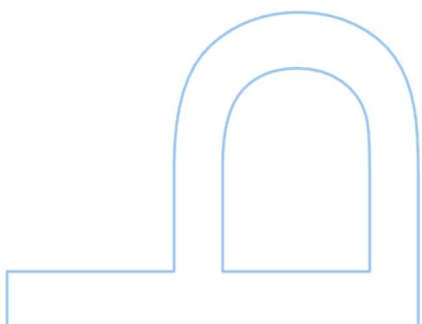
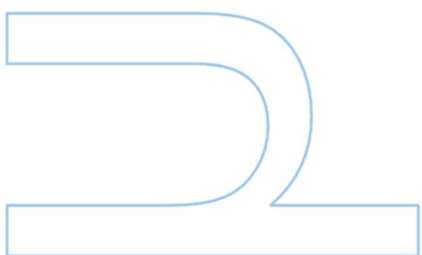
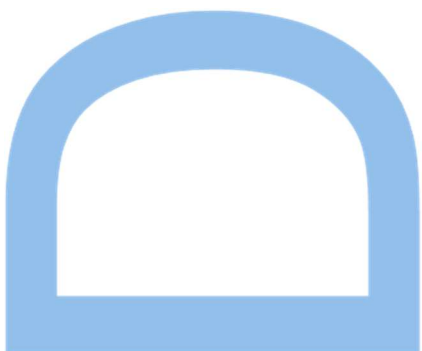




A new paradigm for the estimation of precise stellar radial velocities

André M. Silva
Programa Doutoral em Astronomia
Departamento de Física e Astronomia
Faculdade de Ciências da Universidade do Porto
2024

A new paradigm for the estimation of precise stellar radial velocities

André M. Silva

Programa Doutoral em Astronomia

Departamento de Física e Astronomia

2024

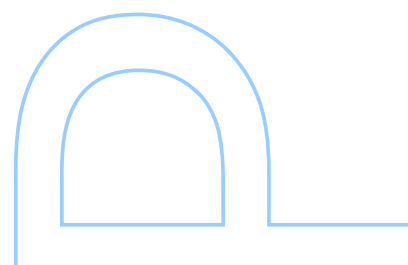
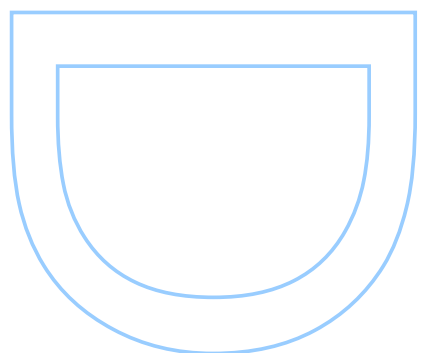
Orientador

Sérgio G. Sousa, Investigador, Instituto de Astrofísica e Ciências do Espaço

Coorientador

Nuno C. Santos, Professor Catedrático, Faculdade de Ciências

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

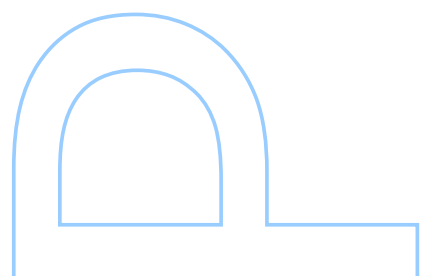
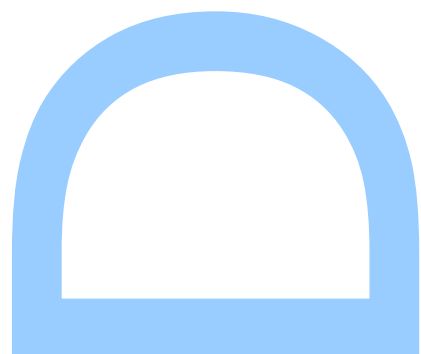




Todas as correções determinadas
pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ____/____/____



UNIVERSIDADE DO PORTO

DOCTORAL THESIS

A new paradigm for the estimation of precise stellar radial velocities

Author:

André M. SILVA

Supervisor:

Sérgio G. SOUSA

Co-supervisor:

Nuno C. SANTOS

Pedro T. P. Viana

*A thesis submitted in fulfilment of the requirements
for the degree of Doctoral Program in Astronomy*

at the

Faculdade de Ciências da Universidade do Porto

Departamento de Física e Astronomia

July 17, 2024

“ I try to avoid having thoughts. They lead to other thoughts, and — if you’re not careful — those lead to actions. Actions make you tired. I have this on rather good authority from someone who once read it in a book. ”

Lightsong in *Warbreaker*, written by Brandon Sanderson

Acknowledgements

The journey that was my PhD was a long and exciting one, only made possible by the contribution of a large number of people.

First and foremost, I extend my deepest appreciation to my supervisors, Sérgio, Nuno and Pedro, for their unwavering support, invaluable guidance, and insightful feedback throughout this journey. Despite their (often very) busy schedule, they always managed to find the time to help me in my scientific and personal endeavours. In the same vein, João Faria was also a key figure as an unofficial supervisor, spending way more than a few afternoons discussing science and other topics. Your combined expertise and dedication have been instrumental in shaping the work that was ultimately compiled in this thesis and my growth both as a scientist and a person.

Second, I want to thank all the people from the institute that were present throughout this incredible journey. The first person that comes to mind might have had a front row ticket to the Big Bang (yes, Jorge, I am talking about you). From discussions of science to software and even co-teaching multiple editions of a python course, I have truly taken a lot of your precious time during the past 4 years! In the same vein, thank you Manuel for always being the voice of reason and making sure that I kept focused. Thank you Pedro Figueira, Vardan, Elisa, Susana and Olivier for all the times that I went to you seeking for your opinion. It is my hope that one day I get fill the shoes of the retired professor that Vardan makes me to be. Elsa, Joana, Argentina thank you for dealing with me nagging you every time I was planning a trip. You have all made my life easier. Last, but not least, one can't forget those that have navigated the doctoral program at the same time as me: Eduardo, Tomás, Ana Rita, Catarina, Bárbara, Nuno, José, Daniel, Pedro (or should I say Cedro Punha?). Our attempt to construct a new node in Lanzarote brought us all together, even though some were not able to take part in the effort. You have brought me happiness and companionship, even though two of you attempted to hide your training to cook lamb in a typical Portuguese style from the region of Monção.

Heartfelt appreciation goes towards those that are closest to me: my parents, and grandparents, for their unwavering love, encouragement, and understanding throughout this journey. Your always-present support and interest in what I do made this journey far more enjoyable. Diana, thank you for always being there and dealing with, it meant the world! It is unfortunate that I lack enough space to mention everyone that is important

to me, both in the university and outside of it. However, do know that I hold dear all of those that are connected to me.

Without the collective support of all these individuals, and the financial support provided by FCT through a fellowship with reference *2020.05387.BD*, this thesis would not have been possible. Once again, thank you for believing in me and contributing to the realization of this milestone in my academic journey.

UNIVERSIDADE DO PORTO

Abstract

Faculdade de Ciências da Universidade do Porto

Departamento de Física e Astronomia

Doctoral Program in Astronomy

A new paradigm for the estimation of precise stellar radial velocities

by [André M. SILVA](#)

The discovery of the first exoplanet orbiting a solar-type star, back in 1995, launched a quest for the discovery of planets – also known as exoplanets – outside our solar system. Since then, more than 5000 exoplanets have been detected, but one of the boldest goals of current-day astrophysics is still standing: the detection and characterization of an Earth-like planet, orbiting a solar-like star. Such an ambitious goal has, in fact, motivated the development of new instruments and pushed towards the development of novel techniques to analyse those unique datasets. One of the most successful techniques for the detection and characterization of exoplanets is the Radial Velocity (RV) method, based on measuring relative variations in the velocity of a star, along our line of sight. This thesis focuses on the exploration of new approaches for the extraction of precise RVs from state-of-the-art stellar spectra and is divided into three different topics.

First, the thesis focuses on the framework and results of **s-BART**, a semi-Bayesian framework for template matching that imposes an assumption of achromaticity on the measured RV shifts. This assumption ensures RVs that are consistent with the expected effect, across different wavelengths, of an orbiting companion. The performance of the pipeline is explored through its application to a large number of targets from ESPRESSO, HARPS-N and CARMENES, as well as the analysis of a few individual systems.

The second focus of the thesis is the exploration of a systematic effect on template matching RVs that are extracted from intra-night observations. This systematic bias can reach the level of a few meters per second over a few hours of observations and can be found in different spectrographs, pipelines and stars, assuming that all observations were carried within the same night. Our analysis was unable to reach a definite conclusion,

albeit our tests suggest the problem results from a still unidentified issue with the creation of the stellar template that is used in template matching.

Lastly, we lay the groundwork for the expansion of the **s**-BART model through the modelling of stellar spectra under a probabilistic framework. This new model is built upon the simultaneous modelling of the stellar spectra whilst determining the optimal RV shift to describe differences between the observations. Gaussian Processes were employed for this modelling effort, leaving the door open for a future expansion: the modelling of spectral distortions, as introduced by stellar activity. The general architecture of this model is described, alongside of a proof of concept on simulated datasets. Despite the current computational limitations that inhibit the application to real datasets, the results are promising.

UNIVERSIDADE DO PORTO

Resumo

Faculdade de Ciências da Universidade do Porto

Departamento de Física e Astronomia

Programa Doutoral em Astronomia

Um novo paradigma para a estimativa de velocidades radiais precisas

por [André M. SILVA](#)

A descoberta, em 1995, do primeiro exoplaneta a orbitar uma estrela como o Sol motivou a procura por planetas fora do nosso sistema solar, também conhecidos por exoplanetas. Desde essa primeira descoberta, já foram detetados mais de 5000 exoplanetas. Contudo, um dos objetivos mais ambiciosos ainda está por ser atingido: a descoberta e a caracterização de um planeta semelhante ao nosso, a orbitar uma estrela como o nosso Sol. Na verdade, esta foi a motivação para a construção de novos instruments, bem como o desenvolvimento de novas técnicas para a análise de dados. Uma das metodologias mais utilizados para a descoberta e caracterização de exoplanetas é o método das Velocidades Radiais (RVs), que baseia na medição de variações relativas na velocidade de uma estrela, ao longo da nossa linha de visão. Esta tese têm como foco principal o desenvolvimento de novas técnicas para a extração de velocidades radiais, sendo que o trabalho pode ser dividido em três tópicos distintos.

Em primeiro lugar, há um foco na metodologia e resultados do **s-BART**, uma reformulação de *template matching* dentro de uma metodologia semi-Bayesiana que assume acromaticidade nas variações de RV que são medidas. Esta suposição garante que os RVs são consistentes com os efeitos, em vários comprimentos de onda, introduzidos pela presença de um companheiro orbital. A pipeline foi validada através da sua aplicação a um número elevado de estrelas observadas com os espectrógrafos ESPRESSO, HARPS-N e CARMENES, bem como através da análise de alguns sistemas individuais.

O segundo foco da tese é a exploração de um efeito sistemático em RVs obtidos com *template matching*. Este efeito aparece em casos onde todas as observações foram obtidas durante a mesma noite, com o efeito a atingir alguns metros por segundo após algumas

horas. Ele foi observado em diferentes estrelas, instrumentos, e algoritmos de *template matching*. A análise detalhada quer é exposta neste documento não permitiu uma conclusão definitiva sobre a causa deste efeito sistemático, embora os resultados aparentem mostrar que está conectado à construção do modelo (template) estelar.

Por último, é apresentada as fundações de uma possível expansão ao modelo **s-BART**, através de uma modelação dos espectros estelares dentro de uma arquitetura probabilística. Este novo modelo consiste numa extração de RVs simultânea à modelação dos espectros estelares. Assim, é possível evitar a criação, *a priori*, de um modelo estelar (template), onde são feitas algumas aproximações. Ainda para mais, a mudança para um modelo deste tipo, abre a porta a uma expansão futura para a modelação direta das distorções dos espectros, causadas por efeitos de atividade estelar. De momento, é apresentada apenas uma prova de conceito em datasets simulados, devido a limitações computacionais. Contudo, os resultados preliminares são promissores.

Table of Contents

Acknowledgements	v
Abstract	vii
Resumo	ix
Table of Contents	xi
List of Figures	xv
1 The search for an Earth analog	1
1.1 The transit method	4
1.2 The radial velocity method	5
1.2.1 Planetary parameters from RV observations	7
1.2.2 Challenges for optimal Radial Velocity extraction	8
1.2.2.1 Earth’s atmosphere	8
1.2.2.2 Stellar activity	12
1.2.3 High-precision spectrographs for exoplanet detection	16
1.2.4 Different approaches for RV extraction	18
1.3 Motivation for this Thesis	24
2 A semi-Bayesian implementation of template matching	27
2.1 Model preparation	28
2.1.1 Pre-processing of the data	28
2.1.2 Creation of a telluric template	30
2.1.3 Creation of a stellar template	32
2.1.3.1 Evaluating the uncertainty associated with the stellar tem- plate	34
2.1.3.2 Removal of outliers in the spectra	37
2.2 A classical template matching approach	39
2.3 Implementation within a Bayesian framework	43
2.3.1 The RV model	44
2.3.2 Characterization of the RV posterior distribution	47
2.4 Benchmarking the s-BART pipeline	50
2.4.1 Benchmarking with HARPS	50
2.4.2 Benchmarking with ESPRESSO data	52

2.4.2.1	The impact of the number of observations in the template .	53
2.4.2.2	Evaluating the validity the RV uncertainties	54
2.4.2.3	Comparison of the RV scatter and precision	55
2.4.2.4	Nightly Zero Point (NKP) variation	57
2.5	Large-scale application of s-BART	60
2.5.1	The KOBE survey from CARMENES	60
2.5.2	Larger ESPRESSO sample	62
2.5.3	Early results with HARPS-N data	63
2.6	Limitations and caveats of the model	65
3	An intra-night RV trend	69
3.1	Differences between single- and multi-night observations	70
3.2	The impact of the atmosphere	72
3.2.1	Change in the transmittance threshold	73
3.2.2	Simulating telluric variability	74
3.2.3	Correction of water features in the data	76
3.3	Evaluating the chromatic evolution of the trend	77
3.3.1	A division by the detectors of the spectrographs	77
3.3.2	Order-by-order RV measurements	78
3.3.3	Wavelength-binning the spectral orders	81
3.4	Evaluating the construction of the stellar template	82
3.4.1	A subset of observations as a template	83
3.4.2	Applying a template from a different night	85
3.4.3	A single observation as a template	86
3.5	Evaluating the interpolation algorithm	88
3.5.1	The change of sampling of the spectral lines	89
3.5.2	Evaluation of a simple Gaussian line	91
3.5.3	Evaluation on a sum of Gaussian functions	94
3.6	Large scale impact of the RV trend and a search for correlations	96
3.7	A multi-night template applied to intra-night data	99
3.8	Overall summary of the RV trend	102
4	A fully-Bayesian RV extraction model	105
4.1	A brief primer in Gaussian Processes	106
4.1.1	Gaussian Processes	107
4.1.2	Application of Gaussian Processes to large datasets	108
4.2	An overview of the components from the model	110
4.2.1	Accounting for the modelling of telluric features	111
4.2.2	Modelling stellar spectra	111
4.2.3	Evaluating the degrees of freedom of the model	113
4.3	In-depth description of the RV-extraction model	114
4.3.1	Choice of the GP model for a large dataset	115
4.3.2	Mean model of the Gaussian Process	118
4.3.3	The parameter vector of Radial Velocities	119
4.3.4	Exploring the posterior distribution of the RVs	120
4.3.5	The final architecture of the model	121
4.4	Application of the method	123

4.4.1 Preliminary results in a ideal case	123
4.4.2 Application to real observations	125
4.5 The caveats and avenues for expansion of the current model	126
5 Conclusions and future work	129
A Three-dimensional orbital parameters	133
B Nightly Zero Point (NZP)	135
C Generation of simulated spectra	137
D Error propagation through a cubic-spline interpolation	139
Bibliography	143

List of Figures

1.1	Yield of confirmed exoplanetary detections over the years, as provided by different methods. Image taken from “ <i>NASA’s exoplanet archive</i> ” in February 2024.	2
1.2	Mass-radius relation of the confirmed exoplanets, in units of Earth’s mass and radius, alongside their associated uncertainties. Exoplanets for which the uncertainty in either measurement, as reported in <i>exoplanet.eu</i> , was higher than 50% were discarded. The colormap represents the year of discovery, since 51 Peg in 1995 up to 2023.	3
1.3	Representation of the flux decrease due to an orbiting companion.	4
1.4	Schematic of the gravitational interaction between a star and a orbiting companion, showcasing the Doppler shift (left side) of the light that reaches the observer (telescope) as the planet orbits the star (right side). As the star moves in our direction, the light is blue-shifted, whilst it is red-shifted as the star moves in the opposite direction.	6
1.5	Doppler-shift of the telluric features, as caused by the BERV correction of different observations, as represented in the different colours. The presented observations were all collected with BERV in the interval [16.9, 26.7] km s ⁻¹ , with a span of ~ 9 km s ⁻¹ . Top: BERV-corrected spectra; Bottom: Spectra without the BERV correction.	10
1.6	Telluric transmittance spectrum for a 25% humidity and a airmass of 1.5, covering the optical and the near infra-red region, highlighting the wavelength coverage of different spectrographs.	11
1.7	Percentage of the wavelength coverage of optical regions from different spectrographs, whilst assuming a threshold of 1% drop on transmittance. The transmittance spectrum was generated with TelFit, assuming different airmasses and water content in the atmosphere for ESPRESSO (black), HARPS-N (red) and the optical arm of CARMENES (blue).	12
1.8	Radial velocity time-series from Earth (blue) compared to solar RVs from the HARPS-N solar telescope (black).	13
1.9	Schematic of the CCF method, with a weighted mask. Left: Cross correlation with the weighted mask. We can see that deep thin lines have larger weights than shallower ones. Right: Final CCF profile, resulting from the cross correlation.	19
1.10	Schematic of the template matching process, for three different Doppler shifts of the stellar template. Left: Alignment between the stellar spectrum and template with different RV shifts. Right: Evolution of the χ^2 curve for the different RV shifts. When the template and spectrum are aligned the curve reaches its minimum value.	21

1.11	Comparison of the M5 CCF mask from ESPRESSO (blue) with a stellar template (black) constructed for a M-type star that uses this mask.	22
2.1	Schematic of the pre-processing of the spectral data, applied before the RV estimation.	28
2.2	Rejection of a wavelength window around the $H\alpha$ line for two observations of an active M-type star.	29
2.3	Comparison, for three spectral regions, between the telluric spectrum generated with Telfit (black) with the continuum level obtained with a median filter (dashed blue line) and the telluric threshold (dotted red line) built from the continuum. Image adapted from Silva et al. (2022).	31
2.4	Gaps in the wavelength coverage introduced by the pre-processing stages (top panel, with an unrealistic gap for ease of visualization) and at the edges of the spectral order (bottom panels). The red lines represent the boundary between the regions where there is an overlap between the 3 observations and those where one observation is missing data.	33
2.5	Histogram of the airmass at the start of 4000 observations of ESPRESSO's WG1 sample. The vertical lines represent a airmass cutoff of 1.2 (red) and 1.5 (dark orange), with the percentages representing the number of observations whose airmasses fall below the cutoffs.	34
2.6	Histogram of the SNR ratio, for the 100th order, of the template and each of the 21 individual observations of an M4 dwarf, obtained after the fibre upgrade of ESPRESSO in mid 2019. In order to compare the impact of the number of observations used to construct the template in its associated uncertainty, one template was built with 8 observations (black curve) and the other with 13 observations (blue curve), selected by the airmass at the start of each observation. For all observations we also highlight by filling the bins with the corresponding colour, the comparison with observations with a SNR of at least 100 for the selected spectral order. The figure was taken from Silva et al. (2022).	36
2.7	Outlier removal routine for two regions of the same order, with respect to an observation for which outlier identification achieved convergence after the first iteration. For representation purposes we normalized the flux in the center and the edge of the order. Top row: Comparison between the stellar template (red line) and the spectra (black line). The blue crosses represent the points that were flagged by the method. Bottom row: Differences between the template and spectra (black points). The blue line is the median value, whilst the dotted red lines represent a 6σ difference from it. Image adapted from Silva et al. (2022).	39
2.8	Schematic of a classical template matching algorithm.	40
2.9	The χ^2 curve (black) of one spectral order from the ESPRESSO spectrograph, compared against the parabolic fit (red). The red circles represent the tentative RV values provided by the Brent minimizer before convergence, whilst the dotted green line is the DRS estimate and the dashed blue line is the RV estimated through Equation 2.3. Figure taken from Silva et al. (2022).	42
2.10	Schematic of the semi-Bayesian approach, considering both an MCMC and Laplace approximation to characterize the posterior distribution.	45

2.11	Laplace approximation (red) applied to two simulated posterior distributions (blue). Top: Normal distribution; Bottom: Normal distribution skewed towards higher values.	48
2.12	Comparison of the RV posterior distribution derived using the MCMC methodology (black curve) with the result of the Laplace approximation (dashed red line), for one observation of the M4 dwarf considered in Section 2.1.3.1. Image adapted from Silva et al. (2022).	49
2.13	Differences between the RV posterior distribution as characterized through the MCMC methodology and the Laplace approximation for the full RV dataset of Proxima Centauri, as used in (Faria et al., 2022). Top: Difference in the RV expected value. Bottom: Difference in the standard deviation of the RV posterior distribution.	50
2.14	RV time-series for 22 observations of Barnard’s star, corrected for the secular acceleration of the star. The black stars and orange crosses are RV estimates obtained with the classical and s -BART methodologies, respectively, with its own mean RV subtracted (for comparison purposes). The blue triangles and red dots are HARPS-TERRA and SERVAL RV estimates, respectively, also with their own mean RV subtracted. Image adapted from Silva et al. (2022).	51
2.15	Evolution of the RV scatter (top panel) and median uncertainty (bottom panel) with respect to 13 ESPRESSO18 observations of a M-type star, as a function of the number of spectra used to construct the template. In black we have the results from the classical methodology and, and in blue those from the s -BART methodology. Image adapted from Silva et al. (2022).	54
2.16	Comparison of the results obtained with our template-matching methodologies and with the CCF-based ESPRESSO DRS, for ESPRESSO18 (black) and ESPRESSO19 (red) observations of the selected targets. Top panel: Ratio of the rms in the template-matching and DRS RV time-series. The results from the classical and s -BART methods are represented by circles and diamonds, respectively. Bottom panel: Median RV uncertainty for each target, as computed by the DRS (crosses), the classical approach (circle) and the s -BART method (diamonds). Image adapted from Silva et al. (2022)..	56
2.17	Comparison of the RV uncertainties obtained for all ESPRESSO18 (top panel) and ESPRESSO19 (bottom panel) observations used in Figure 2.16, separated by spectral type. In blue we have the RV uncertainty estimated by the DRS, in black and red those estimated through the s -BART and classical methodologies, respectively. Image adapted from Silva et al. (2022).	57
2.18	Nightly Zero Point (NZP), black dots, for each night when at least 3 different targets were observed. The date of the fiber link upgrade is highlighted with a dashed red line. The zero-centered RV of each target, observed in the night, is represented with the blue crosses. As derived with results obtained with the following methodologies: ESPRESSO DRS (Top); classical (Middle); s -BART (Bottom). Image adapted from Silva et al. (2022).	59
2.19	RV scatter for the KOBE sample for the CCF method (black), SERVAL (blue) and s -BART (red), highlighting the standard stars and the science targets. The scatter was computed after the correction of the NZP, for both standard and science stars.	61
2.20	Comparison of the s -BART RVs from one HARPS-N target, using data from the old (black) and new (green) DRS versions. For comparison purposes, HARPS-TERRA RVs are also shown (red), as obtained with the old pipeline.	64

3.1	Mean-subtracted RV time-series for one night of ESPRESSO observations of HD40307, a K2 star observed in December 2018 (Pepe et al., 2021). The CCF RVs (black) are compared against those from SERVAL (blue) and s-BART (red).	69
3.2	Variation of the barycentric motion (BERV) in the multiple nights of observations of HD40307.	71
3.3	Impact of the BERV on the location of one selected spectral line on the detector, over 8 hours of observations of HD40307 (the first observation of the night is represented in blue and the last in yellow).	72
3.4	Impact on the RV trend from template matching after creating a telluric mask with different thresholds. The thresholds were applied to the same transmittance spectrum, generated through <i>Telfit</i> .	73
3.5	Residuals of the RV time-series obtained when applying the <i>classical</i> TM algorithm to the synthetic datasets generated with <i>Telfit</i> , using as a basis observations from HD40307 (top row) and a target with higher SNR (bottom row). The title of each panel provides a measurement of the standard deviation of the RV time-series. Left column: RV residuals without injecting any transmittance profile; Center column: RV residuals from the dataset where the atmospheric parameters were kept constant; Right column: RV residuals from the dataset where the atmospheric parameters were those recorded in ESPRESSO's data products.	75
3.6	Top panel: Comparison of the s-BART RVs from ESPRESSO observations of HD40307, when masking (black) and correcting (red) the telluric features; Top panel: Difference between the HD40307 RVs extracted whilst masking the tellurics and those obtained after the telluric correction.	76
3.7	RV residuals extracted from the spectral data of each ESPRESSO (top and middle panel) and HARPS (bottom panel) detector. In black, we have the TM RVs computed with all spectral orders, whilst the colours represent the detector from which the RV time-series was extracted. The middle row is the data from the 5 nights of observations of HD40307.	78
3.8	Comparison of the RVs obtained for two spectral orders in the red detector. The colours group the observations that were taken in the same night. Top: Spectral order 96, with a central wavelength of ~ 537.6 nm; Bottom: Spectral order 113, with a central wavelength of ~ 578.2 nm.	79
3.9	Evolution of the RV (colour-coded) for the central wavelength of each spectral order (XX-axis) across time, measured since the first observation of the night (YY-axis). The order-wise RVs are normalized by the maximum RV measured on that order over the entire night. Each panel corresponds to one of the 5 nights of ESPRESSO observations of HD40307, where the x-axis is the central wavelength of the corresponding spectral order. The two upper panels provide a transmittance profile (for comparison purposes) and a zoom above the 95% level. The percentage of the wavelength coverage (in each spectral order) whose tellurics are deeper than 1% are presented as a red dashed line in the bottom panels. Their values are indexed to the rightmost axis. The variation of RV is encoded as a colour map variation, with the brighter colours representing an higher RV measurement and the scales being independent across the different nights.	80

3.10	Comparison of the RV time-series from the CCF (black) and those from template matching applied all spectral orders. The RVs were derived for the three chunks defined in this Chapter: <i>start</i> in blue, <i>middle</i> in green and <i>End</i> in red.	82
3.11	Comparison of the HD40307 RVs, from each night, extracted with a stellar template constructed from the first 20 observations (red), the 20 observations observations with the lowest airmass (green) and the final 20 observations (blue). For visualization purposes, the RVs of each time-series had their corresponding mean values removed. The CCF RVs are represented as the black crosses, whilst the shaded regions highlight the observations that are present in each of the three templates.	83
3.12	RV variability when constructing stellar templates from an incrementing number of observations (colour coded, from blue to yellow). For visualization purposes, the bottom panel excludes the time series with higher variability, associated with the templates constructed from a smaller number of observations (bluer colours). The red dashed line represents the timestamp associated with the last observation of the last template to be excluded from the top panel.	84
3.13	RVs of the 5 consecutive nights of HD40307, when using stellar templates constructed from each night of the dataset. Each row corresponds to a stellar template constructed from the corresponding night, whilst the columns correspond to the data from which we extract RVs. The data from the first night of observations is presented in the top-left corner. For comparison purposes, the x-axis of each panel represents the number of hours spent since 00:27, local time, selected due to being the hour of the first observation of the first night. For comparison purposes, a sparse representation of CCF RVs is provided as orange circles.	86
3.14	RVs of all nights of HD40307, when using a single observation as the stellar template, taken from the start (red) and end (blue) of the night, as well as the one with the lowest airmass (black). A move along a given column leads to RVs from the same night extracted with different stellar templates, whilst one alongside a row keeps the stellar templates but changes the data to which they were applied. The data from the first night of observations is presented in the top-left corner and the hour at which the template was created is denoted as the dashed lines, coloured with the corresponding colour. For comparison purposes, the x-axis of each panel represents the number of hours spent since 00:27, local time, selected due to being the hour of the first observation of the first night. For comparison purposes, the CCF RVs are plotted as dark-yellow circles.	87
3.15	Impact of the choice of the algorithm used to interpolate the stellar spectra, as seen in the nightly RVs of HD40307. For this comparison, three interpolation algorithms were used: quadratic spline (blue), cubic spline (black) and pchip (red), all implemented in <i>scipy</i> . For comparison purposes the CCF RVs are over-plotted as green circles and each time-series had its own mean value removed.	89

- 3.16 Schematic of the impact of the interpolation on line shape, depending on the sampling of the line. For representation purposes, the sampling is sparser than the one in real observations. The diamonds correspond to different sets (colour coded) of equidistant points selected from the same Gaussian (shown as the black line). The dashed line is the result from different interpolation algorithms, using the same-coloured diamonds as the nodes for the interpolation. The insets provide a zoom-in in the core of the line. **Top:** Linear interpolation; **Bottom:** Cubic-spline interpolation. 90
- 3.17 Gaussian line, built on top of ESPRESSO's wavelength coverage and centred on μ_G , and its first and second numerical derivatives. On the bottom row we have a colormap of the residuals after interpolating the original Gaussian to a given RV shift. The dashed purple lines mark the 480 m s^{-1} , the mean size of one ESPRESSO pixel. 92
- 3.18 **Left:** Residuals of flux (top row), first derivative (middle row), and second derivative (bottom row) for four RV shifts; **Right:** Variability of the residuals from a single pixel, near the core of the line, with the change in the injected RV. The pixel is highlighted with a vertical dashed line on the left column. 93
- 3.19 **Top:** RV dependency of the interpolated flux value of four different points in the gaussian. Two were selected leftwards of the centre (blue colours) and the other rightwards (red colours). **Bottom:** RV shift between the central point of the original μ and the one obtained through a fit to the interpolations of the simplistic line. The dashed purple line represents the RV value for which the deepest wavelength of the line is closest to its neighbours. The missing points were those for which convergence was not achieved. 93
- 3.20 Residuals in the TM time-series, from the creation of the *order of gaussians* using RVs in the interval $[-1, 1] \text{ km s}^{-1}$, with separations of 10 m s^{-1} between points. The dashed lines represent the maximum (red), mean (green) and minimum (blue) RV coverage of the pixels from the selected order. 95
- 3.21 Template matching RVs being measured on the simulated spectral order, albeit with different samplings. The change of sampling of the lines follows the logic of intra-night observations, as it is slowly Doppler-shifted. The dashed lines represent the maximum (red), mean (green) and minimum (blue) RV coverage of the pixels from the selected order 96
- 3.22 Difference between the CCF and the TM RVs as a function of hours spent observing after the first observation of the corresponding night. Each night of observation (of the same target) was processed independently from others. However, all observations of the same target share the same colour. **Top:** Results for all targets. **Bottom:** Zoom-in of the difference. 97
- 3.23 Analysis of the RV slope for all targets discussed in this Chapter, obtained with a first degree fit to the residuals of each night. **Top panel:** Histogram of the slope, across all nights. **Middle and bottom panels:** Correlation between the slope and the median BERV value, the SNR at order 70, the mean stellar RV and the effective temperature. The multiple nights of the same target are represented with the same colour scheme. 99

3.24	Impact of a change in the stellar template in the retrieved RV values. Left: Comparison of the RVs obtained through the <i>full</i> template (red) and <i>nightly</i> templates (black) applied to the 3 nights of observations. Right: Residuals of the RVs from the multi-night observations when using the <i>Full</i> and <i>Nightly</i> templates. Each row corresponds to the usage of a different <i>nightly</i> template.	100
3.25	Comparison of different instrumental metrics for the 4 datasets considered in this (current) analysis. Nights 0 to 2 are represented in black, blue, and green, respectively. The values from the <i>full</i> dataset are represented in red. From top to bottom, the panels present the humidity, airmass, BERV velocity and the ambient temperature.	101
4.1	Comparison between 2 <i>Matérn</i> kernels, one with $\nu = 3/2$ (black) and another with $\nu = 5/2$ (red), with a length scale of 2 (arbitrarily chosen for visualization purposes). Left column: Evolution of the covariance with distance from a given point; Right column: Random samples from each kernel, without conditioning on data.	107
4.2	GP prediction for a small region of stellar spectra, using a <i>Matérn-5/2</i> kernel. The black crosses represent the flux measurements, the dashed blue line represents the mean prediction of the GP, and the red region the 1σ uncertainty of the predictions.	113
4.3	Simplified schematic of the fully-Bayesian RV extraction methodology, modelling the stellar spectra and telluric features to retrieve the relative RV shifts between multiple observations.	114
4.4	Negative log-likelihood evaluated in a grid of possible parameters, using three GP kernels: M32 from <i>celerite</i> (left column) and <i>tinyGP</i> (middle column), and M52 from <i>tinyGP</i> . The GP model was applied to one spectral order of 15 ESPRESSO observations of $\tau - Ceti$, aligned assuming the CCF RVs. The blue values represent those where the log-likelihood is minute, when compared against the optimal one (highlighted in the intersection of the green dashed lines). The white regions in the plot are those for which the <i>celerite</i> implementation failed to produce a valid result, and Table 4.4 presents the values of the hyper-parameters at the maximum-likelihood point. The top row present the results when using the original wavelength solution, whilst the bottom one uses those transformed through Equation 4.5.	117
4.5	Schematic of the entire GP model, considering the different steps presented in this Chapter. For the current implementation of the model, the telluric modelling is fully disabled.	122
4.6	Histogram of the residual scatter obtained in the datasets that were constructed with different <i>amplitudes</i> , without the re-sampling of the flux (left column) and with it (right column). The top row represents the datasets with an injection <i>amplitude</i> of 100 m s^{-1} , the middle those with 10 m s^{-1} and, the bottom one with 1 m s^{-1}	124

4.7	Comparison of the relative RVs obtained through a classical template matching algorithm (black), s-BART (red), CCF (yellow) and the GP model (blue), obtained from the datasets of 5 (left) and 11 (right) HARPS observations. All RVs, except those from the CCF method, were obtained using only the selected spectral order. The observation selected as a reference for each dataset (left and right columns) are not represented and, in both cases, a constant initial guess was used (dashed blue line). The label of each column provides a measurement of the standard deviation of the RVs from the corresponding method.	126
A.1	Representation of the orbit of an exoplanet, with an observer above the Z axis. Image adapted from Murray & Correia (2010).	134
C.1	Gaussian distributions (red line) placed on each pixel of a reference spectrum (black point), from which we will draw random samples from, leading to the creation of a simulated observation.	138
C.2	Spectral line taken from the simulation of $N = 10$ spectra using random draws from Gaussian distributions. The dashed black line represents the original spectrum used to construct the line, with the inset region zooming in on the core of one spectral line.	138

Chapter 1

The search for an Earth analog

For the past few decades, humankind has been looking at the cosmos, searching for planets outside of our solar system. The quest for such planets, also known as exoplanets, started in the 1980's with the first tentative detections with the development of the Radial Velocity (RV) method. In 1988 [Campbell et al. \(1988\)](#) uncovered an object with a minimum mass of $1.7 M_{Jup}^*$ orbiting γ Cep with a period of 2.7 years. One year later, [Latham et al. \(1989\)](#) reported a signal in HD114762 with a periodicity of 84 days and corresponding to a $11M_{Jup}$ companion. At the time, these detections were still putative and the community was split between two explanations: low-mass brown dwarfs or exoplanets. In 1992, [Wolszczan & Frail \(1992\)](#) reported two low mass planets (with masses similar to the one of Earth) around a radio pulsar, with this discovery being confirmed two years later ([Wolszczan, 1994](#)). However, it was not until 3 years later that the exoplanets field truly started flourishing. In 1995, the remarkable detection of *51 Pegasi b* ([Mayor & Queloz, 1995](#)) was the first stepping stone that shaped the future of the exoplanetary field. This was the first detection of an exoplanet orbiting a Sun-like star, with an orbital period of 4.2 days and a minimum mass of $0.47M_{Jup}$. In recognition of the importance of this discovery, it was awarded with the Nobel prize in physics in 2019[†]. Only a few years later, the first system with more than 1 exoplanet was found, as [Butler et al. \(1999\)](#) reported 3 Jupiter-mass exoplanets orbiting Upsilon Andromedae. Around the same time, [Henry et al. \(2000\)](#) reported the first transiting exoplanet, *HD 209458b*, with a radius of $1.42R_{Jup}$ and a orbital period of ~ 3.5 days. A combination of this photometric analysis with one performed through radial velocities made this planet the first to have its true mass measured: $0.62M_{Jup}$.

*Jupiter masses

[†]<https://www.nobelprize.org/prizes/physics/2019/summary/>

Since those early discoveries, 5484 exoplanets have been confirmed and more than 2000 remain as candidates*. This extraordinary increase (see Figure 1.1) has been mainly brought by two techniques: the transit and Radial Velocity (RV) methods. The former is based on measuring the decrease of brightness of a star as its disk is crossed by a companion, whilst the later is based on measuring changes on the velocity of a star, consequence of an orbiting companion (see Chapter 1.2 for a detailed description of the RV method). More importantly, not only can we independently detect exoplanets with different methods, but also they allow us to probe different physical properties of the planet and corresponding orbit.

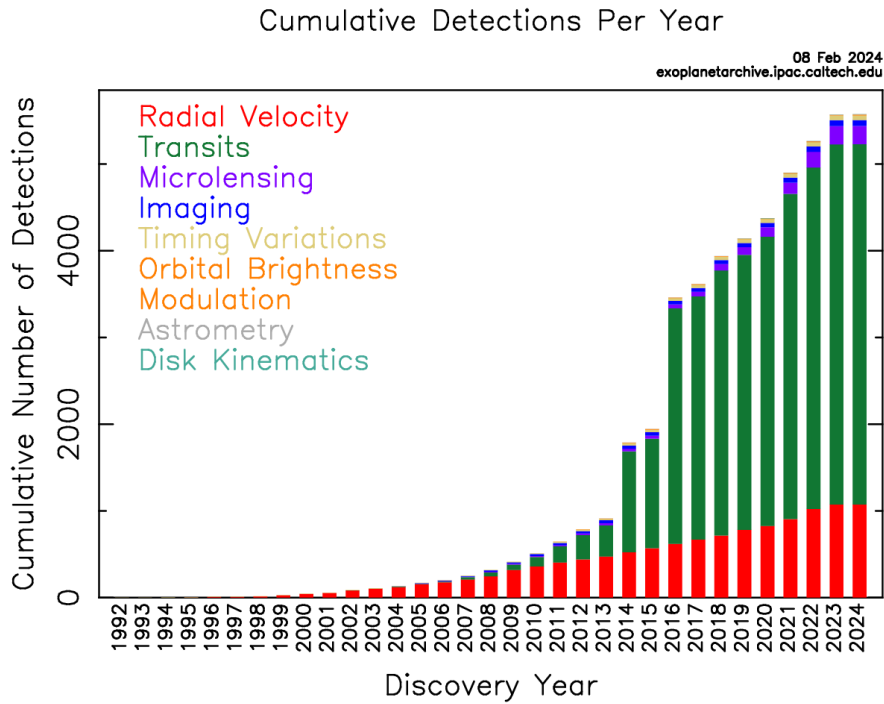


FIGURE 1.1: Yield of confirmed exoplanetary detections over the years, as provided by different methods. Image taken from “NASA’s exoplanet archive” in February 2024.

This increase in the number of discoveries has been a consequence of both advances in instrumentation and the data analysis techniques. Furthermore, and as we can see in Figure 1.2, in recent years we have seen a trend towards the detection of smaller and lighter exoplanets. More remarkably is that we have already managed to find exoplanets smaller and lighter than our own. An example of such planets can be found in Lam et al. (2021), with the discovery of GJ376b, a sub-Earth (radius of $0.71 R_{\oplus}$ and a mass of $0.546 M_{\oplus}$) through TESS transits and HARPS RVs. More recently, Barros et al. (2022) reported a five-planet system, studied with ESPRESSO radial velocities and TESS light-curves, unveiling

*Data taken from *exoplanet.eu*, accessed in August 2023.

a planet with a radius slightly larger than Earth’s and a mass roughly half of the one from Earth. In the same year, [Faria et al. \(2022\)](#) used the ESPRESSO spectrograph to propose a candidate sub-Earth, with a period of 5.1 days and a mass of $0.26M_{\oplus}$ orbiting our closest neighbour, *Proxima Centauri*. As we shall discuss later on, in Chapter 1.2, such discoveries are, currently, only possible in smaller and lighter stars, where these signals are easier to detect with our current approaches.

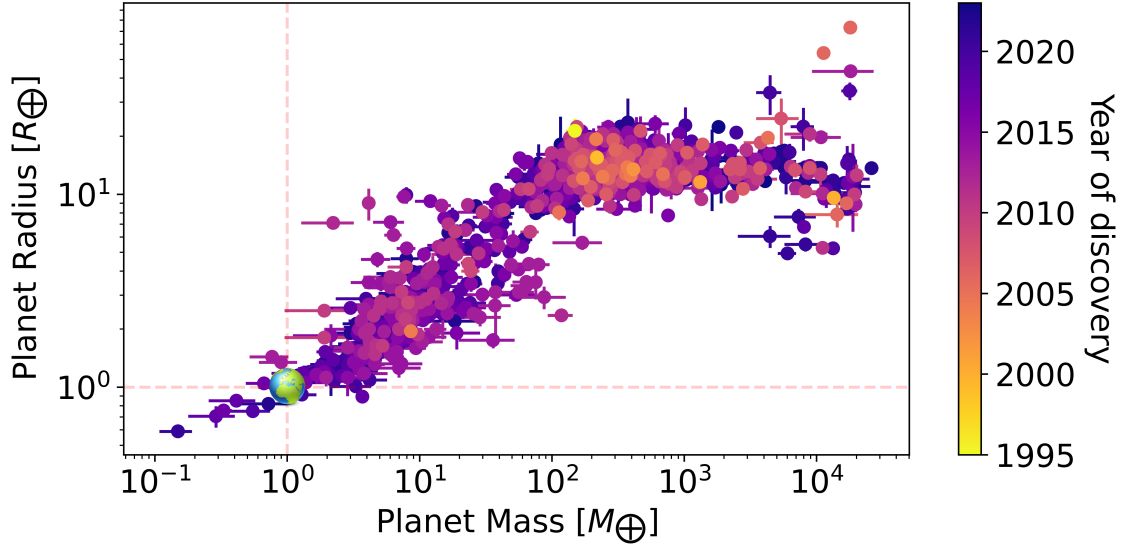


FIGURE 1.2: Mass-radius relation of the confirmed exoplanets, in units of Earth’s mass and radius, alongside their associated uncertainties. Exoplanets for which the uncertainty in either measurement, as reported in *exoplanet.eu*, was higher than 50% were discarded.

The colormap represents the year of discovery, since 51 Peg in 1995 up to 2023.

Despite the impressive results that have been obtained until now, one of the boldest goals of current-day astrophysics is yet to be achieved: the detection and characterization of Earth twins – rocky planets with the physical conditions to hold liquid water on their surface – orbiting a Sun-like star. The discovery that such planets may be very common around solar-type stars, i.e., late F, G and K stars ([Hsu et al., 2019](#), [Mayor et al., 2014](#), [Santos et al., 2004](#)), motivated the development of a new generation of ground and space-based instruments and missions by the main international agencies (e.g., ESA, ESO, NASA), such as HARPS-N ([Cosentino et al., 2012](#)), EXPRES ([Jurgenson et al., 2016](#)), and ESPRESSO ([Pepe et al., 2021](#)). For photometry, Kepler ([Koch et al., 2010](#)) and its successor mission K2 ([Howell et al., 2014](#)), TESS ([Ricker et al., 2014](#)) and CHEOPS ([Benz et al., 2021](#), [Rando et al., 2018](#)), have been the main workhorses with their large scale analysis of stars. More recently, the first light of JWST ([Gardner et al., 2006](#)), opened the door for more precise characterization of exoplanetary atmospheres (e.g., [JWST Transiting](#)

Exoplanet Community Early Release Science Team et al., 2023, Rustamkulov et al., 2023).

On top of the current generation of instruments, the next one is already in preparation, e.g., ANDES (Marconi et al., 2022) in the (under construction) ELT, PLATO (Rauer et al., 2014), ARIEL (Tinetti et al., 2021), among others. Thus, we are currently in the advent of a new era in exoplanetary research, and getting closer and closer to our ambitious goal. In this search, the RV method plays an important role, not only on the detection, but also in the characterization efforts, as it is often used to estimate the mass of exoplanets detected through transits.

In this Chapter we explore the two aforementioned methods: transit photometry and radial velocities, alongside their connections. The core principles, scientific yields, and limitations are discussed, with a greater focus placed on the RV methodology, as it is the main topic of this thesis.

1.1 The transit method

One of the most prolific methods for exoplanet discovery is the transit method. It is based on measuring the decrease of a star's light, introduced by an orbiting body crossing the stellar disk. During the transit event, the companion occults a portion of the stellar disk, creating a dip of the star's light, as depicted in Figure 1.3.

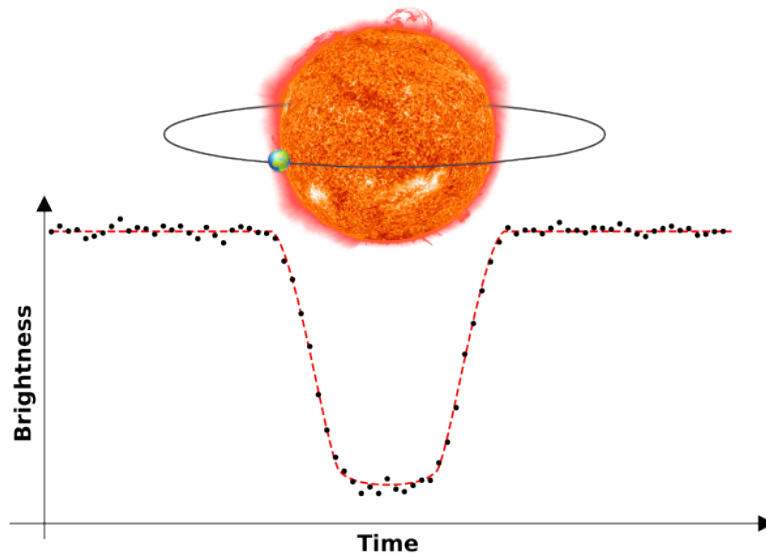


FIGURE 1.3: Representation of the flux decrease due to an orbiting companion.

If we assume a spherical shape for both star and planet, we can estimate the transit depth, ΔF , through:

$$\Delta F = \left(\frac{R_p}{R_s} \right)^2, \quad (1.1)$$

where R_p and R_s are the radius of the planet and star, respectively. From Equation 1.1 it is possible to see that the transiting signal, for a given exoplanet, will be more noticeable when crossing a smaller star.

The *Kepler* and *K2* missions (Howell et al., 2014, Koch et al., 2010) were one of the largest work-horses for transit photometry, providing more than 3000 confirmed exoplanets*. However, the method is prone to false-positives from stellar activity (Simpson et al., 2022) or orbital configurations (Santerne et al., 2013). To mitigate such issues, typically radial velocity measurements are used as a follow-up to validate a given, candidate, exoplanet (e.g., Dalba et al., 2022, Queloz et al., 2009).

1.2 The radial velocity method

The method of Radial Velocities (RV) is one of the most widely used (refer to Figure 1.1), allowing to both detect and characterize exoplanets. It is based on measuring variations of the velocity of a star along our line of sight. The changes in velocity are induced by the gravitational interaction between the star and its orbiting companions. The right side of Figure 1.4 provides a depiction of this interaction.

When a companion is orbiting a star, the latter wobbles around the system's barycenter, due to the gravitational pull of this companion. By measuring the component of that motion along our line of sight, we obtain a measurement of the star's radial velocity (RV).

If we have a photon being emitted by a source, that is moving relative to an observer, it will be detected at a different wavelength than the one it was emitted at. When the source is moving towards the observer, each wave is emitted closer to the observer than the previous one, reducing the time of arrival between waves and, consequently, increasing its frequency. Similarly, if the source is moving away from the observer, we will see an decrease of frequency. The former is called a blue shift, whilst the later is a red shift. The difference in wavelengths, $\Delta\lambda$, can then be written, in its non-relativistic form, as expressed in Eq 1.2.

$$\frac{\Delta\lambda}{\lambda_0} = \frac{V}{c}, \quad (1.2)$$

*Data taken from https://exoplanetarchive.ipac.caltech.edu/docs/counts_detail.html; Accessed: November 2023

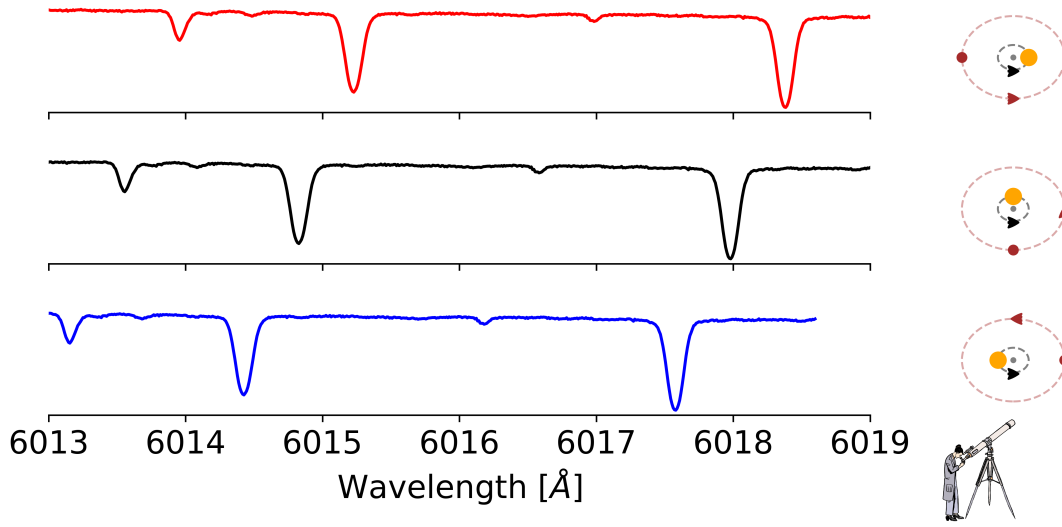


FIGURE 1.4: Schematic of the gravitational interaction between a star and an orbiting companion, showcasing the Doppler shift (left side) of the light that reaches the observer (telescope) as the planet orbits the star (right side). As the star moves in our direction, the light is blue-shifted, whilst it is red-shifted as the star moves in the opposite direction.

where V is the relative velocity between the source and the observer, c is the speed of light and λ_0 is the original wavelength of the emitted photon. If we now apply this concept to a spectra from a star, while the star (source) is moving away from Earth (observer) the spectral lines will be shifted to longer wavelengths (red shift) and, when it is moving toward us, the spectral lines are shifted to shorter wavelengths (blue shift). From here we see that by measuring the displacement of the stellar spectral lines, we will be able to derive the velocity of the star (along our line of sight).

