



**HAL**  
open science

# Planetary systems around low mass stars with SPIRou

Merwan Ould-Elhkim

► **To cite this version:**

Merwan Ould-Elhkim. Planetary systems around low mass stars with SPIRou. Solar and Stellar Astrophysics [astro-ph.SR]. Université Paul Sabatier - Toulouse III, 2023. English. NNT : 2023TOU30335 . tel-04618508

**HAL Id: tel-04618508**

**<https://theses.hal.science/tel-04618508>**

Submitted on 20 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Doctorat de l'Université de Toulouse

préparé à l'Université Toulouse III - Paul Sabatier

---

Systèmes planétaires autour d'étoiles de faible masse avec  
SPIRou

---

Thèse présentée et soutenue, le 6 décembre 2023 par

**Merwan OULD-ELHKIM**

**École doctorale**

SDU2E - Sciences de l'Univers, de l'Environnement et de l'Espace

**Spécialité**

Astrophysique, Sciences de l'Espace, Planétologie

**Unité de recherche**

IRAP - Institut de Recherche en Astrophysique et Planetologie

**Thèse dirigée par**

Claire MOUTOU et Jean-François DONATI

**Composition du jury**

M. Andrew COLLIER CAMERON, Rapporteur, University of St Andrews, School of Physics and Astronomy

M. François BOUCHY, Rapporteur, Université de Genève

Mme Magali DELEUIL, Examinatrice, Aix Marseille Université

Mme Claire MOUTOU, Directrice de thèse, CNRS / IRAP

M. Jean-François DONATI, Co-directeur de thèse, CNRS / IRAP

M. Hervé CARFANTAN, Président, Université Toulouse III - Paul Sabatier



Université  
de Toulouse

# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

---

---

Présentée et soutenue le *06/12/2023* par :

**Merwan Ould-Elhkim**

**Détection d'exoplanètes autour d'étoiles de faible masse par  
vélocimétrie avec le spectropolarimètre infrarouge SPIRou**

-

**Exoplanets detection around low mass stars with SPIRou**

---

---

### JURY

ANDREW COLLIER	Rapporteur	Professeur
CAMERON FRANÇOIS BOUCHY	Rapporteur	Professeur associé
HERVÉ CARFANTAN	Président du jury	Professeur
MAGALI DELEUIL	Examinatrice	Professeure
CLAIRE MOUTOU	Directeur de thèse	Directrice de recherche
JEAN-FRANÇOIS DONATI	Co-directeur de thèse	Directeur de recherche

---

**École doctorale et spécialité :**

*SDU2E : Astrophysique, Sciences de l'Espace, Planétologie*

**Unité de Recherche :**

*Institut de Recherche en Astrophysique et Planétologie (UMR 5277)*

**Directeur(s) de Thèse :**

*Claire Moutou et Jean-François Donati*

**Rapporteurs :**

*Andrew Collier Cameron et François Bouchy*







# Abstract

As of the time of writing this thesis, 5,502 exoplanets have been detected and confirmed. The diversity of those exoplanets challenges and further our knowledge of planet formation and evolution, with implications for the origin of our own Solar system. Moreover, the search for other worlds is closely linked to the quest for extraterrestrial life which begins by identifying planets within the habitable zone (HZ) of their host stars, where liquid water could potentially exist. However, only a small fraction of known Earth-like exoplanets are located within their host star's HZ and many of these HZ planets are also too distant to study their atmospheres and search for signs of life using space-based missions like the James Webb Space Telescope (JWST) or the upcoming Extremely Large Telescope (ELT). Therefore, there is a need to discover nearby planets within the HZ of their host stars, and the radial velocity (RV) method, which indirectly detects planets by monitoring their gravitational tug on their host stars, is one of the most effective methods for this purpose.

M dwarfs, which are abundant in the solar neighborhood and known for hosting a higher number of rocky planets than other stellar types, are prime targets for detecting HZ planets. However, observing M dwarfs in the optical domain is challenging due to their faintness, and their intense magnetic fields can create magnetically induced RV signals that may either overshadow or even mimic planetary signals. To overcome these challenges, the near-infrared (nIR) spectral domain has emerged as a promising avenue, offering increased RV information content and reduced stellar activity signals. Consequently, nIR spectrographs such as SPIRou (SpectroPolarimètre InfraRouge), mounted at the Canada-France-Hawaii Telescope (CFHT) on Mauna Kea, have seen significant development.

This thesis takes place in the context of the SPIRou Legacy Survey - Planet Search (SLS-PS) program, dedicated to detecting and characterizing planetary systems around 57 nearby M dwarfs with observations spanning from February 2019 to June 2022. From those observations spectra are acquired, from which Doppler shifts are computed relative to a reference spectrum, yielding RV time series. These RV time series are then analyzed to detect planetary signals through the identification of periodic variations. However, the presence of systematic errors in the data can give rise to non-planetary signals, they can originate from various sources, including stellar activity, telluric lines, or instrumental artifacts. The primary focus of this thesis is the development and application of a data-driven method to correct these systematic errors, unveiling hidden planetary signals. Applying this method to simulated data and real SPIRou observations from the SLS-PS program demonstrates its effectiveness in correcting for systematic effects, allowing the retrieval of known and previously undetected exoplanetary signals. This research underscores, with systematic corrections in place, that SPIRou has the potential to fulfill its promise of discovering new exoplanetary

systems.

# Résumé

Au moment de la rédaction de cette thèse, 5 502 exoplanètes ont été détectées et confirmées. La diversité des exoplanètes découvertes remet en question et fait progresser nos connaissances sur la formation et l'évolution des planètes, ce qui a des répercussions sur l'origine de notre propre système solaire. En outre, la recherche d'autres mondes est étroitement liée à la quête de la vie extraterrestre, qui commence par l'identification de planètes dans la zone habitable (ZH) de leur étoile hôte, où de l'eau liquide pourrait exister à leur surface. Cependant, seule une petite fraction des exoplanètes connues et semblables à la Terre se trouve dans la ZH de leur étoile hôte et bon nombre de ces planètes situées dans la ZH sont également trop éloignées pour que l'on puisse étudier leur atmosphère et rechercher des signes de vie à l'aide de missions spatiales telles que le JWST ou le futur télescope ELT. Il est donc nécessaire de découvrir des planètes proches dans la ZH de leur étoile hôte, et la méthode des vitesses radiales (RV), qui détecte indirectement les planètes en mesurant leur attraction gravitationnelle sur leur étoile hôte, est l'une des méthodes les plus efficaces à cette fin.

Les naines rouges, qui sont abondantes dans le voisinage solaire et connues pour abriter un plus grand nombre de planètes rocheuses que les autres types d'étoiles, sont des cibles de choix pour la détection de planètes dans la ZH de leur étoile. Cependant, l'observation des naines rouges dans l'optique est difficile en raison de leur faible luminosité, et leurs champs magnétiques intenses peuvent créer des signaux RV induits magnétiquement qui peuvent soit éclipser, voire imiter les signaux planétaires. Pour surmonter ces difficultés, le proche infrarouge (nIR) est apparu comme une voie prometteuse, offrant une plus grande quantité d'informations RV et des signaux d'activité stellaire réduits. Par conséquent, les spectrographes infrarouges tels que SPIRou (SpectroPolarimètre InfraRouge), monté au télescope Canada-France-Hawaii (TCFH) sur le Mauna Kea, ont connu un développement important.

Cette thèse s'inscrit dans le cadre du programme SPIRou Legacy Survey - Planet Search (SLS-PS), dédié à la détection et à la caractérisation de systèmes planétaires autour de 57 naines rouges proches, avec des observations allant de février 2019 à juin 2022. Ces observations permettent d'acquérir des spectres, à partir desquels les décalages Doppler sont calculés par rapport à un spectre de référence, ce qui permet d'obtenir des séries temporelles de RV qui sont ensuite analysées pour détecter des signaux planétaires grâce à l'identification de variations périodiques. Cependant, la présence d'erreurs systématiques dans les données peut donner lieu à la détection de signaux non planétaires qui peuvent provenir de diverses sources, notamment l'activité stellaire, des résidus de raies telluriques ou des effets instrumentaux. L'objectif principal de cette thèse est le développement et l'application d'une méthode pour corriger ces faux signaux, dévoilant ainsi les signaux planétaires

qu'ils masquent. L'application de cette méthode à des données simulées et à des observations SPIRou du programme SLS-PS démontre son efficacité dans la correction des effets systématiques, permettant la récupération de signaux exoplanétaires connus et aussi précédemment non détectés. Cette recherche souligne que, moyennant une correction des erreurs systématiques au préalable, SPIRou a le potentiel de remplir sa promesse de découvrir de nouveaux systèmes exoplanétaires.

# Remerciements

Ce manuscrit présente le travail que j'ai accompli au cours de ces trois années de thèse au sein de l'Institut de Recherche en Astrophysique et Planétologie (IRAP). Cela n'aurait jamais été possible sans l'encadrement de ma directrice de thèse, Claire Moutou, et de mon co-directeur de thèse, Jean-François Donati, que je tiens en premier lieu à remercier pour leur soutien et le partage de leur expérience. Leur exemple aura su me marquer et me guidera tout le long de mon parcours.

Je souhaite aussi exprimer ma gratitude envers les autres membres du laboratoire, qu'ils soient doctorants, post-doctorants ou permanents, avec lesquels il a été agréable de partager les couloirs. J'ai également une pensée pour Pascal, parti à la retraite à la fin de ma thèse. Mes remerciements s'étendent à toute l'équipe SPIROU et j'espère avoir l'opportunité de contribuer davantage au sein de cette équipe à l'avenir et de participer à rendre justice à la qualité des données produites par cet instrument. Je remercie bien évidemment l'équipe de Montréal avec laquelle j'ai eu l'opportunité de travailler directement pendant quelques mois et qui m'a chaleureusement accueilli. Et enfin, merci à Nathan pour avoir été mon premier mentor dans ce domaine des exoplanètes.

Je suis vraiment très reconnaissant d'avoir pu faire ma thèse à Toulouse, qui sera toujours une ville spéciale pour moi en raison des rencontres que j'y ai faites et je considère avoir beaucoup de chance dans ma vie de vous connaître. La liste serait beaucoup trop longue et me connaître en retour c'est savoir que s'épancher sur chaque souvenir n'est pas forcément l'exercice dans lequel j'excelle le plus, mais si tu te reconnais en lisant ces lignes c'est que je pense à toi.

Je finis ces remerciements par mes premiers soutiens que sont mes parents, qui m'ont soutenu de toute les manières possibles, dont celles qu'on ne réalise qu'après.



# Foreword

The field of exoplanets has evolved since the the 19th and early 20th century, when it was facing skepticism by astronomers and filled with unsubstantiated claims (Jacob, 1855; See, 1896; Berman, 1932; van de Kamp, 1969). Today, exoplanet research is one of the most active and captivating areas in astrophysics, driven by the enthusiasm that followed the Nobel laureates Michel Mayor and Didier Queloz’s groundbreaking discovery of an exoplanet orbiting a Sun-like star (Mayor & Queloz, 1995).

The first detection of an exoplanet was achieved using the radial velocity method, which indirectly identifies planets by monitoring the gravitational wobble they induce in their host stars. Since that discovery, various other detection techniques have emerged. As of the time of this thesis, a total of 5,502 exoplanets have been discovered and confirmed<sup>1</sup>, and the radial velocity method continues to be the second most successful approach, accounting for 1,056 confirmed exoplanet detections to date.

The collaborative efforts of astronomers over the past few decades have led to all these findings which provide us with the opportunity to compare theoretical models of planet formation and evolution with observed exoplanetary systems (Morbidelli, 2014; Bitsch et al., 2014; Baruteau et al., 2014, 2016; Mordasini, 2018) and study their population (Mayor et al., 2011; Wang et al., 2015; Santerne et al., 2016; Fulton et al., 2017; Dong et al., 2018; Mordasini, 2020; Bashi et al., 2020).

The quest for exoplanets also presents an opportunity to search for extraterrestrial life. While the conditions necessary for a planet to support life are not yet fully understood, ongoing efforts are being made to define them (Cockell et al., 2016). It is generally agreed that certain attributes, such as the presence of carbon, liquid water, and a source of energy, are essential for the possibility of life. Consequently, the first crucial step in potentially discovering life elsewhere is to identify planets orbiting their host stars within the so-called habitable zone (HZ; Kasting et al. (1993); Kopparapu et al. (2014)). The HZ is a region of radial distances from a star where planets could maintain liquid water on their surfaces.

Detecting Earth-like planets within the HZ of their host stars remains a formidable challenge, even for modern high-precision velocimeters. This difficulty arises from the weak RV signatures of such planets (of around  $0.09 \text{ m.s}^{-1}$  for the Earth orbiting the Sun). Consequently, as of now, no planets within the HZ of Sun-like stars have been detected using the radial velocity method.

However, the situation is more favorable when it comes to detecting exoplanets in the HZ of

---

<sup>1</sup><https://exoplanetarchive.ipac.caltech.edu/>

M dwarfs, due to their lower masses. As a result, the list of habitable exoplanets orbiting M dwarfs is continually expanding (Anglada-Escudé et al., 2016; Gillon et al., 2017; Dittmann et al., 2017; Astudillo-Defru et al., 2017a; Suárez Mascareño et al., 2023; Kossakowski et al., 2023). Not only are M dwarfs more favorable for detecting planets in their HZ due to their lower mass, but they are also the most abundant type of stars in our solar neighborhood (Henry et al., 2006; Winters et al., 2019). Furthermore, M dwarfs are known to host a significantly larger number of rocky planets compared to other stellar types (Bonfils et al., 2013; Dressing & Charbonneau, 2015; Sabotta et al., 2021).

These factors make M dwarfs the most promising targets for the detection of exoplanets in the HZ. However, while several habitable planets have been identified, only a handful of them are close enough to enable potential atmospheric characterization using space-based missions such as the James Webb Space Telescope (JWST; Gardner et al. (2006)) or ground-based missions like the Extremely Large Telescope (ELT)<sup>2</sup>. Therefore, detecting new habitable planets around M dwarfs remains a critical endeavor for advancing our understanding of exoplanetary atmospheres and the search for potential signs of extraterrestrial life.

M dwarfs also come with their challenges for exoplanet detection compared to G and K stars, due to their lower effective temperatures and larger convective zones, they tend to exhibit higher levels of magnetic activity. This heightened activity can lead to RV signals associated with various phenomena on the stellar surface, in the atmosphere, and in the corona of these stars. Stellar surface features like spots, faculae, and plage can interfere with RV measurements, potentially masking planetary signals or even mimicking them with periodic variations matching the star's rotational period and its harmonics.

An essential aspect in these activity-induced RV signals is the contrast between the effective temperatures of bright and dark surface features on M dwarfs, as it determines how distinct these features are from the quiet photosphere which has a direct impact on the magnitude of the RV signatures induced by stellar activity (Berdyugina, 2005). Observing in the near-infrared (nIR) can be advantageous because it reduces the contrast between these surface features, which is expected to help mitigate the contributions of activity-related RV variations (Mahmud et al., 2011; Crockett et al., 2012). Additionally, M dwarfs yield the most valuable RV information in the red-optical spectrum while being brightest in the nIR (Reiners et al., 2010, 2018; Reiners & Zechmeister, 2020).

As a result, nIR spectrographs have experienced significant development in recent years such as the one employed in this thesis, SPIRou<sup>3</sup> (Donati et al., 2020). The nIR offers clear advantages for exoplanet detection around M dwarfs, however technology for precise radial velocity measurements in the nIR is relatively new compared to the optical domain and does not benefit from the decades of instrumental and pipeline refinement seen in the optical. Additionally, working in the nIR presents unique challenges, including maintaining instrument stability at low temperatures, achieving wavelength calibration over a new spectral range, addressing H4RG detector persistence, and effectively correcting for telluric effects.

In this thesis, the aim is to introduce a new method for mitigating systematic errors and revealing planetary signals within SPIRou data. After providing the astrophysical context

---

<sup>2</sup><https://www.eso.org/public/france/teles-instr/elt>

<sup>3</sup>SpectroPolarimètre InfraRouge

in chapter 1, the newly developed data-driven approach is presented in chapter 2. This approach uses weighted Principal Component Analysis (wPCA) to estimate and remove systematics from the data through a three-step process, resulting in the method named **Wapiti** (Weighted principAl comPonent analysIs reconsTructIon). In chapter 3, the method is tested on simulations to assess its ability to detect and characterize planetary signals, particularly in the presence of telluric lines in RV time series. Finally, in Chapter 4, the method is applied to three different targets, demonstrating how it can be integrated into the broader context of the SPIRou Legacy Survey Planet Search program.



# Avant-propos

Le domaine des exoplanètes a évolué depuis le 19<sup>e</sup> et début du 20<sup>e</sup> siècle où il était confronté au scepticisme des astronomes et faisait l'objet d'affirmations non fondées ([Jacob, 1855](#); [See, 1896](#); [Berman, 1932](#); [van de Kamp, 1969](#)). Aujourd'hui, la recherche d'exoplanètes est l'un des domaines les plus actifs et les plus captivants de l'astrophysique, stimulée par l'enthousiasme qui a suivi la découverte révolutionnaire d'une exoplanète en orbite autour d'une étoile semblable au Soleil par les lauréats du prix Nobel Michel Mayor et Didier Queloz ([Mayor & Queloz, 1995](#)).

La première détection d'une exoplanète a été réalisée grâce à la méthode des vitesses radiales, qui identifie indirectement les planètes en surveillant l'oscillation gravitationnelle qu'elles induisent sur leur étoile hôte. Au moment de la rédaction de cette thèse, 5 502 exoplanètes ont été découvertes et confirmées, et même si d'autres méthodes de détection ont depuis vu le jour, la méthode des vitesses radiales reste la deuxième méthode la plus efficace, avec 1 056 détections d'exoplanètes confirmées à ce jour.

Le travail et la collaboration des astronomes au cours des dernières décennies a conduit à ces découvertes qui nous donnent aujourd'hui l'occasion de comparer les modèles théoriques de formation et d'évolution des planètes avec les systèmes exoplanétaires observés ([Morbidelli, 2014](#); [Bitsch et al., 2014](#); [Baruteau et al., 2014, 2016](#); [Mordasini, 2018](#)) et d'étudier leur population ([Mayor et al., 2011](#); [Wang et al., 2015](#); [Santerne et al., 2016](#); [Fulton et al., 2017](#); [Dong et al., 2018](#); [Mordasini, 2020](#); [Bashi et al., 2020](#)).

La recherche d'exoplanètes offre également l'opportunité de rechercher la vie extraterrestre. Bien que les conditions nécessaires pour qu'une planète soit propice à la vie ne soient pas encore totalement comprises, des efforts continus sont déployés pour les définir ([Cockell et al., 2016](#)). Il est généralement admis que certaines conditions, tels que la présence de carbone, d'eau liquide et d'une source d'énergie, sont essentielles à l'apparition de la vie. Par conséquent, la première étape cruciale dans la découverte potentielle de la vie ailleurs est d'identifier les planètes en orbite autour de leur étoile hôte dans la zone dite habitable (ZH; [Kasting et al. \(1993\)](#); [Kopparapu et al. \(2014\)](#)). La ZH est une région de distances radiales par rapport à une étoile où les planètes pourraient conserver de l'eau liquide à leur surface.

Détecter des planètes semblables à la Terre dans la ZH de leur étoile hôte reste un formidable défi, même pour les spectrographes modernes de haute précision. Cette difficulté provient de la faible signature RV de ces planètes (de l'ordre de  $0.09 \text{ m.s}^{-1}$  pour la Terre en orbite autour du Soleil). Par conséquent, à ce jour, aucune planète située dans la ZH d'étoiles semblables au Soleil n'a été détectée à l'aide de la technique des vitesses radiales.

En revanche, la situation est plus favorable lorsqu'il s'agit de détecter des exoplanètes dans la ZH d'étoiles de type M (naines rouges) en raison de leur masse plus faible. Par conséquent, la liste des exoplanètes habitables en orbite autour des naines rouges ne cesse de s'allonger (Zieba et al., 2023; Anglada-Escudé et al., 2016; Dittmann et al., 2017; Astudillo-Defru et al., 2017a; Suárez Mascareño et al., 2023; Kossakowski et al., 2023). Non seulement les naines rouges sont plus propices à la détection de planètes dans leur ZH en raison de leur masse plus faible, mais elles constituent également le type d'étoiles le plus abondant dans notre voisinage solaire (Henry et al., 2006; Winters et al., 2019). De plus, les naines rouges sont connues pour abriter un nombre significativement plus important de planètes rocheuses par rapport aux autres types stellaires (Bonfils et al., 2013; Dressing & Charbonneau, 2015; Sabotta et al., 2021).

Ces facteurs font des naines rouges les cibles les plus prometteuses pour la détection d'exoplanètes dans la ZH. Cependant, bien que plusieurs planètes habitables aient été identifiées, seule une poignée d'entre elles sont suffisamment proches pour permettre une caractérisation atmosphérique à l'aide de missions spatiales telles que le télescope spatial James Webb (JWST) (Gardner et al., 2006) ou de missions terrestres telles que l'Extremely Large Telescope (ELT)<sup>4</sup>. La détection de nouvelles planètes habitables autour de naines rouges reste essentielle pour faire progresser notre compréhension des atmosphères exoplanétaires et la recherche de signes potentiels de vie extraterrestre.

Les naines rouges posent également des problèmes pour la détection d'exoplanètes en raison de leurs champs magnétiques puissants, pouvant conduire à des signaux RV (Radial Velocity) associés à divers phénomènes à la surface, dans l'atmosphère et dans la couronne de ces étoiles. Les caractéristiques de la surface stellaire telles que les taches, les facules et les plages peuvent interférer avec les mesures RV, masquant potentiellement les signaux planétaires ou les imitant même par des variations périodiques correspondant à la période de rotation de l'étoile et à ses harmoniques.

Un aspect essentiel de ces signaux RV induits par l'activité est le contraste entre les températures effectives des zones brillantes et sombres à la surface des naines rouges, car il détermine à quel point ces caractéristiques sont distinctes de la photosphère, ce qui a un impact direct sur l'ampleur des signatures RV induites par l'activité stellaire (Berdyugina, 2005). L'observation dans le proche infrarouge (nIR) peut être avantageuse car elle réduit le contraste entre ces caractéristiques de surface, ce qui devrait contribuer à atténuer les contributions des variations RV liées à l'activité (Mahmud et al., 2011; Crockett et al., 2012). En outre, les naines rouges fournissent le plus d'informations RV dans l'infrarouge tout en y étant les plus brillantes (Reiners et al., 2010, 2018; Reiners & Zechmeister, 2020).

En conséquence, les spectrographes nIR ont connu un développement important ces dernières années, comme celui utilisé dans cette thèse, SPIRou<sup>5</sup> (Donati et al., 2020). Le proche infrarouge offre des avantages évidents pour la détection d'exoplanètes autour des naines rouges. Cependant, la technologie pour les mesures précises de la vitesse radiale dans le proche infrarouge est relativement nouvelle par rapport au domaine optique et ne bénéficie pas des décennies de développement instrumental et de pipeline comme dans l'optique. En outre, le travail dans le proche infrarouge présente des défis uniques, notamment le maintien

---

<sup>4</sup><https://www.eso.org/public/france/teles-instr/elt>

<sup>5</sup>SpectroPolarimètre InfraRouge

de la stabilité de l'instrument à basse température, l'étalonnage de la longueur d'onde sur une nouvelle gamme spectrale, la persistance du détecteur H4RG et la correction efficace des effets telluriques.

Dans cette thèse, l'objectif est d'introduire une nouvelle méthode pour atténuer les systématiques et révéler les signaux planétaires dans les données SPIRou. Après avoir présenté le contexte astrophysique dans le chapitre 1, la nouvelle approche est présentée dans le chapitre 2. Cette approche utilise l'analyse en composantes principales pondérées (wPCA) pour estimer et supprimer les systématiques des données grâce à un processus en trois étapes, ce qui aboutit à la méthode appelée `Wapiti` (Weighted principal component analysis reconstruction). Au chapitre 3, la méthode est testée sur des simulations afin d'évaluer sa capacité à détecter et à caractériser les signaux planétaires, en particulier en présence de contamination par les raies telluriques dans les séries temporelles RV. Enfin, dans le chapitre 4, la méthode est appliquée à trois cibles différentes, démontrant comment elle peut être intégrée dans le contexte plus large du programme SPIRou Legacy Survey Planet Search.



# Contents

<b>Abstract</b>	<b>iii</b>
<b>Résumé</b>	<b>v</b>
<b>Remerciements</b>	<b>vii</b>
<b>Foreword</b>	<b>ix</b>
<b>Avant-propos</b>	<b>xiii</b>
<b>Table of Contents</b>	<b>xix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Historical context . . . . .	1
1.1.1 Prehistory of exoplanets . . . . .	1
1.1.2 First detections . . . . .	4
1.2 On the detection of exoplanets . . . . .	6
1.2.1 Radial velocity . . . . .	6
1.3 The era of exoplanets . . . . .	11
1.3.1 A variety of worlds . . . . .	11
1.3.2 Searching for habitable planets . . . . .	12
1.4 The hunt for planets around M dwarfs . . . . .	16
1.4.1 Habitability around M dwarfs . . . . .	16
1.4.2 Stellar activity and its impact on RV time series . . . . .	17
1.4.3 Mitigating stellar activity . . . . .	19
1.4.4 Observing in the near-infrared . . . . .	20
1.5 SPIRou . . . . .	22
1.5.1 Description of the instrument . . . . .	22
1.5.2 The SPIRou Legacy Survey . . . . .	22
1.5.3 A Pipeline to Reduce Observations (APER0) . . . . .	24
1.5.4 Extracting the RV with the line-by-line method . . . . .	24
1.6 Context of the thesis . . . . .	26
1.6.1 Zero-Point correction . . . . .	27
1.6.2 Overview of the thesis . . . . .	28
<b>2 Wapiti: a data-driven approach to correct for systematics in LBL RV data</b>	<b>31</b>

2.1	Overview . . . . .	31
2.2	The <code>Wapiti</code> method . . . . .	32
2.2.1	Permutation test or is the component better than noise ? . . . . .	33
2.2.2	Leave-p-out cross-validation or is the component chromatically stable ? . . . . .	33
2.2.3	Re-ordering or is the component good for reconstruction ? . . . . .	36
2.3	Results . . . . .	39
<b>3</b>	<b>Simulating the impact of telluric absorption lines on radial velocities</b>	<b>41</b>
3.1	Overview . . . . .	41
3.2	The challenges of correcting for telluric lines . . . . .	42
3.3	The impact of telluric absorption lines . . . . .	43
3.3.1	Outline of the method . . . . .	43
3.3.2	Results . . . . .	44
3.4	wPCA and telluric lines . . . . .	45
3.4.1	Principal vectors . . . . .	45
3.4.2	Detection capacity . . . . .	51
3.4.3	Recovering the injected signals . . . . .	56
3.5	Summary . . . . .	58
<b>4</b>	<b>Integrating <code>Wapiti</code> within the SPIRou Legacy Survey - Planet Search program</b>	<b>63</b>
4.1	Overview . . . . .	63
4.2	Confirming signals present in the data with <code>Wapiti</code> . . . . .	64
4.2.1	G1480 . . . . .	64
4.2.2	G1382 . . . . .	64
4.3	Detecting new and hidden signals in the data with <code>Wapiti</code> . . . . .	66
4.3.1	G1725 B . . . . .	66
<b>5</b>	<b>Conclusion and perspectives</b>	<b>75</b>
5.1	Summary of the thesis . . . . .	75
5.1.1	Capacity and limitations of the <code>Wapiti</code> method . . . . .	77
5.2	Perspectives . . . . .	80
5.2.1	On <code>Wapiti</code> . . . . .	80
5.2.2	On other works . . . . .	81
5.2.3	Characterizing planetary systems around M dwarfs . . . . .	82
<b>6</b>	<b>Conclusion et perspectives</b>	<b>85</b>
6.1	Résumé de la thèse . . . . .	85
6.1.1	Capacité et limites de la méthode <code>Wapiti</code> . . . . .	87
6.2	Perspectives . . . . .	90
6.2.1	Sur <code>Wapiti</code> . . . . .	90
6.2.2	Sur d'autres travaux . . . . .	92
6.2.3	Caractériser les systèmes planétaires autour des naines rouges . . . . .	93
	<b>List of Figures</b>	<b>96</b>
	<b>List of Tables</b>	<b>97</b>

<b>Bibliography</b>	<b>121</b>
<b>A Other detection methods for exoplanets</b>	<b>123</b>
A.1 Imaging . . . . .	123
A.2 Astrometry . . . . .	124
A.3 Transit . . . . .	126
A.4 Orbital brightness modulation . . . . .	127
A.5 Timing Variations . . . . .	127
A.5.1 Transit timing variations . . . . .	127
A.5.2 Eclipse timing variations . . . . .	128
A.5.3 Pulsar timing variations . . . . .	128
A.5.4 Pulsation timing variations . . . . .	128
A.6 Microlensing . . . . .	128
<b>B BERV-Invariant periodogram</b>	<b>131</b>
B.1 Overview . . . . .	131
B.2 Partial and semi-partial correlation . . . . .	131
B.3 Semi-partial distance correlation . . . . .	132
B.4 BERV-I periodogram . . . . .	133
<b>C Publications</b>	<b>137</b>
C.1 List of publications . . . . .	137
C.2 First author publications . . . . .	137
C.2.1 <i>Wapiti</i> : A data-driven approach to correct for systematics in RV data	137
C.2.2 (Submitted to A&A) Confirmation of an exoplanet around Gl480 and detection of a candidate exoplanet orbiting Gl382 . . . . .	158



# Chapter 1

## Introduction

### Contents

---

1.1	Historical context . . . . .	1
1.2	On the detection of exoplanets . . . . .	6
1.3	The era of exoplanets . . . . .	11
1.4	The hunt for planets around M dwarfs . . . . .	16
1.5	SPIRou . . . . .	22
1.6	Context of the thesis . . . . .	26

---

## 1.1 Historical context

### 1.1.1 Prehistory of exoplanets

#### Speculating on other worlds

The concept that other planets exist beyond our own might not have been as apparent in the past as it is today, as one needed to realize that other stars are other suns and then imagine that planets could orbit around them. The earliest written records of this notion can be traced back to Epicurus (ca. 342 - 270 BC).

"There are an infinite number of worlds, some like this world, others unlike it."

- Epicurus - *Letter to Herodotus*

This thought can also be found in the writings of Lucretius (ca. 94 - 54 BC) in his work *De Rerum Natura*. However, Aristotle (ca. 384 - 322 BC) emerged as one of the most influential thinkers and his idea that the Earth is unique held sway. Therefore, it was not until the Renaissance that other authors challenged this concept and revisited the possibility of multiple worlds. The first one, perhaps unexpectedly, is a highly influential cardinal, Nicholas de Cusa (1401 - 1464). In his work *De docte ignorantia*, he wrote that the Universe is an infinite sphere and that the Earth does not have to be at the center, which might have inspired Copernicus (1473 - 1543). Giordano Bruno (1548 - 1600) went even further and

wrote in *The Ash Wednesday Supper* that planets orbit the Sun and that others might also orbit stars, possibly even with inhabitants. Coincidentally, he took a doctorate in theology in Toulouse and was elected by students to lecture in philosophy for two years (1580-1581). Hopefully, knowing his fate, speculating on other planets while being in Toulouse will remain the only point of comparison with this thesis.

Other authors continued to explore these ideas, mostly in works that could be considered as early science fiction. For example, Johannes Kepler (1571 - 1630) wrote *Somnium*, which describes life on the Moon. John Wilkins (1614 - 1672) published something similar in *The Discovery of a World in the Moone* and Savinien de Cyrano de Bergerac (1619 - 1655) contributed to French literature with a similar work titled *Comical History of the States and Empires of the Moon*. Even later, in his *Micromégas*, Voltaire (1694 - 1778) told the story of an extraterrestrial being visiting the planets of the Solar system.

However, all these are fiction works and René Descartes (1596 - 1650), possibly influenced by Epicurus, developed a more serious vortex theory in which a vortex is centered on the Sun and carries the planets. He used this theory to explain the formation of the solar system but probably believed it applied to other stars as well. His theory had a significant impact in Europe, and his successors such as Nicolas Bion (1652 - 1733) explicitly depicted other star systems (see figure 1.1). This theory remained popular until Isaac Newton (1643 - 1727) published his theory of universal gravitation in 1687. Newton's theory led Christiaan Huygens (1629 - 1695), who had seen Saturn's rings as evidence for the vortex theory, to change his views. However, he did not give up on the idea of the plurality of worlds, which he explored in his posthumously published work *Cosmotheoros* discussing the possibilities of life in the Universe.

After Huygens, the topic of the plurality of worlds did not advance much since no new observations were contributing to the subject. Some isolated works remained, like William Herschel (1738 - 1822) speculating on the habitability of the Sun or Camille Flammarion (1842 - 1925) who wrote *The Plurality of Inhabited Worlds*. Even though these topics fascinated the public, the era had shifted towards rationalism, and the question of the plurality of worlds was generally viewed unfavorably by astronomers due to the fact that it was difficult to demonstrate their existence.

### **Failed detection attempts**

Indeed, the 19th and early 20th centuries witnessed a proliferation of unsubstantiated claims regarding the detection of exoplanets, which dissuaded many astronomers from further exploring this field. These endeavors predominantly relied on astrometry, a technique that examines the movement of a star in the sky to identify deviations caused by an orbiting companion. While this method proved successful in the case of Bessel, who discovered the white dwarf Sirius B using astrometry (Bessel, 1844), it was beyond the observational capabilities of astronomers of that era to detect the perturbations induced by a planet. Nevertheless, this did not deter some from occasionally asserting the discovery of such companions.

The most well-known example of such false alarms involves the binary system 70 Ophiuci. In 1855, these two stars were nearing the end of their 93-year orbit around each other since their discovery by Herschel in 1779. William S. Jacob from the Madras Observatory suggested that there was an irregularity in their orbit. Representing it as an ellipse proved challenging

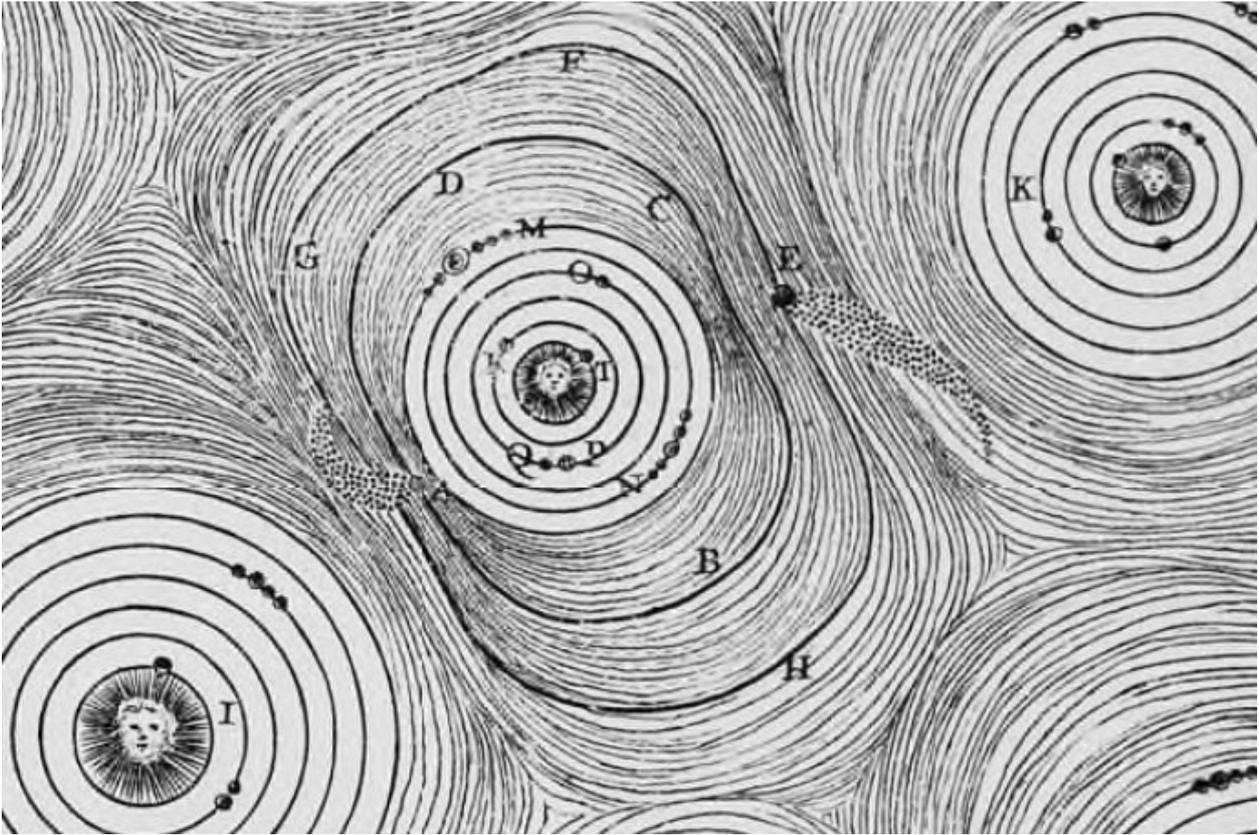


Figure 1.1: Representation of stellar systems in the vortex theory. Credit: *L'Usage des globes céleste et terrestre, et des sphères suivant les différens systèmes du monde* - Nicolas Bion. © Bibliothèque de l'Observatoire de Paris.

and he introduced a third celestial body orbiting the less massive star with a period of 26 years to better explain the observed orbit (Jacob, 1855). This claim has been proven to be incorrect (Heintz, 1988) but it deserves credit for being the initial attempt to use the astrometric method for exoplanet detection.

The series of mistaken claims for this star system does not end there. Forty years later, Thomas J.J. See revisited the measurements and confirmed the existence of the orbiting companion around the least massive star. However, he calculated that this companion would have a longer orbital period of 36 years (See, 1896). This finding was swiftly dismissed by Forest E. Moulton, who not only demonstrated that the third body was unnecessary but also showed that such a three-body system would be extremely unstable (Moulton, 1899). See, upon receiving these results showing that he had drawn an incorrect conclusion, composed a letter to the *Astronomical Journal*. Although the precise contents of the letter remain undisclosed, an attached note from the *Journal* distinctly warned of potential censorship of See's future contributions and was seen as a form of banishment from publication in this journal.

Regarding 70 Ophiuci, the story is not over. Additional claims of a third companion were made, first by Berman in 1932 (Berman, 1932) and later by Reuyl in 1943 (Reuyl & Holmberg, 1943). These claims generated considerable excitement at the time but later observa-

tions have refuted these assertions (Worth & Heintz, 1974).

Similar claims were made for other star systems, including  $\mu$  Draconis (Strand, 1943a),  $\zeta$  Aquarii (Strand, 1942),  $\xi$  Boötis (Strand, 1943b), and 61 Cygni (Strand, 1943c), all by the Danish astronomer Kaj Aa. Strand from the Sproul Observatory. He even confirmed his discovery around 61 Cygni in Strand (1957) but all those claims have now been refuted (Heintz, 1978). He was working at the Sproul Observatory alongside Peter Van de Kamp, who was actively researching Barnard’s Star (Barnard, 1916) until he reported the presence of a planetary companion (van de Kamp, 1969). This discovery would have been of great interest if it had not been disproven in later studies (Hershey, 1973; Gatewood & Eichhorn, 1973). Van de Kamp never acknowledged this as an error and continued to publish about this planetary system until the 1980s (van de Kamp, 1982). The most recent claim of a planet orbiting Barnard’s Star was made in 2018 (Ribas et al., 2018) and was also subsequently found to be incorrect (Lubin & HPF Science, 2022). This illustrates that the issue of false claims is not limited to previous astronomers.

In reality, it became evident that new methods were needed to supplant astrometry as a means of detecting exoplanets. This shift in focus had been proposed by Otto Struve as early as 1952. Struve suggested that precise measurements of stellar radial velocity could be employed to search for planets circling nearby stars. He had the remarkable insight that planets similar to Jupiter could be located as close as 0.02 astronomical units (AU) from their host stars. He also noted that if such close-in planets were ten times the mass of our Jupiter, the change in the star’s velocity due to an edge-on orbit would be about 2 kilometers per second, a measurement achievable with the Doppler precision available at his time (Struve, 1952).

However, Struve’s proposals did not gain much traction, possibly because there was a prevailing belief that other stellar systems would resemble our Solar System, making it seemingly impossible to detect planets from stellar velocity measurements. Nevertheless, over time, the necessary instruments and techniques for such detections were gradually being developed, often not initially intended for exoplanet discovery.

## 1.1.2 First detections

### Detection of a brown dwarf

Towards the end of the 1980s, there was a growing interest in brown dwarfs, failed stars that lack the mass to sustain nuclear fusion. One such brown dwarf, orbiting a star similar to our Sun (HD 114762), was discovered through velocity measurements of the star (Latham et al., 1989) and was found to have a minimum mass of 11 times that of Jupiter.

In parallel with these discoveries, NASA assembled a group of experts to formulate strategies for detecting and studying exoplanetary systems. This effort resulted in the TOPS (Towards Other Planetary Systems) report<sup>1</sup>. However, the impact of this report was limited due to unexpected findings.

---

<sup>1</sup>[https://www.stsci.edu/RAB/MemoRefs/Older/TOPS\\_Report.pdf](https://www.stsci.edu/RAB/MemoRefs/Older/TOPS_Report.pdf)

### First detection of exoplanets around a pulsar

Since the 1970s, radio astronomers have been studying the radio signals emitted by pulsars, which are rapidly rotating stars in the final stages of their lives. These stars emit radio signals at highly stable intervals. If these pulsars are accompanied by planets, the presence of those planets should disrupt this signal.

In 1992, Wolszczan detected such disruptions and announced the discovery of two exoplanets orbiting the pulsar PSR B1257+12 (Wolszczan & Frail, 1992). This discovery was particularly surprising since the presence of exoplanets around a pulsar was unexpected.

### First detection of an exoplanet around a Sun-like star

In the 1990s, astronomers turned to velocimetry to attempt the detection of exoplanets around stars similar to our Sun. We will delve into the details of this method later, but in brief, it involves observing how the presence of a planet affects the velocity of the star it orbits, where astrometry focuses on the position itself. The significant advantage of this method compared to astrometry is its highest capability to detect exoplanets. For instance, in our Solar system, Jupiter induces a variation in the Sun's velocity of approximately  $12.5 \text{ m.s}^{-1}$ , within the reach of available instruments. However, it is important to note that this approach requires long-term observations since, given Jupiter's orbital period of 11.86 years, detecting Jupiter around our Sun would necessitate observations spanning at least 12 years.

As a result, astronomers face the challenge of being theoretically capable of detecting Jupiter-like planets but requiring years of observations to do so. Consequently, they armed themselves with patience and started long-term programs. However, it was during these efforts that they made a serendipitous discovery. While astronomers initially anticipated planetary systems resembling our own, with the exception of Struve's earlier prediction (Struve, 1952), it turns out that there are numerous giant planets located very close to their host stars. This surprising fact paved the way for the first discovery of an exoplanet around a Sun-like star, since it can be shown that the magnitude of the effect of a planet on the star's velocity is most pronounced when dealing with massive and/or closely orbiting bodies, leading to easier detections.

In 1995, Nobel laureates Michel Mayor and Didier Queloz made an announcement that marked a significant milestone: they had discovered the first exoplanet, 51 Pegasi b (Mayor & Queloz, 1995). This discovery was made possible with the ELODIE spectrograph (Baranne et al., 1996) mounted at the Observatoire de Haute-Provence and this planet has roughly half the mass of Jupiter and completes its orbit around its parent star in just 4.2 days. This discovery not only confirmed that our solar system is not unique in the Universe but also highlighted the remarkable diversity of exoplanetary systems, emphasizing that our own system is by no means a universal model.

