *Journal of the American Statistical Association* 90(430). doi: [10.1080/01621459.1995.10476572](https://doi.org/10.1080/01621459.1995.10476572).
- Keenan, P. C. and R. C. McNeil (1989). “The Perkins Catalog of Revised MK Types for the Cooler Stars”. In: 71. doi: [10.1086/191373](https://doi.org/10.1086/191373).

- Keller, C. U. et al. (2004). “On the Origin of Solar Faculae”. In: *ApJ* 607(1). doi: [10.1086/421553](https://doi.org/10.1086/421553).
- Ketu, S. and P. K. Mishra (2020). “Enhanced Gaussian process regression-based forecasting model for COVID-19 outbreak and significance of IoT for its detection”. In: *Applied Intelligence*.
- Kosiarek, M. R. and I. J. M. Crossfield (2020). “Photometry as a Proxy for Stellar Activity in Radial Velocity Analyses”. In: *The Astronomical Journal* 159(6). doi: [10.3847/1538-3881/ab8d3a](https://doi.org/10.3847/1538-3881/ab8d3a).
- Kullback, S. (1959). *Information Theory and Statistics*. Wiley: New York.
- Kung, S. H. (1996). “Proof Without Words: The Difference-Product Identities”. In: *Mathematics Magazine* 69(4). doi: [10.1080/0025570X.1996.11996453](https://doi.org/10.1080/0025570X.1996.11996453).
- Lagrange, A.-M. et al. (2010). “Using the Sun to estimate Earth-like planets detection capabilities - I. Impact of cold spots”. In: *A&A* 512. doi: [10.1051/0004-6361/200913071](https://doi.org/10.1051/0004-6361/200913071).
- Lanza, A. F. et al. (2019). “Sectoral r modes and periodic radial velocity variations of Sun-like stars”. In: *A&A* 623, A50. doi: [10.1051/0004-6361/201834712](https://doi.org/10.1051/0004-6361/201834712).
- Léger, A. et al. (2009). “Transiting exoplanets from the CoRoT space mission. VIII. CoRoT-7b: the first super-Earth with measured radius”. In: *A&A* 506(1). doi: [10.1051/0004-6361/200911933](https://doi.org/10.1051/0004-6361/200911933).
- Levenberg, K. (1944). “A method for the solution of certain non-linear problems in least squares”. In: *Quarterly of applied mathematics* 2(2).
- Li, S. et al. (2020). “Scalable Gaussian Process Regression Networks”. In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*. Ed. by C. Bessiere. Main track. International Joint Conferences on Artificial Intelligence Organization.
- Lin, D. N. C. and J. Papaloizou (1986). “On the Tidal Interaction between Protoplanets and the Protoplanetary Disk. III. Orbital Migration of Protoplanets”. In: *ApJ* 309. doi: [10.1086/164653](https://doi.org/10.1086/164653).
- Lin, D. N. C. et al. (1996). “Orbital migration of the planetary companion of 51 Pegasi to its present location”. In: *Nature* 380(6575). doi: [10.1038/380606a0](https://doi.org/10.1038/380606a0).
- Liu, D. C. and J. Nocedal (1989). “On the limited memory BFGS method for large scale optimization”. In: *Mathematical programming* 45(1).

- Lopez, E. D. and J. J. Fortney (2013). “The Role of Core Mass in Controlling Evaporation: The Kepler Radius Distribution and the Kepler-36 Density Dichotomy”. In: *ApJ* 776(1), 2. doi: [10.1088/0004-637X/776/1/2](https://doi.org/10.1088/0004-637X/776/1/2).
- Löptien, B. et al. (2018). “Global-scale equatorial Rossby waves as an essential component of solar internal dynamics”. In: *Nature Astronomy* 2. doi: [10.1038/s41550-018-0460-x](https://doi.org/10.1038/s41550-018-0460-x).
- Lovis, C. et al. (2011). “The HARPS search for southern extra-solar planets. XXXI. Magnetic activity cycles in solar-type stars: statistics and impact on precise radial velocities”. In: *arXiv e-prints*, arXiv:1107.5325.