Simultaneously to the observations of a given star, it is imperious that we also consider that Earth is moving through space. If this is not accounted for, the measured RV will be a combination of the stellar RV and the Δ RV introduced by Earth's motion and rotation. Over the year, this motion introduces a maximum effect of approximately $\pm 30 \text{ km s}^{-1}$, significantly larger than the typical signals from exoplanets. Furthermore, Earth's rotation also introduces an RV effect at the level of 600 m s^{-1} over 8 hours, making its correction of paramount importance. To remove this effect from our spectra, the wavelengths are converted from the spectrograph's reference frame (in Earth) to an inertial one, which is chosen to be the barycenter (center of mass) of the solar system. This correction, commonly named *BERV* correction, can be calculated with a precision at the level of 1 cm s^{-1} (Wright & Eastman, 2014).

1.2.1 Planetary parameters from RV observations

Now that we have a framework that allows us to convert the displacement of stellar lines into the RV of the star, we must think about the information that we can extract from its evolution over time.

In order to learn about a planetary system, we can solve the two-body problem schematized in Figure 1.4, whilst considering *Newton* laws. This is quite a lengthy derivation and, consequently, it is omitted here. For a full derivation, refer to Chapter 2.1 of [Faria, João Pedro de Sousa \(2018\)](#). Furthermore, we refer the reader to either [Murray & Correia \(2010\)](#), for an overview of the 3-dimensional orbital mechanics, or to Appendix A for a short summary. This process allows us to find an equation that describes the semi-amplitude, K , and the time evolution, $v(t)$, of the RV signal (Equation 1.3).

$$K \approx \left(\frac{2\pi G}{P} \right)^{\frac{1}{3}} \frac{M_{\text{planet}} \sin i}{(M_*)^{2/3} \sqrt{1-e^2}} \quad (1.3)$$

$$v(t) = K[\cos(\omega + f) + e \cos(\omega)],$$

where P is the orbital period of the exoplanet, f is the true anomaly, G is the gravitational constant, M_* is the mass of the star, M_{planet} is the mass of the planet, i is the inclination, e the eccentricity, and ω the longitude of the periastron. An analysis of a RV time-series, consisting in fitting Equation 1.3 to the measured RVs, lets us infer both the orbital period and minimum mass of the planet.

Considering the schematic of the orbit, as presented in Appendix A, it is also possible to infer that its configuration will have large consequences on the RV signal that we measure. First, if the inclination takes values of either 0 or 180 degrees, the planet is moving in a plane perpendicular to our line of sight. This leads to no motion in the radial axis and, consequently, no RV signal. Second, if we don't have a previous measurement of the inclination, the RV method will not be able to measure the true mass of the planet. Instead, we are only able to estimate $M_{\text{planet}} \sin(i)$, i.e., its minimum mass.

Going back to Equation 1.3, it is clear that the RV signatures will have higher amplitudes if we analyse a less massive star and/or are in the presence of a more massive planet. Independently of the case, detecting exoplanets is a tough challenge. To provide context let us consider a star with the same mass as the Sun. In this case, the RV imprint of a planet with mass and period equal to those of Jupiter is $\sim 12.5 \text{ m s}^{-1}$, whilst an Earth-like one would only produce a $\sim 9 \text{ cm s}^{-1}$ signal. If we translate, through Equation

1.2, the RV signal to the displacement of spectral lines, we find that at a wavelength of 500 nm they correspond to a $\Delta\lambda$ of 2.08×10^{-5} and 1.67×10^{-7} , respectively. From here it is already evident that our goal of detecting Earth-like planets is an extremely daunting one. Furthermore, it is only possible to reach such precisions by leveraging the statistical power of combining the information from thousands of spectral lines.

Despite the challenges imposed by the need of measuring such small line displacements, we have already managed to detect signals below the 1 m s^{-1} barrier (Barros et al., 2022, Demangeon et al., 2021, Faria et al., 2022). In fact, advances in instrumentation made it so that the (current) limiting factor is no longer the instrumental precision but, instead, other sources of RV variability. To achieve sensibility to such small signals there is a need to both account for those extra sources of biases (see Chapter 1.2.2) but also the pressing need for the development of novel methodologies to analyse the data (see Chapter 1.2.4).

In order to fully characterize an exoplanet, we can combine the RV method, which provides the minimum mass, with the transit method (Chapter 1.1), which provides its radius. With this, we can thus study its mean density and allow for a study of its interior composition (e.g., Barros et al., 2022, Morello et al., 2023, Motalebi et al., 2015) and formation history.

1.2.2 Challenges for optimal Radial Velocity extraction

As we have seen until now, we have taken strides towards the detection and characterization of low mass planets, with advances in both instrumentation and new data analysis algorithms. However, our techniques are not perfect and we are searching for minute signals hidden in a sea of noise and contaminating effects. In this Chapter we explore the current limitations and the challenges that must be surpassed for the extraction of precise RVs.

1.2.2.1 Earth's atmosphere

High-resolution spectrographs are ground-based instruments, consequence of the large complexity and high monetary investments needed for space-based ones. When observing from Earth we find, as one would expect, that the atmosphere's signature will be imprinted in the spectra, as it absorbs the stellar radiation and creates absorption features, commonly known as telluric lines. From Cunha et al. (2014) and Figueira et al. (2012) we know that this effect strongly depends on wavelength, airmass and weather conditions during

the observations. More worryingly is the fact that the shallower features, also referred to as *micro-telluric* can impact our Radial Velocities measurements at the 10 cm s^{-1} level, which is on par with the signal introduced by a Earth-like planet orbiting a Sun-like star.

In Earth’s referential, the position of the telluric features is constant through time, as the atmosphere is always in the same location, relative to the spectrograph. However, their depth is not constant, as it will depend on the weather conditions and position of our target on the sky. Changes in the concentration of different molecules in the atmosphere will be reflected on the depth of associated features. On top of that, the lower the star is on the sky, the larger the optical path of the light and the deeper the telluric features. Furthermore, and even though the position is fixed on the referential of the spectrograph, we must recall that we change our frame of reference with the barycentric correction. This means that after this correction, the telluric lines will appear Doppler-shifted by the BERV value associated with the observation. In other words, if we observe the same star during one year, the tellurics will fall in a window of $\pm 30 \text{ km s}^{-1}$ (the yearly BERV variation). This effect can be seen in Figure 1.5, where the BERV correction shifts the location of the telluric features (top panel). On the other hand, the observations on the spectrograph referential (bottom panel) are aligned. As, in this case, the stellar RV variability is much smaller than the BERV variation, it is also possible to see their large displacement if the BERV correction is not applied (bottom panel of the image). If such features are not accounted for, they will bias the RV estimation, as their variability is, typically, much larger than the stellar one.

If we now focus on the optical wavelengths, i.e., those that are covered by the spectrographs used in this thesis (Section 1.2.3), Fig. 1.6 shows that this contamination can reach saturation in some regions. In the optical domain, the majority of the absorption features are introduced by water and oxygen. On top of that, it is even clearer that the atmosphere poses a more significant challenge in the NIR regions, e.g., as it is the case for the NIRPS and SPIROU spectrographs.

To account for this source of contamination on our data we can either reject the wavelength regions that are affected by them, or attempt to correct them through models or data-driven approaches:

1. The conservative approach to handle telluric contamination is to create a binary mask from a transmittance spectra, using it to discard all wavelengths that fall within the selected windows. Usually, such masks are constructed by evaluating which

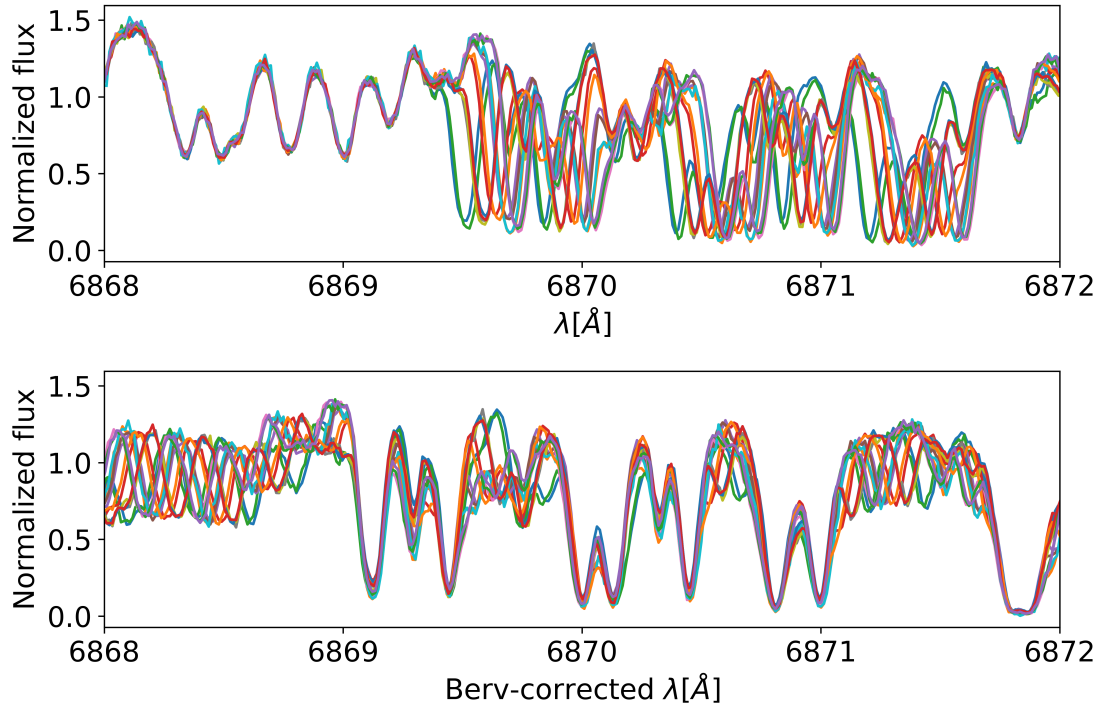


FIGURE 1.5: Doppler-shift of the telluric features, as caused by the BERV correction of different observations, as represented in the different colours. The presented observations were all collected with BERV in the interval $[16.9, 26.7] \text{ km s}^{-1}$, with a span of $\sim 9 \text{ km s}^{-1}$. **Top:** BERV-corrected spectra; **Bottom:** Spectra without the BERV correction.

wavelength regions have a drop in the transmittance larger than a given threshold. The obvious drawback of this approach is the large loss of RV content, especially in the more affected nights. However, it will not introduce any systematics from a possible under or over correction of the stellar spectra.

2. Correction of telluric features: The traditional approach for the correction of telluric features was to divide the spectrum by one of a standard star, selected due to their lower number of stellar lines, (e.g., [Vacca et al., 2003](#)). Standards are typically fast rotating stars which have the desired lower amount of stellar lines, e.g., A-type stars have a small amount of shallow lines and their continuum level can be approximated with a blackbody. This approach inherently leads to a loss of telescope time, a costly trade-off, as it must be spent observing the standard star to build our model. On top of that, the telluric contamination of the standard star is not exactly the same as the one on the target, as the light's path in the atmosphere is different in both cases. The time-evolving nature of the telluric features also mean that this correction model will never be perfect, as long as we don't have simultaneous observations of the two

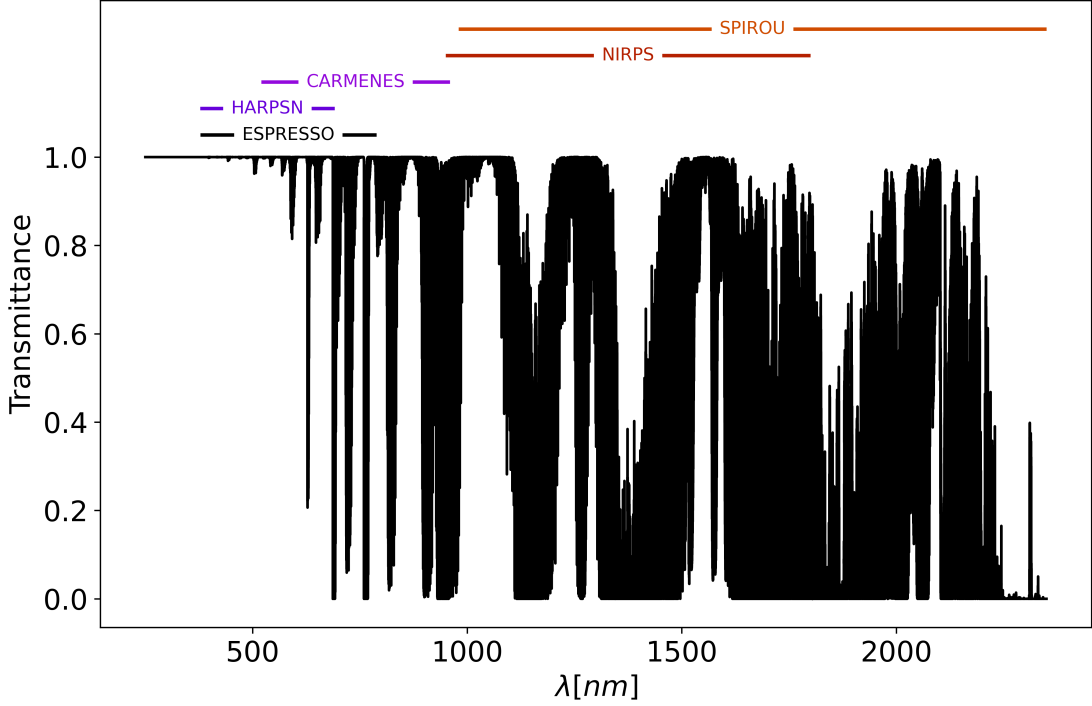


FIGURE 1.6: Telluric transmittance spectrum for a 25% humidity and a airmass of 1.5, covering the optical and the near infra-red region, highlighting the wavelength coverage of different spectrographs.

stars. Alternatives to a standard star are theoretical models, e.g., TelFit (Gullikson et al., 2014), Molecfit (Smette et al., 2015) and, TAPAS (Bertaux et al., 2014), which yield a superior performance for the correction of H_2O lines, but are worse for the O_2 lines (Ulmer-Moll et al., 2019). In the same vein, Allart et al. (2022) uses theoretical models to fit a subset of the water lines in the atmosphere, determining the water content, pressure and temperature that best fit the data. This information is then used to construct a transmittance profile for all water lines and, afterwards, correct the stellar spectra.

3. More recently, data-driven techniques have been developed, attempting to disentangle the telluric spectrum from the stellar spectrum, e.g., as seen in wobble (Bedell et al., 2019) or PCA-based approaches (Artigau et al., 2014). However, for optimal performance they require a large number of observations, spread throughout a large timespan, so that the BERV-induced shift of telluric features is large enough to be captured by the algorithms.

Despite the large number of available approaches their usage will either lead to a significant loss of information or modify the data in an unaccounted manner. First, the available tools for the correction of telluric features often fail to provide a meaningful uncertainty propagation to their estimates for the telluric correction. Second, the rejection of spectral regions can also lead to a large loss of RV content, as seen in Fig 1.7 for a single observation at optical wavelengths.

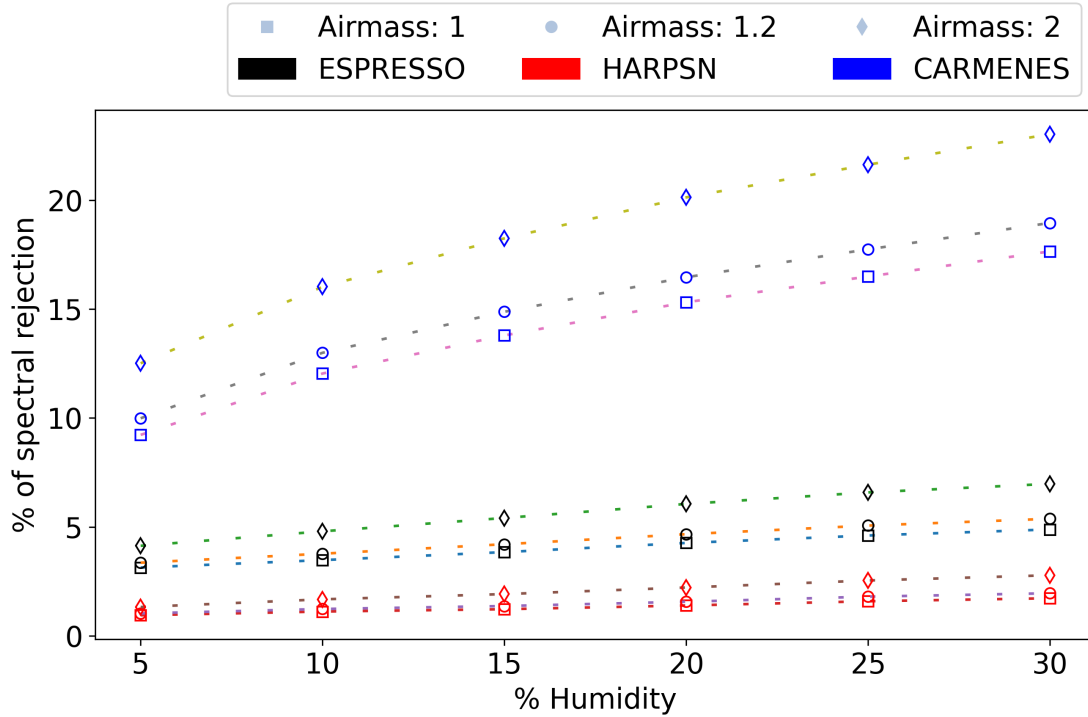


FIGURE 1.7: Percentage of the wavelength coverage of optical regions from different spectrographs, whilst assuming a threshold of 1% drop on transmittance. The transmittance spectrum was generated with TelFit, assuming different airmasses and water content in the atmosphere for ESPRESSO (black), HARPS-N (red) and the optical arm of CARMENES (blue).

If we desire to construct a mask that can be applied to multiple observations, one must consider the maximum BERV span over the year, vastly worsening the rejection of stellar spectra. Furthermore, the option of rejecting telluric features is only viable at optical wavelengths, as the cost of rejecting tellurics based on transmittance would be unbearable in the near infra-red regions.

1.2.2.2 Stellar activity

Unfortunately for our efforts of exoplanet-hunting, stars are neither time-invariant nor uniform. Their multiple processes (e.g., star spots, granulation, plagues, pulsations, magnetic

cycles) will introduce RV jitter that can either hinder our analysis (Affer et al., 2019, Baragán et al., 2023, Faria et al., 2020, Saar & Donahue, 1997, Santos et al., 2000) or, in some cases, even mimic planetary signals (e.g., Forveille et al., 2009, Huelamo et al., 2008, Santos et al., 2014, Simpson et al., 2022).

The issue of stellar activity is exacerbated when we attempt to detect and characterize the small signals that one would expect from Earth-like planets orbiting Sun-like stars. Figure 1.8 shows a comparison of Earth’s RV signal (9 cm s^{-1} signal) with disk-integrated RVs of the Sun (after correcting for the effects of the objects on the Solar System). The Sun-as-a-star RVs were collected over 3 years with the solar telescope (Collier Cameron et al., 2019, Dumusque et al., 2021) connected to HARPS-N.

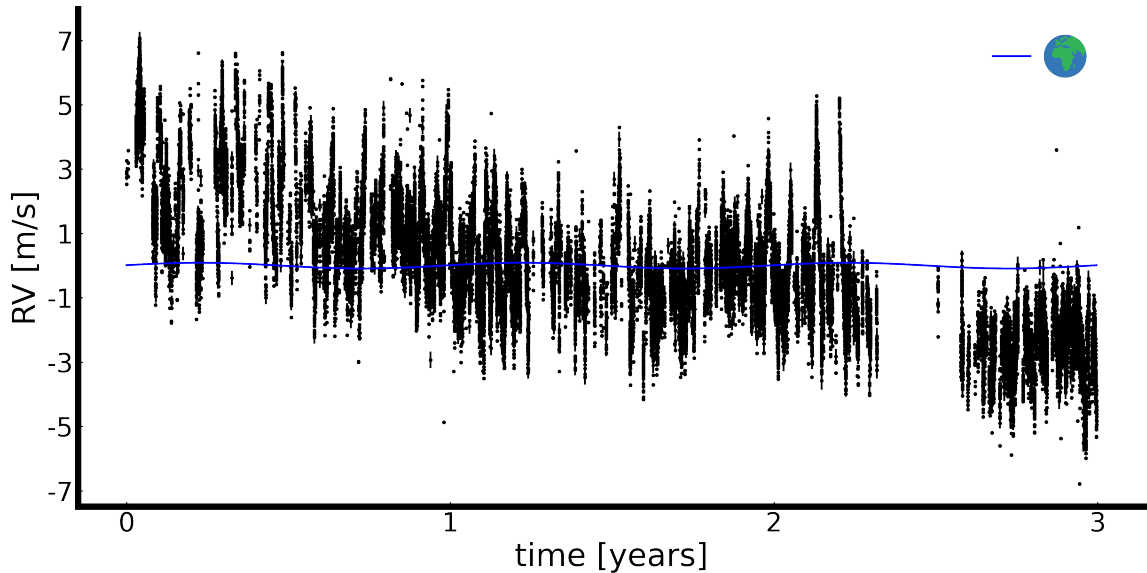


FIGURE 1.8: Radial velocity time-series from Earth (blue) compared to solar RVs from the HARPS-N solar telescope (black).

From here, we clearly see that the signal of our planet is completely drowned out by the amplitude of the solar variability.

Even though we are not yet privy to the internal physics of stars, there are some effects for which we know the timescales and expected amplitudes:

- Stellar oscillations, consequence of interference between multiple p-mode oscillations, can introduce effects at the tens of centimeter per second, at a timescale of a few minutes (see Chapter 1.5.1 of Schrijver & Zwaan, 2000).
- Granulation effects, due to the convective effects found in stellar surfaces, albeit greatly inhibited in the active regions. The timescales for this effect can vary between

minutes and, in the case of super-granulation, reach up to 30 hours (refer to Chapter 1.5.2 of [Schrijver & Zwaan, 2000](#)). The amplitude of such signals is comparable to those from the p-mode oscillations.

- Active regions on the surface of the stars (spots and plages) move across the solar disk as the star rotates, introducing signals with periodicity close to the rotation period of the star. On one hand, the main effect from the hotter plages comes from the suppression of convection in the active regions, introducing an effect between 40 and 60 cm s^{-1} on the RVs ([Meunier et al., 2010](#)), depending on the activity level. On the other hand, the cooler (than the stellar mean temperature) stellar spots will emit less light than the other regions of the star, introducing a distortion on the shape of the lines and an impact on the disk-averaged RV ([Aigrain et al., 2012](#), [Santos, 2008](#)). On the Sun, the effect introduced by spots is slightly smaller than the one introduced by the brighter regions ([Meunier et al., 2010](#)).
- Magnetic activity cycles are a multi-year effect that depends on both the activity level of the star and the number of magnetic regions. These long term effects can introduce RV perturbations at a level of a few meter per second ([Gomes Da Silva et al., 2012](#), and references therein).

Overcoming the RV jitter introduced by the stars is thus of paramount importance, even more so for the smaller planets that we wish to study. The variability present at the smaller timescales (those introduced by oscillations and granulation) can be mitigated through the choice of the observing strategy ([Chaplin et al., 2019](#), [Dumusque, Xavier, 2012](#)), whilst the others must be mitigated through a modelling effort. Commonly, the contribution from stellar activity is modelled through a combination of simultaneous photometry and RV observations, allowing the disentanglement of the two types of signals. In [Aigrain et al. \(2012\)](#) the authors present a method to predict the RV impact from the photometric observations, whilst [Rajpaul et al. \(2015\)](#) presented a Gaussian Process framework to model stellar activity from the RV data and activity indicators, without including photometry data. Activity indicators are proxies for the stellar activity, helping in the disentanglement between stellar and keplerian signals. The most widely used either track changes to the mean spectral line*, e.g., [Lafarga et al. \(2021\)](#), [Lanza et al. \(2018\)](#), [Siegel et al. \(2022\)](#), or measure the emission of certain spectral lines known to be sensitive to activity, e.g., the

*refer to the description of the CCF method in Chapter 1.2.4

$H\alpha$ line and Ca II H&K lines (e.g., [Boisse et al., 2009](#), [da Silva et al., 2022](#), [Schöfer et al., 2019](#)).

Furthermore, RV extraction on a line-by-line basis [Cretignier et al. \(2020a\)](#), [Dumusque \(2018\)](#), [Moulla et al. \(2023\)](#) shows that some spectral lines are more sensitive to such effects. Through a careful selection of the lines to consider, it is then possible to decrease the impact of stellar activity by a factor of ~ 2 . In the same line of thought, some methodologies attempt to directly disentangle RV shifts introduced by changes in line position and line changes. One example of such is [Collier Cameron et al. \(2021\)](#), through a principal-component analysis of the auto-correlation of the CCF profile, allowing to directly retrieve RV time-series for each. The authors showcase the retrieval of keplerian signals with semiamplitudes of 40 cm s^{-1} , injected into solar RVs. In a similar fashion, [Binnenfeld et al. \(2022\)](#) constructs two independent periodograms, allowing to simultaneously retrieve the periodicity of activity and keplerian signals. However it does not provide a framework to retrieve the signals associated with each component. This shows that not only can we tackle these effects at the modelling stage, but also directly during the RV extraction.

Lastly, and as previously described, the design of the observation strategy helps diminish this impact. On one hand, increasing the exposure time of the observations allows to mitigate the small time-scale effects. However, over a period of observations, we will observe the star in different stages of its activity cycle, raising the need to understand the importance of the cadence chosen for the observations. Over the past few years we have seen the advent of different approaches to tackle this issue of the optimal scheduling of observations: ([Sairam & Triaud, 2022](#)) used HARPS-N solar data to predict the timing of solar minima which, if used to construct the cadence of observations, improves the sensitivity to planets. Similarly, [Balsalobre-Ruza et al. \(2022\)](#) implements a Bayesian strategy to decide the next observing date, providing an optimal observation strategy to minimize the number of visits needed to confirm a planetary signal. More recently, [Gupta & Bedell \(2023\)](#) simulated Extreme Precision RV (EPRV) surveys, providing an evaluation of the impact of the selected observation strategy on the yield of detections from the entire sample.

1.2.3 High-precision spectrographs for exoplanet detection

The main goal of this PhD thesis is the development of methodologies that facilitate the extraction of precise RVs. Thus, there is a need for observational data, so that our approaches can be validated. The bulk of the work developed in this thesis was focused in four different, state-of-the-art, instruments: ESPRESSO (Pepe et al., 2021), HARPS (Mayor et al., 2003, Pepe et al., 2002b), HARPS-N (Cosentino et al., 2012), and the optical arm of CARMENES (Quirrenbach et al., 2018). Within this Chapter, a brief introduction is provided to each instrument, alongside a summary of their characteristics in Table 1.1.

TABLE 1.1: Wavelength coverage and expected precision of the spectrographs used in this thesis. For CARMENES we only report information from the optical arm.

Instrument	Wavelength coverage [nm]	Expected precision
HARPS-N	[390, 690]	60 cm s ⁻¹
HARPS	[378, 691]	100 m s ⁻¹
CARMENES (optical arm)	[520, 960]	~ 150 cm s ⁻¹
ESPRESSO	[378.2, 788.7]	10 cm s ⁻¹

For each instrument, we use the data products that are provided by their official Data Reduction Software (DRS), which typically provides a wavelength solution, extracted fluxes and associated uncertainties, after applying the relevant corrections. Furthermore, it also provides access to other products, such as telemetry of the telescope, atmospheric conditions, and quality of the data. The stellar spectra is typically provided in two different formats: in a order by order basis (henceforth *S2D* spectra) or in a 1D format (*S1D* spectra), constructed by merging the order-wise information of the former data product. For this PhD thesis, we focus on the *S2D* data, as it avoids manipulation of the raw data that comes from the instrument.

HARPS

HARPS (Mayor et al., 2003, Pepe et al., 2002b), the “High Accuracy Radial velocity Planet Searcher” is an high resolution spectrograph installed in the 3.6 m telescope of the European Southern Observatory (ESO), in La Silla. It works on optical wavelengths, covering the interval [378, 691] nm with an optical resolution of 115000. It was built to achieve a very accurate RV accuracy ($\sim 1 \text{ m s}^{-1}$) over multiple years. The light that enters the instrument is captured by a mosaic of 2 CCDs which provide a 4k by 4k resolution.

In total we retrieve 72 spectral orders, each spanning 4096 pixels. Due to the geometry of the detectors, there is a loss of light in the interval [530 to 533] nm, as it falls in a gap between the two chips.

HARPS-N

HARPS-N, the “High Accuracy Radial velocity Planet Searcher for the Northern hemisphere” is an high resolution spectrograph in the Telescopio Nazionale Galileo, in the Canary Islands, Spain. This instrument was constructed to the same specifications of the *HARPS* (Pepe et al., 2002b) spectrograph. It works in visible wavelengths, covering the wavelength range of [390, 690] nm with an average resolution of 115000 (Mayor et al., 2003). The light that reaches the two detectors (CCDs) is divided into 69 spectral orders, each with 4096 pixels. To keep it stable and avoid drifts induced by air pressure and temperature changes, it is kept in a vacuum chamber. We refer to (Cosentino et al., 2012) and references therein for more details on the spectrograph and its performance.

CARMENES

CARMENES, the “Calar Alto high-Resolution search for M dwarfs with Exoearths with Near-infrared and optical Échelle Spectrographs” is an high-resolution spectrograph installed in the Calar Alto Observatory, covering optical and Near-Infrared (NIR) wavelengths. The optical arm covers the wavelength interval [520, 960] nm with a resolution of 94600, whilst the NIR covers [960, 1710] nm with a resolution of 80400 (Quirrenbach et al., 2018). In the context of this thesis, we use CARMENES data from the KOBE survey (Lillo-Box et al., 2022), which is only making use of the optical arm. In the optical region, the expected RV precision lies between 1 and 2 m s⁻¹. Due to instrumental effects and calibration issues, the RVs from this spectrograph need to be corrected from the NZP, i.e., the instrumental zero-point (see Appendix B for details), computed with the observation of multiple standard stars.

The optical design of the instrument leads to the creation of 61, independent, spectral orders on the detector, each with 4096 pixels. For more details on the optical design we refer to (Seifert et al., 2016).

ESPRESSO

ESPRESSO, the “Échelle spectrograph for rocky exoplanets and stable spectroscopic observations” is an high-resolution spectrograph installed in ESO’s Very Large Telescope in Cerro Paranal, Chile. It boasts the ambitious goal of reaching a RV precision of 10 cm s^{-1} . This instrument can combine light from up to 4 different Unit Telescopes (UT) of the Very Large Telescope (VLT), achieving the light collecting power of a 16 meter telescope. It covers wavelengths from 380 up to 800 nm, with a resolving power (R) of either 60000, 140000 or 225000, depending on the observation mode (Pepe et al., 2010).

Inside the instrument, the light is separated by a beam splitter, into a blue and red arm where, at its end, is captured by two 90×90 mm CCDs (Pepe et al., 2010). Before reaching the camera, the light passes through a cross-disperser element, separating the light into the different orders (Figure 2 from (Pepe et al., 2010) provides a schematic of the spectral format of each arm). In total, the instrument has 85 spectral orders, with each being duplicated on the detector, leading to a grand total of 170 slices, each with a size of 9111 pixels. For simplicity sake, we refer to those 170 slices as orders, since they are independent measurements from different pixels in the detector. Furthermore, the orders that lie near the upper and lower borders of the CCDs do not manage to fit the entire PSF within the camera. In practice, this means that for orders near 0 and 170 we have missing data at their respective start and end. For more details on the optical layout and design of the instrument we refer to Spanò et al. (2010).

Lastly, due to technical interventions on the instrument, ESPRESSO data is typically divided into three different subsets, ESPRESSO18, ESPRESSO19 and ESPRESSO21, considered to be independent due to changes on the instrumental profile. The first subset, ESPRESSO18, is considered to cover the time period between the first light of the instrument and the upgrade of ESPRESSO fiber link, in 2019 (Pepe et al., 2021). The next subset, ESPRESSO19, covers the data taken between this intervention and the post-COVID ramp-up of the instrument. After this period, the observations are considered to be from ESPRESSO21.

1.2.4 Different approaches for RV extraction

As described until now, during the past few years we have been witnessing the improvement of both instrumentation and the data analysis techniques. In this Chapter, a brief

description will be given of the most widely used methodologies for the extraction of precise radial velocities.

The Cross Correlation Function (CCF)

The initial, major, breakthroughs of the RV methodology were obtained through the Cross Correlation function (CCF) method. In its early stages (Baranne et al., 1996), a binary mask, with non-zero weights attributed to the expected positions of stellar absorption lines was cross-correlated with the spectra, as in Equation 1.4.

$$CCF(v_R) = \int S(\lambda)M(\lambda_{v_R})d\lambda \quad (1.4)$$

where $CCF(v_R)$ is the CCF value at a given Radial Velocity, v_R , S is the spectra and $M(\lambda_{v_R})$ is the Doppler shifted mask. After computing the CCF profile for a given RV interval, a Gaussian fit is made. It's minimum point provides the radial velocity measurement of the data.

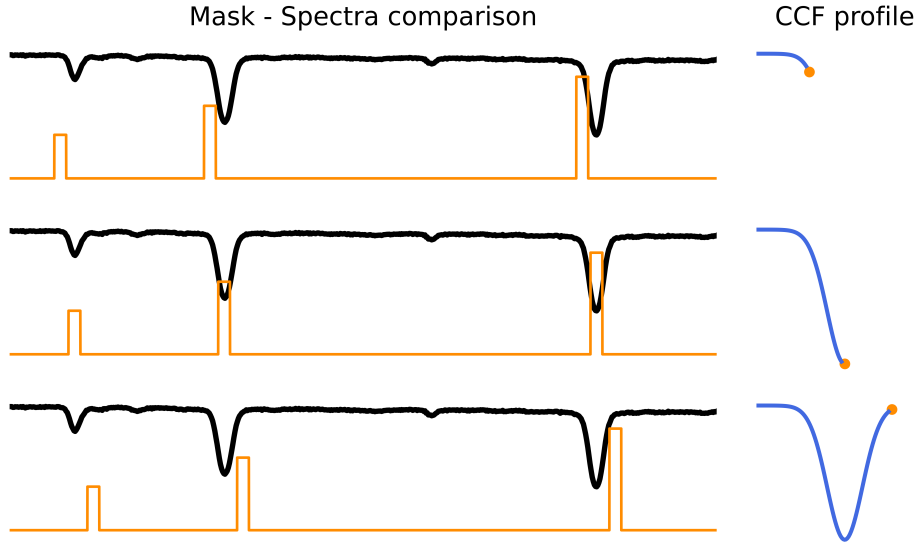


FIGURE 1.9: Schematic of the CCF method, with a weighted mask. **Left:** Cross correlation with the weighted mask. We can see that deep thin lines have larger weights than shallower ones. **Right:** Final CCF profile, resulting from the cross correlation.

The fact that that deep and sharp lines contain more information than broad and shallower ones (Bouchy et al., 2001) allowed the improvement of the method, by attributing different weights to different lines, depending on the RV information present in them. This way, deep lines contribute more to the final CCF profile than shallow ones. In Fig. 1.9 we can find a schematic of the CCF method, using a weighted mask. The weighted masks are

built from either high signal-to-noise spectra, or spectral templates (Baranne et al., 1996, Pepe et al., 2002a, Rainer et al., 2020a).

From the final CCF profile it is also possible to derive metrics that are correlated with the variability of the star, i.e., proxies for stellar activity. As the CCF profile is taken to be the mean spectral line, this is typically done by tracking changes to its shape and symmetry. This is usually done through the asymmetry (bisector span), the central depth (contrast) and the full width half maximum (FWHM) of the CCF profile (e.g., Anglada-Escudé et al., 2016, Lafarga et al., 2020, Lanza et al., 2019, Queloz et al., 2001). Those proxies can then be modelled, simultaneously with the RVs, allowing the disentanglement of keplerian- and activity-induced signals, as previously discussed in Chapter 1.2.2.2.

Despite the impressive results that have already been obtained with the CCF, (e.g., Barros et al., 2022, Castro-González et al., 2023, Santerne et al., 2018), it is not always the optimal approach, consequence of the limitations introduced by the mask and the stellar spectrum. As only a portion of the spectra is used for the RV estimate, there is a loss of RV information from the discarded regions. Furthermore, the masks are typically pre-defined for a given spectral-type, i.e., within the same spectral type, all stars use the same CCF mask. Consequently, the mask will never be a perfect match with the spectra with which it is compared to. On top of that, since the position and depths of the stellar lines are derived from a spectra (either synthetic or from a real observation), their position is not precisely known. Thus, there is an error associated with the stellar masks that is neither propagated, nor accounted for. Lastly, RV extraction can be a challenging task in some situations, such as for M dwarfs. The high number of stellar lines, due to their lower temperatures, results in most of them being blended, impacting the final CCF profile. For such cases, it has been shown that template matching can surpass the CCF method (e.g., Anglada-Escudé & Butler, 2012, Lafarga et al., 2020, Zechmeister et al., 2018). However, current approaches for the construction of the CCF mask has been revealing an improvement for those stars (e.g., Lafarga et al., 2020, 2023, Rainer et al., 2020b).

Template matching

The basis behind template matching is the comparison of the individual observations of a star, with an high-SNR template built from the available observations of the selected star. This means that this data-driven method ensures that the model is as close as possible to the data and maximizes the RV content that we use to inform our extraction. By

aligning the model (template) and data (observations) is thus possible to compute RVs relative to the “zero”-point of the stellar template. The template is created by stacking multiple observations together, usually after being interpolated to a common wavelength solution. Then, it is shifted with different RV values, to find the Doppler shift that better aligns it with the individual spectra, as depicted in Figure 1.10. Typically, the alignment is performed with least-squares fitting for each individual order and, afterwards, the order-wise measurements are combined to build a final RV estimate, as seen in HARPS-TERRA (Anglada-Escudé & Butler, 2012), SERVAL (Zechmeister et al., 2018), NAIRA (Astudillo-Defru, 2015).

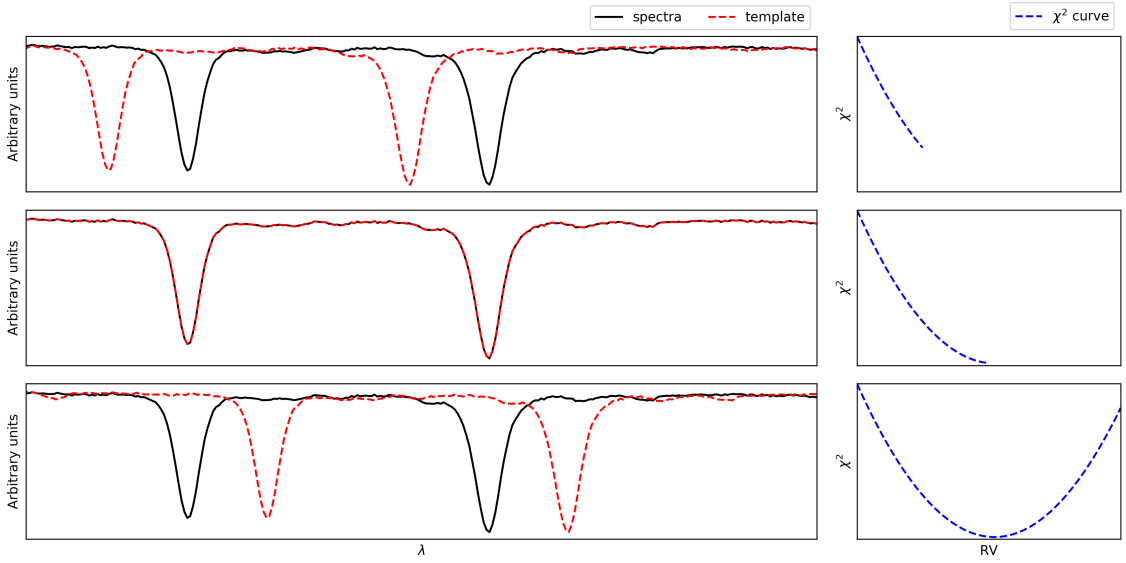


FIGURE 1.10: Schematic of the template matching process, for three different Doppler shifts of the stellar template. **Left:** Alignment between the stellar spectrum and template with different RV shifts. **Right:** Evolution of the χ^2 curve for the different RV shifts.

When the template and spectrum are aligned the curve reaches its minimum value.

As previously discussed, the process of creating a stellar template for the M-type stars is much easier than creating a good CCF mask, consequence of the high number of spectral lines. Thus, a comparison of the two models (CCF mask and template) reveals the reason for such a better performance of template matching in this spectral type. This comparison is shown in Figure 1.11, where it is possible to see the large difference in spectral regions that are included in the stellar template but not in the CCF mask. Thus, from here, we expect to at least find a large gain in the RV precision of the measurements when using the former.

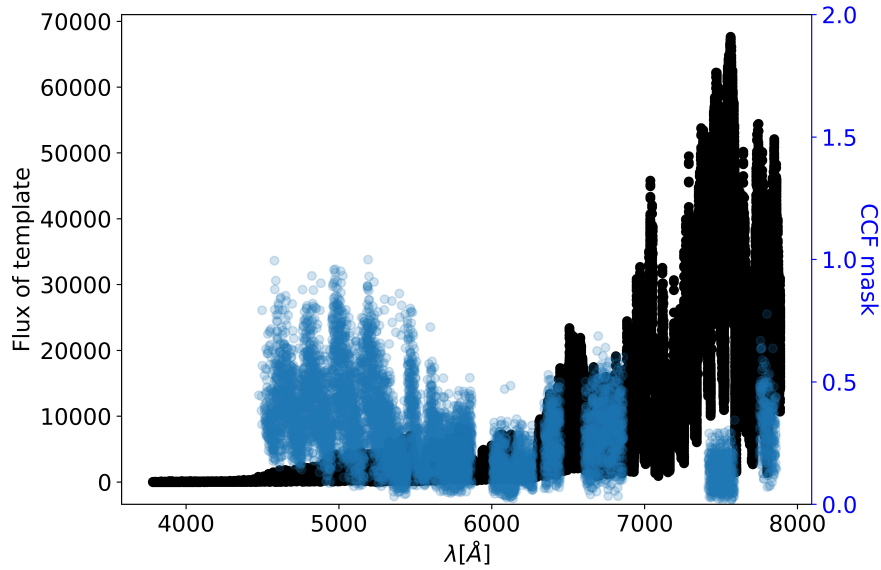


FIGURE 1.11: Comparison of the M5 CCF mask from ESPRESSO (blue) with a stellar template (black) constructed for a M-type star that uses this mask.

The usage of an high-SNR template comes with a few caveats, both in its construction and usage. First, during the creation of the template the RV contributions from the star are removed from the data, which are not known *a priori*. Consequently, the quality of the template will depend on the initial assumptions for the RV, which is typically assumed to be the estimate obtained with a CCF approach. Second, the interpolation routine used to combine all observations in a common wavelength, e.g., cubic spline interpolation in NAIRA, B-spline regression in SERVAL, will also have an impact on the final result, as it introduces assumptions into the underlying shape of the stellar lines. Lastly, the stellar template is assumed to be invariant with time, i.e., each and any observation is assumed to be perfectly reproduced by the stellar template, ignoring the impact of stellar activity. All of those caveats will introduce minute differences in the stellar model, impacting the resulting RVs at an unknown level. One attempt at escaping the interpolation of data can be found in [Holzer et al. \(2021\)](#), where the RVs are calculated through a least-squares fit to the residuals between the stellar spectra and a (fixed) stellar template. The residuals are modelled as a sum of Hermite-Gaussians functions, whose coefficients can be directly mapped to a RV shift. However, this approximation hinges on the fact that the stellar lines have a gaussian shape and that the RV shift is smaller than 500 m s^{-1} .

It is also important to note that the main advantage of template matching also introduces extra sources of variability. By using the majority of the spectral lines, even those

that are more sensitive to stellar variability, it can bias the final RVs. In an attempt to mitigate those issues, we have recently seen the rise of line-by-line (LBL) approaches which, as the name suggests, apply the template matching algorithm at the level of the spectral line, computing individual RVs for each. [Dumusque \(2018\)](#), [Cretignier et al. \(2020a\)](#) often assume that the Doppler shift is smaller than the sampling, allowing to measure the RV shift from changes in flux and the slope of either the template or spectra, expanding on the framework of [Bouchy et al. \(2001\)](#). This individual analysis then allows for the estimation of RVs using a curated set of lines, e.g., selected on their RV scatter, which allows to discard the lines that should be either i) more affected by spectral variability; or ii) micro-telluric features leftover in the spectrum; Not only that, but it also allows to divide the spectral lines into different subsets, based on their formation temperature (e.g., [Moulla et al., 2023](#)). Despite the impressive results, most LBL approaches still lack a consistent process for the selection of which lines to retain in the overall set, with [Artigau et al. \(2022\)](#) implementing a mixture model for this selection process.

Modelling-based approaches

Over the past few years, we have seen a move towards a more general model of the stellar spectra during the process of RV extraction. One such example is WOBBLE ([Bedell et al., 2019](#)), where the spectra is modelled as a linear combination of a time-invariant stellar spectra and a time-variant telluric spectrum. A maximum-likelihood approach of the templates is carried out, coupled with a regularization routine to avoid over-fitting of the model, thus yielding precise RVs and high quality stellar and telluric templates. However, WOBBLE is non-optimal when dealing with faint stars or short exposure times, needs an high number of observations to properly disentangle both components and it assumes a fixed stellar template that does not account for physical effects.

The idea of generating a probabilistic model of stellar spectra is not new and, in fact, it has already been done through the usage of Gaussian Processes. [Czekala et al. \(2017\)](#) used two Gaussian Processes to disentangle the spectra of binary stars, allowing to both retrieve a stellar model of each star and estimate the RVs. The core idea of this model was that we would expect the stellar spectra, at two different moments in time, to be described by the same model but separated by a given RV. Thus, the model was applied to a concatenation of multiple observations, after shifting them to a common rest frame. In

this case, the shift between observations is obtained through a theoretical equation, thus using the parameters of the two binaries as free parameters of the model.

A more recent approach can be found in GRACE, [Rajpaul et al. \(2020\)](#) where each spectra is compared against all other, allowing to compute pairwise RVs. Rather than being limited by a stellar template, in this approach the stellar spectra is modelled with a Gaussian Process (GP) during the pairwise comparison. The relative RV is then extracted through a maximum likelihood approach, inferring both the RV shift and the model hyperparameters simultaneously, yielding an RV measurement for each pair of two different spectra. Then, the individual $N(N - 1)/2$ RV for each observation is obtained through a variance-weighted mean of every pairwise RV.

Despite the promising results of the previous two approaches, they are both limited by the large computational cost associated with the modelling of stellar spectra with a GP. Considering the sizes of each spectral order in the current generation of spectrographs, we see that it can quickly lead to datasets with thousands of data points. Thus, as an attempt to bypass such limitations, they were applied in parallel to small chunks of stellar spectra, thus effectively reducing the size of the dataset and making it computationally tractable.

1.3 Motivation for this Thesis

In this introductory chapter, a brief overview of the evolution of the detection of exoplanets was given, showing an explosive increase since its infancy. Due to improvements in instrumentation, the main limitation of our analysis is no longer the instrumental variability but, rather, spurious signals introduced by our atmosphere and the stars themselves. These limitations are thus calling for novel algorithms to both extract radial velocities from the stellar spectra and to disentangle between the effects injected by stellar variability and those from keplerian sources. With this in mind, the work developed in the context of this thesis was focused on the development of novel techniques for the extraction of precise radial velocities.

In Chapter 2 the first main result of this thesis is presented: **s-BART**, a recast of the traditional template matching algorithm into a semi-Bayesian framework built around an assumption of an achromatic RV shift. This choice allows the model to better describe the expected signal of a (exo-)planetary companion, i.e., wavelength independent shift, whilst providing a consistent methodology to characterize the RV posterior distribution. The performance of this pipeline is showcased on over more than 5000 observations, spread

between 3 different instruments: ESPRESSO, CARMENES, and HARPS-N. Furthermore, highlights from the scientific exploitation of **s-BART** RVs are also described.

In Chapter 3 a transversal issue to template matching algorithms is described (**s-BART** included). We find that whenever all available observations of a given target are collected inside a single night, the template matching RVs reveal a systematic RV trend over the duration of the observations. This trend is approximately linear and it usually has a peak-to-peak amplitude at the the meter per second level, or even higher. Furthermore, it has been detected on other TM pipelines and on multiple stars observed with HARPS-N and ESPRESSO. Despite the wide array of tests applied to the data, the root cause is yet to be identified, thus remaining as an open-question of the thesis.

Following the development of **s-BART** a few sub-optimal characteristics were found, namely the assumptions of i) independence between stellar model (template) and observations; and ii) the time-invariant nature of this model. In an attempt to mitigate those issues, Chapter 4 presents a proof of concept of an RV-extraction model where the stellar template is replaced by a probabilistic model. For this effect, the stellar spectra is modeled under a Gaussian Process framework. In this Chapter, some preliminary results are showcased, alongside possible avenues for further development.

At the end of the thesis, in Chapter 5, the main conclusions will be presented, in tandem with potential directions for future growth.

Chapter 2

A semi-Bayesian implementation of template matching

The past few years have seen the development of a large number of approaches for the extraction of precise RVs from stellar spectra, as previously discussed in Section 1.2.4, with some based on template matching. This subset of methodologies share the same assumption of chromatically-independent RV shifts, i.e., each spectral order is assumed to be an independent measurement of the same underlying phenomenon. However, this assumption is not compatible with our previous knowledge of the expected RV signals in a given star. The presence of planetary companions is expected to introduce periodic variations in the positions of the stellar lines, with an amplitude independent of the wavelength at which it is measured, i.e., we expect an achromatic effect. Furthermore, the implementation of such an approach into a Bayesian context allows for a consistent method to characterize the uncertainties associated with the RV extraction process.

This assumption of achromaticity was the cornerstone of the first major development of this thesis, which led to the implementation of **s-BART**: *Semi-Bayesian Approach for RVs with Template-matching*, published in Silva et al. (2022) and publicly available on github*. In this Chapter we start by detailing the inner workings of this approach, from the pre-processing of the stellar observations up to the estimation of the RV shifts. Then, we describe the performance of the algorithm at the time of publication (Chapter 2.4) and the most recent results (Chapter 2.5), highlighting some of its most noteworthy results.

It is also important to note that whilst the default configurations of **s-BART** are presented, every default value can be customized by the user to fine-tune the performance for

*<https://github.com/iastro-pt/sBART>

a given target. However, for the vast majority of cases, the default values are sufficient to guarantee optimal RV extraction. The pipeline was written in *python*, with the computational bottlenecks being handled by either an efficient parallelization of the code or low-level routines written in *Cython* and compiled to *C*.

2.1 Model preparation

The extraction of precise RVs is preceded by a first stage of pre-processing of the input spectra, where we account for telluric features, create the stellar template and reject some regions of the stellar spectra. In this sub-Chapter we go over those necessary steps, following the order of their application, as schematized in Figure 2.1.

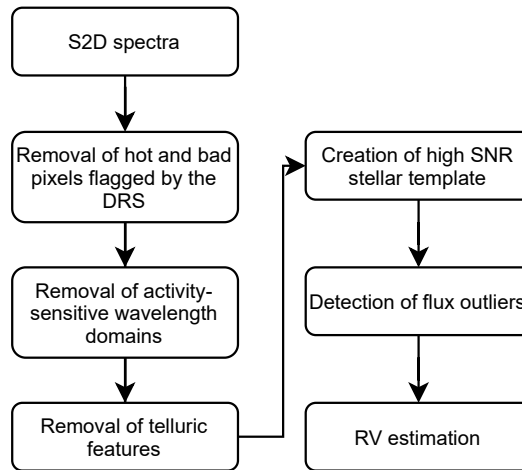


FIGURE 2.1: Schematic of the pre-processing of the spectral data, applied before the RV estimation.