Furthermore, the presence of such a planet challenged existing theories of planetary formation, leading astronomers to revise their models to include the concept of orbital migration (Chambers, 2009; Morbidelli, 2014; Bitsch et al., 2014; Baruteau et al., 2014). It is likely that this planet initially formed at a greater distance from its host star and then moved inward due to interactions with the protoplanetary disk, a scenario that Mayor and Queloz had

already suggested in their original paper (Mayor & Queloz, 1995). This important discovery was soon confirmed (Marcy & Butler, 1995), and other similar findings followed with planets detected around 47 Ursae Majoris (Butler et al., 1996) and 70 Virginis (Marcy & Butler, 1996). These planets had masses about 2.4 and 6.6 times that of Jupiter, respectively, and were located at distances of approximately 0.43 and 2.1 au from their host stars. Once more, those two exoplanets highlight that our Solar system is not the standard model.

The method of velocimetry dominated exoplanet detection until the end of the 20th century, leading to the implementation of systematic research efforts using dedicated spectrographs like ELODIE (Baranne et al., 1996), HIRES<sup>2</sup> (Vogt et al., 1994) and HARPS<sup>3</sup> (Mayor et al., 2003). These increasingly precise instruments facilitated the detection of less massive planets over time, as illustrated in figure 1.2. As the technique improved, it eventually led to the discovery of the first multi-planetary system orbiting around  $\nu$  Andromedae (Butler et al., 1999).

In the same year, the transit method emerged as an other promising technique. This method uses photometric data to detect changes in the brightness of a star when a planet passes in front of it. It demonstrated its effectiveness with the discovery of the first transiting planet orbiting HD 209458 (Charbonneau et al., 2000).

## 1.2 On the detection of exoplanets

As of the writing of this thesis, 5,502 exoplanets have been discovered and confirmed. A range of methods has been developed for their detection and many of the existing techniques primarily rely on observing the effects an exoplanet has on its parent star, making them indirect methods. Directly detecting an exoplanet is a challenging endeavor, but recent technological advancements have made it possible. Two indirect methods dominate the field: the transit method, which has resulted in 4,113 detections, and the radial velocity method, with 1,056 detections. In the following sections, we will delve into greater detail about the latter method, as it is the one employed in this thesis. Nonetheless, we will still provide an overview of the transit and other detection methods, albeit with less detail, in appendix A.

### 1.2.1 Radial velocity

#### General principle

First and foremost, we must derive the radial velocity (RV) equation, which establishes the relationship between an object on its orbit and its radial velocity. In the case of a two-body system, this relationship can be derived using Newton's and Kepler's laws of planetary motion. A straightforward calculation (Lovis & Fischer, 2010) gives that the radial velocity equation is:

$$RV(t) = K_s [\cos(\omega + \nu(t)) + e \cos \omega],$$
$$\text{with } K_s = \left(\frac{2\pi G}{P}\right)^{1/3} \frac{M_p \sin i}{(M_s + M_p)^{2/3} \sqrt{1 - e^2}}, \quad (1.1)$$

---

<sup>2</sup>High Resolution Echelle Spectrometer

<sup>3</sup>High Accuracy Radial Velocity Planet Searcher

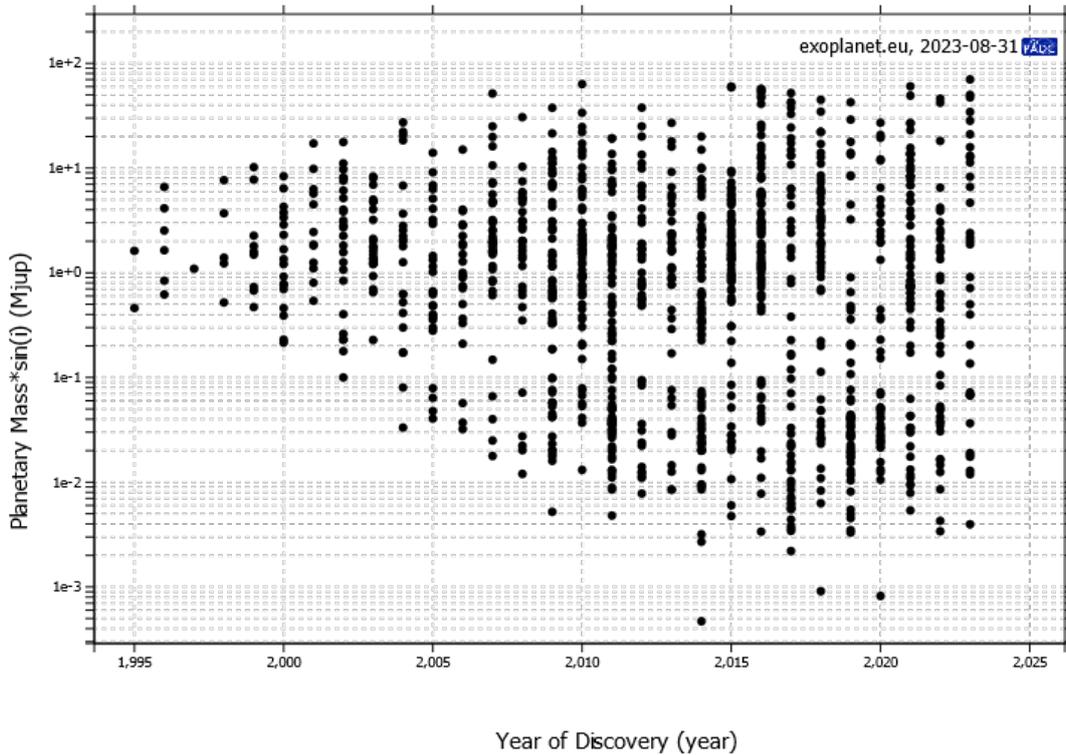


Figure 1.2: Minimum planetary mass confirmed exoplanets over the years. Credit: <http://exoplanet.eu/>

where  $\omega$  represents the argument of periapsis of the stellar orbit,  $e$  is its eccentricity,  $\nu$  denotes the true anomaly (which depends on the orbital period  $P$ , phase, and eccentricity of the planet), and  $K_s$  corresponds to the semi-amplitude of the planet-induced signal. The latter depends on the masses of both the star,  $M_s$ , and the planet,  $M_p$ . Figure 1.3 presents an illustration of a Keplerian orbit, providing a clearer visualization of the orbital parameters.

A crucial aspect of this equation is that the radial velocity is determined by projecting the object's velocity onto the line of sight. Consequently, the semi-amplitude  $K_s$  also depends on the inclination  $i$  (figure 1.3). This degeneracy is fundamental since it means that the RV method can only provide a lower limit for the mass of the orbiting planet. Nevertheless, since inclination angles can be approximated as randomly distributed, the average value of  $\sin i$  is  $\pi/4$  (approximately 0.79). Therefore, this limitation does not significantly impact statistical studies of exoplanets.

However, on certain occasions, a significant discrepancy can be noted. For instance, in the case of HD 154345b, its  $M \sin i$  is approximately one Jupiter mass (Boisse et al., 2012b), while its true mass is ten times greater, about 12 Jupiter masses (Kiefer et al., 2021). Therefore, it is advisable to keep this limitation in mind when evaluating systems detected using the RV method. Table 1.1 provides several examples of RV magnitudes that can be expected for typical mass and semi-major axis values of a planet orbiting a solar-mass star.

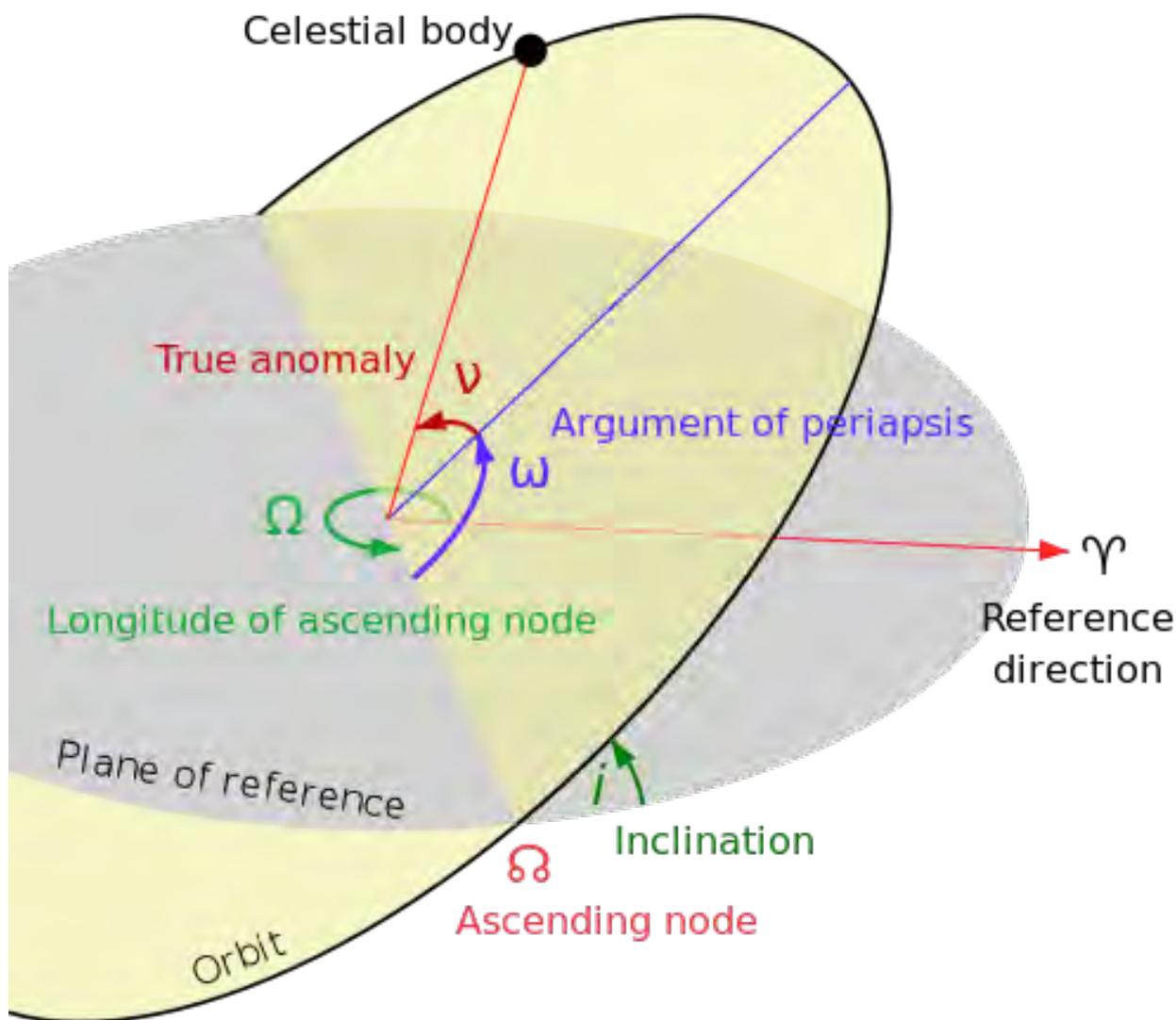


Figure 1.3: Illustration of a planet orbiting a star showing the relevant planetary parameters. Credit: [Wikimedia \(2007\)](#)

### Doppler effect

The fundamental effect upon which the RV technique relies is the Doppler effect. When observing a star in motion along its orbit in relation to an observer on Earth, the electromagnetic spectrum emitted by this star at a wavelength  $\lambda_0$  is observed at a shifted wavelength  $\lambda_{obs}$ . This shift is determined by the Doppler effect equation:

$$V_r = c \frac{\lambda_{obs} - \lambda_0}{\lambda_0}, \quad (1.2)$$

where  $c$  is the speed of light in the vacuum. This equation is an approximation adapted to our case, where we are concerned with low-amplitude RV variations, justifying the neglect of relativistic effects.

Parameters	a (AU)	K (m.s <sup>-1</sup> )
Jupiter	0.1	89.8
Jupiter	1.0	28.4
Jupiter	5.0	12.7
Neptune	0.1	4.8
Neptune	1.0	1.5
Super-Earth (5 M <sub>⊕</sub> )	0.1	1.4
Super-Earth (5 M <sub>⊕</sub> )	1.0	0.45
Earth	0.1	0.28
Earth	1.0	0.09

Table 1.1: Radial velocity signals for different kinds of planets orbiting a solar-mass star. Credit: [Lovis & Fischer \(2010\)](#).

For instance, Jupiter induces a semi-amplitude of 12. m.s<sup>-1</sup> to the Sun which would result in a Doppler shift of approximately  $2.2 \times 10^{-5}$  nm in the optical at  $\lambda_0 = 550$  nm. Dedicated RV spectrographs are therefore necessary to deliver high-precision measurements of the temporal variation of the stellar lines.

### High-precision spectrographs

Such velocimetric measurements are conducted using échelle spectrographs, which can record thousands of lines at a high resolving power  $R = \lambda/\Delta\lambda$  ( $R = 115,000$  in the case of HARPS; [Mayor et al. \(2003\)](#)). These lines are captured by a detector, typically a CCD in the optical or HAWAII 4RG<sup>TM</sup> in the nIR ([Artigau et al., 2018](#)). To enhance the signal-to-noise ratio (S/N) and, consequently, the precision of the RV measurement, a list of stellar lines covering the spectral range of the instrument is used. Traditionally, the cross-correlation (CCF) method ([Mayor & Queloz, 1995](#); [Baranne et al., 1996](#)) is employed to combine these stellar lines and obtain an average line profile for retrieving the RV ([Pepe & Lovis, 2008](#)).

Measuring RV at the necessary precision level to detect Earth-sized planets is a formidable instrumental challenge. Any factor causing a shift in the spectral lines can mimic a Doppler shift effect. For instance, a mere 0.005 nm shift is equivalent to roughly 0.3 m.s<sup>-1</sup> in the optical ([Pepe & Lovis, 2008](#)).

The primary limitation in RV precision is the fundamental constraint imposed by photon noise. This constraint depends on several factors, including the star being observed, the signal-to-noise ratio (S/N) of the observation, the spectral range of the spectrograph, and, of course, the resolving power of the instrument ([Bouchy et al., 2001](#)). A resolving power exceeding 50,000 is necessary to resolve spectral lines when observing a slowly-rotating solar-type star. However, the rotation of the star can broaden the spectral lines which can increase the photon noise significantly. For instance, the photon noise can increase by approximately a factor of  $\sim 32$  between two Sun-like stars with projected rotational velocities of 2 and 20 km.s<sup>-1</sup> ([Lovis & Fischer, 2010](#)).

The second major prerequisite for high-precision RV measurements is the availability of a stable wavelength reference against which the positions of spectral lines can be compared.

Two methods exist:

The first method is the self-calibration technique (Butler et al., 1996), where the incident stellar beam passes through a reference gas absorption cell. However, this approach comes at the cost of significantly reducing the incident flux absorbed by the gas cell (Vogt et al., 1994).

The second method is the simultaneous reference technique, which requires a specific design of the spectrograph to be implemented and was the first one to be used successfully for exoplanet detection with ELODIE (Baranne et al., 1996). In this technique, the stellar light is recorded simultaneously with the light produced by a reference lamp. This reference lamp can be a Fabry-Pérot (FP) etalon combined with a broadband light source such as a U/Ne hollow-cathode lamp for SPIRou (Hobson et al., 2021) or a laser frequency comb (Phillips et al., 2012). The feasibility of this method presupposes that both the science and calibration channels are affected in the same way, which is only the case if the spectrograph is highly stabilized in temperature and pressure (typically below 0.01 K and 0.01 mbar for instruments like HARPS (Mayor et al., 2003)).

All these requirements may appear limiting, but this second method has actually delivered the highest RV precision to date, achieving approximately  $0.1 \text{ m.s}^{-1}$  for instruments like ESPRESSO<sup>4</sup> (Pepe et al., 2021). As a result, this method has become the preferred choice for the next generation of spectrographs, including EXPRES<sup>5</sup> (Jurgenson et al., 2016).

Numerous other instrumental challenges should be noted, with a comprehensive list available in works from Pepe & Lovis (2008) and Lovis & Fischer (2010). These challenges encompass variations in the index of refraction of the air, thermal and mechanical effects, detector-related issues, among others. Some effects can be specific to the instrument in use; for instance, in the case of HARPS, the deformation of spectral lines crossing block stitchings of the CCD resulted in a one-year signal in the data (Dumusque et al., 2015). Additionally, the spectral domain under observation can introduce its own set of challenges. In the nIR, issues such as persistence of infrared arrays are notable. This phenomenon is characterized by the fact that not only is the preceding frame of significance, but the entire illumination history spanning the last few hours holds importance. To illustrate, if bright targets are observed at the beginning of a night, this can have an impact on the observation of fainter targets later in the same night (Artigau et al., 2018).

Another potential source of contamination is related to the measurements being conducted in the geocentric frame. Corrections must be made for the barycentric Earth radial velocity (BERV), which can introduce additional errors into the RV measurements. Additionally, there can be spectrum contamination from various sources that may impact RV precision. One example is the presence of telluric lines from Earth's atmosphere, which can introduce spurious one-year signals in the data. While the visible region is not severely affected, it is known to be a challenge for achieving precise nIR RV measurements (Bean et al., 2010; Cook et al., 2022). Other significant sources of contamination include moonlight and background objects.

Finally, the RV extraction method itself can introduce errors into the RV measurements.

---

<sup>4</sup>Echelle Spectrograph for Rocky Exoplanet and Stable Spectroscopic Observations

<sup>5</sup>The EXtreme PREcision Spectrometer

For example, the CCF method relies on a predetermined set of spectral lines, which may not be well-suited for every star (Mayor & Queloz, 1995). This is why, in recent years, new RV extraction methods have been developed to address this issue such as template matching approaches (Anglada-Escudé & Butler, 2012; Astudillo-Defru et al., 2017b; Bedell et al., 2019a; Silva et al., 2022) or the line-by-line (Dumusque, 2018; Artigau et al., 2022).

## 1.3 The era of exoplanets

### 1.3.1 A variety of worlds

The collaborative efforts of astronomers spanning the last three decades, following the first exoplanet detection, have led to the discovery of a diverse range of exoplanets (as depicted in figure 1.4). This extensive catalog of exoplanets has underscored the fact that the structure of our Solar System is not typical of other stellar systems. Moreover, these discoveries provide a valuable opportunity to compare theoretical models of planet formation and evolution with observed exoplanetary systems (Mordasini, 2018).

Figure 1.4 also illustrates the mass-period distribution of confirmed exoplanets, revealing distinct clusters within the exoplanetary population. The category of most massive exoplanets found in close proximity to their host stars is the hot Jupiters. Their frequent appearance in the diagram is partly due to their easier detectability, as methods like transit photometry or radial velocity measurements are biased towards this class of planets. However, it is essential to note that their actual occurrence rate around Sun-like stars is relatively low, estimated to be in the range of 0.4-1% (Wang et al., 2015; Santerne et al., 2016).

On the other hand, warm and cold Jupiters, which are positioned at greater distances from their host stars, are primarily detected through the radial velocity method, given the reduced likelihood of transit events. Their occurrence rates are notably higher, estimated to be around 6% for Jupiter-mass planets orbiting within 7 astronomical units (AU) of their host stars (Santerne et al., 2016), and possibly as high as 50% for more distant orbits (Bryan et al., 2016).

Super-Earths and Neptunes are also abundant in the population of exoplanets orbiting FGK-type stars, although absent in the Solar system, boasting an occurrence rate in the range of 40-50% (Mayor et al., 2011; Howard et al., 2012; Bashi et al., 2020). These types of planets are detectable using both the transit and radial velocity methods. They are often found in compact planetary systems where orbits can be either far from or close to resonance. The exact reasons behind this observation are still not fully understood and remain a topic of ongoing investigation (Baruteau et al., 2016).

An intriguing feature observed in the mass-period diagram is the presence of a triangular region where very few Neptunes have been detected, known as the "hot Neptune desert." This region encompasses exoplanets with masses similar to Neptune and orbital periods below 4 days (Mazeh et al., 2016; Dong et al., 2018). The upper boundary of this desert is typically attributed to the fact that a gas giant requires a significant amount of mass to circularize its orbit around the parent star without being disrupted by tidal forces. The lower boundary, on the other hand, is often explained by the photoevaporation of the gaseous envelope, primarily composed of helium and hydrogen, due to intense stellar radiation (Owen & Lai, 2018).

Another noteworthy feature observed in the radius-period distribution in figure 1.4 is a gap that appears for close-in exoplanets with orbital periods below 100 days and radii ranging from 1.5 to 2 times that of Earth. This gap is commonly referred to as the "evaporation valley" or alternatively the "Fulton gap," named after the first author who observed it in [Fulton et al. \(2017\)](#). However, this gap was already predicted by simulations of the photoevaporation process affecting the atmospheres of small, close-in planets ([Lopez & Fortney, 2013](#); [Owen & Wu, 2013](#); [Mordasini, 2020](#)).

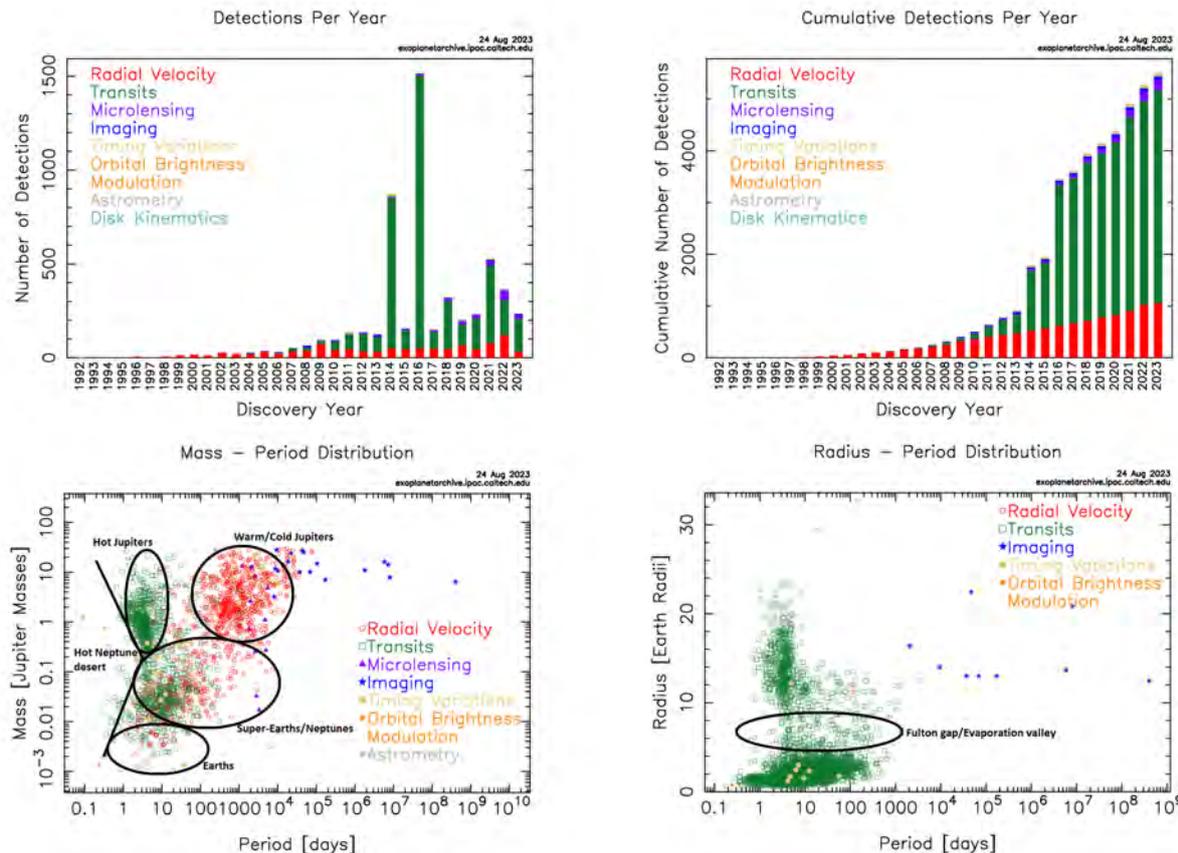


Figure 1.4: Various plots showing the diverse variety of exoplanets. Top-left panel: Detections of exoplanets per year. Top-right panel: Cumulative detections per year. Bottom-left panel: Mass-Period distribution. Bottom-right panel: Radius-Period distribution. Computed from <https://exoplanetarchive.ipac.caltech.edu/>

### 1.3.2 Searching for habitable planets

The search for extraterrestrial life is a significant motivation in the quest for exoplanets. However, assessing a planet's habitability is a complex task that involves multiple parameters, as illustrated in figure 1.5. An extensive review of the conditions required for a planet to potentially support life is presented in [Cockell et al. \(2016\)](#). It is important to note that our understanding of life is currently based on a statistical sample of one—Earth, which is limiting. In general, it is postulated that a planet must possess certain attributes, including carbon, liquid water, and a source of energy, to be considered habitable. The acquisition

of water and carbon is typically not too challenging, as these elements are often abundant beyond the snow line, where they can be accreted by planetary embryos (Robertson & Mahadevan, 2014).

Nevertheless, for a planet to support liquid water at its surface, it must receive an adequate amount of stellar flux through its atmosphere. This concept led to the definition of the habitable zone (HZ; Kasting et al. (1993)), a region of radial distances from the star within which planets could potentially maintain liquid water on their surfaces. The extent of the HZ is influenced by various factors, including the stellar flux received by the planet, which is contingent on the effective temperature of the star. Additionally, the HZ is influenced by factors such as the composition of a planet’s atmosphere (if it has one) and the presence of a sufficiently strong planet magnetosphere to shield the planet atmosphere from harmful stellar radiation.

The inner boundary of the HZ is typically defined as the distance where a planet would reflect all incoming stellar flux, meaning a completely opaque atmosphere. This condition would result in the planet’s surface being cool enough to sustain liquid water, and it is referred to as the runaway greenhouse limit (Kopparapu et al., 2014). Conversely, the outer boundary of the HZ represents the maximum greenhouse limit, where a planet possesses an atmosphere primarily composed of greenhouse gases. This atmospheric composition enables the planet to maintain temperatures suitable for liquid water even at greater distances from its host star (Kopparapu et al., 2014).

As of today, there are 63 exoplanets that are considered potentially habitable, and in figure 1.6, we showcase a variety of these systems. Some well-known examples include TRAPPIST-1 e, f & g (Gillon et al., 2017) and Proxima Centauri b (Anglada-Escudé et al., 2016). Proxima Centauri b, in particular, has garnered significant attention due to its status as the closest habitable exoplanet to Earth, making its habitability a major area of interest (Barnes et al., 2016; Meadows et al., 2018). One could also mention Kepler-22b (Borucki et al., 2012) and Kepler-1638b (Morton et al., 2016) as examples of transiting planets in the habitable zone of Sun-like stars.

The list of habitable exoplanets is continuously evolving. New discoveries, such as GJ 1002 b and GJ 1002 c with ESPRESSO (Suárez Mascareño et al., 2023) or Wolf 1069 b with CARMENES (Kossakowski et al., 2023), are continually expanding the list. Conversely, some candidates previously considered habitable have been inquired or reevaluated, leading to their exclusion from this category. Examples of such revisions include GJ 581 d & g (Robertson et al., 2014), GJ 667C e (Robertson & Mahadevan, 2014), GJ 832 c (Gorini et al., 2022), and Kapteyn b (Bortle et al., 2021). Additionally, habitable planets near the inner or outer boundaries of the HZ may be reclassified due to updates in models. For instance, Kepler-69 c is now considered a super-Venus (Kane et al., 2013).

Overall, these systems represent primary targets for the search for biosignatures, which are signals of potential biological origins within their atmospheres. It is important to keep in mind, however, that unknown abiotic mechanisms could also produce similar signatures. A comprehensive review of biosignatures can be found in Schwieterman et al. (2018). Biosignatures often revolve around detecting thermodynamic anomalies, such as the presence of molecules that should dissipate rapidly in the atmosphere. For example, dioxygen, which transforms into ozone through photo-dissociation but is replenished through photosynthesis,

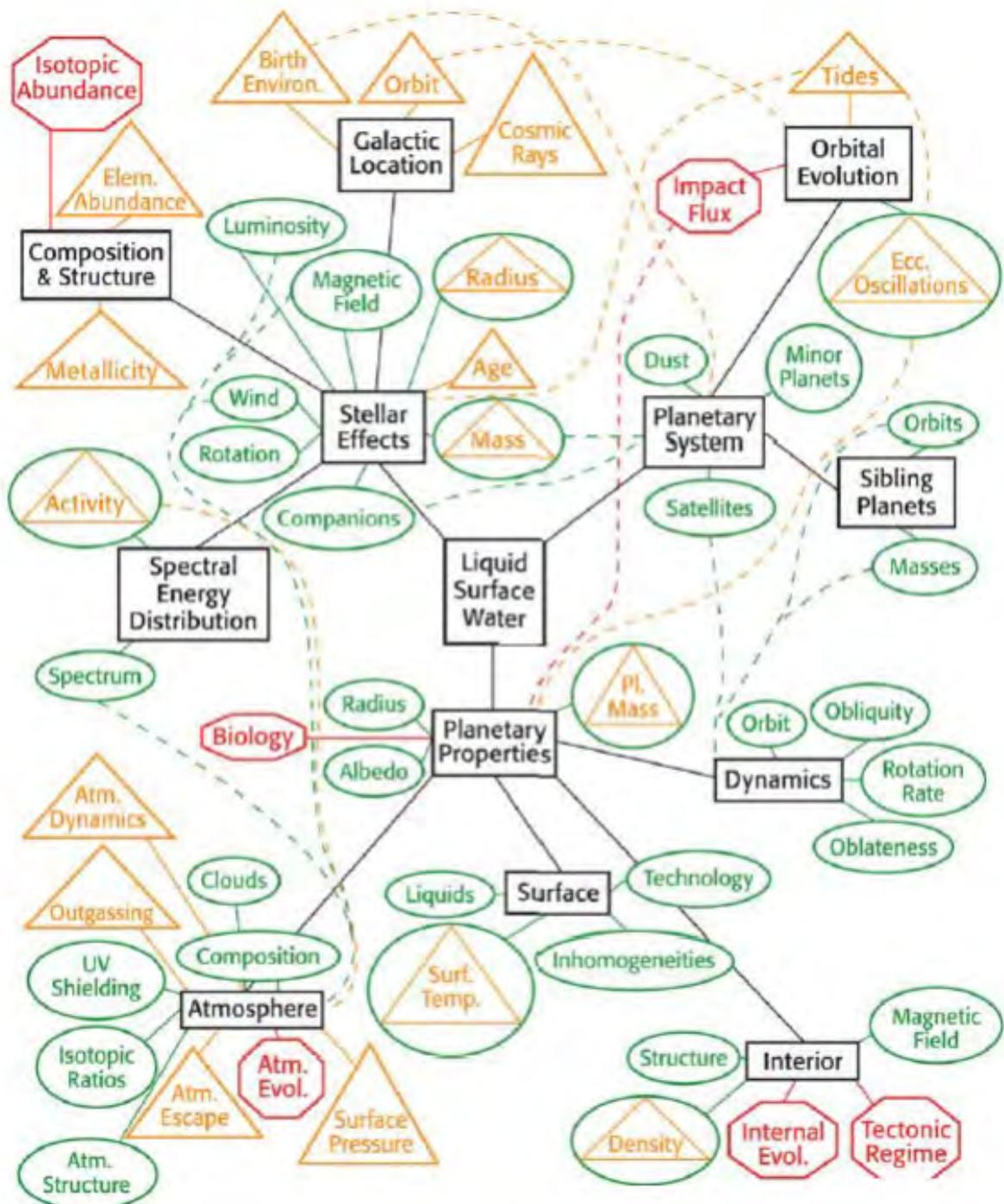


Figure 1.5: The habitability of an exoplanet as a multi-parameter problem. Credit: <https://www.cosmos.esa.int/>.

is one such indicator. Methane is another biosignature as it is unstable in the atmosphere and requires the presence of life to be sustained (Thompson et al., 2022).

It is important to note that relying solely on a single biosignature is not sufficient, as false

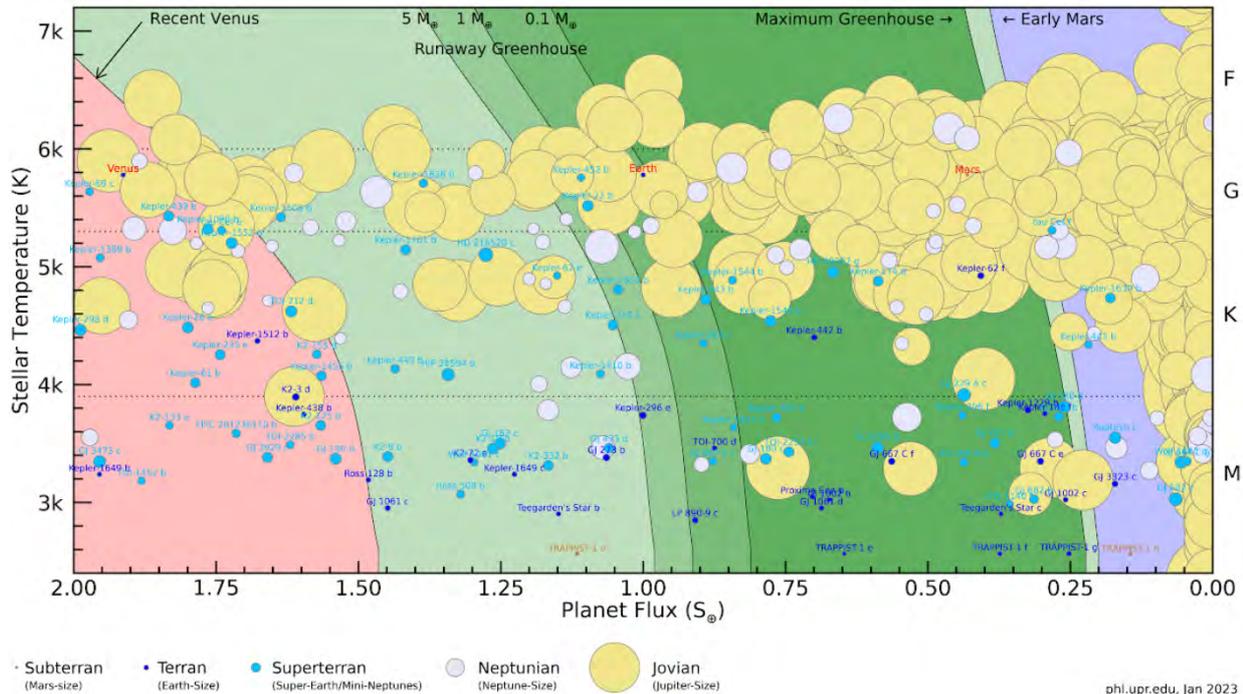


Figure 1.6: All known planets near the habitable zone (the darker green shade is the conservative habitable zone and the lighter green shade is the optimistic habitable zone). Only those planets with less than 10 Earth masses or 2.5 Earth radii are labeled. The different limits of the habitable zone are described in [Kopparapu et al. \(2014\)](#). The size of the circles corresponds to the radius of the planets. They are estimated by the mass-radius relationship of [Chen & Kipping \(2017\)](#) when not available Credit: PHL UPR Arcibo.

positives can occur such as the recent detection of phosphine on Venus ([Clements, 2023](#)). Determining the presence of life on a planet requires a holistic approach that considers the planet's overall context, including its geochemistry and interactions with its host star. Additionally, rocky exoplanets' atmospheres may hold surprises, and we must remain open to unexpected abiotic processes that could produce biosignatures.

The JWST ([Gardner et al., 2006](#)), launched in December 2022, is anticipated to have the capability to detect potential biosignatures in the atmospheres of terrestrial planets that transit very-low-mass stars within 10 parsecs of the Sun ([Morley et al., 2017](#)). However, not all habitable exoplanets meet the criteria for JWST characterization. Among those that do are TRAPPIST-1 e, f & g, as well as LHS 1140 b ([Dittmann et al., 2017](#)), and Luyten b ([Astudillo-Defru et al., 2017a](#)). Looking ahead, collaborative efforts involving high-resolution spectrometers and high-contrast imaging techniques may further enhance our ability to detect potential biosignatures ([Snellen et al., 2015](#); [Mennesson et al., 2016](#)).

Therefore, the discovery of new nearby and temperate Earth-like planets is imperative for upcoming telescopes focused on the search for biosignatures. For this purpose, a spectrographs with ever-increasing precision have emerged, this includes instruments like HARPS ([Mayor et al., 2003](#)) and its twin HARPS-N ([Cosentino et al., 2012](#)), SOPHIE ([Perruchot et al., 2008](#)), and HIRES ([Vogt et al., 1994](#)). Optical spectrographs are now aiming to achieve precisions as high as  $10 \text{ cm.s}^{-1}$ , exemplified by ESPRESSO ([Pepe et al., 2021](#)) and EXPRES

(Jurgenson et al., 2016). Additionally, spectrographs designed for nIR observations have begun to emerge, targeting an RV precision goal of  $1 \text{ m.s}^{-1}$ . These nIR spectrographs are primarily intended for observations of M dwarf stars such as CARMENES-NIR (Quirrenbach et al., 2018), SPIRou (Donati et al., 2020) or NIRPS (Thibault et al., 2022).

## 1.4 The hunt for planets around M dwarfs

While the field of exoplanets initially focused on the study of solar-type stars, its attention has increasingly shifted toward other types of stars, notably M dwarfs ( $0.08\text{-}0.6 M_{\odot}$ ). This shift was prompted by new discoveries from missions like Kepler and early RV surveys, which demonstrated that M dwarfs host a significantly larger number of rocky planets compared to other stellar types (Bonfils et al., 2013; Dressing & Charbonneau, 2015; Gaidos et al., 2016; Hsu et al., 2020; Sabotta et al., 2021). Moreover, M dwarfs are abundantly distributed in our solar neighborhood, constituting at least 75% of stars within 25 parsecs of the Sun (Henry et al., 2006), and possibly up to 90% within 10 parsecs (Winters et al., 2019).

Additionally, their lower masses make them prime candidates for detecting planet-induced RV signals. As evident in Equation 1.1, for an M dwarf with a mass of  $0.1 M_{\odot}$ , all values in table 1.1 can be multiplied by 3.2. Considering that the HZ for these stars is typically located at about 0.1 AU, terrestrial planets in the HZ can induce RV signals on the order of  $1 \text{ m.s}^{-1}$ . This significantly aids in the detection of low-mass planets around M dwarfs (e.g. Anglada-Escudé et al. (2016); Faria et al. (2022); Suárez Mascareño et al. (2023)).

### 1.4.1 Habitability around M dwarfs

Shifting towards M dwarfs aims at detecting more habitable planets in the neighborhood of the Sun, facilitating their atmospheric characterization with JWST or the ELT. However the issue of whether planets orbiting M dwarf stars can provide habitable environments presents a lot of challenges (Ridgway et al., 2023). Several factors could potentially hinder habitability, including a high probability of tidal locking and significant stellar variability (Childs et al., 2022).

Previous assumptions suggested that habitable conditions on planets within the HZ of M dwarfs would be confined to a specific ring-like region at the surface of the planet known as the terminator due to tidal locking (Tarter et al., 2007). It was also thought that tidal locking would lead to intense winds and constant heavy rainfall on the star-facing side of these planets, causing extreme climate disparities (Joshi et al., 1997). However, recent climate models have shown that tidally locked planets could employ mechanisms to mitigate temperature differences between their day and night sides, potentially enhancing their habitability (Merlis & Schneider, 2010; Yang et al., 2013).

Concerning stellar variability in M dwarfs, the occurrence of flares raised concerns about their impact on exoplanetary atmospheres, as flares emit charged particles that might erode the atmosphere (Guinan & Engle, 2013). Active M dwarfs capable of emitting coronal mass ejections (CMEs) could distort a planet's magnetosphere, potentially causing atmospheric erosion and rendering the planet uninhabitable (Khodachenko et al., 2007; Garcia-Sage et al., 2017). Nevertheless, recent research by Vida et al. (2019) suggests that some later-type active

M dwarfs might have a less detrimental impact on exoplanetary systems in terms of CMEs than previously believed.

Furthermore, the presence of a magnetosphere on an exoplanet could potentially deflect particles away from its atmosphere (Alpert, 2005). However, even with the strongest possible dynamo-generated magnetic fields, exoplanets with Earth-like masses would still experience significant atmospheric loss due to erosion caused by CME bursts and extreme ultraviolet (XUV) emissions (Zuluaga et al., 2012; See et al., 2014). Observations of the TRAPPIST-1 exoplanetary system suggest that some planets in this system lack substantial atmospheres, indicating that they are either bare rock planets or possess very thin atmospheres (Zieba et al., 2023). On a more optimistic note, significant flares tend to occur at higher latitudes near the poles of the stellar host, suggesting that exoplanets with orbits aligned with the stellar rotation axis might experience less impact from these flares (Ilin et al., 2021).

### 1.4.2 Stellar activity and its impact on RV time series

The magnetic activity of M dwarfs not only affects the potential habitability of exoplanets but also influences RV measurements. M dwarfs possess magnetic fields with strengths varying from a few tens of G to kG for fully-convective stars (Donati et al., 2006; Morin et al., 2008, 2010). These magnetic fields induce various phenomena collectively referred to as magnetic activity, which are observable on the surface, in the atmosphere, and in the corona of stars. Due to their low effective temperature and large convective zones, M dwarfs exhibit higher activity levels compared to G and K stars. This heightened activity is a concern since it can produce RV signals related to various phenomena.

#### Spots, faculae and plages

Magnetic field tubes can emerge above the stellar photosphere, affecting the star's appearance (Hale, 1908). When these tubes have strong magnetic fields (around a few kG), they suppress convective motion, causing the affected regions to cool and appear darker than the surrounding photosphere, referred to as spots (Strassmeier, 2009). On the other hand, when field tubes are narrower, radiative heating dominates, making these regions brighter than the photosphere, known as faculae. These emerging field tubes also heat the chromospheric plasma, leading to the formation of small, bright areas called plages within the chromosphere through mechanisms not completely understood (Schrijver et al., 2002).

The presence of these dark and bright regions, driven by stellar rotation, can mimic exoplanet signatures with a timescale that corresponds to the star's rotational period and its harmonics (Saar & Donahue, 1997; Queloz et al., 2001; Robertson et al., 2014). The main reason is that when a spot is on half of a star approaching the observer, it causes a global redshift due to a lack of blue-shifted light. Conversely, as it passes on the receding half of the star's surface, it creates a blueshift. The RV signature of spots and plage was studied through simulations (Boisse et al., 2012a; Dumusque et al., 2014a) and observations (Haywood et al., 2016; Collier Cameron et al., 2019) and is known to cause RV signals at the  $\text{m.s}^{-1}$  level in the case of the Sun or similarly active Sun-like stars.

### Long term activity

The long-term evolution of the magnetic field influences the distribution of surface irregularities such as spots and plages. They appear and disappear randomly on the surface of the star and are not stationary with respect to the star's surface. In the case of the Sun, sunspots tend to appear at approximately 30-35 degrees during periods of minimum activity and migrate downwards with an 11-year periodicity. This long-term evolution of the magnetic field, which affects the number and distribution of sunspots, leads to corresponding long-term RV variations (Dumusque et al., 2011). These variations can reach up to  $25 \text{ m.s}^{-1}$  on timescales corresponding to the magnetic cycles, which can range from a few to several years (Lovis et al., 2011). Similar activity cycles have also been reported around M dwarf stars (Gomes da Silva et al. (2012); Route (2016); Mignon et al. (2023) and Lehman et al. (2023), submitted).