- Lubin, D. et al. (2012). “Frequency of Maunder Minimum Events in Solar-type Stars Inferred from Activity and Metallicity Observations”. In: *ApJ* 747(2), L32. doi: [10.1088/2041-8205/747/2/L32](https://doi.org/10.1088/2041-8205/747/2/L32).
- Ma, B. et al. (2018). “The first super-Earth detection from the high cadence and high radial velocity precision Dharma Planet Survey”. In: *MNRAS* 480(2). doi: [10.1093/mnras/sty1933](https://doi.org/10.1093/mnras/sty1933).
- Mackay, D. (1998). “Introduction to Gaussian Processes”. In: *Neural Networks and Machine Learning*. Ed. by C. M. Bishop. Springer-Verlag. ISBN: 978-3540649281.
- Malbet, F. and A. Sozzetti (2018). “Astrometry as an Exoplanet Discovery Method”. In: *Handbook of Exoplanets*. Ed. by H. J. Deeg and J. A. Belmonte, 196. doi: [10.1007/978-3-319-55333-7_196](https://doi.org/10.1007/978-3-319-55333-7_196).
- Maldonado, J. et al. (2010). “A spectroscopy study of nearby late-type stars, possible members of stellar kinematic groups”. In: *A&A* 521, A12. doi: [10.1051/0004-6361/201014948](https://doi.org/10.1051/0004-6361/201014948).
- Marcy, G. et al. (2005). “Observed Properties of Exoplanets: Masses, Orbits, and Metallicities”. In: *Progress of Theoretical Physics Supplement* 158. doi: [10.1143/PTPS.158.24](https://doi.org/10.1143/PTPS.158.24).
- Marcy, G. W. and R. P. Butler (1995). “The Planet around 51 Pegasi”. In: *American Astronomical Society Meeting Abstracts*. Vol. 187. American Astronomical Society Meeting Abstracts, 70.04.
- Marcy, G. W. and R. P. Butler (1996). “A Planetary Companion to 70 Virginis”. In: *ApJ* 464. doi: [10.1086/310096](https://doi.org/10.1086/310096).
- Marois, C. et al. (2006). “Angular Differential Imaging: A Powerful High-Contrast Imaging Technique”. In: *ApJ* 641(1). doi: [10.1086/500401](https://doi.org/10.1086/500401).

- Marquardt, D. W. (1963). “An algorithm for least-squares estimation of nonlinear parameters”. In: *Journal of the society for Industrial and Applied Mathematics* 11(2).
- Martínez-Arnáiz, R. et al. (2010). “Chromospheric activity and rotation of FGK stars in the solar vicinity *** - An estimation of the radial velocity jitter”. In: *A&A* 520. doi: [10.1051/0004-6361/200913725](https://doi.org/10.1051/0004-6361/200913725).
- Mayor, M. (1980). “Metal abundances of F and G dwarfs determined by the radial velocity scanner CORAVEL”. In: *A&A* 87(1-2).
- Mayor, M. et al. (2011). “The HARPS search for southern extra-solar planets XXXIV. Occurrence, mass distribution and orbital properties of super-Earths and Neptune-mass planets”. In: *arXiv e-prints*, arXiv:1109.2497.
- Mayor, M. and D. Queloz (1995). “A Jupiter-mass companion to a solar-type star”. In: *Nature* 378(6555). doi: [10.1038/378355a0](https://doi.org/10.1038/378355a0).
- Mazeh, T. et al. (2016). “Dearth of short-period Neptunian exoplanets: A desert in period-mass and period-radius planes”. In: *A&A* 589, A75. doi: [10.1051/0004-6361/201528065](https://doi.org/10.1051/0004-6361/201528065).
- McArthur, B. E. et al. (2004). “Detection of a Neptune-Mass Planet in the ρ^1 Cancri System Using the Hobby-Eberly Telescope”. In: *ApJ* 614(1). doi: [10.1086/425561](https://doi.org/10.1086/425561).