2.1.1 Pre-processing of the data

The extraction of the spectral orders from the image and necessary calibrations and corrections are handled by the official Data Reduction Software, DRS, of the respective instruments. The DRS also handles the correction of the instrumental profile, also known as the blaze, as well as the computation of the BERV. On top of that, in the case of ESPRESSO, it provides a correction for a difference in the flux in two detectors (blue and red) as introduced by the different atmospheric extinction of the light.

The pre-processing stage starts with a search, on the DRS data products, for hot and bad pixels to reject them, as well as those that have null data, mainly at the borders of the bluer orders. After this point we account for the fact that some spectral lines are more

sensitive to different effects of stellar variability, as previously discussed in Chapter 1.2.2.2, thus being used as activity indicators or proxies. As we want to minimize the impact of those effects on the RV extraction process, we reject a wavelength window around some activity proxies, as prescribed in Table 2.1 and shown in Figure 2.2. The same rejection is applied to all observations analysed through this pipeline.

TABLE 2.1: Central wavelength (measured in vacuum) and size of the spectral regions removed from the spectra.

Line	Wavelength (Vacuum) [Å]	Window [Å]	Reference
CaK	3934.77	0.6	(Robertson et al., 2016)
CaH	3969.59	0.4	(Robertson et al., 2016)
$H\epsilon$	3971.20	0.6	Balmer series ($n = 7 \rightarrow 2$)
$H\delta$	4102.89	1.4	Balmer series ($n = 6 \rightarrow 2$)
$H\gamma$	4341.69	2.0	Balmer series ($n = 5 \rightarrow 2$)
$H\beta$	4862.71	1.8	Balmer series ($n = 4 \rightarrow 2$), (Flores et al., 2016)
Na I D	5891.59	1.4	(Robertson et al., 2016)
Na I D	5897.56	0.9	(Robertson et al., 2016)
$H\alpha$	6564.62	2.0	(Kuerster et al., 2003)
CaI	6574.61	1.8	(Kuerster et al., 2003)

In order to properly reject those regions, we must account for the stellar RV and Earth's motion, so that we find a match between the values of Table 2.1 and lines in the stellar spectra, as show in Figure 2.2

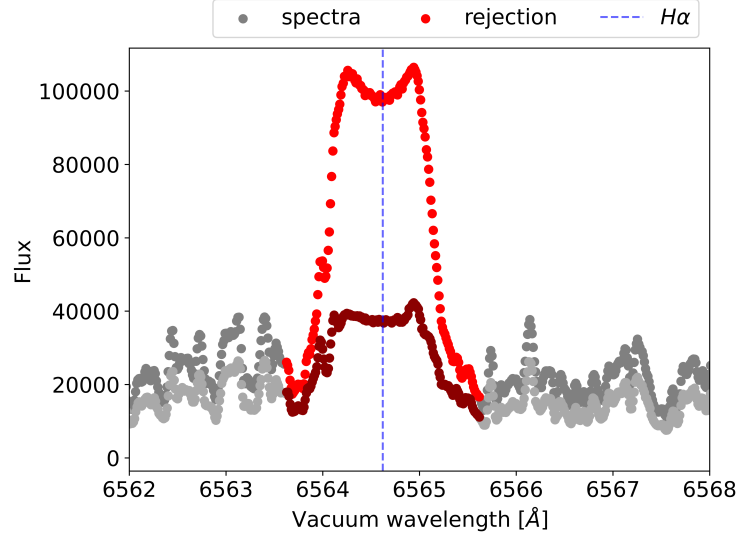


FIGURE 2.2: Rejection of a wavelength window around the $H\alpha$ line for two observations of an active M-type star.

Lastly, after applying the pixel-wise rejections described in this sub-Chapter, we reject any spectral order which had more than 75% of its pixels rejected or has a signal-to-noise

ratio smaller than 5^* .

2.1.2 Creation of a telluric template

In our quest for precise Radial Velocities we must, as discussed in Chapter 1.2.2.1, take into the account the impact of Earth’s atmosphere. Our available stellar spectra will not only have telluric lines in different positions during the year, but also features with different depths, depending on the atmospheric conditions at the time of our observations. Thus, to avoid biasing our RV extraction, we must account for them, either through a correction or masking stage. Despite the impressive results of Allart et al. (2022), it is currently focused towards the lines introduced by the presence of water in the atmosphere, the largest problem at optical wavelengths. For other elements, e.g., oxygen, the model is not fully validated and, consequently, still requires a masking procedure. For simplicity, and to avoid introducing changes to our observed data through a telluric-correction attempt, our approach was to create a binary mask to reject wavelength regions impacted by the atmosphere. This is not the optimal choice, as it leads to a large loss of RV content, but it is a very conservative approach that attempts to minimize the bias introduced by the micro-telluric features. We left for future work the inclusion of a telluric correction stage.

The process of creating a binary mask comprises a series of consecutive steps:

- We start by searching the available observations for the one with the highest water content (humidity) in the atmosphere, as it will provide the atmospheric conditions to maximize the detectability of the water features.
- Using this observation, we use it as a reference to construct a transmittance model through *Telfit* (Gullikson et al., 2014) or *Tapas* (Bertaux et al., 2014). The former is the default choice, as it allows a more fine-tuned control of the model, as well as providing a easier user-interface that can be automatized.
- The continuum level of the transmittance model is estimated through a rolling median filter. To minimize the impact of the deeper telluric features, which can even reach saturation, we select a span of 1000 points. Furthermore, to avoid numerical artifacts near the edges of the model, we locally reduce the window to 50 points, for a improved performance. Otherwise, the continuum model would be very dependent on the boundary conditions selected for the filter.

*Every rejection/threshold imposed by **s-BART** can be customized by the user.

- We flag, as wavelength domains affected by tellurics, those where the transmission is below a given threshold. By default, `s-BART` places it at 99% of the continuum level.

In Figure 2.3 we show the application of the rolling median filter to determine the continuum level of a transmittance spectrum. We see that in the regions with the shallower feature (upper row of the Figure) there is no noticeable impact on the estimation of the continuum. However, it is also clear that the default threshold is not sufficient to properly identify the smaller features. However, selecting a (even) more conservative threshold level would result in a large impact across the entire spectral coverage (refer to the discussion from Chapter 1.2.2.1).

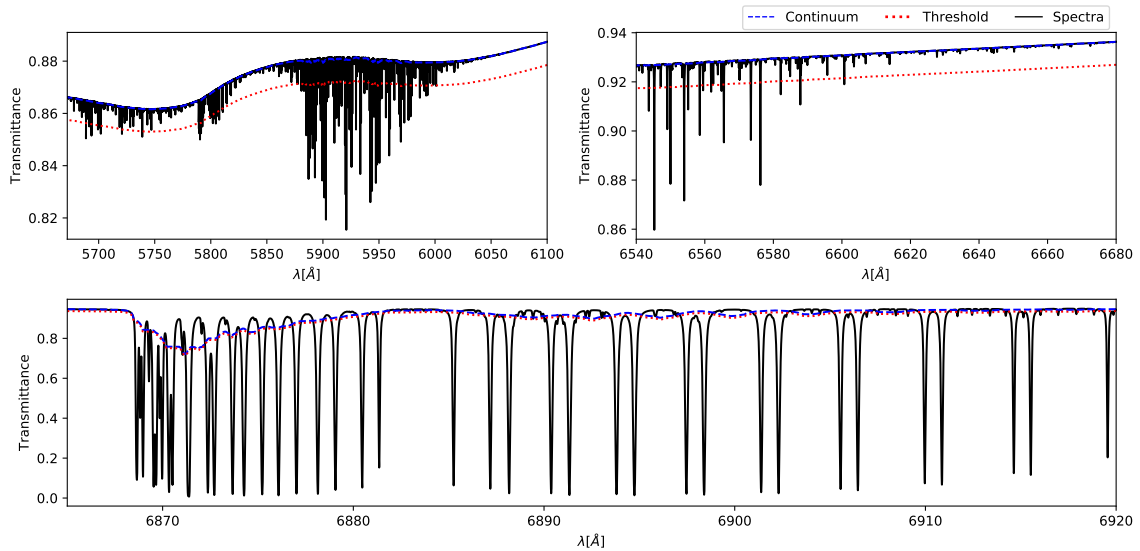


FIGURE 2.3: Comparison, for three spectral regions, between the telluric spectrum generated with Telfit (black) with the continuum level obtained with a median filter (dashed blue line) and the telluric threshold (dotted red line) built from the continuum. Image adapted from [Silva et al. \(2022\)](#).

If we now turn our attention to the heavily contaminated regions (bottom panel), we find a degradation of the quality of the continuum level. However, and despite the worse performance, it is still possible to properly flag them. Thus, our choice of rejecting telluric features at a threshold of 1% is a trade-off between loss of RV content and minimization of bias due to leftover tellurics.

To guarantee that our RV estimation is informed by the same spectral regions, we enforce the usage of a common telluric mask for each observation of our instrument. As previously discussed in Chapter 1.2.2.1, Earth's motion around the barycenter of the Solar System introduces a RV shift, whose correction is usually handled by the DRS of the

instruments. However, as the telluric lines are fixed on the detector, which is in Earth’s reference frame, their position relative to the stellar lines will change, and in the BERV-corrected spectra will appear as shifted by $-BERV$. Thus, for us to reliably construct a binary mask common to multiple observations, we must account for such movement. This process can be done by either:

1. Discarding a wavelength domain around each feature corresponding to the maximum BERV ($\pm 30 \text{ km s}^{-1}$, obtained for stars along the direction of Earth’s orbit).
2. Shifting the binary mask to the BERV associated with each observation, after which we combine the shifted masks into a common one.

Within **s-BART**, the default is to take the more conservative approach of 1), but as it is somewhat rare to have a uniform distribution of observations throughout the year, the second option allows for the usage of a larger amount of spectra, leading to better RV precision on the final measurement.

2.1.3 Creation of a stellar template

The stellar template plays the most critical role in template matching methodologies, as it is assumed to be a high-SNR, time-invariant, model of the stellar spectra. Each observation is assumed to be a perfect match of the template, provided that it is shifted by the correct RV shift. The construction of a stellar template starts from assuming that we have a good first estimate of the stellar RV, either obtained through the CCF method, or a previous application of the template matching algorithm. Using this RV vector, we can use it to Doppler-shift all available observations to a common rest frame, where they can be combined with one another.

For the combination of the observations, we start by selecting one observation to provide the wavelengths at which the template will be defined. By default, this is the observation with the highest signal-to-noise, providing the reference wavelengths λ_{ref}^{rest} . In order to avoid having different signal to noise ratios (SNRs) within the same order of the template – e.g., due to the different observations having different wavelengths at the start and end of the order as shown in Figure 2.4, or due to the pre-processing stages described in Chapter 2.1.1 and 2.1.2 – we also remove regions where one or more regions have a gap, as highlighted in the Figure.

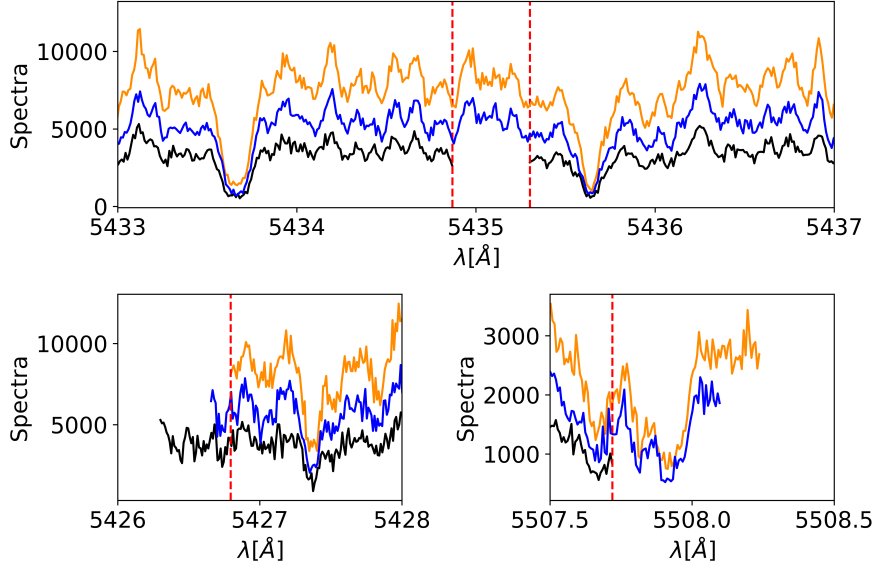


FIGURE 2.4: Gaps in the wavelength coverage introduced by the pre-processing stages (top panel, with an unrealistic gap for ease of visualization) and at the edges of the spectral order (bottom panels). The red lines represent the boundary between the regions where there is an overlap between the 3 observations and those where one observation is missing data.

To combine the multiple observations, we compute the mean spectra, evaluated in wavelengths λ_{ref}^{rest} . To account for the fact that the different observations are sampled at slightly different wavelengths (consequence of different wavelength calibrations, BERV correction and instrumental drift corrections), we assume that the spectra is continuous and interpolate each observation to λ_{ref}^{rest} . For this effect, a cubic spline interpolation (a piece-wise algorithm, better described in Appendix D) is used, providing an interpolated spectrum (and template) that is two times differentiable, i.e., providing a continuous second derivative. After the interpolation to the common reference frame, we compute the mean of the fluxes in order to build the high SNR stellar template. The mean was selected in order to maintain the count level at a reasonable value, avoiding possible numerical issues further ahead in the RV extraction.

Lastly, it is also important to note that in the final stellar model we will have the presence of micro-telluric features that are left out of our masking process from sub-Chapter 2.1.2. If our observations are spread throughout the year, i.e., collected with different BERV, it will allow those features to be averaged out of the model, even more if they were collected at lower airmasses as the features are shallower. By default, **s-BART** uses all available observations, regardless of their airmass. However, users are recommended to reject, for the creation of the stellar model, the observations that were collected at airmasses

higher than 1.5. Under nominal observing conditions, the majority of observations are well under this value, as showcased in around 4000 ESPRESSO observations taken from the WG1 sample (Figure 2.5) that will be explored in Chapter 2.5.2.

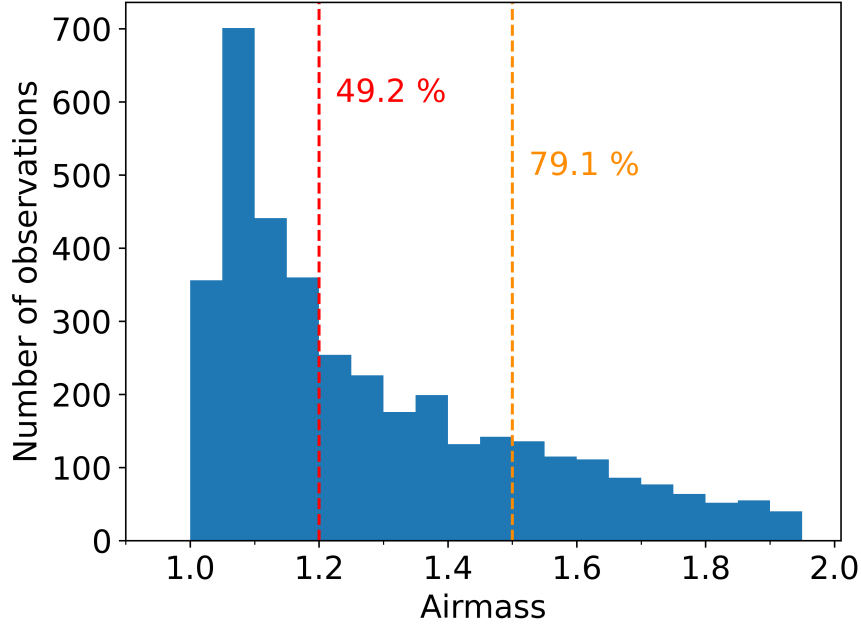


FIGURE 2.5: Histogram of the airmass at the start of 4000 observations of ESPRESSO’s WG1 sample. The vertical lines represent a airmass cutoff of 1.2 (red) and 1.5 (dark orange), with the percentages representing the number of observations whose airmasses fall below the cutoffs.

This criterion thus provides a balance between the number of observations that are rejected, assuming usual conditions for observability, and the number of observations with deeper telluric features that are kept for the construction of the stellar template. On top of that, it is also motivated by the correction of atmospheric dispersion being worse at higher airmasses (Wehbe et al., 2019), which might lead to non-homogeneity in the data used for the template.

2.1.3.1 Evaluating the uncertainty associated with the stellar template

The stellar template will be affected by some uncertainty, given that it is built as a mean of N spectra, all affected by flux measurement uncertainties. If each spectrum had the same SNR, then we would expect the SNR of the template to be approximately equal to $\sqrt{N}$ times the SNR of each observation. Under this assumption, we would need 100 observations to achieve a mean uncertainty (standard deviation) per flux one order of magnitude smaller than that associated with any single observation. Since many targets

will have fewer observations, it is thus important to keep in mind that whilst our stellar model will have a very high SNR, it will not be error free. Thus, the uncertainties of this model will be taken into account in the final RV estimate, as described in Chapters 2.2 and 2.3.

This process is not straightforward, as we have to take into account that both the spectral data and the template are interpolated with cubic splines in two different stages: during the creation of the stellar template and in the RV extraction procedure. In Appendix D we describe the analytical uncertainty propagation through the cubic spline interpolation algorithm. To showcase the need of accounting for this source of uncertainty, we selected the available observations of an M4 star ($N_{spectra}=21$) from the sample used in Chapter 2.4.2. The chosen observations were collected after ESPRESSO’s fiber upgrade in June 2019 [Pepe et al. \(2021\)](#) and we selected data from the 100th spectral order (central wavelength of 541 nm).

In order to evaluate the impact of the number of spectra used to construct the template the observations were subdivided into two sets: those with an airmass below 1.2 ($N = 8$); and those with an airmass below 1.5 ($N=13$). After creating the stellar templates we align them both with each observation and interpolate them to the wavelength solution of the observation, also evaluating the uncertainties with the process. Then, we compare them against the ones associated with the spectra, as computed by the ESO pipeline. A direct comparison is not possible, as observations with lower flux values also have lower photon noise and, consequently, smaller absolute flux uncertainties. Instead, we compute the SNR of the template and spectra, for each pixel of the order, and compute their ratio. It is also important to note that we should expect to see different results if we compare the template uncertainties with those from spectra with higher or smaller flux uncertainties. Thus, we distinguish the comparison with high SNR spectra (arbitrarily defined as those with a flux SNR of at least 100) and lower SNR spectra (flux SNR smaller than 100).

As expected, an increase in the number of observations used to construct the template translates into an increase in its SNR, when compared to the ones from the individual observations, as shown in Figure 2.6. If we only consider the lower SNR spectra available in this dataset (non-filled bins) we see that both templates lead to uncertainties close to one order of magnitude smaller than the ones found on the observations, albeit not quite there. If we turn our focus to the higher SNR data, the gain is not as evident, which hints towards the fact that we should not ignore the noise in the stellar model. For a

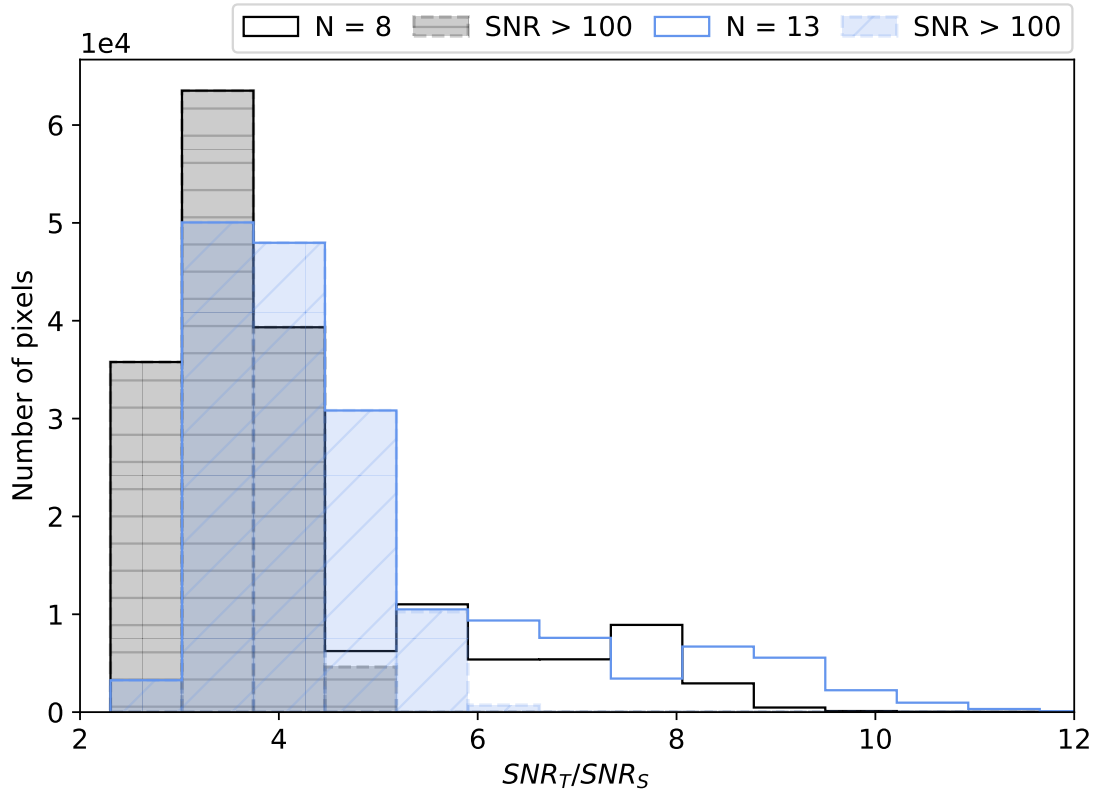


FIGURE 2.6: Histogram of the SNR ratio, for the 100th order, of the template and each of the 21 individual observations of an M4 dwarf, obtained after the fibre upgrade of ESPRESSO in mid 2019. In order to compare the impact of the number of observations used to construct the template in its associated uncertainty, one template was built with 8 observations (black curve) and the other with 13 observations (blue curve), selected by the airmass at the start of each observation. For all observations we also highlight by filling the bins with the corresponding colour, the comparison with observations with a SNR of at least 100 for the selected spectral order. The figure was taken from [Silva et al. \(2022\)](#).

more detailed analysis, Table 2.2 presents the comparison of each template against three different sets of observations: with all observations; with only the observations used in the construction of the template; with observations whose SNR is at least 100. In all sets, the median value of the SNR is above $\sqrt{N}$, which is explained by the selected observations having different SNRs. Considering the results presented in the Table and Figure, we can thus conclude that we should not ignore the uncertainties in the template, pressing the need of having a realistic propagation of uncertainties in the construction of the template and the subsequent interpolations that it will be subjected to.

The main problem with our uncertainty propagation procedure, fully described in Appendix D, lies in its computational cost, as it depends on the inversion of large matrices. Whilst it is feasible to use it for the propagation of uncertainties during the construction of the stellar template, it will place a large bottleneck in the interpolation of the stellar

TABLE 2.2: Analysis of the SNR ratio between the two templates and three different subsets of the available observations. The first criteria, "All", compares the two templates against all available observations. The second, "Template", presents the comparison against only those that were used to construct the corresponding stellar template. The last one, "SNR ≥ 100 ", represents the comparison against all observations that have a SNR, in the 100th order, higher than 100. In the last column we report the median and standard deviation of the SNR ratios across fluxes.

Airmass	N	$\sqrt{N}$	Observations	SNR ratio
≤ 1.2	8	2.83	All	3.68 ± 2.11
			Template	3.37 ± 0.45
			SNR ≥ 100	3.32 ± 0.50
≤ 1.5	13	3.61	All	4.28 ± 1.88
			Template	4.09 ± 0.92
			SNR ≥ 100	3.92 ± 0.61

template for each tentative RV shift that is tested during the RV extraction procedure.

To mitigate the extent of this problem, we evaluated the validity of also interpolating the uncertainties, rather than applying the costly analytical propagation. To do so, we selected the two templates from Figure 2.6 and compared the uncertainties obtained through a cubic spline interpolation of the originals and the result of the analytical propagation. The main result of this comparison is that the interpolation of flux uncertainties results in their overestimation by a factor of 1.07 ± 0.05 across the selected observations. Despite this small increase in the uncertainty, it is not deemed to be problematic, especially when taking into account the already high SNR of the template itself (when comparing with the majority of the available observations). With this in mind we propagate, during the RV extraction, the uncertainties in the stellar template by interpolating them to the desired wavelength solution.

2.1.3.2 Removal of outliers in the spectra

In the template matching algorithm, the stellar template is taken to be a very high SNR model of the stellar spectra, assumed to be capable of describing any given observation of our star. In the vast majority of cases, this leads to a very good agreement between the observations and the template. However, we can find spectral regions where such assumption does not hold, e.g., as seen in the top row of Figure 2.7, thus raising the need to remove flux outliers before starting the RV estimation procedure. On top of such obvious outliers, this rejection process might also be able to flag some micro-telluric features, given that the BERV spacing is enough for such features to be averaged-out of

the stellar template. It is important to note that any point that was previously discarded in Chapter 2.1.1 will be ignored during this process.

The search for outliers is done individually for each observation, starting the alignment of the stellar template and observation using the previous guess for the RV that was used for the construction of the model. Then, to account for differences in the continuum level, a first degree polynomial is fitted, with slope m and intercept b , to the ratio between the spectrum and template. Lastly, we compute the logarithm of the ratio between spectrum and template and use it as a metric to flag mismatch regions:

$$metric_{\lambda_i} = \log \left(\frac{S_{\lambda_i}}{p(m, b)_{\lambda_i} T_{\lambda_i}} \right) \quad (2.1)$$

where S_{λ_i} is the flux of the stellar spectrum, $p(m, b)_{\lambda_i}$ the first degree polynomial and T_{λ_i} the interpolated stellar template, all evaluated at wavelength λ_i . The logarithm was selected, rather than the ratio, to mitigate the impact of larger differences appearing in the lower SNR regions closer to the edges of the spectral orders, thus ensuring a reasonable behaviour for the large dynamic range of fluxes seen inside a single spectral order. From this metric, we then flag each point more than $6\sigma^*$ away from the median metric (of the entire order) as outliers. This large window ensures that only the larger flux differences are rejected, rather than over-flagging on the noisier portion of the spectra, as using lower thresholds would result in excessive flagging in both edges of the order, as the differences (in absolute values) between spectra and template are larger.

The performance of this algorithm is showcased in Figure 2.7, where it is applied to two different spectral chunks in different regions of a spectral order. In the left panel, minimal (relative) differences are found between the spectrum and the normalized template, with the exception of the clear outlier that is picked by the metric. On the noisier part of the spectrum (right panel) the match between spectrum and template is worse, but the tolerance window is large enough to avoid flagging the points, as there are no clear outliers.

*We use σ to refer to the standard deviation of the metric across the entire order.

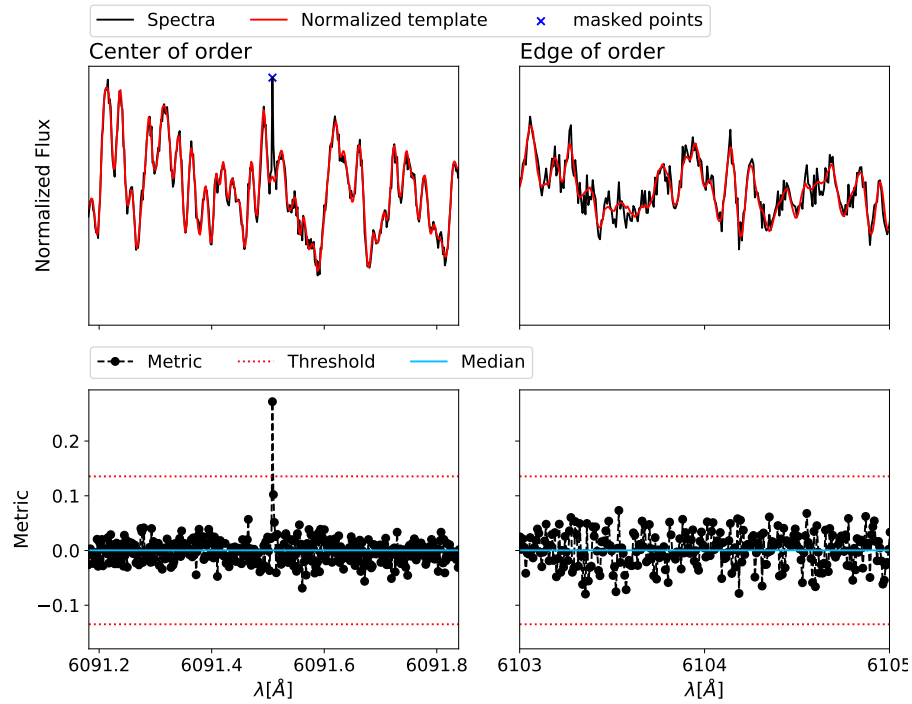


FIGURE 2.7: Outlier removal routine for two regions of the same order, with respect to an observation for which outlier identification achieved convergence after the first iteration. For representation purposes we normalized the flux in the center and the edge of the order. **Top row:** Comparison between the stellar template (red line) and the spectra (black line). The blue crosses represent the points that were flagged by the method. **Bottom row:** Differences between the template and spectra (black points). The blue line is the median value, whilst the dotted red lines represent a 6σ difference from it. Image adapted from [Silva et al. \(2022\)](#).

2.2 A classical template matching approach

The core concept of a classical template matching algorithm is the determination of the RV shift that best aligns a given stellar model with the spectra. This process is depicted in Figure 2.8 and further described within this Chapter. In the previous sub-Chapters we have built the basic building blocks of the method, which will now be used to fully implement it. The main motivation behind the re-implementation of a classical approach is to create a tool to properly benchmark **s-BART** (introduced in Chapter 2.3) against a classical method, in the case where both approaches apply the same pre-processing steps. Furthermore, it also provided a avenue to compute template matching RVs for spectrographs in which, at the time, had no official implementation of the other existing approaches.

The classical template matching approaches are built on top of the assumption that the different spectral orders are independent measurements that can be Doppler-shifted by different amounts. Thus, they compute independent RVs for each spectral order, henceforth

order-wise RVs, which are then combined, e.g., through a weighted mean, to build the final RV estimate for the entire observation. This core assumption is also shared by this re-implementation of the TM algorithm.

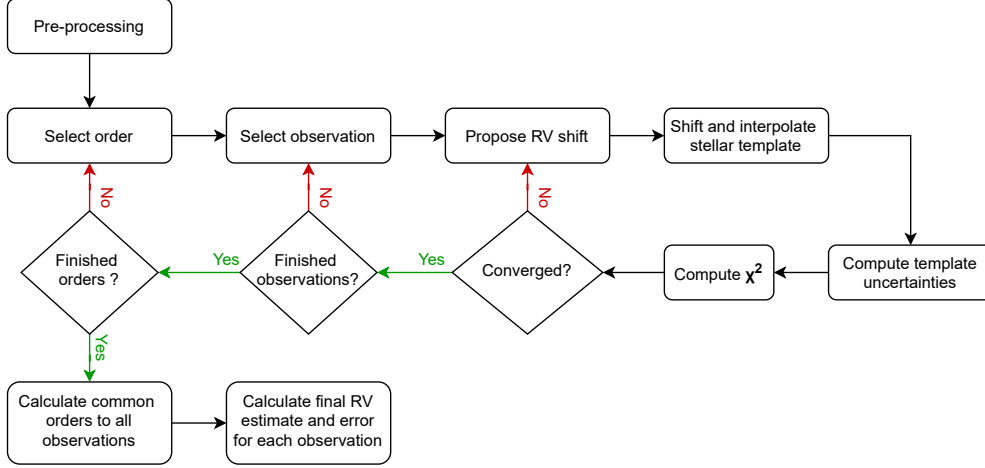


FIGURE 2.8: Schematic of a classical template matching algorithm.

One important issue to keep in mind is that by applying different Doppler-shifts to the stellar template, some of its lines might no longer overlap with the wavelength coverage of the spectral orders that we are processing at that time. This will mean that the spectral regions that inform our RV estimation will be different, depending on the RV proposal that our alignment algorithm will propose. To avoid such issue, it is important to determine the spectral regions that will be common to any RV shift that the template might suffer, i.e., we must define *a priori* an interval of RV shifts that our algorithm can propose. Furthermore, not only is this an issue at the edges of the orders, but also in any gap that might suffer from either the pre-processement of the data, or the telluric masking (i.e., Chapters 2.1.1 and 2.1.2).

To account for the gaps, we start by determining the edges of each continuous wavelength interval, which are then Doppler-shifted by the maximum and minimum blue and red shifts. The wavelengths that are not common to any valid shift, are rejected from the RV extraction process. By default, both our classical implementation and **s-BART** consider a window of $\pm 200 \text{ m s}^{-1}$ centred in the previous RV estimate.

After this point, we can compute radial velocities for any given order, o in our data, through a comparison of the stellar template ($T_{\lambda_i}^o$), with the individual observations ($S_{\lambda_i}^o$). However, there is no guarantee that the two will share the same continuum level, as discussed in Chapter 2.1.3.2, since the continuum of the stellar template will have the mean flux

TABLE 2.3: The default RV steps that represent the separation between the minimum of the χ^2 and the neighbouring points that will be used to perform a numerical fit of the parabola.

Spectrograph	$\Delta RV [cm/s]$
ESPRESSO	10
HARPS / HARPS-N	50
CARMENES	50

correction of different effects. To account for this difference we fit a polynomial, $p(m, b)_{o, \lambda_i}$ to the ratio of the stellar template, and the spectra. This will allow the scaling of the stellar template to the level of any given observation, as a mismatch in the continuum levels might introduce a bias in the RV estimation. The order-wise RV shift is then estimated through a χ^2 minimization, as shown in Equation 2.2.

$$\chi_o^2 = \sum_i^{N_{pixels}} \frac{1}{\sigma_S^2 + \sigma_T^2} \left[S_{\lambda_i}^o - p(m, b)_{o, \lambda_i} * T_{\lambda_i}^o \right]^2, \quad (2.2)$$

where σ_S and σ_T are the 1σ uncertainties of the stellar spectra and template, all evaluated at the wavelength λ_i . This minimization is performed through Brent's method, implemented within *scipy* (Virtanen et al., 2020), and bounded inside the previously mentioned RV interval, that is constructed for each individual observation. Each RV shift proposed by the algorithm will trigger a Doppler-shift of the stellar template, followed by an interpolation to the wavelength solution of the observation. As mentioned in Chapter 2.1.3.1, we apply a cubic-spline algorithm (described in Appendix D) and the flux uncertainties are also interpolated through the same algorithm. Similarly to Zechmeister et al. (2018) we numerically fit a parabola to the χ^2 curve after the convergence criterion of the minimizer is met, using the optimal RV shift and its two nearest neighbours, separated by $\pm \Delta RV$. Due to the different precision level of different spectrographs, this ΔRV must not be constant. Otherwise, the shape of the χ^2 curve may either: i) not be sensitive to such a small shift; or ii) be too sensitive to a large RV step and degrade the quality of the RV estimation; In Table 2.3 we present the steps that are typically considered for the different spectrographs.

This numerical fit of a parabola follows the prescription from Section 10.2 of Press (1992), with it being used to correct the order-wise estimation of the radial velocity and associated uncertainty with Equation 2.3.

$$\begin{aligned}
RV_{order} &= RV_{min} - \frac{\Delta RV}{2} \frac{\chi_{m+1}^2 - \chi_{m-1}^2}{\chi_{m+1}^2 + \chi_{m-1}^2 - 2\chi_m^2} \\
\sigma_{RV}^2 &= \frac{2\Delta RV^2}{\chi_{m-1}^2 - 2\chi_m^2 + \chi_{m+1}^2},
\end{aligned} \tag{2.3}$$

where RV_{order} is the final RV estimate, σ_{RV} is the measurement (1σ) uncertainty, RV_{min} is the RV value that minimises Equation 2.2, while $m-1$ and $m+1$ identify the two RV values selected at a distance ΔRV from RV_{min} . Figure 2.9 showcases the good match of the parabola to the χ^2 curve obtained in one ESPRESSO observation, coupled with the quick convergence of the selected minimization method.

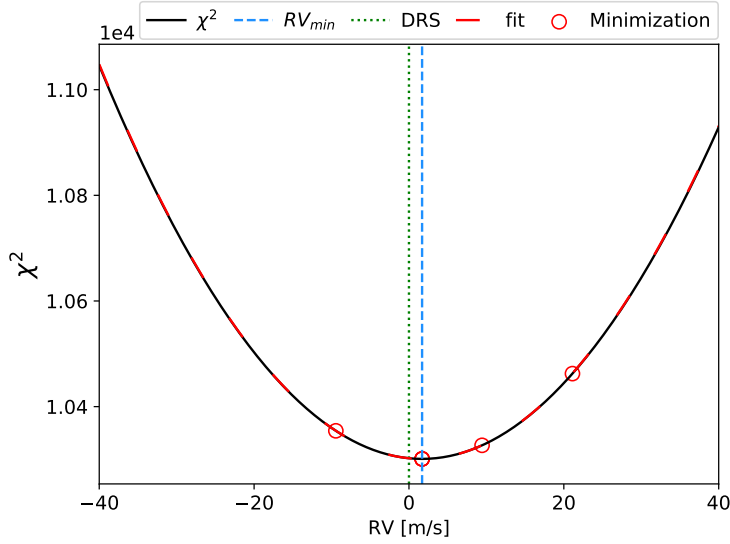


FIGURE 2.9: The χ^2 curve (black) of one spectral order from the ESPRESSO spectrograph, compared against the parabolic fit (red). The red circles represent the tentative RV values provided by the Brent minimizer before convergence, whilst the dotted green line is the DRS estimate and the dashed blue line is the RV estimated through Equation 2.3. Figure taken from Silva et al. (2022).

As the minimization procedure is carried out in a fixed RV window, we might find that the optimal RV lies near the edges, implying that there is no guarantee that we found the true minimum value of the χ^2 curve. That said, differences larger than 200 m s^{-1} between the current and previous RV estimations are not expected and atypical. Thus, we consider such cases to be outliers and completely reject the entire order from further consideration if the optimal RV lies within $5 \Delta RV$ of any edge of the interval.

After computing a RV with respect to each spectral order of each observation, they are then combined to construct the final RV estimate and associated uncertainty. This is done

through a weighted mean with inverse variance weights [Schmelling \(1995\)](#), as prescribed in Eq. 2.4:

$$\begin{aligned} v &= \frac{\sum \sigma_{v_i}^{-2} v_i}{\sum \sigma_{v_i}^{-2}} \\ \sigma_v &= \sqrt{\frac{1}{\sum 1/\sigma_{v_i}^2}}, \end{aligned} \tag{2.4}$$

where v_i is the RV of order i and σ_{v_i} is its associated (1σ) uncertainty, while N is the total number of orders for which we estimated an RV. The value of σ_{v_i} is estimated whilst ignoring any correlations that might exist between the spectral orders, assumed to be independent.

During the entire RV extraction process, i.e., this Chapter and Chapter 2.1.1, some spectral orders can be rejected either due to a lack of convergence within the specified interval, or too aggressive data rejection. Thus, it is very likely that the different observations will have RVs estimated with different spectral orders, resulting in varying levels of presence from different sources of noise (e.g., instrumental systematics, stellar activity) in the final time-series, possibly introducing extra RV variability. To ensure consistency in information, only RVs from orders that are common to every observation, in the dataset that we are using, will be considered for the weighted mean of Equation 2.4. This will obviously lead to a loss of RV precision, as a single observation will drive the rejection of one spectral order in all others. It also follows that, at some point, the preferable choice might be to reject the entire observation, rather than keeping it and losing one order in all others. Currently, this decision must be done manually, in a case-by-case basis, as the algorithm lacks a decision criteria to do so automatically.

2.3 Implementation within a Bayesian framework

The current approaches to template matching assume Doppler-shifts associated with the different spectral orders are independently generated. However, this is clearly not the case for the stellar RV component induced by orbiting bodies, such as planets or companion stars. In fact, those shifts are independent of the wavelength at which we measure them, i.e., they are achromatic. Since our main goal is the characterization of the masses and orbits of orbiting bodies, consistency suggests that our RV model should reflect this expectation of achromaticity through a single RV shift, common to all orders, to describe the differences

between stellar template and spectra. Deviations from this expectation, i.e., effects whose nature is stellar activity or instrumental systematics, should be dealt with explicitly, either through modelling or exclusion of affected spectral data.

The implementation of such a RV model into a Bayesian framework would also provide an avenue for consistent and straightforward characterization of the RV (posterior) probability for any observation. It will also allow to easily include into our model the uncertainties associated with the modelling of the continuum that are typically ignored in traditional template matching algorithms. Even more so, as we are not interested in the parameters associated with that modelling stage, we can marginalize them and keep the RV shift as the single free parameter of the model.

The main drawback of the application in a Bayesian framework is that in such cases all aspects of the model need to be specified prior to the actual data analysis, i.e., the information contained in the data cannot be used twice, for building the model (prior specification) and also for comparison with its predictions (through the likelihood). Unfortunately, in the context of template matching, the available spectra are used to specify the template (model building) as well as to estimate the RVs at the times of spectra acquisition (data analysis). However, as the template is usually built from a large number of observations, the information of a single observation will be diluted, reason for which we ignore the problem.

This is the motivation to describe the template matching approach for RV estimation described schematically in Figure 2.10 and with greater detail in this sub-Chapter as semi-Bayesian. This approach has been implemented in a pipeline capable of processing HARPS, HARPS-N, CARMENES and ESPRESSO data, was named **s-BART**: *Semi-Bayesian Approach for RVs with Template-matching*.

2.3.1 The RV model

The RV model is applied independently to each observation, containing only one parameter of interest: the relative RV shift that separates the observation from the stellar template. We thus desire to characterize the RV posterior probability distribution given each observed spectrum, which is proportional to the product of the RV prior probability distribution by the likelihood of the spectral data. The prior probability is assumed to be a uniform probability distribution, i.e., an uninformative prior was selected. This model still keeps an assumption of independence between pixels, based on the fact that the spectral data is

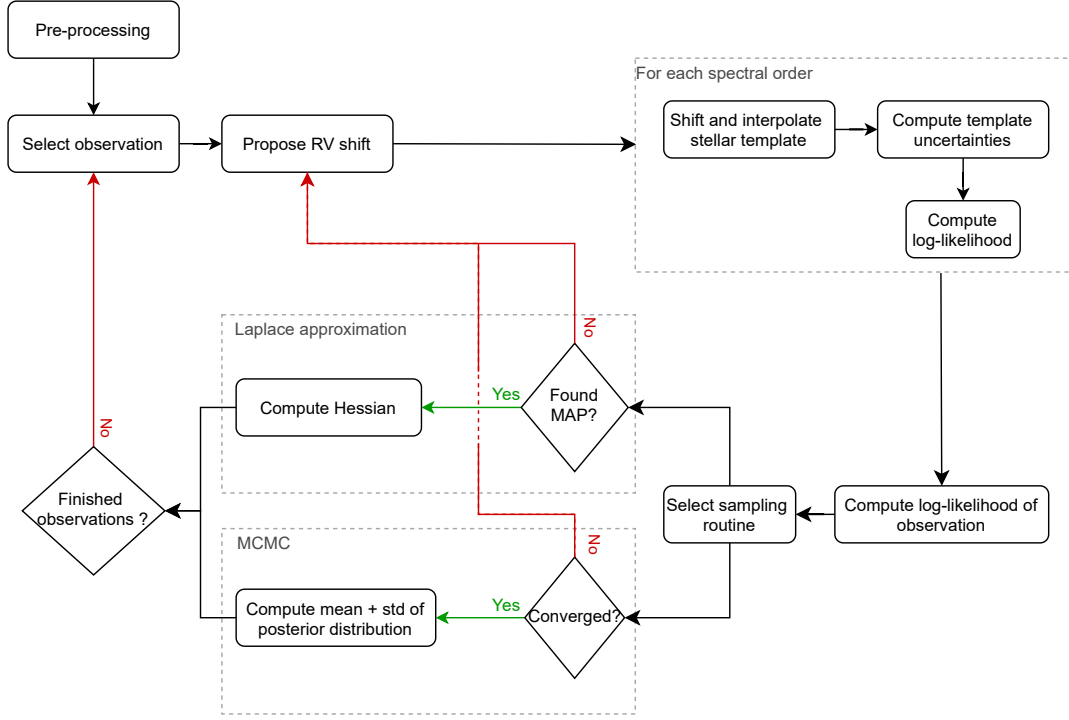


FIGURE 2.10: Schematic of the semi-Bayesian approach, considering both an MCMC and Laplace approximation to characterize the posterior distribution.

collected from independent physical pixels in the detector. In other words, we assume the flux measurements to be independent, leading to the likelihood of the full spectral order, conditioned on a given RV value, being equal to the product of the likelihoods of the fluxes measured for each pixel. As we assume the pixels to be independent, it follows that the different spectral orders are also taken to be independent, allowing us to first calculate the likelihood of the spectral data for each order ($\mathcal{L}_{order}$) and then multiply these to obtain the full likelihood ($\mathcal{L}_{OBS}$) for the entire observation:

$$\mathcal{L}_{OBS} = \prod_{order=0}^{N_{orders}} \mathcal{L}_{order} \quad (2.5)$$

As we saw in Chapter 2.2, the RV model will also comprise nuisance parameters that we wish to marginalize. Such parameters are those related with the stellar template:

- The slope, m , and intercept, b , of the first degree polynomial that is used to adjust the continuum level of the template spectrum to that of each spectrum under analysis;
- The flux associated with each pixel in the template, relative to the continuum.

The latter are effectively latent variables, affected by some uncertainty, characterized in Subsection 2.1.3. Each spectral order has its own set of independent nuisance parameters, thus the marginalization procedure can be applied order by order. We assume the joint prior distribution with respect to all model parameters to be separable, i.e., can be written as the product of prior distributions specifically associated with each parameter. Thus, in practice, the marginalization procedure involves the integration of the product of the prior distributions with respect to the nuisance parameters by the likelihood as a function of all parameters.

The likelihood of a given observed spectrum, S , conditioned on an assumed RV value, is thus given by

$$P(S|RV) = \prod_{i=1}^{N_{\text{orders}}} \int P(m_i, b_i) \prod_{j=1}^{N_i} P(T_{\lambda_{i,j}}) \times P(S_{\lambda_{i,j}}|RV, m_i, b_i, T_{\lambda_{i,j}}) dm db dT_{\lambda_{i,j}} \quad (2.6)$$

where N_{orders} is the total number of orders in the spectrum that are not discarded as a result of the procedure discussed in Section 2.2, N_i is the number of data points in order i (generally smaller than the number of pixels in each order, consequence of the masking of spectral regions in the pre-processing stages), $T_{\lambda_{i,j}}$ and $S_{\lambda_{i,j}}$ are the fluxes associated with pixel $\lambda_{i,j}$ in the (interpolated) template and spectrum, respectively. The prior probabilities, $P(m_i, b_i)$ and $P(T_{\lambda_{i,j}})$, are taken to be uninformative Gaussian distributions. Given that the last probability is also a Gaussian, with an expected value that is a linear function of $\lambda_{i,j}$, i.e., $(b + m\lambda_{i,j})T_{\lambda_{i,j}}$, in the limit of infinite variance for the prior probabilities, the integral is equal to a Gaussian and

$$\log P(S|RV) = \sum_{i=1}^{N_{\text{orders}}} \left(-\frac{1}{2} S_i^T K_i^{-1} S_i + \frac{1}{2} S_i^T C S_i - \frac{1}{2} \log |K_i| - \frac{1}{2} \log |A| - \frac{n_i - n_H}{2} \log 2\pi \right) \quad (2.7)$$

where S_i is a vector with the flux measurements for all pixels in order i , A is equal to $H_i K_i^{-1} H_i^T$, C is given by $K_i^{-1} H_i^T A_i^{-1} H_i K_i^{-1}$, and n_i is the number of data points in order i (for more details see Section 2.7 of [Rasmussen & Williams, 2006](#)). The $2 \times n_i$ matrix H_i contains the values of the $n_H = 2$ basis functions associated with the linear model $E[S_{\lambda_{i,j}}] = (b + m\lambda_{i,j})T_{\lambda_{i,j}}$ for each data point, i.e., 1 and $\lambda_{i,j}$ with the associated parameters $bT_{\lambda_{i,j}}$ and $mT_{\lambda_{i,j}}$, respectively. Finally, the variance-covariance matrix K_i is diagonal with

entries $\sigma_{S_{\lambda_{ij}}}^2 + \sigma_{T_{\lambda_{ij}}}^2$, where $\sigma_{S_{\lambda_{ij}}}$ and $\sigma_{T_{\lambda_{ij}}}$ are the standard deviations associated with the Gaussian probability distributions that describe the uncertainties, respectively, in the flux measurement and template construction (including interpolation) processes.

The construction of the RV estimate is done through the characterization of the RV posterior distribution, further discussed in the upcoming sub-Chapter 2.3.2. Its mean value and standard deviation will provide estimates of the RV value and uncertainty, respectively. The mean value was selected as it minimizes the quadratic loss function associated with this estimation, whilst the uncertainty accounts for noise introduced by the interpolation of the spectra, noise in the observed spectra and stellar template and, order-wise deviations of our assumption of a common RV shift.

2.3.2 Characterization of the RV posterior distribution

To characterize the RV posterior probability distribution associated with each observation we employ a *Markov Chain Monte Carlo* (MCMC) methodology as implemented in the *emcee* package (Foreman-Mackey et al., 2013). Convergence of the model is assessed through several criteria that must be simultaneously met by the chains:

- The chain length must be at least 50 times larger than the autocorrelation time (τ);
- The value of τ cannot change more than 1% after each iteration;
- The mean and standard deviation of the chains cannot change by more than 2 % after each iteration.

The strict convergence criteria, coupled with the inversion of large matrices due to the number of points in each spectral order, leads to a large computational burden. For context, achieving convergence for a single ESPRESSO observation takes ~ 10 minutes on a 24 cores@2.3GHz and 128 GB RAM server, limiting the applicability of the method to stars with a large number of available observations.

One solution to this problem is to approximate the posterior distribution for the RV shift, as it is approximately Gaussian (due to the large amount of information in the data). By performing a second order Taylor expansion around the posterior distribution maximum we can approximate it with a Gaussian distribution centred at the posterior's mode (also known as MAP - *maximum a posteriori*) and variance equal to the inverse of the Hessian of the log-posterior evaluated at the mode. For further details of the method, we refer back

to Section 3.4 of (Rasmussen & Williams, 2006) and Section 11.3 of (Gregory, 2005). This method is known as Laplace approximation and it transforms the characterization of the posterior into a optimization problem: the minimization of the negative log likelihood*. The main drawback that is associated with Laplace’s approximation is that it is only driven by the values near the MAP of the posterior distribution. Thus, if the posterior does not closely follow a Gaussian distribution, the approximation will not be accurate, as seen in the bottom panel of Figure 2.11. If the posterior is skewed, our estimation of the RVs will be biased and the associated uncertainties will either be under or over-estimated.