### Other sources of stellar activity

As mentioned earlier, M dwarfs are known for their occasional flares and CMEs which could make the surface of exoplanets not suitable for life due to atmosphere erosion (Khodachenko et al., 2007). They can also generate RV signals reaching tens of meters per second, which can complicate the detection of exoplanets (Reiners, 2009). Those events are relatively frequent for M dwarfs which can be separated by only a few days (Mondrik et al., 2019; Glazier et al., 2020; Günther et al., 2020).

There are other sources of stellar activity in RV signals that are not related to the magnetic field, such as acoustic oscillations. Acoustic oscillations are the propagation of acoustic waves in the convective zone of stars (Leighton et al., 1962). These waves can cause radial oscillations in Sun-like stars and produce RV variations of a few  $\text{m.s}^{-1}$  on timescales of a few minutes (Bouchy & Carrier, 2002). However, the amplitude of these oscillations scales with the ratio of stellar luminosity to stellar mass (Christensen-Dalsgaard, 2004), leading to a significant decrease in this effect for M dwarf stars. As a result, these oscillations remain undetected in the RV signals of M dwarfs (Rodriguez-Lopez et al., 2015).

Another effect to consider is granulation. In the convective zone of a star, hot gas rises until it reaches the photosphere, where it cools down and moves back towards the star's center. This means that the gas moving toward the observer has, on average, a higher temperature than the gas moving away, resulting in an overall blueshift of the star's surface. The timescale of this effect can range from a few minutes for small granules to a few days for supergranules (Del Moro, 2004). Stochastic variations in the granulation process can induce RV fluctuations at the  $\text{m/s}$  level for solar-like stars (Meunier et al., 2017). However, the amplitude of these fluctuations scales with the velocity at which convection operates, and this velocity is significantly lower for M dwarf stars, making RV perturbations from granulation less impactful (Baroch et al., 2020).

Additionally, while our primary focus has been on the impact of stellar activity on RV measurements and its effect on exoplanet detection since RVs are the main focus of this work, it is important to add that other detection methods can also be influenced by stellar activity. For instance, spots on a star's surface can mimic the signal of a transiting exoplanet, and flares can disrupt transit detections. In the following section, we will discuss various methods for mitigating the effects of stellar activity in the context of RV data.

### 1.4.3 Mitigating stellar activity

Filtering out stellar activity-induced RV signals represents one of the primary challenges encountered in high-precision velocimetry over recent decades. To address this challenge effectively, the initial step involves the detection and characterization of stellar activity. This task is often conducted through the use of stellar activity indicators.

A natural proxy for stellar activity can be derived from stellar photometric curves, which are also influenced by surface inhomogeneities. All the more so as if contemporaneous measurements are available (Aigrain, 2012; Haywood et al., 2014). Another effective activity indicator is to directly probe the magnetic field and the longitudinal magnetic field, measured from circularly-polarized Zeeman signatures, which has been demonstrated to be a reliable indicator of the rotation period of active stars (Hébrard et al., 2016; Donati et al., 2023a; Fouqué et al., 2023; Donati et al., 2023b).

Other activity indicators can be derived from the data products of the RV extraction process itself. For instance, in the case of CCF extraction, we can analyze the width and depth of the CCF using parameters like the Full Width at Half Maximum (FWHM) and contrast. This is because the shape of the average line can be influenced by stellar activity (Queloz et al., 2001). Another method to assess the shape of the average line is the Bisector Inverse Slope (BIS), which measures the velocity span between the top and the bottom of the CCF. RV techniques like template matching, such as the SERVAL method, provide access to other stellar activity indicators, the differential line width (dLW) and the chromatic velocity slope (Zechmeister et al., 2018). The dLW indicator can be related to the FWHM, while the chromatic velocity slope aims to identify signals that exhibit chromaticity, which is typically not the case for planetary signals.

In the optical domain, several well-known spectral lines correlated with stellar activity are available for analysis, such as Ca II H & K, He I D3, Na I D1 & D2, and H $\alpha$  (Boisse et al., 2009; Gomes da Silva et al., 2012, 2021). However, in the nIR, the identification and characterization of spectral lines related to activity are still areas of active research. Currently, some of the lines under investigation in the nIR include the Ti I line, the Helium triplet He I, the iron Fe I line, the potassium K I line, and the aluminum Al I line (Schöfer et al., 2019; Yana Galarza et al., 2019; Spina et al., 2020; Terrien et al., 2022; Cortes-Zuleta et al., 2023).

Among these activity indicators, one can examine whether a periodicity detected in the RV data also appears in one or more activity proxies, potentially suggesting its stellar origin. When it is determined that the detected signal in the RV data is not of planetary origin, simply filtering out the activity through a linear fit can be challenging due to phase shifts between RVs and activity indicators, which can hinder correlation with the RV signal (Dumusque et al., 2014b; Collier Cameron et al., 2019).

A common approach to mitigate the impact of stellar activity is to employ Gaussian Processes (GPs; Rasmussen & Williams (2006)). GPs can be seen as a way to extend probability distributions to functions and determine which function is the most likely representation of the data. The choice of the GP's behavior is influenced by selecting a kernel, which essentially guides the GP toward a specific function that best fits the data. However, one of the challenges in using GPs is accurately selecting the appropriate kernel. Additionally,

these kernels are defined by a set of hyperparameters, and estimating these parameters can be a complex task. Given that RV variations typically follow a pattern influenced by stellar rotation, quasi-periodic kernels are commonly employed and have demonstrated effectiveness in practice (Donati et al., 2023a; Cortes-Zuleta et al., 2023). Furthermore, employing multidimensional GPs by incorporating activity indicators can enhance the GP’s ability to model stellar activity while preserving planetary signals (Rajpaul et al., 2015).

Additional mitigation techniques exist for addressing stellar activity. One approach involves separating Doppler shift and shape variations in the CCF (Collier Cameron et al., 2021), allowing for the distinction between planetary and stellar activity signals. Another technique uses the varying sensitivities of spectral lines to stellar activity to minimize their contributions to the RV signal (Dumusque, 2018).

#### 1.4.4 Observing in the near-infrared

The contrast between the effective temperatures of bright and dark surface features on M dwarfs determines their relative difference from the quiet photosphere and has an impact on the magnitude of the induced RV signatures. For M dwarfs, this contrast is typically of  $\sim 300$  K while it is of  $\sim 1500$  K for the Sun (Berdyugina, 2005).

When observing in the nIR, this reduction in contrast is expected to help mitigate the contributions of activity-related RV variations (Mahmud et al., 2011; Crockett et al., 2012) and it was observed in the case of AD Leo (Carmona et al., 2023) for which the signal at the rotational period of the star due to activity was not observed in the nIR (figure 1.7). Furthermore, in contrast to G stars, M dwarfs provide the most valuable RV information in the red-optical spectrum and are brightest in the nIR (Reiners et al., 2010, 2018; Quirrenbach et al., 2018; Reiners & Zechmeister, 2020).

As a result, nIR spectrographs have experienced significant development in recent years such as iSHELL (Rayner et al., 2022), CARMENES-nIR (Quirrenbach et al., 2018; Bauer et al., 2020), IRD (Kotani et al., 2014; Hirano et al., 2020), HPF (Mahadevan et al., 2014; Metcalf et al., 2019), GIANO (Carleo et al., 2016; Claudi et al., 2016), SPIRou<sup>6</sup> (Donati et al., 2020), NIRPS (Thibault et al., 2022) and CRIRES+ (Dorn et al., 2023).

The nIR offers clear advantages for exoplanet detection around M dwarfs, although it also presents its unique set of challenges. Firstly, there is the instrumental challenge, where cutting-edge devices like HAWAII 4RGTM (H4RG) detectors were required for spectrographs like SPIRou (Artigau et al., 2018). Another challenge, as previously discussed, involves spectral contamination by telluric lines originating from Earth’s atmosphere which do not exhibit the same Doppler shift as stellar lines. In the nIR, addressing telluric contamination to achieve high RV precision has long been acknowledged as a substantial hurdle (Bean et al., 2010; Artigau et al., 2014; Cook et al., 2022). This challenge primarily arises due to the abundance and depth of telluric absorption lines, notably those stemming from O<sub>2</sub>, H<sub>2</sub>O, CO<sub>2</sub>, CH<sub>4</sub>, N<sub>2</sub>O and O<sub>3</sub>.

Two main approaches have been employed to mitigate the impact of telluric lines on RV measurements. The first approach involves using observations of telluric standard stars,

---

<sup>6</sup>SpectroPolarimètre InfraRouge

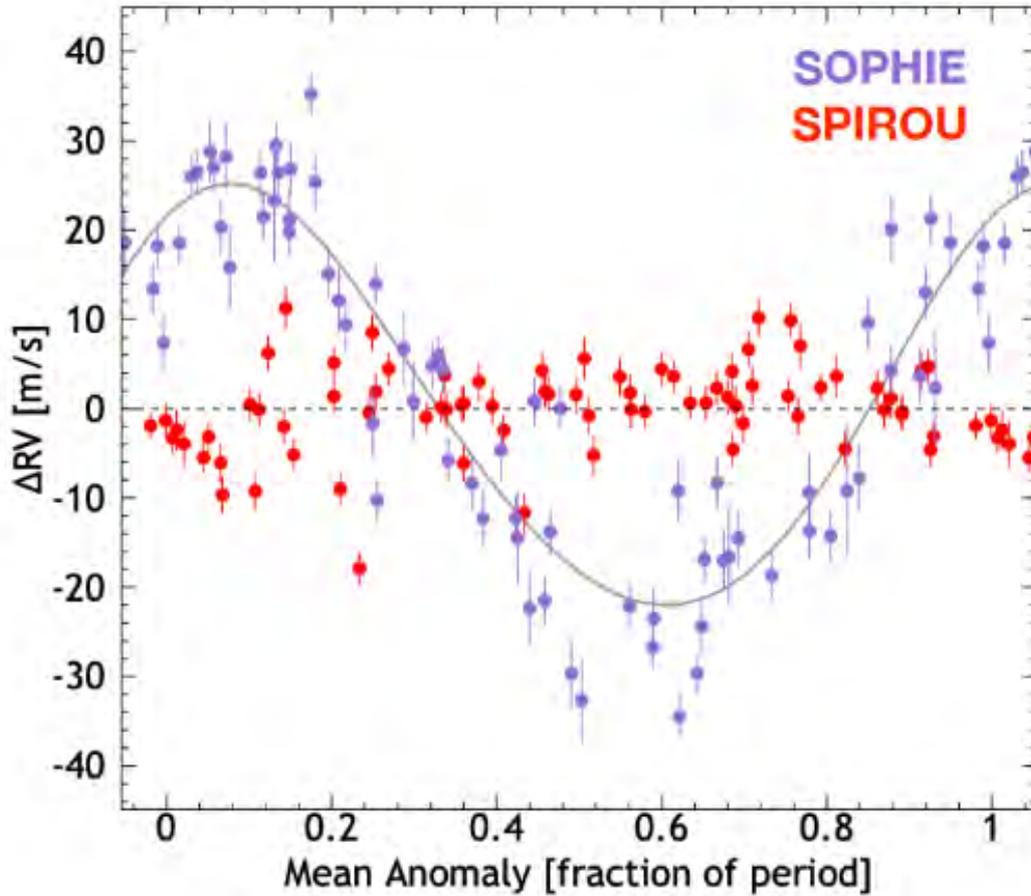


Figure 1.7: Phase-folded RV measurements of AD Leo with SPIROU in the near-IR (red dots) and with SOPHIE (blue dots) in the optical. The gray line represents the Keplerian model with a period of 2.23 days and  $K = 23.6 \text{ m.s}^{-1}$  derived from the SOPHIE data. Credit: [Carmona et al. \(2023\)](#).

which exhibit relatively flat continua, to evaluate and correct for these telluric absorption lines ([Vacca et al., 2003](#); [Artigau et al., 2014](#)). Alternatively, software tools can be used to rectify telluric absorption lines through the use of synthetic spectra ([Bertaux et al., 2014](#); [Smette et al., 2015](#); [Allart et al., 2022](#)). In the most favorable scenarios, both methods result in telluric residuals in the nIR falling within the range of 2% ([Sameshima et al., 2018](#); [Ulmer-Moll et al., 2019](#)). While this residual level might not initially appear significant, it is sufficient to introduce  $\sim 2 \text{ m.s}^{-1}$  systematic errors in the RV data for both G and M type stars in the nIR ([Latouf et al., 2022](#)), giving rise to spurious signals that are modulated on a yearly timescale ([Wang et al., 2022](#); [Latouf et al., 2022](#)).

## 1.5 SPIRou

### 1.5.1 Description of the instrument

SPIRou<sup>7</sup> (SpectroPolarimètre InfraRouge; Donati et al. (2018, 2020)) is a nIR high-resolution spectropolarimeter installed at the Canada France Hawaii Telescope (CFHT). SPIRou draws inspiration from the success of existing instruments, including the ESPaDOnS (Echelle SpectroPolarimetric Device for the Observation of Stars) spectropolarimeter (Donati, 2003) and the HARPS spectrograph (Mayor et al., 2003), which have demonstrated their efficiency. SPIRou comprises three primary units: the Cassegrain module, the calibration unit, and the cryogenic high-resolution spectrograph (figure 1.8).

SPIRou has a wide spectral range, covering 0.98 to 2.35  $\mu\text{m}$ , and offers a high spectral resolving power of around 70,000 (Donati et al., 2020). The cryogenic spectrograph operates under vacuum conditions at a temperature of 73 Kelvin, stabilized to a sub-milliKelvin level (Challita et al., 2018). It incorporates precise light injection mechanisms (Parès et al., 2012) and uses ZnSe quarter-wave rhombs along with a Wollaston prism located in its Cassegrain unit to split incoming light into two orthogonal polarization states and carry out the requested polarimetric analysis of the stellar light. SPIRou also employs two science fibers, designated as A and B, along with a calibration fiber labeled as C (Micheau et al., 2018). The detector used is a 15-micron science-grade HAWAII 4RG<sup>TM</sup> (H4RG) from Teledyne Systems, operating in its up-the-ramp readout mode (Artigau et al., 2018).

### 1.5.2 The SPIRou Legacy Survey

The primary scientific objectives of SPIRou are encompassed within the SPIRou Legacy Survey<sup>8</sup>) (SLS), a 310-night program spanning from the beginning of 2019 to late June 2022. This program is divided into five components, called work packages (WP). The first three packages pertain to key questions regarding M dwarfs and pre-main sequence stars:

- **WP 1:** The SLS-Planet Search (SLS-PS) aims to identify and characterize planetary systems around M dwarfs. It was allocated 153 nights, and the selection of targets was based on the input catalog of Moutou et al. (2017) and Fouqué et al. (2018), with the criterion that these stars were likely to host terrestrial planets. Realistic Monte-Carlo simulations on a sample of 100 nearby M dwarf population and their properties have indicated that SPIRou should be capable of detecting approximately 60 new exoplanets, including 25 Earth-mass planets, with 6 of them orbiting within the habitable zone of their host stars (Cloutier et al., 2018).

The actual sample was downsized to 57 stars to manage the reduced allocation time while maximizing the number of visits per stars, and details can be found on the sample and the observation strategy in Moutou et al. (2023). In summary, out of the 57 stars, 45 are located within 10 parsecs of the Sun, representing a quarter of the population of M dwarfs within this distance (Reylé, 2021). Most of the targets in the sample are known to have a low activity level (Moutou et al., 2017), and planetary signals have already been detected around sixteen of them (Table 1 in Moutou et al. (2023)).

---

<sup>7</sup><http://spirou.irap.omp.eu/>

<sup>8</sup><https://spirou-legacy.irap.omp.eu/doku.php/>

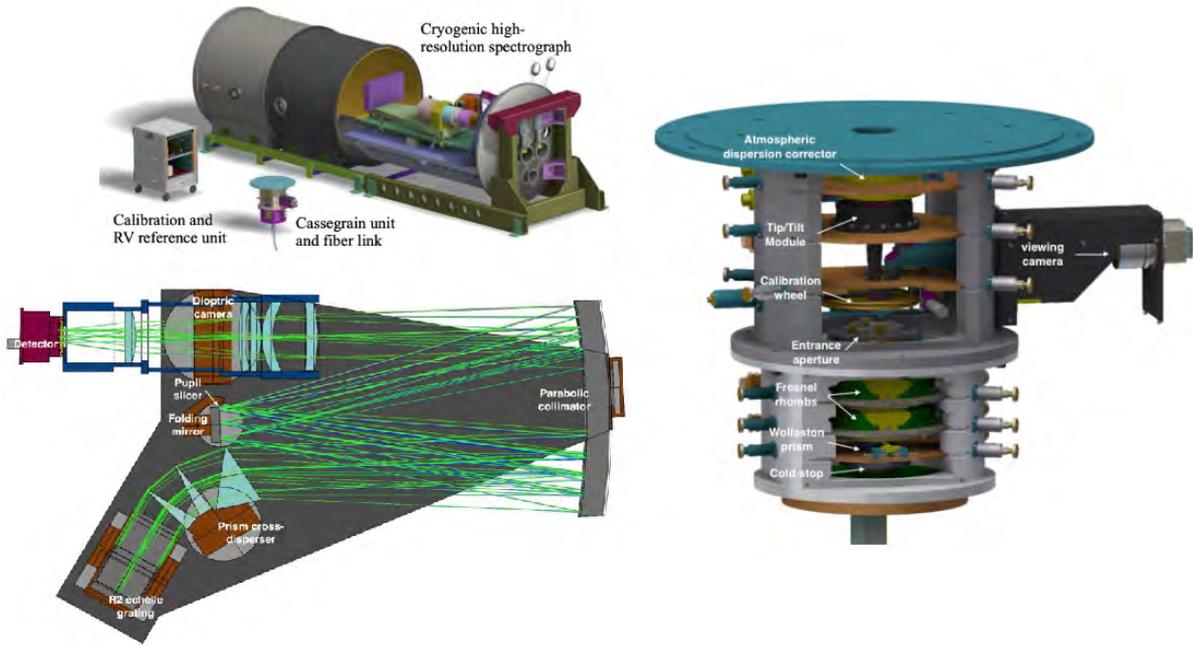


Figure 1.8: Computer-aided design (CAD) views of the main SPIROU units (top-left) and the Cassegrain unit (right), the optical layout of the SPIROU spectrograph is also displayed (bottom-left). Credit: [Donati et al. \(2020\)](#).

- **WP 2:** The SLS-Transit Follow-up (SLS-TF) was granted 76 nights and is specifically designed for observing planet candidates previously detected by space-based telescopes like TESS or K2. The objective is to validate these planets and measure their mass, a crucial parameter for atmospheric analysis. Some follow-up investigations can be found in the works of [Martioli et al. \(2022\)](#); [Cadieux et al. \(2022\)](#); [Kiefer et al. \(2023\)](#); [Almenara et al. \(2023\)](#).
- **WP 3:** The SLS-Magnetized Star/Planet Formation (SLS-MP) initiative was granted 76 nights and is specifically focused on using SPIROU's spectropolarimetric capabilities to investigate magnetic fields through the continuous monitoring of pre-main sequence stars.

The remaining two work packages are complementary to these three WPs. **WP 4** focuses on the development of new observational and theoretical methods to enhance the results of **WP 1**, specifically aimed at improving the detection and characterization of newly discovered exoplanets. As for **WP 5**, it encompasses additional analyses that can be conducted using SPIROU data. This includes for instance the development of tools for modeling star spectra and establishing automated spectral classifications.

### 1.5.3 A PipelinE to Reduce Observations (APER0)

APER0 (A PipelinE to Reduce Observations, [Cook et al. \(2022\)](#)) is an open-source Python 3 package freely accessible to the public<sup>9</sup> and it has been specifically designed as the official data reduction pipeline for SPIRou at CFHT. The main objective of APER0 is to process raw data acquired from the telescope, applying necessary calibration and correction procedures to mitigate instrumental and systematic effects. The pipeline also performs spectral extraction, generating both 2D and 1D spectra before and after telluric correction. Furthermore, APER0 provides RV estimates and facilitates the calculation of polarimetry data as needed. This analysis is divided into different stages (see [Cook et al. \(2022\)](#) for exhaustive details on the DRS).

#### Telluric correction

As previously mentioned, the majority of telluric correction methods fall into two categories: model-based ([Gullikson, 2014](#); [Smette et al., 2015](#); [Allart et al., 2022](#)), or data-driven ([Artigau et al., 2014](#); [Sameshima et al., 2018](#); [Bedell et al., 2019b](#); [Cretignier et al., 2021](#)). APER0, on the other hand, employs a hybrid three-step telluric correction approach. We provide a concise summary here, with a comprehensive description to be presented in forthcoming work by Artigau et al. (in preparation) and we also display results from this correction in figure 1.9 for a random observation.

Firstly, the extracted spectrum undergoes a preliminary correction by fitting it with a TAPAS ([Bertaux et al., 2014](#)) atmospheric model. Subsequently, any remaining residuals are eliminated through the use of observations from fast-rotating hot stars, which were observed with SPIRou under varying atmospheric conditions. Lastly, finite resolution effects originating from telluric modeling ([Wang et al., 2022](#)) are mitigated, resulting in final residuals at a 1 – 2% level consistent with the Principal Component Analysis (PCA) method outlined in [Artigau et al. \(2014\)](#).

### 1.5.4 Extracting the RV with the line-by-line method

One of the critical requirements for SPIRou is the precise measurement of RVs. A commonly employed technique in the field is the CCF method ([Mayor & Queloz, 1995](#)). Initially, within the APER0 DRS, this method served as the default tool for deriving RV measurements ([Cook et al., 2022](#)), alongside an independent implementation ([Martioli et al., 2020](#)). However, an alternative approach, the line-by-line method (LBL; [Artigau et al. \(2022\)](#)), has since been implemented and has yielded the most accurate RV measurements obtained from SPIRou to date. All RV measurements in this thesis have been computed using the LBL method.

This method, which is rooted in the framework initially developed by [Bouchy et al. \(2001\)](#), was proposed by [Dumusque \(2018\)](#) and further explored by [Cretignier et al. \(2020\)](#) within the context of optical RV observations using HARPS. The LBL approach calculates Doppler shifts for individual spectral lines, resulting in a velocity spectrum for each measurement epoch (figure 1.10). To accomplish this, a low-noise template is essential, as the RVs are derived by comparing the residuals of the observed spectrum with the template, along with the derivative of the template. In practice, a high S/N combined spectrum is employed as a

---

<sup>9</sup><https://github.com/njcuk9999/apero-drs>

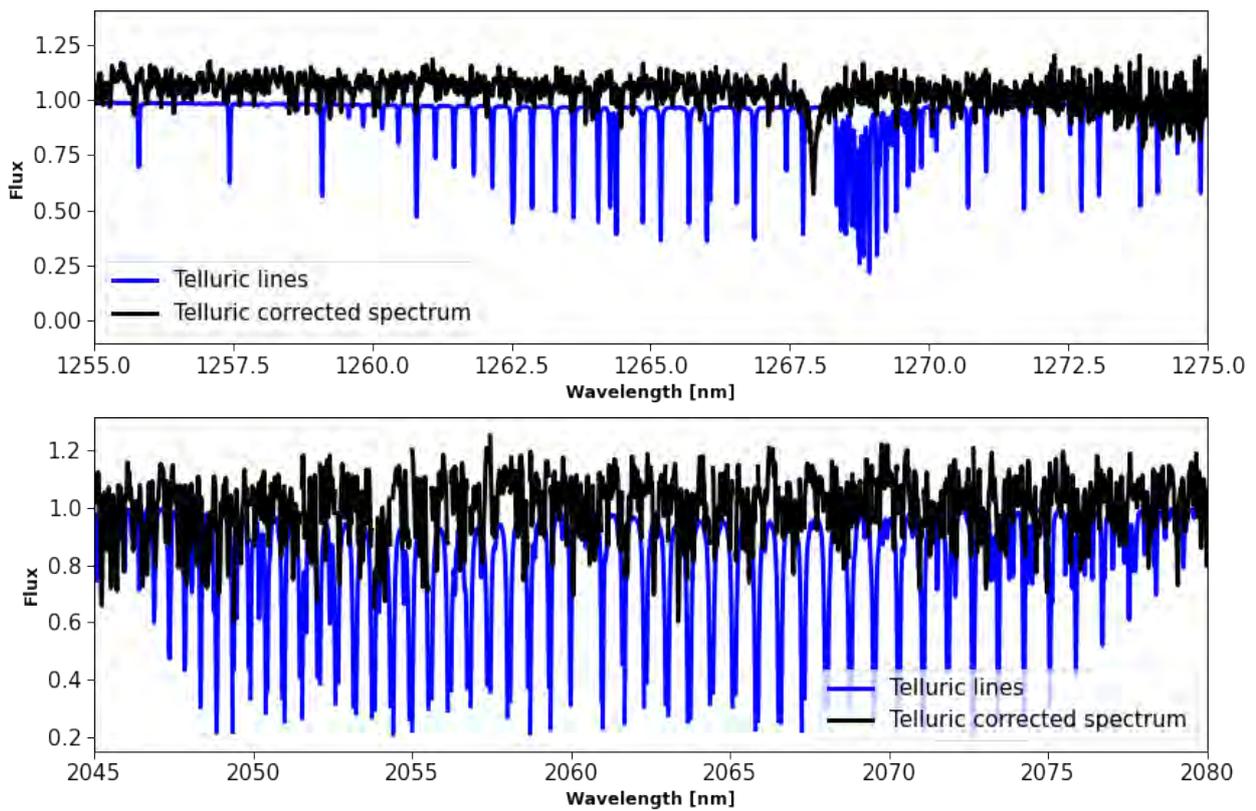


Figure 1.9: Example of the telluric correction in APERO. The telluric corrected spectrum is depicted in black while the telluric lines remove from the spectrum are in blue. Top panel shows a 2D spectrum retrieved from APERO from 1255 to 1275 nm while the bottom panel shows the same spectrum from 2045 to 2080 nm.

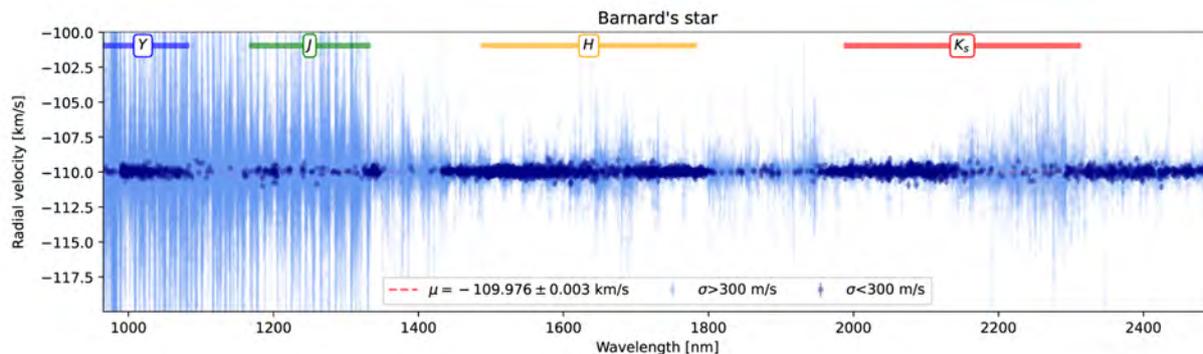


Figure 1.10: Sample line-by-line velocity spectrum for Barnard’s star showing the scatter of velocity measurements over a wide velocity range; lines with uncertainties smaller than  $300 \text{ m.s}^{-1}$  are shown in dark blue to improve readability. Credit: [Artigau et al. \(2022\)](#).

template for a specific target, ensuring that any remaining noise is negligible when compared to an individual spectrum.

The velocity spectrum can then be averaged to obtain a single RV measurements per epochs. The LBL method has been subject to comparative studies against the CCF and template matching methods such as NAIRA ([Astudillo-Defru et al., 2017b](#)) in [Martoli et al. \(2022\)](#) and [Artigau et al. \(2022\)](#). Presently, it is widely applied in most RV analyses involving SPIRou (e.g; [Ould-Elhkim et al. \(2023\)](#); [Moutou et al. \(2023\)](#)).

## 1.6 Context of the thesis

Despite the impressive precision of the LBL method, it fails to correct for systematic errors when their cumulative effect becomes significant when averaging the velocity spectrum. These systematic errors can arise from various sources, including stellar activity, telluric lines, or instrumental errors. To illustrate this point, we present an RV time series obtained from SPIRou data in figure 1.11 featuring observations of Gl 251. This particular star is known to host a habitable-zone planet with an orbital period of 14.2 days ([Stock et al., 2020](#)). The time series comprises 176 data points, processed with the latest APERO version (v0.7286), using the most recent version of the LBL (v0.60) for the extraction, and corrected for the FP drift. Details on the RV measurements can be found in appendix C.2.1. To detect periodicities in the time series, we employ a generalized Lomb-Scargle (GLS) periodogram ([Zechmeister & Kürster, 2009](#)). The periodogram spans a grid of 100,000 points, ranging from 1.1 days to 1,000 days, using the `astropy` package ([The Astropy Collaboration, 2018](#)). The false alarm probability (FAP) associated with the maximum peak is determined using the analytical formula outlined in [Baluev \(2008\)](#).

As observed in figure 1.11, the planetary signal, while discernible, lacks statistical significance. Instead, we can discern a long-term signal originating from the instrument (even after accounting for nightly drift using the FP correction, the instrument can exhibit long-term variations). Additionally, we can readily identify other peaks at 180 days and 1 year, marked by vertical dashed lines, which are attributable to tellurics.

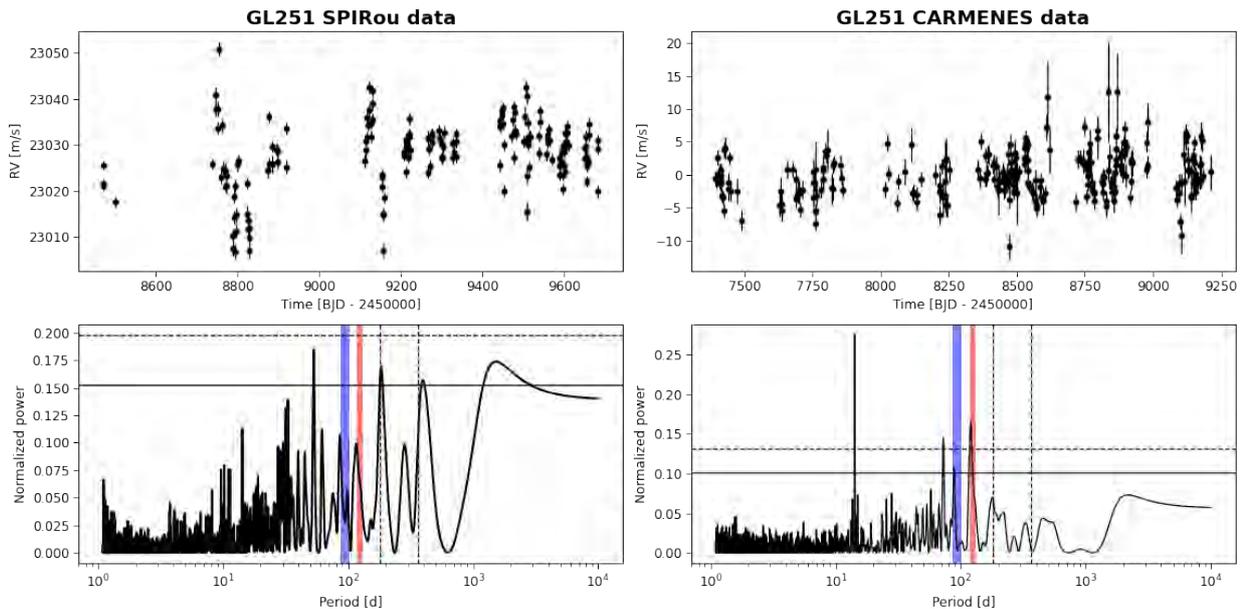


Figure 1.11: Comparison between SPIRou (left) and CARMENES (right) RV data of Gl 251. Top panels show the RV time series and bottom panels the respective GLS periodograms. The 180 days and 1 year periods are highlighted with vertical dashed black lines. The red and blue vertical bands represent respectively the rotational period obtained from photometry and longitudinal magnetic field data. The significance levels of  $FAP = 10^{-3}$  and  $FAP = 10^{-5}$  are indicated by the solid and dashed horizontal lines, respectively

Furthermore, photometric analysis has revealed a rotational period of  $124.2^{+4.8}_{-5.1}$  days (Stock et al., 2020), and an analysis of the longitudinal magnetic field time series has unveiled a rotational period of  $93.4 \pm 6.6$  days (Donati et al., 2023b), depicted in red and blue, respectively, and both visible in the periodogram even if not significant. Lastly, the origin of the most prominent peak at 53 days remains unclear; it does not align with harmonics of stellar activity periods, nor is it a 1-year alias of those periods. Recent work has shown that this peak could correspond to correlated structures in the pixels, which will soon be corrected in APERO.

In any case, the absence of detection is not satisfying, especially when considering that the 14.2-day signal in the optical domain is easily discernible. For comparison, we show data from CARMENES (Quirrenbach et al., 2018), obtained from a recent data release (Ribas et al., 2023) and processed with SERVAL (Zechmeister et al., 2018). The planetary peak stands out as the most prominent feature, even though we can note that the activity signal (from photometric data) appears more pronounced in the optical domain compared to the nIR. Given these observations, it becomes apparent that in addition to the LBL method and the FP correction, there is a need for supplementary corrections to make the most of the data provided by SPIRou.

### 1.6.1 Zero-Point correction

One initial correction that can be applied is addressing the nightly zero-point (NZP) of the instrument, which is responsible for the long term signal observed in the data. This

correction can be performed relatively straightforwardly by selecting several targets observed with SPIRou that exhibit good time coverage and are known to be stable over time.

Once these stable stars are identified, the NZP can be computed through several methods. One approach involves calculating the NZP by averaging their RVs using a rolling median (Trifonov et al., 2020). Alternatively, as demonstrated in figure 1.12, a GP can be fitted to the data using the `celerite` module (Foreman-Mackey et al., 2017) with a Matérn-3/2 kernel.

As depicted in figure 1.12, this correction effectively mitigates the long-term signal, but the planetary signal remains hidden by other systematic effects. Furthermore, the 1-year signal has become more pronounced. An alternative, yet unpublished method that has emerged during this thesis is the notion of extending the correction beyond the time domain and into the "Vtot" space. This space represents the difference between the systemic velocity of the targets and the BERV.

The underlying idea is that systematic errors originating from telluric sources may exhibit similar behavior in this "Vtot" space. As illustrated in Figure 1.12, applying a GP indeed captures a discernible pattern, suggesting the potential effectiveness of this approach.

The application of this correction gives rise to what is referred to as the "bervzp" time series. While the precise details of this method are yet unpublished, these time series are now readily available among the SLS collaboration and are considered suitable for planetary search purposes. Examining figure 1.12, we can observe that, even with the application of this most recent correction, the planetary signal remains elusive within the data, and the 1-year signal continues to dominate.

## 1.6.2 Overview of the thesis

The primary objective of this thesis is to introduce an improved method for mitigating systematic errors and uncovering planetary signals within SPIRou data obtained using the LBL method. This innovative approach is encapsulated in the `Wapiti` method (Weighted principal component analysis reconstruction), elaborated upon in Chapter 1, which uses weighted Principal Component Analysis (wPCA) to estimate and correct for the systematics present in the data.

Chapter 2 of the thesis is dedicated to simulating the impact of telluric lines on LBL RV data and subsequently applying wPCA to understand how this method behaves in a controlled environment with specific challenges such as telluric lines. Lastly, in Chapter 3, the thesis provides examples on three targets illustrating how `Wapiti` can seamlessly integrate itself into the broader context of the SLS-PS program, demonstrating its potential utility within this program.

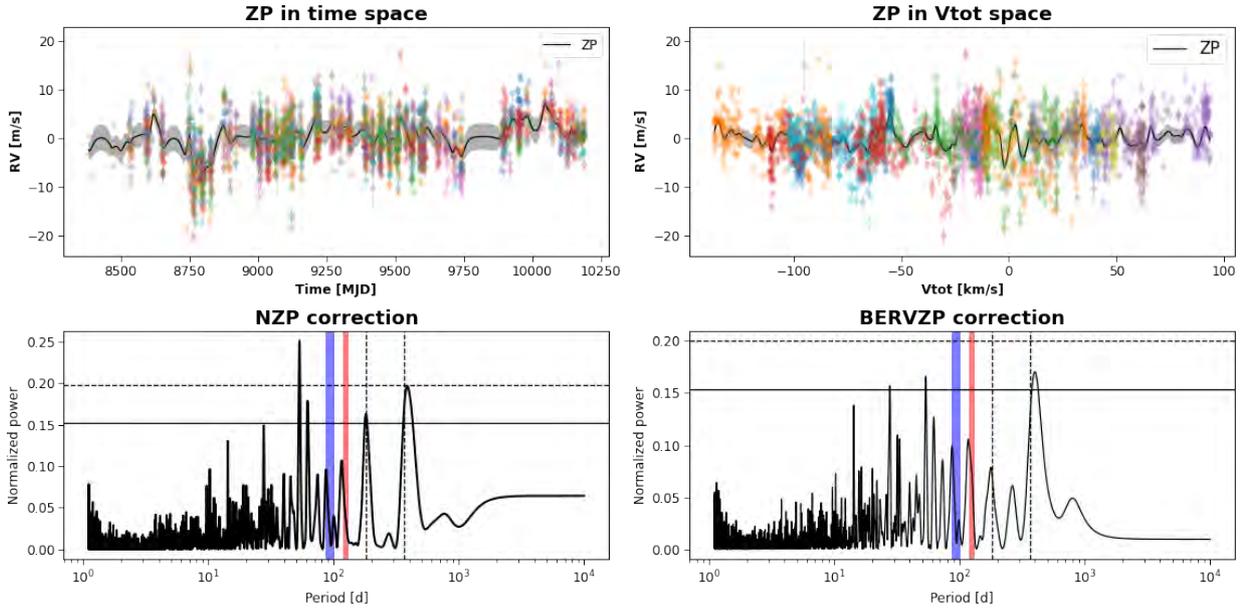


Figure 1.12: Illustration of the ZP corrections. Top panels show the ZP correction in time (left) and  $v_{\text{tot}}$  (right) spaces. The bottom panels show the respective GLS periodograms when applied to correct the RV of Gl251. The 180 days and 1 year periods are highlighted with vertical dashed black lines. The red and blue vertical bands represent respectively the rotational period obtained from photometry and longitudinal magnetic field data. The significance levels of  $\text{FAP} = 10^{-3}$  and  $\text{FAP} = 10^{-5}$  are indicated by the solid and dashed horizontal lines, respectively



# Chapter 2

## Wapiti: a data-driven approach to correct for systematics in LBL RV data

### Contents

---

<b>2.1</b>	<b>Overview</b>	<b>31</b>
<b>2.2</b>	<b>The Wapiti method</b>	<b>32</b>
<b>2.3</b>	<b>Results</b>	<b>39</b>

---

### 2.1 Overview

In this chapter, we present a data-driven approach called **Wapiti** (Weighted principal component analysis reconstruction) to address the challenges of detecting exoplanetary systems around M-type stars using near-infrared (nIR) observations. Recent advancements in precise radial velocity (RV) instruments like SPIRou have opened up new opportunities to study planetary systems around M-type stars. However as previously mentioned, the nIR domain poses unique challenges due to various planet-independent signals leading to systematic errors in RV data.

To overcome these challenges, we introduce **Wapiti**<sup>1</sup> (Ould-Elhkim et al., 2023), which uses a dataset of per-line RV time-series generated by the line-by-line (LBL) algorithm and employs a weighted principal component analysis (wPCA; Delchambre (2015)) technique to reconstruct the original RV time-series. A carefully designed multi-step process is used to estimate the systematics with the appropriate components, with the goal of removing them from the original data.

The effectiveness of the **Wapiti** method is demonstrated through its application to the star Gl 251, introduced earlier. This method was initially presented and detailed in Ould-Elhkim et al. (2023) for the same target, and readers can refer to the specifics provided in appendix

---

<sup>1</sup><https://github.com/HkmMerwan/wapiti>

**C.2.1.** Since its initial publication, the method has undergone certain refinements and adjustments, and the current version is presented in this thesis. Consequently, there might be some disparities compared to the original method presented in the earlier publication.

## 2.2 The Wapiti method

Our approach involves employing PCA on the per-line RV time series generated by the LBL algorithm of the shape lines  $\times$  epochs, meaning that each row is a per-line RV timeseries. The main goal of the **Wapiti** method is to estimate the systematics within the data by using the principal vectors from the wPCA as a proxy.

PCA is a widely used statistical technique in the field of data analysis and dimensionality reduction (Jolliffe, 1986). It is a mathematical tool that aims to simplify complex data sets while retaining the most important information. PCA helps in identifying patterns, correlations, and underlying structures in high-dimensional data by transforming the original variables into a new set of uncorrelated variables, known as principal components.

The importance of PCA lies in its ability to reduce the dimensionality of data, which is crucial when dealing with large datasets with numerous variables. PCA addresses these issues by compressing the data into a lower-dimensional space, capturing the dominant patterns and variations in the original data. Data analysis is not the only application of PCA; it also plays a significant role in data reconstruction. By retaining the most relevant principal components, PCA enables the reconstruction of the original data with minimal loss of information.

PCA has many applications in a wide variety of astronomical domains from the classification of the Sloan Digital Sky Survey (SDSS) quasar spectra and their redshift determination (Pâris et al., 2014) to the study of the point spread function variation in lensing surveys (Jarvis & Jain, 2004). Regarding RV research, PCA-based techniques have also been investigated for correcting telluric contamination (Artigau et al., 2014) and addressing stellar activity effects (Collier Cameron et al., 2021; Cretignier et al., 2022; Donati et al., 2023b).

### Weighted-PCA

There have been a number of extensions to PCA proposed to handle datasets with noise and containing missing data such as LBL RV data. In this work, we use the wPCA method introduced in Delchambre (2015). Mathematically, the weights correspond to the inverse of the Gaussian variances associated with each data point. In the case of missing data the weights are equal to zero. When the weights are equal, the outcome of the process simplifies to that of standard PCA.

### Data preprocessing

Before applying wPCA to the data, several preliminary steps are undertaken. First, the correction for the Fabry-Perot (FP) drift is applied to all RV measurements. Subsequently, all the spectral lines are grouped into night-binned datasets. Following this, nightly-averaged individual line RV time series with more than 50% of missing data are excluded from further analysis, reducing the initial count of 27 272 lines to 17 701. Missing data in this context can

happen when the RV value deviates by more than five standard deviations from the line’s RMS as outlined in [Artigau et al. \(2022\)](#). This step is crucial to ensure the proper behavior of wPCA, as some lines even have 100% missing data (7 625 of them in the case of G1251). However, discarding all lines with missing data would underutilize wPCA’s ability to handle missing data while reducing the line count to 12 900. Therefore, removing lines with 50% or more missing data seems like a reasonable balance, ensuring that wPCA has adequate data, while maintaining a substantial number of lines.

Subsequently, a nightly zero-point correction in the time domain is computed and subtracted from all remaining per-line RV time series to eliminate long-term instrumental variations. The list of standard stars used to compute this ZP is GJ 905, GJ 699, GJ 1012, GJ 1289, GJ 3378, GJ 752 A, GJ 1002, GJ 1286, GJ 1103 and PMJ21463+3813.

Following this, the weighted-averaged RV time series is subtracted from each per-line RV time series to ensure that Keplerian or any achromatic signals, such as trends, remain unaffected by wPCA and are not misinterpreted as systematic errors. Finally, the time series are standardized by using their inverse-variance weighted mean and standard deviation.

With these steps completed, the objective now is to estimate the optimal components that can effectively replicate the systematics within the data. This is done through a three-step procedure.