- McIntosh, S. W. et al. (2014). “Deciphering Solar Magnetic Activity. I. On the Relationship between the Sunspot Cycle and the Evolution of Small Magnetic Features”. In: *ApJ* 792(1), 12. doi: [10.1088/0004-637X/792/1/12](https://doi.org/10.1088/0004-637X/792/1/12).
- Melia, F. and M. K. Yennapureddy (2018). “Model selection using cosmic chronometers with Gaussian Processes”. In: 2018(2), 034. doi: [10.1088/1475-7516/2018/02/034](https://doi.org/10.1088/1475-7516/2018/02/034).
- Meunier, N. and A. M. Lagrange (2020). “The effects of granulation and super-granulation on Earth-mass planet detectability in the habitable zone around F6-K4 stars”. In: *A&A* 642, A157. doi: [10.1051/0004-6361/202038376](https://doi.org/10.1051/0004-6361/202038376).
- Meunier, N. and Lagrange, A.-M. (2019). “Unexpectedly strong effect of super-granulation on the detectability of Earth twins orbiting Sun-like stars with radial velocities”. In: *A&A* 625. doi: [10.1051/0004-6361/201935099](https://doi.org/10.1051/0004-6361/201935099).
- Meunier, N. et al. (2010). “Using the Sun to estimate Earth-like planets detection capabilities - II. Impact of plages”. In: *A&A* 512. doi: [10.1051/0004-6361/200913551](https://doi.org/10.1051/0004-6361/200913551).

- Morbidelli, A. et al. (2005). “Chaotic capture of Jupiter’s Trojan asteroids in the early Solar System”. In: *Nature* 435(7041). doi: [10.1038/nature03540](https://doi.org/10.1038/nature03540).
- Murray, C. D. and A. C. M. Correia (2010). “Keplerian Orbits and Dynamics of Exoplanets”. In: *Exoplanets*. Ed. by S. Seager.
- Muterspaugh, M. W. et al. (2010). “The Phases Differential Astrometry Data Archive. V. Candidate Substellar Companions to Binary Systems”. In: ??*JnAJ* 140(6). doi: [10.1088/0004-6256/140/6/1657](https://doi.org/10.1088/0004-6256/140/6/1657).
- Nandy, D. et al. (2020). “Sunspot Cycle 25 is Brewing: Early Signs Herald its Onset”. In: *Research Notes of the American Astronomical Society* 4(2), 30. doi: [10.3847/2515-5172/ab79a1](https://doi.org/10.3847/2515-5172/ab79a1).
- Neal, R. M. (2000). “Slice Sampling”. In: *arXiv e-prints*, physics/0009028.
- Newton, I. (1833). *Philosophiae naturalis principia mathematica*. Vol. 1. G. Brookman.
- Nguyen, T. and E. Bonilla (2013). “Efficient Variational Inference for Gaussian Process Regression Networks”. In: *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*. Ed. by C. M. Carvalho and P. Ravikumar. Vol. 31. Proceedings of Machine Learning Research. PMLR: Scottsdale, Arizona, USA.
- Nguyen, T. V. (2015). “Scalable and automated inference for gaussian process models”. PhD thesis. Australian National University. Research School of Computer Science.
- Nirwan, R. S. and N. Bertschinger (2020). “Applications of Gaussian Process Latent Variable Models in Finance”. In: *Intelligent Systems and Applications*. Ed. by Y. Bi et al. Springer International Publishing: Cham. ISBN: 978-3-030-29513-4.
- NIST (2019). *The NIST Reference on Constants, Units, and Uncertainty*. National Institute of Standards and Technology, USA.
- Noyes, R. W. et al. (1984). “Rotation, convection, and magnetic activity in lower main-sequence stars”. In: *ApJ* 279. doi: [10.1086/161945](https://doi.org/10.1086/161945).
- Osborne, M. A. (2010). “Bayesian Gaussian processes for sequential prediction, optimisation and quadrature”. PhD thesis. Oxford University, UK.
- Oshagh, M. (2018). “Noise Sources in Photometry and Radial Velocities”. In: *Asteroseismology and Exoplanets: Listening to the Stars and Searching for*

- New Worlds*. Ed. by T. L. Campante et al. Springer International Publishing: Cham. ISBN: 978-3-319-59315-9.
