A semi-Bayesian implementation of template matching for precise radial Velocities

André M. Silva[1,2], João P. Faria[1], Nuno C. Santos[1,2], Sérgio G. Sousa[1], Pedro T. P. Viana[1,2]

1 - Instituto de Astrofísica e Ciências do Espaço
2 - Departamento de Física e Astronomia, Faculdade de Ciências

Abstract

The detection and characterization of an increasing variety of exoplanets and, in particular, the derivation of their masses, has been possible thanks to the continuous development of high-resolution, stable spectrographs, and using the Doppler radial-velocity (RV) method. Despite the large success of the CCF methodology, the traditional approach for computing RVs, it is not the optimal approach to study M-type stars. For such cases, the more recent template matching methodology is capable of delivering more precise RVs. State of the art template matching approaches consider the different spectral orders to be independent, derive individual RVs for each and, afterwards, combine all of the information available to build a final RV estimate. However, under normal conditions, i.e. without stellar activity impacting the individual measurements or instrumental effects, we do not expect to find a chromatic dependence of the measured radial-velocity.

In this work we discuss a Bayesian framework where a single RV shift describes the Doppler shift suffered by the entire spectra of a given observation. This approach, similarly to template matching techniques, is based in aligning an high SNR stellar template, with the individual observations, yielding relative RVs. The key differences lie in inclusion of a single parameter in the model, a RV shift that is equal for all orders of the same observation, and the usage of a Bayesian framework, easing the error propagation throughout the model to the final RV estimate. To benchmark this novel framework we compare it against two different implementations of the classical template matching approach., using archival HARPS data of Barnard's star.

The quest for precise Radial Velocities

- Finding and characterizing other Earths, rocky planets with the physical conditions to hold liquid water on their surface is one of the boldest goals of present-day astrophysics. When a star is orbited by a planet, the gravitational pull from the orbiting companion makes the former move about the center of mass of the star-planet system. This motion can be observed as a variation in the radial-velocity (RV, the velocity in the direction of the line-of-sight) of the star as a function of time, holding information about the orbital parameters and mass of the planet.
- However, the RV amplitude induced by the presence of an Earth-like planet orbiting a solar-type star is only of the order of 10 cm/s. Measuring the velocity of a distant star with such a precision is thus an impressive achievement. In practice, the RV of a star is measured from the Doppler shift of stellar spectral lines caused by the star's motion, as depicted in Figure 1.
- The first detection of a planet outside our Solar System around a solar-like star, 51 Pegasi b (Mayor & Queloz 1995), was achieved with radial velocities computed using the Cross Correlation Function (CCF) method. In this method, a weighted mask, with non-zero weights attributed to the expected positions of stellar absorption lines, is cross-correlated with the spectra (e.g. Baranne et al. 1996, Pepe et al 2002).

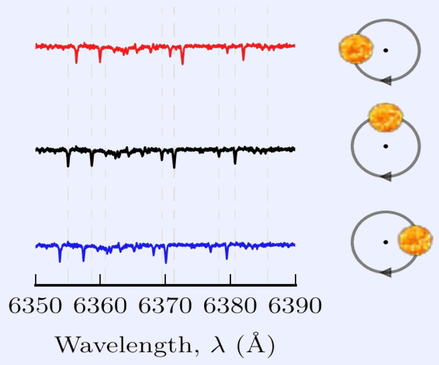

Figure 1: Schematic of the Doppler shift impacting the stellar spectra. Image taken from (Faria J., 2018).

- Despite the CCF method being one of the most prolific exoplanet discovery methods, the construction of the masks poses a challenge under some conditions. This is more evident in the case of M dwarfs (e.g. Rainer et al. 2020; Lafarga et al. 2020), a consequence of:
  a. The high number of stellar spectral lines, due to the lower temperatures of M-type stars, results in most of them being blended, hardening the construction of the CCF mask;
  b. M dwarfs exhibit high levels of stellar activity, creating distortions in the spectral lines which impact the RV measurements.
- For this type of stars, template matching has been shown to be able to surpass the CCF approach (e.g. Anglada-Escudé & Butler 2012; Zechmeister et al. 2018; Lafarga et al. 2020). This is a data-driven method where RVs are computed by aligning a stellar template with each individual observation (Figure 2). Traditional approaches consider the spectral orders independently, deriving individual RVs for each, which are then combined.