### 2.2.1 Permutation test or is the component better than noise ?

We use a permutation test ([Vieira, 2012](#)) to determine the components in the wPCA that are better than noise. This test involves independently shuffling the columns of the dataset and then conducting a wPCA on the shuffled dataset, remembering that each column represent the value of all lines at a given epoch. We repeat this process 100 times and compare the explained variance of the original dataset with that of the permuted versions. To identify the relevant components, we examine the p-values obtained from the permutation test and only select those with low p-values ( $\leq 10^{-5}$ ). and as a result, we find that only the first 24 components cannot be explained by noise. An illustration of this process is provided in [figure 2.1](#), where we also display the explained variance ratio of components 1, 21, and 35 from the permutations in comparison to the one derived from the actual data. The variance explained by components 1 and 21 is higher than what would be expected if they were purely noise, and inversely for component 35.

### 2.2.2 Leave-p-out cross-validation or is the component chromatically stable ?

To evaluate the significance of the number of components retained after the permutation test, we conduct a leave-p-out cross-validation. This approach was used in [Cretignier et al. \(2022\)](#) to select relevant components, albeit with variations in the methodology. In this study, we generate 100 subsets of the per-lines RV dataset, each containing 80% of the lines randomly selected. We then recalculate the wPCA for each of these N datasets to assess the robustness of the 24 remaining components. Following this process, we are left with a set of  $V_{i,k}$  principal vectors, with the index  $i$  denoting the component number and the index  $k$

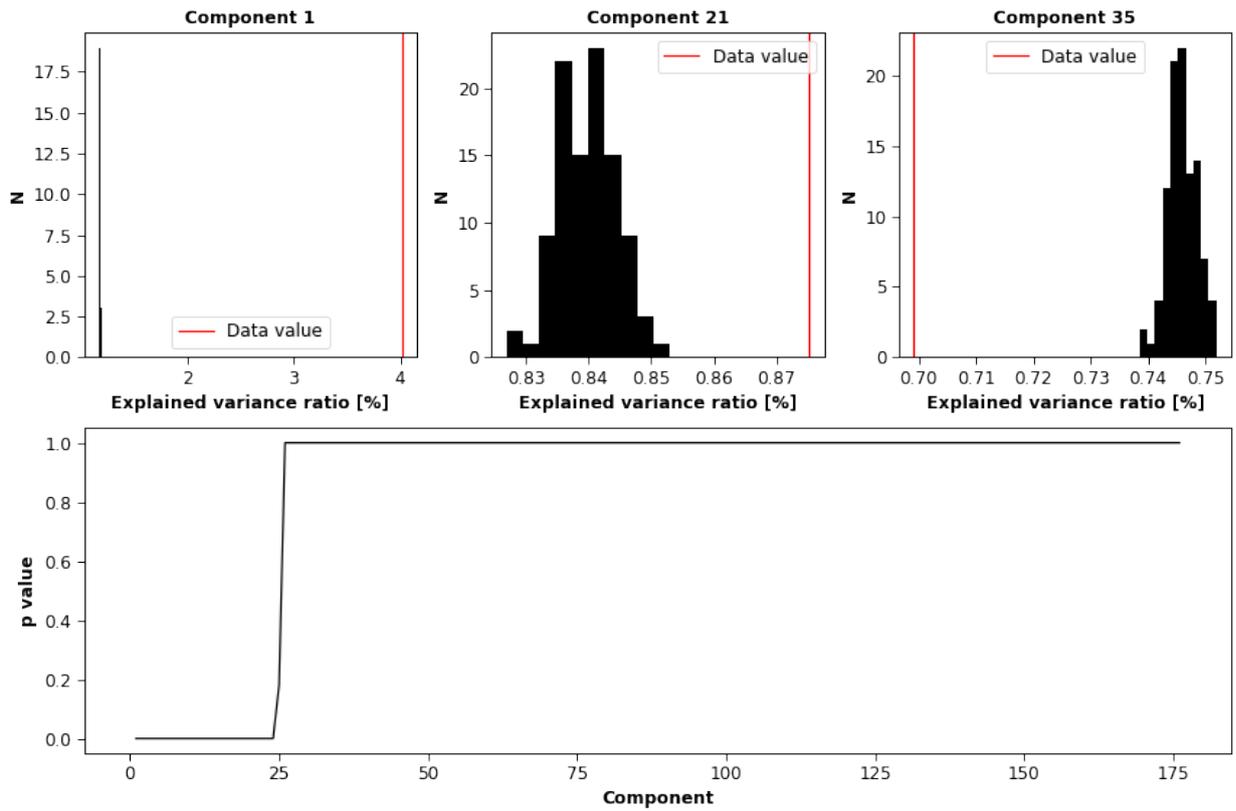


Figure 2.1: p-values of each component as determined by the permutation test. Only the components with a p-value below a certain threshold ( $\leq 10^{-5}$ ) are considered to have a significant contribution, rather than being just random noise.

ranging from 0 to  $N$ , where  $k = 0$  corresponds to the original principal vector when wPCA is applied to the entire dataset.

The underlying idea involves assessing the stability of the principal vectors, by analyzing their response when certain lines are removed. If a principal vector shows excessive sensitivity to the lines, we discard it. To evaluate this, we use the Pearson coefficient of correlation, as explained in the work of [Cretignier et al. \(2022\)](#), due to its computational efficiency, bounded nature, and insensitivity to vector amplitudes. We expect stable wPCA vectors  $V_{i,k}$  from the subsets to be highly correlated with  $V_{i,0}$ . Since the correlation may change sign, we take the absolute value of the coefficient.

We can also visualize the correlation matrix ([Cretignier et al., 2020, 2023](#)) derived from the leave-p-out subsets in figure 2.2, which serves as a valuable tool to explain how we use this test to eliminate components. To clarify, it is important to understand how to interpret this correlation matrix: the value at the 103<sup>rd</sup> row and 804<sup>th</sup> column represents the absolute Pearson coefficient between the second component of the 3<sup>rd</sup> subset and the 9<sup>th</sup> component of the 4<sup>th</sup> subset, and so on.

In the visualization, distinct diagonal blocks become evident, each of size  $N \times N$ . This outcome is expected as it illustrates that the  $n^{\text{th}}$  component among all subsets tends to exhibit correlation with the same components across the subsets.

However, some of these blocks are less well-defined than others, indicating components that may be considered chromatically unstable. To quantify this observed sensitivity, an approach employed in [Cretignier et al. \(2023\)](#) and in the original version of Wapiti (as shown in appendix C.2.1) is to compute the median Pearson coefficient  $\text{med}(\rho)$  of these blocks and subsequently reject components falling below a certain arbitrary threshold, respectively 0.6 and 0.95, see figure 2.2.

Nevertheless, this method has several limitations. First, determining the threshold value is subjective, and there is no clear definition of what constitutes a "good" Pearson value. Furthermore, these values tend to vary among different targets, making the choice of the threshold prone to fine-tuning for each specific target, which is not ideal. Additionally, this approach does not fully use the wealth of information available from the subsets, as it condenses all Pearson values into a single data point. However, the distribution of these Pearson coefficients is equally valuable information.

As illustrated in figure 2.2, where we plot the 16th and 84th quantiles as error bars, the distribution of these Pearson coefficients can vary significantly. Having similar median Pearson values does not necessarily imply that the components are equivalent. For example, the 11<sup>th</sup> component may have a similar median Pearson value as the 12<sup>th</sup> component, but their Pearson coefficients are more dispersed, indicating distinct characteristics.

Therefore, we aim to fully leverage the distribution of Pearson coefficients while quantifying the distinctions between these coefficient distributions. To achieve this, we introduce the Pearson distribution  $\rho_{i,k}$ , which represents the Pearson coefficients computed when comparing all principal vectors  $V_{i,j}$  with  $j \neq k$  to  $V_{i,k}$ , some example of those distributions are depicted in figure 2.2. Our goal is to compare each normalized distribution  $\rho_{i,k}$  to  $\rho_{i,0}$ . For this purpose, a natural choice is the Kullback-Leibler divergence ([Kullback & Leibler, 1951](#)), denoted as  $\text{DKL}(\rho_{i,k} || \rho_{i,0})$ . This measure quantifies how similar  $\rho_{i,k}$  is to  $\rho_{i,0}$ , with values

closer to 0 indicating greater similarity between the distributions. This calculation is performed for all  $k$  values and we define the chromatic sensitivity (CS) of a component  $i$  as the median of these divergence values,

$$CS(i) = \text{med}_k (\text{DKL}(\rho_{i,k}^{\hat{}} || \rho_{i,0}^{\hat{}})) . \quad (2.1)$$

Admittedly, the chromatic sensitivity, as introduced here, is more complex than a straightforward median calculation. However, it represents a necessary refinement, and while its derivation may appear challenging, it effectively quantifies a crucial aspect of the problem: how the distribution of Pearson coefficients for a component behaves in response to changes in line selection. This complexity is a justified trade-off for the valuable insights it provides.

In figure 2.2, we present the chromatic sensitivity of each component in a logarithmic scale. Notably, the error bars associated with this metric are considerably less scattered compared to those observed when using the median of their Pearson coefficients. Additionally, it is evident that a more dispersed distribution corresponds to a higher value of chromatic sensitivity (CS), aligning with our intended interpretation. Therefore, this metric can be readily understood: the closer the CS value is to 0, the less sensitive the component is to the line selection. We illustrate several examples in figure 2.2, where we display the normalized Pearson distribution  $\rho_{i,0}^{\hat{}}$  in black for four components (i.e., 1, 10, 18, and 21). Alongside these distributions, we present the first three  $\rho_{i,k}^{\hat{}}$  distributions. Additionally, the chromatic sensitivity of each component is indicated at the top of these plots.

Subsequently, we proceed to eliminate all components that fall below a chosen threshold of  $10^{-2}$ . While this threshold may appear arbitrary, it possesses the advantage of not relying on the exact value of the Pearson coefficient but rather considers their distribution. This approach is more robust and permits the consistent application of the same threshold across all targets, aligning with our earlier argument against arbitrarily determined thresholds when taking the median.

### 2.2.3 Re-ordering or is the component good for reconstruction ?

After the preceding steps, we are left with 13 components, which may be an excessive number, considering that not all principal components equally contribute to the underlying systematics in the data. This arises from the arrangement of principal components, which does not inherently align with our intended objective. Our goal is to optimize the reconstruction of the RV time series, but the principal components are initially sorted in descending order of their contribution to the variance of the LBL RVs.

As a result, we proceed to restructure the principal components by rearranging them in descending order of their Bayesian Information Criterion (BIC; Schwarz (1978)) values. During this process, we identify the component that, when incorporated, minimizes the BIC. This component is designated as the new primary component. Subsequently, we iterate this procedure to determine the second component, which, when added to the primary component, further minimizes the BIC, and so forth. This reordering yields a new sequence of principal vectors, denoted as  $\hat{V}_i$ . Using the mathematical framework from Cretignier et al. (2023) this translates to computing the BIC when fitting the following model to the data when considering the  $k$ -th component:

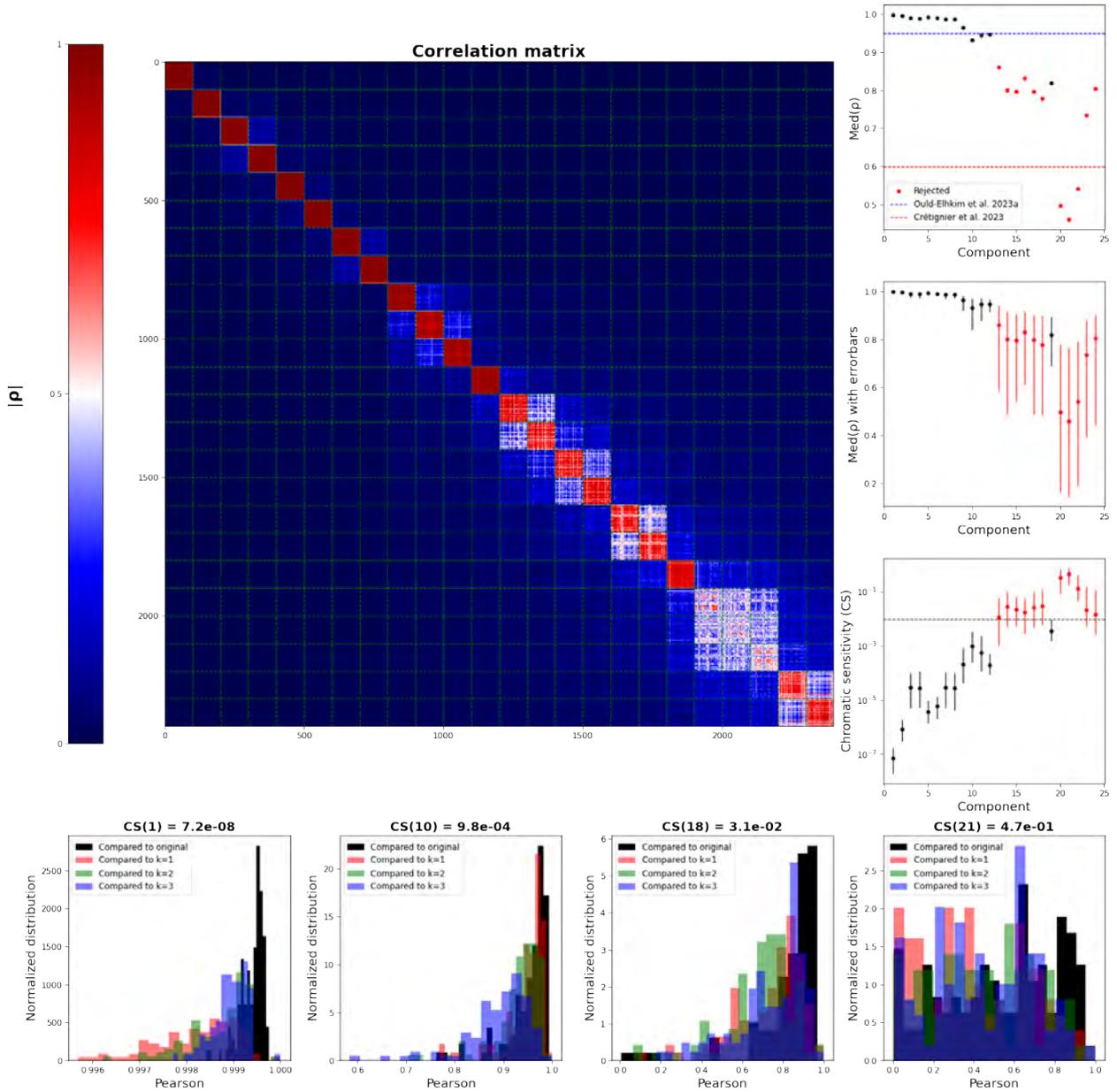


Figure 2.2: Leave-p-out cross-validation with G1251. The top-left panel shows the correlation matrix from the leave-p-out cross-validation. Top and middle right panels show the median Pearson of each diagonal blocks, with and without error bars respectively. The chromatic sensitivity defined in equation 2.1 is shown just below with the threshold at  $10^{-2}$  with a black line. Bottom panels display the normalized Pearson distribution  $\rho_{i,0}$  in black for four components (i.e., 1, 10, 18, and 21) alongside the first three  $\rho_{i,k}$  distributions. The chromatic sensitivity of each component is indicated in their title.

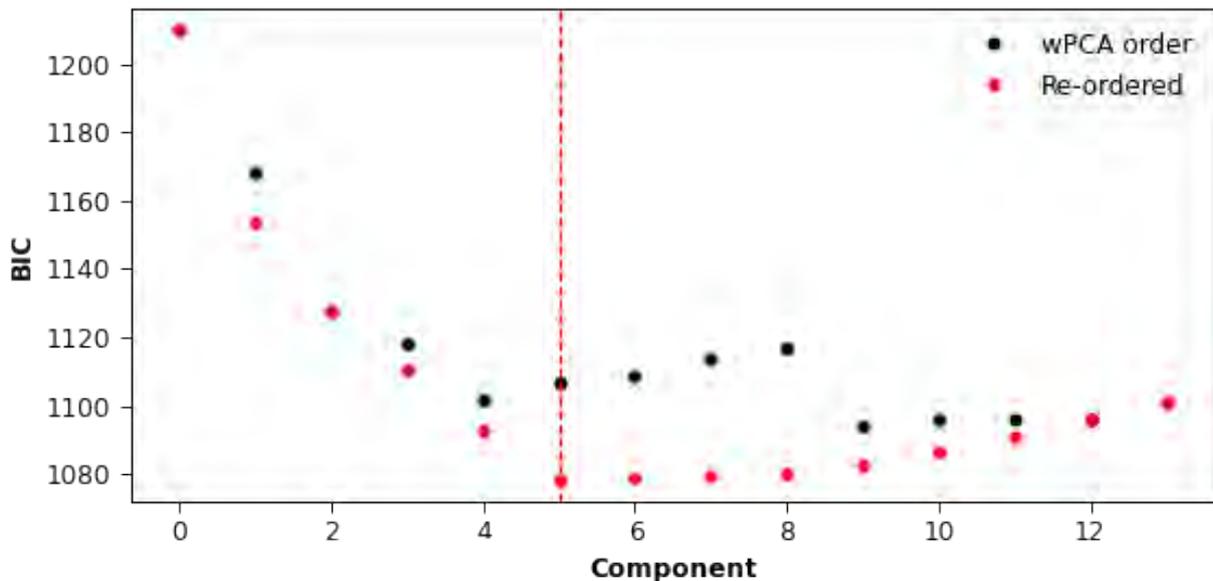


Figure 2.3: Re-ordering the wPCA components in decreasing BIC order for Gl 251. In black, the BIC when fitting the model from equation 2.2 using the principal vectors in the wPCA order  $V_i$ . In red, the BIC when using the  $\hat{V}_i$  principal vectors re-ordered in decreasing BIC order.

$$RV(t) = \gamma_S + \sum_{i=1}^N a_i V_i. \quad (2.2)$$

Where we included an offset term  $\gamma_S$  and the coefficients  $a_i$  to fit the principal vectors  $V_i$  in the model. The parameter  $N$  represents the number of components used in the reconstruction, and it can be set to 0, corresponding to the data without applying *Wapiti* correction. We fitted these models through weighted least-squares regression using the `statsmodels` module (Seabold & Perktold, 2010), which also provided their BIC value.

To determine the final number of selected components, we simply choose the number of components that results in the lowest BIC value, as shown in Figure 2.3. This reordering method is similar to the  $\chi^2$  reordering technique employed in Collier Cameron et al. (2021), with the only difference being our use of the BIC as the criterion to penalize using too many components. Consequently, this process leads to a reduction in the number of components employed from 13 to 5. The first five identified  $\hat{V}_i$  components, in order, are  $V_2$ ,  $V_1$ ,  $V_9$ ,  $V_4$ , and  $V_3$ . This underscores the fact that explaining the highest variability in the per-line RVs does not necessarily imply that these components are the most effective at reconstructing the data. To be clear, by variability we mean that some principal components may have high amplitudes of RV variation, but their scaled contribution to the final goodness of fit is too small to have a significant effect.

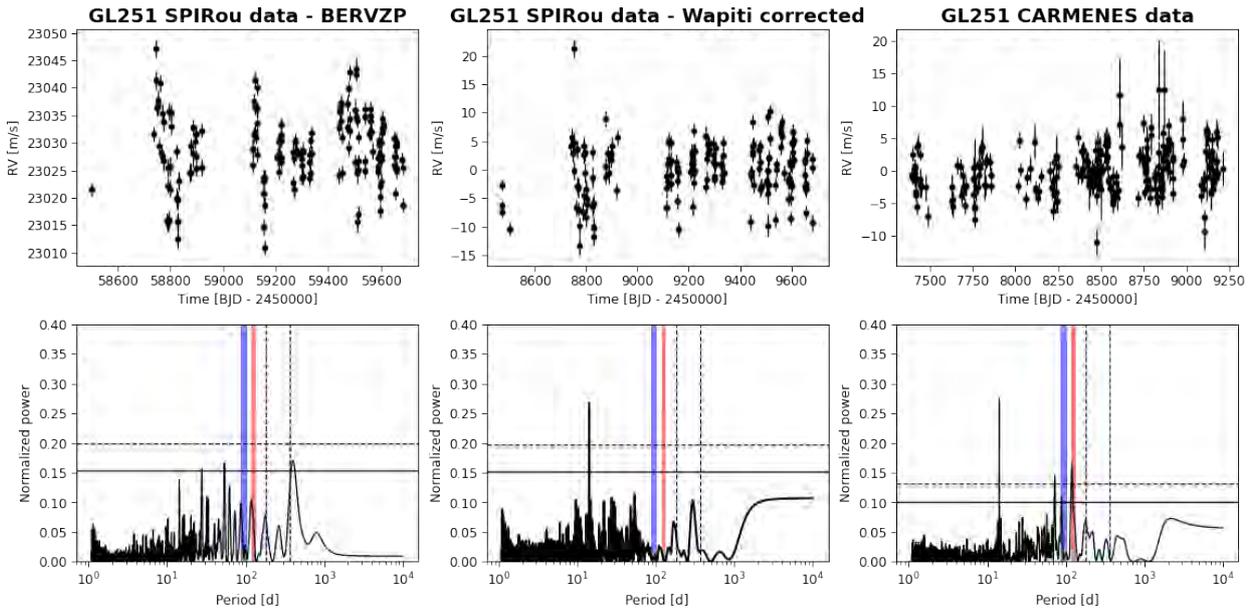


Figure 2.4: Comparison between SPIRou data using `bervzp` (left) or `Wapiti` (middle) correction and CARMENES (right) data of Gl 251. Top panels show the RV time series and bottom panels the respective GLS periodograms. The 180 days and 1 year periods are highlighted with vertical dashed black lines. The red and blue vertical bands represent respectively the rotational period obtained from photometry and longitudinal magnetic field data. The significance levels of  $\text{FAP} = 10^{-3}$  and  $\text{FAP} = 10^{-5}$  are indicated by the solid and dashed horizontal lines, respectively

## 2.3 Results

The three-step process detailed in the previous section enables us to identify the principal vectors from wPCA that are most suitable for reconstructing the data. These vectors serve as proxies to model the systematic errors, and we eliminate them by fitting equation 2.2 to the RV. As a result, we obtain the time series showcased in figure 2.4, which we present alongside SPIRou data that underwent the `bervzp` correction and CARMENES data, as shown in the introduction.

The results demonstrate that the `Wapiti` method successfully recovered the 14.2-day planet signal in SPIRou data. The power of this signal in the periodogram is comparable to the one in CARMENES data, while simultaneously eliminating systematic errors. To be specific, the FAP associated with the 14.2 days signal in CARMENES data is at  $2.2 \times 10^{-16}$ , and  $5.0 \times 10^{-9}$  in the `Wapiti` corrected data. This allowed for the first detection of this planet in the nIR using SPIRou, while the comprehensive analysis and characterization of this signal fall beyond the scope of this chapter, interested readers can find detailed information in appendix C.2.1.

Nevertheless, there are two important considerations to qualify this result. Firstly, the `Wapiti` method was initially developed with the specific objective of recovering the already known 14.2-day signal in Gl 251. Therefore, it is imperative to evaluate the `Wapiti` method's capacity to detect other known signals in different targets to demonstrate that it is not just

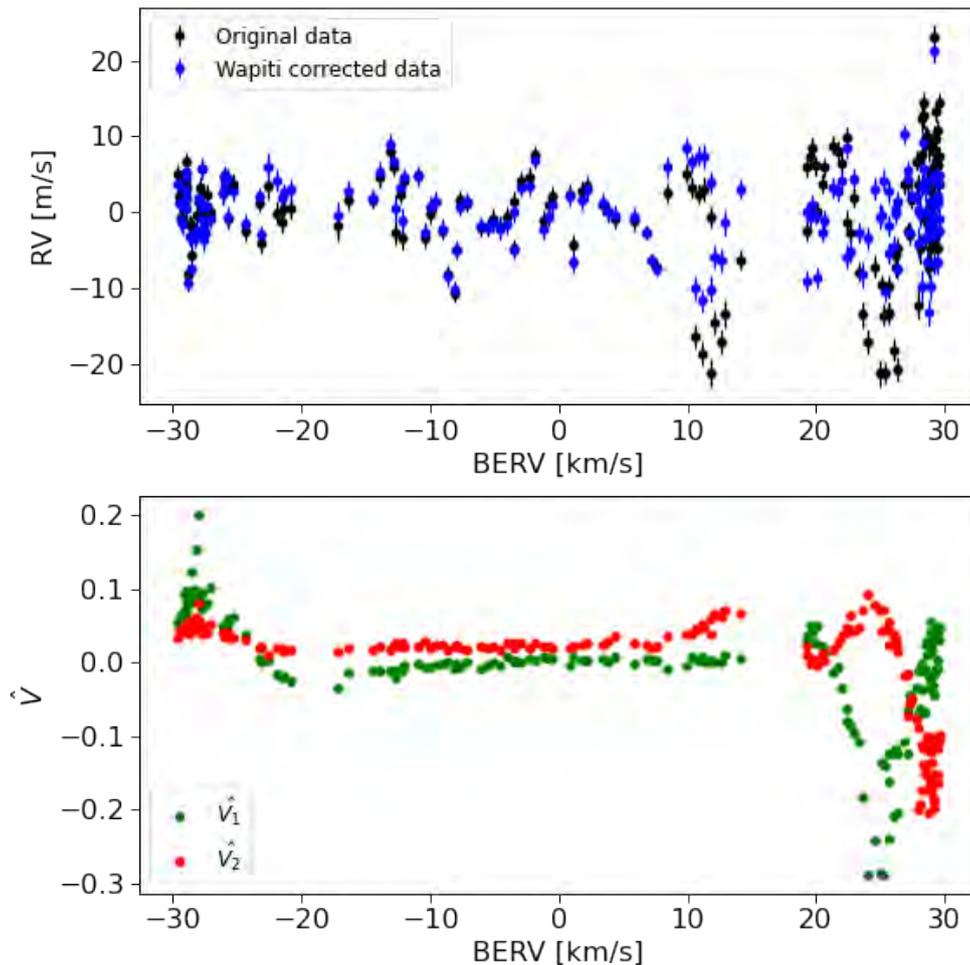


Figure 2.5: Variability of RV data and principal vectors in the BERV space. In the top panel, the original data’s variability with respect to BERV is depicted, along with how the **Wapiti** method effectively removed this dependency in the corrected RV. The bottom panel showcases the behavior of the first two principal vectors with respect to the BERV.

tailored for Gl251 and instill confidence in the method if it were to identify unknown signals in SPIRou data.

Secondly, due to its data-driven nature, there is a need for a deeper understanding of the physical aspects that the **Wapiti** method corrects. By visualizing in figure 2.5 the original and corrected data in the BERV space and examining the first two components used for reconstruction,  $\hat{V}_1$  and  $\hat{V}_2$ , it becomes evident that **Wapiti** may correct for a signal related to the BERV which could be due to tellurics.

From this observation, we were motivated to conduct a more in-depth investigation into the influence of telluric absorption on LBL RV data using simulations. By precisely controlling the amount of injected telluric residuals during the simulation, we can isolate their effects on the RV data. This will enable us to assess the effectiveness of the **Wapiti** method in reducing telluric-related systematic errors and gain a deeper comprehension of what it captures.

# Chapter 3

## Simulating the impact of telluric absorption lines on radial velocities

### Contents

---

3.1	Overview . . . . .	41
3.2	The challenges of correcting for telluric lines . . . . .	42
3.3	The impact of telluric absorption lines . . . . .	43
3.4	wPCA and telluric lines . . . . .	45
3.5	Summary . . . . .	58

---

### 3.1 Overview

In this chapter, our focus is directed towards investigating the influence of telluric absorption lines on radial velocities (RVs) in the near-infrared (nIR) by employing simulations. As discussed in chapter 2, systematic errors originating from tellurics exert an impact on SPIRou data. However, comprehending real data becomes challenging due to their susceptibility to various forms of contamination, such as unusual observations, stellar activity, suboptimal signal-to-noise (S/N) ratios, and more. Consequently, the use of simulations becomes essential for isolating and enhancing our understanding of the impact of telluric lines in our data, particularly those arising from telluric absorption lines.

We will also leverage these simulations to assess the effectiveness of the `Wapiti` method (Ould-Elhkim et al., 2023) in correcting induced systematic errors and recovering injected signals. As expounded upon in chapter 2, the method effectively corrected the systematic errors in the context of GJ 251 data and successfully retrieved the planetary signal. This success suggests a high likelihood of the method’s capability to rectify telluric-related systematic errors. However, the controlled nature of the injected tellurics and the simplicity of the data will significantly facilitate the interpretation of the outcomes produced by the `Wapiti` method or any other wPCA-base method for the matter.

Knowing that telluric lines are responsible for the systematic issues might make one wonder why we can not just remove these lines from the spectra before calculating the RVs. There are actually different ways to try fixing the spectra by dealing with these lines, but as we will detail in the next section, none of them work perfectly. Even in the best-case scenario, there are still some remaining telluric residuals, and these are causing the observed systematics.

## 3.2 The challenges of correcting for telluric lines

The issue of telluric contamination has long been acknowledged as a barrier to achieving improved precision in RV measurements in the nIR domain (Bean et al., 2010). This problem arises since the nIR region contains more and deeper telluric absorption lines, particularly from water, methane, carbon dioxide, nitrous dioxide, and oxygen.

Minimizing the impact of telluric lines on RV measurements poses a significant challenge, especially when tackling telluric absorption lines. In earlier investigations, scientists commonly used observations of telluric standard stars (hot stars with a mostly flat continuum) to understand and remove these telluric absorption lines (Artigau et al., 2014). Noteworthy is the result achieved by Sameshima et al. (2018) using this approach, attaining low telluric residuals of 2% for water lines in the nIR range, specifically within the 900 to 1350 nm wavelength interval.

In the research conducted by Ulmer-Moll et al. (2019), a comparison was made among different methods and commonly used software tools for addressing telluric absorption lines using synthetic spectra, such as the TAPAS model (Bertaux et al., 2014). The findings indicated that the most effective results typically showed residuals ranging between 3% and 7%. In the most successful instances, these residuals were decreased to levels as minimal as 1% to 2%.

Therefore, effectively addressing the challenge presented by telluric contamination is of great significance, especially when aiming for precise nIR velocimetry. In the study conducted by Wang et al. (2022), the impact of telluric absorption lines on RV time series was replicated through simulations. This was accomplished using synthetic telluric spectra and a synthetic spectrum of a G-type star, covering a broad spectral range from 350 to 2500 nm. Expanding on this, Latouf et al. (2022) further investigated the case of an M dwarf. They explored various mitigation techniques, ultimately identifying the most effective approach. Notably, in the most favorable scenario, they achieved a minimal RV scatter of  $1.6 \text{ m.s}^{-1}$  attributed to telluric contamination in the nIR range.

This chapter introduces a novel approach for simulating and mitigating the influence of telluric absorption lines within the context of SPIRou RV data. This work builds upon the SPIRou data reduction system (DRS), APERO (Cook et al., 2022), and employs the line-by-line (LBL) framework to extract data from the telluric-corrected spectra (Dumusque, 2018; Artigau et al., 2022). By accurately reproducing telluric-related systematics, we conduct the first test of the Wapiti method (Ould-Elhkim et al., 2023) on simulated data. Furthermore, given that the efficacy of telluric correction hinges on the Barycentric Earth radial velocity (BERV) coverage (Cretignier et al., 2021), we divide our simulations into two scenarios: one with wide BERV coverage and the other with narrow BERV coverage.

## 3.3 The impact of telluric absorption lines

### 3.3.1 Outline of the method

In order to assess the influence of telluric absorption lines on RV data, we employed the subsequent approach:

To begin with, we opted for a star exhibiting a notable BERV excursion and an extensive dataset. This selection ensured that the template, denoted as  $T_*$ , would exhibit minimal residuals originating from telluric effects. The star of choice for our analysis was Gl 905 (HH And and Ross 248), classified as a M5.0V dwarf (Alonso-Floriano et al., 2015). The maximum BERV value (BERVMAX) associated with this star is  $23 \text{ km.s}^{-1}$  according to the data reduction system DRS header. Note that the star template  $T_{Gl905}$  is in the reference frame of the star itself.

Gl 905 was observed using the SPIRou instrument in a series of circular polarization sequences. Each observation night comprised 4 distinct sub-exposures. The observational span extended from April 26, 2019, to June 09, 2023, encompassing a cumulative count of 1073 sub-exposures, which were organized into groups of 4 per visit. However, considering the objectives of our simulations, our interest lies in the full-night observations rather than sub-exposures. Therefore, we opt to focus solely on the last observation of each night, resulting in a reduced dataset containing 253 nights.

To investigate the scenario of a narrow BERV coverage, we decided to use observations of the M dwarf Gl 725 B, which we know to be close to the north ecliptic pole and to have a BERVMAX value of  $4.6 \text{ km.s}^{-1}$  based on the DRS header information. Gl 725 B was observed using the SPIRou instrument through a series of circular polarization sequences. Each night of observation consisted of 4 separate sub-exposures. The observational campaign spanned from February 14, 2019, to April 19, 2022, comprising a total of 860 sub-exposures, organized into sets of 4 per visit. For similar reasons to our approach when using observations of Gl 905, we limit our analysis to the last observation of each night reducing the number of epochs to 210.

Subsequently, we substituted each recorded spectrum with the designated template, to which we introduced photon noise. This process yielded distinct spectra for each epoch, labeled as  $T(t) = T_{Gl905} + \sigma_{photon}(t)$ . In order to replicate the influence of telluric effects, we created a two-dimensional (2D) telluric template. This was accomplished using the output of APERO, which furnishes the 2D telluric spectra corresponding to each observation via the telluric correction process (Cook et al., 2022). By computing their median within the Earth's reference frame, we mitigated residuals stemming from the star itself, yielding a telluric template denoted as  $T_{\oplus}$ .

Once the telluric template was obtained, we proceeded by adjusting the noisy star template to align it with the Earth's reference frame, using the BERV value associated with the specific epoch. Subsequently, for each individual observation, the noisy star template was multiplied with the telluric template. This multiplication process was controlled by a parameter denoted as  $\alpha$ . To mimic the variability of telluric effects across different epochs, the parameter  $\alpha$  was sampled from a normal distribution denoted as  $\mathcal{N}(\mu_{\alpha}, \sigma_{\alpha})$ . Herein, we assigned a value of 3% to  $\mu_{\alpha}$ , and 1% to  $\sigma_{\alpha}$ , for plausible and practical values. The parameter  $\alpha$  may assume

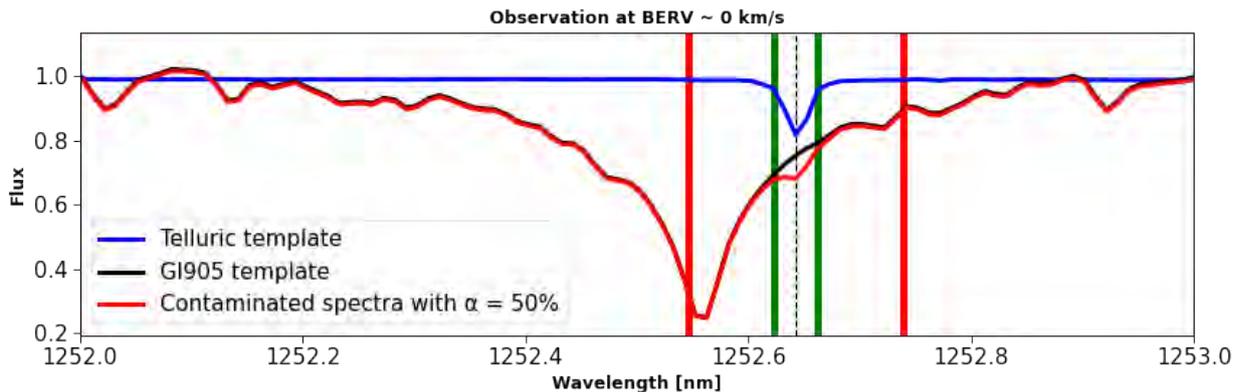


Figure 3.1: Illustration of the simulations to inject tellurics in the star template with  $\alpha = 50\%$ . The telluric template is in blue, the original star template in black and the resulting telluric contaminated spectrum in red. The green and red vertical bands represent the excursion limit of the telluric line in the spectrum in the case of a narrow and wide BERV coverage, respectively.

either positive or negative values, allowing for over or under-correction. In mathematical terms, to introduce telluric contamination for a given epoch, the computation is as follows:

$$T_{contaminated} = T(t) \times T_{\oplus}^{\alpha(t)}. \quad (3.1)$$

In figure 3.1, we illustrate the injection of telluric lines into the stellar template. We have arbitrarily displayed it from 1252 to 1253 nm, with an  $\alpha$  value of 50% to ensure its visibility. Additionally, we show the difference in BERV excursion between the two cases by highlighting the excursion extent of the telluric line in the template. In the case of a narrow BERV coverage, it is represented by green bands, while in the case of a wide BERV coverage, it is indicated in red. We show it for an epoch when the BERV is close to  $0 \text{ km.s}^{-1}$  ( $\text{BERV} = -0.008 \text{ km.s}^{-1}$ )

### 3.3.2 Results

We generated the RV time series employing the LBL algorithm. To ensure realism, we refrained from directly employing the mask and template from G1905. Instead, we recalculated a star mask and template based on the simulated  $T_{contaminated}(t)$ . However, for reasons that are not entirely clear, the LBL algorithm rejected three nights in the case of the wide BERV coverage at 2458752.9, 2460097.1, and 2460098.1 BJD, as well as in the case of the narrow BERV coverage at 2458601.1, 2458789.7, and 2458790.7 BJD. Fortunately, these rejections have minimal impact on our simulations.

We conducted three simulations for both wide and narrow BERV coverage and found negligible differences between them. Consequently, for better clarity in presenting the results, we only display the RVs from the first simulation in each case. After deriving the velocity spectra for each simulated observation using the LBL method outlined by Artigau et al. (2022), we excluded individual line RV time series with over 50% missing data, enabling the computation of the final RV time series. The corresponding results are depicted in figure

3.2, where we display the RVs in BERV instead of time, as BERV is the parameter of primary interest in our analysis. For the detection of periodicities within the time series, we employed a generalized Lomb-Scargle (GLS) periodogram (Zechmeister & Kürster, 2009). This periodogram was computed through a grid comprising 100,000 data points, ranging from 1.1 days to 1,000 days. The analysis was carried out using the `astropy` package (The Astropy Collaboration, 2018). To assess the significance of the highest peak, we determined its false alarm probability (FAP) using the analytical formula outlined in Baluev (2008).

In figure 3.2, we can observe that in the case of a wide BERV coverage, we find a RMS similar to the simulations of Latouf et al. (2022), however we do not see any periodicities at 1 year or any harmonics in the periodograms. This shows that in this case the wide BERV coverage was sufficient to mitigate the impact of tellurics in the data when computing RVs with the LBL. On the other hand in the case of a narrow BERV coverage, the RMS sky-rockets to a value of  $8.86 \text{ m.s}^{-1}$  and signals at 1 year and other telluric harmonics are prominent in the periodogram. As anticipated, the BERV coverage was not sufficient to correct for the systematic induced by tellurics which is visible from the behavior of the RVs in the BERV space.

We also computed the LBL RVs of the same simulated spectra using a different template to assess how employing a potentially improved template might mitigate the impact of telluric contamination. In our scenario, the "better" template was the original Gl 905 template and the results are displayed in red in figure 3.2. However, in the context of real data, the most suitable choice is yet to be determined. Nonetheless, essential considerations would include selecting a star with similar stellar properties, a similar or even greater number of observations, and a broader BERV coverage. In our case, the RMS of the time series decreases by half which is not negligible, but the signals of telluric origin still maintain their significance in the data. This underscores the necessity for a correction method, even if we were to have a perfect template.

## 3.4 wPCA and telluric lines

The context of the subsequent work is to assess the effectiveness of `Wapiti` in correcting for systematic errors originating from telluric absorption lines. However, the results we will present are general to any wPCA-based method designed to correct for telluric systematics in LBL RV data.

### 3.4.1 Principal vectors

From the previous work, we have three datasets of LBL RVs: one with wide BERV coverage, one with narrow BERV coverage, and the last one with narrow BERV coverage but using a perfect template to compute the RV. Our initial focus is on studying the principal vectors obtained when applying wPCA to these datasets, specifically identifying those that are relevant. We thus used the permutation test (Vieira, 2012; Ould-Elhkim et al., 2023) introduced in chapter 2 to ensure that a component is better than simple noise. We only applied this test to the wide BERV coverage and narrow BERV coverage simulations, which was sufficient for drawing conclusions. As a result, we selected 11 components in the case

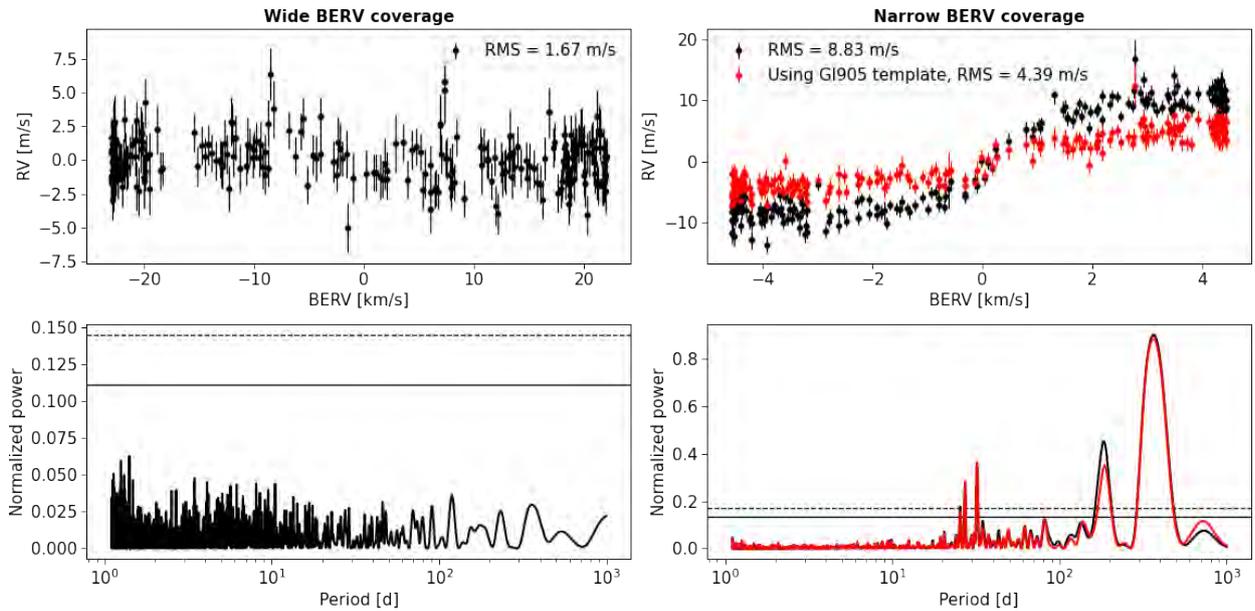


Figure 3.2: Simulation results when injecting telluric at a  $3 \pm 1\%$  level. Top panels display the resulting RV timeseries from the simulations in the BERV space, while the bottom panels present their respective GLS periodograms. Left panels depict the results for a wide BERV coverage, while the right panels show the outcomes for a narrow BERV coverage. The RVs computed with the original G1905 template are indicated in red. The significance levels of  $FAP = 10^{-3}$  and  $FAP = 10^{-5}$  are indicated by the solid and dashed horizontal lines, respectively

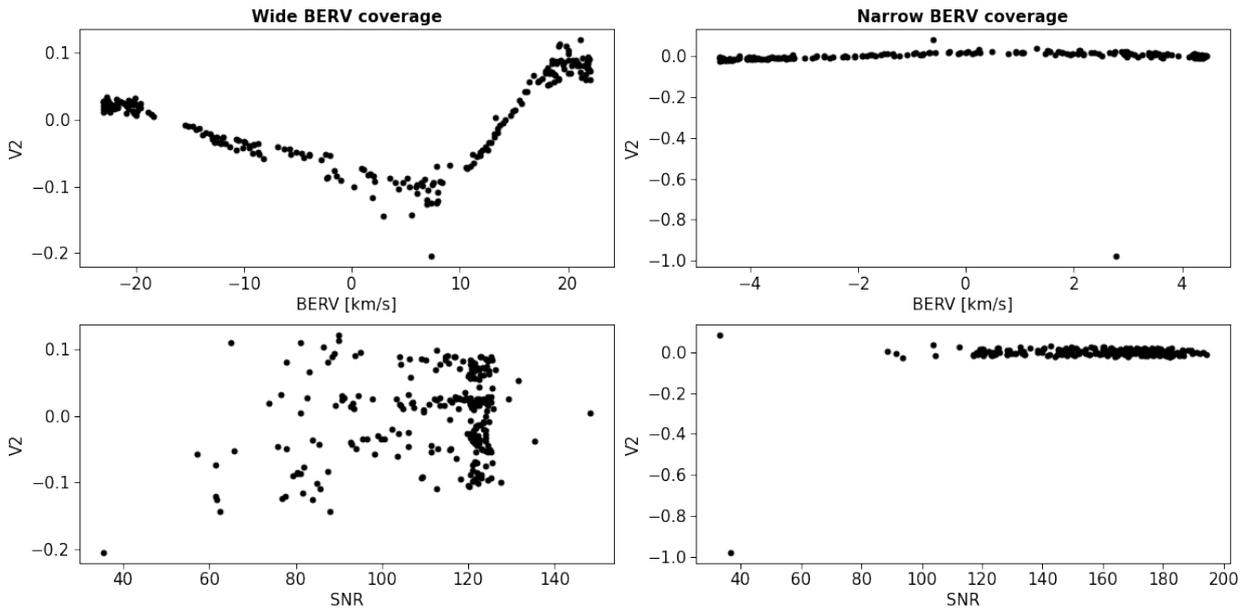


Figure 3.3: Principal vectors from the simulations without rejecting low SNR observations. Top panels display the second principal vector of the simulations in the BERV space, while the bottom panels present them as a function of the SNR. Left panels depict the results for a wide BERV coverage, while the right panels show the outcomes for a narrow BERV coverage. The principal vectors are clearly affected by low SNR observations, showing the sensitivity of wPCA to outliers.

of the wide BERV coverage simulation and 4 components in the narrow BERV coverage simulation.