- Owen, J. E. and D. Lai (2018). “Photoevaporation and high-eccentricity migration created the sub-Jovian desert”. In: *MNRAS* 479(4). doi: [10.1093/mnras/sty1760](https://doi.org/10.1093/mnras/sty1760).
- Owen, J. E. and Y. Wu (2013). “Kepler Planets: A Tale of Evaporation”. In: *ApJ* 775(2), 105. doi: [10.1088/0004-637X/775/2/105](https://doi.org/10.1088/0004-637X/775/2/105).
- Pagano, I. (2013). “Stellar Activity”. In: *Planets, Stars and Stellar Systems: Volume 4: Stellar Structure and Evolution*. Ed. by T. D. Oswalt and M. A. Barstow. Springer Netherlands: Dordrecht. ISBN: 978-94-007-5615-1. doi: [10.1007/978-94-007-5615-1_10](https://doi.org/10.1007/978-94-007-5615-1_10).
- Pasquini, L. et al. (2010). “CODEX: An Ultra-stable High Resolution Spectrograph for the E-ELT”. In: *The Messenger* 140.
- Patel, J. and C. Read (1996). *Handbook of the Normal Distribution, Second Edition*. 2nd. Statistics: A Series of Textbooks and Monographs. Taylor & Francis. ISBN: 9780824793425.
- Pepe, F et al. (2002). “HARPS: ESO’s coming planet searcher. Chasing exoplanets with the La Silla 3.6-m telescope”. In: *The Messenger* 110.
- Pepe, F. et al. (2002). “The CORALIE survey for southern extra-solar planets VII. Two short-period Saturnian companions to <ASTROBJ>HD 108147</ASTROBJ> and <ASTROBJ>HD 168746</ASTROBJ>”. In: *A&A* 388. doi: [10.1051/0004-6361:20020433](https://doi.org/10.1051/0004-6361:20020433).
- Pepe, F. et al. (2014). “ESPRESSO: The next European exoplanet hunter”. In: *Astronomische Nachrichten* 335(1). doi: [10.1002/asna.201312004](https://doi.org/10.1002/asna.201312004).
- Pepe, F. et al. (2021). “ESPRESSO at VLT. On-sky performance and first results”. In: *A&A* 645, A96. doi: [10.1051/0004-6361/202038306](https://doi.org/10.1051/0004-6361/202038306).
- Pepe, F. A. and C. Lovis (2008). “From HARPS to CODEX: exploring the limits of Doppler measurements”. In: *Physica Scripta Volume T* 130, 014007. doi: [10.1088/0031-8949/2008/T130/014007](https://doi.org/10.1088/0031-8949/2008/T130/014007).
- Pepe, F. et al. (2014). “Instrumentation for the detection and characterization of exoplanets”. In: *Nature* 513(7518). doi: [10.1038/nature13784](https://doi.org/10.1038/nature13784).
- Pepe, F. et al. (2018). “High-Precision Spectrographs for Exoplanet Research: CORAVEL, ELODIE, CORALIE, SOPHIE and HARPS”. In: *Handbook of Exoplanets*. Ed. by H. J. Deeg and J. A. Belmonte. Springer International

- Publishing: Cham. ISBN: 978-3-319-55333-7. doi: [10.1007/978-3-319-55333-7_190](https://doi.org/10.1007/978-3-319-55333-7_190).
- Perrakis, K. et al. (2014). “On the use of marginal posteriors in marginal likelihood estimation via importance sampling”. In: *Computational Statistics Data Analysis* 77. doi: <https://doi.org/10.1016/j.csda.2014.03.004>.
- Perryman, M. (2011). *The Exoplanet Handbook*. Cambridge University Press. doi: [10.1017/CBO9780511994852](https://doi.org/10.1017/CBO9780511994852).
- Petigura, E. A. et al. (2017). “The California-Kepler Survey. I. High-resolution Spectroscopy of 1305 Stars Hosting Kepler Transiting Planets”. In: [??jnlAJ 154\(3\), 107](https://doi.org/10.3847/1538-3881/aa80de). doi: [10.3847/1538-3881/aa80de](https://doi.org/10.3847/1538-3881/aa80de).