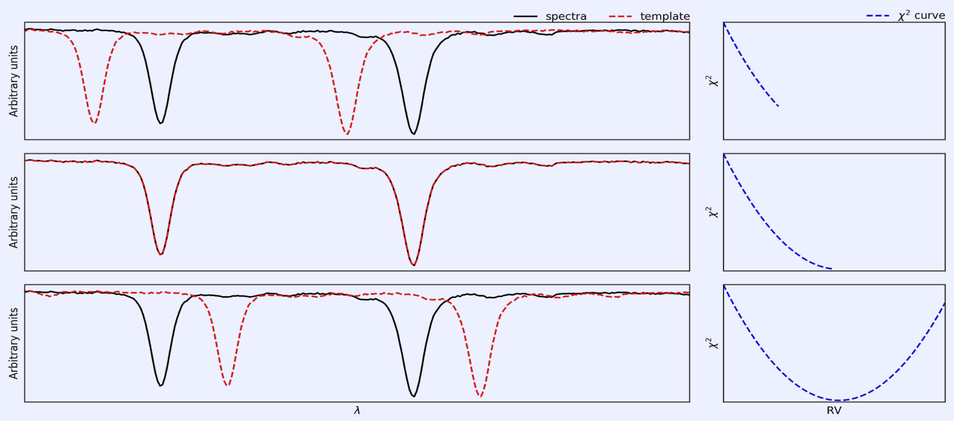

Figure 2: Schematic of the alignment of a stellar template (left pannel, colored in red) with the stellar spectra (left pannel, colored in black), by Doppler-shifting it by different RVs. The alignment is typically done through a least-squares minimization (right panel).

Data pre-processing

- **Removal of telluric features**: Earth's atmosphere absorbs radiation, imprinting telluric absorption features in spectra acquired with ground-based high-resolution spectrographs. This phenomenon strongly depends on the wavelength, airmass of the observations and weather conditions (e.g. Figueira et al. 2012; Cunha et al. 2014). If not corrected it can lead to biased and less precise RV estimates. Even shallow telluric lines, or micro-telluric lines, can induce a significant bias, about 10-20 cm/s (Cunha et al. 2014), on par with, or larger than the signal produced by an Earth-like world around a solar-type star.
- **Stellar template:** The stellar template is the most important component of our model, as it is assumed to be a high signal-to-noise model spectrum that represents very accurately the stellar spectrum at all times. Any observed spectrum is thus assumed to differ from this template only as a result of a Doppler-shift induced by the stellar RV. This high signal-to-noise template is built by combining the information of multiple observations of the same star, on an order-by-order basis, i.e. we construct an independent stellar template for each spectral order. The template is constructed by placing all observations in the rest frame (i.e. we remove contributions from stellar RV and BERV) and computing their mean value. We calculate the mean, instead of summing, in order to keep the count level at a reasonable value and avoid possible numerical issues further ahead.
- **Outlier removal:** Even though the stellar template is generally a good match with the stellar spectra of any given observation, there are some regions where that assumption does not hold. We start by aligning the stellar template and spectra using the initial guess for the RV. Then, to adjust the continuum levels of them both we fit a first-degree polynomial to the ratio between the spectra and template. Then, we compute the logarithm of the ratio between spectra and template and use it as a metric to flag mismatch regions. We consider all points whose metric is more than 6σ away from the median metric (of the entire order) to be outliers, as seen in Figure 3.