However, if we display as an example the second principal vector in figure 3.3, it is clear that outliers are influencing our results. Such outliers can introduce bias to the wPCA results since the method is sensitive to data variability, potentially leading the permutation test to erroneously conclude that a component is better than noise. This issue is highlighted when we plot the principal vectors alongside the SNR of the observations. Consequently, we opted to exclude all epochs with an SNR below 100 to ensure that our components will be well-behaved and to take advantage of the fact that we have a high number of observations. This reduced the number of components in the case of the wide BERV coverage simulation to 9 components and in the case of the narrow BERV coverage simulation to only 2 components. In practical applications with real data, we would strongly recommend considering the removal of low SNR observations if feasible, as long as it does not result in the loss of too many data points.

### Principal vectors in the case of a narrow BERV coverage

To comprehend the behavior of the first principal vector  $V_1$  in the BERV space, we calculated what we refer to as an indicator of telluric contamination denoted as  $d1_{\oplus}$ . This indicator is determined by computing the Pearson correlation between the telluric 2D spectra produced by the APERO DRS and the first derivative of the star template. As a reminder, in the LBL algorithm, the RV is determined by subtracting the star template from the observed spectrum and then projecting the residuals onto the first derivative of the star template

(Artigau et al., 2022). Therefore, the use of Pearson correlation is a natural choice since it allows us to investigate the linear correlation of this derivative with the telluric lines. By averaging these individual time series, we derived the telluric indicator  $d1_{\oplus}$ .

However, calculating this correlation for all lines would be unnecessary since not all lines contribute to the systematic effects with the same importance. To identify the lines most responsible for the systematics, we performed a linear fit of each spectral line  $j$  to the principal vectors  $V_i$  and obtained their linear coefficients  $a_{i,j}$ . Using these coefficients, we can then compute the z-scores  $z_i$  for each component  $i$ , defined as  $\frac{a_{i,j} - \mu_i}{\sigma_i}$ , where  $\mu_i$  and  $\sigma_i$  represent the weighted mean and standard deviation of the coefficients across the spectral lines.

As demonstrated in figure 3.4, the lines primarily responsible for these systematics are concentrated within the region of strong telluric absorption lines. Consequently, we computed  $d1_{\oplus}$  by only calculating the Pearson coefficients for the top 5% of lines based on the absolute  $z_1$  values. Figure 3.4, shows that  $d1_{\oplus}$  and  $V_1$  exhibit a similar pattern and are, in fact, highly correlated with a Pearson coefficient of  $r = 0.96$ . Consequently, our interpretation of  $V_1$  is that it serves as an effective proxy for assessing the extent to which telluric absorption lines contaminate the first derivative of the template. This contamination contributes to the emergence of a 1-year signal in the data and upon fitting it to the RV time series, we find that it exhibits an RMS of  $8.65 \text{ m.s}^{-1}$ .

Concerning  $V_2$ , we followed a similar procedure to compute  $d2_{\oplus}$ , which represents the mean linear correlation between telluric lines and the second derivative of the template. Notably,  $d2_{\oplus}$  and  $V_2$  display a high correlation, with a Pearson coefficient of  $r = -0.62$ . This observation suggests that  $V_2$  serves as a quantification of the impact of telluric contamination on the second derivative of the template, which, in turn, introduces a 180-day signal in the data. Although this principal vector contributes significantly to the variability in the LBL RVs, its net effect remains relatively small, with an RMS of  $0.29 \text{ m.s}^{-1}$ .

### Principal vectors in the case of a wide BERV coverage

In the scenario of a wide BERV coverage, we have identified 9 components that cannot be explained by noise according to the permutation test. The associated principal vectors are displayed in figure 3.5 with their respective periodograms. Interpreting so many principal vectors is less straightforward, but it can be attempted. The first principal vector,  $V_1$ , still exhibits a strong negative correlation with  $d1_{\oplus}$  ( $\rho = -0.89$ ), while the correlation of  $V_2$  with  $d2_{\oplus}$  is not as strong ( $\rho = 0.27$ ). However, the impact of these principal vectors on the data, namely  $0.2 \text{ m.s}^{-1}$  for  $V_1$  and  $0.029 \text{ m.s}^{-1}$  for  $V_2$ , could potentially explain these lower correlations due to their less pronounced impact on the data.

Regarding the remaining components, their correlation with the other  $di_{\oplus}$  telluric indicators breaks, but they can be interpreted as resembling standing waves in a cavity with a size of 2BERVMAX. In simpler terms, wPCA recognized the periodic pattern in the BERV space and tried to decompose the telluric signals in the data in a manner akin to a Fourier decomposition. These components have minimal impact on the data, typically in the range of  $10^{-3}$  to  $10^{-1} \text{ m.s}^{-1}$ , but they are nonetheless captured by the wPCA.

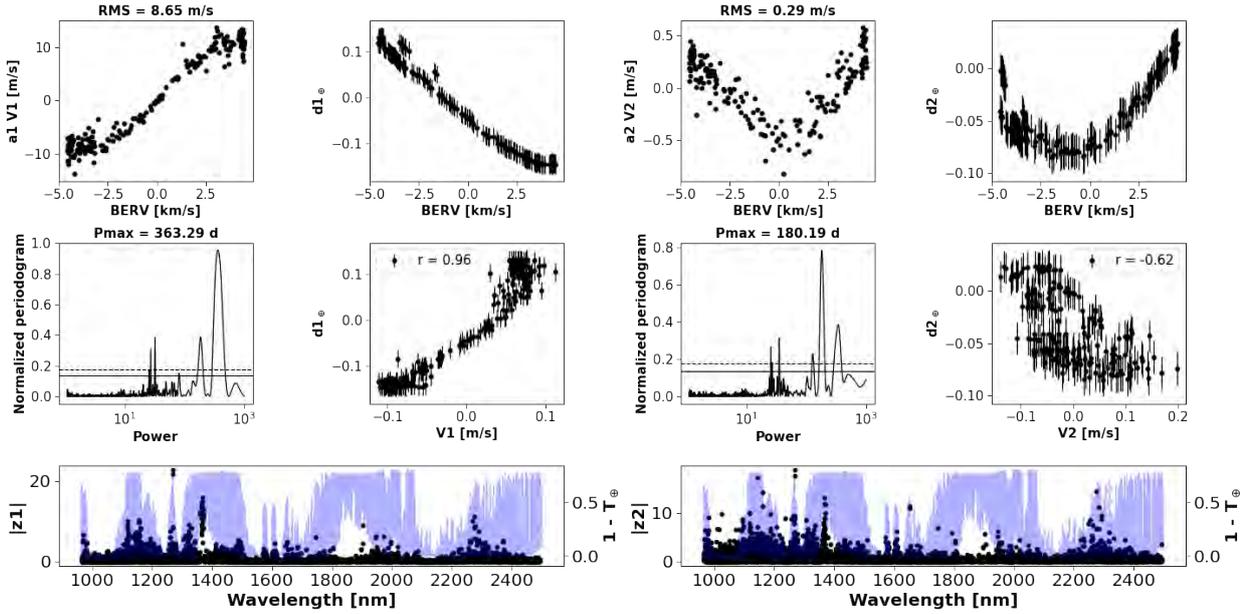


Figure 3.4: Characterizing the impact of tellurics on wPCA in the case of a narrow BERV coverage. The top panels illustrate the principal vectors in the BERV space alongside their periodogram. For each principal vector  $V_i$ , its corresponding telluric indicator  $d_i^\oplus$  is shown to the right in the BERV space, while below, its correlation with  $V_i$  is presented. Bottom panels show the absolute z-score values  $|z_i|$  in the wavelength space, with the telluric absorption lines from the telluric template overlaid as emission lines for reference.

### Variability in LBL data and systematic errors

At first, it might appear surprising that there are more principal components associated with tellurics in the case of a wide BERV coverage compared to a narrow BERV coverage, especially considering that telluric impact is more pronounced in the latter and negligible in the former. However, this observation can be explained by the same underlying reason. As illustrated in figure 3, the wider the BERV coverage, the more a telluric line can span across lines in the data. Remember that in the context of LBL, "lines" represent the interval between two consecutive maxima (Artigau et al., 2022). Consequently, a telluric line can affect more lines in the case of wide BERV coverage, albeit not frequently. This explains why the net effect is negligible when per-line RVs are averaged out while simultaneously having high variability. On the other hand, a narrower BERV coverage confines telluric lines to a smaller "territory," leading to less variability in the LBL data but a non-negligible net impact on RVs.

This result highlights a crucial distinction: the variability in LBL data and the net impact of systematics on the RV time series are two distinct aspects. In other words, having many principal components does not automatically imply the presence of strong systematics. This insight is a key takeaway from these simulations and underscores a critical aspect of the Wapiti method. It justifies the method's approach of reordering the wPCA components in decreasing Bayesian Information Criterion (BIC; Schwarz (1978)) order to ensure their relevance for the purpose of correcting systematics.

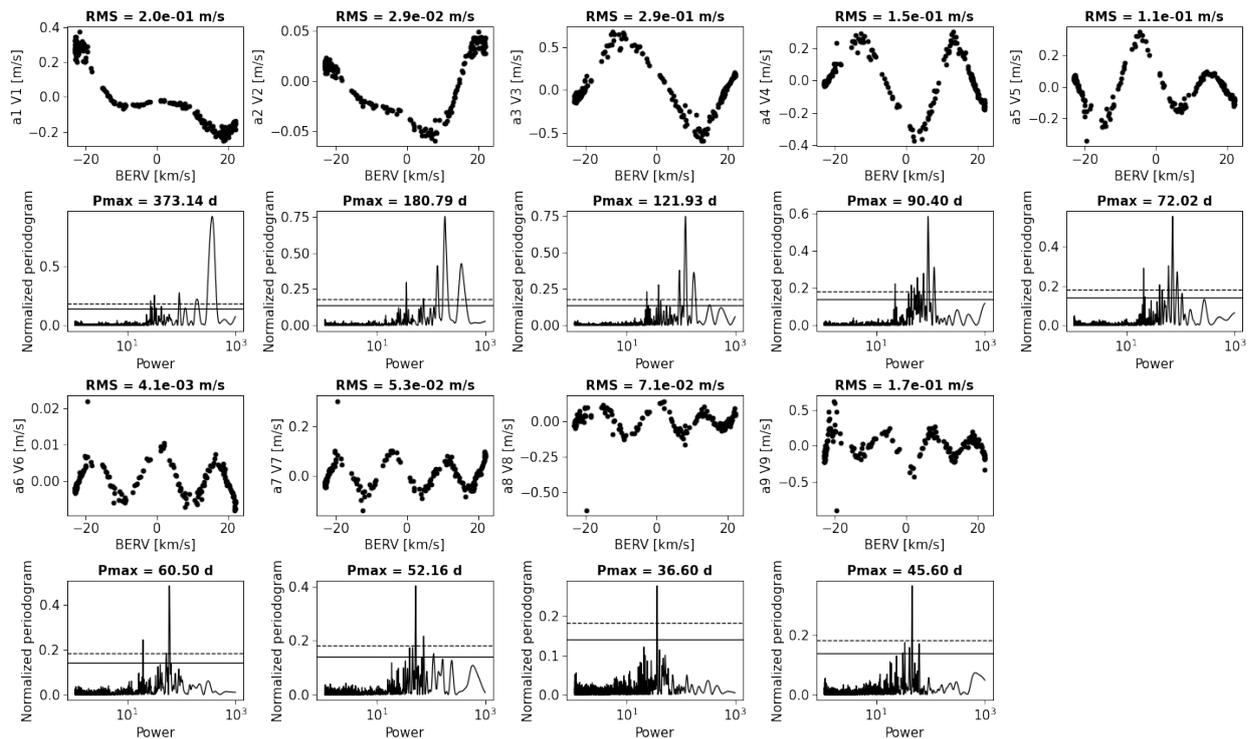


Figure 3.5: Characterizing the impact of tellurics on wPCA in the case of a wide BERV coverage. The top panels illustrate the principal vectors in the BERV space alongside their periodogram. For each principal vector  $V_i$ , its corresponding telluric indicator  $d_i^{\oplus}$  is shown to the right in the BERV space, while below, its correlation with  $V_i$  is presented. Bottom panels show the absolute z-score values  $|z_i|$  in the wavelength space, with the telluric absorption lines from the telluric template overlaid as emission lines for reference.

### Contamination of stellar activity indicators

For this very same reason, the simulations also show that stellar activity indicators can be contaminated by tellurics. This is notably the case for the differential line width (dLW) (Zechmeister et al., 2018). The dLW, computed through LBL analysis, quantifies alterations in the average width of line profiles in relation to the template (Artigau et al., 2022). In our simulations, we observed that at the 3% level of telluric contamination, this indicator is not affected in the case of a wide BERV coverage by telluric-induced systematics while it becomes affected in the case of a narrow BERV coverage. This is evident in figure 3.6, where a one-year signal emerges due to this contamination level. This situation could pose challenges when attempting to identify and characterize stellar activity within the data, as it compromises the reliability of dLW as a dependable stellar activity indicator.

Moreover, it is possible to extend the application of wPCA to the per-line dLWs, resulting in the computation of the first principal vector, denoted as  $W_1$ . This principal vector has shown promise as a stellar activity indicator, as demonstrated when studying the young active M dwarf AU Mic (Donati et al., 2023a). However, the  $W_1$  indicator is even more susceptible to telluric contamination compared to dLW. This is evident from its behavior with the BERV, showing that it is contaminated even in the case of a wide BERV coverage.

While telluric effects may not be immediately evident in the dLW time series, the  $W_1$  indicator can reveal their influence even at low levels of telluric contamination. This is achieved by capturing the component responsible for the most variance in the per-line RVs, even if the effect on the mean RV is negligible, as observed in the case of dLW for a wide BERV coverage. This finding aligns entirely with our previous results with the LBL RVs in which we emphasized that variability in the per-lines data and the net effect in the averaged data are two different things. The first cause of variability in the data does not have to be stellar activity (in fact, it may not create any at all), and this makes the  $W_1$  indicator particularly vulnerable to tellurics. Therefore, caution should be exercised when using it for activity characterization.

If stellar activity does create variability in the per-lines dLWs, it would be more advisable to check for activity signals in all the relevant  $W_i$  instead of just  $W_1$ . In the case of AU Mic, its success could be attributed to the elevated activity level of this young M dwarf, surpassing the influence of tellurics and contributing to the indicator's effectiveness by being responsible for the most variability in the data.

#### 3.4.2 Detection capacity

Now that we have a better understanding of the behaviour of the wPCA in the presence of telluric absorption lines when applied to per-line RVs, we are interested in exploring the detection capacity of this method. To do so, we first integrate a planetary signal into our simulations by adding a circular signal to the simulated per-line time series.

$$K \sin\left(\frac{2\pi t}{P} + \phi\right), \quad (3.2)$$

with  $K$  the semi-amplitude,  $P$  the orbital period and  $\phi$  the phase of the signal.

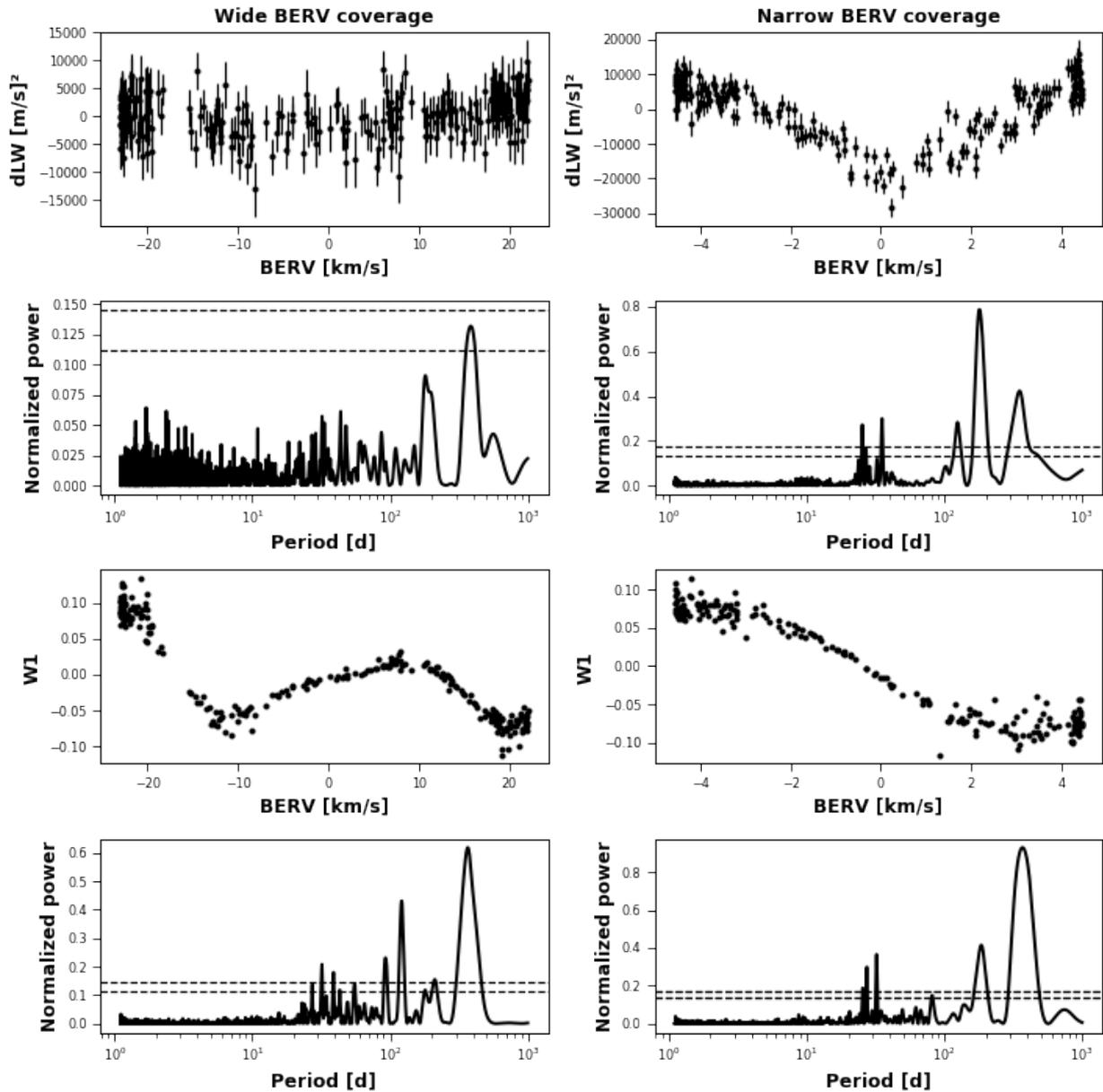


Figure 3.6: Telluric contamination of stellar activity indicators. Top plots shows the  $dLW$  versus the BERV and the respective periodograms. Bottom plots shows the  $W_1$  versus the BERV and the respective periodograms. Left and right correspond to the simulations at a 3% telluric contamination level in the case of a wide and narrow BERV coverage, respectively.

The `Wapiti` method aims to identify the most appropriate principal vectors for modeling and correcting systematics in the data through a multi-step process. However, for the purpose of gaining insights from our simulations, we find it sufficient to repeatedly use the first 2, 3, or 4 components. Specifically, for the three cases involving wide BERV coverage, narrow BERV coverage, and narrow BERV coverage with a perfect template, we compute the first 4 principal vectors, considering them as a good approximation of the telluric-induced systematics.

To explore the recovery capabilities of any wPCA-based method when it comes to detecting injected signals across a wide range of semi-amplitudes and periods, we implemented an injection-recovery test. This test involved a grid of periods  $P$  ranging from 2 to 400 days with a step of 2.5 days and a grid of semi-amplitudes  $K$  ranging from 1 to 5  $\text{m.s}^{-1}$  with 21 evenly spaced values. For each fixed combination of  $P$  and  $K$ , we performed the injection test ten times, each time introducing the signal with a random phase. The injected signal was considered detected if, after fitting the model from equation 2.2 to the data using 2, 3, or 4 components, the most prominent peak in the periodogram corresponded to the injected period with a FAP below  $10^{-5}$ . We evaluated whether this criterion was met for each of the 10 time series and computed the detection success rate for all combinations of  $K$  and  $P$ . The results are displayed in figure 3.7.

From this simple test, we can extract valuable insights. Firstly, it is important to note that the high detection rates of the injected signals near 1 year in the absence of wPCA correction for narrow BERV coverage are artifacts. These detections are the result of identifying the telluric signal rather than the planetary signal. Now, let us focus on the narrow BERV coverage cases.

As expected, we can observe that most of the injected signals can be recovered with wPCA correction, and the quality of the template has an impact on the minimum detectable amplitudes. However, there are still signals that remain undetectable, particularly those located close to periods of 1 year and 180 days. This occurs because the wPCA, at those periods, incorrectly removes the injected planetary signals, which are similar to the signals present in the principal vectors.

In a similar fashion, when applying wPCA correction to the wide BERV coverage data, we observe an inability to detect signals that have periods close to the signals in the principal vectors being used. This is why, as more components are employed, an increasing number of telluric harmonics become undetectable. In the case of a narrow BERV coverage, one might wonder why we do not see the same pattern, with undetectable signals remaining close to only 1 year and 180 days with increasing number of components. The reason for this lies in what was mentioned earlier, that principal vectors above the second component are essentially noise in this case, devoid of any associated signals.

In our scenarios, since we explore the wPCA method in the context of telluric contamination, it generates principal vectors with periodicities only corresponding to harmonics of 1 year. However, this outcome is not limited to telluric contamination; it is a general observation. Any planetary signal with an orbital period close to the periods associated with a principal vector will be erroneously subtracted when applying equation 2.2.

This presents a clear limitation in the case of a narrow BERV coverage, but it may seem not

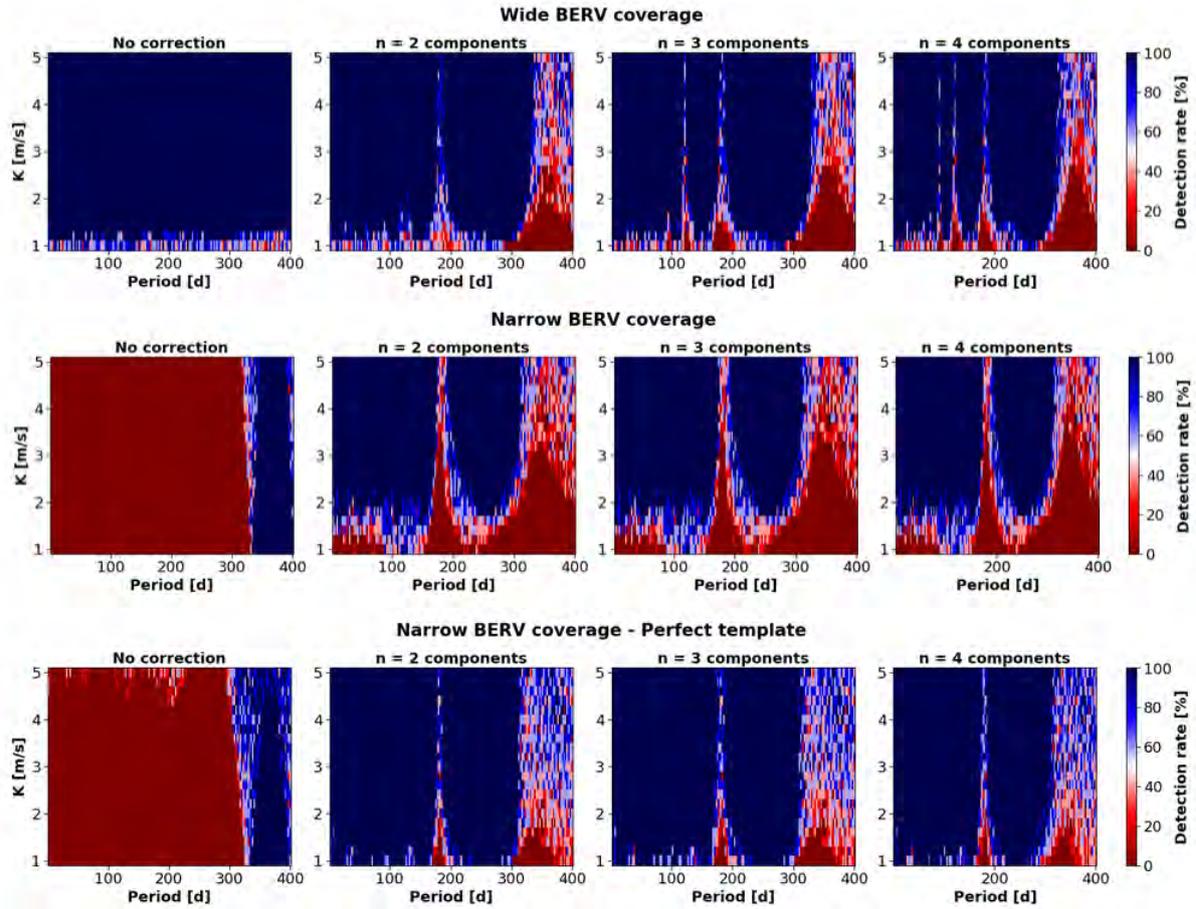


Figure 3.7: Detection capacity of the wPCA in case of telluric contamination using GLS periodograms. Top panels: detection rate from the injection-recovery test in the wide BERV coverage case using 2, 3 or 4 components. Middle panels: detection rate from the injection-recovery test in the narrow BERV coverage case using 2, 3 or 4 components. Top panels: detection rate from the injection-recovery test in the narrow BERV coverage case with a perfect template using 2, 3 or 4 components.

as critical for a wide BERV coverage since no wPCA correction is needed in the first place. However, we would contend that it remains a significant drawback. If a planetary signal at 1 year or its harmonics were detected in RV data, the most cautious assumption would be to suspect that it is of telluric origin, regardless of the quality of the telluric correction or the width of the BERV excursion.

Therefore, the wPCA method might still be applied, and as we have emphasized, even in cases of negligible impact on RV data from telluric lines, wPCA can capture telluric signals due to their high variability in the case of a wide BERV coverage. Consequently, the signal suspected of being of telluric origin would be corrected, potentially causing a real planetary signal to remain undetected. Nevertheless, this potential issue can be alleviated by changing our detection method and opting for an alternative periodogram.

#### Optimizing the detection method in the wPCA framework

This issue arises because subtracting the model in equation 2.2 (referred to as  $\mu_0$ ) causes the systematic error to substitute itself for the planetary signal. Thus, the most straightforward method is to give a better model to the data, namely by also fitting the following model  $\mu_P$  for each period P.

$$\gamma_S + \sum_{i=1}^N a_i V_i + A \sin(2\pi t/P) + B \sin(2\pi t/P). \quad (3.3)$$

As an example, let us consider three scenarios: A) no planetary signal, B) a planetary signal injected with a semi-amplitude  $K = 2 \text{ m.s}^{-1}$  and orbital period  $P_b = 20$  days, and C) a planetary signal injected with a semi-amplitude  $K = 2 \text{ m.s}^{-1}$  and orbital period  $P_b = 365.25$  days (both with a phase  $\phi = 0$ ). We will fit the two models (with  $P = P_b$ ) to the three cases using only the first principal vector, which has a 1-year periodicity. We separate cases B and C into B0/C0 and BP/CP, representing the parameters obtained when fitting  $\mu_0$  and  $\mu_P$  to B and C, respectively. The parameter values obtained from weighted-least squares regression are available in table 3.1.

We can see that in the case of a wide BERV coverage,  $a_1$  normally has a value of  $-2.7 \text{ m.s}^{-1}$  (which corresponds to the RMS of  $0.2 \text{ m.s}^{-1}$  in the data previously seen in figure 3.5). This value is retained when fitting either  $\mu_0$  or  $\mu_P$  to B. However, we can see that it goes up to  $-23 \text{ m.s}^{-1}$  when fitting  $\mu_0$  to C, showing that indeed the systematic error is wrongly considered as the origin of the 1-year signal. When fitting  $\mu_P$  to the data, it becomes lower - although still not at the original value - when fitting  $\mu_P$  to the data. We will return to the issue of obtaining the true values of the parameters, but for our detection purposes, this result is already significant.

We can define a Bayesian periodogram (Delisle et al., 2018) that compares the two models  $\mu_0$  and  $\mu_P$  at each period by computing their respective BIC ( $BIC_0$  and  $BIC(P)$ ) and computing the logarithm of the Bayes Factor  $\log BF(P) = \frac{BIC_0 - BIC(P)}{2}$ . A value above 5 can be considered statistically significant. We compute this periodogram in the context of scenario C, as shown in figure 3.8, and we can see that compared to using the GLS periodogram, we are still able to detect the 1-year signal when applying wPCA correction.

Parameters	A	B0	BP	C0	CP
$a_1$ (m. s <sup>-1</sup> )	-2.7	-2.7	-2.7	-23	8.7
$A$ (m. s <sup>-1</sup> )	-	-	1.92	-	3.1
$B$ (m. s <sup>-1</sup> )	-	-	0.09	-	-0.08

Table 3.1: Table of the parameters for scenarios A, B and C in the simulations. The parameters are estimated through WLS regression.

We thus perform the same injection-recovery test again, but this time using the log BF periodograms, we consider having a detection if the injected signal has a log BF above 5. The results can be seen in figure 3.8, and they show that the issue of being unable to detect planetary signals with orbital periods close to the signals in the systematics is almost completely resolved besides low-amplitude signals that would remain elusive.

This demonstrates the overall success of wPCA-based methods in mitigating telluric signals in LBL data. However, we can still observe that 1-year signals of planetary origin may remain undetectable in some cases. In the case of a wide BERV coverage for instance, the more components we use, the less capable we are of detecting a 1-year signal of planetary origin. Therefore, the effectiveness of any wPCA-based method, such as *Wapiti*, depends on its ability to identify and use the minimal number of appropriate components.

Furthermore, the intended use of wPCA is to explain variability, and we are repurposing it for correcting systematics. So, besides telluric-induced systematics, if there are systematics affecting the RV time series that does not induce significant variability within the LBL RV data, wPCA will not capture it, and consequently, it will not be corrected. An achromatic signal would be an example of such a systematic, and we will see more examples in the next chapter.

### 3.4.3 Recovering the injected signals

We have shown the high detection capacity of wPCA-based methods but in addition to detecting the signal, our goal is to accurately characterize it by recovering the correct injected parameters. As shown in table 3.1, when the planetary signal and the systematics share a similar periodicity, even if the coefficients adjust themselves, the resulting value is significantly different from the injected value (approximately 3 m.s<sup>-1</sup> instead of the injected 2 m.s<sup>-1</sup>).

To test our characterization capacity when using wPCA, we injected signals into in the wide BERV coverage scenario for 21 periods linearly sampled from 2 to 400 days. The injected signals had fixed semi-amplitude and phase values of 2 m.s<sup>-1</sup> and 0, respectively.

To retrieve the parameters, we used the adaptive MCMC sampler *samsam*<sup>1</sup>. The sampler was configured to perform a total of 100,000 iterations, and we used the final 75,000 iterations for statistical analysis. The prior of the semi-amplitude was set to an uniform distribution ranging from 10<sup>-1</sup> to 10<sup>5</sup> m.s<sup>-1</sup>.

---

<sup>1</sup><https://gitlab.unige.ch/Jean-Baptiste.Delisle/samsam>

### 3.4. WPCA AND TELLURIC LINES

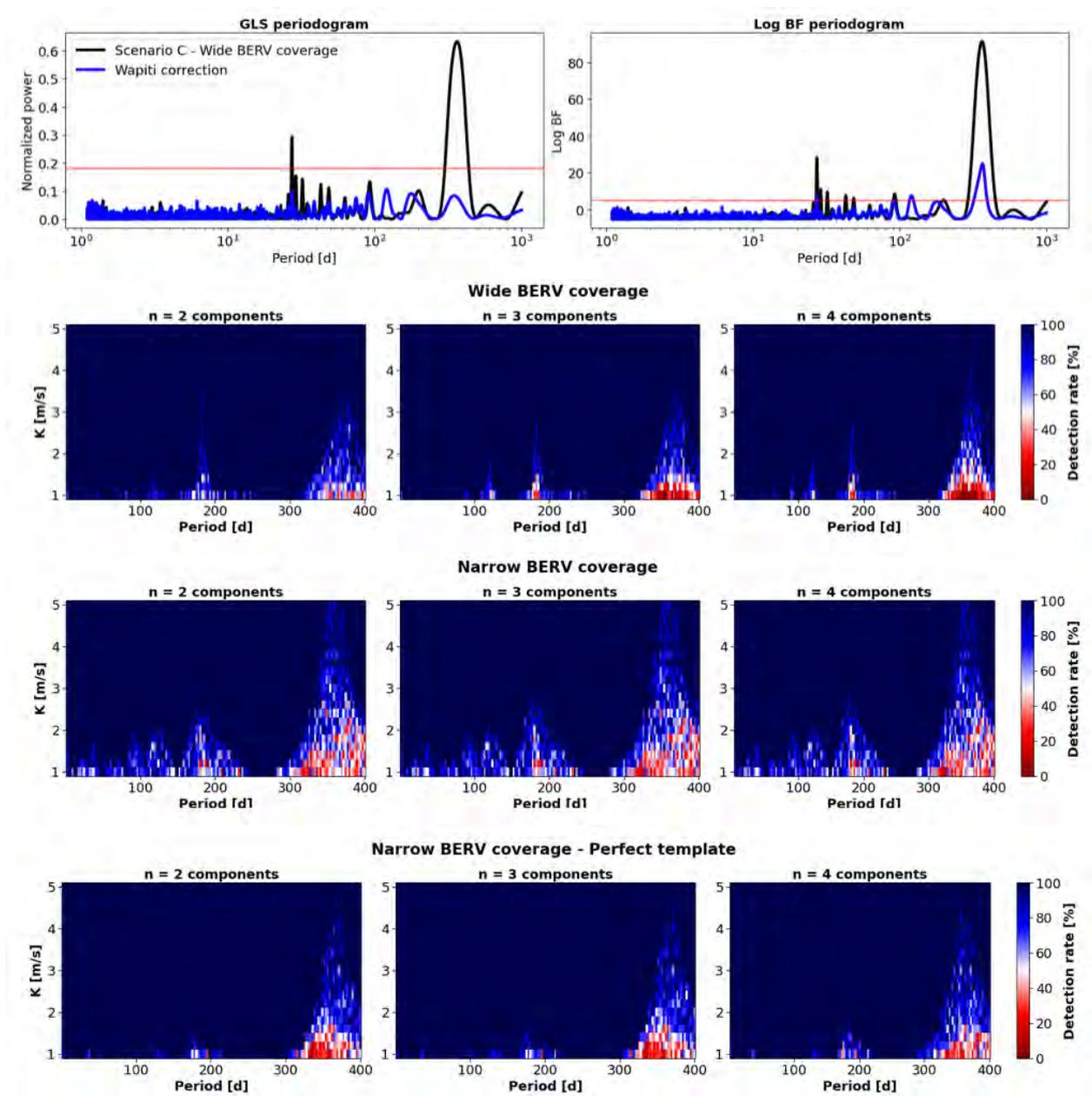


Figure 3.8: Detection capacity of the wPCA in case of telluric contamination using log BF periodograms. Top panels: detection rate from the injection-recovery test in the wide BERV coverage case using 2, 3 or 4 components. Middle panels: detection rate from the injection-recovery test in the narrow BERV coverage case using 2, 3 or 4 components. Bottom panels: detection rate from the injection-recovery test in the narrow BERV coverage case with a perfect template using 2, 3 or 4 components.

The results are displayed in figure 3.9. As expected, the recovered semi-amplitude tends to skyrocket when the injected signal has a period near 1 year and 180 days. For that reason, when characterizing signals, we change the  $V_i$  basis into  $V_i^*$  with a simple procedure: once we have our detected signals, we check if any combination of them is present within a systematic  $V_i$ . If it is the case, it is subsequently subtracted from the systematic, leading to a new  $V_i^*$  vector. We can see that in the case of a wide BERV coverage in figure 3.9, this results in a recovered semi-amplitude which is stable and much closer to the one injected.

Finally, to further assess our characterization capacity, we tested our ability to recover different semi-amplitude values at fixed periods, 20 and 365.25 days, to test the two cases of having a period not present in the systematic and conversely a period in the systematic. We did that for 11 values of semi-amplitudes linearly sampled from 2 to 30 m.s<sup>-1</sup>, and the results can be seen in figure 3.9 using the  $V_i^*$  vectors. We can observe that in the case of a 20-day injected period, the semi-amplitude is recovered within uncertainties. However, in the case of a 1-year signal, even if the retrieved semi-amplitude became stable from substituting  $V_i$  with  $V_i^*$ , there still exists an offset  $\leq 1$  m.s<sup>-1</sup>.

We conducted the exact same experiment in the case of a narrow BERV coverage, and the results are shown in figure 3.10. Even when using the updated  $V_i^*$  vectors, we still observe an increase in the retrieved semi-amplitude in the region of telluric harmonics. This is due to the fact that in this case, the systematics have a non-negligible impact on the RV time series. When we inject a fixed period at 1 year, our characterization is limited by the semi-amplitude of the 1 year period in the  $V_1$  vector, and the retrieved values stagnate at this value if the injected semi-amplitude is lower.

This demonstrates that even though wPCA methods have a high detection capacity, it is still crucial to correct for systematic effects beforehand as best as possible since their presence in the data can limit our ability to accurately characterize signals.

## 3.5 Summary

The simulations provided valuable insights into the application of wPCA to per-line RVs in the context of telluric contamination, as well as more general lessons.

In summary, one of the most important takeaways, as previously emphasized, is that wPCA is primarily designed to capture variability in the data rather than systematics in the averaged data. Therefore, when using wPCA to estimate systematics, it is crucial to keep this in mind to avoid misinterpreting the results. For example, a high number of components does not necessarily indicate that a large number of systematics are affecting the data. Additionally, outliers and low SNR observations can contaminated the principal vectors from the wPCA. As a rule of thumb, before applying wPCA to the data, it would be advisable to ensure that at least the first few principal vectors are not contaminated by outliers. If they are, those epochs should be rejected from the analysis.

However, wPCA-based methods demonstrate a high detection capacity, and when used in conjunction with log BF periodograms, even in the case of a narrow BERV coverage, it becomes possible to detect almost all injected signals, including those near telluric harmonics. Nevertheless, while detection is feasible, characterizing the signal is more challenging, espe-

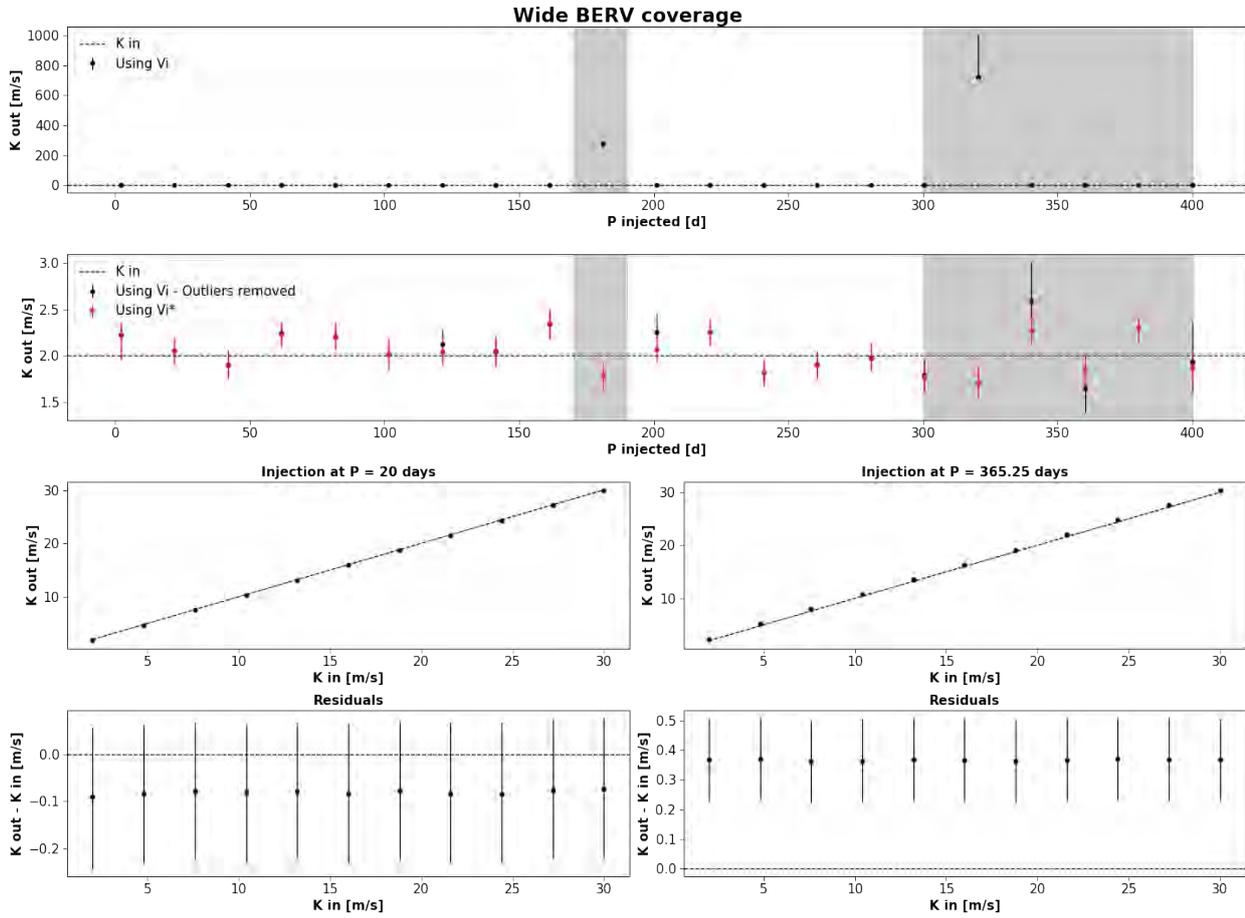


Figure 3.9: Characterization capacity of the wPCA in case of telluric contamination in the wide BERV coverage scenario. Top panel shows the recovered semi-amplitude from the MCMC sampling for different injected periods when using  $V_i$ . Below is the same plot but with the outliers removed for clarity, and the retrieved semi-amplitudes when using the  $V_i^*$  vectors are also overlaid in red. The region near 180 days and 1 year are highlighted in grey. The third row shows the retrieved semi-amplitude when injecting different semi-amplitudes at 20 days (left) and 365.25 days (right), with the residuals shown below it.