- Pizzolato, N. et al. (2003). “The stellar activity-rotation relationship revisited: Dependence of saturated and non-saturated X-ray emission regimes on stellar mass for late-type dwarfs”. In: *A&A* 397. doi: [10.1051/0004-6361:20021560](https://doi.org/10.1051/0004-6361:20021560).
- Plagemann, C. et al. (2008). “Nonstationary Gaussian Process Regression Using Point Estimates of Local Smoothness”. In: *Machine Learning and Knowledge Discovery in Databases*. Ed. by W. Daelemans et al. Springer Berlin Heidelberg: Berlin, Heidelberg. ISBN: 978-3-540-87481-2.
- Pollack, J. B. et al. (1996). “Formation of the Giant Planets by Concurrent Accretion of Solids and Gas”. In: 124(1). doi: [10.1006/icar.1996.0190](https://doi.org/10.1006/icar.1996.0190).
- Queloz, D. et al. (2001). “No planet for HD 166435”. In: *A&A* 379(1). doi: [10.1051/0004-6361:20011308](https://doi.org/10.1051/0004-6361:20011308).
- Queloz, D. (1995). “Echelle Spectroscopy by a CCD at Low Signal-to-Noise Ratio”. In: *Symposium - International Astronomical Union* 167. doi: [10.1017/S0074180900056473](https://doi.org/10.1017/S0074180900056473).
- Rajpaul, V. et al. (2015). “A Gaussian process framework for modelling stellar activity signals in radial velocity data”. In: *Monthly Notices of the Royal Astronomical Society* 452(3). doi: [10.1093/mnras/stv1428](https://doi.org/10.1093/mnras/stv1428).
- Rajpaul, V. M. (2017). “Gaussian process tools for modelling stellar signals and studying exoplanets”. PhD thesis. University of Oxford.
- Rasmussen, C. E. and C. K. I. Williams (2006). *Gaussian Processes for Machine Learning*. eng. 1st. MIT Press. ISBN: 0-262-18253-X.
- Reiners, A. (2009). “Activity-induced radial velocity jitter in a flaring M dwarf”. In: *A&A* 498(3). doi: [10.1051/0004-6361/200810257](https://doi.org/10.1051/0004-6361/200810257).

- Reiners, A. et al. (2012). “A Catalog of Rotation and Activity in Early-M Stars”. In: *??jnlAJ* 143(4), 93. doi: [10.1088/0004-6256/143/4/93](https://doi.org/10.1088/0004-6256/143/4/93).
- Reinhold, T. et al. (2019). “Transition from spot to faculae domination. An alternate explanation for the dearth of intermediate Kepler rotation periods”. In: *A&A* 621, A21. doi: [10.1051/0004-6361/201833754](https://doi.org/10.1051/0004-6361/201833754).
- Rieutord, M. and F. Rincon (2010). “The Sun’s Supergranulation”. In: *Living Reviews in Solar Physics* 7(1), 2. doi: [10.12942/lrsp-2010-2](https://doi.org/10.12942/lrsp-2010-2).
- Rivera, E. J. et al. (2005). “A $\sim 7.5M_{\oplus}$ Planet Orbiting the Nearby Star, GJ 876”. In: *The Astrophysical Journal* 634(1). doi: [10.1086/491669](https://doi.org/10.1086/491669).
- Robertson, P. et al. (2015). “Stellar Activity Mimics a Habitable-zone Planet around Kapteyn’s Star”. In: *ApJ* 805(2), L22. doi: [10.1088/2041-8205/805/2/L22](https://doi.org/10.1088/2041-8205/805/2/L22).
- Rowe, J. F. et al. (2014). “Validation of Kepler’s Multiple Planet Candidates. III. Light Curve Analysis and Announcement of Hundreds of New Multi-planet Systems”. In: *ApJ* 784(1), 45. doi: [10.1088/0004-637X/784/1/45](https://doi.org/10.1088/0004-637X/784/1/45).