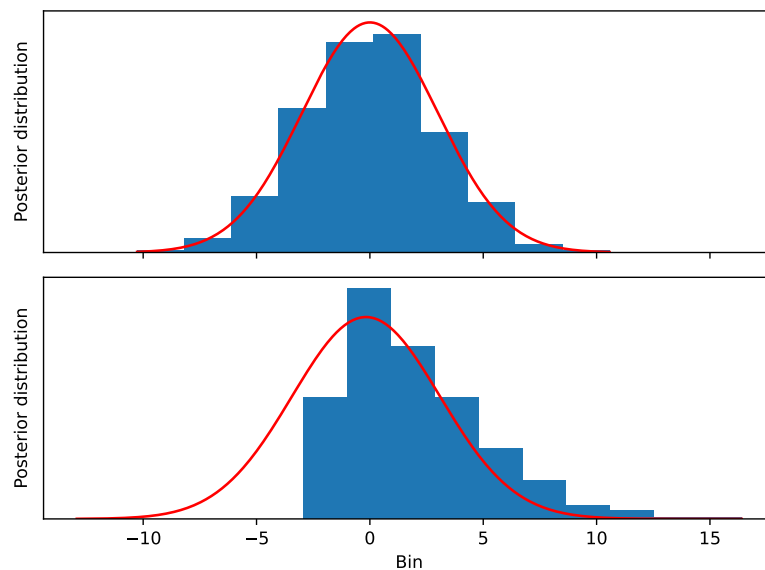


FIGURE 2.11: Laplace approximation (red) applied to two simulated posterior distributions (blue). **Top:** Normal distribution; **Bottom:** Normal distribution skewed towards higher values.

Figure 2.12 showcases the application of the approximation to one ESPRESSO observation, suggesting that the posterior does not exhibit an apparent skewness. Furthermore, and as we shall see further ahead, the results from the posterior and its Gaussian approximation agree at the 1σ level. However, if this was not the case, a more sophisticated algorithm should be preferred, such as Variational Inference (Blei et al., 2017, Gunapati et al., 2022).

For the practical implementation of the method we, once again, apply the Brent method bounded to the same RV interval as in a classical template matching method. This reduces the computational time to ~ 30 seconds per observation on the same machine.

*They are equivalent, as a uniform prior was taken for the RV

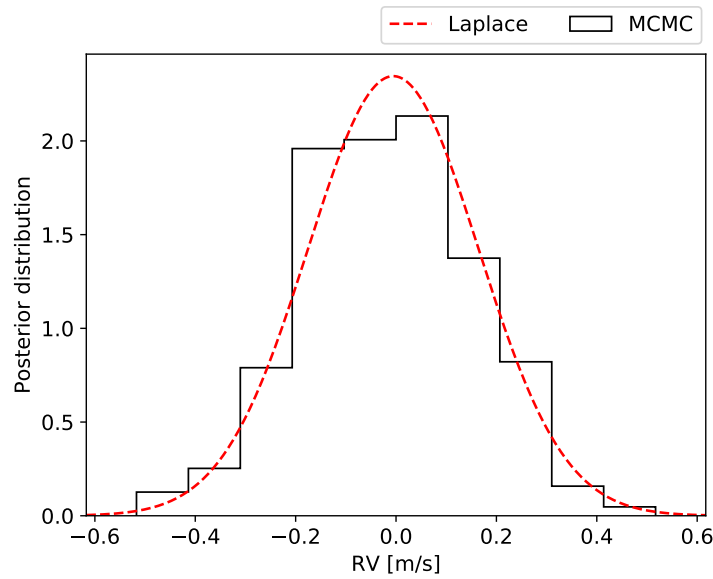


FIGURE 2.12: Comparison of the RV posterior distribution derived using the MCMC methodology (black curve) with the result of the Laplace approximation (dashed red line), for one observation of the M4 dwarf considered in Section 2.1.3.1. Image adapted from [Silva et al. \(2022\)](#).

In order to compare RV estimates obtained through the MCMC methodology and the Laplace approximation we ran the MCMC algorithm and the Laplace approximation on the ESPRESSO dataset of Proxima that was utilized in [Faria et al. \(2022\)](#). The comparison was divided in the different sub-instruments of ESPRESSO, creating individual stellar and telluric templates for each dataset, and is shown in Figure 2.13. We see that both the RV shifts and associated uncertainties are in agreement, with differences at the level of the cm s^{-1} level, well below the expected 10 cm s^{-1} precision of ESPRESSO, thus showing the validity of our assumption of a gaussian posterior for the RV distribution. Furthermore, in the original paper of **s-BART**, the comparison was performed with two targets, a K-type star with 27 ESPRESSO18 observations and a M-type star with 21 ESPRESSO19 observations reduced with the old DRS, leading to the same conclusions.

With this, the preferred approach to run the **s-BART** algorithm is to use the Laplace approximation to estimate the RV and associated uncertainty, but both options are available to the user.

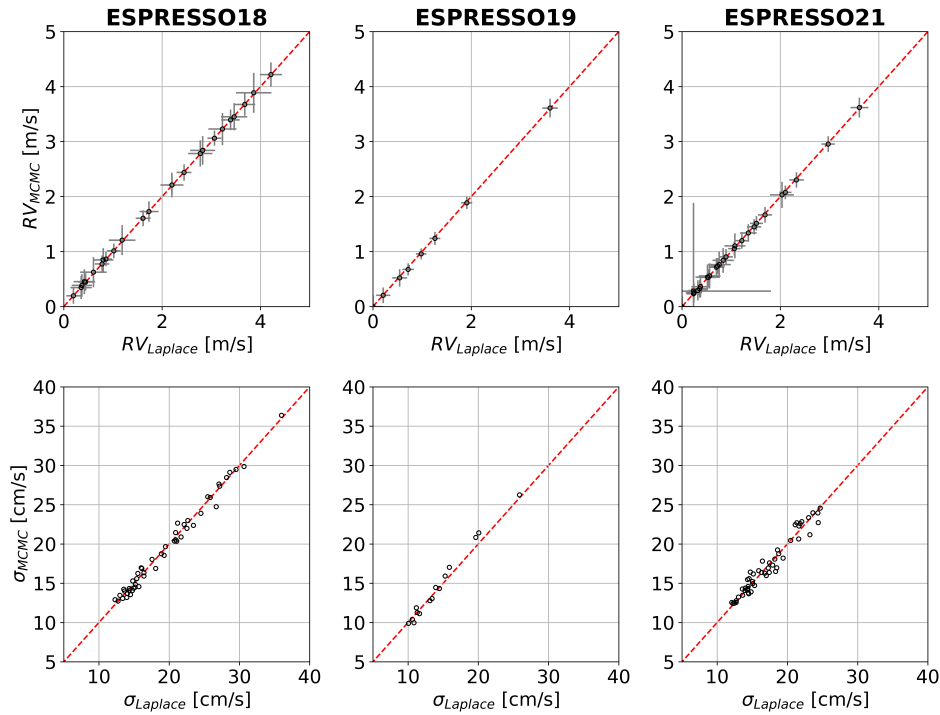


FIGURE 2.13: Differences between the RV posterior distribution as characterized through the MCMC methodology and the Laplace approximation for the full RV dataset of Proxima Centauri, as used in (Faria et al., 2022). **Top:** Difference in the RV expected value. **Bottom:** Difference in the standard deviation of the RV posterior distribution.

2.4 Benchmarking the s-BART pipeline

With the development of the s-BART pipeline there was a need to confirm that the results are comparable to those from other template matching algorithms and that our core assumptions for the stellar template hold out when tested with real data. For this effect, in Silva et al. (2022) we compared the pipeline with the CCF method, SERVAL (Zechmeister et al., 2018), and HARPS-TERRA (Anglada-Escudé & Butler, 2012). Throughout this Chapter we shall adopt the following naming convention: ‘classical’ will refer to results obtained with the χ^2 methodology discussed in Chapter 2.2, ‘s-BART’ to refer to those from the semi-Bayesian methodology (coupled with the Laplace approximation) from Chapter 2.3 and, ‘DRS’ to those from the CCF provided by the instrument’s pipeline.

2.4.1 Benchmarking with HARPS

The first real test of s-BART was a comparison of the obtained RVs with those from SERVAL and HARPS-TERRA. As the latter pipeline is not, to the best of our knowledge, publicly available, we searched the literature for a suitable target. Our selection was Barnard’s star

(GJ699), where we obtained 22 HARPS observations of Barnard’s star (GJ699), collected between 2007-04-04 and 2008-05-02, with program ID 072.C-0488(E). Figure 2.14 presents the results obtained for HARPS-TERRA, SERVAL and our two template-matching methodologies. The HARPS-TERRA time-series was extracted from Table 6 of [Anglada-Escudé & Butler \(2012\)](#)^{*} and the SERVAL time-series was derived by using the most recent, at the time, public version of SERVAL[†]. For comparison purposes the RV estimates had their own mean value subtracted, to avoid issues with the reference RV of each method. A visual comparison of the different time-series allows to verify that the RV measurements show the same trends in all cases.

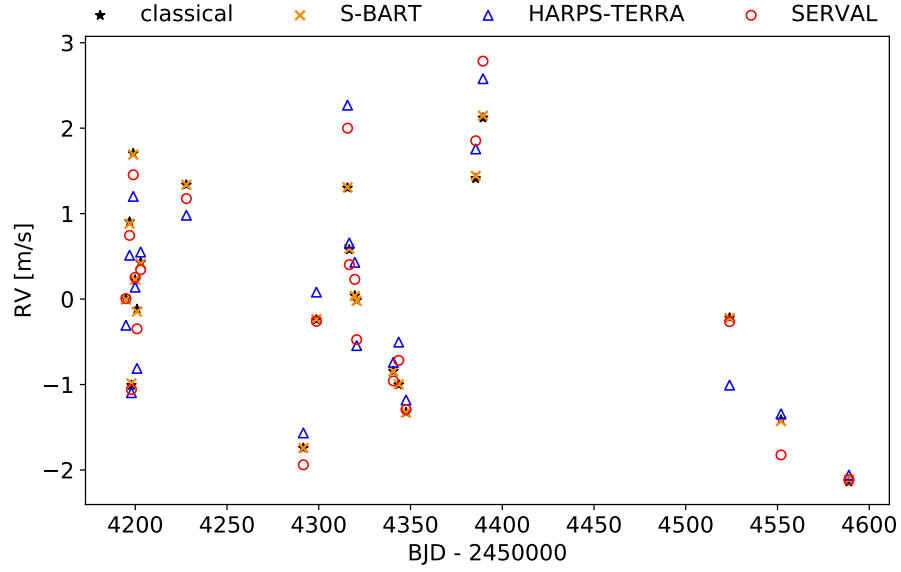


FIGURE 2.14: RV time-series for 22 observations of Barnard’s star, corrected for the secular acceleration of the star. The black stars and orange crosses are RV estimates obtained with the classical and **s**-BART methodologies, respectively, with its own mean RV subtracted (for comparison purposes). The blue triangles and red dots are HARPS-TERRA and SERVAL RV estimates, respectively, also with their own mean RV subtracted. Image adapted from [Silva et al. \(2022\)](#).

Table 2.4 summarizes the performance of the different techniques through the standard deviation of the RVs, which provide a measure of their scatter. For completion purposes, the RV scatter reported in ([Zechmeister et al., 2018](#)) is also included, as SERVAL-PAPER. We find that both methodologies from this thesis reach the meter per second precision on HARPS data, achieving smaller scatter than with the HARPS-TERRA and SERVAL pipelines. The CCF-based HARPS pipeline leads to more scattered RV estimates, as expected given

^{*}We used the RVs obtained with the entire spectrum.

[†]<https://github.com/mzechmeister/serval>; commit d31a918

that Barnard’s star belongs to the M spectral class. We refrain from comparing the estimates for the RV uncertainties as the different template-matching algorithms use different estimators for their calculation, thus making a direct comparison not a straightforward endeavour.

TABLE 2.4: Time-series RV scatter obtained with different template-based methodologies when applied to Barnard’s star. The HARPS-TERRA and SERVAL-PAPER results were obtained from [Anglada-Escudé & Butler \(2012\)](#) and [Zechmeister et al. \(2018\)](#), respectively. The SERVAL and DRS-HARPS results were obtained by using one of the publicly available version of SERVAL (Commit d31a918) and with the HARPS pipeline, respectively. Finally, the results obtained with our methodologies are labelled as classical and s-BART. Table adapted from [Silva et al. \(2022\)](#).

Method	std RV [m s^{-1}]
DRS-HARPS	1.51
HARPS-TERRA	1.22
SERVAL-PAPER	1.30
SERVAL	1.28
classical	1.14
s-BART	1.14

We find that our results are slightly less scattered than those obtained with other template-matching methods. Despite the small differences, they are concordant with the others, suggesting that our methodologies are working as intended. However, it is also important to stress out that the comparison of the RV scatter does not allow to directly compare the performance of different pipelines. As the different algorithms are sensitive to different effects on the data, it is not clear what a smaller scatter entails. Furthermore, even if the larger scatter is a consequence of a method being more affected by activity, it might also lead to an easier modelling during the analysis of the RVs time-series.

2.4.2 Benchmarking with ESPRESSO data

Following the good results obtained for HARPS (refer to the previous Chapter) the template-based algorithms were then applied to an instrument which, at the time, had no public implementation of a template matching algorithm. In this sub-Chapter we thus present the results from applying both template-based methodologies to the ESPRESSO dataset, comparing them against the CCF method implemented in the DRS. For this initial comparison, as presented in [Silva et al. \(2022\)](#), we used a sub-sample of the ESPRESSO WG1 dataset, reduced with version 2.2.8 of the official pipeline*. In order to perform this comparison the selected sample was comprised of 33 targets, where 16 are M-type stars, 13

*<https://www.eso.org/sci/software/pipelines/>

K-type stars and 4 G-type stars. In total, approximately 1000 observations, distributed between ESPRESSO18 and ESPRESSO19, were used, as specified in Table 2.5.

TABLE 2.5: Number of observations, of each spectral type (ST), obtained before and after ESPRESSO’s fiber link upgrade for the WG1 subsample.

ST	Targets	ESPRESSO18	ESPRESSO19	Total
M	16	176	133	309
K	13	249	158	407
G	4	251	79	330

The construction of the stellar template does not include any observation that has an airmass greater than 1.5, and the only targets that were considered were those in which we have at least 5 valid observations to construct the model. The choice of the latter criterion is motivated in Chapter 2.4.2.1.

2.4.2.1 The impact of the number of observations in the template

The first step towards the benchmark of **s-BART** on the ESPRESSO sample is to evaluate the conditions under which the stellar template is created. For this effect, we assess the impact of the number of observations, included in the stellar model, on the final RV estimates. We selected 24 ESPRESSO18 observations of an M-type star, from which the first 11 were set aside to be used in the construction of stellar templates, whilst the other 13 served to evaluate the performance of the templates. It is also important to refer that there is no mixture of observations between the subset of data that is used to construct the template and the one that is used to estimate the RVs. If this was not the case, the templates constructed from a low number of observations would be too similar to the data it is being compared to, which will impact the overall performance.

The 11 observations selected for the construction of the template cover a BERV region that starts at 25 km s^{-1} , in the first observation, and ends at -19 km s^{-1} in the last one. Different stellar templates are constructed, by sequentially selecting observations based on their BERV, starting from largest to smallest. With each template, RVs were extracted for the 13 observations and their scatter and uncertainties were evaluated.

In Figure 2.15 the results of the comparison are shown, showing an improvement in both scatter and median uncertainty by both methodologies with the increase of the SNR of the template. It is also possible to see that the final RV scatter plateaus after the

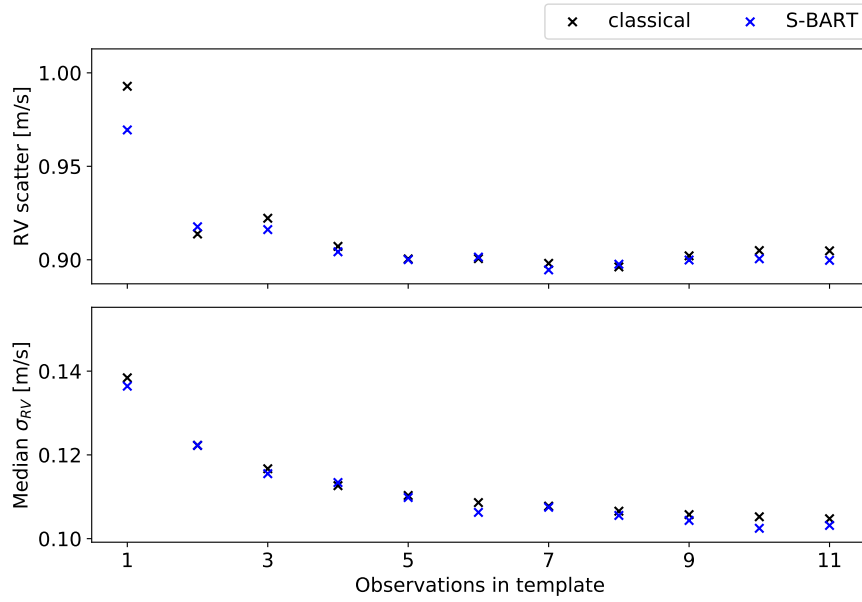


FIGURE 2.15: Evolution of the RV scatter (top pannel) and median uncertainty (bottom pannel) with respect to 13 ESPRESSO18 observations of a M-type star, as a function of the number of spectra used to construct the template. In black we have the results from the classical methodology and, and in blue those from the **s**-BART methodology. Image adapted from [Silva et al. \(2022\)](#).

inclusion of 5 observations, which hence becomes the recommended minimum number of observations to include in the generation of the stellar template. However, increasing the number of observations will still result in an improved RV precision.

2.4.2.2 Evaluating the validity the RV uncertainties

The next step in the evaluation of **s**-BART is to evaluate if the uncertainties are consistent with the information that is present in the data. To do so, we simulated 100 observations following the recipe from Appendix C and applied both the classical, **s**-BART and CCF methods. In total, 2 observations of an M5 star were used as reference, one from ESPRESSO18 and another from ESPRESSO19. For each dataset, the uncertainties of the reference spectrum were compared with the scatter and median uncertainty of the series. Table 2.6 shows an agreement inside each methodology, where the σ_{RV} of the original spectrum is similar with the mean uncertainty and RV scatter of the simulated datasets. It is also clearly seen the gain in precision for the template-based methodologies, as their metrics are approximately half of those obtained through the CCF

To complement this analysis, we also made a comparison with the expected RV precision ([Bouchy et al., 2001](#)), as implemented in **eniric** ([Neal & Figueira, 2019](#)), revealing that

TABLE 2.6: Results, in cm s^{-1} , of the application of template matching and CCF to two simulated datasets where white noise was injected: one built from an ESPRESSO18 and another from an ESPRESSO19 observation, both from the same M5 star. Table adapted from [Silva et al. \(2022\)](#).

ESPRESSO	method	σ_{RV} reference	Simulated data	
			std	median σ_{RV}
18	classical	18.9	11.8	13.0
	s-BART	12.8	11.8	12.9
	CCF	24.6	26.2	24.6
19	classical	15.1	14.0	14.9
	s-BART	14.8	14.1	14.9
	CCF	28.0	29.3	28.0

the median **s-BART** uncertainties from each spectral type are just a few cm s^{-1} above the corresponding photon noise limit.

2.4.2.3 Comparison of the RV scatter and precision

Following the positive tests found on the ESPRESSO dataset, we now compare the RV scatter and uncertainties obtained through the classical and **s-BART** methodologies. In order to achieve this, we apply both of our template matching algorithms to the roughly 1000 observations that compose our ESPRESSO sample (previously detailed in Table 2.5).

The outcome from this comparison is presented in both Figure 2.16 and Table 2.7, which is compatible with previous implementations of template matching. The largest improvement is seen for the M-type stars, where the template-based methods consistently provide smaller RV scatters. This decrease in RV scatter is clearer within the ESPRESSO19 dataset ($\sim 10\%$ smaller) than in the ESPRESSO18 one ($\sim 8\%$ smaller). A similar decrease is also found in the K-type stars, albeit smaller, of $\sim 4\%$ for the two datasets. Lastly, for G-type stars, we find an increase of $\sim 0.5\%$, a result that should be taken with caution due to the very limited sample size of this spectral type. It is also important to mention that the very similar RV estimates of the template-matching methodologies were expected, as both use the same information, i.e., the same spectral regions, and the same model (the template).

If we now turn our attention to the RV uncertainties, we find that the difference between the median value of **s-BART** and the classical approach is below the cm s^{-1} mark. This can be better seen in Figure 2.17, with the histograms of the individual RV uncertainty estimates for the observations, separated by spectral type, for each ESPRESSO dataset.

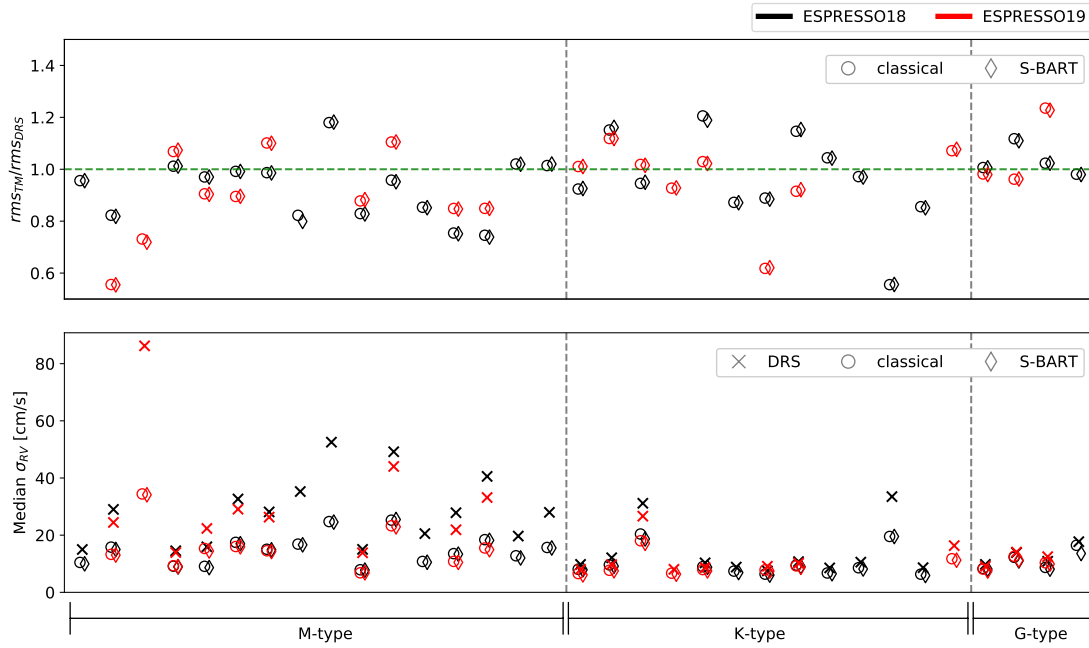


FIGURE 2.16: Comparison of the results obtained with our template-matching methodologies and with the CCF-based ESPRESSO DRS, for ESPRESSO18 (black) and ESPRESSO19 (red) observations of the selected targets. **Top panel:** Ratio of the rms in the template-matching and DRS RV time-series. The results from the classical and s-BART methods are represented by circles and diamonds, respectively. **Bottom panel:** Median RV uncertainty for each target, as computed by the DRS (crosses), the classical approach (circle) and the s-BART method (diamonds). Image adapted from Silva et al. (2022)..

TABLE 2.7: Mean ratio between the scatter of the RV time series as derived through the two template-based methodologies and the CCF-based ESPRESSO DRS, separated by spectral type, ST, and methodology. Table adapted from Silva et al. (2022).

Dataset	Method	M-type	K-type	G-type
ESPRESSO18	classical	0.928	0.960	1.032
	s-BART	0.923	0.961	1.029
ESPRESSO19	classical	0.894	0.964	1.060
	s-BART	0.893	0.964	1.057

The comparison of the uncertainties is also summarized in Table 2.8. We see that for M-type stars both template-matching implementations yields a median RV uncertainty $\sim 13 \text{ cm s}^{-1}$ smaller than with the ESPRESSO DRS, corresponding to almost half of the median CCF RV uncertainty. For K- and G-type stars the gain in the median RV precision, in comparison with the CCF, is below 5 cm s^{-1} . Once again, the improvements fall in line with those obtained with TM algorithms in other spectrographs. This comparison, however, must be taken with caution as the CCF profile is constructed through the combination of lines with different profiles. This means that that RV content is slightly diluted and

the error propagation is not equivalent to one that is done in the context of comparing template and spectra.

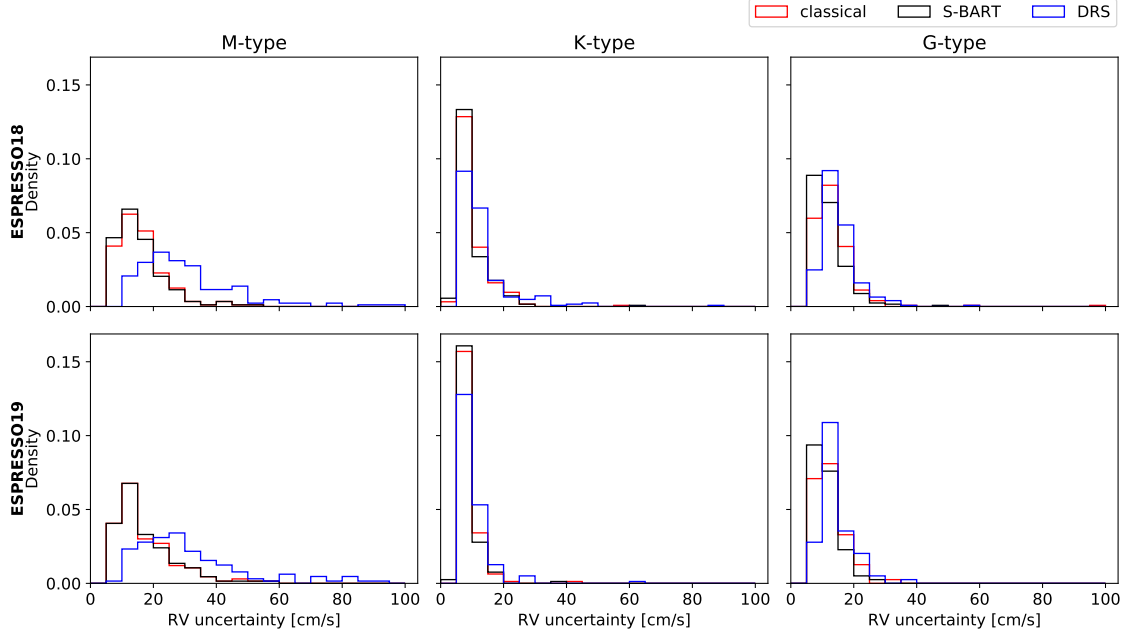


FIGURE 2.17: Comparison of the RV uncertainties obtained for all ESPRESSO18 (top panel) and ESPRESSO19 (bottom panel) observations used in Figure 2.16, separated by spectral type. In blue we have the RV uncertainty estimated by the DRS, in black and red those estimated through the **s**-BART and classical methodologies, respectively. Image adapted from Silva et al. (2022).

Despite the impressive results, it is also important to note that the sample from ESPRESSO’s WG1, defined in Hojjatpanah et al. (2019), is mainly composed by quiet stars, i.e., ones without high levels of stellar activity. From this analysis, it is thus impossible to infer the performance of **s**-BART in active stars with this stellar sample. Even with a sample with more active stars, the evaluation of the performance is not straightforward, as it would involve the modelling of both the activity signal and a unknown number of keplerian signals.

2.4.2.4 Nightly Zero Point (NZP) variation

Following the extraction of RVs from such a large number of observations collected with ESPRESSO, we can attempt to estimate the achievable RV precision of our instrument. To do so, we analyse the nightly zero point (NZP) of each RV estimation procedure, following the methodologies presented in Courcol et al. (2015), Tal-Or et al. (2019) and described in detail in Appendix B. To enforce a balance of the number of observations between pre and post fibre link upgrade data, we do not consider the G-type stars, as they only consist

TABLE 2.8: Comparison of the median RV uncertainties, in cm s^{-1} , as obtained through the three methodologies. The values were derived for the two ESPRESSO datasets (ESPRESSO18 and ESPRESSO19) used in Figure 2.16.

Dataset	Method	M-type	K-type	G-type
ESPRESSO18	DRS	26.7	10.4	13.4
	classical	14.8	8.6	10.6
	s-BART	14.3	8.1	10.6
ESPRESSO19	DRS	27.5	9.0	12.5
	classical	14.7	7.6	10.9
	s-BART	14.4	7.2	10.1

of 4 targets, with the majority of observations taken before the fibre link upgrade (Table 2.5). After this cut, we still find that the ESPRESSO18 observations represent $\sim 60\%$ of the overall sample.

The NZP time-series is shown in Figure 2.18, as derived from the RVs obtained with the ESPRESSO DRS as well as with our two template-matching methodologies. A visual comparison of the NZPs from the different estimation methods shows no significant differences between them, albeit a smaller scatter seems to be found for the ESPRESSO19 dataset. If we now compute the weighted standard deviation of the NZPs, as show in Table 2.9, we now find a smaller scatter in the template-based methods, as compared with the CCF. Interestingly, the classical approach leads to the smaller scatter, especially in the ESPRESSO18 data. This Table also allows to see that the weighted variability of the ESPRESSO19 dataset is $\sim 10 \text{ cm s}^{-1}$ smaller than the one from ESPRESSO18, independently of the methodology employed to extract RVs.

TABLE 2.9: Weighted standard deviation, in cm s^{-1} , of the derived NZPs for observations obtained before and after the 2019 ESPRESSO fiber link upgrade. The results were obtained with the data from M and K-type stars, with RVs estimated through the three different methodologies. Table adapted from Silva et al. (2022).

Method	ESPRESSO18	ESPRESSO19
ESPRESSO DRS	76.0	59.9
classical	69.6	57.2
s-BART	71.0	57.1

As a final note, it is important to keep in mind that the RVs used in the estimation of the NZPs include the contribution of stellar activity, planetary signals, and photon noise. Thus, the results must not be taken as a estimation of the stability of the instrument, but rather as an upper bound for the achievable stability of ESPRESSO. Nonetheless, it is still

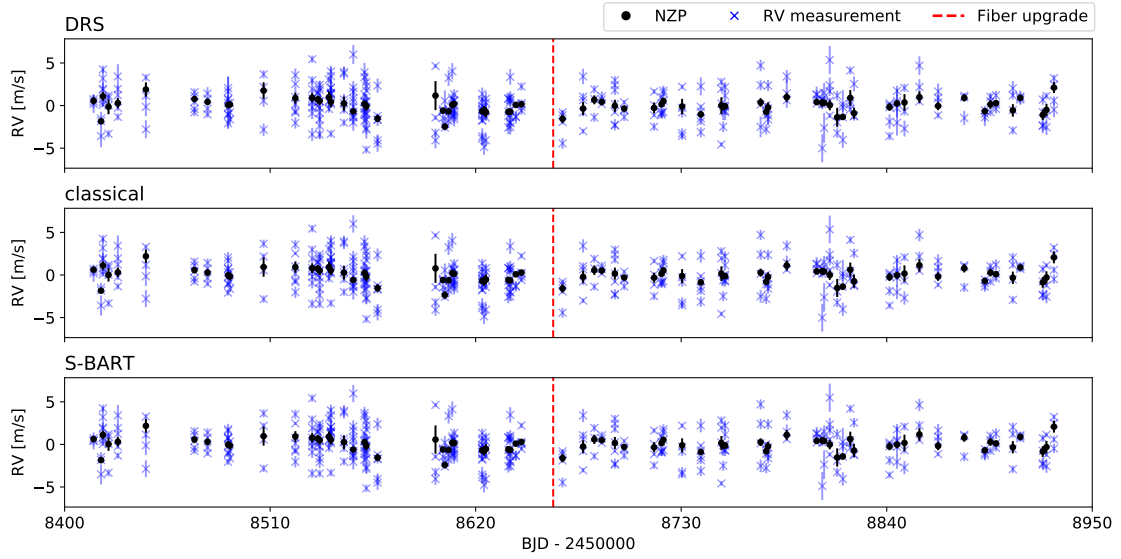


FIGURE 2.18: Nightly Zero Point (NZP), black dots, for each night when at least 3 different targets were observed. The date of the fiber link upgrade is highlighted with a dashed red line. The zero-centered RV of each target, observed in the night, is represented with the blue crosses. As derived with results obtained with the following methodologies: ESPRESSO DRS (**Top**); classical (**Middle**); **s**-BART (**Bottom**). Image adapted from [Silva et al. \(2022\)](#).

noteworthy that, even under such circumstances, we find a NZP scatter below the meter per second mark, both before and after the fiber link upgrade.

2.5 Large-scale application of s-BART

Following the publication of the paper and the promising initial results from Chapter 2.4, the s-BART pipeline has continued being developed. The algorithm was adapted to different instruments and applied to larger datasets, whose results will be described in this sub-Chapter. In particular, we used a larger number of observations for the WG1 sample of ESPRESSO (Chapter 2.5.2) adapted the code to handle CARMENES data in the KOBE survey (Chapter 2.5.1) and started an adaptation to the HARPSN spectrograph (Chapter 2.5.3).

2.5.1 The KOBE survey from CARMENES

The first survey in which s-BART was included was the KOBE (Lillo-Box et al., 2022) survey, from CARMENES. The main goal of this survey is the detection of planets in the Habitable Zone of K-dwarfs, thus proposing to observe 50 K-dwarf targets with the expected yield of 1.68 ± 0.25 planets per star, with a minimum semi-amplitude of the planets being 1.08 m s^{-1} . To account for the stability of the instrument, the survey also observes standard stars for calibration purposes, through the calculation of NZPs (see Chapter B for details) which are then used to correct the RV time-series.

Currently, we have available approximately 2900 observations and the survey is extracting RVs with 3 different pipelines: SHAQ which provides CCF RVs, and both s-BART and SERVAL for template matching RVs. The criterion used to select the RV pipeline of this survey was the performance on the standard stars, after the NZP correction has been applied to them.

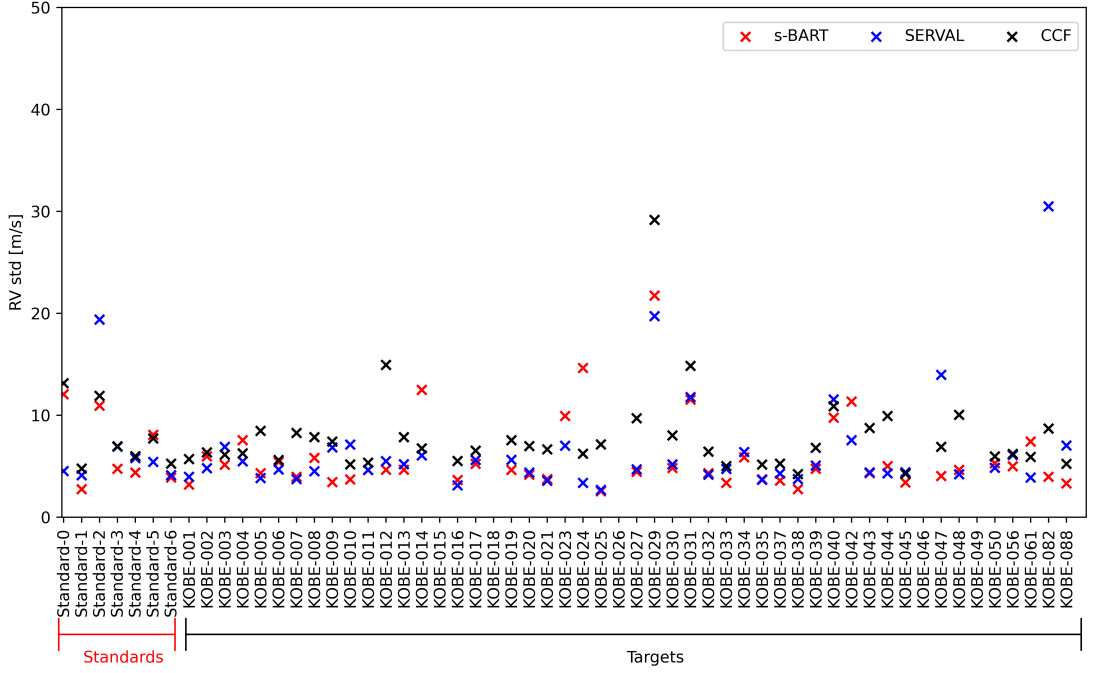


FIGURE 2.19: RV scatter for the KOBE sample for the CCF method (black), **SERVAL** (blue) and **s-BART** (red), highlighting the standard stars and the science targets. The scatter was computed after the correction of the NZP, for both standard and science stars.

From Figure 2.19 we see that for the standard stars (left side of the image, highlighted with a red line), the RV scatter is usually smaller for **s-BART** than in the alternatives. This smaller RV scatter has been consistent since the early stages of the survey, which lead to the selection of **s-BART** as the main pipeline for RV extraction. If we now turn our attention to the NZP-corrected science targets (right side of the image), we find that in most cases, **s-BART** also presents the smaller RV scatter, however it is not always. However, it is important to note that a closer look at the NZPs of all methods reveals corrections at the level of several m s^{-1} . This is clearly a worrisome value, especially when coupled with the caveats of the NZP correction and considering the amplitude of the expected signals.

Despite the large sources of variability in the RV time series, the preliminary analysis of **s-BART** RVs has started to reveal, in some targets, hints of periodic signals near the Habitable Zone. One available option to mitigate this variability would be to simultaneously model the RVs from the science star and those from the standard stars. However, such endeavour lies outside the scope of this thesis and it is currently being developed.

2.5.2 Larger ESPRESSO sample

Since the publication of **s-BART** paper, the ESPRESSO GTO dataset has both increased (due to new observations) and was entirely re-reduced with a new version of its official pipeline. This new version introduced long-term flux corrections and a new set of CCF masks, leading to changes in both the CCF RVs and the stellar spectra. The new CCF masks not only change the number and location of lines, but also change their weight.

Such changes raised the need of a new evaluation of **s-BART** performance on the newer dataset, which now comprises 3439 observations spread across 37 different targets. Similarly to the previous dataset, we consider independent subsets of data in ESPRESSO, linked to the different instrumental interventions in the instrument (ESP18, ESP19 and ESP21). For each subset of observations we computed RVs, creating different stellar and telluric templates for each.

TABLE 2.10: Description of the ESPRESSO dataset, alongside a comparison of the ratio of the rms and median uncertainties of **s-BART** and the CCF implemented in the official pipeline.

ST	Dataset	Number of stars	Total number of stars	Number of Observations	$\frac{rms_{sbart}}{rms_{CCF}}$	σ_{RV}^{sbart} [m/s]	σ_{RV}^{CCF} [m/s]
M	ESP18	18	20	224	0.949	0.143	0.244
	ESP19	11		163	0.955	0.160	0.274
	ESP21	12		637	0.943	0.147	0.270
K	ESP18	11	13	253	0.958	0.082	0.108
	ESP19	9		196	0.990	0.075	0.096
	ESP21	9		549	1.029	0.086	0.110
G	ESP18	5	4	276	1.007	0.098	0.126
	ESP19	4		271	1.020	0.089	0.117
	ESP21	5		870	1.002	0.095	0.120

Table 2.10 presents an overview of the final sample, as well as a direct comparison with the RVs of the CCF. Similarly to the results from Section 2.4.2, we find a smaller RV scatter and improved precision for the different spectral types, albeit smaller than before. Whilst our initial results lead to a decrease of $\sim 10\%$ on M-type stars, the current improvement is closer to $\sim 5\%$, which is attributed to the improved CCF masks used within the new ESPRESSO pipeline. However, we still encounter a large gain in RV precision in the M-type stars (almost double, at the level of 10 cm s^{-1}) and a small improvement on K-type stars ($\sim 2 \text{ cm s}^{-1}$). The small sample size of the G-type stars doesn't allow to draw meaningful conclusions, but appears to lead to comparable results.

Since the publication of the paper where **s-BART** is described, its ESPRESSO RVs have been present in two peer-reviewed publications, albeit a few more are currently being prepared:

- Proxima d: [Faria et al. \(2022\)](#) proposes a candidate planet around our nearest neighbour, *Proxima Centauri* (M-type star). This sub-Earth ($M \approx 0.25 M_{\oplus}$) with a very short-period planet ($P \approx 5.2$ days) was detected with ESPRESSO RVs. The analysis was made through a simultaneous GP fit to both **s-BART** RVs and the FWHM of the CCF, as an activity proxy. To probe the planetary nature of this signal, we also computed **s-BART** RVs in three wavelength bins with no overlaps, where it was possible to independently detect the planetary signal, and ensuring the achromatic nature of the signal (and its keplerian nature).
- GJ1002: [Mascareño et al. \(2023\)](#) performed the main analysis of this system with SERVAL, but the results were also reproduced by the author using **s-BART** RVs;

2.5.3 Early results with HARPS-N data

The motivation behind the adaptation of **s-BART** to HARPS-N was a major update in the DRS of the instrument, as the new pipeline is an adaptation of the ESPRESSO one. The data from the old pipeline of the instrument had RVs from both the CCF and TM, through HARPS-TERRA. With the advent of the new DRS, the HARPS-N GTO gave preference to an adaptation of **s-BART**, rather than HARPS-TERRA. During the adaptation stage, **s-BART** was made to accept not only data from the new pipeline, but also the old one, allowing us to draw direct comparisons between our RVs in the two versions of the pipeline, as well as validate them with HARPS-TERRA.

Before a large scale application of the algorithm, we started by selecting one M-type target whose data had been reduced with both the old and new DRS versions. Then, we extracted **s-BART** RVs for both versions and HARPS-TERRA RVs for the old DRS data, allowing us to validate the implementation.

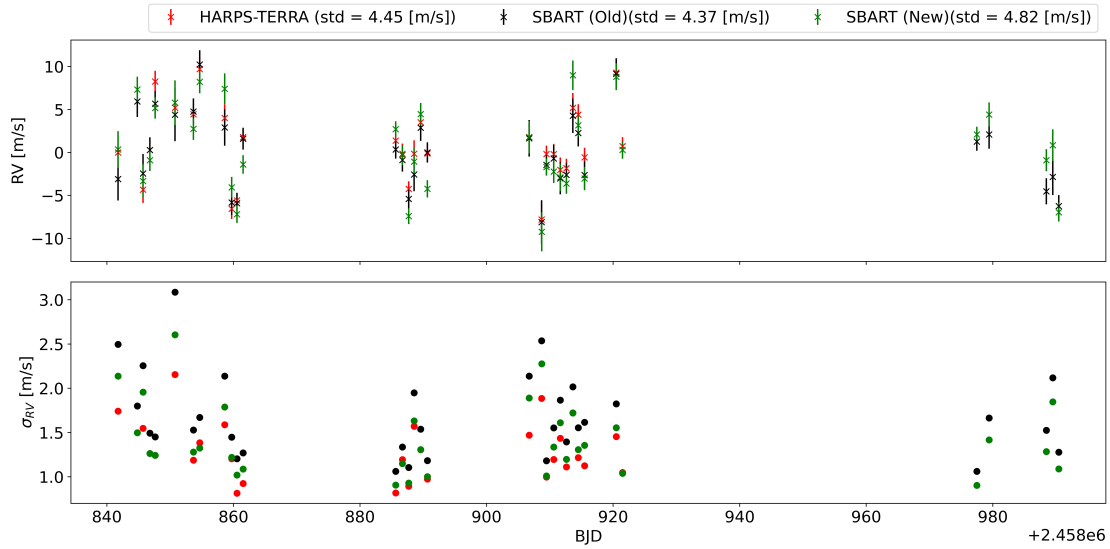


FIGURE 2.20: Comparison of the **s**-BART RVs from one HARPS-N target, using data from the old (black) and new (green) DRS versions. For comparison purposes, HARPS-TERRA RVs are also shown (red), as obtained with the old pipeline.

From Figure 2.20 we see that there is an increase in the RV precision (bottom panel) with the new DRS, as compared with the old one. However, the scatter of **s**-BART RVs is $\sim 40 \text{ cm s}^{-1}$ larger in the newer dataset, with a few observations whose RV are separated by more than 1σ . If we also look at the HARPS-TERRA RVs, which do not exist for the entire dataset, we find RV scatter similar to the one from **s**-BART applied to the same spectral data, albeit with smaller uncertainties. Similarly to before, a direct comparison of the uncertainties associated with each observation is not straightforward as the two pipelines do not estimate them in the same manner. Despite the small differences, we find that **s**-BART implementation is yielding comparable results to the previous TM pipeline of the instrument, allowing us to proceed with the analysis of the GTO sub-sample reduced with the new DRS.

From the GTO dataset we selected a total of 2554 observations from 30 stars of M, K, and G spectral types. For each target we created a transmittance spectra through *TelFit* and extracted RVs with **s**-BART. The overall results of this analysis are presented in Table 2.11, following the same structure as the analysis from the ESPRESSO dataset (see Chapter 2.5.2).

The overall comparison to the CCF, performed through **s**-BART, is similar to what we found before. In terms of the RV scatter, we find a decrease of about 5% in M-type stars and $\sim 2\%$ in the K-type stars, comparable with the results of the newer version of the DRS applied to ESPRESSO. Once again, the **s**-BART methodology yields an improved RV

TABLE 2.11: Description of the subsample from the HARPS-N GTO that was used to benchmark the adaptation of **s-BART**, coupled with a comparison of its RVs with those from the CCF.

ST	Number of stars	Number of Observations	$\frac{rms_{sbart}}{rms_{CCF}}$	σ_{RV}^{sbart} [m/s]	σ_{RV}^{CCF} [m/s]
M	16	1445	0.957	1.065	1.642
K	12	973	0.983	0.501	0.621
G	2	136	1.044	0.492	0.506

precision, reaching $\sim 50 \text{ cm s}^{-1}$ on M-dwarfs and 10 cm s^{-1} on the K-stars. For the very limited sampler of G-type stars, we report a minor gain in the RV precision, and similar behaviour to the DRS.

The validation that **s-BART** is performing as expected after being adapted to handle HARPS-N data allowed us to start the process of fully automatizing the data reduction of this instrument, in collaboration with the *GTO* consortium of the instrument. Furthermore, and in the same context of the *GTO* program, there are currently two papers in preparation using **s-BART** RVs, with more expected to follow in the future.

2.6 Limitations and caveats of the model

Despite the impressive results that have been obtained with **s-BART** – as described in Chapters 2.4 and 2.5 – it still introduces a fair number of assumptions, alongside a lack of a simultaneous modelling of the telluric features of each observation. In this Chapter we discuss some of the largest caveats and lacking features of **s-BART**, discussing possible avenues for a solution.

First, the largest assumption of **s-BART** is that it ignores the sharing of information between the model and the data to which it is compared. Even though the high SNR is expected to drown the contribution of individual observations, this is not fully compatible with a Bayesian framework. This means that our RV extraction and estimation of associated uncertainties is not as robust as we would like it to be. The most straightforward solution to this problem would be to reserve a set of spectra solely for the purpose of constructing the template, but at the cost of losing RV estimates at the times those spectra were acquired. A superior alternative is to discard the usage of a stellar template, directly using a probabilistic framework to model the stellar spectra, as implemented to some degree in *wobble* (Bedell et al., 2019) and, using Gaussian Processes, in *GRACE* (Rajpaul et al., 2020). This will be explored in Chapter 4.

Second, the algorithm shares a flaw that is intrinsic to the framework of template matching: we assume a time invariant model of the stellar spectra, in the form of the stellar template. We know that the activity level of stars changes through time, impacting the shape and position of spectral lines, as discussed in Chapter 1.2.2.2, which is not compatible with this formulation of the spectral model. Furthermore, we also know that the presence of stellar activity and instrumental systematics will introduce chromatic effects on the stellar spectra, introducing a wavelength dependence of the measured RV. As we have discussed during the formulation of the model, **s-BART** ignores this issue in an attempt to closely follow the expected behaviour of RV shifts from keplerian sources. However, it is unclear at which level this chromaticity will impact the gaussianity of the posterior distribution and the quality of our Laplace approximation to it. From here, **s-BART** can be expanded in different directions: i) we either perform a simultaneous modelling of the RV and an activity model; ii) we directly model the variations of the flux; iii) some form of data selection, namely at the spectral line level, e.g., [Cretignier et al. \(2020a\)](#), [Dumusque \(2018\)](#), given that correctly modelling such complex effects seems a much more daunting task. The first approach would provide extra flexibility to the model but it would still be limited by restrictions associated with the construction and usage of a stellar template. The second solution would in fact call for a complete re-structuration of **s-BART** itself as we would need a different, non-static, stellar template that would need to be adjusted to each observation. Furthermore, there isn't a straightforward solution to model the flux variations, over time, in a computationally efficient manner. The third solution would involve the development of a criterion, within a Bayesian framework, capable of identifying the spectral regions affected by tellurics and activity. One possibility is the usage of a Bayesian data re-weighting scheme, e.g., ([Yixin Wang and Alp Kucukelbir and David M. Blei, 2017](#)). However, this would increase the number of free parameters in the model – as each weight would have to be selected by the model – and, consequently, increase the computational cost of the RV extraction.

Third, as discussed in Chapter 2.2, **s-BART** depends on the classical template matching implementation for the selection of the spectral orders to use. As discussed in the Chapter, we are currently using those for which the classical template-matching procedure was able to satisfy the RV convergence criteria. However, it is likely that we could be discarding more (or less) information than needed. One possible solution to this problem could be selection of orders through a maximization of the information gain between the (Gaussian)

RV posterior and the (uniform) RV prior, using the Kullback-Leibler (KL) divergence (Kullback & Leibler, 1951). In this case it would be equivalent to minimizing the RV uncertainty (standard deviation) through the rejection of spectral orders. Similarly to what is done in the classical procedure, this approach would have to select the orders for all available observations simultaneously, but now such selection involves the full RV results, not only those calculated at the level of the spectral order. Consequently, for targets with a large number of observations this procedure would not be applicable.

Lastly, the approximation of the RV posterior probability distribution using the Laplace approximation might be sub-optimal under certain conditions, as previously discussed in this Chapter. This method only uses the information around the MAP estimate to build the approximation and, consequently, is unable to account for any skewness in the posterior. To mitigate such issues, one could either run the MCMC or move towards a more complex technique to approximate the posterior distribution. One such example is Variational Inference (for a detailed explanation refer to Blei et al., 2017, Jain et al., 2018), which uses information from the entire posterior to build a more realistic approximation and, consequently, is more robust to skewness in the posterior.

Chapter 3

An intra-night RV trend

During the development and subsequent application of **s**-BART to multiple targets, a transversal issue to template matching techniques was found: if all observations available were taken within the same night, then a quasi-linear trend in RV appears over the night, with the vast majority of nights presenting a negative slope. This issue was found in observations taken with ESPRESSO and HARPS, with different amplitudes reported for different stars, as we shall see later on the Chapter. Furthermore, it seems to be present in different RV pipelines (**s**-BART, our classical implementation and **SERVAL**), with comparison being shown in Figure 3.1 for one night of HD40307, a K2 star observed during ESPRESSO's commissioning in December 2018 (Pepe et al., 2021).