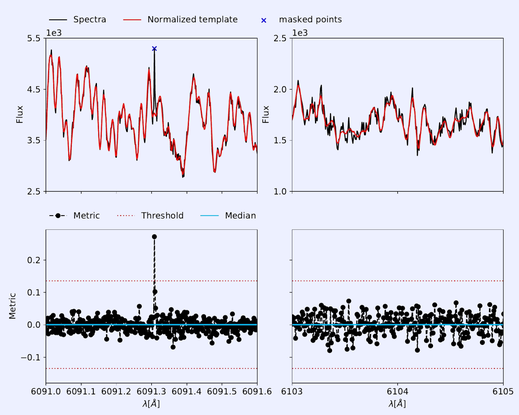

Figure 3: Outlier removal routine for two regions of the same order.Top row: Comparison between the stellar template (red line) and the spectra (black line). The blue crosses represent the points that were flagged by the method. Bottom row: Differences between the template and spectra (black points). The blue line is the median value, whilst the dotted red lines represent a 6σ difference from it.

A semi-Bayesian template matching framework

- The spectra of the stars are collected with specialized instruments - spectrographs - and it is in the form of images, as seen in Figure 5. From the images, the spectral orders - created by the path of the light through the different optical components - are extracted by the official Data Reduction Software (DRS) of the instrument.
- The current approaches to template matching assume the Doppler-shifts associated with the different spectral orders to be independently generated. However, this is clearly not the case when those shifts are the result of the stellar RV component induced by orbiting bodies, like planets or companion stars. In fact, such shifts are achromatic, i.e. their magnitude is independent of the wavelengths associated with the shifted spectral features.
- Given that relative RV estimates, as those obtained through template matching, are primarily used to detect orbiting bodies and characterize their masses and orbits, consistency then suggests one should use a single RV shift to describe simultaneously the differences for all orders between a given spectrum and the template .

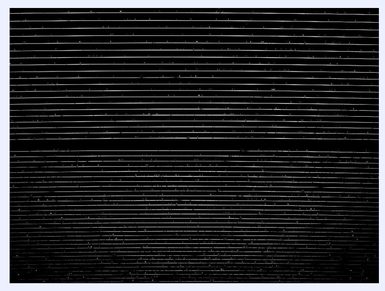

Figure 5: Raw image from the HARPS spectrograph. Each horizontal white line is a spectral order.
Credit: ESO https://www.eso.org/public/images/eso0308c/)

- The casting of RV estimation through template matching into a Bayesian statistical framework would allow for consistent and straightforward characterization of the RV (posterior) probability for any observation. Within a Bayesian framework 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, the latter takes place in the context of template matching, which is why we call semi-Bayesian to the template matching approach for RV estimation described schematically in Figure 6.
- Our RV model contains only one parameter of interest, the relative RV with respect to the frame for which the template RV is zero. We thus wish 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. We will assume the first to be uninformative, taking the form of an uniform probability distribution. The likelihood of the full spectral data, conditioned on a given RV value, is assumed to equal the product of the likelihoods of the fluxes measured for each pixel, i.e. the flux measurements for all pixels are considered to be independent.

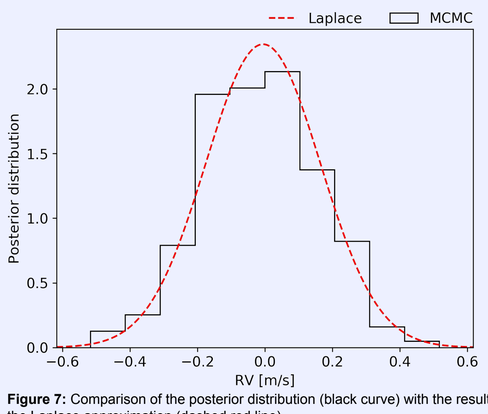

Figure 7: Comparison of the posterior distribution (black curve) with the result of the Laplace approximation (dashed red line).