- Ruane, G. et al. (2019). “Reference Star Differential Imaging of Close-in Companions and Circumstellar Disks with the NIRC2 Vortex Coronagraph at the W. M. Keck Observatory”. In: *??jnlAJ* 157(3), 118. doi: [10.3847/1538-3881/aafec2](https://doi.org/10.3847/1538-3881/aafec2).
- Rumelhart, D. E. et al. (1986). “Learning representations by back-propagating errors”. In: *Nature* 323(6088).
- Rutten, R. and G. Severino (2012). *Solar and Stellar Granulation*. Nato Science Series C: Springer Netherlands. ISBN: 9789400909113.
- Saar, S. H. (2009). “The Radial Velocity Effects of Stellar Surface Phenomena”. In: *15th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun*. Ed. by E. Stempels. Vol. 1094. American Institute of Physics Conference Series. doi: [10.1063/1.3099086](https://doi.org/10.1063/1.3099086).
- Saar, S. H. and R. A. Donahue (1997). “Activity-Related Radial Velocity Variation in Cool Stars”. In: *The Astrophysical Journal* 485(1). doi: [10.1086/304392](https://doi.org/10.1086/304392).
- Sahli Costabal, F. et al. (2019). “Machine learning in drug development: Characterizing the effect of 30 drugs on the QT interval using Gaussian process regression, sensitivity analysis, and uncertainty quantification”. In: *Computer Methods in Applied Mechanics and Engineering* 348. doi: [10.1016/j.cma.2019.01.033](https://doi.org/10.1016/j.cma.2019.01.033).

- Santerne, A. et al. (2015). “*pastis*: Bayesian extrasolar planet validation – II. Constraining exoplanet blend scenarios using spectroscopic diagnoses”. In: *Monthly Notices of the Royal Astronomical Society* 451(3). doi: [10.1093/mnras/stv1080](https://doi.org/10.1093/mnras/stv1080).
- Santerne, A. et al. (2016). “SOPHIE velocimetry of Kepler transit candidates. XVII. The physical properties of giant exoplanets within 400 days of period”. In: *A&A* 587, A64. doi: [10.1051/0004-6361/201527329](https://doi.org/10.1051/0004-6361/201527329).
- Santos, N. C. et al. (2002). “The CORALIE survey for southern extra-solar planets. IX. A 1.3-day period brown dwarf disguised as a planet”. In: *A&A* 392. doi: [10.1051/0004-6361:20020876](https://doi.org/10.1051/0004-6361:20020876).
- Santos, N. C. et al. (2004a). “Spectroscopic [Fe/H] for 98 extra-solar planet-host stars. Exploring the probability of planet formation”. In: *A&A* 415. doi: [10.1051/0004-6361:20034469](https://doi.org/10.1051/0004-6361:20034469).
- Santos, N. C. et al. (2004b). “The HARPS survey for southern extra-solar planets. II. A 14 Earth-masses exoplanet around μ Arae”. In: *A&A* 426. doi: [10.1051/0004-6361:200400076](https://doi.org/10.1051/0004-6361:200400076).
- Santos, N. C. et al. (2014). “The HARPS search for southern extra-solar planets - XXXV. The interesting case of HD41248: stellar activity, no planets?” In: *A&A* 566. doi: [10.1051/0004-6361/201423808](https://doi.org/10.1051/0004-6361/201423808).
- Santos, N. C. and J. P. Faria (2018). “Exoplanetary Science: An Overview”. In: *Asteroseismology and Exoplanets: Listening to the Stars and Searching for New Worlds*. Ed. by T. L. Campante et al. Springer International Publishing: Cham. ISBN: 978-3-319-59315-9.
- Sauer, T. (2008). “Nova Geminorum 1912 and the origin of the idea of gravitational lensing”. In: *Archive for History of Exact Sciences* 62(1).
- Schrijver, C. J. and C. Zwaan (2000). *Solar and Stellar Magnetic Activity*. Cambridge Astrophysics. Cambridge University Press. doi: [10.1017/CBO9780511546037](https://doi.org/10.1017/CBO9780511546037).