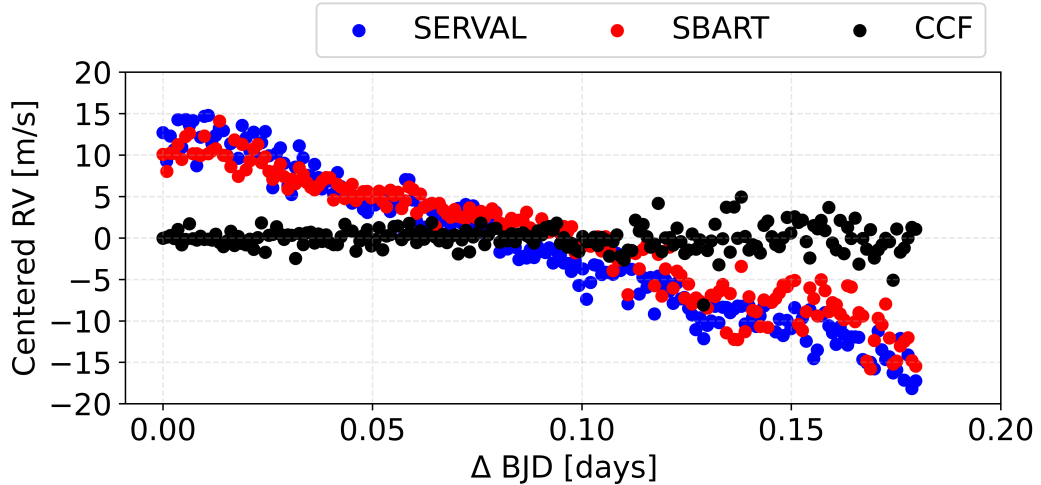


FIGURE 3.1: Mean-subtracted RV time-series for one night of ESPRESSO observations of HD40307, a K2 star observed in December 2018 (Pepe et al., 2021). The CCF RVs (black) are compared against those from **SERVAL** (blue) and **s**-BART (red).

On top of that, if we analyse the metrics from the CCF method that track changes to the shape of the spectral lines, i.e., FWHM, contrast and bisector span, we find no clear variability of the lines that could motivate such a response from the TM algorithm. Nonetheless, this result was expected, as the CCF RVs do not present this systematic effect. At a first stage, we can also rule out stellar activity, as we do not expect to find such a large variation in the activity level of a star, over a single night.

The fact that this trend is found across two instruments, as well in different algorithms, hints towards a systematic effect at play that is biasing our RV estimates. More worrisome is that not only do we have to consider the impact on intra-night observations, but also understand how it translates to RV extraction in multi-night observations. From the on-going RV extraction, as detailed in Chapter 2, we fail to find an evident effect on those cases, but it is possible that a small effect is present.

Understanding the root cause behind this trend is of paramount importance and the main focus of this Chapter. For computational purposes, the RVs were derived only with our classical template matching implementation, with the bulk of the analysis of this Chapter using data from 5 consecutive nights of observations of HD40307 with ESPRESSO. The RV trend is present if we use all observations to construct a single stellar template or if we construct one for each night. In order to ease our analysis of possible effects that occur at timescales of hours, we consider each night to be independent from all others. Thus, for each night of observations, we construct a stellar and telluric template, which are then used for the RV extraction process.

3.1 Differences between single- and multi-night observations

Before delving deeper into the multiple tests that have been carried out in our quest to understand this RV trend, one must first understand the differences between the sets of observations that were collected in a single night, and those that are spread over multiple ones. Henceforth, we shall refer to them as single- and multi-night observations, respectively. These mainly lie in the data products that are used for the calibration of the instrument.

ESPRESSO's DRS makes use of multiple calibration products to transform the raw images into the extracted spectra, e.g., to calibrate the wavelength solution, to correct the CCD response (ESO, 2023). The cadence at which calibration frames are taken varies between nightly to bi-yearly, depending on their type and purpose. Furthermore, different

observing modes will require different calibration products, allowing for the correction of different effects. Usually, fiber “A” of ESPRESSO collects light from the science object, whilst fiber-B will either be placed on the sky (to allow for the correction of its contribution) or on a calibration light-source (FP-laser or a Laser Frequency Comb). It is however important to note that the RV trend was found for data with the two configurations of fiber-B, i.e., simultaneous calibration or sky.

At the moment in which the DRS is applied to the raw images, it selects the calibration frames from the closest night. This means that for our cases, of intra-night observations, the DRS will reduce every observation (of the same night) using the same set of calibration frames. One important step of the DRS extraction is the wavelength calibration, where each pixel in the CCD is attributed to a given wavelength. This is done accordingly to a map constructed from a calibration frame taken with a nightly cadence. After this point, the wavelength solution is corrected from the BERV corresponding to that given moment in time. Within a single night, the BERV correction is mainly accounting for the effect of Earth’s rotation, leading to the smooth variation depicted in Figure 3.2.

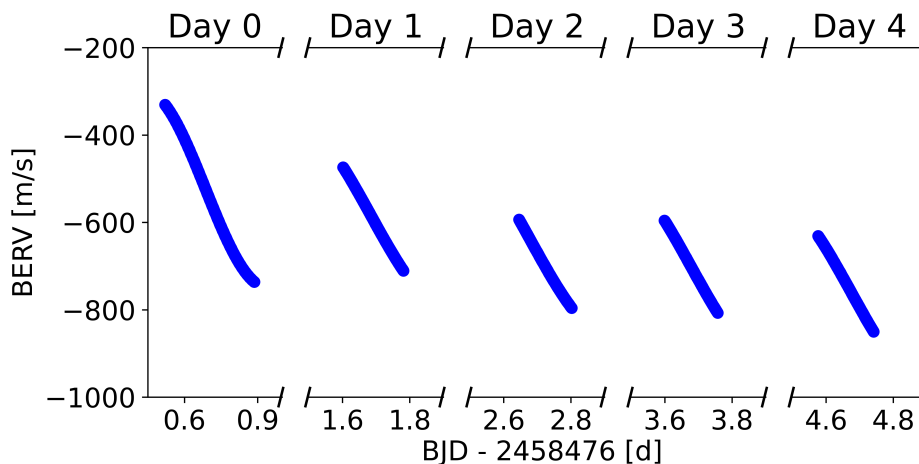


FIGURE 3.2: Variation of the barycentric motion (BERV) in the multiple nights of observations of HD40307.

If we think about the physical consequences of the BERV change during a night, we see that this will translate into a slow drift of the spectral lines over the detector. This physical drift will be driven by the difference in BERV between consecutive exposures which, within a single night, is at a level of the meters per second. In other words, during one single night of observations, the points in which we sample the spectral lines slowly change, as depicted in the ~ 8 hours of HD40307 observations in Figure 3.3.

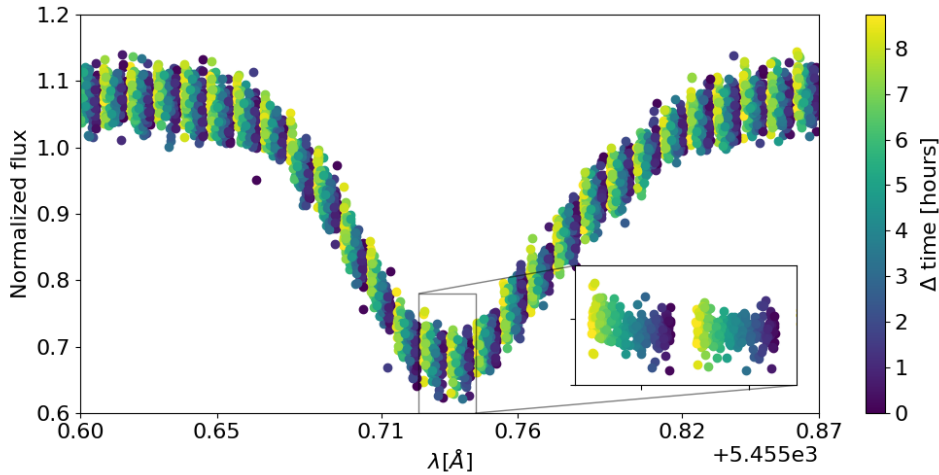


FIGURE 3.3: Impact of the BERV on the location of one selected spectral line on the detector, over 8 hours of observations of HD40307 (the first observation of the night is represented in blue and the last in yellow).

We can see, in this Figure, that during the night there is a slow change of the location where we actually sample the lines, introducing minute effects on our RV extraction, further discussed in Chapter 3.5.1 with simulated data. It is also important to note that the same process is applied in the case of multi-night observations. However, as there is a larger BERV variability we will not find such a smooth evolution of the points at which the spectral lines are sampled. Whilst this change is tied with the BERV correction, it is important to note that a possible, sub-optimal, correction is not expected to be the cause of the trend. If this was the case, the same trend would be measured on the CCF RVs, which is not the case.

3.2 The impact of the atmosphere

At a first glance, a possible contamination from leftover telluric features seemed to be the straightforward explanation for this RV effect. As all of our data is collected within the same night, we will not have a large enough BERV separation to allow them to be averaged out of the spectral template. Thus, this means that the telluric features that are not rejected from the binary mask (refer to Section 2.1.2 for the details) will be kept in the stellar model. If we couple this, with the knowledge that the lines will be shifted in the same direction as the night progresses, one could argue that the RV shift is driven by the leftover, blended, telluric features present in the template. The fact that the RV variation is smaller than the maximum BERV variation in a night, follows from the fact that the

majority of lines are not injected by our atmosphere, thus dampening this effect over an entire spectral order. Three tests were devised to evaluate this impact:

- Evaluate the RV trend when using different thresholds for the masking of telluric features (see Chapter 3.2.1);
- Inject a synthetic transmittance spectrum, shifted by different BERVES, in one observation, which is then followed by the evaluation of the extracted RV (see Chapter 3.2.2);
- Compute the RV time-series from telluric-corrected observations (see Chapter 3.2.3).

3.2.1 Change in the transmittance threshold

A change in the threshold under which we reject atmosphere-induced features leads to a change in the number of telluric features that bypass our rejection. In other words, by relaxing the threshold we will see an increase in both the number and depth of telluric features that are not rejected. Thus, if the RV trend is introduced by leftover tellurics in the stellar template, its effect should increase when the number of telluric features increases. To evaluate this hypothesis, we selected one night of HD40307 observations and constructed telluric masks using seven (7) different thresholds. Then, we extracted RVs whilst using each template, as shown in Figure 3.4.

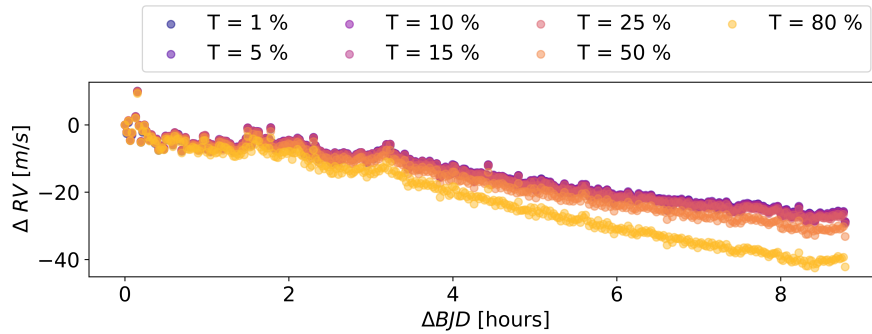


FIGURE 3.4: Impact on the RV trend from template matching after creating a telluric mask with different thresholds. The thresholds were applied to the same transmittance spectrum, generated through *Telfit*.

From Figure 3.4 we see that a change in the transmittance threshold translates into a different RV effect, mainly for the less conservative thresholds (i.e., those that allow deeper features). It is also clear that at a given point, around a threshold of 15%, the impact on the peak-to-peak amplitude of the RV trend is minimal. This result suggests that either

the telluric features are not the driving cause of the trend, or that the *telfit* model is missing a large number of telluric lines. The later hypothesis can be explored with a chromatic analysis of the RVs, which shall be done further ahead, in Chapter 3.3.

3.2.2 Simulating telluric variability

Following the results from a change in the telluric threshold, doubts were raised about our hypothesis of the telluric features driving the RV variability. Thus, to better understand if the “non-averaged out” telluric features are enough to translate into this RV variability, we attempted to simulate this behaviour through the injection of transmittance profiles on the stellar spectra. Our simplistic approach for this comprises a series of consecutive steps:

1. For each observation in one night of observations of a selected target we retrieve the BERV value, alongside the atmospheric parameters as recorded in the ESPRESSO data products. Alongside we construct a *telfit* transmittance profile, using the retrieved values;
2. Generate a set of N – where N is the number of retrieved observations from the selected night – synthetic spectra $\{S_i, \dots, S_N\}$ following the approach specified in Appendix C. The base spectra used in the simulation process is not telluric corrected. However, such features will be fixed at a radial velocity of zero, relative to the stellar lines.
3. Each BERV value is used to Doppler-shift the transmittance profile created with the corresponding observation, thus mimicking the displacement of telluric features throughout the night;
4. The shifted transmittance profile is interpolated (with a cubic-spline) to the wavelength solution of the reference frame, after which it is used to multiply the stellar spectra.

Even though this injected transmittance profile will not reflect the instrumental profile, it should be enough to probe the impact of the intra-night RV variability injected from a small number of moving features. This process was repeated whilst assuming for each synthetic spectra the exact same atmospheric conditions, taken from the first observation of the night. Furthermore, for control purposes, a smaller dataset was constructed without any transmittance profile being injected. This provides us with three synthetic datasets to

which we will apply our template matching algorithm. During this process, we also create a telluric mask following the process laid out in Chapter 2.1.2. This ensures that this test will probe the impact of the telluric features that lie under the threshold selected by our algorithm.

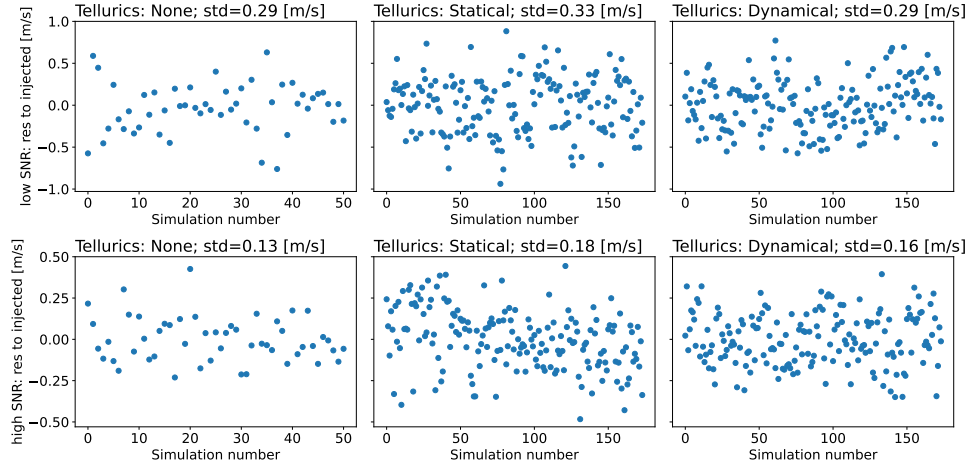


FIGURE 3.5: Residuals of the RV time-series obtained when applying the *classical* TM algorithm to the synthetic datasets generated with *Telfit*, using as a basis observations from HD40307 (top row) and a target with higher SNR (bottom row). The title of each panel provides a measurement of the standard deviation of the RV time-series. **Left column:** RV residuals without injecting any transmittance profile; **Center column:** RV residuals from the dataset where the atmospheric parameters were kept constant; **Right column:** RV residuals from the dataset where the atmospheric parameters were those recorded in ESPRESSO’s data products.

The results from this application are presented in Figure 3.5, where we applied the aforementioned recipe using two observations, from different targets, as the basis. We see that in both cases the RV residuals, from the injection of the transmittance profile, are comparable to those from the dataset where only noise was added. However, it is important to note a small downwards trend on the high-SNR dataset with a synthetic spectrum generated through the same parameters. Despite that, the measured peak-to-peak amplitude is negligible, when compared against the amplitude of the effect measured on real data. Once again, this result hints towards the fact that the telluric features, small enough to bypass the binary mask, will not have a large enough weight to significantly bias the final RV estimate. However, it is important to note that the magnitude of the residuals is higher in the lower-SNR dataset, but no trend can be clearly discerned.

3.2.3 Correction of water features in the data

After the initial tests raised questions on both the validity of the transmittance model and the role played by telluric features, we attempted to correct them whilst using a different underlying model. For this purpose, we collaborated with Romain Allart to apply his telluric-correction tool (Allart et al., 2022) to one of the nights from HD40307. Concurrently, the **s**-BART pipeline was adapted to handle the outputs of this tool and applied the classical template matching approach to this set of observations.

A crucial aspect to consider is that the correction was only applied to *water* features, as described in the paper of the methodology. The other telluric features were masked with a transmittance spectra, generated through *Tapas*, assuming a water-free atmosphere. The change from *Telfit* was motivated by implementation constraints. However, the transmittance spectra of *Telfit* and *Tapas* were compared, revealing the same features in both. To account for intra-night BERV variability, **s**-BART’s telluric mask rejected any possible position of those lines during the year.

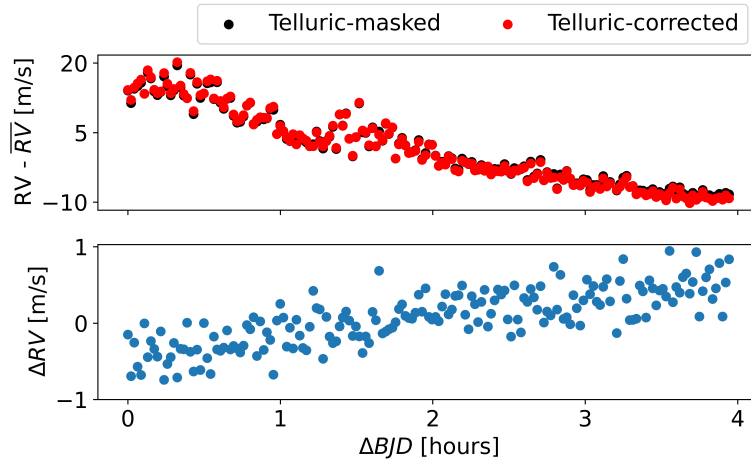


FIGURE 3.6: **Top panel:** Comparison of the **s**-BART RVs from ESPRESSO observations of HD40307, when masking (black) and correcting (red) the telluric features; **Top panel:** Difference between the HD40307 RVs extracted whilst masking the tellurics and those obtained after the telluric correction .

The results from this analysis are presented in Figure 3.6, where we find surprising results. A change from the telluric-masked spectra to the telluric corrected one actually results in an increased slope. One possible explanation for this is a RV bias introduced by residuals in the telluric correction. However, it is not clear how this translates into a quasi-linear effect over the 4 hours of observations. Nonetheless, if we look at the amplitude of the residuals, we find that it is one order of magnitude smaller than the peak-to-peak

amplitude of the RV variation. A comparison of the RV uncertainties associated with each method reveals, as one would expect, a difference. The median uncertainty of the telluric-corrected dataset is $\sim 4 \text{ cm s}^{-1}$ smaller than the masked one. This comes naturally from the usage of a larger amount of RV information due to the larger wavelength region in use.

Summarizing these findings, we reach the same conclusions as in the previous tests: we are either missing lines in the telluric models, or the tellurics are not the root cause. The minute changes between the two time-series do not hint towards the tellurics being the driving force of this effect. Furthermore, as the underlying telluric models are different for the correction and removal of tellurics, it is unlikely that they are both missing the same sets of features.

3.3 Evaluating the chromatic evolution of the trend

An investigation into the wavelength dependence of the RV trend might shed some light into its nature. For this effect, we propose two different tests:

1. The nightly evolution of the RVs in each detector of the spectrograph, coupled with an analysis of the order-by-order RVs. This first test will allow us to probe the wavelength dependence of the variability. If an achromatic variability is found, i.e., similar effect across wavelengths, it hinders the possibility of tellurics, as those are not expected to introduce a uniform impact across the spectral range.
2. The wavelength dependence within each spectral order. This will allow us to understand if the issue has connections with the instrumental profile if we compare the centre of the spectral order with its edges.

3.3.1 A division by the detectors of the spectrographs

The first attempt at searching for a chromatic dependence was to compute individual RV time-series for each of the two detectors (*blue* and *red*). For this, we selected 4 nights of ESPRESSO observations of two targets and 1 night of HARPS observations of a different target. For each, we computed the RV time-series for the spectral orders associated with each detector of the corresponding spectrograph, ignoring all orders from the non-selected one.

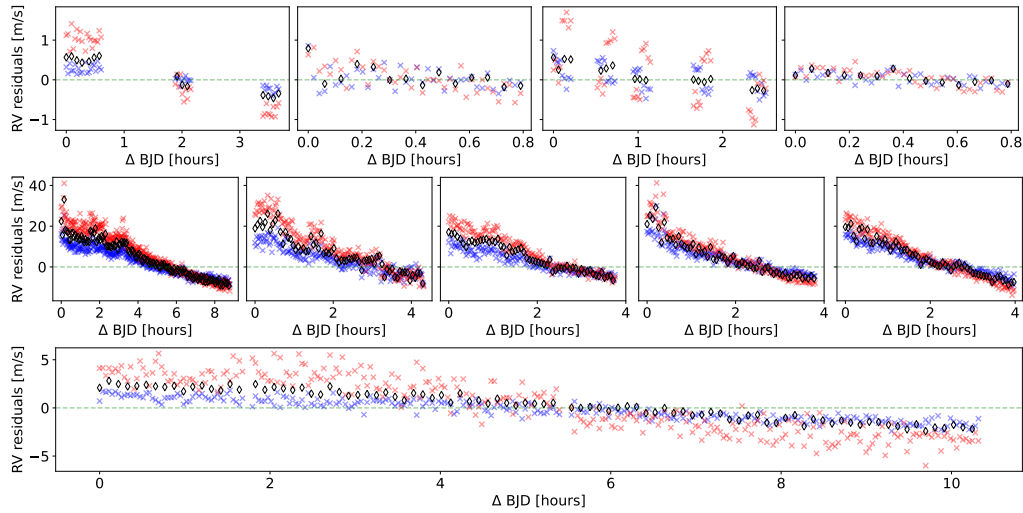


FIGURE 3.7: RV residuals extracted from the spectral data of each ESPRESSO (top and middle panel) and HARPS (bottom panel) detector. In black, we have the TM RVs computed with all spectral orders, whilst the colours represent the detector from which the RV time-series was extracted. The middle row is the data from the 5 nights of observations of HD40307.

Firstly, Figure 3.7 allows us to determine that the RV trend is present in multiple targets, albeit at different peak-to-peak amplitudes. If we now focus on each individual target, it is also possible to see that the amplitude is not the same, as some nights appear to be more affected than others. Chapter 3.22 will further explore this effect on a larger number of targets.

Secondly, we find a slight chromatic dependence, as the magnitude of the effect is larger on the red detector, when compared against the blue detector. As the average sampling in the two detectors is not the same, it might lead to this disparity between them. However, and despite the differences, the overall behaviour of the two time-series is similar, suggesting that whilst there might be some chromatic impact, it is not the main driving factor of the RV trend that we measure. Thus, the study of this connection between the amplitude of the trend and the sampling of each detector is left for future work.

3.3.2 Order-by-order RV measurements

As the division by detector revealed hints of a chromatic dependence, the next step is to look at the individual spectral orders, evaluating the variability of their RV throughout the night. For this purpose, we selected two spectral orders from the red detector and evaluated their evolution.

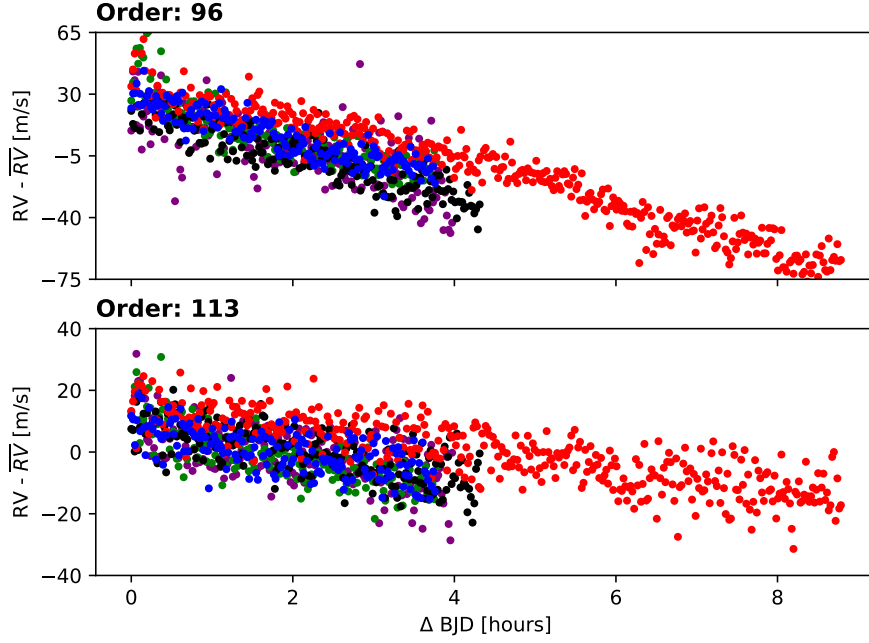


FIGURE 3.8: Comparison of the RVs obtained for two spectral orders in the red detector. The colours group the observations that were taken in the same night. **Top:** Spectral order 96, with a central wavelength of $\sim 537.6 \text{ nm}$; **Bottom:** Spectral order 113, with a central wavelength of $\sim 578.2 \text{ nm}$.

Figure 3.8 shows the behaviour for each night of observations, revealing a clearly linear trend over the entire night. On top of that, one can also see that: i) the evolution, over the multiple nights, is quite similar; and ii) Even within the same physical detector, different orders showcase different amplitudes. Thus, the final RV trend that we recover after combining all spectral orders is the mean of the effect as spread over the wavelength coverage of the instrument. As this Figure clearly shows the RV variability, we can now evaluate how it behaves across all spectral orders.

In Figure 3.9 we find the results of this comparison for the 5 nights of observation of HD40307, alongside a comparison with the percentage of pixels in each order where the transmittance profile falls under 1%. From this image it is clear that we are in the presence of a RV effects that appears to span across the wavelength coverage of the instrument, albeit with a greater impact in some selected regions. Similarly to the results from previous tests, the hypothesis of telluric features appears to be disproved, as the majority of the vertical columns with the largest variability are not those with the largest drops of the transmittance profile.

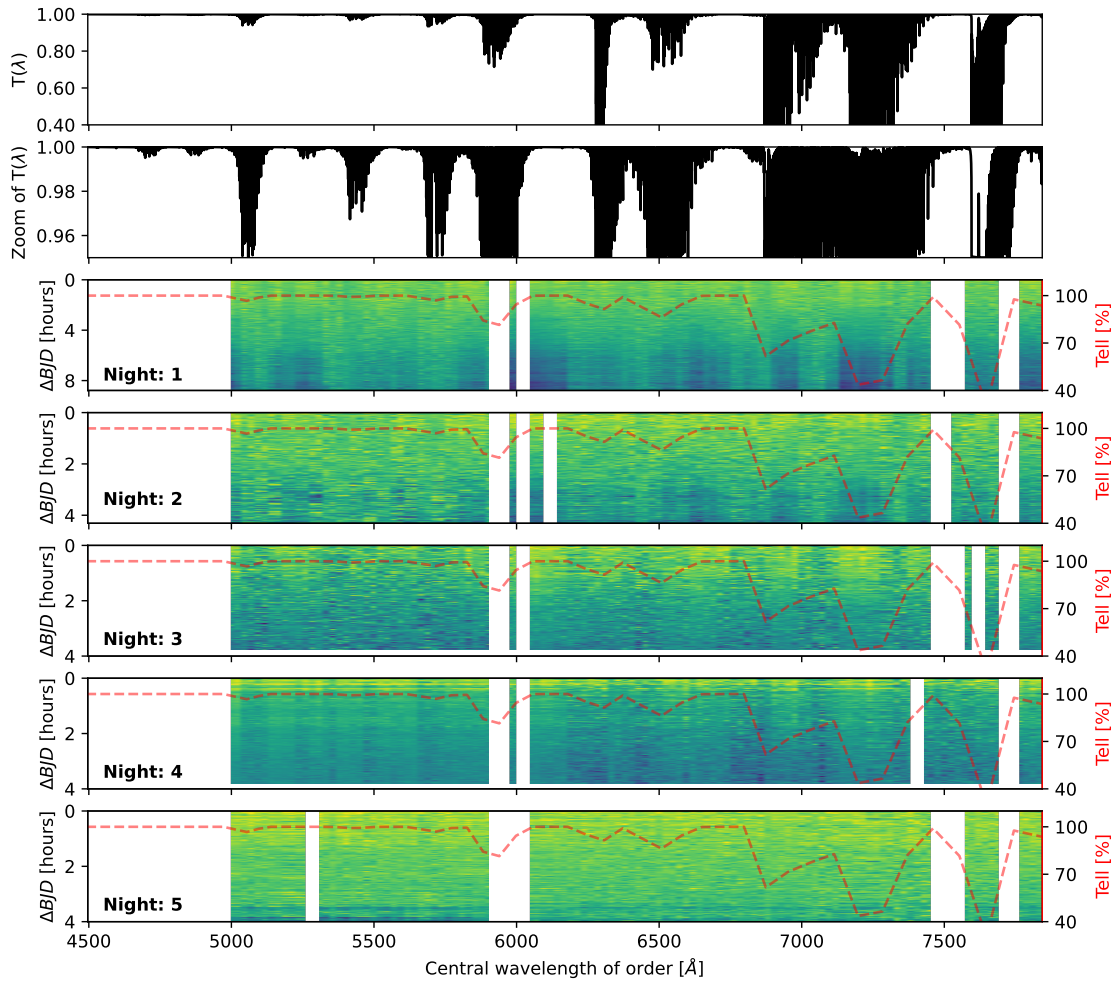


FIGURE 3.9: Evolution of the RV (colour-coded) for the central wavelength of each spectral order (XX-axis) across time, measured since the first observation of the night (YY-axis). The order-wise RVs are normalized by the maximum RV measured on that order over the entire night. Each panel corresponds to one of the 5 nights of ESPRESSO observations of HD40307, where the x-axis is the central wavelength of the corresponding spectral order. The two upper panels provide a transmittance profile (for comparison purposes) and a zoom above the 95% level. The percentage of the wavelength coverage (in each spectral order) whose tellurics are deeper than 1% are presented as a red dashed line in the bottom panels. Their values are indexed to the rightmost axis. The variation of RV is encoded as a colour map variation, with the brighter colours representing an higher RV measurement and the scales being independent across the different nights.

Furthermore, and in agreement with the previous analysis with the independent RVs per detector, the orders with the largest variability are located in the redder regions and for the different nights the more affected orders seem to be common. The same can be seen for the bluer region, albeit not as evident.

3.3.3 Wavelength-binning the spectral orders

As seen in Chapter 3.3.2, the RV trend is present in all spectral orders, leading to an (almost) achromatic behaviour across ESPRESSO detectors. One next step to take is the analysis of different regions of the spectral orders, across full wavelength coverage of ESPRESSO. To do so, we identify three regions:

1. *Start* - The first 3000 consecutive points (pixel) inside a spectral order
2. *Middle* - The 3000 pixels starting right after the previous chunk.
3. *End* - The remaining 3111 pixels in the order.

It is obvious that this division will naturally lead to a few problems: i) Across the wavelength coverage of the instrument, this strategy will lead to the two edges having a varying number of valid data points. In ESPRESSO case this is a consequence of the blue and red edges of the stellar spectra falling outside of the detectors; ii) The flux signal-to-noise ratio of the *middle* chunk will always be much higher than the other two, as it covers the central region of the spectral order; iii) The wavelength coverage of each chunk is not constant across the wavelength coverage of the instrument.

Following this recipe, we split the data of the nights from HD40307 and compute individual RVs for each chunk. To diminish the computational cost of this test we selected the stellar template constructed with data from the entire night, but only extracted RVs from a subset of observations selected within each night to probe the behaviour at the start, middle and end of the night. With this selection of observations, we can infer the behaviour of the intra-night RVs and evaluate if the trend is present or not.

From Figure 3.10 we see that the effect is present in the three chunks, despite the aforementioned issues with their definition. Not only that, but the amplitude of the variability is larger on the edges on the spectral orders (blue and red) than on the central region (green). The outliers that are found in the “Start” chunk of two of the nights are attributed to numerical issues introduced by the usage of a small region of the spectral orders, coupled with generally smaller SNR in such regions. Coming back to the intra-night variability, this test allows us to hypothesize that the trend is being introduced by our approach to the stellar continuum. Near the central region of the order, the blaze correction has a much smaller impact than on the edges, which might be playing a role in this.

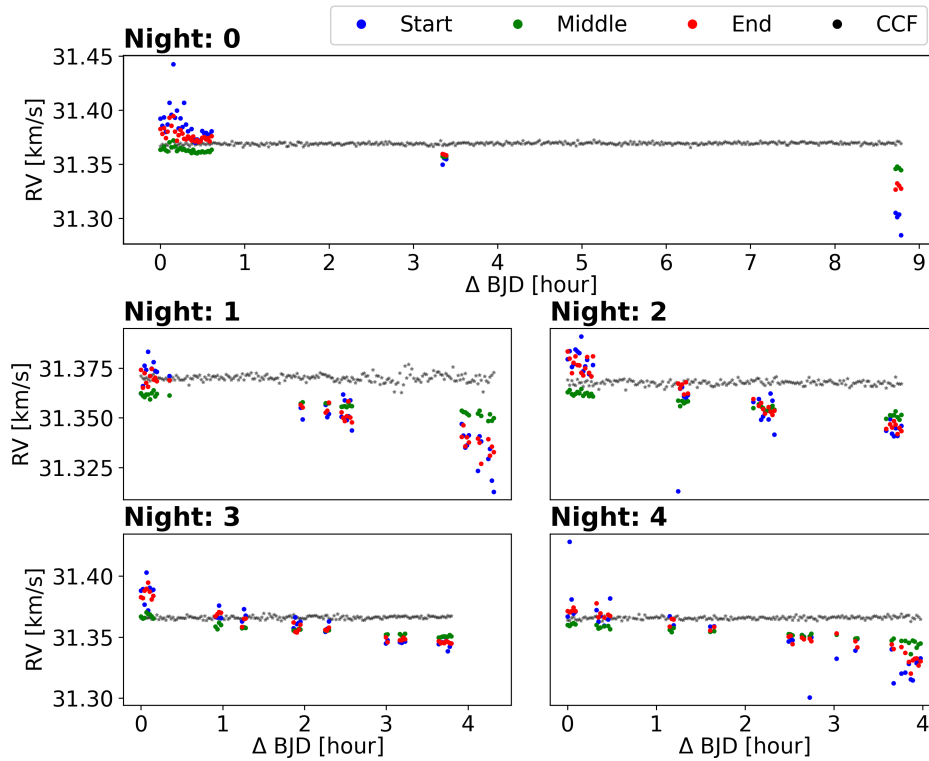


FIGURE 3.10: Comparison of the RV time-series from the CCF (black) and those from template matching applied all spectral orders. The RVs were derived for the three chunks defined in this Chapter: *start* in blue, *middle* in green and *End* in red.

Furthermore, it also connects back with the calibration of the instrument for single-night observations (refer back to Chapter 3.1), as we are using the same calibrations for all observations. However, ESPRESSO is a very stable spectrograph, which raises doubt about a continuum variation large enough to inject meter-per-second signals on template matching radial velocities. That said, if the source of the trend lies in the continuum, it would explain it not appearing on the CCF, as it only uses the information from the stellar lines. Unfortunately, at the time of writing this thesis, a test capable of evaluating the impact of changes in the spectral continuum is yet to be devised and implemented.

3.4 Evaluating the construction of the stellar template

A crucial part of the template matching algorithm is the model with which we compare the individual observations, i.e., the stellar template. Not only do we have to consider the method that is employed to construct the template, but also the data that is selected. As discussed until now (see Chapter 3.1 and 3.2), when considering intra-night observations we have different effects than when observations are spread through multiple nights. Not

only that, but the interval between two consecutive exposures is much smaller. Thus, it might be possible that the stellar template is contaminated by a possible variability of the stellar spectra. This would also point towards such an effect being averaged-out whenever the observations are spread throughout multiple nights. At shorter time-scales, i.e., within a single night, this possible effect would thus be present in the stellar model. By changing the observations that are selected for the construction of the template, it should be possible to seek out any intra-night variability that is driving the RV extraction.

3.4.1 A subset of observations as a template

The first test for the short-scale variability, is to construct the stellar template from only a sub-set of available observations, using it to compute RVs for the entire night. For this, we selected one night of observations of HD40307 and chosen 3 sub-sets of 20 observations each. Two sub-sets were selected from the start and end of the night, and the last one comprised the 20 observations with the lowest airmass in each night. Each template was used to extract RVs for the entire night, as depicted in Figure 3.11.

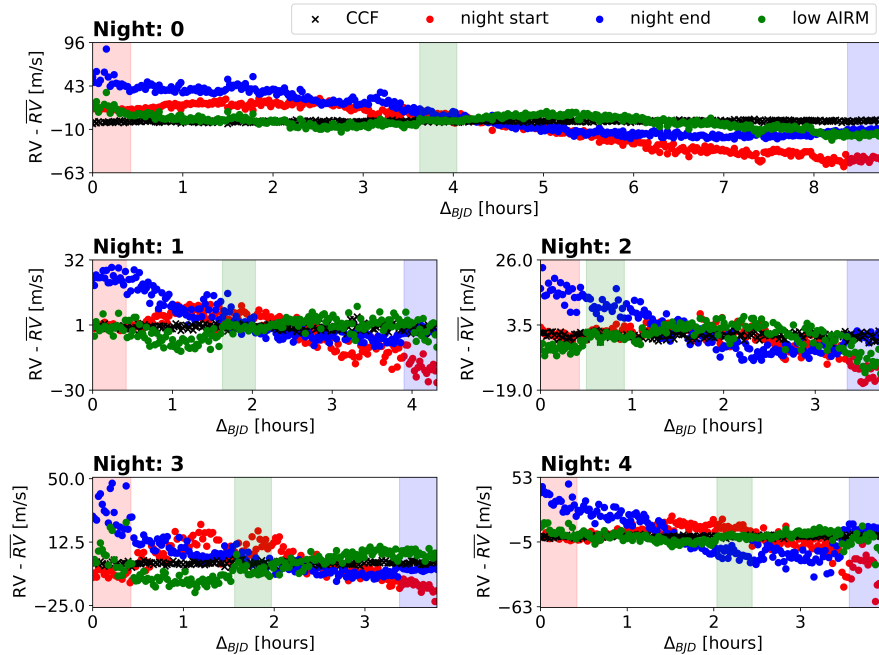


FIGURE 3.11: Comparison of the HD40307 RVs, from each night, extracted with a stellar template constructed from the first 20 observations (red), the 20 observations observations with the lowest airmass (green) and the final 20 observations (blue). For visualization purposes, the RVs of each time-series had their corresponding mean values removed. The CCF RVs are represented as the black crosses, whilst the shaded regions highlight the observations that are present in each of the three templates.

It is clear, from this analysis, that the selection of observations for the template has a strong correlation with the overall shape of the RV trend, as we find different variability patterns on each template of each night. However, it is important to note that there is a consistent behaviour of the “night start” and “night end” RVs across the different nights. The stellar template constructed from the lower airmass observations (green) appears to present a periodic behaviour. One exception can be found on night 2, where the lower airmass observations are close to the start of the observations. This, apparently, periodic variability might hint towards an intra-night correlation that ends up being averaged over when all observations are used to construct the stellar template, as previously seen.

As we have just seen that the selection of observations is playing a role in the final RV trend that we measure, we decided to derive RVs whilst incrementally adding new observations to the construction of the template. The initial template is one constructed from the first 7 observations of the night and all subsequent templates are built by including all previous observations and the 7 following ones. Each template constructed in this manner is then used in the extraction of RVs for the entire night of observations.

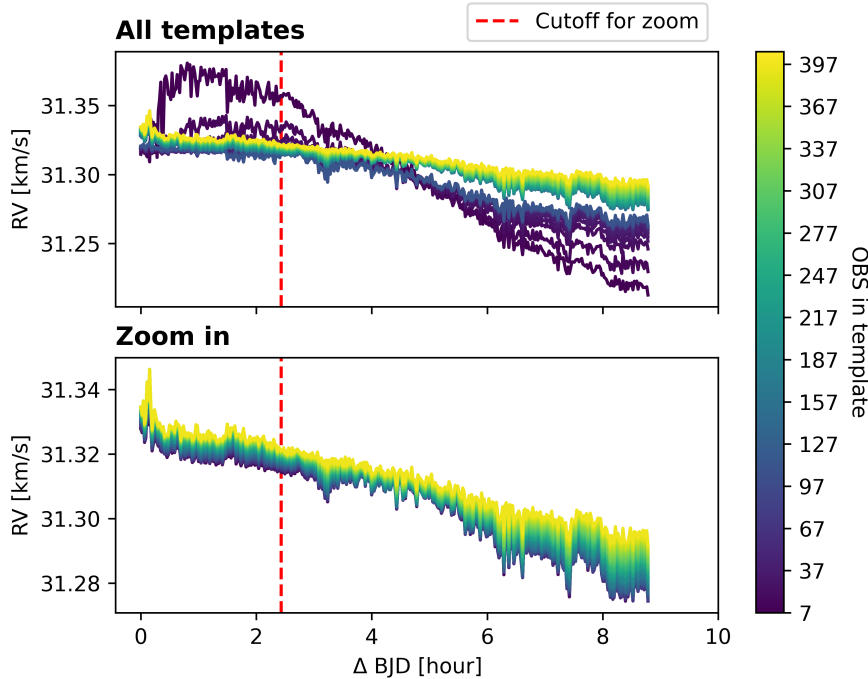


FIGURE 3.12: RV variability when constructing stellar templates from an incrementing number of observations (colour coded, from blue to yellow). For visualization purposes, the bottom panel excludes the time series with higher variability, associated with the templates constructed from a smaller number of observations (bluer colours). The red dashed line represents the timestamp associated with the last observation of the last template to be excluded from the top panel.

Figure 3.12 presents the results of this test, allowing us to see that as we increase the number of observations in the template the effect slowly decreases, until it converges in the final trend that we have been trying to understand in this Chapter. Independently of the number of observations used to construct the stellar template, we retrieve a negative slope in the RVs, albeit with different values. The observations taken in the first 2 hours of the night, result in a increased variability. After that point, the evolution is much smoother, until eventually converging to the final trend. Once again, this is compatible with a short-scale variability being averaged in the stellar template as more information (observations) is used in its construction.

3.4.2 Applying a template from a different night

As we have been seeing within this sub-Chapter, the generation of the stellar model from different subsets of observations from the same night are leading to clear changes in the measured spurious signal. One hypothesis that was brought up, through the selection of different observations for the construction of the template (see Chapter 3.4.1), was correlated noise with a short time-scale of evolution. Given this situation, the application of a template to a night different than the one used in its construction should result in similar behaviour. This reasoning is motivated by the fact that inside a night we keep the flux variability (in the observations) that is being compared with a fixed model (stellar template), where a similar averaged effect is present.

From Figure 3.13 we find that despite the stellar template being constructed from different nights, the underlying issue is still present. However, it is noteworthy to mention that a change in the data used to construct the stellar template leads to different variability patterns in the nightly RVs, as we can see across the different columns. These results lead us to the hypothesis that there is some intra-night variability in the fluxes, albeit with an unclear source. However, it is not clear if it is exacerbated during the construction of the stellar model.

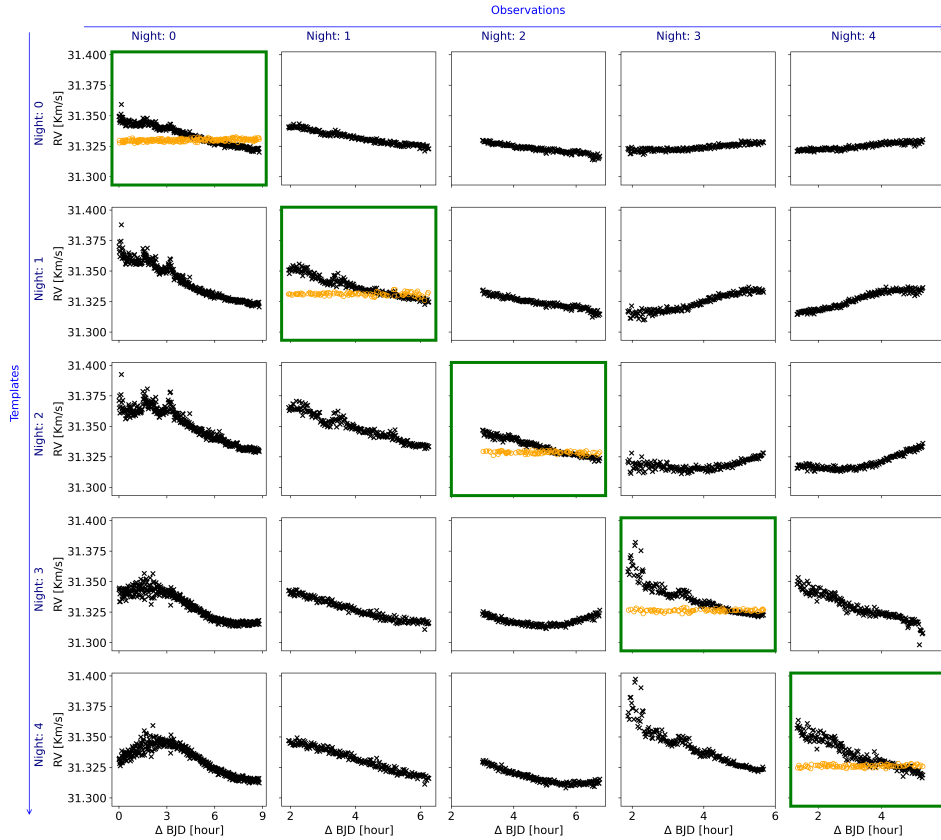


FIGURE 3.13: RVs of the 5 consecutive nights of HD40307, when using stellar templates constructed from each night of the dataset. Each row corresponds to a stellar template constructed from the corresponding night, whilst the columns correspond to the data from which we extract RVs. The data from the first night of observations is presented in the top-left corner. For comparison purposes, the x-axis of each panel represents the number of hours spent since 00:27, local time, selected due to being the hour of the first observation of the first night. For comparison purposes, a sparse representation of CCF RVs is provided as orange circles.

3.4.3 A single observation as a template

Following the previous results from Chapter 3.4.1 and 3.4.2, we see that the choice of observations for the template plays an important role in the overall effect, with an open question regarding the methodology under which we create the stellar template. Thus, to disentangle between an issue on the creation of the template and variability on the data itself, we extract RVs using a single observation as the stellar template. For this effect, and to provide a larger baseline of comparison, we select all nights of observations of HD40307, selecting three observations from each: the first, the last and the one with the lowest airmass. Each observation is then used as the template for all other observations, even those from different nights.

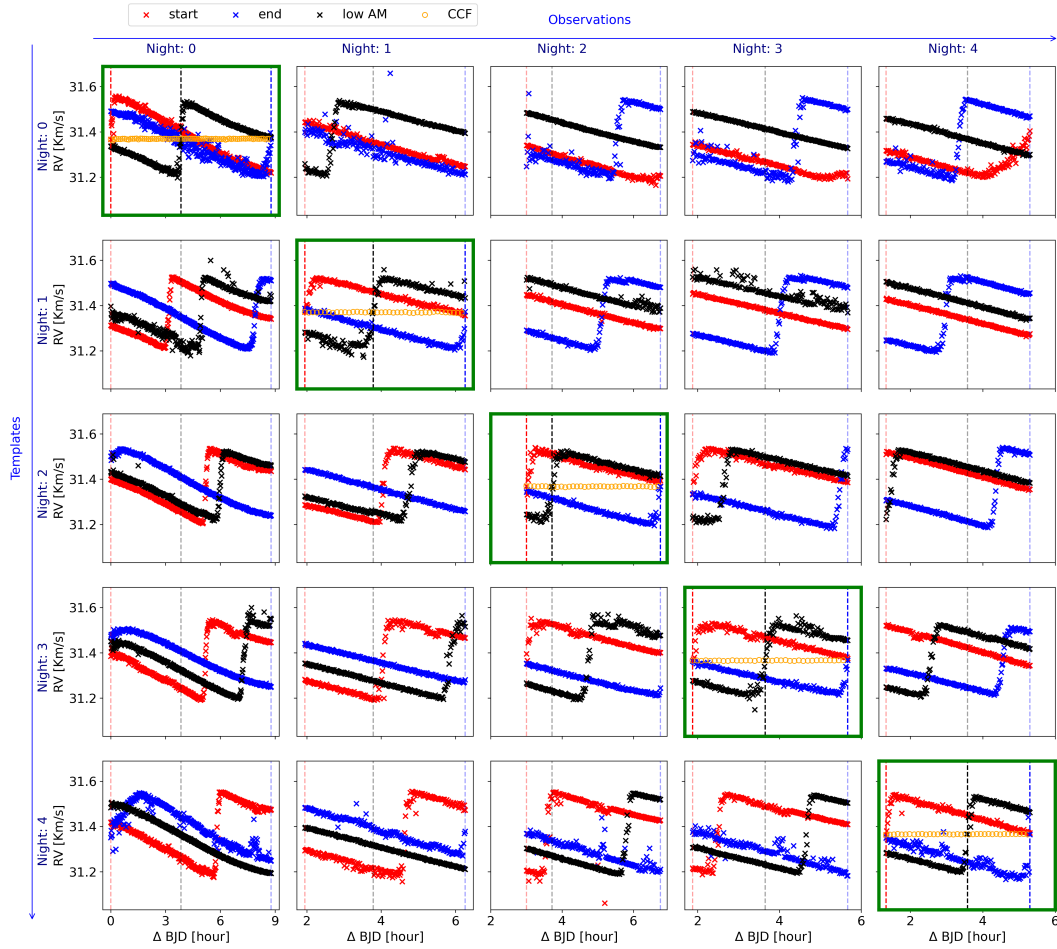


FIGURE 3.14: RVs of all nights of HD40307, when using a single observation as the stellar template, taken from the start (red) and end (blue) of the night, as well as the one with the lowest airmass (black). A move along a given column leads to RVs from the same night extracted with different stellar templates, whilst one alongside a row keeps the stellar templates but changes the data to which they were applied. The data from the first night of observations is presented in the top-left corner and the hour at which the template was created is denoted as the dashed lines, coloured with the corresponding colour. For comparison purposes, the x-axis of each panel represents the number of hours spent since 00:27, local time, selected due to being the hour of the first observation of the first night.

For comparison purposes, the CCF RVs are plotted as dark-yellow circles.

The results from this analysis are presented in Figure 3.14, where we apply each of the 15 (in total) stellar templates to all available nights. If we start by focusing on the diagonal of the matrix, highlighted in green, we observe that despite the RV trend being present, the RVs reveal a large jump as we approach the hour of the observation selected to be the stellar template. In the majority of cases, the peak-to-peak amplitude of this effect seems to be constant, as irrespectively of using the first or last observations (red and blue curves) they meet each other. The “low airmass” template also reveals that

the same slanted behaviour is found before and after we cross the hour of the selected observation.

If we now turn our attention to the variation across different stellar templates (i.e., across rows, in the same column) we find that as we increase the number of days between data and template, the RV jump occurs at different hours within the night. This behaviour is clear in the first night, where the jump for the “low airmass” curve occurs later in the dataset, until it eventually disappears. Simultaneously, the same can be seen for the other two curves (red and blue).