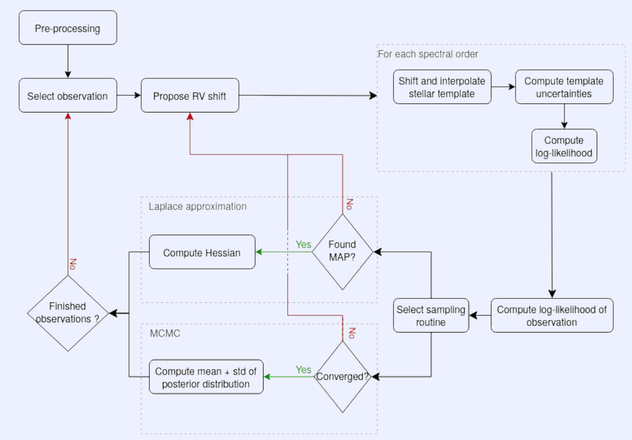

Figure 6: Schematic of the semi-Bayesian approach, considering both an MCMC and Laplace approximation to explore the posterior distribution.

- Through the characterization of the posterior distribution we can estimate the RV shift (mean value) and its uncertainty (standard deviation). However, we found that the application of an MCMC approach, to characterize the posterior distribution, is not viable due to the high computational cost of evaluating the log-likelihood. However, as the posterior for the RV shift is approximately Gaussian, it is possible to use Laplace's approximation to estimate it. This method approximates the posterior distribution with a Gaussian (Figure 7) centered in the posterior's mode with a standard deviation equal to the Hessian of the negative log-posterior evaluated in the MAP estimate (Section 3.4 of Rasmussen & Williams 2006).
- For more details on the methodology we refer to (Silva et al, in prep).

Application to HARPS data

In order to benchmark our algorithm against other template matching pipelines we selected 22 HARPS observations of Barnard's star (GJ699), obtained between 2007-04-04 and 2008-05-02, with program ID 072.C-0488(E). The observations were all reduced with the version 3.5 of the official pipeline and we present the results obtained for HARPS-TERRA, SERVAL and our semi-Bayesian methodology:

- Figure 8 shows the RV timeseries for the three template-based algorithms. The HARPS-TERRA time-series was obtained from Table 6 of Anglada-Escudé & Butler (2012) and the one from SERVAL was through our own application of the most recent public version of that pipeline. For comparison purposes we show the semi-Bayesian measurements after subtracting its own mean RV. A visual comparison of the different time-series allows to see that the RV measurements present the same trends in all cases, with the errors being generally smaller in the HARPS-TERRA pipeline.

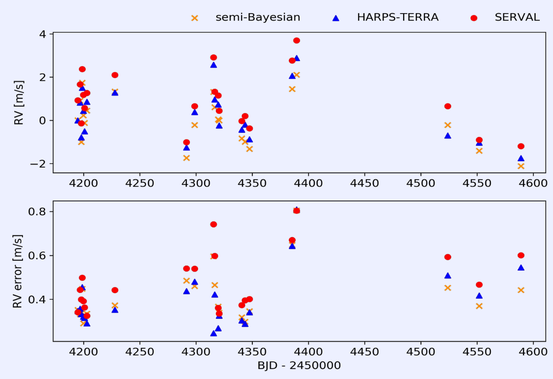

Figure 8: RV time-series of the 22 observations of Barnard's star. The orange crosses are the results from the semi-Bayesian methodology. The blue triangles are the measurements from Table 6 of the HARPS-TERRA paper and the red dots the data from our own application of the most recent, publicly available, version of the SERVAL pipeline. **Top:** Measured RVs. **Bottom:** Uncertainty of each observation.

Method	std RV $[m\ s^{-1}]$	median σ_{RV} $[m\ s^{-1}]$
DRS	1.51	0.54
HARPS-TERRA	1.22	0.35
SERVAL-PAPER	1.30	—
SERVAL	1.28	0.44
semi-Bayesian	1.14	0.37

Table 1: Comparison of the standard deviation and median error of different template-based methodologies when applied to Barnard's star. "HARPS-TERRA" and "SERVAL-PAPER" results were obtained from Anglada-Escudé & Butler (2012) and Zechmeister et al. (2018), respectively. "SERVAL" was obtained by using the latest (publicly) available version of the pipeline, "DRS" from the official pipeline, and "semi-Bayesian" from the methodology discussed within this poster.