- Seager, S. and G. Mallén-Ornelas (2003). “A Unique Solution of Planet and Star Parameters from an Extrasolar Planet Transit Light Curve”. In: *ApJ* 585(2). doi: [10.1086/346105](https://doi.org/10.1086/346105).
- Serrano, L. M. et al. (2018). “Distinguishing the albedo of exoplanets from stellar activity”. In: *A&A* 611, A8. doi: [10.1051/0004-6361/201731206](https://doi.org/10.1051/0004-6361/201731206).
- Sher, D. (1968). “The Relativistic Doppler Effect”. In: 62.

- Sivia, D. and J. Skilling (2006). *Data Analysis: A Bayesian Tutorial*. Oxford science publications. OUP Oxford. ISBN: 9780191546709.
- Skilling, J. (2006). “Nested sampling for general Bayesian computation”. In: *Bayesian Anal.* 1(4). doi: [10.1214/06-BA127](https://doi.org/10.1214/06-BA127).
- Sokal, A. (1997). “Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms”. In: *Functional Integration: Basics and Applications*. Ed. by C. DeWitt-Morette et al. Springer US: Boston, MA. ISBN: 978-1-4899-0319-8. doi: [10.1007/978-1-4899-0319-8_6](https://doi.org/10.1007/978-1-4899-0319-8_6).
- Solanki, S. K. (1993). “Smallscale Solar Magnetic Fields - an Overview”. In: 63(1-2). doi: [10.1007/BF00749277](https://doi.org/10.1007/BF00749277).
- Solanki, S. K. (2003). “Sunspots: An overview”. In: *A&A Rev.* 11(2-3). doi: [10.1007/s00159-003-0018-4](https://doi.org/10.1007/s00159-003-0018-4).
- Sousa, S. G. et al. (2008). “Spectroscopic parameters for 451 stars in the HARPS GTO planet search program. Stellar [Fe/H] and the frequency of exo-Neptunes”. In: *A&A* 487(1). doi: [10.1051/0004-6361:200809698](https://doi.org/10.1051/0004-6361:200809698).
- Sousa, S. G. et al. (2011). “Spectroscopic stellar parameters for 582 FGK stars in the HARPS volume-limited sample. Revising the metallicity-planet correlation”. In: *A&A* 533, A141. doi: [10.1051/0004-6361/201117699](https://doi.org/10.1051/0004-6361/201117699).
- Speagle, J. S. (2020). “DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences”. In: *MNRAS* 493(3). doi: [10.1093/mnras/staa278](https://doi.org/10.1093/mnras/staa278).
- Spruit, H. C. (1976). “Pressure equilibrium and energy balance of small photospheric fluxtubes.” In: *Sol. Phys.* 50(2). doi: [10.1007/BF00155292](https://doi.org/10.1007/BF00155292).
- Stein, M. (1999). “Interpolation of Spatial Data: Some Theory for Kriging”. In.
- Styan, G. P. (1973). “Hadamard products and multivariate statistical analysis”. In: *Linear Algebra and its Applications* 6. doi: [https://doi.org/10.1016/0024-3795\(73\)90023-2](https://doi.org/10.1016/0024-3795(73)90023-2).
- Szabó, G. M. and L. L. Kiss (2011). “A Short-period Censor of Sub-Jupiter Mass Exoplanets with Low Density”. In: *ApJ* 727(2), L44. doi: [10.1088/2041-8205/727/2/L44](https://doi.org/10.1088/2041-8205/727/2/L44).
- Teh, Y.-W. et al. (2005). “Semiparametric Latent Factor Models”. In: *Artificial Intelligence and Statistics 10*.
- Tsiganis, K. et al. (2005). “Origin of the orbital architecture of the giant planets of the Solar System”. In: *Nature* 435(7041). doi: [10.1038/nature03539](https://doi.org/10.1038/nature03539).

- Udalski, A. et al. (1997). “Optical Gravitational Lensing Experiment. OGLE-2 – the Second Phase of the OGLE Project”. In: 47.
- Udry, S. and N. C. Santos (2007). “Statistical Properties of Exoplanets”. In: *Annual Review of Astronomy and Astrophysics* 45(1). doi: [10.1146/annurev.astro.45.051806.110529](https://doi.org/10.1146/annurev.astro.45.051806.110529).