With this, we see that the source of the RV trend is not the construction of the stellar template itself, as the RV trend is still recovered in this extreme case of using a single observation as the stellar template. Despite the much lower SNR of the model, when compared with the individual observations, we are able to retrieve the spurious signal that we have been trying to diagnose within this Chapter. Thus, from here it follows that the issue either lies in a different part of the TM algorithm itself, or in the usage of a common set of calibration files inside a given night, as previously discussed.

3.5 Evaluating the interpolation algorithm

As the previous tests yielded no clear indication of what might be driving this structured RV variability, we decided to go over the principles of the template matching algorithm, as well as the assumptions that it enforces. One glaring assumption is that at two distinct points we assume the shape of the stellar spectra between pixels, as the data is interpolated to a given wavelength solution. This interpolation process is carried out with a cubic-spline interpolation. Furthermore, it is carried out both during the creation of the stellar model and subsequent RV extraction. Contrastingly, the CCF method does not perform an interpolation of the stellar spectra.

It is thus possible that this assumption is introducing a bias in RV that is only appearing on single-night observations due to the smooth change in sampling. Fortunately, testing this hypothesis is straightforward, as we only need to change the interpolating algorithm, recompute the RVs, and search for the presence of meaningful differences.

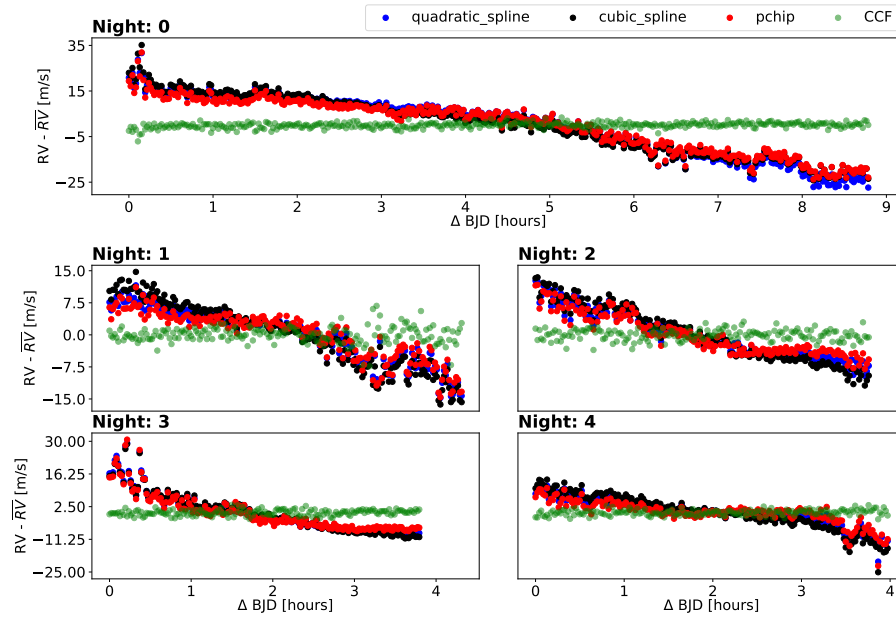


FIGURE 3.15: Impact of the choice of the algorithm used to interpolate the stellar spectra, as seen in the nightly RVs of HD40307. For this comparison, three interpolation algorithms were used: quadratic spline (blue), cubic spline (black) and pchip (red), all implemented in *scipy*. For comparison purposes the CCF RVs are over-plotted as green circles and each time-series had its own mean value removed.

For this preliminary test, we selected different interpolation algorithms, as implemented in *scipy*, constructing new stellar templates with each and, afterwards, extracted RVs using the same interpolation algorithm where applicable. The result of this analysis is shown in Figure 3.15, where we find that the different algorithms are leading to noticeable changes in the overall trend. However, despite the measurable difference in behaviour, the trend is still present. One hypothesis for this behaviour is that all of the presented algorithms are piecewise methods and, consequently, could be introducing systematics due to slight changes in the nodes, as discussed in sub-Chapter 3.5.1. This hypothesis will be further explored, through simulations, later on.

3.5.1 The change of sampling of the spectral lines

Recalling back from Section 3.1, the BERV variation during the night is slowly shifting the position of the spectra, relative to the pixels of the CCF. In other words, during the course of the night, we are probing different parts of the spectral line, as the spectra drifts over the detector. If we now ignore the effect of any stellar RV, one would not expect that a change of the points in which we sample a given stellar line would result in a RV shift, as it is the exact same line.

During the RV extraction process, we have already seen that both observations and template are interpolated for the desired wavelength solutions (refer back to Chapter 2.1.3 and 2.2, respectively.). The interpolation that is employed within **s-BART** is a piece-wise cubic-spline, meaning that it depends only on the nearest neighbours (see Appendix D for more details). If we now have a variation of the nodes of the interpolations (i.e., the neighbours), the RV trend could arise as a systematic impact of the interpolation on the line shapes. In fact, if we exaggerate the issue, we can easily see it on the Gaussian from Figure 3.16.

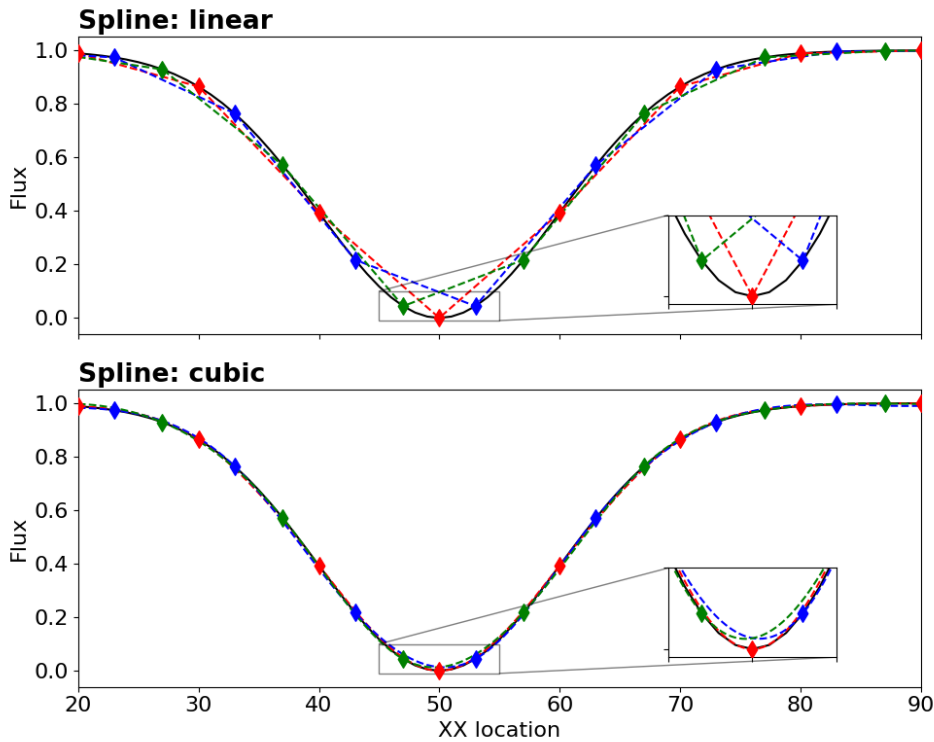


FIGURE 3.16: Schematic of the impact of the interpolation on line shape, depending on the sampling of the line. For representation purposes, the sampling is sparser than the one in real observations. The diamonds correspond to different sets (colour coded) of equidistant points selected from the same Gaussian (shown as the black line). The dashed line is the result from different interpolation algorithms, using the same-coloured diamonds as the nodes for the interpolation. The insets provide a zoom-in in the core of the line. **Top:** Linear interpolation; **Bottom:** Cubic-spline interpolation.

In Figure 3.16 we demonstrate the impact of the interpolation on the line shape and position, depending on the sampling of the original function. From an initial Gaussian (black line) 3 sets of equidistant points were defined (each represented by the same-coloured diamonds). Then, for each set of points, we interpolate back to the original grid (dashed lines), comparing the output against the original one. It is thus clear that by dealing with

finite data, we can have a drastic difference on the final line, depending on the parts of the lines that were sampled. Not only do we see a difference in the central location of the line, but also on its depth, with the different interpolating algorithm introducing a different effect.

3.5.2 Evaluation of a simple Gaussian line

As we have just seen in Chapter 3.5.1, the exact sampling locations of the spectral line can introduce a systematic bias in the interpolated spectrum. To measure this effect we attempt to simulate it on a noise-free Gaussian line, which is constructed by the following recipe:

1. Arbitrarily select one observation of ESPRESSO, from which we select the 100-th spectral order. From this spectral order, we select 100 consecutive wavelengths, henceforth λ_{gauss} , which will be used to generate our Gaussian line;
2. Define a Gaussian function, $\mathcal{G}_{\mu=\bar{\lambda}}^{\sigma=0.05}$, that we will be evaluated at different wavelengths to construct the noise-free line. The central location of the Gaussian, μ , was set to be the mean value of the two central pixels of the λ_{gauss} interval, $\bar{\lambda}$. Its standard deviation, σ , was chosen to take the value of 0.05 Å. This arbitrary choice produces a line with a width similar to the ones from ESPRESSO.
3. The “noise-free” line, F_{gauss} , is then created by evaluating the Gaussian at the selected wavelengths, $F_{gauss} = \mathcal{G}_{\mu=\bar{\lambda}}^{\sigma=0.05}(\lambda_{gauss})$

The evaluation of the effect of the interpolation follows a simple procedure: i) we Doppler-shift λ_{gauss} by a given RV shift, creating λ'_{gauss} ; ii) Interpolate F_{gauss} to the new wavelength solution, obtaining F'_{gauss} ; iii) Compare F'_{gauss} with the original function evaluated at the new wavelengths, $\mathcal{G}_{\mu=\bar{\lambda}}^{\sigma=0.05}(\lambda'_{gauss})$ wavelengths. As discussed before, the interpolation algorithm is a cubic-spline algorithm, a piece-wise interpolation. To avoid a large change in the interpolating nodes, we limit the range of possible RV shifts to an interval of $[-500, 500] \text{ m s}^{-1}$, slightly larger than the mean RV-size of one ESPRESSO pixel.

Our comparison is show in Figure 3.17 where the residuals are computed for the “flux” value. On top of that, a numerical computation of first and second derivatives, through second order central differences of both the original and interpolated lines, allowed for comparison of variations in RV content. From the colormap we can clearly note that as the

RV shift leads us away from the nodes of the interpolation, the residual steadily increases, with the largest contamination occurring near the centre of the line. More importantly, we can see a slight RV offset between the location of the maximum residuals from positive and negative RV shifts, which is explained by the non-uniform distance between the nodes of the interpolation and the usage of different nodes for the interpolation.

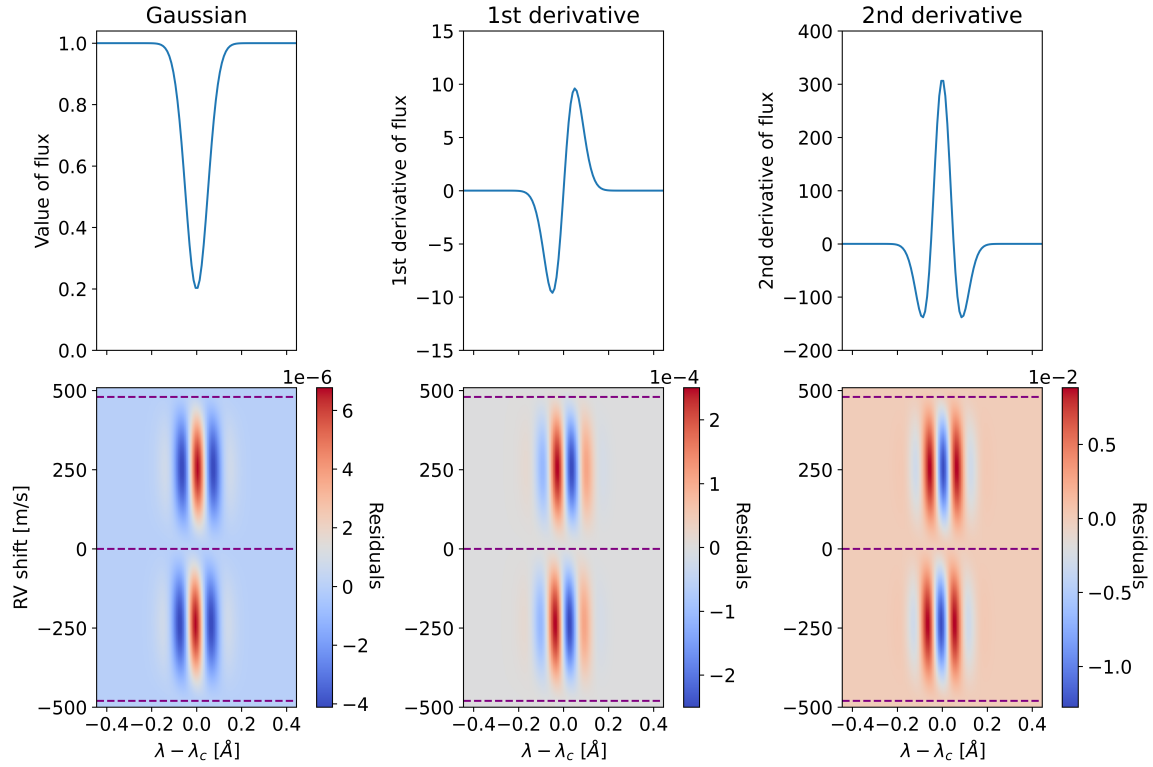


FIGURE 3.17: Gaussian line, built on top of ESPRESSO’s wavelength coverage and centred on μ_G , and its first and second numerical derivatives. On the bottom row we have a colormap of the residuals after interpolating the original Gaussian to a given RV shift. The dashed purple lines mark the 480 m s⁻¹, the mean size of one ESPRESSO pixel.

The wavelength asymmetry in the residuals leads us to take a closer look for only for a few selected RV shifts and a specific pixel, as shown in Figure 3.18. In this image, it is clear that there is an offset between the residuals of the interpolated lines generated through a positive RV shift, and those from a negative shift. On top of that, it can also be seen that the RV shift is also correlated with the amplitude of the residuals from the first derivative, with negative values leading to the larger impact.

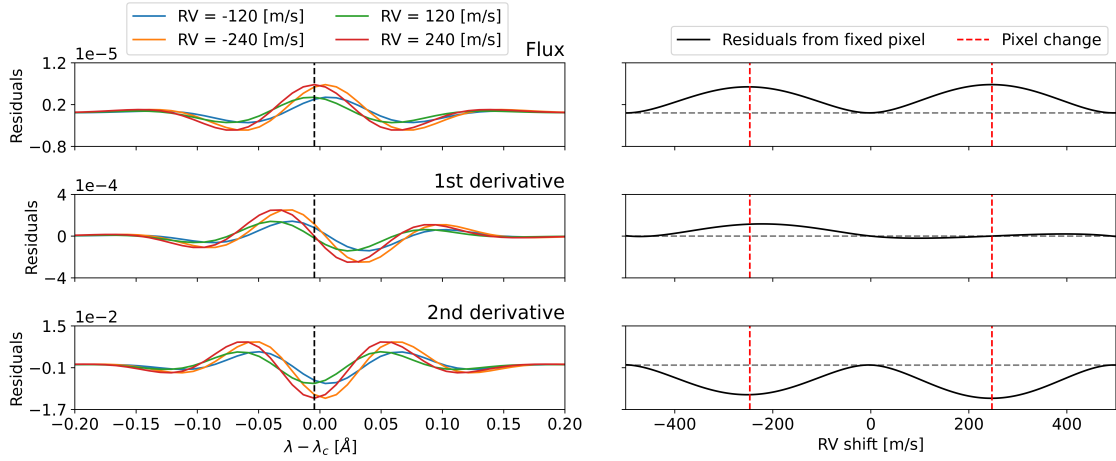


FIGURE 3.18: **Left:** Residuals of flux (top row), first derivative (middle row) , and second derivative (bottom row) for four RV shifts; **Right:** Variability of the residuals from a single pixel, near the core of the line, with the change in the injected RV. The pixel is highlighted with a vertical dashed line on the left column.

Until now, our tests have revealed that an interpolation of this simplistic Gaussian line is leading to wavelength asymmetries in the residuals, raising questions about its direct impact on the RVs. Thus, to estimate this impact, we fit a Gaussian to the interpolated line, F'_{gauss} , through a χ^2 minimization. Afterwards, the central location of the fitted Gaussian is compared with the original μ , and converted from a difference in wavelength to one in RV space.

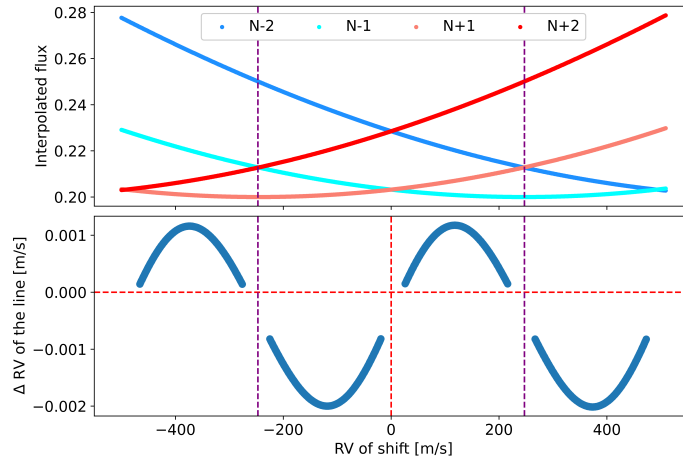


FIGURE 3.19: **Top:** RV dependency of the interpolated flux value of four different points in the gaussian. Two were selected leftwards of the centre (blue colours) and the other rightwards (red colours). **Bottom:** RV shift between the central point of the original μ and the one obtained through a fit to the interpolations of the simplistic line. The dashed purple line represents the RV value for which the deepest wavelength of the line is closest to its neighbours. The missing points were those for which convergence was not achieved.

In Figure 3.19 we see the RV impact as we interpolate the line farther away from its original position, rejecting those for which we were not able to achieve convergence. For RVs that takes us close to the interpolation nodes (i.e., $0, \pm 480 \text{ m s}^{-1}$) the differences are minute, and thus the minimization algorithm is dominated by numerical instabilities and fails to achieve convergence. However, after that is no longer the case, we find a small peak-to-peak variation of $\sim 3 \text{ mm s}^{-1}$, 2 orders of magnitude smaller than the expected precision of ESPRESSO. This leads us to the conclusion that, rather than the root cause behind the RV trend, we are seeing an effect that can not be measured with the current precision of our spectrographs. However, it is not clear how the effect translates from a single line to a full spectral order.

3.5.3 Evaluation on a sum of Gaussian functions

In the multiple tests carried out within this Chapter we have seen that a change of the interpolation algorithm was translated into a measurable, albeit extremely small, difference in the RV trend. However, in real observations, we have more than 1 spectral line per order which, at the end of the day, might contribute to the measured variability. Thus, we set out to transform our model of a single gaussian line, into one comprising a spectral order where each line is a Gaussian. The generation of such dataset follows a simple recipe:

1. Select one order of a given observation, once again chosen to be the 100th spectral order, from which we will retrieve the wavelength solution. The continuum of the order is normalized by the median flux level;
2. Determine the central position, λ_i , of each spectral line, through a search of local minima in the order;
3. To avoid always sampling the exact centre of the line, each central location is Doppler-shifted by a given RV shift.
4. For each detected line centre, create a Gaussian using it as its mean value and a standard standard deviation of $0.05 \text{ \AA}$. Each Gaussian is evaluated in the desired wavelength coverage, λ_i , and then they are all summed together. Henceforth, the result from this comparison shall be referred to as *order of gaussians*.

With this recipe in hand we create observations with RV shifts in the interval $[-1, 1] \text{ km s}^{-1}$, with separations of 10 m s^{-1} between points. Then, we apply our classical template

matching algorithm to it, using the observation with an RV shift of zero as the stellar template, providing an avenue to compare the injected and retrieved RV shift.

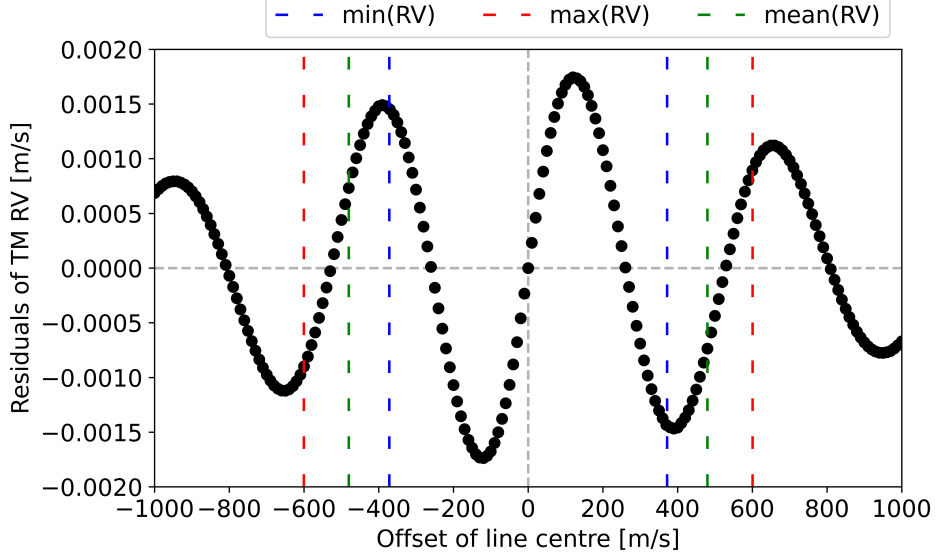


FIGURE 3.20: Residuals in the TM time-series, from the creation of the *order of gaussians* using RVs in the interval $[-1, 1]$ km s^{-1} , with separations of 10 m s^{-1} between points. The dashed lines represent the maximum (red), mean (green) and minimum (blue) RV coverage of the pixels from the selected order.

The result from this analysis is presented in Figure 3.20 where, similarly to the previous test with a single line, we find an asymmetrical effect in the RV residuals around the RV of 0 m s^{-1} . Once again, we find that the peak to peak amplitude of the signal in the residuals is at the level of the mm s^{-1} , well below what we can measure with ESPRESSO.

For comparison purposes, we compute for each pixel, in this order, the RV needed for the wavelength λ_i to be equal to λ_{i+1} and plot the minimum, mean and maximum value for it. We find that the RVs such that one pixel is transformed into its neighbour match some of the inflection points, whilst the two closest peaks actually occur at an RV near half of the one covered by the median ESPRESSO pixel. One issue with this test is that it is not a good translation of reality, as we are not probing for a change in the sampling of the line but, instead, we are shifting its centre.

To probe for the variation in the line sampling, we now fix the line displacement to a single RV and evaluate the previously described *order of gaussians* in a different wavelength solution, constructed by Doppler shifting the original wavelength solution, λ_i by a certain RV shift. As this process does not change the line centres, it is only introducing slight change in the wavelengths at which the spectral line will be present. Following the same

approach as before, we use the original spectrum as the stellar template and extract RVs under a classical TM methodology.

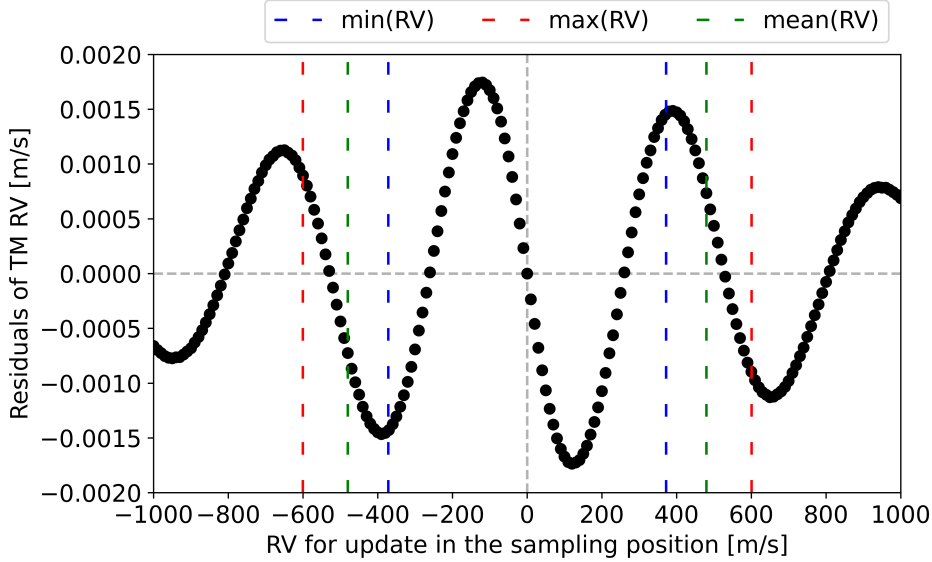


FIGURE 3.21: Template matching RVs being measured on the simulated spectral order, albeit with different samplings. The change of sampling of the lines follows the logic of intra-night observations, as it is slowly Doppler-shifted. The dashed lines represent the maximum (red), mean (green) and minimum (blue) RV coverage of the pixels from the selected order

From Figure 3.21 we find a very similar behaviour as before, with a periodic variability around the starting position of the sampling. Not only that, but the amplitude of the retrieved residuals is similar to the one from before, thus casting doubts about the hypothesis of the interpolating algorithm. More importantly, we are able of retrieving a bias introduced by our choice of interpolation algorithm, but its amplitude is not enough to realistically impact measurements with current, state-of-the-art, spectrographs.

3.6 Large scale impact of the RV trend and a search for correlations

Following the results from our tests, showcased within this Chapter, and the lack of an obvious conclusion for the RV trend, we set out to increase the sample of stars with intra-night observations so that we could search for possible correlations and similarities between different targets. In total, 25 targets were selected, with 1 being from HARPS data and 24 from ESPRESSO. The vast majority of the ESPRESSO observations being motivated by transiting events and collected during the *GTO* period. For simplicity sake,

we manually searched through each RV time-series, rejecting any observations where we observed a RV variation associated with the transiting event (i.e., Rossiter-McLaughlin profiles, similar to those from Figure 2 of Santos et al. (2020)). Thus, for each night, we constructed a stellar template with out-of-transit observations and applied the classical template matching algorithm to those same observations.

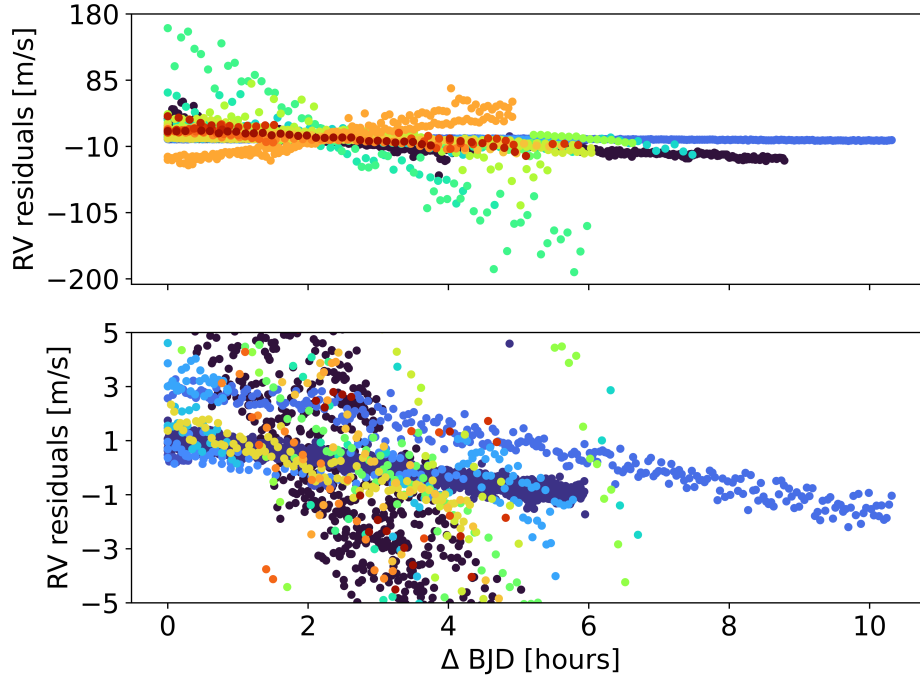


FIGURE 3.22: Difference between the CCF and the TM RVs as a function of hours spent observing after the first observation of the corresponding night. Each night of observation (of the same target) was processed independently from others. However, all observations of the same target share the same colour. **Top:** Results for all targets. **Bottom:** Zoom-in of the difference.

A comparison of the RV time-series, as shown in Figure 3.22, allows to see that there is a great similarity between the different targets, albeit with a few clear outliers. Among the ESPRESSO dataset, we see that the RV slope appears to be the same, with the largest differences occurring for the HD40307 target, the target selected for the bulk of the analysis of this Chapter. The fact that we see a similar behaviour in the different spectrographs and targets – obtained at different absolute values of BERV – suggests that this effect is connected with the template matching methodology, rather than the spectrograph or individual targets.

The most noteworthy results appear on the two nights of one ESPRESSO target, where we see a slope with a opposite signal than all others. A direct comparison, of the different

instrumental and atmospheric parameters, between those two nights and the ones from other targets, leads to no clear conclusion for the reason of the different behaviour. The only notable difference is the amplitude of RV variation from the CCF, which is larger than 500 m s^{-1} . If the true culprit for the RV trend is the sampling of the stellar lines, hypothesis that Chapter 3.5.1 raised doubts on, then it might be connected with the sizes of the ESPRESSO pixels, as this RV variation is larger than the mean RV-coverage of one pixel.

Despite this different behaviour for one target, we can already search for a possible correlation between the slope of the RV difference and some instrumental parameters. For that, we fit a first degree polynomial to each time-series of residuals and compare them against:

1. The median BERV value, in an attempt to evaluate a year-long dependency;
2. The mean CCF RV;
3. The flux SNR at the end of ESPRESSO's blue detector (order 70, middle wavelength of $\sim 480 \text{ nm}$);
4. The effective temperature of the star, as taken from the *SweetCat* catalogue (Sousa et al., 2021) if available. Otherwise, it was taken from the literature.

To simplify the comparison, we ignored the HARPS target, drawing comparisons only between the ESPRESSO ones. From Figure 3.23 we first see that the majority of the targets have a similar RV slope, between -0.250 and zero $\text{m s}^{-1} \text{ day}^{-1}$, with only a few outliers. Once again, this result points towards a systematic issue with the method itself, despite the presence of a relatively large window for the slope values across targets. A search for possible correlations with the selected metrics also yield no evident conclusion. More interestingly, we find that the target with a positive slope does not appear to differ from the others (in the selected metrics).

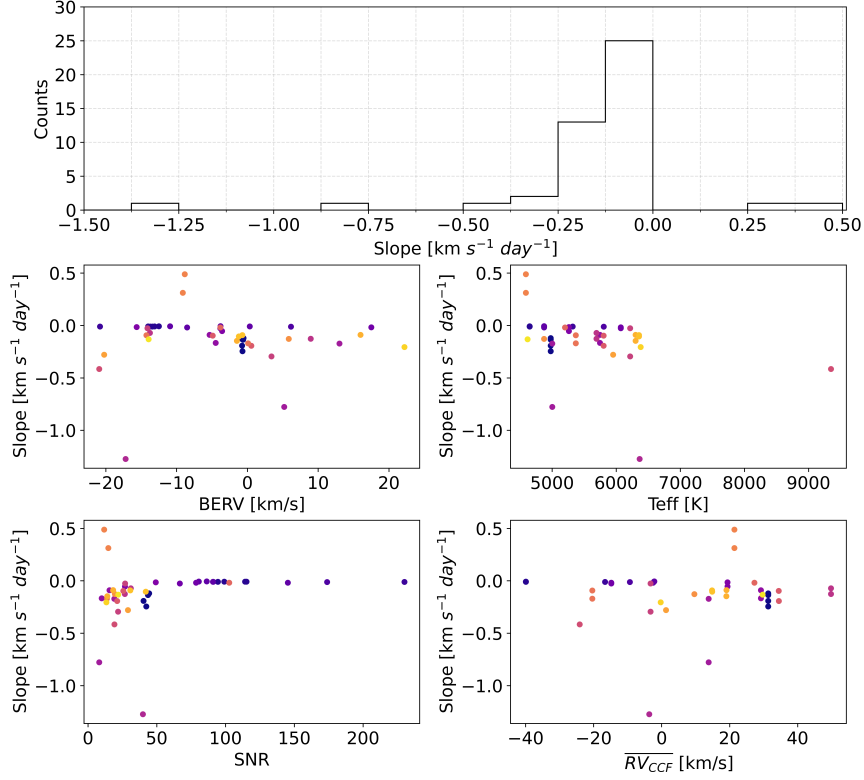


FIGURE 3.23: Analysis of the RV slope for all targets discussed in this Chapter, obtained with a first degree fit to the residuals of each night. **Top panel:** Histogram of the slope, across all nights. **Middle and bottom panels:** Correlation between the slope and the median BERV value, the SNR at order 70, the mean stellar RV and the effective temperature. The multiple nights of the same target are represented with the same colour scheme.

3.7 A multi-night template applied to intra-night data

In our quest to diagnose issues in the construction of the stellar template, we select ESPRESSO data from GJ9827. This star, unlike others analysed within this Chapter, was observed in two different settings within the ESPRESSO GTO:

- Characterization of the transiting planet as it crosses the stellar disk. In total, it was observed in three nights, spanning a total of 5.6, 7.6 and 4.1 hours. We shall refer to them as night *0*, *1* and *2*.
- Further studies of the planetary system, following results previously obtained through the transit method. In this setting, the system was observed in 2019 for 5 months, with the campaign starting on the 3rd of July and ending on the 2nd of December. For simplicity, we shall refer to this set of observations as “multi-night”.

This system thus presents a unique opportunity to probe the construction of a stellar template in this intra-night setting. From this data, we can construct stellar templates from the single- and multi-night observations and apply it to the different datasets. For our analysis, we construct 4 stellar templates:

- *Nightly* templates: As usual in this Chapter, we consider each night to be independent, thus creating 3 independent stellar templates;
- *Full* template: Constructed by only using the *multi*-night observations that were obtained in the context of the transit follow-up observations.

The *nightly* templates are used to extract RVs from the corresponding night, as well as from the multi-night observations. Similarly, the *Full* template was used to extract RVs from its own observations and the *nightly* ones.

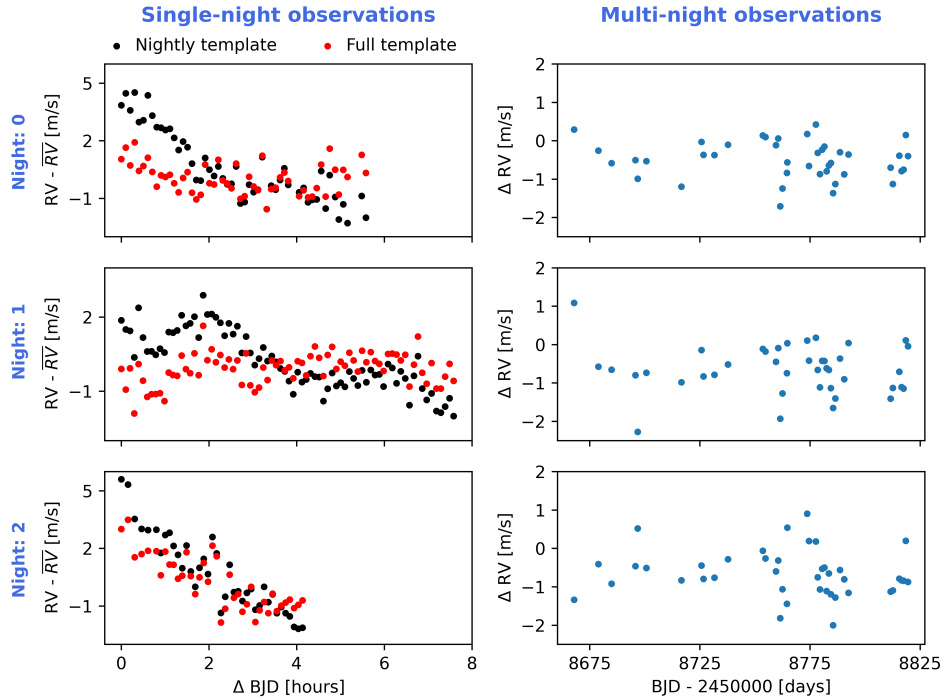


FIGURE 3.24: Impact of a change in the stellar template in the retrieved RV values. **Left:** Comparison of the RVs obtained through the *full* template (red) and *nightly* templates (black) applied to the 3 nights of observations. **Right:** Residuals of the RVs from the multi-night observations when using the *Full* and *Nightly* templates. Each row corresponds to the usage of a different *nightly* template.

From Figure 3.24 we find a surprising result: when we use the *Full* template (the one constructed from multiple nights of observations) we no longer find the RV trend that is retrieved with the *Nightly* template. It is also important to note that on night 3, this

template only translates into a decrease of the amplitude of the trend, rather than a full correction of its effect. Furthermore, the application of the 3 *nightly* templates to the multi-night dataset reveals an increased scatter, compared to the *full* template. Their behaviour is similar, albeit with a few observations presenting outliers in the RV extraction.

As the most striking result of this analysis is the disappearance of the trend, we set to understand the differences between the 4 sets of observations. To do so, we look into different metrics to compare night 2 with the other datasets. From Figure 3.25 we find that the different datasets do not present a clear difference in their observational metrics. A noteworthy point is that night 2 has BERV values similar to the majority of the observations from the *full* dataset. This is not the case for the other nights. Furthermore, the atmospheric humidity in this night is slightly higher than the other ones. However, the difference is small and, following previous results, the impact of leftover telluric features do not seem to be the driving effect of this trend.

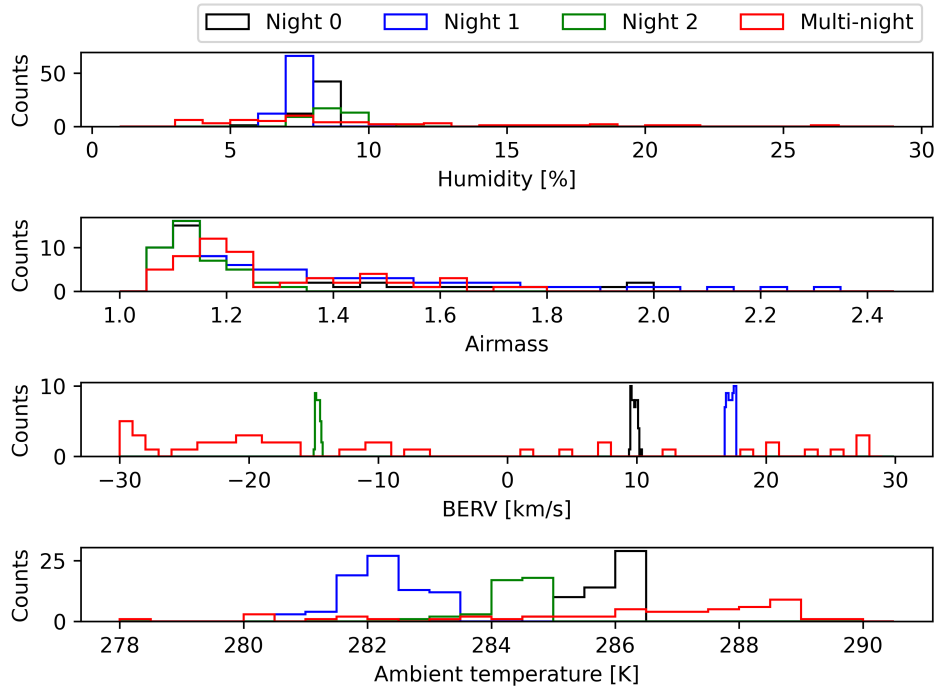


FIGURE 3.25: Comparison of different instrumental metrics for the 4 datasets considered in this (current) analysis. Nights 0 to 2 are represented in black, blue, and green, respectively. The values from the *full* dataset are represented in red. From top to bottom, the panels present the humidity, airmass, BERV velocity and the ambient temperature.

The major takeaway messages from this analysis is that the data selected for the stellar template plays a major role on the RV effect that we measure. Even though it is not clear why the same improvement is not retrieved in the three nights of observations, we find

that the usage of *multi*-night data allows for a drastic improvement. Previous analysis in this Chapter had already allowed to rule out possible systematics introduced by the flux interpolation. However, we have yet to rule out any modulation of the stellar flux that is present in the *nightly* templates. This analysis was left for future work.

3.8 Overall summary of the RV trend

During the course of this Chapter we have seen the effects of a RV bias that appears in template matching RVs as long as all observations were taken within a single observation night. Furthermore, the presence of such a large systematic bias raises questions around the impact that it is introducing into the non-nightly observations, typically used for the detection and characterization of exoplanets. From practice, we have yet to see a similar systematic effect on such observations, but it is possible that it is present at smaller peak-to-peak amplitudes.

Our initial attempts to diagnose and correct the issue have failed, but it has already allowed us to rule out some hypothesis. The most straightforward explanation – tellurics – was discarded, even if we do assume that the current models for Earth’s transmittance are missing molecular lines. The application of methods to both reject and correct tellurics, without a significant change in the measured variability, motivated their rejection as the main drivers of this effect. Furthermore, a chromatic analysis of the RVs over the night reveals a systematic shift across the wavelength coverage of the instrument. Notably, a larger amplitude was detected for a few spectral orders. However, through a division of each spectral order into three chunks, we find that the edges present a larger variability, when compared against the central region. This result can be connected with a variability of the stellar continuum, impacting the TM measurements but not those from the CCF. This would be explained by the fact that the CCF does not use information from the continuum, contrarily to template matching.

An analysis of the construction of the stellar template also reveals that the choice of observations plays an important role in the final pattern of variability. This is especially true when we consider a stellar template constructed from observations at either the start, middle, or end of the night. Under such conditions, we find a larger variability with the templates constructed from start and end of the night. In fact, our results seem to suggest that the template constructed with the observations near the middle of the night leads to the retrieval of an average of the effect seen on the others. However, if we consider a single

observation as our stellar template, we find that the extracted RV time-series presents a sharp jump near the hours of the selected observation, whilst the overall trend seems to be similar before and after such time. The fact that we see, in this case, a variability pattern which such behaviour might hint towards an issue with the interpolation. As the observations get closer to the hour of the observations used to construct the template, the line sampling of observation and template get closer, thus minimizing the impact of this bias on the measured RV. However, as we have seen with simplistic tests with Gaussian lines (both a single line and a collection of multiple lines) we estimate that the impact of the interpolation, coupled with the change of sampling in the lines, leads to effects at the order of the mm s^{-1} . Considering current, state-of-the-art instruments, this level of biases lay well below their expected precision. Furthermore, the RV trend has an impact at the meter per second level, well above the output from our simulations. From here we can draw one of two conclusions: i) our simplistic spectral model is not accurate enough to be able to properly retrieve the variability from real observations; or ii) our tests allow us to retrieve a RV effect that is not measurable in current spectrographs, but might come into play whenever we breach the cm s^{-1} barrier. Currently, we have yet to disentangle between the two explanations.

We have also analysed a large number of targets observed with ESPRESSO, allowing to search for correlations with both different observational parameters and properties of the host stars. Once again we find that the majority of the targets present a negative slope over the entire night, with its values being comprised in the interval $[-250, 0] \text{ m s}^{-1} \text{ day}^{-1}$. Not only do we find this large discrepancy, but we also fail to find any evident correlation with different parameters of the selected stars. Contrastingly with the majority of the stars in this dataset, one shows a positive slope in two nights. An initial comparison of the different instrumental parameters of this target, against those from HD40307, revealed no obvious differences. The sole exception is the stellar RV as, in this case, the CCF RV has a peak-to-peak variation of approximately 500 m s^{-1} , larger than ESPRESSO mean pixel size. This amplitude of RV variability will lead to the stellar spectra being measured in different physical pixels throughout the night. It is not yet clear if this is connected with the different sign of the slope.

Lastly, the analysis of the GJ9827 system proved itself a critical factor in our understanding of this issue. It allowed to determine that a template, constructed from multiple nights, appears to yield no RV trend when applied to single-night data. However, this is

only the case in two of the three available nights. Currently, we are yet to determine the difference between the nights, leading to some being affected and others not. One possible cause for such behaviour might be a modulation of the stellar flux, that is present during the course of one night of observations. We are currently evaluating the impact of the known effects that impact the flux during one night of observations (i.e., correction of the instrumental profile and the atmospheric extinction).

Despite our efforts to understand this issue, this is still an ongoing analysis, with the current hypothesis being:

1. A bias introduced by the interpolation algorithm that is applied to the spectra during the TM process, despite the inconclusive results from our tests. To further evaluate this hypothesis, we plan to test different interpolating algorithms based on the modelling of the entire spectral order, rather than just the closest neighbours.
2. A short-scale variability from the spectral continuum, that is biasing the stellar RVs. In this line of thought, we can explore both an *a priori* normalization of the continuum, or the usage of a different model to describe the differences between the template and the individual observations.

Chapter 4

A fully-Bayesian RV extraction model

In Chapter 2 we have seen the good results presented by the **s**-BART methodology on different spectrographs. However, and in spite of its performance, it still has some flaws that can be tackled, previously discussed in Chapter 2.6. In this Chapter we present the proof-of-concept of a methodology that tackles (or provides avenues to do so) those issues:

- The usage of a stellar template raises problems with both its assumption of time-invariability (common across all template matching approaches) and the sharing of information between model and data (in **s**-BART). By transforming the stellar template in a generative model of the stellar spectra we can both take into account flux changes and disentangle the model and the data with which it is compared with;
- Through the usage of a Bayesian framework we can achieve a consistent characterization of the uncertainties associated with the RV estimates;
- The inclusion of a simultaneous modelling of Earth’s transmittance (i.e., telluric features) will allow the correction of this contamination, allowing for the usage of more RV information. Furthermore, it will also provide an avenue to properly propagate the uncertainties associated with that process.

In this Chapter we describe the first steps towards the implementation of this rough outline of the model, with some core features being left for future work. We start, in Chapter 4.1 with a description of the generative model that we will use – Gaussian Processes – alongside a discussion on how they can be applied to large datasets. Then, in Chapter

4.2, the process under which the stellar and telluric spectra are modelled is described. Afterwards, in Chapter 4.3 the current implementation of the model is discussed, with detailed descriptions of the different parts of the RV extraction. With a full description of the model, Chapter 4.4 presents the results from the model when applied to both simulated and real datasets. Lastly, in Chapter 4.5 we discuss the current limitations of the model, the possible paths that can be taken to mitigate them, and the plans for future work.

4.1 A brief primer in Gaussian Processes

In a Bayesian mindset, we observe the existing data, trying to assess the likelihood of a given hypothesis holding true. In a more formal approach, Bayesian inference is the process of fitting a probability model to a set of data and summarizing the result by a probability distribution on the parameters of the model and on unobserved quantities, such as predictions for new observations (Gregory, 2005). Within this framework, we have to specify a prior distribution, reflecting previous information that is available, alongside our beliefs about the parameters before looking at the observations.

From Bayes' theorem we can calculate, through Equation 4.1, the posterior probability $Pr(H|D)$, for the hypothesis H , given data D . By using the prior probability of the hypothesis and data, $Pr(H)$ and $Pr(D)$, respectively, and the probability of data given the hypothesis, $Pr(D|H)$ we find the probability distribution of the unobserved parameter, conditioned on the existing data.

$$Pr(H|D) = \frac{Pr(D, H)}{Pr(D)} = \frac{Pr(D|H)Pr(H)}{Pr(D)} \quad (4.1)$$

Typically, Bayesian inference is based on random samples from the posterior. For that, we can use *Markov Chain Monte Carlo* algorithms, henceforth MCMC. In a summarized way, this method starts a chain in a location inside the parameter space and proposes a new point, which is accepted or rejected, depending on the posterior density. The proposals are made in such a way that convergence to the posterior distribution is guaranteed, albeit in some situations it may take longer to achieve it (Parviainen, 2018), depending on the complexity of the model. For a more in-depth overview we refer to Qian et al. (2003).

4.1.1 Gaussian Processes

Gaussian Processes (GPs) are a powerful tool to tackle regression and classification problems. On a regression context GPs can be seen as a distribution over functions that will be constrained by the N data points used as input data (Rasmussen & Williams, 2006). They are defined by a kernel, a $N \times N$ covariance matrix, where each entry describes the correlations between any two given points, and the median function, that describes the deterministic component of the data. The elements of the mean function and the covariance matrix are controlled by their respective parameters, also referred to as hyper-parameters. An example of a covariance function is the *Matérn* kernel, presented in the following equation.

$$K(\lambda, \lambda') = h^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu} |\lambda - \lambda'|}{\rho} \right)^\nu \mathbb{B} \left(\frac{\sqrt{2\nu} |\lambda - \lambda'|}{\rho} \right) \quad (4.2)$$

,where Γ is the gamma function, $\mathbb{B}$ is the modified Bessel function of second order, and ν, h, ρ are the kernel hyper-parameters. Typically, ν takes values of $(\frac{1}{2}, \frac{3}{2}, \frac{5}{2})$, with higher values resulting in smoother families of functions.

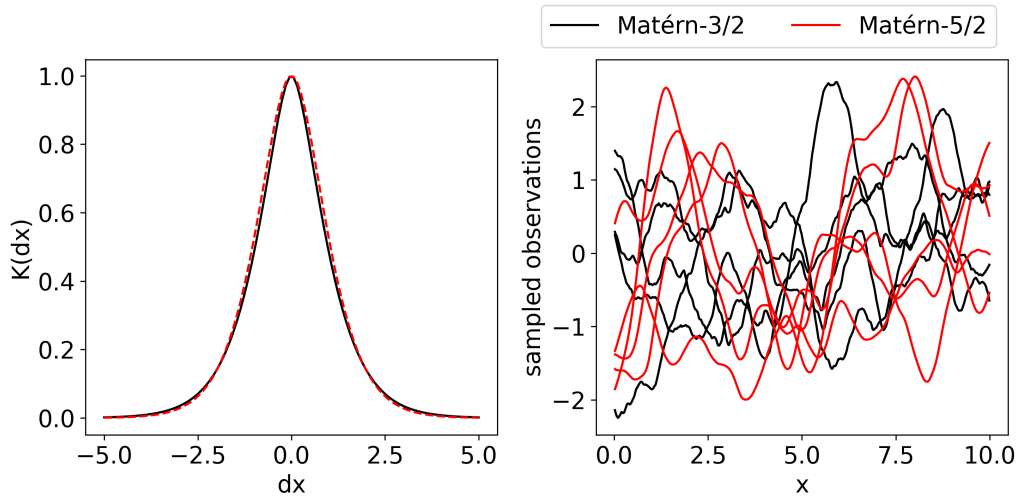


FIGURE 4.1: Comparison between 2 *Matérn* kernels, one with $\nu = 3/2$ (black) and another with $\nu = 5/2$ (red), with a length scale of 2 (arbitrarily chosen for visualization purposes). **Left column:** Evolution of the covariance with distance from a given point; **Right column:** Random samples from each kernel, without conditioning on data.

In Figure 4.1 we find a comparison between the *Matérn* 3/2 and 5/2 kernels which, in order to simplify the discussion, shall be referred as M32 and M52, respectively. The 7/2 kernel is missing as it is not currently implemented in the libraries that will be selected for the implementation stage of our model (*celerite* and *tinyGP*). From a first glance of the

Figure we find that the evolution of the covariance profile (left column) shows no significant differences, albeit the M52 kernel has a slightly higher correlation for an equal separation. The right side of the figure allows us to see some samples from the GP without conditioning it on data, i.e., a draw from the prior, allowing us to confirm that the family of functions from a GP with an M52 kernel is smoother than the ones associated with an M32 kernel.