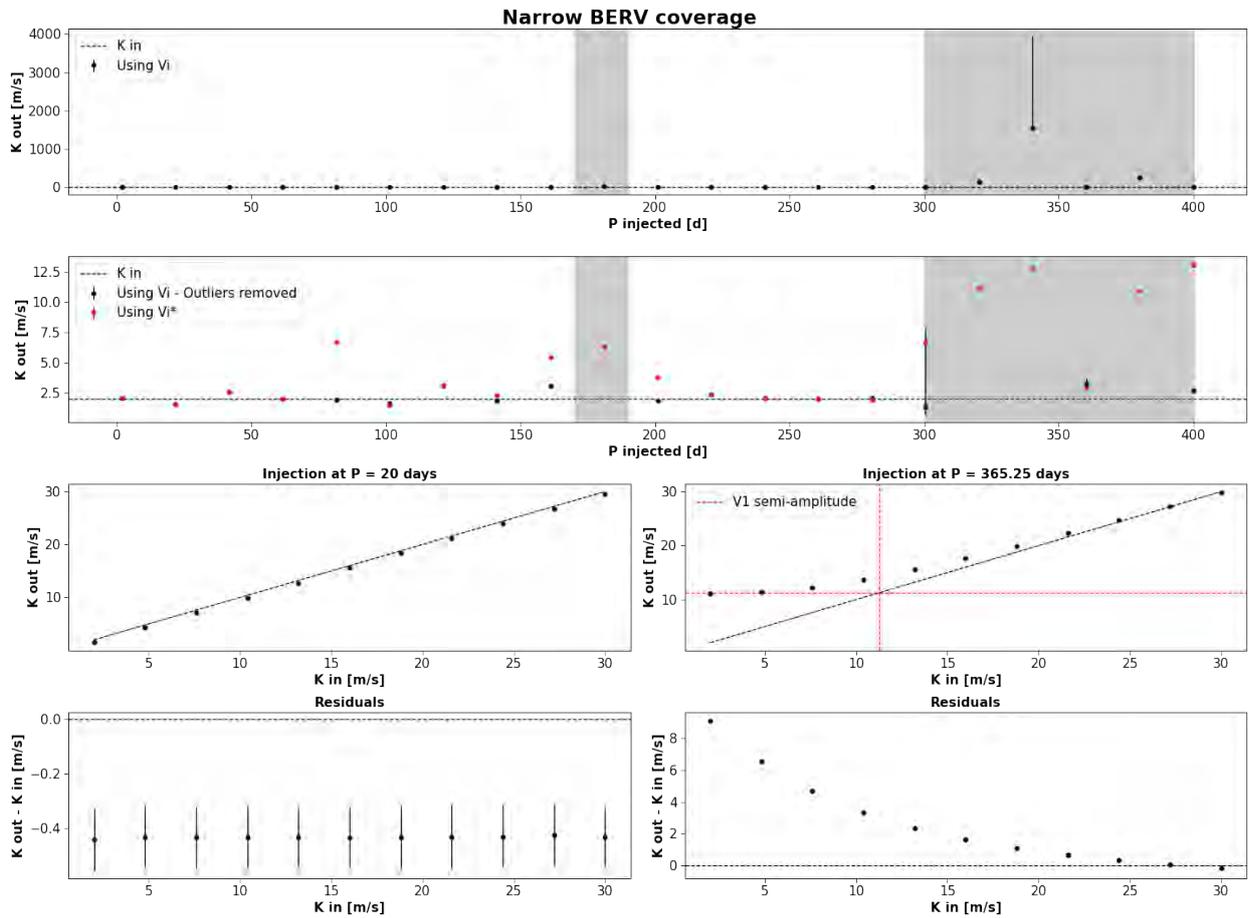


Figure 3.10: Characterization capacity of the wPCA in case of telluric contamination in the narrow BERV coverage scenario. Top panel shows the recovered semi-amplitude from the MCMC sampling for different injected periods when using  $V_i$ . Below is the same plot but with the outliers removed for clarity, and the retrieved semi-amplitudes when using the  $V_i^*$  vectors are also overlaid in red. The region near 180 days and 1 year are highlighted in grey. The third row shows the retrieved semi-amplitude when injecting different semi-amplitudes at 20 days (left) and 365.25 days (right), with the residuals shown below it.

cially if the signal is at a period present within one of the systematics. This may result in inaccurate estimations of the semi-amplitude. Even if a better estimate can be obtained by subtracting the signals from the systematics in which they are present, the characterization remains limited by the effective impact of the systematic on the data. This underscores the importance of mitigating systematics as effectively as possible before relying solely on wPCA for their correction and detection.

It is also essential to acknowledge that our simulations represent a simplified model for studying wPCA under controlled conditions. These simulations do not account for various complexities present in real observations, such as the potential seasonal or chromatic variations in the tellurics through the  $\alpha$  which we took to be random, the influence of OH and Na emission lines, the effects of Moon contamination (Crepp et al., 2016), or changes in the blaze function that can be caused by instrument alterations and varying incident light conditions (Halverson et al., 2015). Additionally, we simplified the star spectrum as simply the star template with photon noise, neglecting the possibility of more complex changes occurring within observations and leading to a more intricate interplay with telluric lines. Finally, other sources of systematics could be present in real data and the principal components could be a mixture of those effects with tellurics since systematics are not necessarily mutually orthogonal with each other, hindering the effectiveness of wPCA-based methods.



# Chapter 4

## Integrating Wapiti within the SPIRou Legacy Survey - Planet Search program

### Contents

---

4.1	Overview . . . . .	63
4.2	Confirming signals present in the data with Wapiti . . . . .	64
4.3	Detecting new and hidden signals in the data with Wapiti . . .	66

---

### 4.1 Overview

In the previous chapters, we have developed the `Wapiti` method, aimed at effectively mitigating the influence of systematics on RV measurements. Specifically, we found in chapter 3 using simulations that wPCA-based methods (`Wapiti` included) are successful in correcting systematics stemming from telluric lines. This capability is particularly valuable when analyzing time-series data in the near-infrared (nIR) due to the abundance of tellurics in this domain (Bean et al., 2010), and even the most advanced telluric correction methods leave telluric residuals, giving rise to telluric-related systematics (Sameshima et al., 2018; Ulmer-Moll et al., 2019).

Nevertheless, real data are more complex than our simulations and now that the theoretical effectiveness of the `Wapiti` method has been demonstrated, it needs to be more widely integrated within the analysis of other targets from the SPIRou Legacy Survey - Planet Search (SLS-PS) program than Gl 251.

Consequently, we applied this tool to various targets from the SPIRou Legacy Survey - Planet Search program (SLS-PS)<sup>1</sup>. In this manuscript, we will present the results obtained with the current version of `Wapiti` and of `APER0` (v.0.2786) on targets that effectively demonstrate its capabilities and limitations.

---

<sup>1</sup><https://spirou-legacy.irap.omp.eu/doku.php>

## 4.2 Confirming signals present in the data with Wapiti

In order to ensure that **Wapiti** behaves as intended, the next step after applying it to Gl 251 and unveiling its known 14.2d planet (Stock et al., 2020) was to apply it to data in which a signal of planetary origin was already discernible in non-corrected data and verify that **Wapiti** does not suppress Keplerian signals, which is what we mean by "confirming" signals.

We applied it on two different targets, Gl 480 and Gl 382, which are the subjects of a submitted paper in A&A (Ould-Elhkim et al. 2023b), and details on the RV measurements can be found in appendix C.2.2.

### 4.2.1 Gl 480

Gl 480 (Wolf 433) is a red dwarf star with spectral type M3.5V (Lépine et al., 2013) that is located  $14.2625 \pm 0.0060$  pc away from Earth (Gaia Collaboration, 2020). A planet orbiting at a  $9.567 \pm 0.005$  d orbital period was claimed to orbit Gl 480 from the analysis of HIRES and HARPS data (Feng et al., 2020) The signal has a semi-amplitude  $K = 6.80 \pm 0.87$  m.s<sup>-1</sup> with an estimated  $M_p \sin i$  of  $13.2 \pm 1.7 M_{\oplus}$ .

#### Wapiti correction

We applied the **Wapiti** method as detailed in chapter 2, including the permutation test, chromatic sensitivity, and re-ordering of the components. The results are displayed in figure 4.1, which showcases all the steps of the **Wapiti** correction, including the permutation test, chromatic sensitivity, and re-ordering of the components. Additionally, we include the median of the absolute Pearson coefficients to highlight how this value is not as well-behaved in comparison to the chromatic sensitivity.

We also present three time series in figure 4.1: the original RV time series without correction, the one resulting from the bervzp correction, and the **Wapiti**-corrected RV time series. We employed the log BF periodograms introduced in chapter 3 to identify periodicities in the time series. In all cases, the known 9.5-day signal is detected in the data, even without applying correction. However, we can observe that its significance decreased after bervzp correction while it increased after applying the **Wapiti** correction. This demonstrates that the **Wapiti** correction not only preserves existing planetary signals in the data but can also enhance their statistical significance by mitigating systematics. This enhancement is illustrated by the disappearance of various peaks in the periodograms compared to the other two time series.

This target is an important step in demonstrating the proper performance of the **Wapiti** method on real data sets. It showcases the ability of the method to increase the statistical significance of an already known signal, ultimately resulting in its first detection in the nIR. A more detailed RV analysis of this target using data from other instruments can be found in appendix C.2.2.

### 4.2.2 Gl 382

Gl 382 (V\* AN Sex) is a red dwarf of spectral type M2.0V (Kirkpatrick et al., 1991) located  $7.707 \pm 0.0015$  pc away from Earth (Gaia Collaboration, 2020). In Sabotta et al. (2021), a

## 4.2. CONFIRMING SIGNALS PRESENT IN THE DATA WITH WAPITI

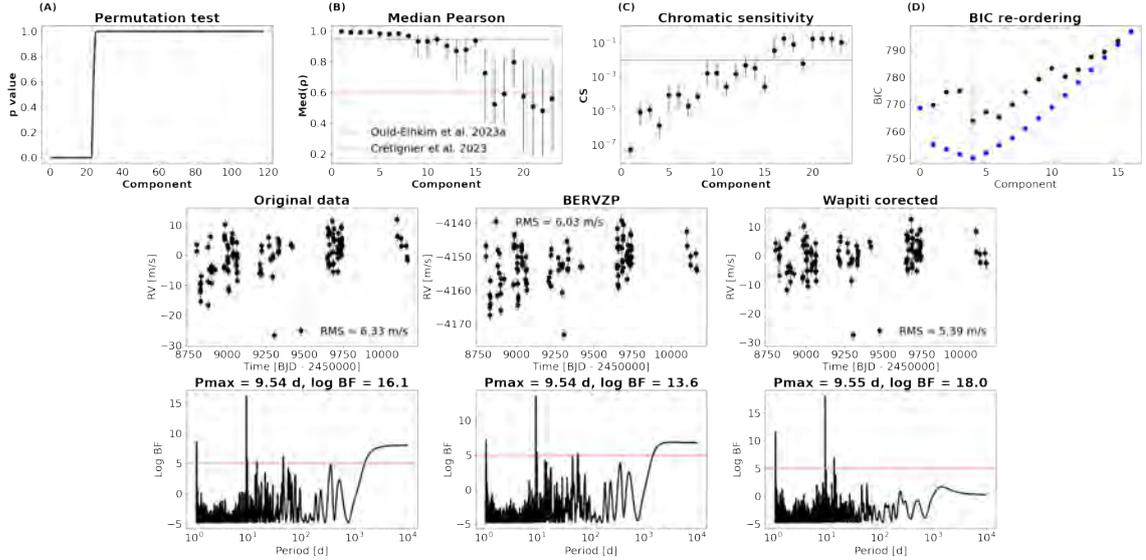


Figure 4.1: **Wapiti** correction of Gl 480. Panel A: Results from the permutation test. Panel B: Median absolute Pearson coefficients from the leave-p-out cross-validation test. Panel C: Chromatic sensitivity of each component from the leave-p-out cross-validation test. Panel D: Components re-ordered in decreasing BIC order. Bottom panels show the time series with their respective log BF periodograms below it for the original time series, the bervzp corrected time series, and the **Wapiti** corrected RV time series.

signal with a period of 10.65 days was detected in CARMENES optical RV data. However, the signal was automatically flagged as spurious due to this period being close to the first harmonic of the rotational period. This star was also part of the activity study conducted by [Mignon et al. \(2023\)](#) from HARPS spectra, and they did not find any signs of an activity cycle in the chromospheric activity indices.

### Wapiti correction

We applied the **Wapiti** correction on this target, and the outcomes are presented in figure 4.2. It is notable that the 10.4-day signal is already discernible in the original RV dataset. The signal’s presence in the nIR with a strength comparable to that in the optical data could provide evidence that this signal likely originates from a planetary source rather than being attributed to stellar activity.

Regarding the two correction methods, the bervzp correction diminishes the signal’s statistical significance while leaving various spurious peaks in the data. In contrast, the **Wapiti** method substantially boosts the signal’s statistical significance and cleans the data by eliminating these additional peaks. This confirms the method’s effectiveness compared to the bervzp correction and affirmed the presence of a planetary signal in the data which had been claimed previously. The comprehensive joint nIR and optical RV analysis of this target is accessible in appendix C.2.2.

The successful application of the **Wapiti** method to Gl 480 and Gl 382 served as a demonstration of its reliability on real data other than Gl 251. Consequently, we proceeded to the next step beyond confirming signals, which is the detection of previously unknown signals.

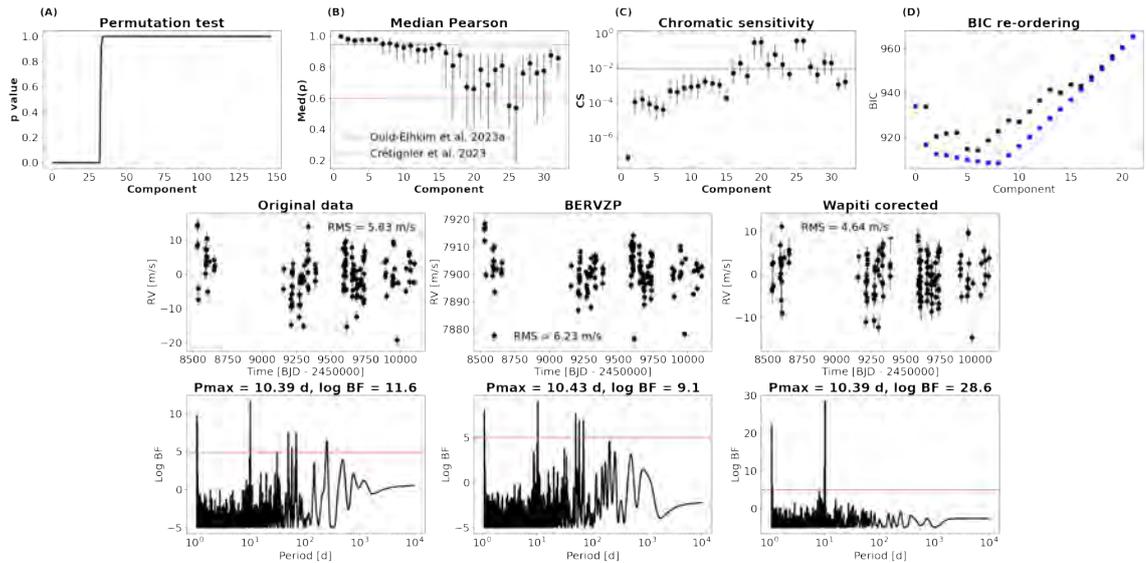


Figure 4.2: Same as figure 4.1 for Gl 382.

### 4.3 Detecting new and hidden signals in the data with Wapiti

We then proceeded to apply the `Wapiti` method to Gl 725 B, which, as discussed in chapter 3, is a target with a narrow BERV coverage ( $\leq 5 \text{ km.s}^{-1}$ ). This characteristic renders this target susceptible to telluric contamination, providing a good opportunity for assessing the effectiveness of the `Wapiti` method.

#### 4.3.1 Gl 725 B

Gl 725 B (Struve 2398 B, HD 173740) is a M5.0V dwarf star (Kirkpatrick et al., 1991), positioned at a distance of  $3.5231 \pm 0.0003 \text{ pc}$  from Earth, as reported by (Gaia Collaboration, 2020). A summary of its physical properties is available in Table 4.1. Additionally, this star has been the subject of longitudinal magnetic field ( $B_\ell$ ) time-series data collection, conducted using circularly polarized spectropolarimetry. These observations resulted in the identification and characterization of a rotational period for Gl 725 B, estimated to be  $135 \pm 15 \text{ days}$  (Donati et al., 2023b).

Gl 725 B was observed using the SPIROU instrument, which employed a series of circular polarization sequences. The observational campaign extended from February 14, 2019, to April 19, 2022. After night-binning the data, we acquired 208 epochs of measurements. However, we decided to exclude measurements with a SNR below 100 for the reason detailed in chapter 3, resulting in a final dataset consisting of 203 measurements.

It is important to note that Gl 725 B is part of a binary system, with its companion being Gl 725 A, classified as an M3V dwarf star (Kirkpatrick et al., 1991). The separation between these two stars is estimated to be on average 56 AU and an orbital period of 295 years (Heintz, 1987). A potential planet with a period of 2.7 days was previously reported around Gl 725 B. However, it is important to note that this signal was considered to be comparable

Parameters	Gl 725 B	Ref.
$T_{\text{eff}}$ (K)	$3379 \pm 31$	1
$[M/H]$	$-0.28 \pm 0.10$	1
$[\alpha/\text{Fe}]$	$0.14 \pm 0.04$	1
$\log g$	$4.82 \pm 0.06$	1
Radius ( $R_{\odot}$ )	$0.280 \pm 0.005$	1
Mass ( $M_{\odot}$ )	$0.25 \pm 0.02$	1
$\log (L/L_{\odot})$	$-2.038 \pm 0.003$	1
$P_{\text{rot}}$ (d)	$135 \pm 15$	2

Table 4.1: Stellar parameters of Gl 725 B. **1:** [Cristofari et al. \(2022\)](#), **2:** [Donati et al. \(2023b\)](#).

to the noise floor ([Berdiñas et al., 2016](#)).

### Wapiti correction

We applied the `Wapiti` correction to Gl 725 B dataset, and the results can be found in figure 4.3. The correction process was similar to the previous targets, with one difference. During the BIC re-ordering process, we accounted for the presence of the orbiting companion Gl 725 A by incorporating a trend in the base model. This led to the selection of 12 components, a considerable increase compared to the 2 components identified in our simulations. This suggests that there are other sources of variability in the data beyond those arising from simulated telluric absorption lines.

Nevertheless, in figure 4.3, the RV data still exhibit contamination from telluric signals at periods of 180 days, 1 year, and 120 days. Notably, the `bervzp` correction manages to partially reduce the statistical significance of the 180 days signal, but it remains insufficient. Conversely, the `Wapiti` correction effectively eliminates these telluric signals, revealing a new 4.77-day signal in the data. Based on our extensive work with simulations, which demonstrated the detection capabilities of `wPCA`, combined with the success in recovering known signals in Gl 251, Gl 480, and Gl 382, we interpret this newly discovered signal as being of planetary origin.

Further analysis of this signal will be presented in a forthcoming paper ([Ould-Elhkim et al. 2023c](#), in preparation). However, for the purposes of this manuscript, we will conduct an RV analysis and characterization of this target to showcase the utility of the `Wapiti` method in the context of real data.

### Signals detection in Gl 725 B

We adopt an iterative approach to detect signals in Gl 725 B, where we systematically add significant signals in the model and then recompute a log BF periodogram. We continue this process until the maximum peak in the periodogram remains above 5 ([Delisle et al., 2018](#)). To be more precise, with  $N_p$  representing the number of fitted planets, we compare the models:

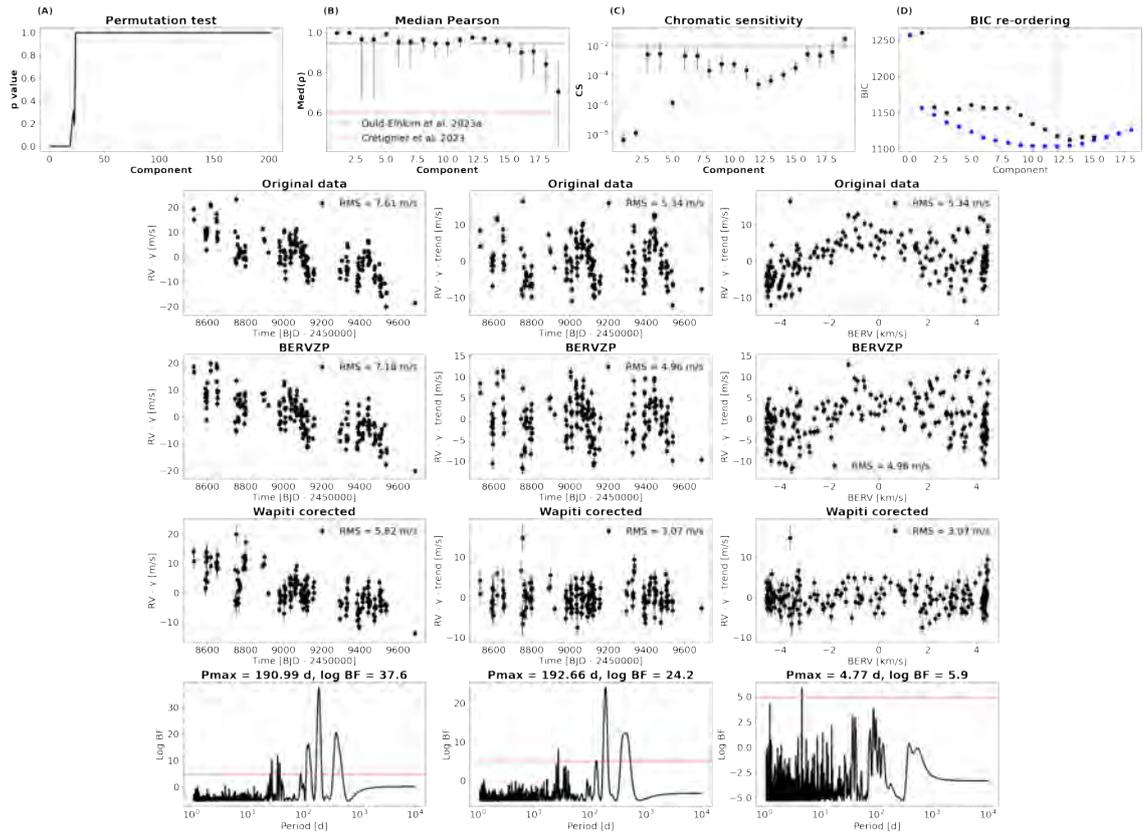


Figure 4.3: Same as figure 4.1 for Gl 725 B and trend correction included.

$$\begin{aligned}
 M_0 &= \gamma_S + \alpha \cdot t \\
 M_{0,N_p} &= M_0 + \sum_{i=1}^{N_{comp}} \hat{V}_i + \sum_{i=1}^{N_p} A_i \sin(2\pi t/P_i) + B_i \sin(2\pi t/P_i) \\
 M_{P,N_p} &= M_{0,N_p} + A \sin(2\pi t/P) + B \sin(2\pi t/P).
 \end{aligned} \tag{4.1}$$

In this equation,  $M_0$  represents the base model, which includes a constant term  $\gamma_S$  and a trend parameter  $\alpha$ . The  $\hat{V}_i$  vectors are the  $N_{comp}$  principal vectors determined by the `Wapiti` method, selected based on their ability to minimize the Bayesian Information Criterion (BIC) when added to the base model  $M_0$ .

Consequently, we included the 4.77-day signal in the model  $M_{0,N_p}$  and compared it to the model  $M_{P,N_p}$ , resulting in a new log BF periodogram (figure 4.4). This revealed another signal at 38 days, prompting us to add this signal to the model as well. With these additions, no more residuals remained in the data. However, the `Wapiti` method does not stop at this stage. Once the two signals were detected, we incorporated them into the base model  $M_0$  when reordering the components to potentially obtain a better estimate of the systematics. This refinement reduced the number of components from 12 to 10. We then applied this updated model to search for any remaining signals, but no statistically significant signals were found in the residuals.

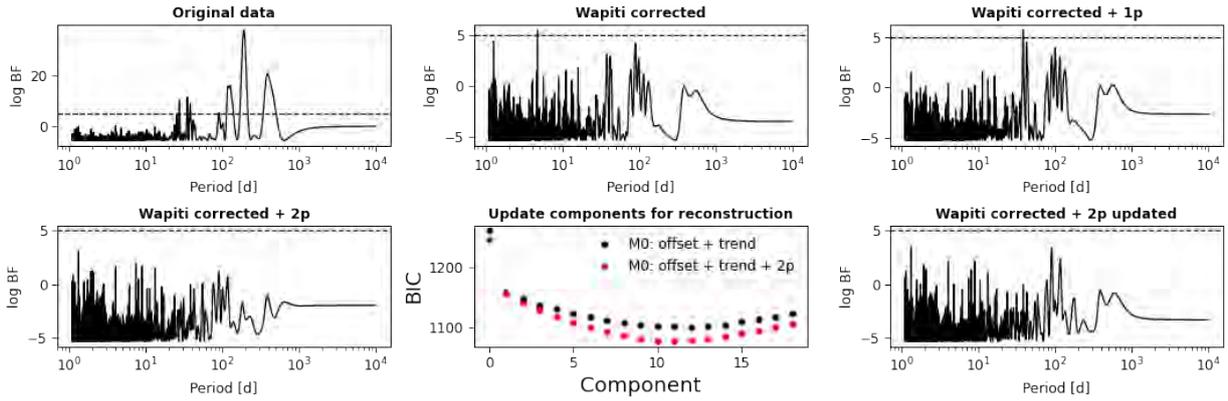


Figure 4.4: Iterative search for signals in Gl 725 B. From left to right: Periodograms of the original data; periodograms of the *Wapiti* corrected data with a 4.77 days signal as the most prominent peak; periodograms of the *Wapiti* corrected data with the first signal included and with a 38 days signal as the most prominent peak; periodograms of the residuals with the two signals included with no more significant signal; re-ordering of the components when the two signals are included in the base model; periodograms of the updated model with still no remaining significant signals in the data.

### Signals characterization in Gl 725 B

To characterize the detected signals, we followed the procedure outlined in chapter 3. We examined whether these signals were present in any of the 10 systematics represented by  $V_i$ . We did not detect the 4.77-day signal in any of the systematics. However, the 38-day signal was detected in  $V_8$ . As a result, we proceeded to fit and subtract this signal from this principal vector, resulting in a basis  $V_i^*$  better adapted for characterization.

To explore the architecture of this multiplanetary system, we used the adaptive MCMC sampler *samsam* (Delisle et al., 2018). We tried two models for the planets: one with eccentric orbits and another with circular orbits. For both models, we ran the sampler for a total of 100,000 iterations and analyzed the final 75,000 iterations for statistical purposes. The results are presented in table 4.2 for the eccentric model and in table 4.3 for the circular model.

From the MCMC results we can observe that in the eccentric model the eccentricities of the planets are not well estimated in this analysis, which is understandable considering their low amplitudes relative to the precision capacity of SPIRou. Nevertheless, it is probable that tidal dissipation have circularized their orbits, and we can note that the orbital parameters are compatible in both the eccentric and circular models. Therefore, we opt to use this model to showcase the final model in figure 4.5, where we include the phase-folding of the RV curve by using the parameter values maximizing the likelihood of this model. We also show the architecture of the system using the module *rebound* (Rein & Liu, 2012).

The minimum mass of these planets increases inside-out, ranging from  $1.5 \pm 0.3 M_\oplus$  for Gl 725 Bb to  $3.4 \pm 0.7 M_\oplus$  Gl 725 Bc. According to the classification by Mishra et al. (2023), the architecture of this system falls into the "ordered" class. It is worth noting that this class was identified as relatively rare in simulations, with a 1.5% occurrence rate. In contrast, their

<b>Wapiti Parameters</b>			
$a_1$ (m. s <sup>-1</sup> )	$44.4^{+3.2}_{-3.0}$	-	$\mathcal{U}(-\infty, \infty)$
$a_2$ (m. s <sup>-1</sup> )	$-18.5^{+3.0}_{-3.1}$	-	$\mathcal{U}(-\infty, \infty)$
$a_3$ (m. s <sup>-1</sup> )	$14.7^{+3.2}_{-3.0}$	-	$\mathcal{U}(-\infty, \infty)$
$a_4$ (m. s <sup>-1</sup> )	$-16.7^{+3.1}_{-3.1}$	-	$\mathcal{U}(-\infty, \infty)$
$a_5$ (m. s <sup>-1</sup> )	$-13.1^{+3.1}_{-3.0}$	-	$\mathcal{U}(-\infty, \infty)$
$a_6$ (m. s <sup>-1</sup> )	$-11.2^{+3.0}_{-3.0}$	-	$\mathcal{U}(-\infty, \infty)$
$a_7$ (m. s <sup>-1</sup> )	$-11.1^{+3.0}_{-3.1}$	-	$\mathcal{U}(-\infty, \infty)$
$a_8$ (m. s <sup>-1</sup> )	$-12.8^{+3.4}_{-3.2}$	-	$\mathcal{U}(-\infty, \infty)$
$a_9$ (m. s <sup>-1</sup> )	$-9.6^{+3.5}_{-3.2}$	-	$\mathcal{U}(-\infty, \infty)$
$a_{10}$ (m. s <sup>-1</sup> )	$-10.2^{+3.5}_{-3.2}$	-	$\mathcal{U}(-\infty, \infty)$
<b>Instrumental Parameters</b>			
$\gamma_{SPIROU}$ (m. s <sup>-1</sup> )	$235.6^{+24.5}_{-37.8}$	-	$\mathcal{U}(-\infty, \infty)$
$\sigma_{SPIROU}$ (m. s <sup>-1</sup> )	$2.72^{+0.18}_{-0.15}$	-	$\mathcal{U}(10^{-3}, 10^2)$
<b>Stellar companion</b>			
$\alpha$ (m. s <sup>-1</sup> .d <sup>-1</sup> )	$-0.0198^{+0.0005}_{-0.0008}$	-	$\mathcal{U}(-\infty, \infty)$
<b>Eccentric model</b>			
Orbital parameters	b	c	
$P$ (days)	$4.765^{+0.003}_{-0.003}$	$38.23^{+0.09}_{-0.14}$	$\mathcal{U}(10^{-1}, 10^5)$
$K$ (m. s <sup>-1</sup> )	$1.52^{+0.33}_{-0.36}$	$1.72^{+0.33}_{-0.36}$	$\mathcal{U}(10^{-1}, 10^5)$
$\lambda(0)$ (deg)	$3.1^{+33.4}_{-34.3}$	$4.3^{+28.4}_{-17.7}$	$\mathcal{U}(0, 360)$
$e \cos \omega$	$-0.018^{+0.080}_{-0.141}$	$0.046^{+0.220}_{-0.078}$	-
$e \sin \omega$	$-0.054^{+0.075}_{-0.174}$	$0.037^{+0.161}_{-0.071}$	-
<b>Derived Parameters</b>			
$e$	$0.14^{+0.18}_{-0.10}$	$0.13^{+0.21}_{-0.10}$	$\mathcal{B}(0.867, 3.03)^{(1)}$
$\omega$ (deg)	$-81^{+172}_{-62}$	$29^{+62}_{-99}$	$\mathcal{U}(0, 360)$
$M_p \sin i$ (M <sub>⊕</sub> )	$1.5^{+0.3}_{-0.4}$	$3.5^{+0.7}_{-0.7}$	-
$a$ (AU)	$0.0350^{+0.0009}_{-0.0010}$	$0.140^{+0.004}_{-0.004}$	-
T <sub>eq</sub> (K)	$430^{+8}_{-8}$	$215^{+4}_{-4}$	-

Table 4.2: Orbital and derived planetary parameters from the MCMC algorithm applied on RV data of Gl 725 B. (1): [Kipping \(2013\)](#)

<b>Wapiti Parameters</b>			
$a_1$ (m. s <sup>-1</sup> )	$44.9^{+3.1}_{-3.3}$	-	$\mathcal{U}(-\infty, \infty)$
$a_2$ (m. s <sup>-1</sup> )	$-17.9^{+3.1}_{-3.0}$	-	$\mathcal{U}(-\infty, \infty)$
$a_3$ (m. s <sup>-1</sup> )	$14.5^{+3.2}_{-3.2}$	-	$\mathcal{U}(-\infty, \infty)$
$a_4$ (m. s <sup>-1</sup> )	$-16.5^{+3.2}_{-3.2}$	-	$\mathcal{U}(-\infty, \infty)$
$a_5$ (m. s <sup>-1</sup> )	$-13.0^{+3.0}_{-3.1}$	-	$\mathcal{U}(-\infty, \infty)$
$a_6$ (m. s <sup>-1</sup> )	$-11.2^{+3.1}_{-3.1}$	-	$\mathcal{U}(-\infty, \infty)$
$a_7$ (m. s <sup>-1</sup> )	$-11.0^{+3.1}_{-3.0}$	-	$\mathcal{U}(-\infty, \infty)$
$a_8$ (m. s <sup>-1</sup> )	$-12.1^{+3.4}_{-3.3}$	-	$\mathcal{U}(-\infty, \infty)$
$a_9$ (m. s <sup>-1</sup> )	$-9.8^{+3.4}_{-3.5}$	-	$\mathcal{U}(-\infty, \infty)$
$a_{10}$ (m. s <sup>-1</sup> )	$-10.5^{+3.1}_{-3.2}$	-	$\mathcal{U}(-\infty, \infty)$
<b>Instrumental Parameters</b>			
$\gamma_{SPIRou}$ (m. s <sup>-1</sup> )	$220.8^{+40.7}_{-41.2}$	-	$\mathcal{U}(-\infty, \infty)$
$\sigma_{SPIRou}$ (m. s <sup>-1</sup> )	$2.74^{+0.18}_{-0.15}$	-	$\mathcal{U}(10^{-3}, 10^2)$
<b>Stellar companion</b>			
$\alpha$ (m. s <sup>-1</sup> .d <sup>-1</sup> )	$-0.0201^{+0.0009}_{-0.0009}$	-	$\mathcal{U}(-\infty, \infty)$
<b>Circular model</b>			
Orbital parameters	b	c	
$P$ (days)	$4.765^{+0.003}_{-0.003}$	$38.21^{+0.17}_{-0.16}$	$\mathcal{U}(10^{-1}, 10^5)$
$K$ (m. s <sup>-1</sup> )	$1.43^{+0.31}_{-0.32}$	$1.62^{+0.33}_{-0.34}$	$\mathcal{U}(10^{-1}, 10^5)$
$\lambda(0)$ (deg)	$2.5^{+39.1}_{-37.3}$	$8.2^{+32.1}_{-32.8}$	$\mathcal{U}(0, 360)$
-			
<b>Derived Parameters</b>			
$M_p \sin i$ (M <sub>⊕</sub> )	$1.5^{+0.3}_{-0.3}$	$3.4^{+0.7}_{-0.7}$	-
$a$ (AU)	$0.0350^{+0.0009}_{-0.0010}$	$0.140^{+0.004}_{-0.004}$	-
T <sub>eq</sub> (K)	$430^{+8}_{-8}$	$215^{+4}_{-4}$	-

Table 4.3: Same as table 4.2 for a circular model.

observations sample indicates a higher prevalence, with 37% of observed systems falling into this class. Therefore, our findings support the idea that ordered planetary systems may be more common than predicted by simulations.

Finally, based on the work of [Kopparapu et al. \(2014\)](#) and using the stellar parameters from table 4.2, we can calculate that for an Earth-mass planet, the habitable zone (HZ) of this star spans from 0.09 to 0.19 AU. Consequently, Gl 725 Bc, which is located at a distance of  $0.140 \pm 0.004$  AU, falls within the habitable zone of its host star, making it a potentially habitable exoplanet.

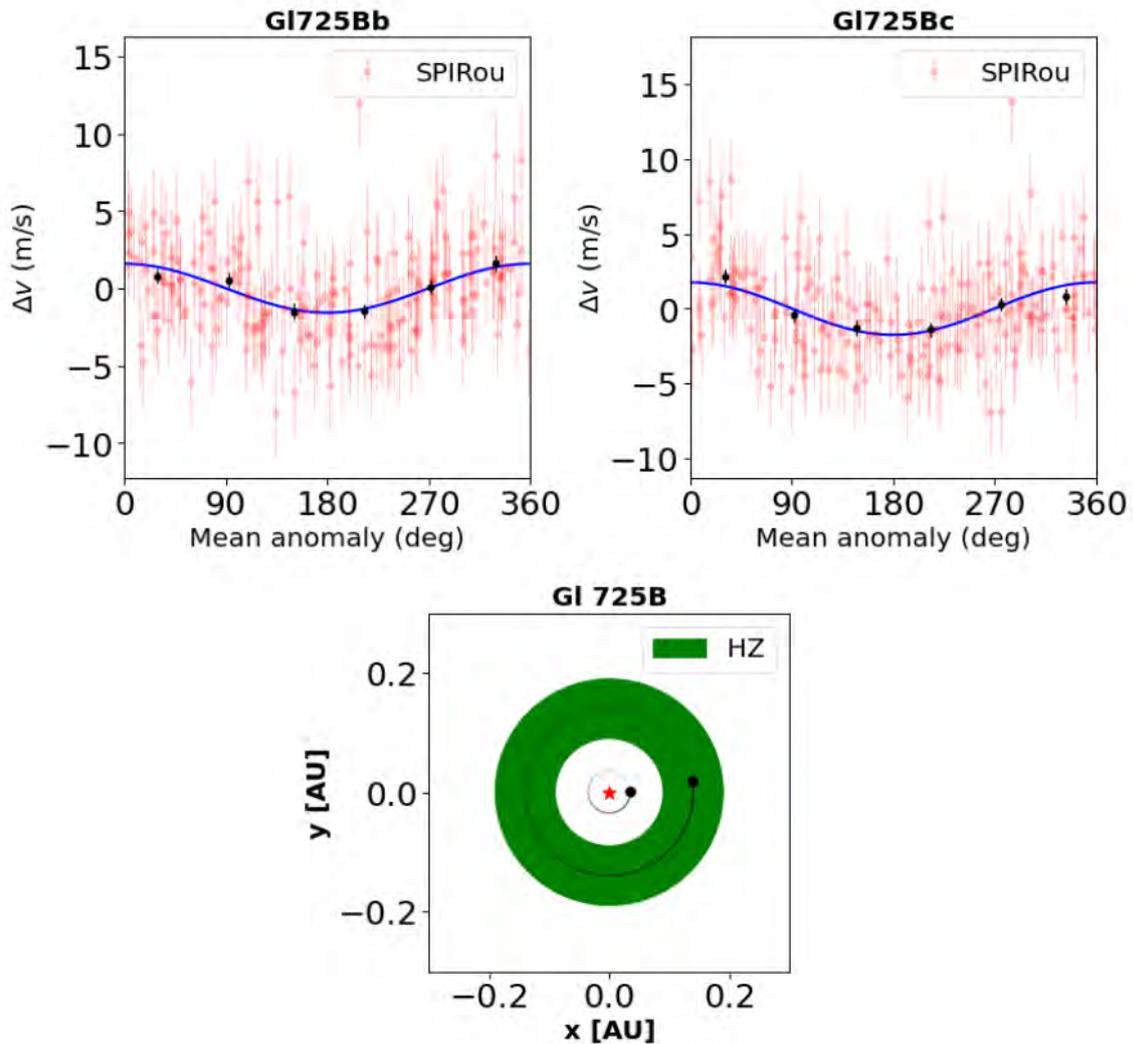


Figure 4.5: Phase-folding of the SPIRou RV data of Gl 725 B. Top panels show the phase-folding of SPIRou LBL RVs data at the orbital period of Gl 725 Bb (left) and Gl 725 Bc (right). The bottom panel is a simple visualisation of the system architecture with `rebound` using the parameters from the circular model.

In summary, the application of the `Wapiti` method to Gl 725 B demonstrated its effectiveness in mitigating telluric systematics in SPIRou data, even in challenging cases with small BERV coverage. Additionally, it successfully detected new signals in this target, illustrating how the

### 4.3. DETECTING NEW AND HIDDEN SIGNALS IN THE DATA WITH WAPITI

---

**Wapiti** method can be seamlessly integrated into the SLS-PS program to fulfill its promise of discovering exoplanets with  $\text{m.s}^{-1}$  precision.



# Chapter 5

## Conclusion and perspectives

### Contents

---

5.1 Summary of the thesis . . . . .	75
5.2 Perspectives . . . . .	80

---

### 5.1 Summary of the thesis

This thesis focused on enhancing the accuracy of RV measurements in the SPIRou instrument and addressing systematic errors, particularly those caused by telluric lines. This effort resulted in the development of the `Wapiti` algorithm, which is designed to identify and correct for the underlying systematic effects in the data using wPCA.

We introduced the current version of the algorithm and tested it on simulations as well as four targets from the SLS-PS program, achieving encouraging results. The algorithm successfully revealed the previously concealed 14.2-day signal in Gl 251, which was hidden by systematic errors in SPIRou data. Additionally, our application on Gl 480 and Gl 382 showed that even when signals were already apparent in the data, `Wapiti` effectively corrected for systematic errors while preserving the planetary signal and enhancing its statistical significance.

Lastly, we applied the `Wapiti` method to Gl 725 B. This target was posing a significant challenge for RV analysis due to its minimal BERV excursion, making it particularly unfavorable for telluric correction methods. Consequently, the RV data were affected by strong systematics arising from telluric residuals. However, the `Wapiti` method effectively addressed these systematics, simplifying the process of detecting and characterizing previously undiscovered planetary signals, including one located within the habitable zone of its host star.

The confirmed and detected planetary signals from this thesis are summarized in the mass-period plot shown in figure 6.1. In this plot, we have also included the mass and periods of known planets within the SLS sample, as listed in Table 1 of [Moutou et al. \(2023\)](#). The planet parameters for this plot were taken from the mentioned references, with the exceptions of Gl 1148 and Gl 876 ([Moutou et al., 2023](#)), Gl 251 ([Ould-Elhkim et al., 2023](#)), Gl 382

and Gl 480 (from the submitted paper in appendix C.2.2). Additionally, the parameters of Gl 725 Bb and Gl 725 Bc are derived from Table 4.3.