- van de Kamp, P. (1982). “The planetary system of Barnard’s star”. In: *Vistas in Astronomy* 26(2). doi: [10.1016/0083-6656\(82\)90004-6](https://doi.org/10.1016/0083-6656(82)90004-6).
- Vaughan, A. H. and G. W. Preston (1980). “A survey of chromospheric CA II H and K emission in field stars of the solar neighborhood.” In: *PASP* 92. doi: [10.1086/130683](https://doi.org/10.1086/130683).
- Vaughan, A. H. et al. (1978). “Flux measurements of CA II H and K emission”. In: *Publications of the Astronomical Society of the Pacific* 90. doi: [10.1086/130324](https://doi.org/10.1086/130324).
- Virtanen, P. et al. (2020). “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17. doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
- Wendland, H. (2004). *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press. ISBN: 9781139456654.
- West, R. G. et al. (2019). “NGTS-4b: A sub-Neptune transiting in the desert”. In: *MNRAS* 486(4). doi: [10.1093/mnras/stz1084](https://doi.org/10.1093/mnras/stz1084).
- Wilcox, J. M. (1972). “Divers Solar Rotations”. In: *Cosmic Plasma Physics*. Ed. by K. Schindler. Springer US: Boston, MA. ISBN: 978-1-4615-6758-5.
- Wilson, A. G. (2014). “Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes”. PhD thesis. University of Cambridge.
- Wilson, A. G. et al. (2012). “Gaussian Process Regression Networks”. In: *Proceedings of the 29th International Conference on International Conference on Machine Learning*. ICML’12. Omnipress: Edinburgh, Scotland. ISBN: 978-1-4503-1285-1.
- Wilson, O. C. (1962). “Relationship Between Colors and Spectra of Late Main-Sequence Stars.” In: *ApJ* 136. doi: [10.1086/147437](https://doi.org/10.1086/147437).
- Wilson, O. C. (1978). “Chromospheric variations in main-sequence stars”. In: *ApJ* 226. doi: [10.1086/156618](https://doi.org/10.1086/156618).

- Winn, J. et al. (2005). “Variational message passing.” In: *Journal of Machine Learning Research* 6(4).
- Winn, J. N. (2018). “Planet Occurrence: Doppler and Transit Surveys”. In: *Handbook of Exoplanets*. Ed. by H. J. Deeg and J. A. Belmonte, 195. doi: [10.1007/978-3-319-55333-7\195](https://doi.org/10.1007/978-3-319-55333-7_195).
- Wolszczan, A. and D. A. Frail (1992). “A planetary system around the millisecond pulsar PSR1257 + 12”. In: *Nature* 355(6356). doi: [10.1038/355145a0](https://doi.org/10.1038/355145a0).
- Wolszczan, A. (1994). “Confirmation of Earth-Mass Planets Orbiting the Millisecond Pulsar PSR B1257+12”. In: *Science* 264(5158). doi: [10.1126/science.264.5158.538](https://doi.org/10.1126/science.264.5158.538).
- Wright, J. T. et al. (2012). “The Frequency of Hot Jupiters Orbiting nearby Solar-type Stars”. In: *ApJ* 753(2), 160. doi: [10.1088/0004-637X/753/2/160](https://doi.org/10.1088/0004-637X/753/2/160).
- Wright, J. T. and B. S. Gaudi (2013). “Exoplanet Detection Methods”. In: *Planets, Stars and Stellar Systems: Volume 3: Solar and Stellar Planetary Systems*. Ed. by T. D. Oswalt et al. Springer Netherlands: Dordrecht. ISBN: 978-94-007-5606-9. doi: [10.1007/978-94-007-5606-9_10](https://doi.org/10.1007/978-94-007-5606-9_10).
- Zhao, L. et al. (2020). “The EXPRES Stellar-signals Project. I. Description of Data”. In: *Research Notes of the AAS* 4(9). doi: [10.3847/2515-5172/abb8d0](https://doi.org/10.3847/2515-5172/abb8d0).