As the model should be able to describe the input data we calculate, through Equation 4.3, the log marginal likelihood ($\mathcal{L}$), i.e., the probability of observing the data given the model.

$$\log(\mathcal{L}) = -\frac{1}{2}r^TK^{-1}r^T - \frac{1}{2}\log(\det(K)) - \frac{N}{2}\log(2\pi) \quad (4.3)$$

where r is the residuals of the input data and the mean function and K is the covariance matrix. The inversion of the covariance matrix, an $\mathcal{O}(N^3)$ operation, makes the evaluation of the marginal likelihood a computationally expensive process, particularly for the large data sets that we use in our case. In Chapter 4.1.2 we go over some of the available tools to make this problem computationally tractable.

The selection of the model, i.e., the choice of the optimal values for the GP hyperparameters, is typically performed through either a maximization of the log marginal likelihood function or a characterization of the posterior distribution through an MCMC approach. For further details on the theory behind Gaussian Processes we refer back to [Rasmussen & Williams \(2006\)](#).

4.1.2 Application of Gaussian Processes to large datasets

During the past years there has been a large focus on the application of GPs to large datasets, as their application is greatly limited by their high time ($\mathcal{O}(N^3)$) and memory ($\mathcal{O}(N^2)$) costs. For our desired use case - modelling of concatenated spectral orders – with just 10 ESPRESSO observations we reach a dataset with a size of $\sim 10^5$, assuming a single spectral order. As we have 170 orders, this scaling quickly turns into a major bottleneck in the definition of our RV model. The possible avenues for the application of GPs to large datasets can be roughly divided into two categories:

- Global approximations, where the GPs are applied to the entire dataset (albeit with different degrees of approximations to make them computationally viable)

- Local approximations, where the GPs are applied to "chunks" of data, artificially decreasing N .

The local approximations for the application of GPs often run into convergence issues and, in some cases, result in less accurate predictions, (e.g., [Gramacy & Apley, 2014](#), [Gramacy & Haaland, 2016](#)). Due to this, we focus on the global approximations, which also provide multiple possible frameworks that we can explore. There have been numerous efforts put into the optimization of GPs through the selection of a subset (of size d) of the full subset, also known as the active subset. Within this framework, [Quinonero-Candela & Rasmussen \(2005\)](#) presented a unifying framework for the major approaches that preceded it (e.g., [Csató & Oppier, 2002](#), [Smola & Bartlett, 2001](#), [Snelson & Ghahramani, 2006](#)). Despite the impressive performance of this type of algorithms, they often neglect the tuning of the hyperparameters, only focusing on the predictive performance. On top of that, another drawback with this methodology is that it is not always clear how to select the inducing points (or, in some cases, variables (e.g., [Hensman et al., 2013](#), [Titsias, 2009-04-16/2009-04-18](#))). As the predictive performance will not play a role in our model, we decided to discard this methodology.

If we are expecting short correlations in the input coordinates, one available option is to create a sparse matrix by setting the covariance to zero after it falls under a given threshold. This can then allow to apply algorithms optimized for handling sparse functions, both in terms of memory and computational costs. Such approaches can be found in ([Gneiting, 2002](#), [Melkumyan & Ramos, 2009](#)) but there isn't a readily available *python* implementation and they are often prone to capture local patterns in the data ([Liu et al., 2020](#)).

More interestingly for our use cases are the fast matrix-vector multiplication (*MVM*) algorithms, which typically exploit a known structure on the covariance matrix to speed-up operations. The usage of such algorithms can imply either approximations or limitations on the kernels that we can use ([Foreman-Mackey et al., 2017](#), [Gardner et al., 2018](#), [Gilboa et al., 2013](#), e.g.), or even limitations of the structure of the input coordinates to the GP ([Cunningham et al., 2008](#), [Gray, 2004](#), [Shen et al., 2006](#)). We refer back to [Chalupka et al. \(2012\)](#) for a comparison of the predictive performance of some algorithms and to [Liu et al. \(2020\)](#) for an up-to-date review of the available frameworks.

For this work we shall explore two MVM methods that rely on the structure of the covariance matrix. One of them is *celerite* ([Foreman-Mackey, 2018](#), [Foreman-Mackey et al., 2017](#)) that can be used to employ scalable kernels as long as they are composed by complex

exponentials. Furthermore, some kernels can also be approximated using this methodology, e.g., the *Matérn-3/2* can be approximated through:

$$k(\tau) = \sigma^2 \left[(1 + 1/\epsilon) e^{-(1-\epsilon)\sqrt{3}\tau/\rho} (1 - 1/\epsilon) e^{-(1+\epsilon)\sqrt{3}\tau/\rho} \right], \quad (4.4)$$

where σ is the amplitude of the kernel, ρ its length-scale, and ϵ the quality of the approximation. In the limit of $\epsilon \rightarrow 0$ we retrieve the exact expression for the *Matérn-3/2* kernel. The second method that we evaluate is *tinyGP* (Foreman-Mackey, 2023). In it, some kernels can be expressed as quasiseparable matrices (Vandebriel et al., 2008) that allow for the algebraic operations to scale linearly with the size of the dataset.

4.2 An overview of the components from the model

Now that we have the statistical tools needed to start implementing this RV-extraction model, we can divide the work in two major areas:

- Design a framework capable of accounting for telluric features, described in Chapter 4.2.1.
- Model the stellar spectra using Gaussian Processes, a process that shall be discussed in Chapter 4.2.2;

In this sub-Chapter we strike to discuss a general framework under which those two components will be implemented, with Chapter 4.3 providing a full description of the model and its inner details. To ease the discussion, Table 4.1 presents the notation adopted for the remainder of this Chapter to refer to the different free-parameters of the model.

TABLE 4.1: Notation used to described the GP model that is presented in this Chapter.

Parameter	Description
N	Number of observations
M	Number of orders in each observation
S_i^j	Original spectra of j -th order of the i -th observation
S^j	Original spectra of the j -th order across multiple observations
$\phi = [\phi_0, \dots, \phi_N]$	Parameters of the telluric model from the i -th observation
$\vartheta = [\vartheta_0, \dots, \vartheta_M]$	Parameters of the kernel used to model S^j
$\gamma = [\gamma_0, \dots, \gamma_N]$	Parameters of the mean function that models S^j
$\nu = [\nu_0, \dots, \nu_N]$	Relative RVs between all observations

4.2.1 Accounting for the modelling of telluric features

As we are working within a Bayesian framework, we can account for the uncertainty in the modelling stage if the parameters of such model are taken to be free-parameters in our RV model. From Allart et al. (2022) we know that it is possible to fit a telluric model to a given molecule using a carefully selected set of the deeper telluric lines associated with it. Furthermore, it is also shown that the generation of such a model only calls for the introduction of two free-parameters per molecule that we wish to include: molecule content along our line of sight; formation pressure of the lines. In the optical domain, i.e., the one covered by ESPRESSO, the majority of such features are introduced by two molecules: O_2 and H_2O . However, the methodology presented in (Allart et al., 2022) was only tested and validated for the water absorption features. The oxygen features were not accounted for and, instead, are masked with a binary mask.

We can thus correct telluric features by applying the same methodology, following a set of steps for each observation:

1. Use *Telfit* or *Tapas* to generate a synthetic transmittance model, T_i^j , for each spectral order (S_i^j), using ϕ_i ;
2. Compare the selected set of lines in the template with those from the spectra, to evaluate the likelihood of the telluric model;
3. Generate a corrected stellar spectrum from the synthetic transmittance model (S_i^j/T_i^j);

4.2.2 Modelling stellar spectra

The generation of a probabilistic model of the stellar spectra will follow a similar methodology to the ones from Czekala et al. (2017), Rajpaul et al. (2020) - Modelling the stellar spectra using Gaussian Processes (GP). Even though the usage of a GP is common to both works, it is applied differently: in Czekala et al. (2017) a GP is used to model the same spectral order throughout time, S^j , aligned in the same reference frame. This time-evolving spectral order is concatenated and modelled with a Gaussian Process with shared hyper-parameters for all observations. On the other hand, Rajpaul et al. (2020) performs a pair-wise modelling and comparison of the spectral orders, S_i^j . As a consequence of the $\mathcal{O}(n^3)$ time scaling of GPs the two methods share a common flaw: a prohibitive computational cost that does not allow to model more than a few small chunks of stellar spectra.

As discussed in Chapter 4.1.2, during the past years there has been a push towards the application of GPs to big data, allowing their usage on cases with millions, or even billions of data points. This means that the previously costly operations can now be applied at a relatively cheaper cost. Thus, we will implement a model similar to the one from Czekala et al. (2017). Under such model, the size of our input data set (for the GP of each order) will be $N_{\text{observations}} * N_{\text{pixels per order}}$. To contextualize, for ESPRESSO* we would need 110 observations to reach a dataset with one million points. As this size is well below the limits reported in the literature, we can write down the rules that will guide our probabilistic model of the stellar spectra:

- We assume that the same spectral order will be at all times (i.e., in each observation) reproducible by a covariance structure with the same amplitude (σ) and length-scale (ρ).
- Pixels in the same local neighbourhood are strongly correlated with one another, with a reasonably short length scale, i.e., far away pixels are very weakly correlated. To describe this correlation structure we select a *Matérn* kernel as our covariance function; This kernel has two parameters, the amplitude and the length-scale, individually defined for each spectral order. We showcase the modelling of one spectral order with this kernel, in Figure 4.2, where we see that its predictions can capture the stellar features.
- The mean function of our model will be a low-order polynomial, used to describe any variability in the continuum between any given observations. We select a first degree polynomial for this purpose, following the good results from the **s**-BART methodology. This choice implies usage of blaze-corrected observations, i.e., the observations that have had the instrumental profile corrected, to ensure a flat continuum level.

It is also important to note that this preliminary approach keeps the assumption of a time-invariant stellar spectra, i.e., there is no flexibility of the covariance structure to accommodate time variability of the data. Furthermore, this model could also be expanded to take into account the known correlation lengths of the instrument and even include a wavelength-dependent white-noise term. This was left for future work and in Chapter 4.5 we further describe the options to adapt the model to account for stellar activity, either through a direct modelling of the fluxes or through the definition of the model for the RVs.

*Each ESPRESSO order has 9111 pixels

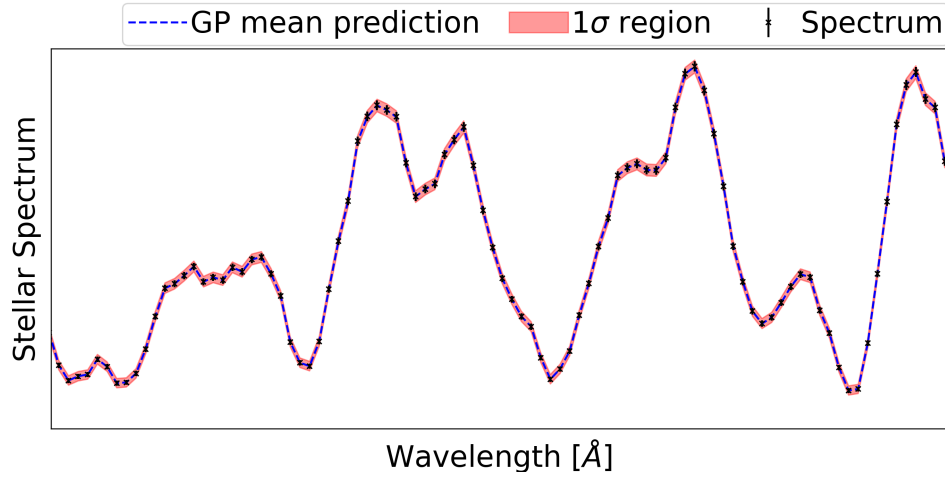


FIGURE 4.2: GP prediction for a small region of stellar spectra, using a *Matérn-5/2* kernel. The black crosses represent the flux measurements, the dashed blue line represents the mean prediction of the GP, and the red region the 1σ uncertainty of the predictions.

4.2.3 Evaluating the degrees of freedom of the model

At this point of the design of the RV-extraction model, we can already compute the free parameters that we will have to characterize. From Chapter 4.2.1 we saw that we can handle the telluric features with a minimum of 2 free parameters per observation, if we only consider water absorption features. Then, from Chapter 4.2.2 we find that the modelling of stellar spectra will need: i) 2 free parameters per order, for the GP kernel; ii) 2 free parameters per order of each observation, to use a first degree polynomial as the mean model. Lastly, if we consider an achromatic RV shift per observation, it will add one free-parameter per observation. The assumption of achromatic RV shift is kept from **s-BART** as it reflects the expected RV signal from a planetary companion. However, it will be further discussed in Chapter 4.3.3.

TABLE 4.2: Number of free parameters in the model, as introduced by its different components, for a different number of observations and spectral orders.

Source	$N_{OBS} = 15$		$N_{OBS} = 25$	
	$N_{order} = 10$	$N_{order} = 20$	$N_{order} = 10$	$N_{order} = 10$
Telluric model	30		50	
RV	15		25	
continuum	300	600	500	1000
kernel	20	40	20	40
total	365	685	595	1115

With this in hand, Table 4.2 compiles the evolution of the free-parameters with the number of observations (N_{OBS}) and the number of spectral orders that are being modelled (N_{orders}). From there, we see that even with a modest number of observations ($N_{OBS} = 15$) and a small wavelength coverage (only 10 spectral orders considered) we already have more than 300 free parameters. This will harden the convergence of any technique that we use to characterize the posterior distribution. Furthermore, we see that our parameter of interest (the RVs) is the smallest contributor to the number of free-parameters in the entire model. This fact is used, in Chapter 4.3, to motivate the usage of data pre-processing to remove some of them.

4.3 In-depth description of the RV-extraction model

Now that Chapter 4.2 presented the two main components of this RV model, we can go into an in-depth description of how it works, coupled with the assumptions and simplifications that must be applied to ensure that we are left with a tractable model. As previously mentioned, the notation of Table 4.1 will be used when necessary.

The process of extracting RVs will be characterized by the ν_i RV shifts between observations, such that if we Doppler-shift all observations by the corresponding ν_i , they will be placed in a common rest frame, i.e., the stellar lines will be aligned. Our initial idea for the RV model is schematized in Fig 4.3 and can be summarized as a sequence of sequential steps for each set of free-parameters explored in our characterization of the posterior distribution:

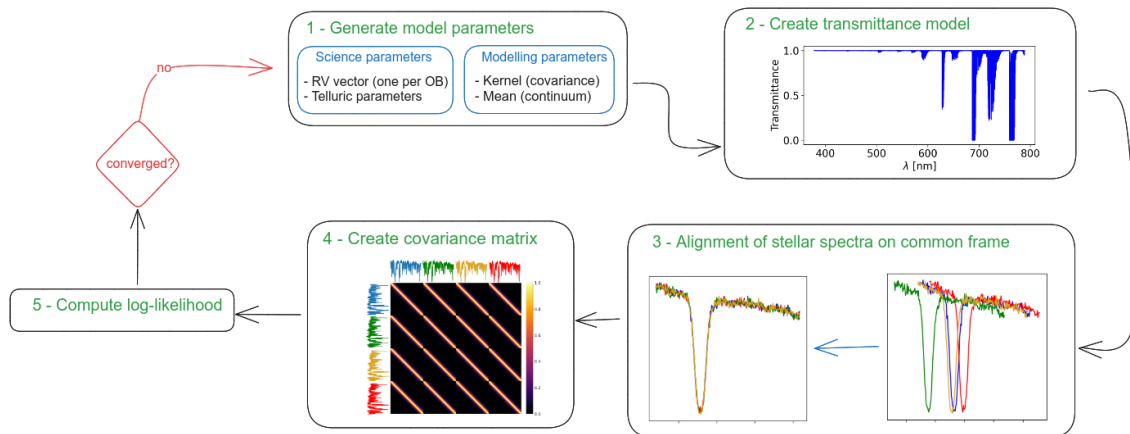


FIGURE 4.3: Simplified schematic of the fully-Bayesian RV extraction methodology, modelling the stellar spectra and telluric features to retrieve the relative RV shifts between multiple observations.

1. A set of tentative values (for each of the free-parameters) is provided by a given methodology (to be discussed in Chapter 4.3.4);
2. We apply the telluric model generation and correction process from Sec 4.2.1, using the current ϕ_i values;
3. Each telluric-corrected spectral order is Doppler-shifted by the proposed relative radial velocity, ν_i , and the concatenated observations are modelled.
4. This sequence of steps is repeated until a given convergence criterion is met.

In this Chapter we shall start by discussing the selection of the methodology to apply a GP to such a large dataset (Chapter 4.3.1), followed by a discussion on the mean model (Chapter 4.3.2), the assumptions behind the RV vector (Chapter 4.3.3), and how to explore the parameter space (Chapter 4.3.4). The characterization of the posterior distribution is discussed in Chapter 4.3.4 and Chapter 4.3.5 presents the final architecture of the system, after accounting for the approximations and simplifications detailed in this Chapter.

4.3.1 Choice of the GP model for a large dataset

With a framework for our RV-extraction model, we can now consider the details of its implementation. From the discussion within this Chapter, we have already seen that the major bottleneck lies in the usage of Gaussian Processes, due to the large sizes of the concatenated spectral order. Consequently, and as briefly discussed in Chapter 4.1.2, we must make a proper choice of methodology to handle such data sizes.

In this Chapter we evaluate two methodologies for the application of GPs to large datasets: *celerite* (Foreman-Mackey, 2018, Foreman-Mackey et al., 2017) and *tinyGP* (Foreman-Mackey, 2023). In this sub-Chapter we compare the computational efficiency and the parameter space that both approaches can provide, when handling datasets with sizes similar to our use case.

The first test between the methodologies is to evaluate the computational efficiency of one log-likelihood call, as that will be the critical portion of our model. Table 4.3 presents the average time per call, assuming 1 spectral order from up to 15 ESPRESSO observations (recall that each order has 9111 pixels). From here, we can see that the *celerite* methodology produces results more than 200 times faster than *tinyGP*.

TABLE 4.3: Average and standard deviation of the time, in seconds, spent on evaluating 15 times the log-likelihood for datasets with different sizes. The datasets considered have sizes that are multiples of one ESPRESSO order (9111 pixels) and the log-likelihood was computed through *celerite* and *tinyGP*.

Method	Number of ESPRESSO observations		
	5	10	15
<i>celerite</i>	0.0046 ± 0.0001	0.0087 ± 0.0004	0.0119 ± 0.0002
<i>tinyGP</i>	1.3843 ± 0.0038	2.7807 ± 0.0094	4.1703 ± 0.0365

This obviously translates into a large penalty to performance, with a further investigation into it being left for future work. However, it is also important to mention that we were not able to retrieve the same computational performance that is showcased on *tinyGP*'s documentation. This discrepancy was attributed to a misconfiguration of the CPU-bound version on our computational cluster* and solving it was also left for future work.

celerite provides an exact computation of a few different kernels, whilst others can be built as a sum of multiple *celerite* terms (for further details on their implementation we refer to the corresponding papers). The *Matérn-5/2* (M52) kernel, that we desire to use due to its smoothness, is not available in the library but, instead, we have access to the *Matérn-3/2* (M32) one. Thus, in our evaluation of the optimal GP approach, we use the *celerite* implementation of the M32 kernel, which we shall test for the validity of the approximation and associated numerical instabilities than can occur. Within *tinyGP* some kernels can be represented by quasiseparable matrices, which allows for a linear scaling in the algebraic operations. From this library, we have access to both kernels (M32 and M52), which we use for benchmarking against the values yielded by *celerite*. Furthermore, as the range of the wavelengths values will be much larger than the one from the normalized fluxes, we also evaluate the effect of rescaling the wavelength solution of each order(λ_o):

$$\lambda'_o = \frac{\lambda_o - \bar{\lambda}}{10} \quad (4.5)$$

, where λ' is the rescaled wavelength and $\bar{\lambda}$ is the mean wavelength of all observations of that order. We have to use a common value to every observation as, otherwise, we would be introducing an RV shift.

*Issues with the CPU-bound performance of *jax* is the topic of some *issues* in its repository;

The tests of the three kernels are carried out with 15 ESPRESSO observations of $\tau - Ceti$ which are concatenated with the procedure from Chapter 4.3, taking the CCF RV for the alignment on a common referential. Then, we explore the parameter space of the hyper-parameters associated with the spectral modelling. We evaluate the amplitude and length-scale of each kernel on a grid of possible values, setting $\epsilon = 1e - 20$ in the *celerite* kernel, and using both the original wavelength solution and the one modified through Equation 4.5. Figure 4.4 presents the parameter space for the three kernels, allowing us to find that the M32 kernel from *celerite* can run into some numerical issues relatively close to the maximum-likelihood point (indicated in the intersection of the two dashed lines). However, it is important to note that the re-parametrization of the wavelength solution through Equation 4.5 slightly increases the values for which we are able to retrieve a finite likelihood. Furthermore, the choice of ϵ was driven by the (manual) minimization of the number of numerical errors of the employed approximation.

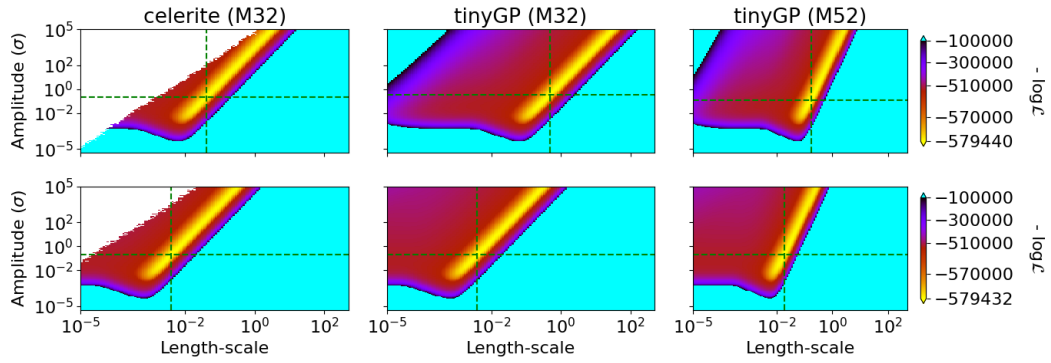


FIGURE 4.4: Negative log-likelihood evaluated in a grid of possible parameters, using three GP kernels: M32 from *celerite* (left column) and *tinyGP* (middle column), and M52 from *tinyGP*. The GP model was applied to one spectral order of 15 ESPRESSO observations of $\tau - Ceti$, aligned assuming the CCF RVs. The blue values represent those where the log-likelihood is minute, when compared against the optimal one (highlighted in the intersection of the green dashed lines). The white regions in the plot are those for which the *celerite* implementation failed to produce a valid result, and Table 4.4 presents the values of the hyper-parameters at the maximum-likelihood point. The top row present the results when using the original wavelength solution, whilst the bottom one uses those transformed through Equation 4.5.

Most striking is that we find a correlation between the kernel's amplitude (σ) and length-scale (ρ) that appears in all kernels and in the two wavelength solutions. If we now compare the hyper-parameters of *celerite* with those from *tinyGP*, we see a significant difference, at least when considering the original wavelength solution (see Table 4.4). It is unclear the cause behind this difference, but we posit that it is attributed to the numerical instability of this specific kernel. Fortunately, if we now consider the case where

the wavelength solution is transformed through Equation 4.5, there is a very good agreement between the two implementations. Not only do we retrieve the same values at the maximum-likelihood point, but even the parameter space (refer back to Figure 4.4) appears to be very similar.

TABLE 4.4: Comparison of the hyper-parameters (kernel amplitude and length scale) associated with the maximum log-likelihood in the grid search presented in Figure 4.4.

λ type	celerite2 (M32)		tinyGP (M32)		tinyGP (M52)	
	amplitude (σ)	LS (ρ)	amplitude (σ)	LS (ρ)	amplitude (σ)	LS (ρ)
Original	0.2226	0.0407	0.3294	0.4874	0.1158	0.1714
Centred	0.1953	0.0039	0.1953	0.0039	0.2226	0.0186

In spite of the computational cost associated with the *tinyGP* method, its application to datasets is still feasible. Thus, it will be the selected approach for the remainder of this Chapter. This is mainly motivated by the option to use the M52 kernel without facing the numerical issues that are brought by the *celerite* method, keeping our model in line with the ones that already exist.

4.3.2 Mean model of the Gaussian Process

The previous analysis of the number of free-parameters of the model (refer to Table 4.2) has shown that the continuum model is their largest source. This means that if we consider a first-degree polynomial as the mean, and wish to simultaneously model multiple observations, we are trying to achieve convergence in over than more than 300 parameters. This will be a daunting task and, consequently, we attempt to avoid it.

For this purpose, we run a pre-processing stage of normalization of the stellar spectra, through the application of a custom implementation of RASSINE (Cretignier et al., 2020b). This methodology uses the alpha-shape technique to determine the continuum level of the stellar spectrum. We use stellar spectra in an order-by-order basis, which is supported by our implementation of the RASSINE algorithm, named *Spectra Normalization Tool* (SNT). This implementation was written by a bachelor student, Diogo Marques, under the supervision of Dr. Nuno Santos and I.

Thus, the normalization of the stellar spectra occurs in an order-by-order basis, following a set of consecutive steps:

1. Apply SNT to each individual spectra, S_i , determining its continuum level, C_i ;

2. Normalize the stellar spectra by the corresponding continuum, $S_i^{norm} = S_i/C_i$, updating the flux uncertainties, σ_i , to account for the flux change: $\sigma_i^{norm} = \sigma_i/C_i$;
3. To ensure that we don't have different mean levels in different observations, we divide S_i^{norm} by its own mean, ensure that all resulting spectra has the same mean level of 1.

Through the application of this methodology, we now start our RV-extraction process with normalized stellar spectra, thus removing the need of including free-parameters to account for a varying spectral continuum across time and allowing us to set the mean value of the GP to a fixed value (1, in this case). This process, albeit helpful for achieving convergence, comes with a few caveats and drawbacks. Firstly, it hardens the applicability of the model to M-dwarf stars, where it is much harder to determine the continuum level and, consequently, normalize it. Lastly, as this process is being applied as a pre-processing stage, it means that uncertainties associated with the modelling of the continuum will not be accounted in the uncertainty of the estimated radial velocity.

4.3.3 The parameter vector of Radial Velocities

In the quest for convergence, our model will evaluate multiple iterations of the RV vector, ν , which will contain one RV per observation. As the observations are not compared against a fixed model (i.e., a template) we are only able of retrieving relative RV shifts, as there is no defined reference. The consequences of this limitation is that we are unable to have individual RVs for each spectral order, as there is no guarantee that they are indexed to the same zero-point. Thus, by design, we are limited to the same core assumption around which **s-BART** was built: the RV shifts that we are trying to characterize with our model should be independent of the wavelength at which they are measured, assuming that they are keplerian in nature.

There is one important detail to keep in mind: the GP libraries in use assume (both *tinyGP* and *celerite*) that input coordinates are sorted in increasing order. As our model evaluates different combination of RVs it is likely that the incremental order is not maintained. Thus, each concatenated spectral order is sorted after the original wavelengths are Doppler-Shifted by the corresponding ν_i , ensuring that the association between wavelengths and fluxes are maintained after the procedure.

4.3.4 Exploring the posterior distribution of the RVs

The characterization of the RV posterior distribution is an important step of the RV extraction process. Ideally, we would apply an MCMC approach to properly characterize the posterior distributions. However, such an approach faces two significant challenges that must be surpassed:

1. The current implementation of the model, albeit computationally tractable, is yet unable to handle the number of calls of the log-likelihood that would be needed to achieve convergence, taking into account the number of *walkers* and iterations needed to achieve convergence.
2. The comparison of the RV vector across multiple iterations is not a straightforward endeavour. One must consider that using the RVs as a free parameter leads to an high number of degenerate solutions. This issue is clear when we think about two RV vectors, where one differs, from the other, only by an offset. This suggests that the relative shift between observations can be the same, whilst the minimization procedure considers them to be distinct.

In order to construct a model that we are able to test, under controlled conditions, we avoid using an *MCMC* algorithm. Instead, the preliminary implementation of our model is based on the minimization of the posterior distribution, which is carried out in two steps:

1. First, we use a differential evolution method (Storn & Price, 1997) for a preliminary exploration of the parameter space. This method constructs a population of possible solutions, that are iteratively improved, using the posterior distribution as the optimization metric. A population size of 60 candidate solutions is used, with the initial values being distributed within the parameter space to be explored. On top of that, we swap one of the initial candidates by a user-given solution, to which we shall refer as the “*initial guess*” for the minimization stage. The remainder parameters of the algorithm are kept the same to the default values of *scipy*’s implementation.
2. The optimal solution provided by the differential evolution method is then used as the initial guess for a bounded minimization, carried out through Brent’s method. The bounds for the minimization are defined on a case-by-case basis, constraining the possible values of the RV shifts for which the log-likelihood will be computed.

The comparison of distinct RV vectors is the most pressing issue, and two distinct solutions are hypothesized:

1. We use one of the observations as the reference, fixing its RV to zero. Thus, by default, the model should prefer to align the other observations with this one. The drawback of this approach is, obviously, the loss of RV information from one observations (i.e., from N observations, we can only extract $N - 1$ RVs.)
2. Instead of using the RVs as free parameters, we attempt to directly model the number of keplerian signals in the RV time-series. Thus, the RV vector would be constructed in each iteration, using: i) a free parameter to control the number of orbital companions, and ii) the free-parameters that would be associated with the orbital parameters of each companion. Furthermore, using such a framework, the number of free-parameters in the model would be agnostic to the number of observations allowing its applicability to larger datasets.

Ideally, we would adopt the second solution, as it is the most complete one. However, due to a lack of time, it was not possible to implement it. Thus, for the remainder of this Chapter, we use the highest-SNR observation in our dataset as the reference for the RV vector, with the RV vector being given by:

$$\nu = [0, \nu_1, \dots, \nu_N] \quad (4.6)$$

Lastly, it is also important to note that the subsequent analysis will refrain from estimating the uncertainties associated with each observation. One option would be to blindly apply a Laplace approximation to the posterior distribution. However, as described here, we are yet to truly characterize this distribution, leading to no guarantee of its shape. Consequently, we have no basis to impose an assumption of gaussianity. This is left for future work, after solving the problems related to the comparison of RV vectors across optimizations

4.3.5 The final architecture of the model

After implementing all of the details until now, we are left with a model where the only free parameters are the radial velocities and the kernel hyper-parameters. For the sake of clarity, Figure 4.5 presents a schematic of the full process. Furthermore, it also accounts

for a future expansion of the model (box 5b) that will allow to remove the parameters associated with the kernel. Chapter 4.5 will discuss this process and its benefits.

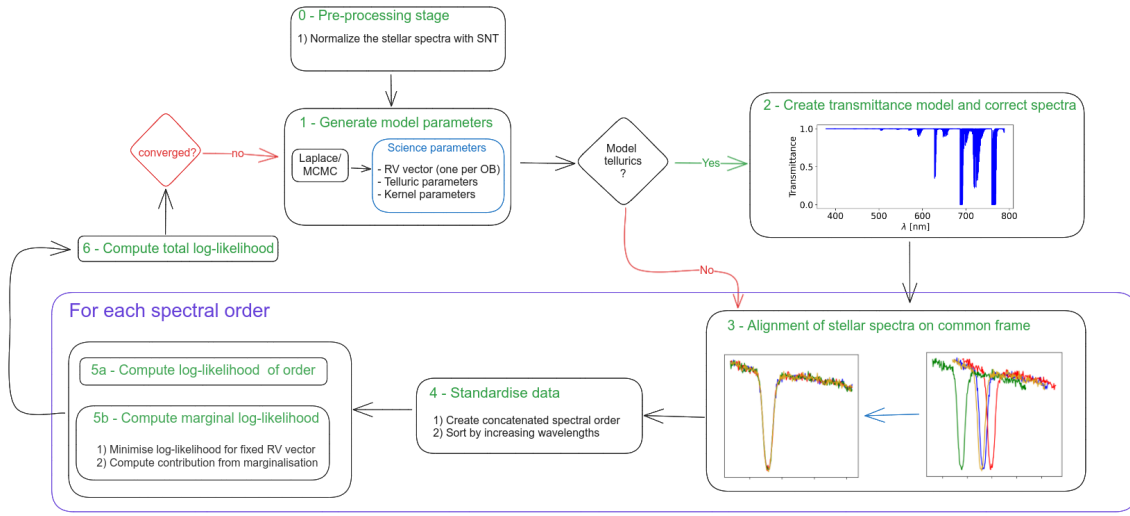


FIGURE 4.5: Schematic of the entire GP model, considering the different steps presented in this Chapter. For the current implementation of the model, the telluric modelling is fully disabled.

The process is also fully described here:

1. The model starts with the pre-processing of the observations, normalizing them with SNT.
2. Our approach to characterize the posterior distribution (detailed on Chapter 4.3.4) provides a tentative vector ν , with the relative radial velocities between observations. It also provides the kernel hyper-parameters, θ , for each spectral order. If the telluric-modelling is enabled*, it also provides the ϕ vector.
3. If the telluric-modelling is enabled, *telfit* is applied to generate the transmittance spectra associated with each observation. This spectra is then used to correct the corresponding observations. This step could also be improved by making use of telluric residuals obtained in hot stars (telluric standards) to account for features that are not present in the synthetic models.
4. For each spectral order, the following steps are then applied:
 - (a) Doppler-shift the wavelengths by the ν_i value associated with the observation, placing each observation in the rest frame defined by ν . This implies the assumption of an achromatic RV shift across all spectral orders of the same observation;

*As mentioned in this Chapter, the telluric modelling is currently disabled.

- (b) Sort the Doppler-shifted wavelengths (ensuring that the association to the fluxes and uncertainties is not changed), so that the GP will be evaluated in strictly increasing points.
5. The log-likelihood of each spectral order is summed, to compute the one associated with the tentative RV vector ($\mathcal{L} = \sum_{i=0}^{N_{orders}} \mathcal{L}_i$);
6. The convergence of the model is evaluated, repeating back from step 2 if the criteria was not met.
7. The RV of each observation, and the associated uncertainty, are estimated through the mean and standard deviation, respectively, of the posterior distribution of the observation. Currently, as we are not approximating the posterior distribution, the RV of each observation is obtained directly from the ν vector, without any estimation of uncertainties

4.4 Application of the method

The model that has been presented up until this point promises to be one that is capable of tackling the majority of the most pressing issues in the quest for precise RVs. However, it has a high complexity and we must evaluate its feasibility. Furthermore, as we have seen until now, multiple simplifications and approximations were applied. Thus, within this Chapter, we shall refer to this model as the *s-GP* model.

4.4.1 Preliminary results in a ideal case

Our first test of the *s-GP* model is through a synthetic dataset, under controlled conditions. Its generation is done in accordance with the following prescription:

1. We generate N_{OBS} synthetic spectra following the methodology presented on Appendix C.
2. Generate N_{OBS} RVs, drawing random samples from a Gaussian distribution centred on 0 and a standard deviation equal to a given *amplitude*. The RVs are then injected into the synthetic spectra.
3. This process is then repeated 15 times, proving 15 independent datasets, each with N_{OBS} spectra.

To test the methodology under different strenghts of signals, the process was repeated for an array of 3 different amplitudes (100, 10, 1 m s^{-1}). Furthermore, for control reasons, the prescription was also applied without the flux-resampling (i.e., the same observation was copied multiple times, with an RV-shift being injected in each “copy”). To decrease the computational burden of this test, it was carried over with one spectral order from the HARPS spectrograph.

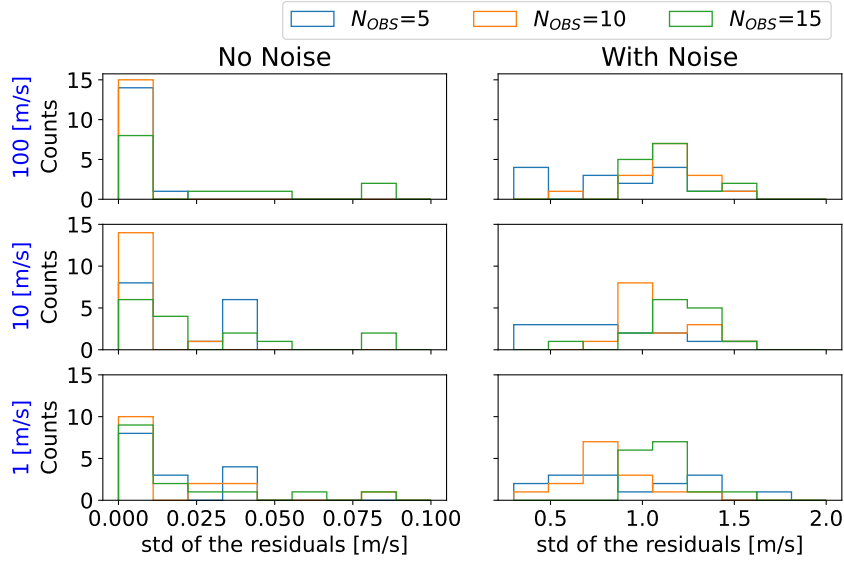


FIGURE 4.6: Histogram of the residual scatter obtained in the datasets that were constructed with different *amplitudes*, without the re-sampling of the flux (left column) and with it (right column). The top row represents the datasets with an injection *amplitude* of 100 m s^{-1} , the middle those with 10 m s^{-1} and, the bottom one with 1 m s^{-1} .

For each dataset, we apply our model to it and compare the retrieved RVs with the ones that were injected, as shown in Figure 4.6. Uniform priors were selected, covering the window of $[-400, 400] \text{ m s}^{-1}$, with the initial guess of the RVs being a random draw from a Gaussian distribution ($\mathcal{N}(\mu = 0, \sigma = 50 \text{ m s}^{-1})$). Starting with the datasets for which the flux was kept the same across simulated observations we find that we are able to closely retrieve the injected signals. This can be attributed to the lack of noise in the data, easing the retrieval of the signal. The datasets in which the spectral flux was resampled (right column) paint a different picture. The small changes in the stellar spectra are reflected in a larger discrepancy between the retrieved and injected signals. That said, we find that the performance of the model is consistent in the three injected amplitudes. Furthermore, the expected RV precision on this spectral order is of 1.3 m s^{-1} , which can explain the obtained scatters. It is also noteworthy that across all datasets, the scatter of the residuals is slightly higher for those comprising 15 observations. The larger number of free-parameters in the

model hardens the convergence towards the optimal solution and, consequently, leads to higher errors in the retrieval of injected signals.

Despite the strong assumptions enforced in the model, this simplistic analysis paints a promising scenario for the applicability of this model to real datasets.

4.4.2 Application to real observations

As previously described in this Chapter, the current computational challenges limits the application of the current implementation of the model to more than one spectral order. Thus, its applicability to a real dataset poses a significant challenge. For such purpose, we once again selected HARPS observations of *51 Pegasi* and extracted RVs from one spectral order. Three different approaches were employed for the RV extraction: i) the CCF RVs provided by the DRS, computed using the entire spectra; ii) **s-BART** RVs from the selected spectral order; iii) RVs from the GP model. Furthermore, to limit the number of free parameters for the *minimization* stage of the model, we select small samples of the available observations.

Two datasets were constructed, one from 5 and another from 11 observations, as presented in Figure 4.7. To allow for a comparison between the different methods, we remove the RV that it extracted from the observation selected to be the reference . Thus, every time-series is relative to the same observation, allowing for a direct comparison of their behaviours.

Similarly as before, we find that on the smaller dataset the retrieved RVs are closer to those from the other methods. However, in both cases, we find that there is an offset across all methods (GP, CCF and template matching). That said, we retrieve the same behaviour of RVs in each method, with this being clearer on the larger data-set. Furthermore, a comparison of the RV scatter also shows similar dispersions across methods, with the closeness being larger in the smaller dataset. We find that the GP model provides RVs with the same tendencies as **s-BART**, albeit with a vertical offset. This can also be interpreted as a disagreement of the RVs that each method extracts from the reference observation. It is also important to note that the GP model yields a slightly larger scatter on the larger dataset, possibly a consequence of difficulties in achieving convergence in the model. This issue will be revisited in the future, after improving our characterization of the posterior distribution.

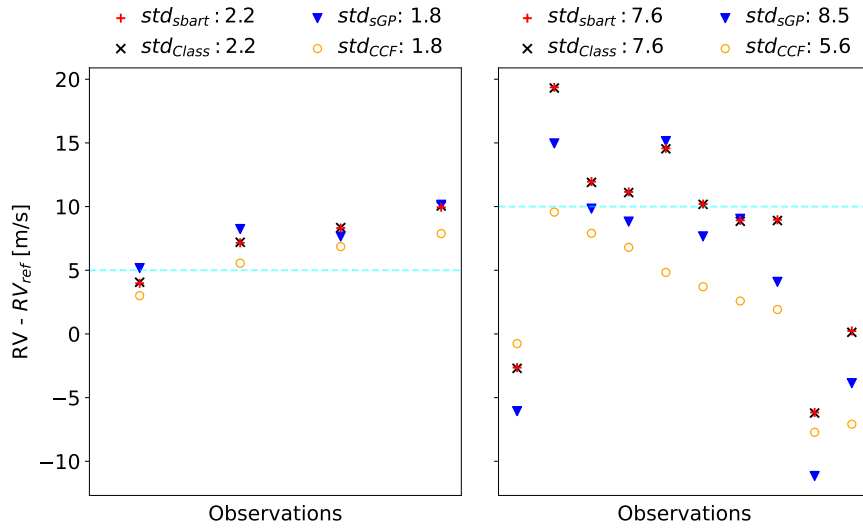


FIGURE 4.7: Comparison of the relative RVs obtained through a classical template matching algorithm (back), **s-BART** (red), CCF (yellow) and the GP model (blue), obtained from the datasets of 5 (left) and 11 (right) HARPS observations. All RVs, except those from the CCF method, were obtained using only the selected spectral order. The observation selected as a reference for each dataset (left and right columns) are not represented and, in both cases, a constant initial guess was used (dashed blue line). The label of each column provides a measurement of the standard deviation of the RVs from the corresponding method.

To conclude, and similarly to what was found in the simulated datasets, our simplistic GP model was able to retrieve RVs that appear to be compatible with those from **s-BART** and the CCF. Obviously, and despite the good preliminary results, a larger number of facets of the model are left for future work.

4.5 The caveats and avenues for expansion of the current model

With the preliminary results presented in this Chapter, we are moving towards the completion of this idealized model. However, some parts of it are still under development, whilst others are left for an even further future.

Accounting for stellar activity

Accounting for stellar activity will be a daunting task, with two possible avenues that we can take: i) Model the time-variability of the stellar spectra; ii) Change the tentative ν vector to a sum of keplerian and activity driven signals. If we decide to model the flux

impact of the stellar activity, we will either have to properly simulate it, or add a time-dependency on our GP kernel. Neither of those two options are straightforward and, on top of that, the addition of a new dimension to the GP modelling will bring extra challenges both on the convergence of the model and on the time complexity associated with the log-likelihood. The later option ii), would revolve around keeping the assumption of no time variability but, instead, model the activity signal directly on the RVs. Thus, the RVs would be given by:

$$RV_{final}(t) = RV_{OBS} + SA(t) \quad (4.7)$$

, where $RV_{final}(t)$ is the RV used for the alignment of the observations, RV_{OBS} is the current ν vector, and $SA(t)$ is a model for stellar activity that would introduce extra free parameters. Undertaking such an approach would in fact remove one step in the *a posteriori* analysis of the RV series, as we would no longer be summarizing the spectral information into RVs before searching for keplerian signals. It is yet to be tested if this direct search for signals would translate into a meaningful improvement. Such an endeavour will be further explored in the future.

A need to further decrease the number of free parameters

Secondly, even though we managed to decrease the number of free parameters to be the same as the number of observations, we might run into issues with exploring the posterior space. Due to the low signals that we are trying to search for, which are also hidden beneath the stellar contribution, we often need more than 50 observations of a given target to properly characterize it. One option to decrease the number of parameters is to model directly the number of planets and their physical parameters, using those to construct the estimated Radial Velocity, as previously described. This will be tested in the future, both in the context of this fully-Bayesian model and in **s-BART** itself.

Furthermore, we also have the free parameters associated with the GP kernel, amounting to two free parameters per spectral order. As our main goal lies in the extraction of precise RVs, the parameters associated with the modelling of the stellar spectra can be considered to be nuisance parameters. Thus, we should be able to marginalize over them, leaving only the RVs as the free-parameters in this model. If we apply the Laplace approximation to the hyper-parameters (σ and ρ) of the selected kernel, we will be able to

marginalize over them. Following Section 11.3 of [Gregory \(2005\)](#) we can write an expression of the global likelihood ($p(D|M, i)$) after marginalization:

$$p(D|M, i) \approx p(\hat{\theta}|m, I) \mathcal{L}(\hat{\theta}) (2\pi)^{M/2} \det(\mathbf{I})^{-1/2} \quad (4.8)$$

, $p(\hat{\theta}|M, I)$ is the prior distribution, $\hat{\theta}$ represents the parameters associated with the MAP, $\mathbf{I}$ is the Fisher information matrix evaluated at that same point and M is the size of the dataset.

This process of marginalization of the global likelihood would be performed for each vector, v , of tentative radial velocities, using the wavelengths that were placed in a common rest frame. The determination of $\hat{\theta}$ is done through the minimization of Equation 4.8, using the *L-BFGS-B* method implemented in *scipy*. Lastly, the Fisher matrix, $\mathbf{I}$, is obtained through the Hessian of the negative $\mathcal{L}(\hat{\theta})$, often provided by the minimization procedure.

Improving the modelling of the stellar spectra

The current approach to handle the stellar continuum is sub-optimal, limiting the stars that we can analyse with this methodology and leads to the loss of RV information as we use the *SID* spectral format. A more principled model for the stellar spectra, e.g., using *PHOENIX* ([Husser et al., 2013](#)) synthetic spectra as the mean model, would allow the application to M-type stars. It could also help with the modelling of the stellar spectra, as the GP would only need to capture the variations between the synthetic model and the real observations. Furthermore, the simultaneous extraction of RVs and modelling of telluric features is still an on-going effort that was not completed before the delivery of the manuscript. Its inclusion will not only allow for the usage of larger wavelength regions but, also, provide a straightforward way of characterizing the uncertainties associated with this process (which the current telluric-modelling tools are lacking).

Chapter 5

Conclusions and future work

The 1995 discovery of *51 Pegasi B* – the first exoplanet orbiting a solar-type star – launched a revolution in the field of astrophysics. The realization that there could be worlds similar to ours, orbiting a star like our own, triggered a multi-decade quest for the detection and characterization of such elusive planets. This far-reaching goal not only motivated the development of new instruments, but also pushed forward the development of novel data analysis techniques. It is in this context that we can place the work developed during this thesis. More specifically, in the development of new tools to extract precise Radial Velocities from state-of-the-art stellar spectra.

The cornerstone of this thesis lies in **s-BART**, a semi-Bayesian reformulation of the classical template matching algorithm. It imposes an assumption of a common RV shifts across all wavelengths, i.e., for all spectral orders. This imposition, compatible with the effects induced by orbital companions, thus leads to a model that is more consistent with the systems that we are attempting to study. Since the development of this methodology it has been used in 2 peer-reviewed papers, with a few more in the preparation stages. This pipeline has been used to analyse data from multiple state-of-the-art instruments, such as ESPRESSO, HARPS-N, HARPS and CARMENES, with it being selected as the preferred RV-extraction tool in several of their different programs.

The impressive results presented by **s-BART** are however limited by a systematic effect that appears under specific conditions – in cases where all available observations were taken during the same night – and not present in the CCF RVs. In fact, this effect is detected in the data of multiple stars, observed with multiple instruments. More worrisome is the amplitude of such effect, capable of reaching a level of tens of meters per second, over only a few hours of expositions. Furthermore, it is also detected in different pipelines that are

based on the comparison of an high-SNR template and the stellar spectra. A wide array of tests has been applied, with little success in unveiling the root cause. Our analysis appears to suggest that the issue is being introduced by the stellar template, albeit from an unknown source.

The last main output of this thesis comes in the proof-of-concept of a Gaussian Process based model for RV extraction that should cover some of the flaws and assumptions brought by the **s-BART** framework. This model determines the RV separation whilst simultaneously modelling the stellar spectra in a probabilistic framework. For testing purposes, a simplified version of the proposed model was implemented, yielding promising results both on simulated and real data. We find that, when using a single spectral order, that this model leads to RVs consistent with those from the other approaches, despite the large number of simplifications that were imposed. More importantly, is the fact that such a framework will open the door for multiple options to tackle the effect of stellar activity on the RV extraction process.

Future work

Despite the amount of work that has been described in this manuscript, we find that much more remains to be done, leaving multiple avenues for future exploitation.

The implementation of the full GP model, idealized in Chapter 4.3.5, remains as an important step to be taken. It will most likely require the re-implementation of some parts of the model, namely the evaluation of the log-likelihood and its parallelization. It is highly likely that this effort will be made through a GPU-based algorithm, as they provide significant improvements, when compared to the CPU-bound methods. From there, it follows that the adoption of a time-varying covariance structure will provide a way to directly model the contribution of stellar activity, despite i) the expected larger computational cost of doing so; ii) the need to properly model such contributions. The later issue will be mitigated with the advent of new solar telescopes, allowing for a deeper understanding of the effects of stellar activity on the stellar spectra. For this, the data from the *PoET* solar telescope (Leite et al., 2022) – which will be connected to the ESPRESSO spectrograph – will allow the use of better informed stellar activity models..

Unveiling the root cause of the systematic RV effect that results from analysing data gathered within just one night, is also an important future goal. In particular, it is not clear to what extent this – often large – effect, only present in template-based methodologies,

could be present also in results derived from data obtained over several nights. Fixing this problem could thus lead to an improvement of the current s-BART methodology across all application cases. The extensive tests performed suggest the problem arises from the construction of the stellar template and/or the calibration files that are used during the extraction of the stellar spectra from the raw images of the instrument.”

Lastly, directly extracting the RVs from the stellar spectra might not be the optimal approach. The move towards directly modelling the number of exoplanets, and associated parameters will allow to skip the summarization of information (spectra into RVs) before the analysis. With this, we can remove one set of assumptions in our analysis process, e.g., in s-BART case we would no longer be assuming the achromatic RV, as we could directly model the companions and activity signal. Exploring this avenue might lead to improvements in our analysis stage, albeit it is yet to be tested.

Appendix A

Three-dimensional orbital parameters

In this Appendix we briefly describe the different parameters that describe the motion of an exoplanet, as it orbits a star. We refer back to [Murray & Correia \(2010\)](#) for a full derivation of the equations that govern the two-body problem, schematized in Figure [A.1](#), and that leads to an elliptical orbit.

This orbit, alongside the planet's position in it, can be traced with 6 parameters:

1. semi-major axis, a , of the ellipse.
2. eccentricity, e , that quantifies the deviation of an object's orbit from a perfectly circular path. A value of 0 indicates a circular orbit, while values within the range of $]0,1[$ signify an elliptical orbit. A value of 1 represents a parabolic escape orbit, and values exceeding 1 imply a hyperbolic trajectory.
3. Inclination of the orbit, i , that measures the angle between the normal orbital plane and the line of sight. It is only defined in the range $[0, 180]$ degrees;
4. Longitude of the ascending node, Ω , angle between the radius vector at the ascending node and the periastron of the orbit. This parameter, together with the inclination, define the orientation of the orbital plane.
5. Argument of the perihelion, ω , angle between the perihelion (the closest point to the star) and the ascending node, measured in the direction of motion;
6. Time of the periastron passage, t_0 , denoting the time at which the object is closest to its star.