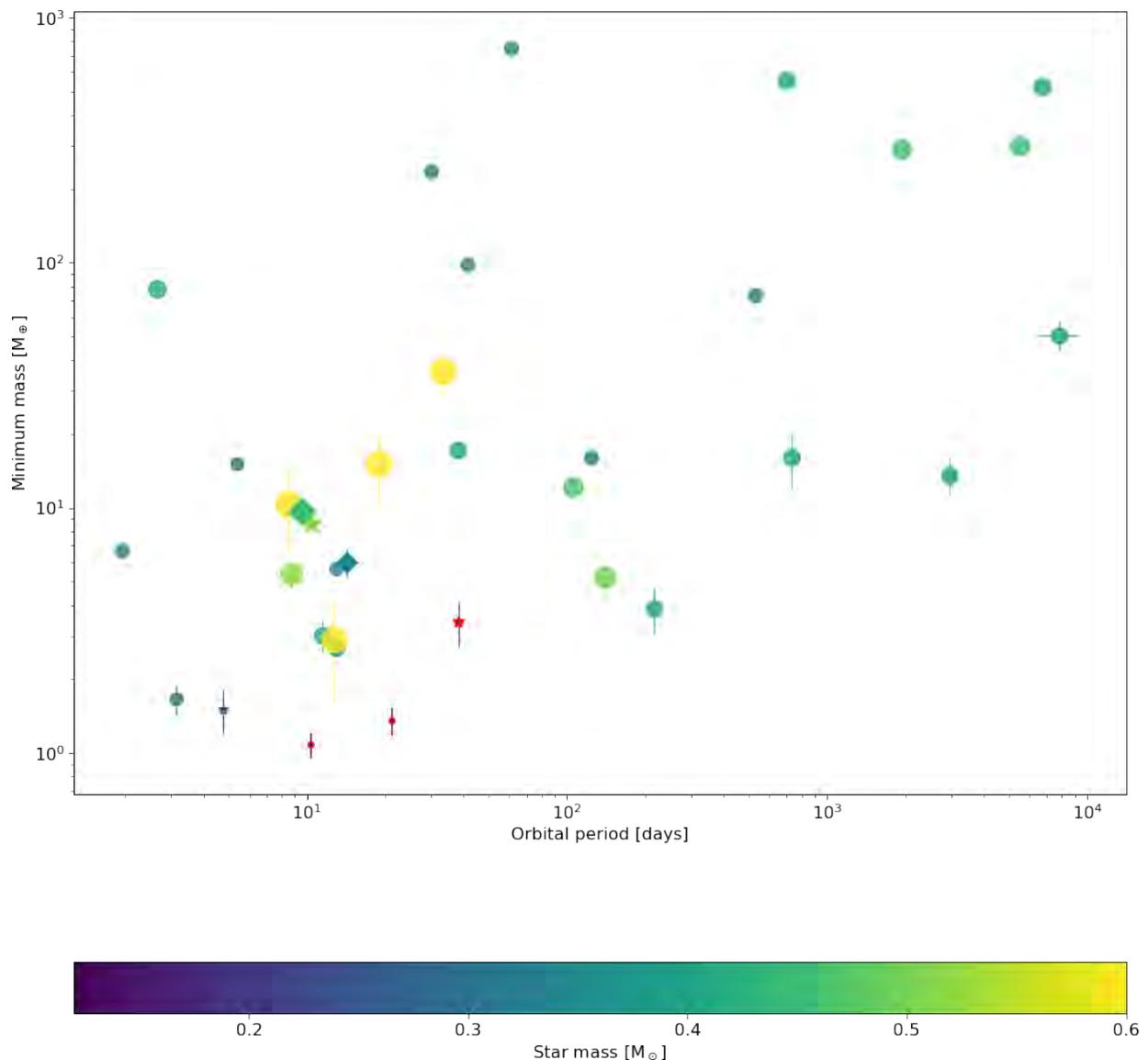


Figure 5.1: Detected and confirmed signals in this thesis in the context of the SLS-PS program. Confirmed planetary detections from this work are represented by diamond-shaped markers, while previously unclaimed detections are indicated by star-shaped markers and planets in the HZ of their star are outlined in red. The markers are color-coded with respect to the mass of the host star and their size is also proportionate to this same mass.

In the mass-period plot depicted in figure 6.1, confirmed planetary detections from this thesis are represented by two diamond-shaped markers, while previously unclaimed detections are indicated by three star-shaped markers. Notably, only two planets (Gl 1002b and Gl 1002c) were originally located within the HZ of their host stars in the SLS-PS sample. Thus, the detection of Gl 725 Bc is of significant importance as it increases the number of habitable

worlds in this sample (Planets in the HZ of their star are outlined in red in figure 6.1).

Furthermore, it is observed that Gl 725 Bb is one of the least massive planets in this sample and has one of the shortest orbital periods, allowing for further exploration of this parameter space. Regarding the other detections, they remain at relatively low masses ( $\leq 10 M_{\oplus}$ ). This bias is due to the focus of this study on stars that exhibit high levels of systematics, which tend to conceal low-amplitude planets. Targets with strong planetary signals that surpass the influence of systematics were not thoroughly investigated in this study.

Drawing definitive conclusions about the exoplanet population within the SLS-PS sample based on our new detections would be obviously premature. Nevertheless, this thesis concluded at an exciting stage: the `Wapiti` method can now be readily applied to the targets of the SLS-PS sample and help in their individual analysis. Ongoing efforts are in progress to apply this method to targets beyond those discussed in this study, with the aim of uncovering new planetary systems in upcoming publications (A. Carmona et al. in preparation and P. Cortes-Zuleta et al. in preparation). In the future, thanks to the `Wapiti` method developed during this thesis, more planets will hopefully be added to the parameter space in figure 6.1, furthering our knowledge of nearby worlds in the neighborhood of the Sun. The code is accessible on GitHub<sup>1</sup>, accompanied by a tutorial to assist users in its application.

As a way to conclude with what we started, we applied `Wapiti` to all the targets that constitute the nightly zero point. As shown in figure 6.2, the behavior initially observed in the  $V_{\text{tot}}$  space, which led to the development of the BERVZP correction, has now disappeared. Consequently, it appears that the BERVZP correction may no longer be required, although it remains valuable for comparison with `Wapiti`-corrected data. All the more so as the `Wapiti` method itself has inherent limitations, as outlined in the next section.

### 5.1.1 Capacity and limitations of the `Wapiti` method

Despite the achievements outlined, it is important to acknowledge certain limitations associated with this method, which typically arise from its reliance on wPCA as its core technique.

#### wPCA and systematics

The `Wapiti` method uses wPCA to estimate data systematics but wPCA is meant for capturing variance rather than these effects. For instance in our simulations, with wide BERV coverage, telluric influences led to many principal components and increased data variability, even when systematics were negligible. In contrast, in the scenario with limited BERV coverage, data variability decreased, leading to fewer components, while the systematics became more noticeable at  $10 \text{ m.s}^{-1}$  level.

This is a key aspect to consider when using our method, since most of its capability and limitation stems from it. For example, due to this tendency to capture variability, wPCA is highly sensitive to outliers and low SNR observations and it may be necessary to exclude some measurements in certain situations.

Furthermore, if there is a systematic effect in the data that does not introduce significant variability, wPCA will not detect it, and as a result, `Wapiti` will not correct for it. For

---

<sup>1</sup><https://github.com/HkmMerwan/wapiti>

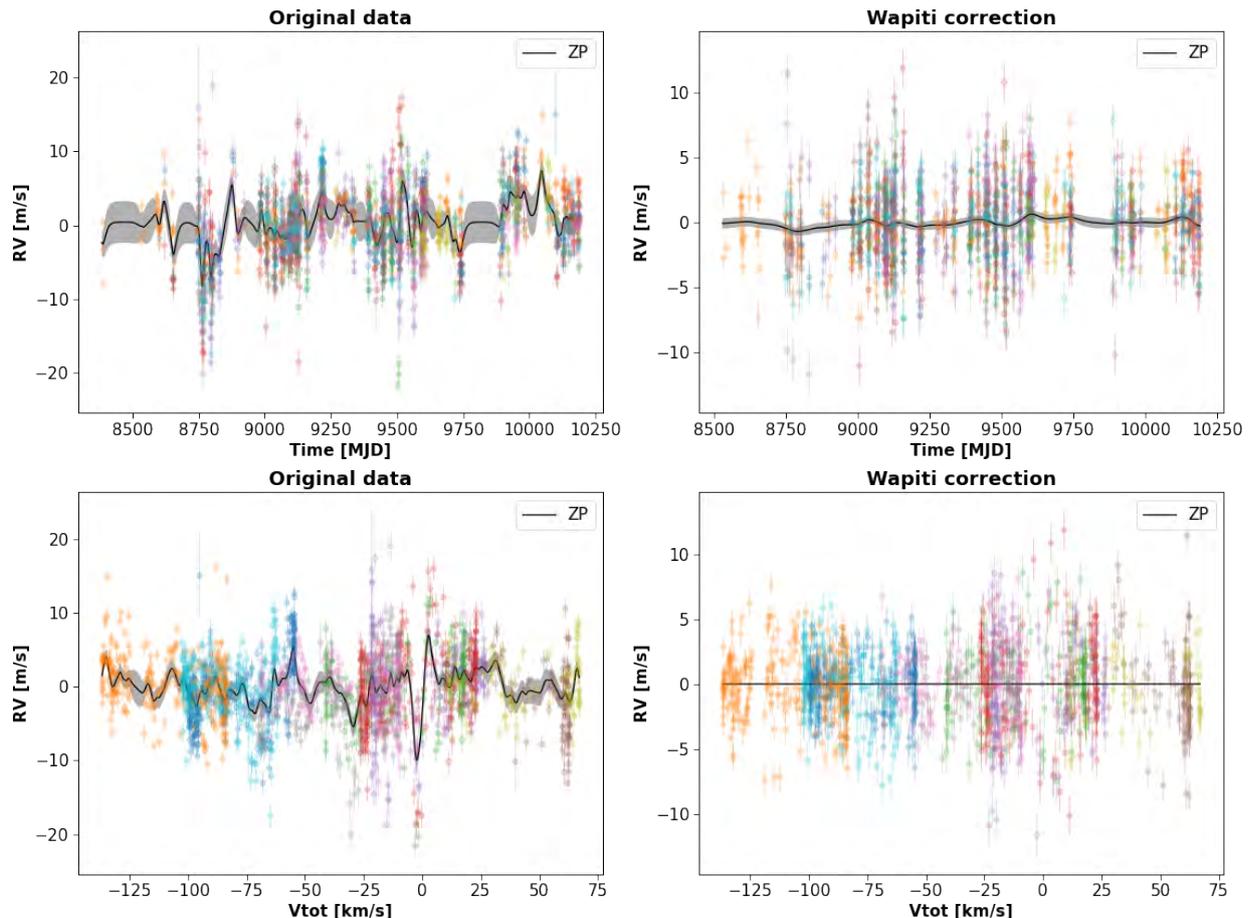


Figure 5.2: RV standards in time and  $V_{\text{tot}}$  spaces before and after *Wapiti* correction. Top panels show the RV of standards stars that make up the zero-point in night space, while bottom panels show them in the  $V_{\text{tot}}$  space. The left panels are the original RV data and the right panels are the RV after *Wapiti* correction.

example, achromatic errors fall into this category, which is why it is still necessary to address instrumental effects such as the FP drift or the nightly zero point before using this method.

Lastly, the wPCA process constructs orthogonal principal components by definition. However, real-life systematic effects may not be orthogonal, and this can lead to the wPCA mixing different systematics together, making them harder to interpret.

### Characterization of detected signals

The detection ability of *Wapiti*, and wPCA correction methods in general, has demonstrated high effectiveness in simulations and when applied to real data. However, we encountered a challenge in the simulations when the detected signal had the same period as the systematic effects. Even after removing this period from the systematic effects, our ability to estimate the amplitude of the injected signal remained constrained by the amplitude of the systematic effects.

This highlights that while *Wapiti* can reduce systematics and reveal signals in the RV

timeseries, it is still preferable to address systematics beforehand if possible to limit their impact on the data. This underscores the importance of a high-quality telluric correction, such as the one used in **APER0**.

This is crucial because it directly implies that in the worst-case scenario where a planetary signal shares its orbital period with a signal present in the systematics, the determination of its mass and other orbital parameters could be biased by the systematics.

### **Wapiti and stellar activity**

The question of whether the **Wapiti** method can effectively mitigate stellar activity is perhaps one of its most intriguing aspects. We did not encounter such a scenario in the targets discussed in this manuscript, as stellar activity tends to be relatively weak in near-infrared RV data. However, we did apply this correction method to other targets from the SLS-PS program where activity could be observed in the RV time series. Although this work is still in its early stages, a trend is emerging. Ultimately, the answer to whether **Wapiti** can address stellar activity boils down once again to the distinction between identifying components responsible for variability and systematic effects.

We came across three different scenarios. In the first case, we observed that the method successfully corrected for stellar activity, and the components were displaying signals related to activity. In the second case, despite the presence of stellar activity, the **Wapiti** method could not correct it, and it identified zero components suitable for data reconstruction. Lastly, the third case combined aspects of the previous two, where the **Wapiti** method found components displaying activity signals that it deemed appropriate for reconstruction. However, even after removing these components from the data, complete removal of activity from the corrected RV timeseries remained a challenge.

These three scenarios can be explained by remembering that stellar activity is a term that encapsulates various effects, as discussed in the introduction. If the effect responsible for the stellar activity signal introduces variability in the LBL RV data, wPCA will capture it, and the **Wapiti** method will recognize it as suitable for reconstruction. Conversely, if stellar activity does not induce variability in the per-line RVs, the **Wapiti** method will not detect this effect in the data, potentially resulting in the use of zero components, and thus, activity will not be corrected. In cases where stellar activity in the data arises from a combination of different effects, some causing variability and others not, the method will capture the variable components and partially mitigate stellar activity but may not completely correct it.

Hence, **Wapiti** may not always correct stellar activity, but by addressing other systematic sources, it should at least leave stellar activity as the remaining concern in the data. Additionally, in cases where activity introduces variability in the LBL RV data, the principal vectors displaying activity signals could serve as indicators of activity and be used for activity correction using multi-dimensional GPs. Moreover, the z-scores introduced in chapter 3 could help identify lines most sensitive to stellar activity.

However, due to the wPCA computing orthogonal components, systematics can become intertwined, potentially mixing stellar activity with other systematic effects and making attempts to characterize activity challenging. Lastly, as seen earlier, characterizing planetary

signals can be complicated when their orbital periods are close to periods present in the systematics. Similarly, the same challenge may arise in characterizing stellar activity signals when the star has a rotational period close to periods present in the systematics. For example, some stars in the SLS-PS sample have rotational periods close to telluric harmonics of 90, 120, or 180 days (Donati et al., 2023b).

## 5.2 Perspectives

### 5.2.1 On Wapiti

#### With SPIRou

The *Wapiti* method developed during this thesis will be employed on targets observed with SPIRou as part of the SLS-PS program, with the potential to confirm, disprove, or discover new planetary systems. Moreover, its capability to reduce systematics in SPIRou data extends beyond SLS-PS targets and can be readily applied to any targets observed with SPIRou as part of other programs, including the WP2 follow-up program. However these targets have lower SNR observations, potentially introducing variability that could affect the effectiveness of wPCA in capturing systematics. On the other hand, the orbital period of the planet is already known from transit observations for these targets, offering prior information that could be used to guide the *Wapiti* method in addressing systematic errors.

#### With other instruments

The method presented in this thesis relies on LBL data for its application and this algorithm has been made compatible with numerous instruments, including those operating in the optical range, such as HARPS, HARPS-N, ESPRESSO, CARMENES, or SOPHIE<sup>2</sup>. Consequently, the *Wapiti* method can be readily used with these instruments and in the near future, it will be interesting to assess how *Wapiti* performs in the optical domain and explore the potential for joint nIR and optical RV analyses using the *Wapiti* framework.

As for the nIR domain, there are other instruments available on which the LBL can be used, such as NIRPS (which will even have its data processed with the same DRS APERO). The *Wapiti* method's ability to correct for telluric-origin systematic effects in nIR SPIRou data will be highly valuable for mitigating similar effects in NIRPS data. Additionally, another instrument set to benefit from the work done on SPIRou is its upcoming twin, SPIP<sup>3</sup>, located at the Pic du Midi observatory. The *Wapiti* method should also be applied to SPIP RV data to facilitate their analysis.

#### With more than RVs

In this study, our focus was on reducing the impact of systematics on RV data. However, the *Wapiti* method has the potential for broader applications beyond per-line RV data. In the future, it could be interesting, for instance, to compute a wPCA on the per-line differential line widths (dLWs) in addition to the per-line RVs. This could provide additional components

---

<sup>2</sup><https://github.com/njcuk9999/lbl/tree/main/>

<sup>3</sup><https://www5.obs-mip.fr/wp-content-omp/uploads/sites/10/2016/10/spip2-tbl.pdf>

for the `Wapiti` method to identify those most suitable for data reconstruction. This approach would align with the philosophy of a method like `SCALPELS` (Collier Cameron et al., 2019), which mitigates stellar activity by disentangling between shift and shape systematics. In this case, it would be achieved by using `wPCA` components associated with both the shift (per-line RVs) and the width (per-line `dLWS`) of the spectral lines.

Moreover, the `Wapiti` method is not restricted to focusing solely on reconstructing RVs; it could also be applied to per-line `dLWs` to eliminate systematics from the `dLW` time series itself, potentially revealing hidden activity signals. Furthermore, it is not limited to LBL data alone neither: as long as the objective is to remove systematics from any type of data using `wPCA`, the `Wapiti` method can be effectively employed. For instance, in the context of atmospheric spectroscopy, the `ATMOSPHERIX` pipeline employs a `PCA` on the pixels for this purpose (Klein et al., 2023; Debras et al., 2023). It could prove valuable to apply `Wapiti` in this context to assess whether it can enhance signal detection.

### Beyond `Wapiti`

We saw that most of the limitations of the `Wapiti` method came from its use of `wPCA` to identify systematics, even though that is not the primary purpose of this technique. Therefore, a natural next step would involve exploring the use of autoencoders for RV reconstruction and mitigating systematics. Substituting `PCA` with autoencoders is a common practice, and during this thesis, efforts were made to implement autoencoders in the `ATMOSPHERIX` pipeline (Klein et al., 2023; Debras et al., 2023). This experience provided insights into potential challenges when applying autoencoders to per-line RVs, including considerations such as neural network architecture, choice of loss or activation functions, and training methods.

Autoencoders offer the advantage of capturing non-linear relationships in the data, whereas `PCA` is inherently linear. However, this comes at the expense of interpretability, as the features in the latent space may not be orthogonal or ordered in a readily interpretable manner, such as by decreasing variance. Therefore, the application of `wPCA` to the data using the `Wapiti` method remains a simpler, more intuitive, and computationally efficient approach. Further research will be needed to evaluate the benefits of using an autoencoder compared to `wPCA` on per-line RV time series.

### 5.2.2 On other works

While the development of the `Wapiti` method constituted a significant portion of this thesis, other tools were also created during this research. Although they did not reach a stage where they could be integrated into the main text of the manuscript, we are presenting one of them in its preliminary form in more detail in appendix B and it will be subjects of further research in the future.

### `BERV-I` periodogram

As efficient as the `Wapiti` method is in correcting for telluric systematics in RV data, it depends on the RVs to be computed using the LBL method. While it is increasingly applied to various instruments beyond `SPIRou`, other instruments may still employ alternative methods

such as the CCF method or template matching techniques such as `harps-terra` (Anglada-Escudé & Butler, 2012), `naira` (Astudillo-Defru et al., 2017b), `wobble` (Bedell et al., 2019a), or `serval` (Zechmeister et al., 2018).

For these reasons, we developed a tool capable of distinguishing between signals of telluric origins and those from other sources. To achieve this, we took advantage of the characteristic behavior of telluric signals, which is closely linked to the BERV. Consequently, to filter out signals originating from telluric lines, we tried to design a periodogram that remains invariant to BERV variations.

The details of the BERV-I periodogram can be found in appendix B. While it showed promising results when applied to the simulations by successfully revealing injected planetary signals without requiring systematic corrections, its application to real data was less effective. This highlights the need for further investigation and refinement for it to function as intended.

### 5.2.3 Characterizing planetary systems around M dwarfs

While the development of these tools is inherently interesting and exciting, the ultimate objective is to use them to reduce systematic errors, in order to detect and characterize planetary systems revolving around M dwarf stars.

Advancements in the detection and characterization of exoplanets around low-mass stars hold significant importance. They not only contribute to our knowledge of planetary formation and evolution but also have implications for missions aiming to characterize exoplanet atmospheres. These missions include ongoing projects such as the JWST as well as future missions (ARIEL, E-ELT).

Moreover, improving our ability to accurately determine exoplanet masses allows for more precise constraints on orbital and atmospheric parameters. For exoplanets located within the habitable zone, this can potentially lead to the detection of biosignatures, furthering the quest to identify potentially habitable worlds.

Additionally, while atmospheric characterization of exoplanets traditionally relies on transmission spectroscopy, which requires planets to transit their host stars, this is not the only method for studying exoplanetary atmospheres. In recent years, there has been a proposal to combine high-resolution spectrographs with high angular resolution imaging as an alternative approach to investigate the atmospheres of nearby exoplanets (Lovis et al., 2017; Snellen et al., 2019).

In the near future, a proposed visitor instrument for the Very Large Telescope (VLT) known as RISTRETTO (Blind et al., 2022) will make it possible to conduct high dispersion coronagraphy (HDC) spectroscopy of exoplanets. This development is expected to enable the study of exoplanetary atmospheres and serve as a proof of concept for ANDES (Marconi et al., 2022), a high-resolution spectrograph designed for the European Extremely Large Telescope (ELT). ANDES will offer  $R \sim 100,000$  stabilised Échelle spectroscopy, covering a wavelength range of at least 400 to 1800 nm. It is anticipated to conduct HDC spectroscopy of exoplanets, similar to RISTRETTO, with the aim of detecting biomarkers such as the oxygen band at 1300 nm (Suárez Mascareño et al., 2023).

With all these ongoing effort, the years to come will likely witness a multitude of claims

## 5.2. PERSPECTIVES

---

related to the potential detection of biosignatures in exoplanets, such as the recent one regarding K2-18b ([Madhusudhan et al., 2023](#)). Hopefully, one of these candidates for hosting life will owe its detection to the collective contributions of the SPIRou team and other dedicated collaborators.



# Chapter 6

## Conclusion et perspectives

### Contents

---

<b>6.1</b>	<b>Résumé de la thèse . . . . .</b>	<b>85</b>
<b>6.2</b>	<b>Perspectives . . . . .</b>	<b>90</b>

---

### 6.1 Résumé de la thèse

Cette thèse s’est concentrée sur l’amélioration de la précision des mesures RV de SPIRou et sur le traitement des erreurs systématiques, en particulier celles causées par les raies telluriques. Cet effort a abouti au développement de l’algorithme `Wapiti`, qui est conçu pour identifier et corriger les effets systématiques sous-jacents dans les données via une wPCA.

Nous avons présenté la version actuelle de l’algorithme et l’avons testé sur des simulations ainsi que sur quatre cibles du programme SLS-PS, obtenant des résultats prometteurs. L’algorithme a révélé avec succès le signal de 14,2 jours précédemment dissimulé dans Gl 251, qui était caché par des erreurs systématiques dans les données SPIRou. En outre, notre application sur Gl 480 et Gl 382 a montré que même lorsque les signaux étaient déjà apparents dans les données, `Wapiti` corrigeait efficacement les erreurs systématiques tout en préservant le signal planétaire et en améliorant sa signification statistique.

Enfin, nous avons appliqué la méthode `Wapiti` à Gl 725 B. Cette cible représentait un défi important pour l’analyse RV en raison de sa faible excursion en BERV, ce qui la rendait particulièrement défavorable aux techniques de correction tellurique. Par conséquent, les données RV sont affectées par de fortes systématiques provenant des résidus telluriques. Cependant, la méthode `Wapiti` a permis de traiter efficacement ces systématiques, simplifiant ainsi le processus de détection et de caractérisation de signaux planétaires non encore découverts, dont un situé dans la zone habitable de son étoile hôte.

Les signaux planétaires confirmés et détectés dans le cadre de cette thèse sont résumés dans le graphique masse-période présenté dans la figure 6.1. Dans ce graphique, nous avons également inclus la masse et les périodes des planètes connues dans l’échantillon SLS, comme

indiqué dans le tableau 1 de [Moutou et al. \(2023\)](#). Les paramètres des planètes pour ce graphique ont été tirés des références mentionnées, à l'exception de Gl 1148 et Gl 876 ([Moutou et al., 2023](#)), Gl 251 ([Ould-Elhkim et al., 2023](#)), Gl 382 et Gl 480 (tirés de l'article soumis en annexe C.2.2). En outre, les paramètres de Gl 725 B et Gl 725 Bc sont dérivés du tableau 4.3.

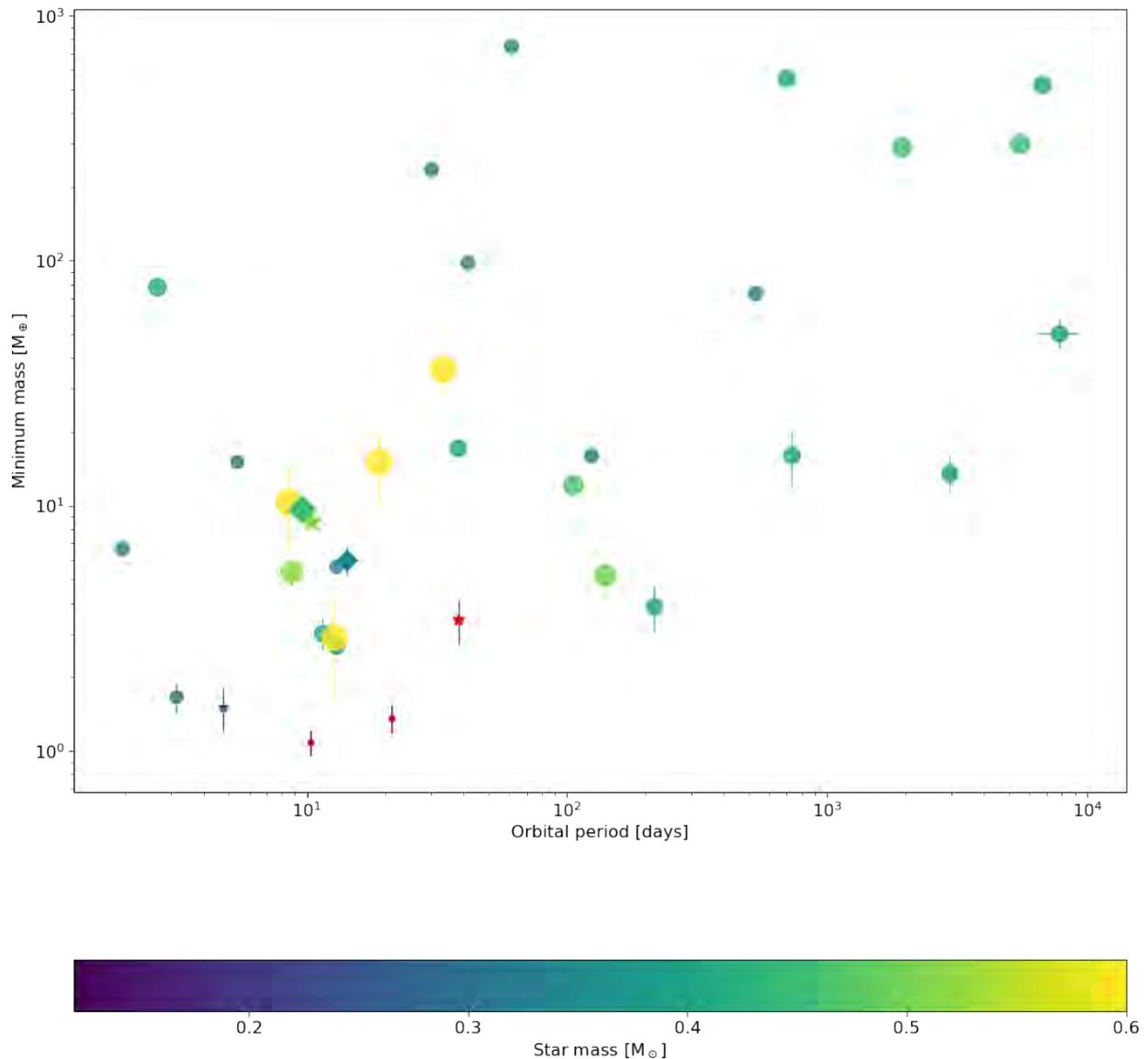


Figure 6.1: Signaux détectés et confirmés dans cette thèse dans le cadre du programme SLS-PS. Les détections de planètes confirmées par ce travail sont représentées par des marqueurs en forme de diamant, tandis que les détections non reportées précédemment sont indiquées par des marqueurs en forme d'étoile et les planètes dans la ZH de leur étoile sont soulignées en rouge. Les marqueurs sont codés par couleur en fonction de la masse de l'étoile hôte et leur taille est proportionnelle à cette même masse.

Dans le diagramme masse-période représenté dans la figure 6.1, les détections de planètes

confirmées par cette thèse sont représentées par des marqueurs en forme de diamant, tandis que les détections non reportées précédemment sont indiquées par des marqueurs en forme d'étoile. Notamment, seules deux planètes (Gl 1002b et Gl 1002c) étaient initialement situées dans la ZH de leur étoile hôte dans l'échantillon SLS-PS. Ainsi, la détection de Gl 725 Bc est d'une importance significative car elle augmente le nombre de mondes habitables dans cet échantillon (les planètes situées dans la ZH de leur étoile sont soulignées en rouge dans la figure 6.1).

De plus, on observe que Gl 725 Bb est l'une des planètes les moins massives de cet échantillon et qu'elle possède l'une des périodes orbitales les plus courtes, ce qui permet de poursuivre l'exploration de cet espace de paramètres. En ce qui concerne les autres détections, elles restent à des masses relativement faibles ( $\leq 10 M_{\oplus}$ ). Ce biais est dû au fait que cette étude se concentre sur les étoiles qui présentent des niveaux élevés de systématiques, ce qui tend à dissimuler les planètes de faible amplitude. Les cibles présentant des signaux planétaires forts qui dépassent l'influence des systématiques n'ont pas été examinées en détail dans cette étude.

Tirer des conclusions définitives sur la population d'exoplanètes au sein de l'échantillon SLS-PS sur la base de nos nouvelles détections serait évidemment prématuré. Néanmoins, cette thèse s'achève sur une étape passionnante : la méthode `Wapiti` peut maintenant être facilement appliquée aux cibles de l'échantillon SLS-PS et contribuer à leur analyse individuelle. Des efforts continus sont en cours pour appliquer cette méthode à des cibles autres que celles discutées dans cette étude, dans le but de découvrir de nouveaux systèmes planétaires dans les publications à venir (A. Carmona et al. en préparation et P. Cortes-Zuleta et al. en préparation). Dans le futur, grâce à la méthode `Wapiti` développée au cours de cette thèse, d'autres planètes seront, nous l'espérons, ajoutées à l'espace des paramètres de la figure 6.1, approfondissant ainsi notre connaissance des mondes proches dans le voisinage du Soleil. Le code est accessible sur GitHub<sup>1</sup>, accompagné d'un tutoriel pour aider les utilisateurs dans son application.

Pour conclure avec ce par quoi nous avons commencé, nous avons appliqué `Wapiti` à toutes les cibles qui constituent le point zéro discuté en introduction. Comme le montre la figure 6.2, le comportement initialement observé dans l'espace  $V_{\text{tot}}$ , qui a conduit au développement de la correction BERVZP, a maintenant disparu. Par conséquent, il semble que la correction BERVZP ne soit plus nécessaire, même si elle reste utile pour la comparaison avec les données corrigées par `Wapiti`. D'autant plus que la méthode `Wapiti` elle-même présente des limites inhérentes, comme indiqué dans la section suivante.

### 6.1.1 Capacité et limites de la méthode `Wapiti`

Malgré les résultats obtenus, il est important de reconnaître certaines limites associées à cette méthode, qui découlent généralement de sa dépendance à l'égard de la wPCA en tant que technique de base.

---

<sup>1</sup><https://github.com/HkmMerwan/wapiti>

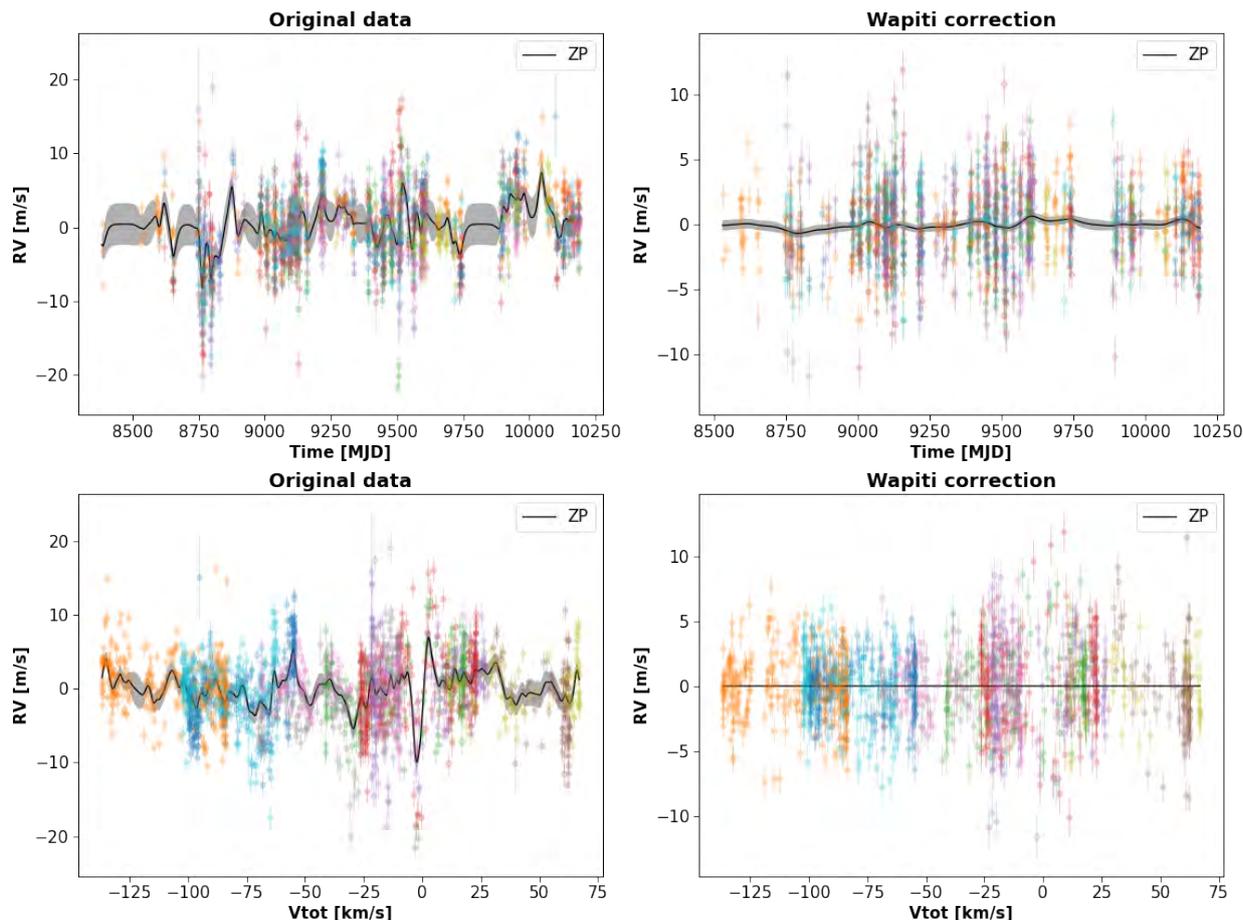


Figure 6.2: Standard RVs en fonction du temps et de  $V_{\text{tot}}$  avant et après la correction *Wapiti*. Les figures du haut montrent la RV des étoiles standards qui constituent le pont zéro, tandis que les figures du bas les montrent dans l'espace  $V_{\text{tot}}$ . Les panneaux de gauche sont les données RV originales et les panneaux de droite sont les RV après la correction *Wapiti*.

### wPCA et systématique

La méthode *Wapiti* utilise la wPCA pour estimer les erreurs systématiques dans les données, mais la wPCA est conçue pour capturer la variabilité plutôt que ces effets. Par exemple, dans nos simulations, avec une large couverture en BERV, les influences telluriques ont conduit à de nombreuses composantes principales et ont augmenté la variabilité des données, alors même que les systématiques étaient négligeables. En revanche, dans le scénario avec une couverture en BERV limitée, la variabilité des données a diminué, conduisant à moins de composantes, tandis que les systématiques sont devenues plus perceptibles au niveau de  $10 \text{ m.s}^{-1}$ .

Il s'agit d'un aspect essentiel à prendre en compte lors de l'utilisation de notre méthode, car la plupart de ses capacités et de ses limites en découlent. Par exemple, en raison de cette tendance à capturer la variabilité, la wPCA est très sensible aux valeurs aberrantes et aux observations à faible RSB et il peut être nécessaire d'exclure certaines mesures dans certaines situations.

En outre, s'il existe un effet systématique dans les données qui n'introduit pas de variabilité significative, la wPCA ne le détectera pas et, par conséquent, **Wapiti** ne le corrigera pas. Par exemple, les erreurs achromatiques entrent dans cette catégorie, c'est pourquoi il est toujours nécessaire de traiter les effets instrumentaux tels que la dérive du FP ou le point zéro avant d'utiliser cette méthode.

Enfin, la wPCA construit des composantes principales orthogonales par définition. Cependant, les effets systématiques dans les données réelles peuvent ne pas être orthogonaux, ce qui peut conduire la wPCA à mélanger différentes systématiques, les rendant ainsi plus difficiles à interpréter.

### Caractérisation des signaux détectés

La capacité de détection de **Wapiti**, et les méthodes de correction wPCA en général, ont démontré une grande efficacité dans les simulations et lorsqu'elles sont appliquées aux données réelles. Cependant, nous avons rencontré un problème dans les simulations lorsque le signal détecté avait la même période que les effets systématiques. Même après avoir supprimé cette période des effets systématiques, notre capacité à estimer l'amplitude du signal injecté est restée limitée par l'amplitude des effets systématiques.

Cela montre que même si **Wapiti** peut réduire les effets systématiques et révéler des signaux dans les séries temporelles RV, il est préférable de traiter les effets systématiques au préalable afin de limiter leur impact sur les données. Cela souligne l'importance d'une correction tellurique de haute qualité, telle que celle utilisée dans **APER0**.

Ce point est crucial car il implique directement que, dans le pire des cas où un signal planétaire partage sa période orbitale avec un signal présent dans les systématiques, la détermination de sa masse et d'autres paramètres orbitaux pourrait être biaisée par les systématiques.

### **Wapiti** et activité stellaire

La question de savoir si la méthode **Wapiti** peut effectivement atténuer l'activité stellaire est peut-être l'un de ses aspects les plus intrigants. Nous n'avons pas rencontré un tel scénario dans les cibles discutées dans ce manuscrit, car l'activité stellaire a tendance à être relativement faible dans les données RV dans le proche infrarouge. Cependant, nous avons appliqué cette méthode de correction à d'autres cibles du programme SLS-PS où l'activité pouvait être observée dans les séries temporelles RV. Bien que ce travail en soit encore à ses débuts, une tendance se dessine. En fin de compte, la réponse à la question de savoir si **Wapiti** peut traiter l'activité stellaire se résume une fois de plus à la distinction entre l'identification des composants responsables de la variabilité et les effets systématiques.

Nous avons rencontré trois scénarios différents. Dans le premier cas, nous avons observé que la méthode corrigeait avec succès l'activité stellaire et que les composantes affichaient des signaux liés à l'activité. Dans le deuxième cas, malgré la présence d'une activité stellaire, la méthode **Wapiti** n'a pas pu la corriger, et elle a identifié zéro composante convenant à la reconstruction des données. Enfin, le troisième cas combine les aspects des deux précédents, la méthode **Wapiti** ayant trouvé des composantes affichant des signaux d'activité qu'elle

a jugées appropriées pour la reconstruction. Cependant, même après avoir supprimé ces composantes des données, l'activité stellaire n'a pas pu être complètement retirée.

Ces trois scénarios peuvent être expliqués en se rappelant que l'activité stellaire est un terme qui englobe divers effets, comme nous l'avons vu dans l'introduction. Si l'effet responsable du signal d'activité stellaire introduit de la variabilité dans les données RV du LBL, la wPCA le capturera et la méthode **Wapiti** le reconnaîtra comme approprié à la reconstruction. Inversement, si l'activité stellaire n'induit pas de variabilité dans les RVs, la méthode **Wapiti** ne détectera pas cet effet dans les données, ce qui pourrait entraîner l'utilisation de zéro composantes, et l'activité ne sera donc pas corrigée. Dans les cas où l'activité stellaire dans les données provient d'une combinaison de différents effets, certains causant de la variabilité et d'autres non, la méthode capturera les composantes variables et atténuera partiellement l'activité stellaire, mais ne la corrigera peut-être pas complètement.

Par conséquent, **Wapiti** ne corrige pas toujours l'activité stellaire, mais en traitant d'autres sources systématiques, il devrait au moins laisser l'activité stellaire comme la préoccupation restante dans les données. En outre, dans les cas où l'activité introduit une variabilité dans les données RV du LBL, les vecteurs principaux affichant des signaux d'activité pourraient servir d'indicateurs d'activité et être utilisés pour la correction de l'activité à l'aide de GPs multidimensionnels. En outre, les z-scores introduits dans le chapitre 3 pourraient aider à identifier les lignes les plus sensibles à l'activité stellaire.

Cependant, du fait que la wPCA calcule des composantes orthogonales, les systématiques peuvent s'entremêler, mélangeant potentiellement l'activité stellaire avec d'autres effets systématiques et rendant les tentatives de caractérisation de l'activité difficiles. Enfin, comme nous l'avons vu précédemment, la caractérisation des signaux planétaires peut être compliquée lorsque leurs périodes orbitales sont proches des périodes présentes dans les systématiques. De même, la caractérisation des signaux d'activité stellaire peut poser le même problème lorsque la période de rotation de l'étoile est proche des périodes présentes dans les systématiques. Par exemple, certaines étoiles de l'échantillon SLS-PS ont des périodes de rotation proches des harmoniques telluriques de 90, 120 ou 180 jours.

## 6.2 Perspectives

### 6.2.1 Sur **Wapiti**

#### Avec SPIRou

La méthode **Wapiti** développée au cours de cette thèse sera employée sur des cibles observées avec SPIRou dans le cadre du programme SLS-PS, avec la possibilité de confirmer, d'infirmer ou de découvrir de nouveaux systèmes planétaires. De plus, sa capacité à réduire les systématiques dans les données SPIRou s'étend au-delà des cibles SLS-PS et peut être facilement appliquée à toutes les cibles observées avec SPIRou dans le cadre d'autres programmes, y compris le programme de suivi WP2. Cependant, ces cibles sont observées avec un SNR plus faible, introduisant potentiellement une variabilité qui pourrait affecter l'efficacité de la wPCA à capturer les systématiques. Cependant, la période orbitale de la planète est déjà connue à partir des observations de transit pour ces cibles, offrant une information préalable qui pourrait être utilisée pour guider la méthode **Wapiti** dans le traitement des erreurs

systematiques.