- From Table 1 we find that our approach reaches the meter per second precision on HARPS data, with a scatter significantly smaller (~ 14 cm/s) than those from HARPS-TERRA and SERVAL. Regarding the median RV error we find that the semi-Bayesian method leads to slightly smaller errors than those from SERVAL, but larger than the ones from HARPS-TERRA. We nevertheless note that the differences in the median error are of a few cm/s. Lastly, a comparison of our results with those from the official pipeline show that, similarly to the other template matching approaches, the scatter and median RV of our two methods are smaller than those from the CCF method.
- Overall, we find that our results are slightly better than those from the other pipelines, showing that our methodology is working properly. Nonetheless, there is still a need of a subsequent analysis of the datasets to understand the impact that our new RVs have on the determination of the planet parameters. For a more in-depth analysis of this implementation we refer to (Silva et al, in prep).

Key points

- In this poster we presented a new methodology to derive precise Radial Velocities. Contrarily to current techniques, we do not assume independent stellar orders but, instead, use a singular RV value to describe the entirety of the observation;
- We work in a semi-Bayesian framework, allowing for a more robust characterization of the radial velocity;
- Testing our semi-Bayesian approach in HARPS observations of Barnard's star revealed:
  a. Smaller scatter and median uncertainty than the CCF method of ESO's pipeline;
  b. Smaller scatter (~14 cm/s) and similar median precision to the RVs from the HARPS-TERRA and SERVAL pipelines;

Acknowledgements

This work was supported by FCT - Fundação para a Ciência e a Tecnologia through national funds and by FEDER through COMPETE2020 - Programa Operacional Competitividade e Internacionalização by these grants: UID/FIS/04434/2019; UIDB/04434/2020; UIDP/04434/2020; PTDC/FIS-AST/32113/2017 & POCI-01-0145-FEDER-032113; PTDC/FIS-AST/28953/2017 & POCI-01-0145-FEDER-028953. A.M.S acknowledges support from the Fundação para a Ciência e a Tecnologia (FCT) through the Fellowship 2020.05387.BD. and POCH/FSE (EC). J.P.F. is supported in the form of a work contract funded by national funds through FCT with reference DL57/2016/CP1364/CT0005. S.G.S acknowledges the support from FCT through the contract nr. CEECIND/00826/2018 and POPH/FSE (EC)

References

- Anglada-Escudé, G. & Butler, R. P. 2012, ApJS, 200, 15
- Baranne, A., Queloz, D., Mayor, M., et al. 1996, Astron. Astrophys. Suppl. Ser.,119, 373
- Cunha, D., Santos, N. C., Figueira, P., et al. 2014, A&A, 568, A35
- Figueira, P., Pepe, F., Lovis, C., & Mayor, M. 2010, A&A, 515, A106, arXiv:1003.0541
- João Pedro de Sousa Faria (2018). Exoplanet detection in metal-poor stars. PhD, Faculdade de Ciências da Universidade do Porto.
- Lafarga, M., Ribas, I., Lovis, C., et al. 2020, arXiv:2003.07471 [astro-ph], arXiv:2003.07471
- Mayor, M. & Queloz, D. 1995, Nature, 378, 355
- Pepe, F., Mayor, M., Galland, F., et al. 2002, A&A, 388, 632
- Rainer, M., Borsa, F., & Affer, L. 2020, Exp Astron, 49, 73, arXiv: 2002.06844
- Rasmussen, C. E. & Williams, C. K. I. 2006, Gaussian processes for machine learning, Adaptive computation and machine learning (Cambridge, Mass:MIT Press),
- Zechmeister, M., Reiners, A., Amado, P. J., et al. 2018, A&A, 609, A12