It is also important to also mention the true anomaly, f , which defines the orbital position of the planet at any given time. It measures the angular distance between perihelion and the current position of the planet.

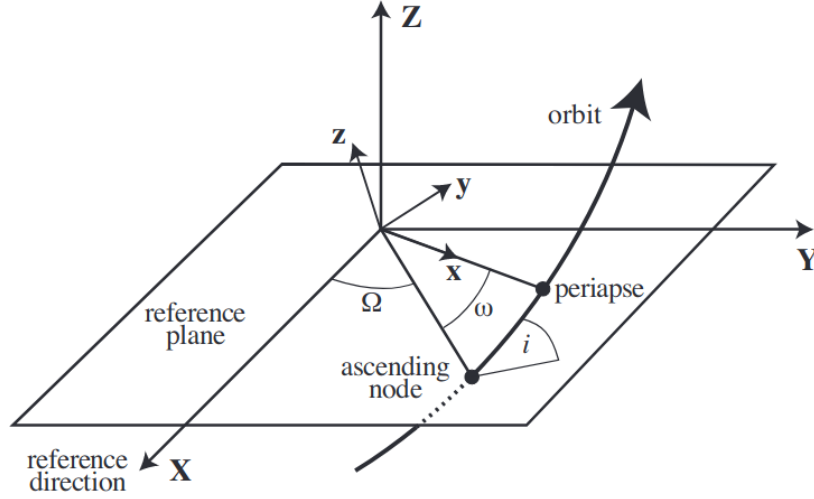


FIGURE A.1: Representation of the orbit of an exoplanet, with an observer above the Z axis. Image adapted from [Murray & Correia \(2010\)](#).

Appendix B

Nightly Zero Point (NZP)

The Nightly Zero Point (NZP) serves as a proxy for the stability of a spectrograph, by providing an estimation of the zero-RV point for any given night. Thus, by tracking its variation through multiple nights, it is possible to estimate the effect of the instrumental response on the RV measurements. However, if the instrument is stable and the known drift factors (e.g., instrumental drift, BERV) are properly corrected, the zero-point variability should be very small.

In the context of this PhD thesis, a need arose for the calculation of this metric, as mentioned in Chapter 2. For this purpose, the methodology proposed by [Tal-Or et al. \(2019\)](#) was closely followed. It starts by subtracting, from the time series of each target, its own error-weighted average, thus centring all time-series around an RV of 0 m s^{-1} . If a target has multiple observations in the same night, they were replaced by its median value. The NZPs were computed for all nights in which at least 3 targets were observed, through a weighted average of the RVs, using weights equal to the inverse of the RV variances. The uncertainty in the NZP measurement is taken to be the maximum value between the propagated (through the weighted mean) RV uncertainty and the RV scatter of the night in question.

Over the years, multiple works have been making use of this technique, ([Courcol et al., 2015](#), [Trifonov et al., 2020](#), e.g.), with some moderate success. More notably, it is also widely used in the CARMENES spectrograph, where the NZP can take values at the order of tens of meters per second. For such an application, there is a relatively large expenditure of telescope time to observe standard stars, those that we know to be less active and have no planetary companions. The RV time-series of such stars are used to compute the NZP, which are then used to correct the RVs from scientifically interesting targets.

Despite the widespread usage, this procedure can introduce extra variability in the data, as the NZP is calculated without removing the contribution from both stellar activity and planetary companions. The core assumption is that by having multiple measurements within the same night such signals, which are already to be small due to the selection of standard stars, will be averaged out. However, different nights will have a different number of points, from different targets, which can introduce some time-evolving variability in the final NZP timeseries. The impact from this is yet to be understood and, consequently, the NZP value is taken to be an estimation of the upper precision level of an instrument, rather than an actual measurement of its stability.

Appendix C

Generation of simulated spectra

In order to determine whether our template matching estimates for the radial velocities uncertainties are consistent with the information present in the spectra, we often required simulated spectra using real observations as a basis. As this task was needed for benchmarking both `s-BART` and the fully Bayesian RV model, a small *Python* library was written^{*}, easing the process and ensuring that the final result from this simulation is compatible with the input format of our methodologies. The generation of the new spectrum is done accordingly to the following recipe:

1. We, the user, start by selecting one reference observation from the available data, following a given criterion;
2. A Gaussian distribution is placed in each pixel of the reference spectrum, as shown in Figure C.1. Its mean and standard deviation are equal to the flux value and uncertainty of the reference spectrum, respectively;
3. A random sample is drawn from each distribution, leading to the generation of a new, simulated, spectrum. By default, the uncertainties associated with the simulated spectrum are the same as the ones from the reference observation.
4. If requested, a RV signal is injected in the timeseries, either assuming a Keplerian source, or random draws of a user-specified distribution. By default, and unless specifically mentioned, this step is not carried out and, consequently, each simulated observation is expected to have a RV shift of zero m s^{-1} , relative to the reference.

^{*}Currently not publicly available

5. Each spectrum is then stored to disk, following the format of the original data, so that our RV extraction methodologies can easily process them.

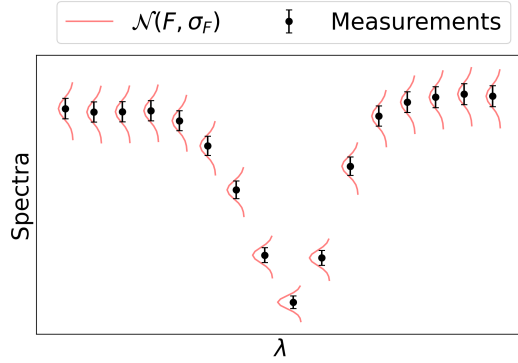


FIGURE C.1: Gaussian distributions (red line) placed on each pixel of a reference spectrum (black point), from which we will draw random samples from, leading to the creation of a simulated observation.

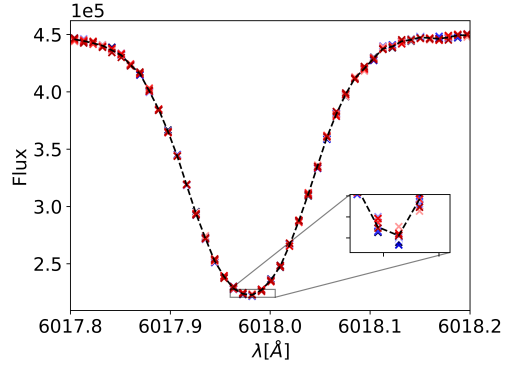


FIGURE C.2: Spectral line taken from the simulation of $N = 10$ spectra using random draws from Gaussian distributions. The dashed black line represents the original spectrum used to construct the line, with the inset region zooming in on the core of one spectral line.

Figure C.2 shows the output of 10 simulations following the methodology described in this Appendix, focusing on one spectral line, without any introducing any RV signal.

Appendix D

Error propagation through a cubic-spline interpolation

The cubic-spline interpolation used to calculate a value, y , at position x in the interval $[x_i, x_{i+1}]$ can be written (see p. 113 of [Press, 1992](#)) in the following form:

$$y = Ay_i + By_{i+1} + Cy_i'' + Dy_{i+1}'', \quad (\text{D.1})$$

where the double apostrophe represents the second derivative with respect to x and

$$\begin{aligned} A &= \frac{x_{i+1} - x}{x_{i+1} - x_i} \\ B &= \frac{x - x_i}{x_{i+1} - x_i} \\ C &= \frac{1}{6}(A^3 - A)(x_{i+1} - x_i)^2 \\ D &= \frac{1}{6}(B^3 - B)(x_{i+1} - x_i)^2 \end{aligned} \quad (\text{D.2})$$

The computation of the second derivatives requires choosing proper boundary conditions. As discussed throughout this thesis, we have chosen to remove the wavelength regions that are not common to all observations. Thus, as we do not interpolate in the edges, we can use natural boundary conditions, where the second derivative takes a value of zero at the edges of the input data. Following the notation of [Press \(1992\)](#) we can write down a general expression for the second derivatives:

$$\begin{pmatrix} \frac{y_3 - y_2}{x_3 - x_2} - \frac{y_2 - y_1}{x_2 - x_1} \\ \vdots \\ \frac{y_{i+1} - y_i}{x_{i+1} - x_i} - \frac{y_i - y_{i-1}}{x_i - x_{i-1}} \\ \vdots \\ \frac{y_N - y_{N-1}}{x_N - x_{N-1}} - \frac{y_{N-1} - y_{N-2}}{x_{N-1} - x_{N-2}} \end{pmatrix} = h_{i,j} \begin{pmatrix} y_2'' \\ \vdots \\ y_i'' \\ \vdots \\ y_{N-1}'' \end{pmatrix}, \quad (D.3)$$

$$y_1'' = y_N'' = 0$$

where h is a symmetric tridiagonal matrix given by:

$$h_{i,j} = \begin{pmatrix} \frac{x_3 - x_1}{3} & \frac{x_3 - x_2}{6} & 0 & 0 \\ \frac{x_i - x_{i-1}}{6} & \frac{x_{i+1} - x_{i-1}}{3} & \frac{x_{i+1} - x_i}{6} & 0 \\ 0 & \ddots & \ddots & \ddots \\ 0 & 0 & \frac{x_{N-1} - x_{N-2}}{6} & \frac{x_N - x_{i-2}}{3} \end{pmatrix}. \quad (D.4)$$

An expression for the propagation of uncertainties in cubic splines was derived in [Gardner \(2003\)](#), with the covariance between any two given interpolated points, $u(y_m, y_n)$, given by:

$$u(y_m, y_n) = \begin{pmatrix} A_i & B_i & C_i & D_i \end{pmatrix} \begin{pmatrix} u(y_i, y_j) & u(y_i, y_{j+1}) & u(y_i, y_j'') & u(y_i, y_{j+1}'') \\ u(y_{i+1}, y_j) & u(y_{i+1}, y_{j+1}) & u(y_{i+1}, y_j'') & u(y_{i+1}, y_{j+1}'') \\ u(y_i'', y_j) & u(y_i'', y_{j+1}) & u(y_i'', y_j'') & u(y_i'', y_{j+1}'') \\ u(y_{i+1}'', y_j) & u(y_{i+1}'', y_{j+1}) & u(y_{i+1}'', y_j'') & u(y_{i+1}'', y_{j+1}'') \end{pmatrix} \begin{pmatrix} A_j \\ B_j \\ C_j \\ D_j \end{pmatrix} \quad (D.5)$$

with

$$\begin{aligned} u(y_m'', y_n'') &= Q_m^T U_y Q_n \\ u(y_m'', y_n) &= Q_m^T U_y g_n' \end{aligned} \quad (D.6)$$

and where U_y is a $N \times N$ covariance matrix of the input data, g_n is a column vector of length N with a 1 in the n -th column and 0 elsewhere, Q_i is a column-vector of sensitivity coefficients of the second-derivatives in relation to the input values:

$$Q_m = \begin{cases} \frac{\partial y_m''}{\partial y_k} & k \in \{2, \dots, N-1\} \\ 0 & \text{otherwise} \end{cases}. \quad (\text{D.7})$$

In order to calculate the partial derivative (of the second derivatives) we start by selecting a given row from Equation D.3 and re-arranging the summation limits:

$$y_i'' = - \sum_{j=2}^{N-1} h_{i,j}^{-1} \left(\frac{1}{x_{j+1} - x_j} + \frac{1}{x_j - x_{j-1}} \right) y_j + \sum_{j=3}^N h_{i,j-1}^{-1} \frac{y_j}{x_j - x_{j-1}} + \sum_{j=1}^{N-2} h_{i,j+1}^{-1} \frac{y_j}{x_{j+1} - x_j}. \quad (\text{D.8})$$

The partial derivative then follows from Equation D.8:

$$\frac{\partial y_i''}{\partial y_j} = -h_{i,j}^{-1} \left(\frac{1}{x_{j+1} - x_j} + \frac{1}{x_j - x_{j-1}} \right) + \frac{h_{i,j-1}^{-1}}{x_j - x_{j-1}} + \frac{h_{i,j+1}^{-1}}{x_{j+1} - x_j}. \quad (\text{D.9})$$

Special care must be taken due to the summation limits in Equation D.8, as not all terms in Equation D.9 exist for all indexes. As we are only interested in the variances of the interpolated values, i.e., we will not consider the effect of the covariances, we only have to evaluate Equation D.5 for the $m = n$ case.

Lastly, as we are mainly dealing with ESPRESSO data, the computation of the second derivatives through Eq. D.3 implies the inversion of a matrix with size $N - 2 \times N - 2$, with $N = 9111$, which represents a large computational burden. However, as we are dealing with a symmetric tridiagonal matrix of the form:

$$h = \begin{pmatrix} x_1 & y_1 & & & \\ z_1 & x_2 & \ddots & & \\ & \ddots & \ddots & y_{N-1} & \\ & & z_{N-1} & y_N & \end{pmatrix}, \quad (\text{D.10})$$

where $z_n = y_n$, we can invert it with an explicit formula, using backwards continued fractions (Kılıç, 2008):

$$h_{ij}^{-1} = \begin{cases} \frac{1}{C_i^b} + \sum_{k=i+1}^N \left(\frac{1}{C_k^b} \prod_{t=i}^{k-1} \frac{y_t^2}{(C_t^b)^2} \right) & \text{if } i = j \\ (-1)^{i+j} \prod_{t=j}^{i-1} \frac{y_t}{C_t^b} h_{ii}^{-1} & \text{otherwise} \end{cases}, \quad (\text{D.11})$$

with C_n^b given by Eq 4 of the aforementioned paper:

$$C_n^b = \begin{cases} x_1 & \text{if } n = 1 \\ x - n + \frac{-y_{n-1}z_{n-1}}{C_{n-1}^b} & \text{otherwise} \end{cases}. \quad (\text{D.12})$$

Bibliografia

- Affer, L., Damasso, M., Micela, G., et al. 2019, HADES RV Program with HARPS-N at the TNG: IX. A Super-Earth around the M Dwarf Gl 686, *Astronomy & Astrophysics*, 622, A193 [Cited on page [13](#).]
- Aigrain, S., Pont, F., & Zucker, S. 2012, A Simple Method to Estimate Radial Velocity Variations Due to Stellar Activity Using Photometry, *Monthly Notices of the Royal Astronomical Society*, 419, 3147 [Cited on page [14](#).]
- Allart, R., Lovis, C., Faria, J., et al. 2022, Automatic Model-Based Telluric Correction for the ESPRESSO Data Reduction Software: Model Description and Application to Radial Velocity Computation, *Astronomy & Astrophysics*, 666, A196 [Cited on pages [11](#), [30](#), [76](#), and [111](#).]
- Anglada-Escudé, G. & Butler, R. P. 2012, THE HARPS-TERRA PROJECT. I. DESCRIPTION OF THE ALGORITHMS, PERFORMANCE, AND NEW MEASUREMENTS ON A FEW REMARKABLE STARS OBSERVED BY HARPS, *The Astrophysical Journal Supplement Series*, 200, 15 [Cited on pages [20](#), [21](#), [50](#), [51](#), and [52](#).]
- Anglada-Escudé, G., Tuomi, M., Arriagada, P., et al. 2016, NO EVIDENCE FOR ACTIVITY CORRELATIONS IN THE RADIAL VELOCITIES OF KAPTEYN'S STAR, *The Astrophysical Journal*, 830, 74 [Cited on page [20](#).]
- Artigau, É., Astudillo-Defru, N., Delfosse, X., et al. 2014, in *SPIE Astronomical Telescopes + Instrumentation*, ed. A. B. Peck, C. R. Benn, & R. L. Seaman, Montréal, Quebec, Canada, 914905 [Cited on page [11](#).]
- Artigau, É., Cadieux, C., Cook, N. J., et al. 2022, Line-by-Line Velocity Measurements: An Outlier-resistant Method for Precision Velocimetry, *The Astronomical Journal*, 164, 84 [Cited on page [23](#).]

- Astudillo-Defru, N. 2015, Search for Earth-like Planets in the Habitable Zone of M-dwarfs, PhD thesis, Université Grenoble, Alpes [Cited on page 21.]
- Balsalobre-Ruza, O., Lillo-Box, J., Berihuete, A., et al. 2022, `\texttt{KOBESim}`: A Bayesian Observing Strategy Algorithm for Planet Detection in Radial Velocity Blind-Search Surveys [Cited on page 15.]
- Baranne, A., Queloz, D., Mayor, M., et al. 1996, ELODIE: A Spectrograph for Accurate Radial Velocity Measurements, *Astronomy and Astrophysics Supplement Series*, 119, 373 [Cited on pages 19 and 20.]
- Barragán, O., Gillen, E., Aigrain, S., et al. 2023, Revisiting K2-233 Spectroscopic Time-Series with Multidimensional Gaussian Processes [Cited on page 13.]
- Barros, S. C. C., Demangeon, O. D. S., Alibert, Y., et al. 2022, HD 23472: A Multi-Planetary System with Three Super-Earths and Two Potential Super-Mercuries,, *Astronomy & Astrophysics*, 665, A154 [Cited on pages 2, 8, and 20.]
- Bedell, M., Hogg, D. W., Foreman-Mackey, D., Montet, B. T., & Luger, R. 2019, WOBBLE : A Data-driven Analysis Technique for Time-series Stellar Spectra, *The Astronomical Journal*, 158, 164 [Cited on pages 11, 23, and 65.]
- Benz, W., Broeg, C., Fortier, A., et al. 2021, The CHEOPS Mission, *Experimental Astronomy*, 51, 109 [Cited on page 3.]
- Bertaux, J. L., Lallement, R., Ferron, S., Boonne, C., & Bodichon, R. 2014, TAPAS, a Web-Based Service of Atmospheric Transmission Computation for Astronomy, *Astronomy & Astrophysics*, 564, A46 [Cited on pages 11 and 30.]
- Binnenfeld, A., Shahaf, S., Anderson, R. I., & Zucker, S. 2022, New Periodograms Separating Orbital Radial Velocities and Spectral Shape Variation, *Astronomy & Astrophysics*, 659, A189 [Cited on page 15.]
- Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. 2017, Variational Inference: A Review for Statisticians, *Journal of the American Statistical Association*, 112, 859 [Cited on pages 48 and 67.]
- Boisse, I., Moutou, C., Vidal-Madjar, A., et al. 2009, Stellar Activity of Planetary Host Star HD 189733, *Astronomy & Astrophysics*, 495, 959 [Cited on page 15.]

- Bouchy, F., Pepe, F., & Queloz, D. 2001, Fundamental Photon Noise Limit to Radial Velocity Measurements, *Astronomy & Astrophysics*, 374, 733 [Cited on pages 19, 23, and 54.]
- Butler, R. P., Marcy, G. W., Fischer, D. A., et al. 1999, Evidence for Multiple Companions to ν Andromedae, *The Astrophysical Journal*, 526, 916 [Cited on page 1.]
- Campbell, B., Walker, G. A. H., & Yang, S. 1988, A Search for Substellar Companions to Solar-Type Stars, *The Astrophysical Journal*, 331, 902 [Cited on page 1.]
- Castro-González, A., Demangeon, O. D. S., Lillo-Box, J., et al. 2023, An Unusually Low-Density Super-Earth Transiting the Bright Early-Type M-dwarf GJ 1018 (TOI-244), *Astronomy & Astrophysics*, 675, A52 [Cited on page 20.]
- Chalupka, K., Williams, C. K. I., & Murray, I. 2012, A Framework for Evaluating Approximation Methods for Gaussian Process Regression, arXiv:1205.6326 [cs, stat] [[arXiv]1205.6326] [Cited on page 109.]
- Chaplin, W. J., Cegla, H. M., Watson, C. A., Davies, G. R., & Ball, W. H. 2019, Filtering Solar-Like Oscillations for Exoplanet Detection in Radial Velocity Observations, *The Astronomical Journal*, 157, 163 [Cited on page 14.]
- Collier Cameron, A., Ford, E. B., Shahaf, S., et al. 2021, Separating Planetary Reflex Doppler Shifts from Stellar Variability in the Wavelength Domain, *Monthly Notices of the Royal Astronomical Society*, 505, 1699 [Cited on page 15.]
- Collier Cameron, A., Mortier, A., Phillips, D., et al. 2019, Three Years of Sun-as-a-star Radial-Velocity Observations on the Approach to Solar Minimum, *Monthly Notices of the Royal Astronomical Society*, 487, 1082 [Cited on page 13.]
- Cosentino, R., Lovis, C., Pepe, F., et al. 2012, in *SPIE Astronomical Telescopes + Instrumentation*, ed. I. S. McLean, S. K. Ramsay, & H. Takami, Amsterdam, Netherlands, 84461V [Cited on pages 3, 16, and 17.]
- Courcol, B., Bouchy, F., Pepe, F., et al. 2015, The SOPHIE Search for Northern Extrasolar Planets VIII. A Warm Neptune Orbiting HD164595, *Astronomy & Astrophysics*, 581, A38 [Cited on pages 57 and 135.]

- Cretignier, M., Dumusque, X., Allart, R., Pepe, F., & Lovis, C. 2020a, Measuring Precise Radial Velocities on Individual Spectral Lines: II. Dependence of Stellar Activity Signal on Line Depth, *Astronomy & Astrophysics*, 633, A76 [Cited on pages 15, 23, and 66.]
- Cretignier, M., Francfort, J., Dumusque, X., Allart, R., & Pepe, F. 2020b, RASSINE: Interactive Tool for Normalising Stellar Spectra I. Description and Performance of the Code, *Astronomy & Astrophysics*, 640, A42 [Cited on page 118.]
- Csató, L. & Oppenheimer, M. 2002, Sparse On-Line Gaussian Processes, *Neural Computation*, 14, 641 [Cited on page 109.]
- Cunha, D., Santos, N. C., Figueira, P., et al. 2014, Impact of Micro-Telluric Lines on Precise Radial Velocities and Its Correction, *Astronomy & Astrophysics*, 568, A35 [Cited on page 8.]
- Cunningham, J. P., Shenoy, K. V., & Sahani, M. 2008, in *Proceedings of the 25th International Conference on Machine Learning - ICML '08* (Helsinki, Finland: ACM Press), 192–199 [Cited on page 109.]
- Czekala, I., Mandel, K. S., Andrews, S. M., et al. 2017, Disentangling Time Series Spectra with Gaussian Processes: Applications to Radial Velocity Analysis, *The Astrophysical Journal*, 840, 49 [Cited on pages 23, 111, and 112.]
- da Silva, J. G., Bensabat, A., Monteiro, T., & Santos, N. C. 2022, Optimising the H α Index for the Identification of Activity Signals in FGK Stars. Improvement of the Correlation between H α and CaII H&K [Cited on page 15.]
- Dalba, P. A., Kane, S. R., Dragomir, D., et al. 2022, The TESS-Keck Survey. VIII. Confirmation of a Transiting Giant Planet on an Eccentric 261 Day Orbit with the Automated Planet Finder Telescope*, *The Astronomical Journal*, 163, 61 [Cited on page 5.]
- Demangeon, O. D. S., Osorio, M. R. Z., Alibert, Y., et al. 2021, Warm Terrestrial Planet with Half the Mass of Venus Transiting a Nearby Star, *Astronomy & Astrophysics* [[arXiv]2108.03323] [Cited on page 8.]
- Dumusque, X. 2018, Measuring Precise Radial Velocities on Individual Spectral Lines: I. Validation of the Method and Application to Mitigate Stellar Activity $\star, \star\star$, *Astronomy & Astrophysics*, 620, A47 [Cited on pages 15, 23, and 66.]

- Dumusque, X., Cretignier, M., Sosnowska, D., et al. 2021, Three Years of HARPS-N High-Resolution Spectroscopy and Precise Radial Velocity Data for the Sun, *Astronomy & Astrophysics*, 648, A103 [Cited on page [13](#).]
- Dumusque, Xavier. 2012, Mitigating Stellar Signals in the Quest for Other Earths, PhD thesis [Cited on page [14](#).]
- ESO. 2023, ESPRESSO Pipeline User Manual [Cited on page [70](#).]
- Faria, J. P., Adibekyan, V., Amazo-Gómez, E. M., et al. 2020, Decoding the Radial Velocity Variations of HD 41248 with ESPRESSO, *Astronomy & Astrophysics*, 635, A13 [Cited on page [13](#).]
- Faria, J. P., Mascareño, A. S., Figueira, P., et al. 2022, A Candidate Short-Period Sub-Earth Orbiting Proxima Centauri, *Astronomy & Astrophysics*, 658, A115 [Cited on pages [xvii](#), [3](#), [8](#), [49](#), [50](#), and [63](#).]
- Faria, João Pedro de Sousa. 2018, Exoplanet Detection in Metal-Poor Stars, PhD thesis, Faculdade de Ciências da Universidade do Porto [Cited on page [7](#).]
- Figueira, P., Kerber, F., Chacon, A., et al. 2012, Comparing Radial Velocities of Atmospheric Lines with Radiosonde Measurements, *Monthly Notices of the Royal Astronomical Society*, 420, 2874 [Cited on page [8](#).]
- Flores, M., González, J. F., Arancibia, M. J., Buccino, A., & Saffe, C. 2016, Discovery of an Activity Cycle in the Solar-Analog HD 45184. Exploring Balmer and Metallic Lines as Activity Proxy Candidates, *Astronomy & Astrophysics*, 589, A135 [Cited on page [29](#).]
- Foreman-Mackey, D. 2018, Scalable Backpropagation for Gaussian Processes Using Celestite, *Research Notes of the AAS*, 2, 31 [Cited on pages [109](#) and [115](#).]
- Foreman-Mackey, D. 2023, Dfm/Tinygp: The Tiniest of Gaussian Process Libraries, *Zenodo* [Cited on pages [110](#) and [115](#).]
- Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, Fast and Scalable Gaussian Process Modeling with Applications to Astronomical Time Series, *The Astronomical Journal*, 154, 220 [Cited on pages [109](#) and [115](#).]
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, Emcee: The MCMC Hammer, *Publications of the Astronomical Society of the Pacific*, 125, 306 [Cited on page [47](#).]

- Forveille, T., Bonfils, X., Delfosse, X., et al. 2009, The HARPS Search for Southern Extra-Solar Planets: XIV. Gl 176b, a Super-Earth Rather than a Neptune, and at a Different Period, *Astronomy & Astrophysics*, 493, 645 [Cited on page [13](#).]
- Gardner, J. 2003, Uncertainties in Interpolated Spectral Data, *Journal of Research of the National Institute of Standards and Technology*, 108, 69 [Cited on page [140](#).]
- Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, The James Webb Space Telescope, *Space Science Reviews*, 123, 485 [Cited on page [3](#).]
- Gardner, J. R., Pleiss, G., Wu, R., Weinberger, K. Q., & Wilson, A. G. 2018, Product Kernel Interpolation for Scalable Gaussian Processes, arXiv:1802.08903 [cs, stat] [[arXiv]1802.08903] [Cited on page [109](#).]
- Gilboa, E., Saatçi, Y., & Cunningham, J. P. 2013, Scaling Multidimensional Inference for Structured Gaussian Processes, *IEEE transactions on pattern analysis and machine intelligence*, 37, 424 [Cited on page [109](#).]
- Gneiting, T. 2002, Compactly Supported Correlation Functions, *Journal of Multivariate Analysis*, 83, 493 [Cited on page [109](#).]
- Gomes Da Silva, J., Santos, N. C., Bonfils, X., et al. 2012, Long-Term Magnetic Activity of a Sample of M-dwarf Stars from the HARPS Program: II. Activity and Radial Velocity ★★ ★★ ★★, *Astronomy & Astrophysics*, 541, A9 [Cited on page [14](#).]
- Gramacy, R. B. & Apley, D. W. 2014, Local Gaussian Process Approximation for Large Computer Experiments, arXiv:1303.0383 [stat] [[arXiv]1303.0383] [Cited on page [109](#).]
- Gramacy, R. B. & Haaland, B. 2016, Speeding Up Neighborhood Search in Local Gaussian Process Prediction, *Technometrics*, 58, 294 [Cited on page [109](#).]
- Gray, A. G. 2004 [Cited on page [109](#).]
- Gregory, P. 2005, *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica® Support*, 1st edn. (Cambridge University Press) [Cited on pages [48](#), [106](#), and [128](#).]
- Gullikson, K., Dodson-Robinson, S., & Kraus, A. 2014, Correcting for Telluric Absorption: Methods, Case Studies, and Release of the TelFit Code, *The Astronomical Journal*, 148, 53 [Cited on pages [11](#) and [30](#).]

- Gunapati, G., Jain, A., Srijith, P. K., & Desai, S. 2022, Variational Inference as an Alternative to MCMC for Parameter Estimation and Model Selection, Publications of the Astronomical Society of Australia, 39, e001 [Cited on page 48.]
- Gupta, A. F. & Bedell, M. 2023, Fishing for Planets: A Comparative Analysis of EPRV Survey Performance in the Presence of Correlated Noise [Cited on page 15.]
- Henry, G. W., Marcy, G. W., Butler, R. P., & Vogt, S. S. 2000, A Transiting “51 Peg-like” Planet, The Astrophysical Journal, 529, L41 [Cited on page 1.]
- Hensman, J., Fusi, N., & Lawrence, N. D. 2013, Gaussian Processes for Big Data, 9 [Cited on page 109.]
- Hojjatpanah, S., Figueira, P., Santos, N. C., et al. 2019, Catalog for the ESPRESSO Blind Radial Velocity Exoplanet Survey, Astronomy & Astrophysics, 629, A80 [Cited on page 57.]
- Holzer, P. H., Cisewski-Kehe, J., Fischer, D., & Zhao, L. 2021, A Hermite–Gaussian Based Exoplanet Radial Velocity Estimation Method, The Annals of Applied Statistics, 15 [Cited on page 22.]
- Howell, S. B., Sobeck, C., Haas, M., et al. 2014, The K2 Mission: Characterization and Early Results, Publications of the Astronomical Society of the Pacific, 126, 398 [Cited on pages 3 and 5.]
- Hsu, D. C., Ford, E. B., Ragozzine, D., & Ashby, K. 2019, Occurrence Rates of Planets Orbiting FGK Stars: Combining Kepler DR25, Gaia DR2 and Bayesian Inference, The Astronomical Journal, 158, 109 [Cited on page 3.]
- Huelamo, N., Figueira, P., Bonfils, X., et al. 2008, TW Hydrae: Evidence of Stellar Spots Instead of a Hot Jupiter, Astronomy & Astrophysics, 489, L9 [Cited on page 13.]
- Husser, T.-O., von Berg, S. W., Dreizler, S., et al. 2013, A New Extensive Library of PHOENIX Stellar Atmospheres and Synthetic Spectra, Astronomy & Astrophysics, 553, A6 [Cited on page 128.]
- Jain, A., Srijith, P. K., & Desai, S. 2018, Variational Inference as an Alternative to MCMC for Parameter Estimation and Model Selection, arXiv:1803.06473 [astro-ph, physics:physics] [[arXiv]1803.06473] [Cited on page 67.]

- Jurgenson, C., Fischer, D., McCracken, T., et al. 2016, in *SPIE Astronomical Telescopes + Instrumentation*, ed. C. J. Evans, L. Simard, & H. Takami, Edinburgh, United Kingdom, 99086T [Cited on page [3](#).]
- JWST Transiting Exoplanet Community Early Release Science Team, Ahrer, E.-M., Alderson, L., et al. 2023, Identification of Carbon Dioxide in an Exoplanet Atmosphere, *Nature*, 614, 649 [Cited on page [3](#).]
- Kılıç, E. 2008, Explicit Formula for the Inverse of a Tridiagonal Matrix by Backward Continued Fractions, *Applied Mathematics and Computation*, 197, 345 [Cited on page [141](#).]
- Koch, D. G., Borucki, W. J., Basri, G., et al. 2010, *KEPLER MISSION DESIGN, REALIZED PHOTOMETRIC PERFORMANCE, AND EARLY SCIENCE*, *The Astrophysical Journal*, 713, L79 [Cited on pages [3](#) and [5](#).]
- Kuerster, M., Endl, M., Rouesnel, F., et al. 2003, The Low-Level Radial Velocity Variability in Barnard’s Star (=GJ 699). Secular Acceleration, Indications for Convective Redshift, and Planet Mass Limits, *Astronomy & Astrophysics*, 403, 1077 [Cited on page [29](#).]
- Kullback, S. & Leibler, R. A. 1951, On Information and Sufficiency, *The Annals of Mathematical Statistics*, 22, 79 [Cited on page [67](#).]
- Lafarga, M., Ribas, I., Lovis, C., et al. 2020, The CARMENES Search for Exoplanets around M Dwarfs: Radial Velocities and Activity Indicators from Cross-Correlation Functions with Weighted Binary Masks, *Astronomy & Astrophysics*, 636, A36 [Cited on page [20](#).]
- Lafarga, M., Ribas, I., Reiners, A., et al. 2021, The CARMENES Search for Exoplanets around M Dwarfs: Mapping Stellar Activity Indicators across the M Dwarf Domain, *Astronomy & Astrophysics*, 652, A28 [Cited on page [14](#).]
- Lafarga, M., Ribas, I., Zechmeister, M., et al. 2023, The CARMENES Search for Exoplanets around M Dwarfs. Line-by-line Sensitivity to Activity in M Dwarfs [Cited on page [20](#).]
- Lam, K. W. F., Csizmadia, S., Astudillo-Defru, N., et al. 2021, GJ 367b: A Dense, Ultrashort-Period Sub-Earth Planet Transiting a Nearby Red Dwarf Star, *Science*, 374, 1271 [Cited on page [2](#).]

- Lanza, A. F., Cameron, A. C., & Haywood, R. D. 2019, Reducing Activity-Induced Variations in a Radial-Velocity Time Series of the Sun as a Star, *Monthly Notices of the Royal Astronomical Society*, 486, 3459 [Cited on page 20.]
- Lanza, A. F., Malavolta, L., Benatti, S., et al. 2018, The GAPS Programme with HARPS-N at TNG XVII. Line Profile Indicators and Kernel Regression as Diagnostics of Radial-Velocity Variations Due to Stellar Activity in Solar-like Stars, *Astronomy & Astrophysics*, 616, A155 [Cited on page 14.]
- Latham, D. W., Mazeh, T., Stefanik, R. P., Mayor, M., & Burki, G. 1989, The Unseen Companion of HD114762: A Probable Brown Dwarf, *Nature*, 339, 38 [Cited on page 1.]
- Leite, I., Cabral, A., Abreu, M., & Santos, N. 2022, Imaging Sensors for PoET: A Spatially Resolved Solar Spectroscopy Instrument, *Journal of Physics: Conference Series*, 2407, 012020 [Cited on page 130.]
- Lillo-Box, J., Santos, N. C., Santerne, A., et al. 2022, The KOBE Experiment: K-dwarfs Orbited By Habitable Exoplanets: Project Goals, Target Selection, and Stellar Characterization, *Astronomy & Astrophysics*, 667, A102 [Cited on pages 17 and 60.]
- Liu, H., Ong, Y.-S., Shen, X., & Cai, J. 2020, When Gaussian Process Meets Big Data: A Review of Scalable GPs, *IEEE Transactions on Neural Networks and Learning Systems*, 31, 4405 [Cited on page 109.]
- Marconi, A., Abreu, M., Adibekyan, V., et al. 2022, in *Ground-Based and Airborne Instrumentation for Astronomy IX*, ed. C. J. Evans, J. J. Bryant, & K. Motohara (Montréal, Canada: SPIE), 75 [Cited on page 4.]
- Mascareño, A. S., González-Álvarez, E., Osorio, M. R. Z., et al. 2023, Two Temperate Earth-mass Planets Orbiting the Nearby Star GJ1002, *Astronomy & Astrophysics*, 670, A5 [Cited on page 63.]
- Mayor, M., Lovis, C., & Santos, N. C. 2014, Doppler Spectroscopy as a Path to the Detection of Earth-like Planets, *Nature*, 513, 328 [Cited on page 3.]
- Mayor, M., Pepe, F., Queloz, D., et al. 2003, Setting New Standards with HARPS, *The Messenger*, 114, 20 [Cited on pages 16 and 17.]
- Mayor, M. & Queloz, D. 1995, A Jupiter-mass Companion to a Solar-Type Star, *Nature*, 378, 355 [Cited on page 1.]

- Melkumyan, A. & Ramos, F. 2009, 1936–1942 [Cited on page [109](#).]
- Meunier, N., Desort, M., & Lagrange, A.-M. 2010, Using the Sun to Estimate Earth-like Planets Detection Capabilities: II. Impact of Plages, *Astronomy and Astrophysics*, 512, A39 [Cited on page [14](#).]
- Morello, G., Parviainen, H., Murgas, F., et al. 2023, TOI-1442 b and TOI-2445 b: Two Potentially Rocky Ultra-Short Period Planets around M Dwarfs, *Astronomy & Astrophysics*, 673, A32 [Cited on page [8](#).]
- Motalebi, F., Udry, S., Gillon, M., et al. 2015, The HARPS-N Rocky Planet Search I. HD219134b: A Transiting Rocky Planet in a Multi-Planet System at 6.5 Pc from the Sun, *Astronomy & Astrophysics*, 584, A72 [Cited on page [8](#).]
- Moulla, K. A., Dumusque, X., & Cretignier, M. 2023, Measuring Precise Radial Velocities on Individual Spectral Lines. IV. Stellar Activity Correlation with Line Formation Temperature [Cited on pages [15](#) and [23](#).]
- Murray, C. D. & Correia, A. C. M. 2010, Keplerian Orbits and Dynamics of Exoplanets [Cited on pages [xxii](#), [7](#), [133](#), and [134](#).]
- Neal, J. & Figueira, P. 2019, Eniric: Extended NIR Information Content, *Journal of Open Source Software*, 4, 1053 [Cited on page [54](#).]
- Parviainen, H. 2018, in *Handbook of Exoplanets*, ed. H. J. Deeg & J. A. Belmonte (Cham: Springer International Publishing), 1567–1590 [Cited on page [106](#).]
- Pepe, F., Cristiani, S., Rebolo, R., et al. 2021, ESPRESSO@VLT – On-sky Performance and First Results, *Astronomy & Astrophysics*, 645, A96 [Cited on pages [xviii](#), [3](#), [16](#), [18](#), [35](#), and [69](#).]
- Pepe, F., Mayor, M., Galland, F., et al. 2002a, The CORALIE Survey for Southern Extra-Solar Planets VII: Two Short-Period Saturnian Companions to HD 108147 and HD 168746, *Astronomy & Astrophysics*, 388, 632 [Cited on page [20](#).]
- Pepe, F., Mayor, M., Rupprecht, G., et al. 2002b, HARPS: ESO’s Coming Planet Searcher. Chasing Exoplanets with the La Silla 3.6-m Telescope, *The Messenger*, 110, 9 [Cited on pages [16](#) and [17](#).]

- Pepe, F. A., Cristiani, S., Rebolo Lopez, R., et al. 2010, in *SPIE Astronomical Telescopes + Instrumentation*, ed. I. S. McLean, S. K. Ramsay, & H. Takami, San Diego, California, USA, 77350F [Cited on page [18](#).]
- Press, W. H., ed. 1992, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd edn. (Cambridge ; New York: Cambridge University Press) [Cited on pages [41](#) and [139](#).]
- Qian, S. S., Stow, C. A., & Borsuk, M. E. 2003, On Monte Carlo Methods for Bayesian Inference, *Ecological Modelling*, 159, 269 [Cited on page [106](#).]
- Queloz, D., Bouchy, F., Moutou, C., et al. 2009, The CoRoT-7 Planetary System: Two Orbiting Super-Earths, *Astronomy & Astrophysics*, 506, 303 [Cited on page [5](#).]
- Queloz, D., Henry, G. W., Sivan, J. P., et al. 2001, No Planet for HD 166435, *Astronomy and Astrophysics*, 379, 279 [Cited on page [20](#).]
- Quinonero-Candela, J. & Rasmussen, C. E. 2005, A Unifying View of Sparse Approximate Gaussian Process Regression, *The Journal of Machine Learning Research*, 6, 1939 [Cited on page [109](#).]
- Quirrenbach, A., Amado, P. J., Ribas, I., et al. 2018, in *Ground-Based and Airborne Instrumentation for Astronomy VII*, ed. H. Takami, C. J. Evans, & L. Simard (Austin, United States: SPIE), 32 [Cited on pages [16](#) and [17](#).]
- Rainer, M., Borsa, F., & Affer, L. 2020a, Stellar Masks and Bisector’s Shape for M-type Stars Observed in the GAPS Programme with HARPS-N at TNG, *Experimental Astronomy*, 49, 73 [Cited on page [20](#).]
- Rainer, M., Borsa, F., & Affer, L. 2020b, Stellar Masks and Bisector’s Shape for M-type Stars Observed in the GAPS Programme with HARPS-N at TNG, *Experimental Astronomy*, 49, 73 [Cited on page [20](#).]
- Rajpaul, V., Aigrain, S., Osborne, M. A., Reece, S., & Roberts, S. 2015, A Gaussian Process Framework for Modelling Stellar Activity Signals in Radial Velocity Data, *Monthly Notices of the Royal Astronomical Society*, 452, 2269 [Cited on page [14](#).]
- Rajpaul, V. M., Aigrain, S., & Buchhave, L. A. 2020, A Robust, Template-Free Approach to Precise Radial Velocity Extraction, *Monthly Notices of the Royal Astronomical Society*, 492, 3960 [Cited on pages [24](#), [65](#), and [111](#).]

- Rando, N., Asquier, J., Corral Van Damme, C., et al. 2018, in *Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave*, ed. H. A. MacEwen, M. Lystrup, G. G. Fazio, N. Batalha, E. C. Tong, & N. Siegler (Austin, United States: SPIE), 19 [Cited on page [3](#).]
- Rasmussen, C. E. & Williams, C. K. I. 2006, *Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning* (Cambridge, Mass: MIT Press) [Cited on pages [46](#), [48](#), [107](#), and [108](#).]
- Rauer, H., Catala, C., Aerts, C., et al. 2014, The PLATO 2.0 Mission, *Experimental Astronomy*, 38, 249 [Cited on page [4](#).]
- Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, Transiting Exoplanet Survey Satellite, *Journal of Astronomical Telescopes, Instruments, and Systems*, 1, 014003 [Cited on page [3](#).]
- Robertson, P., Bender, C., Mahadevan, S., Roy, A., & Ramsey, L. W. 2016, Proxima Centauri as a Benchmark for Stellar Activity Indicators in the Near Infrared, *The Astrophysical Journal*, 832, 112 [Cited on page [29](#).]
- Rustamkulov, Z., Sing, D. K., Mukherjee, S., et al. 2023, Early Release Science of the Exoplanet WASP-39b with JWST NIRSpec PRISM, *Nature*, 614, 659 [Cited on page [4](#).]
- Saar, S. H. & Donahue, R. A. 1997, Activity-Related Radial Velocity Variation in Cool Stars, *The Astrophysical Journal*, 485, 319 [Cited on page [13](#).]
- Sairam, L. & Triaud, A. H. M. J. 2022, The Need for a Public Forecast of Stellar Activity to Optimize Exoplanet Radial Velocity Detections and Transmission Spectroscopy, *Monthly Notices of the Royal Astronomical Society*, 514, 2259 [Cited on page [15](#).]
- Santerne, A., Brugger, B., Armstrong, D. J., et al. 2018, An Earth-sized Exoplanet with a Mercury-like Composition, *Nature Astronomy*, 2, 393 [Cited on page [20](#).]
- Santerne, A., Fressin, F., Díaz, R. F., et al. 2013, The Contribution of Secondary Eclipses as Astrophysical False Positives to Exoplanet Transit Surveys, *Astronomy & Astrophysics*, 557, A139 [Cited on page [5](#).]
- Santos, N. C. 2008, Extra-Solar Planets: Detection Methods and Results, *New Astronomy Reviews*, 52, 154 [Cited on page [14](#).]

- Santos, N. C., Cristo, E., Demangeon, O., et al. 2020, Broadband Transmission Spectroscopy of HD 209458b with ESPRESSO: Evidence for Na, TiO, or Both, *Astronomy & Astrophysics*, 644, A51 [Cited on page [97](#).]
- Santos, N. C., Israelian, G., & Mayor, M. 2004, Spectroscopic [Fe/H] for 98 Extra-Solar Planet-Host Stars: Exploring the Probability of Planet Formation, *Astronomy & Astrophysics*, 415, 1153 [Cited on page [3](#).]
- Santos, N. C., Mayor, M., Naef, D., et al. 2000, The CORALIE Survey for Southern Extra-Solar Planets [Cited on page [13](#).]
- Santos, N. C., Mortier, A., Faria, J. P., et al. 2014, The HARPS Search for Southern Extra-Solar Planets XXXV. The Interesting Case of HD41248: Stellar Activity, No Planets?, *Astronomy & Astrophysics*, 566, A35 [Cited on page [13](#).]
- Schmelling, M. 1995, Averaging Correlated Data, *Physica Scripta*, 51, 676 [Cited on page [43](#).]
- Schöfer, P., Jeffers, S. V., Reiners, A., et al. 2019, The CARMENES Search for Exoplanets around M Dwarfs: Activity Indicators at Visible and near-Infrared Wavelengths, *Astronomy & Astrophysics*, 623, A44 [Cited on page [15](#).]
- Schrijver, C. J. & Zwaan, C. 2000, Solar and Stellar Magnetic Activity, Cambridge Astrophysics Series No. 34 (Cambridge, U.K. ; New York: Cambridge University Press) [Cited on pages [13](#) and [14](#).]
- Seifert, W., Xu, W., Stahl, O., et al. 2016, in *SPIE Astronomical Telescopes + Instrumentation*, ed. C. J. Evans, L. Simard, & H. Takami, Edinburgh, United Kingdom, 990865 [Cited on page [17](#).]
- Shen, Y., Seeger, M., & Ng, A. 2006, in *Advances in Neural Information Processing Systems*, ed. Y. Weiss, B. Schölkopf, & J. Platt, Vol. 18 (MIT Press) [Cited on page [109](#).]
- Siegel, J. C., Rubenzahl, R. A., Halverson, S., & Howard, A. W. 2022, Into the Depths: A New Activity Metric for High-precision Radial Velocity Measurements Based on Line Depth Variations, *The Astronomical Journal*, 163, 260 [Cited on page [14](#).]
- Silva, A. M., Faria, J. P., Santos, N. C., et al. 2022, A Novel Framework for Semi-Bayesian Radial Velocities through Template Matching, *Astronomy & Astrophysics* [Cited on pages [xvi](#), [xvii](#), [27](#), [31](#), [36](#), [39](#), [42](#), [49](#), [50](#), [51](#), [52](#), [54](#), [55](#), [56](#), [57](#), [58](#), and [59](#).]

- Simpson, E. R., Fetherolf, T., Kane, S. R., et al. 2022, Revisiting BD-06 1339b: A Likely False Positive Caused by Stellar Activity, *The Astronomical Journal*, 163, 215 [Cited on pages [5](#) and [13](#).]
- Smette, A., Sana, H., Noll, S., et al. 2015, Molecfit: A General Tool for Telluric Absorption Correction. I. Method and Application to ESO Instruments, *Astronomy & Astrophysics*, 576, A77 [Cited on page [11](#).]
- Smola, A. & Bartlett, P. 2001, in *Advances in Neural Information Processing Systems*, ed. T. Leen, T. Dietterich, & V. Tresp, Vol. 13 (MIT Press) [Cited on page [109](#).]
- Snelson, E. & Ghahramani, Z. 2006, in *Advances in Neural Information Processing Systems*, ed. Y. Weiss, B. Schölkopf, & J. Platt, Vol. 18 (MIT Press) [Cited on page [109](#).]
- Sousa, S. G., Adibekyan, V., Delgado-Mena, E., et al. 2021, SWEET-Cat 2.0: The Cat Just Got SWEETer: Higher Quality Spectra and Precise Parallaxes from *Gaia* eDR3, *Astronomy & Astrophysics*, 656, A53 [Cited on page [98](#).]
- Spanò, P., Mégevand, D., Herreros, J. M., et al. 2010, in *SPIE Astronomical Telescopes + Instrumentation*, ed. I. S. McLean, S. K. Ramsay, & H. Takami, San Diego, California, USA, 77350K [Cited on page [18](#).]
- Storn, R. & Price, K. 1997, A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces, *Journal of Global Optimization*, 11, 341 [Cited on page [120](#).]
- Tal-Or, L., Trifonov, T., Zucker, S., Mazeh, T., & Zechmeister, M. 2019, Correcting HIRES/Keck Radial Velocities for Small Systematic Errors, *Monthly Notices of the Royal Astronomical Society: Letters*, 484, L8 [Cited on pages [57](#) and [135](#).]
- Tinetti, G., Eccleston, P., Haswell, C., et al. 2021, Ariel: Enabling Planetary Science across Light-Years [Cited on page [4](#).]
- Titsias, M. 2009-04-16/2009-04-18, in *Proceedings of Machine Learning Research*, Vol. 5, *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, ed. D. van Dyk & M. Welling (Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA: PMLR), 567–574 [Cited on page [109](#).]
- Trifonov, T., Tal-Or, L., Zechmeister, M., et al. 2020, A Public HARPS Radial Velocity Database Corrected for Systematic Errors, *Astronomy & Astrophysics*, 636, A74 [Cited on page [135](#).]

- Ulmer-Moll, S., Figueira, P., Neal, J. J., Santos, N. C., & Bonnefoy, M. 2019, Telluric Correction in the Near-Infrared: Standard Star or Synthetic Transmission?, *Astronomy & Astrophysics*, 621, A79 [Cited on page [11](#).]
- Vacca, W. D., Cushing, M. C., & Rayner, J. T. 2003, A Method of Correcting Near-Infrared Spectra for Telluric Absorption, *Publications of the Astronomical Society of the Pacific*, 115, 389 [Cited on page [10](#).]
- Vandebril, R., van Barel, M., & Mastronardi, N. 2008, *Matrix Computations and Semiseparable Matrices* (Baltimore: Johns Hopkins University Press) [Cited on page [110](#).]
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nature Methods*, 17, 261 [Cited on page [41](#).]
- Wehbe, B., Cabral, A., Figueira, P., & Avila, G. 2019, in *Fourth International Conference on Applications of Optics and Photonics*, ed. M. F. P. Martins Costa (Lisbon, Portugal: SPIE), 65 [Cited on page [34](#).]
- Wolszczan, A. 1994, Confirmation of Earth-Mass Planets Orbiting the Millisecond Pulsar PSR B1257 + 12, *Science*, 264, 538 [Cited on page [1](#).]
- Wolszczan, A. & Frail, D. A. 1992, A Planetary System around the Millisecond Pulsar PSR1257 + 12, *Nature*, 355, 145 [Cited on page [1](#).]
- Wright, J. T. & Eastman, J. D. 2014, Barycentric Corrections at 1 Cm/s for Precise Doppler Velocities, *Publications of the Astronomical Society of the Pacific*, 126, 838 [Cited on page [6](#).]
- Yixin Wang and Alp Kucukelbir and David M. Blei. 2017, Robust Probabilistic Modeling with {B}ayesian Data Reweighting, *Proceedings of the 34th International Conference on Machine Learning*, 70, 3646 [Cited on page [66](#).]
- Zechmeister, M., Reiners, A., Amado, P. J., et al. 2018, Spectrum Radial Velocity Analyser (SERVAL): High-precision Radial Velocities and Two Alternative Spectral Indicators, *Astronomy & Astrophysics*, 609, A12 [Cited on pages [20](#), [21](#), [41](#), [50](#), [51](#), and [52](#).]