### Avec d'autres instruments

La méthode présentée dans cette thèse s'appuie sur les données du LBL pour son application et cet algorithme a été rendu compatible avec de nombreux instruments, y compris ceux opérant dans le domaine optique, tels que HARPS , HARPS-N, ESPRESSO, CARMENES, ou SOPHIE<sup>2</sup>. Par conséquent, la méthode `Wapiti` peut être facilement utilisée avec ces instruments et, dans un avenir proche, il sera intéressant d'évaluer les performances de `Wapiti` dans le domaine optique et d'explorer le potentiel d'analyses RV dans l'optique et dans le proche infrarouge en utilisant `Wapiti`.

En ce qui concerne le domaine du proche infrarouge, il existe d'autres instruments sur lesquels le LBL peut être utilisé, comme NIRPS (dont les données seront même traitées avec le même DRS APERO). La capacité de la méthode `Wapiti` à corriger les effets systématiques d'origine tellurique dans les données SPIRou sera très utile pour atténuer des effets similaires dans les données NIRPS. En outre, un autre instrument qui bénéficiera des travaux effectués sur SPIRou est son jumeau à venir, SPIP<sup>3</sup>, situé à l'observatoire du Pic du Midi. La méthode `Wapiti` devrait également être appliquée aux données RV de SPIP afin de faciliter leur analyse.

### Avec plus que des RVs

Dans cette étude, nous nous sommes concentrés sur la réduction de l'impact de la systématique sur les données RVs. Cependant, la méthode `Wapiti` a le potentiel pour des applications plus larges au-delà des données RVs raies par raies. À l'avenir, il pourrait être intéressant, par exemple, de calculer une wPCA sur les dLWs en plus des RVs. Cela pourrait fournir des composantes supplémentaires à la méthode `Wapiti` pour identifier celles qui conviennent le mieux à la reconstruction des données. Cette approche serait conforme à la philosophie d'une autre méthode, SCALPELS (Collier Cameron et al., 2019), qui atténue l'activité stellaire en démêlant les systématiques liées au déplacement et à la forme de la CCF. Dans ce cas, cela serait réalisé en utilisant des composantes wPCA associées à la fois au déplacement (RVs) et à la largeur (dLWs) des raies spectrales.

En outre, la méthode `Wapiti` n'est pas limitée à la reconstruction des RVs ; elle pourrait également être appliquée aux dLWs pour éliminer les systématiques de la série temporelle dLW elle-même, révélant potentiellement des signaux d'activité cachés. En outre, `Wapiti` n'est pas non plus limitée aux seules données LBL : tant que l'objectif est d'éliminer les systématiques de n'importe quel type de données à l'aide d'une wPCA, la méthode `Wapiti` peut être employée. Par exemple, dans le contexte de la spectroscopie atmosphérique, le pipeline ATMOSPHERIX utilise une PCA sur les pixels à cette fin (Klein et al., 2023; Debras et al., 2023). Il pourrait s'avérer utile d'appliquer `Wapiti` dans ce contexte afin d'évaluer s'il peut améliorer la détection des signaux.

---

<sup>2</sup><https://github.com/njcuk9999/lbl/tree/main/>

<sup>3</sup><https://www5.obs-mip.fr/wp-content-omp/uploads/sites/10/2016/10/spip2-tbl.pdf>

### Au-delà de Wapiti

Nous avons vu que la plupart des limites de la méthode Wapiti provenaient de son utilisation de la wPCA pour identifier les systématiques, même si ce n'est pas l'objectif premier de cette technique. Par conséquent, la prochaine étape naturelle consisterait à explorer l'utilisation d'autoencodeurs pour la reconstruction du RV et l'atténuation des systématiques. Remplacer la pCA par des autoencodeurs est une pratique courante, et au cours de cette thèse, des efforts ont été faits pour mettre en œuvre des autoencodeurs dans le pipeline ATMOSPHERIX (Klein et al., 2023; Debras et al., 2023). Cette expérience a permis de comprendre les défis potentiels lors de l'application des autoencodeurs aux vitesses radiales, y compris des considérations telles que l'architecture du réseau de neurones, le choix des fonctions de perte ou d'activation, et les méthodes d'entraînement.

Les autoencodeurs offrent l'avantage de capturer les relations non linéaires dans les données, alors que la PCA est intrinsèquement linéaire. Toutefois, cela se fait au détriment de l'interprétabilité, car les caractéristiques de l'espace latent peuvent ne pas être orthogonales ou ordonnées d'une manière facilement interprétable (par exemple par variance décroissante pour la PCA). Par conséquent, l'application de la wPCA aux données à l'aide de la méthode Wapiti reste une approche plus simple et plus intuitive. Plus de travail sera nécessaire pour évaluer les avantages de l'utilisation d'un autoencodeur par rapport à la wPCA sur des séries temporelles de vitesses radiales

### 6.2.2 Sur d'autres travaux

Si le développement de la méthode Wapiti a constitué une part importante de cette thèse, d'autres outils ont également été créés au cours de cette recherche. Bien qu'ils n'aient pas atteint un stade où ils pouvaient être intégrés dans le texte principal du manuscrit, nous présentons l'un d'entre eux dans sa forme préliminaire plus en détail dans l'annexe B et il fera l'objet d'une recherche plus approfondie à l'avenir.

#### périodogramme BERV-I

Aussi efficace que soit la méthode Wapiti pour corriger les systématiques telluriques dans les données RV, elle nécessite que les RV soient calculées à l'aide de la méthode LBL. Bien qu'elle soit de plus en plus appliquée à divers instruments au-delà de SPIRou, d'autres instruments peuvent encore utiliser des méthodes alternatives telles que la méthode CCF ou des techniques de template-matching telles que `harps-terra` (Anglada-Escudé & Butler, 2012), `naira` (Astudillo-Defru et al., 2017b), `wobble` (Bedell et al., 2019a), ou `serval` (Zechmeister et al., 2018).

Pour ces raisons, nous avons développé un outil capable de distinguer les signaux d'origine tellurique de ceux provenant d'autres sources. Pour ce faire, nous avons tiré parti du comportement caractéristique des signaux telluriques, qui est étroitement lié au BERV. Par conséquent, pour filtrer les signaux provenant des lignes telluriques, nous avons essayé de concevoir un périodogramme qui reste invariant aux variations du BERV.

Les détails du périodogramme BERV-I se trouvent dans l'annexe B. Bien qu'il ait donné des résultats prometteurs lorsqu'il a été appliqué aux simulations en révélant avec succès les signaux planétaires injectés sans nécessiter de corrections, son application aux données

réelles s'est avérée moins efficace. Cela souligne la nécessité de poursuivre les recherches et d'affiner cet outil pour qu'il fonctionne comme prévu.

### 6.2.3 Caractériser les systèmes planétaires autour des naines rouges

Bien que le développement de ces outils soit intéressant et passionnant par eux-mêmes, l'objectif reste de les utiliser pour réduire les erreurs systématiques, afin de détecter et de caractériser les systèmes planétaires orbitant autour de naines rouges.

Les progrès réalisés dans la détection et la caractérisation des exoplanètes autour des étoiles de faible masse sont d'une grande importance. Elles contribuent non seulement à notre connaissance de la formation et de l'évolution des planètes, mais ont également des implications pour les missions visant à caractériser les atmosphères des exoplanètes. Ces missions comprennent des projets en cours tels que le JWST ainsi que des missions futures (ARIEL, E-ELT).

En outre, l'amélioration de notre capacité à déterminer avec précision la masse des exoplanètes permet d'obtenir des contraintes plus précises sur les paramètres orbitaux et atmosphériques. Pour les exoplanètes situées dans la zone habitable, cela peut potentiellement conduire à la détection de biosignatures, nous aidant un peu plus dans la quête de mondes potentiellement habitables.

En outre, si la caractérisation atmosphérique des exoplanètes repose traditionnellement sur la spectroscopie de transmission, qui nécessite que les planètes transitent, ce n'est pas la seule méthode d'étude des atmosphères exoplanétaires. Ces dernières années, il a été proposé de combiner des spectrographes à haute résolution avec une imagerie à haute résolution angulaire comme approche alternative pour étudier les atmosphères des exoplanètes proches (Lovis et al., 2017; Snellen et al., 2019).

Dans un avenir proche, un instrument proposé pour le Very Large Telescope (VLT), connu sous le nom de RISTRETTO (Blind et al., 2022), permettra de réaliser une spectroscopie par coronagraphie à haute dispersion (HDC) des exoplanètes. Ce développement devrait permettre l'étude des atmosphères exoplanétaires et servir de preuve de concept pour ANDES (Marconi et al., 2022), un spectrographe à haute résolution conçu pour l'ELT. ANDES offrira une spectroscopie d'échelle stabilisée avec une résolution de  $R \sim 100\,000$ , couvrant une gamme de longueurs d'onde d'au moins 400 à 1800 nm. Il est prévu qu'il réalise de la spectroscopie HDC des exoplanètes, similaire à RISTRETTO, dans le but de détecter des biomarqueurs tels que la bande d'oxygène à 1300 nm (Suárez Mascareño et al., 2023).

Avec tous ces efforts en cours, les années à venir verront probablement une multitude d'annonces liées à la détection potentielle de biosignatures dans les exoplanètes, comme celle récente concernant K2-18b (Madhusudhan et al., 2023). Espérons que l'un de ces candidats à la vie sera détecté grâce aux contributions collectives de l'équipe SPIROU et d'autres collaborateurs dévoués.



# List of Figures

1.1	Representation of stellar systems in the vortex theory . . . . .	3
1.2	Minimum planetary mass of confirmed exoplanets over the years . . . . .	7
1.3	Illustration of a planet orbiting a star showing the relevant planetary parameters . . . . .	8
1.4	Various plots showing the diverse variety of exoplanets. . . . .	12
1.5	The habitability of an exoplanet as a multi-parameter problem. . . . .	14
1.6	All known planets near the habitable zone. . . . .	15
1.7	Phase-folded RV measurements of AD Leo with SPIRou and SOPHIE. . . . .	21
1.8	Computer-aided design (CAD) views of the main SPIRou units and the Cassegrain unit, the optical layout of the SPIRou spectrograph is also displayed. . . . .	23
1.9	Example of the telluric correction in APERO . . . . .	25
1.10	Sample line-by-line velocity spectrum for Barnard’s star showing the scatter of velocity measurements over a wide velocity range . . . . .	26
1.11	Comparison between SPIRou and CARMENES RV data of Gl 251 . . . . .	27
1.12	Illustration of the ZP corrections . . . . .	29
2.1	Illustration of the permutation test with Gl 251 . . . . .	34
2.2	Leave-p-out cross-validation with Gl 251 . . . . .	37
2.3	Re-ordering the wPCA components in decreasing BIC order for Gl 251 . . . . .	38
2.4	Comparison between SPIRou data using bervzp or Wapiti correction and CARMENES data of Gl 251 . . . . .	39
2.5	Variability of RV data of Gl 251 and its principal vectors in the BERV space . . . . .	40
3.1	Illustration of the simulations to inject tellurics in the star template with $\alpha = 50\%$ . . . . .	44
3.2	Simulation results when injecting telluric at a $3 \pm 1\%$ level . . . . .	46
3.3	Principal vectors from the simulations without rejecting low SNR observations . . . . .	47
3.4	Characterizing the impact of tellurics on wPCA in the case of a narrow BERV coverage . . . . .	49
3.5	Characterizing the impact of tellurics on wPCA in the case of a wide BERV coverage . . . . .	50
3.6	Telluric contamination of the stellar activity indicators . . . . .	52
3.7	Detection capacity of the wPCA in case of telluric contamination using GLS periodograms . . . . .	54
3.8	Detection capacity of the wPCA in case of telluric contamination . . . . .	57
3.9	Characterization capacity of the wPCA in case of telluric contamination in the wide BERV coverage scenario. . . . .	59

---

3.10	Characterization capacity of the wPCA in case of telluric contamination in the narrow BERV coverage scenario. . . . .	60
4.1	Wapiti correction of Gl 480 . . . . .	65
4.2	Wapiti correction of Gl 382 . . . . .	66
4.3	Wapiti correction of Gl 725 B . . . . .	68
4.4	Iterative search for signals in Gl 725 B . . . . .	69
4.5	Phase-folding of the SPIRou RV data of Gl 725 B . . . . .	72
5.1	Detected and confirmed signals in this thesis in the context of the SLS-PS program . . . . .	76
5.2	RV standards in time and $V_{\text{tot}}$ spaces before and after Wapiti correction . . . . .	78
6.1	Signaux détectés et confirmés dans cette thèse dans le cadre du programme SLS-PS . . . . .	86
6.2	Standard RVs en fonction du temps et de $V_{\text{tot}}$ avant et après la correction Wapiti . . . . .	88
A.1	HR8799 direct imaging planet detections . . . . .	124
A.2	Best-fitted solution including the astrometric signal due to the planetary companion GJ 896Ab . . . . .	125
A.3	Illustration of the orbit of a transiting planet and the corresponding light curve	126
A.4	Simulations of exoplanets that the Roman Space Telescope should be capable of detecting . . . . .	130
B.1	BERV-I periodogram applied on simulations with a narrow BERV coverage . . . . .	135
B.2	BERV-I periodogram applied on Gl 725 B RV timeseries . . . . .	135

# List of Tables

1.1	Radial velocity signals for different kinds of planets orbiting a solar-mass star	9
3.1	Table of the parameters for scenarios A, B and C in the simulations . . . . .	56
4.2	Orbital and derived planetary parameters from the MCMC algorithm applied on RV data of Gl 725 B. (1): <a href="#">Kipping (2013)</a> . . . . .	70
4.3	Same as table <a href="#">4.2</a> for a circular model. . . . .	71



# Bibliography

- [1] Aigrain, S. 2012, in *New Horizons in Time Domain Astronomy*, ed. E. Griffin, R. Hanisch, & R. Seaman, Vol. 285, 105–105, doi: [10.1017/S1743921312000361](https://doi.org/10.1017/S1743921312000361)
- [2] 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 and Astrophysics*, 666, A196, doi: [10.1051/0004-6361/202243629](https://doi.org/10.1051/0004-6361/202243629)
- [3] Almenara, J. M., Bonfils, X., Bryant, E. M., et al. 2023, TOI-4860 b, a short-period giant planet transiting an M3.5 dwarf, arXiv e-prints, arXiv:2308.01454, doi: [10.48550/arXiv.2308.01454](https://doi.org/10.48550/arXiv.2308.01454)
- [4] Alonso-Floriano, F. J., Morales, J. C., Caballero, J. A., et al. 2015, CARMENES input catalogue of M dwarfs. I. Low-resolution spectroscopy with CAFOS, *Astronomy and Astrophysics*, 577, A128, doi: [10.1051/0004-6361/201525803](https://doi.org/10.1051/0004-6361/201525803)
- [5] Alpert, M. 2005, Red Star Rising, *Scientific American*, 293, 28, doi: [10.1038/scientificamerican1105-28](https://doi.org/10.1038/scientificamerican1105-28)
- [6] 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, *ApJ Supplements*, 200, 15, doi: [10.1088/0067-0049/200/2/15](https://doi.org/10.1088/0067-0049/200/2/15)
- [7] Anglada-Escudé, G., Amado, P. J., Barnes, J., et al. 2016, A terrestrial planet candidate in a temperate orbit around Proxima Centauri, *Nature*, 536, 437, doi: [10.1038/nature19106](https://doi.org/10.1038/nature19106)
- [8] Artigau, É., Saint-Antoine, J., Lévesque, P.-L., et al. 2018, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 10709, High Energy, Optical, and Infrared Detectors for Astronomy VIII, ed. A. D. Holland & J. Beletic, 107091P, doi: [10.1117/12.2314475](https://doi.org/10.1117/12.2314475)
- [9] Artigau, É., Astudillo-Defru, N., Delfosse, X., et al. 2014, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 9149, Observatory Operations: Strategies, Processes, and Systems V, ed. A. B. Peck, C. R. Benn, & R. L. Seaman, 914905, doi: [10.1117/12.2056385](https://doi.org/10.1117/12.2056385)
- [10] 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, doi: [10.3847/1538-3881/ac7ce6](https://doi.org/10.3847/1538-3881/ac7ce6)

- [11] Astudillo-Defru, N., Forveille, T., Bonfils, X., et al. 2017a, The HARPS search for southern extra-solar planets. XLI. A dozen planets around the M dwarfs GJ 3138, GJ 3323, GJ 273, GJ 628, and GJ 3293, *Astronomy and Astrophysics*, 602, A88, doi: [10.1051/0004-6361/201630153](https://doi.org/10.1051/0004-6361/201630153)
- [12] —. 2017b, The HARPS search for southern extra-solar planets. XLI. A dozen planets around the M dwarfs GJ 3138, GJ 3323, GJ 273, GJ 628, and GJ 3293, *Astronomy and Astrophysics*, 602, A88, doi: [10.1051/0004-6361/201630153](https://doi.org/10.1051/0004-6361/201630153)
- [13] Auvergne, M., Bodin, P., Boisnard, L., et al. 2009, The CoRoT satellite in flight: description and performance, *Astronomy and Astrophysics*, 506, 411, doi: [10.1051/0004-6361/200810860](https://doi.org/10.1051/0004-6361/200810860)
- [14] Bachelet, E., Specht, D., Penny, M., et al. 2022, Euclid-Roman joint microlensing survey: Early mass measurement, free floating planets, and exomoons, *Astronomy and Astrophysics*, 664, A136, doi: [10.1051/0004-6361/202140351](https://doi.org/10.1051/0004-6361/202140351)
- [15] Bakos, G., Noyes, R. W., Kovács, G., et al. 2004, Wide-Field Millimagnitude Photometry with the HAT: A Tool for Extrasolar Planet Detection, , 116, 266, doi: [10.1086/382735](https://doi.org/10.1086/382735)
- [16] Ballard, S., Fabrycky, D., Fressin, F., et al. 2011, The Kepler-19 System: A Transiting  $2.2 R_{\oplus}$  Planet and a Second Planet Detected via Transit Timing Variations, *The Astrophysical Journal*, 743, 200, doi: [10.1088/0004-637X/743/2/200](https://doi.org/10.1088/0004-637X/743/2/200)
- [17] Baluev, R. V. 2008, Assessing the statistical significance of periodogram peaks, *Monthly Notices of the Royal Astronomical Society*, 385, 1279, doi: [10.1111/j.1365-2966.2008.12689.x](https://doi.org/10.1111/j.1365-2966.2008.12689.x)
- [18] Baranne, A., Queloz, D., Mayor, M., et al. 1996, ELODIE: A spectrograph for accurate radial velocity measurements., , 119, 373
- [19] Barnard, E. E. 1916, A small star with large proper motion, *Popular Astronomy*, 24, 504
- [20] Barnes, R., Deitrick, R., Luger, R., et al. 2016, The Habitability of Proxima Centauri b I: Evolutionary Scenarios, arXiv e-prints, arXiv:1608.06919, doi: [10.48550/arXiv.1608.06919](https://doi.org/10.48550/arXiv.1608.06919)
- [21] Baroch, D., Morales, J. C., Ribas, I., et al. 2020, The CARMENES search for exoplanets around M dwarfs. Convective shift and starspot constraints from chromatic radial velocities, *Astronomy and Astrophysics*, 641, A69, doi: [10.1051/0004-6361/202038213](https://doi.org/10.1051/0004-6361/202038213)
- [22] Baruteau, C., Bai, X., Mordasini, C., & Mollière, P. 2016, Formation, Orbital and Internal Evolutions of Young Planetary Systems, *Space Science Review*, 205, 77, doi: [10.1007/s11214-016-0258-z](https://doi.org/10.1007/s11214-016-0258-z)
- [23] Baruteau, C., Crida, A., Paardekooper, S. J., et al. 2014, in *Protostars and Planets VI*, ed. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 667–689, doi: [10.2458/azu\\_uapress\\_9780816531240-ch029](https://doi.org/10.2458/azu_uapress_9780816531240-ch029)
- [24] Bashi, D., Zucker, S., Adibekyan, V., et al. 2020, Occurrence rates of small planets from

- HARPS. Focus on the Galactic context, *Astronomy and Astrophysics*, 643, A106, doi: [10.1051/0004-6361/202038881](https://doi.org/10.1051/0004-6361/202038881)
- [25] Bauer, F. F., Zechmeister, M., Kaminski, A., et al. 2020, The CARMENES search for exoplanets around M dwarfs. Measuring precise radial velocities in the near infrared: The example of the super-Earth CD Cet b, *Astronomy and Astrophysics*, 640, A50, doi: [10.1051/0004-6361/202038031](https://doi.org/10.1051/0004-6361/202038031)
- [26] Bean, J. L., Seifahrt, A., Hartman, H., et al. 2010, The CRIRES Search for Planets Around the Lowest-mass Stars. I. High-precision Near-infrared Radial Velocities with an Ammonia Gas Cell, *The Astrophysical Journal*, 713, 410, doi: [10.1088/0004-637X/713/1/410](https://doi.org/10.1088/0004-637X/713/1/410)
- [27] Beaulieu, J. P., Bennett, D. P., Batista, V., et al. 2010, in *Astronomical Society of the Pacific Conference Series*, Vol. 430, *Pathways Towards Habitable Planets*, ed. V. Coudé du Foresto, D. M. Gelino, & I. Ribas, 266, doi: [10.48550/arXiv.1001.3349](https://doi.org/10.48550/arXiv.1001.3349)
- [28] Bedell, M., Hogg, D. W., Foreman-Mackey, D., Montet, B. T., & Luger, R. 2019a, WOBBLE: A Data-driven Analysis Technique for Time-series Stellar Spectra, *The Astronomical Journal*, 158, 164, doi: [10.3847/1538-3881/ab40a7](https://doi.org/10.3847/1538-3881/ab40a7)
- [29] —. 2019b, WOBBLE: A Data-driven Analysis Technique for Time-series Stellar Spectra, *The Astronomical Journal*, 158, 164, doi: [10.3847/1538-3881/ab40a7](https://doi.org/10.3847/1538-3881/ab40a7)
- [30] Berdiñas, Z. M., Amado, P. J., Anglada-Escudé, G., Rodríguez-López, C., & Barnes, J. 2016, High-cadence spectroscopy of M dwarfs - I. Analysis of systematic effects in HARPS-N line profile measurements on the bright binary GJ 725A+B, *Monthly Notices of the Royal Astronomical Society*, 459, 3551, doi: [10.1093/mnras/stw906](https://doi.org/10.1093/mnras/stw906)
- [31] Berdyugina, S. V. 2005, Starspots: A Key to the Stellar Dynamo, *Living Reviews in Solar Physics*, 2, 8, doi: [10.12942/lrsp-2005-8](https://doi.org/10.12942/lrsp-2005-8)
- [32] Berman, L. 1932, 70 Ophiuchi as a triple system, *Lick Observatory Bulletin*, 443, 24, doi: [10.5479/ADS/bib/1932LicOB.16.24B](https://doi.org/10.5479/ADS/bib/1932LicOB.16.24B)
- [33] Bertaux, J.-L., Lallement, R., Ferron, S., & Boone, C. 2014, in *13th International HITRAN Conference*, 8, doi: [10.5281/zenodo.11110](https://doi.org/10.5281/zenodo.11110)
- [34] Bessel, F. W. 1844, On the variations of the proper motions of Procyon and Sirius, *Monthly Notices of the Royal Astronomical Society*, 6, 136, doi: [10.1093/mnras/6.11.136](https://doi.org/10.1093/mnras/6.11.136)
- [35] Beuzit, J. L., Vigan, A., Mouillet, D., et al. 2019, SPHERE: the exoplanet imager for the Very Large Telescope, *Astronomy and Astrophysics*, 631, A155, doi: [10.1051/0004-6361/201935251](https://doi.org/10.1051/0004-6361/201935251)
- [36] Binnenfeld, A., Shahaf, S., Anderson, R. I., & Zucker, S. 2022, New periodograms separating orbital radial velocities and spectral shape variation, *Astronomy and Astrophysics*, 659, A189, doi: [10.1051/0004-6361/202141406](https://doi.org/10.1051/0004-6361/202141406)
- [37] Binnenfeld, A., Shahaf, S., & Zucker, S. 2020, USuRPER: Unit-sphere representation periodogram for full spectra, *Astronomy and Astrophysics*, 642, A146, doi: [10.1051/0004-6361/202039001](https://doi.org/10.1051/0004-6361/202039001)

- [38] Bitsch, B., Morbidelli, A., Lega, E., & Crida, A. 2014, Stellar irradiated discs and implications on migration of embedded planets. II. Accreting-discs, *Astronomy and Astrophysics*, 564, A135, doi: [10.1051/0004-6361/201323007](https://doi.org/10.1051/0004-6361/201323007)
- [39] Blind, N., Chazelas, B., Kühn, J., et al. 2022, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 12185, *Adaptive Optics Systems VIII*, ed. L. Schreiber, D. Schmidt, & E. Vernet, 1218573, doi: [10.1117/12.2628320](https://doi.org/10.1117/12.2628320)
- [40] Boisse, I., Bonfils, X., & Santos, N. C. 2012a, SOAP. A tool for the fast computation of photometry and radial velocity induced by stellar spots, *Astronomy and Astrophysics*, 545, A109, doi: [10.1051/0004-6361/201219115](https://doi.org/10.1051/0004-6361/201219115)
- [41] Boisse, I., Moutou, C., Vidal-Madjar, A., et al. 2009, Stellar activity of planetary host star HD 189 733, *Astronomy and Astrophysics*, 495, 959, doi: [10.1051/0004-6361:200810648](https://doi.org/10.1051/0004-6361:200810648)
- [42] Boisse, I., Pepe, F., Perrier, C., et al. 2012b, The SOPHIE search for northern extrasolar planets. V. Follow-up of ELODIE candidates: Jupiter-analogs around Sun-like stars, *Astronomy and Astrophysics*, 545, A55, doi: [10.1051/0004-6361/201118419](https://doi.org/10.1051/0004-6361/201118419)
- [43] Bond, I. A., Udalski, A., Jaroszyński, M., et al. 2004, OGLE 2003-BLG-235/MOA 2003-BLG-53: A Planetary Microlensing Event, *ApJ Letters*, 606, L155, doi: [10.1086/420928](https://doi.org/10.1086/420928)
- [44] Bonfils, X., Delfosse, X., Udry, S., et al. 2013, The HARPS search for southern extrasolar planets, *Astronomy and Astrophysics*, 549, A109, doi: [10.1051/0004-6361/201014704](https://doi.org/10.1051/0004-6361/201014704)
- [45] Bortle, A., Fausey, H., Ji, J., et al. 2021, A Gaussian Process Regression Reveals No Evidence for Planets Orbiting Kapteyn’s Star, *The Astronomical Journal*, 161, 230, doi: [10.3847/1538-3881/abec89](https://doi.org/10.3847/1538-3881/abec89)
- [46] Borucki, W. J., Koch, D., Basri, G., et al. 2010, Kepler Planet-Detection Mission: Introduction and First Results, *Science*, 327, 977, doi: [10.1126/science.1185402](https://doi.org/10.1126/science.1185402)
- [47] Borucki, W. J., Koch, D. G., Batalha, N., et al. 2012, Kepler-22b: A 2.4 EARTH-RADIUS PLANET IN THE HABITABLE ZONE OF A SUN-LIKE STAR, *The Astrophysical Journal*, 745, 120, doi: [10.1088/0004-637x/745/2/120](https://doi.org/10.1088/0004-637x/745/2/120)
- [48] Bouchy, F., & Carrier, F. 2002, The acoustic spectrum of alpha Cen A, *Astronomy and Astrophysics*, 390, 205, doi: [10.1051/0004-6361:20020706](https://doi.org/10.1051/0004-6361:20020706)
- [49] Bouchy, F., Pepe, F., & Queloz, D. 2001, Fundamental photon noise limit to radial velocity measurements, *Astronomy and Astrophysics*, 374, 733, doi: [10.1051/0004-6361:20010730](https://doi.org/10.1051/0004-6361:20010730)
- [50] Brandl, B. R., Lenzen, R., Pantin, E., et al. 2012, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 8446, *Ground-based and Airborne Instrumentation for Astronomy IV*, ed. I. S. McLean, S. K. Ramsay, & H. Takami, 84461M, doi: [10.1117/12.926057](https://doi.org/10.1117/12.926057)
- [51] Bryan, M. L., Knutson, H. A., Howard, A. W., et al. 2016, Statistics of Long Period

- Gas Giant Planets in Known Planetary Systems, *The Astrophysical Journal*, 821, 89, doi: [10.3847/0004-637X/821/2/89](https://doi.org/10.3847/0004-637X/821/2/89)
- [52] Butler, R. P., Marcy, G. W., Fischer, D. A., et al. 1999, Evidence for Multiple Companions to  $\nu$  Andromedae, *The Astrophysical Journal*, 526, 916, doi: [10.1086/308035](https://doi.org/10.1086/308035)
- [53] Butler, R. P., Marcy, G. W., Williams, E., et al. 1996, Attaining Doppler Precision of 3 M s<sup>-1</sup>, , 108, 500, doi: [10.1086/133755](https://doi.org/10.1086/133755)
- [54] Cadieux, C., Doyon, R., Plotnykov, M., et al. 2022, TOI-1452 b: SPIRou and TESS Reveal a Super-Earth in a Temperate Orbit Transiting an M4 Dwarf, *The Astronomical Journal*, 164, 96, doi: [10.3847/1538-3881/ac7cea](https://doi.org/10.3847/1538-3881/ac7cea)
- [55] Carleo, I., Sanna, N., Gratton, R., et al. 2016, High precision radial velocities with GIANO spectra, *Experimental Astronomy*, 41, 351, doi: [10.1007/s10686-016-9492-3](https://doi.org/10.1007/s10686-016-9492-3)
- [56] Carmona, A., Delfosse, X., Bellotti, S., et al. 2023, Near-IR and optical radial velocities of the active M dwarf star Gl 388 (AD Leo) with SPIRou at CFHT and SOPHIE at OHP: A 2.23 day rotation period and no evidence for a corotating planet. <https://arxiv.org/abs/2303.16712>
- [57] Carter, A. L., Hinkley, S., Kammerer, J., et al. 2023, The JWST Early Release Science Program for Direct Observations of Exoplanetary Systems I: High-contrast Imaging of the Exoplanet HIP 65426 b from 2 to 16  $\mu$ m, *ApJ Letters*, 951, L20, doi: [10.3847/2041-8213/acd93e](https://doi.org/10.3847/2041-8213/acd93e)
- [58] Challita, Z., Reshetov, V., Baratchart, S., et al. 2018, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 10702, *Ground-based and Airborne Instrumentation for Astronomy VII*, ed. C. J. Evans, L. Simard, & H. Takami, 1070262, doi: [10.1117/12.2311999](https://doi.org/10.1117/12.2311999)
- [59] Chambers, J. E. 2009, Planetary Migration: What Does It Mean for Planet Formation?, *Annual Review of Earth and Planetary Sciences*, 37, 321, doi: [10.1146/annurev.earth.031208.100122](https://doi.org/10.1146/annurev.earth.031208.100122)
- [60] Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, Detection of Planetary Transits Across a Sun-like Star, *ApJ Letters*, 529, L45, doi: [10.1086/312457](https://doi.org/10.1086/312457)
- [61] Charbonneau, D., Berta, Z. K., Irwin, J., et al. 2009, A super-Earth transiting a nearby low-mass star, *Nature*, 462, 891, doi: [10.1038/nature08679](https://doi.org/10.1038/nature08679)
- [62] Charpinet, S., Fontaine, G., Brassard, P., et al. 2011, A compact system of small planets around a former red-giant star, *Nature*, 480, 496, doi: [10.1038/nature10631](https://doi.org/10.1038/nature10631)
- [63] Chen, J., & Kipping, D. 2017, Probabilistic Forecasting of the Masses and Radii of Other Worlds, *The Astrophysical Journal*, 834, 17, doi: [10.3847/1538-4357/834/1/17](https://doi.org/10.3847/1538-4357/834/1/17)
- [64] Childs, A. C., Martin, R. G., & Livio, M. 2022, Life on Exoplanets in the Habitable Zone of M Dwarfs?, *ApJ Letters*, 937, L41, doi: [10.3847/2041-8213/ac9052](https://doi.org/10.3847/2041-8213/ac9052)
- [65] Christensen-Dalsgaard, J. 2004, Physics of solar-like oscillations, *Solar Physics*, 220, 137, doi: [10.1023/B:SOLA.0000031392.43227.7d](https://doi.org/10.1023/B:SOLA.0000031392.43227.7d)

- [66] Claudi, R., Benatti, S., Carleo, I., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, ed. C. J. Evans, L. Simard, & H. Takami, 99081A, doi: [10.1117/12.2231845](https://doi.org/10.1117/12.2231845)
- [67] Clements, D. L. 2023, Venus, Phosphine and the Possibility of Life. <https://arxiv.org/abs/2301.05160>
- [68] Cloutier, R., Artigau, É., Delfosse, X., et al. 2018, Predictions of Planet Detections with Near-infrared Radial Velocities in the Upcoming SPIRou Legacy Survey-planet Search, *The Astronomical Journal*, 155, 93, doi: [10.3847/1538-3881/aaa54e](https://doi.org/10.3847/1538-3881/aaa54e)
- [69] Cockell, C. S., Bush, T., Bryce, C., et al. 2016, Habitability: A Review, *Astrobiology*, 16, 89, doi: [10.1089/ast.2015.1295](https://doi.org/10.1089/ast.2015.1295)
- [70] 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, doi: [10.1093/mnras/stz1215](https://doi.org/10.1093/mnras/stz1215)
- [71] 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, doi: [10.1093/mnras/stab1323](https://doi.org/10.1093/mnras/stab1323)
- [72] Cook, N. J., Artigau, É., Doyon, R., et al. 2022, APERO: A PipelinE to Reduce Observations, *Astrophysics Source Code Library*, record ascl:2211.019. <http://ascl.net/2211.019>
- [73] Cortes-Zuleta, P., Boisse, I., Klein, B., et al. 2023, Optical and near-infrared stellar activity characterization of the early M dwarf GL\_205 with SOPHIE and SPIRou, arXiv e-prints, arXiv:2301.10614, doi: [10.48550/arXiv.2301.10614](https://doi.org/10.48550/arXiv.2301.10614)
- [74] Cosentino, R., Lovis, C., Pepe, F., et al. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV, ed. I. S. McLean, S. K. Ramsay, & H. Takami, 84461V, doi: [10.1117/12.925738](https://doi.org/10.1117/12.925738)
- [75] Crepp, J. R., Crass, J., King, D., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, ed. C. J. Evans, L. Simard, & H. Takami, 990819, doi: [10.1117/12.2233135](https://doi.org/10.1117/12.2233135)
- [76] Cretignier, M., Dumusque, X., Aigrain, S., & Pepe, F. 2023, YARARA V2: Reaching sub m/s precision over a decade using PCA on line-by-line RVs, arXiv e-prints, arXiv:2308.11812, doi: [10.48550/arXiv.2308.11812](https://doi.org/10.48550/arXiv.2308.11812)
- [77] Cretignier, M., Dumusque, X., Allart, R., Pepe, F., & Lovis, C. 2020, Measuring precise radial velocities on individual spectral lines. II. Dependence of stellar activity signal on line depth, *Astronomy and Astrophysics*, 633, A76, doi: [10.1051/0004-6361/201936548](https://doi.org/10.1051/0004-6361/201936548)
- [78] Cretignier, M., Dumusque, X., Hara, N. C., & Pepe, F. 2021, YARARA: Significant im-

- provement in RV precision through post-processing of spectral time series, *Astronomy and Astrophysics*, 653, A43, doi: [10.1051/0004-6361/202140986](https://doi.org/10.1051/0004-6361/202140986)
- [79] Cretignier, M., Dumusque, X., & Pepe, F. 2022, Stellar activity correction using PCA decomposition of shells, *Astronomy and Astrophysics*, 659, A68, doi: [10.1051/0004-6361/202142435](https://doi.org/10.1051/0004-6361/202142435)
- [80] Cristofari, P. I., Donati, J. F., Masseron, T., et al. 2022, Estimating the atmospheric properties of 44 M dwarfs from SPIRou spectra, *Monthly Notices of the Royal Astronomical Society*, 516, 3802, doi: [10.1093/mnras/stac2364](https://doi.org/10.1093/mnras/stac2364)
- [81] Crockett, C. J., Mahmud, N. I., Prato, L., et al. 2012, A Search for Giant Planet Companions to T Tauri Stars, *The Astrophysical Journal*, 761, 164, doi: [10.1088/0004-637X/761/2/164](https://doi.org/10.1088/0004-637X/761/2/164)
- [82] Curiel, S., Ortiz-León, G. N., Mioduszewski, A. J., & Sanchez-Bermudez, J. 2022, 3D Orbital Architecture of a Dwarf Binary System and Its Planetary Companion, *The Astronomical Journal*, 164, 93, doi: [10.3847/1538-3881/ac7c66](https://doi.org/10.3847/1538-3881/ac7c66)
- [83] Davies, R., Schubert, J., Hartl, M., et al. 2016, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 9908, *Ground-based and Airborne Instrumentation for Astronomy VI*, ed. C. J. Evans, L. Simard, & H. Takami, 99081Z, doi: [10.1117/12.2233047](https://doi.org/10.1117/12.2233047)
- [84] Debras, F., Klein, B., Donati, J.-F., et al. 2023, ATMOSPHERIX: II- Characterising exoplanet atmospheres through transmission spectroscopy with SPIRou, *Monthly Notices of the Royal Astronomical Society*, doi: [10.1093/mnras/stad2608](https://doi.org/10.1093/mnras/stad2608)
- [85] Del Moro, D. 2004, Solar granulation properties derived from three different time series, *Astronomy and Astrophysics*, 428, 1007, doi: [10.1051/0004-6361:20040466](https://doi.org/10.1051/0004-6361:20040466)
- [86] Delchambre, L. 2015, Weighted principal component analysis: a weighted covariance eigendecomposition approach, *Monthly Notices of the Royal Astronomical Society*, 446, 3545, doi: [10.1093/mnras/stu2219](https://doi.org/10.1093/mnras/stu2219)
- [87] Delisle, J. B., Ségransan, D., Dumusque, X., et al. 2018, The HARPS search for southern extra-solar planets. XLIII. A compact system of four super-Earth planets orbiting HD 215152, *Astronomy and Astrophysics*, 614, A133, doi: [10.1051/0004-6361/201732529](https://doi.org/10.1051/0004-6361/201732529)
- [88] Dittmann, J. A., Irwin, J. M., Charbonneau, D., et al. 2017, A temperate rocky super-Earth transiting a nearby cool star, *Nature*, 544, 333, doi: [10.1038/nature22055](https://doi.org/10.1038/nature22055)
- [89] Donati, J. F. 2003, in *Astronomical Society of the Pacific Conference Series*, Vol. 307, *Solar Polarization*, ed. J. Trujillo-Bueno & J. Sanchez Almeida, 41
- [90] Donati, J. F., Howarth, I. D., Bouret, J. C., et al. 2006, Discovery of a strong magnetic field on the O star HD 191612: new clues to the future of  $\theta^1$  Orionis C\*, *Monthly Notices of the Royal Astronomical Society*, 365, L6, doi: [10.1111/j.1745-3933.2005.00115.x](https://doi.org/10.1111/j.1745-3933.2005.00115.x)
- [91] Donati, J.-F., Kouach, D., Lacombe, M., et al. 2018, in *Handbook of Exoplanets*, ed. H. J. Deeg & J. A. Belmonte, 107, doi: [10.1007/978-3-319-55333-7\\_107](https://doi.org/10.1007/978-3-319-55333-7_107)

- [92] Donati, J. F., Kouach, D., Moutou, C., et al. 2020, SPIRou: NIR velocimetry and spectropolarimetry at the CFHT, *Monthly Notices of the Royal Astronomical Society*, 498, 5684, doi: [10.1093/mnras/staa2569](https://doi.org/10.1093/mnras/staa2569)
- [93] Donati, J. F., Cristofari, P. I., Finocciety, B., et al. 2023a, The magnetic field and multiple planets of the young dwarf AU Mic, *Monthly Notices of the Royal Astronomical Society*, doi: [10.1093/mnras/stad1193](https://doi.org/10.1093/mnras/stad1193)
- [94] Donati, J. F., Lehmann, L. T., Cristofari, P. I., et al. 2023b, Magnetic fields & rotation periods of M dwarfs from SPIRou spectra, arXiv e-prints, arXiv:2307.14190, doi: [10.48550/arXiv.2307.14190](https://doi.org/10.48550/arXiv.2307.14190)
- [95] Dong, S., Xie, J.-W., Zhou, J.-L., Zheng, Z., & Luo, A. 2018, LAMOST telescope reveals that Neptunian cousins of hot Jupiters are mostly single offspring of stars that are rich in heavy elements, *Proceedings of the National Academy of Science*, 115, 266, doi: [10.1073/pnas.1711406115](https://doi.org/10.1073/pnas.1711406115)
- [96] Dorn, R. J., Bristow, P., Smoker, J. V., et al. 2023, CRIRES<sup>+</sup> on sky at the ESO Very Large Telescope. Observing the Universe at infrared wavelengths and high spectral resolution, *Astronomy and Astrophysics*, 671, A24, doi: [10.1051/0004-6361/202245217](https://doi.org/10.1051/0004-6361/202245217)
- [97] Dressing, C. D., & Charbonneau, D. 2015, The Occurrence of Potentially Habitable Planets Orbiting M Dwarfs Estimated from the Full Kepler Dataset and an Empirical Measurement of the Detection Sensitivity, *The Astrophysical Journal*, 807, 45, doi: [10.1088/0004-637X/807/1/45](https://doi.org/10.1088/0004-637X/807/1/45)
- [98] Dumusque, X. 2018, Measuring precise radial velocities on individual spectral lines. I. Validation of the method and application to mitigate stellar activity, *Astronomy and Astrophysics*, 620, A47, doi: [10.1051/0004-6361/201833795](https://doi.org/10.1051/0004-6361/201833795)
- [99] Dumusque, X., Boisse, I., & Santos, N. C. 2014a, SOAP 2.0: A Tool to Estimate the Photometric and Radial Velocity Variations Induced by Stellar Spots and Plages, *The Astrophysical Journal*, 796, 132, doi: [10.1088/0004-637X/796/2/132](https://doi.org/10.1088/0004-637X/796/2/132)
- [100] Dumusque, X., Lovis, C., Udry, S., & Santos, N. C. 2011, in *The Astrophysics of Planetary Systems: Formation, Structure, and Dynamical Evolution*, ed. A. Sozzetti, M. G. Lattanzi, & A. P. Boss, Vol. 276, 530–532, doi: [10.1017/S1743921311021090](https://doi.org/10.1017/S1743921311021090)
- [101] Dumusque, X., Pepe, F., Lovis, C., & Latham, D. W. 2015, Characterization of a Spurious One-year Signal in HARPS Data, *The Astrophysical Journal*, 808, 171, doi: [10.1088/0004-637X/808/2/171](https://doi.org/10.1088/0004-637X/808/2/171)
- [102] Dumusque, X., Bonomo, A. S., Haywood, R. D., et al. 2014b, The Kepler-10 Planetary System Revisited by HARPS-N: A Hot Rocky World and a Solid Neptune-Mass Planet, *The Astrophysical Journal*, 789, 154, doi: [10.1088/0004-637X/789/2/154](https://doi.org/10.1088/0004-637X/789/2/154)
- [103] Esteves, L. J., De Mooij, E. J. W., & Jayawardhana, R. 2015, Changing Phases of Alien Worlds: Probing Atmospheres of Kepler Planets with High-precision Photometry, *The Astrophysical Journal*, 804, 150, doi: [10.1088/0004-637X/804/2/150](https://doi.org/10.1088/0004-637X/804/2/150)

## BIBLIOGRAPHY

---

- [104] Faria, J. P., Suárez Mascareño, A., Figueira, P., et al. 2022, A candidate short-period sub-Earth orbiting Proxima Centauri, *Astronomy and Astrophysics*, 658, A115, doi: [10.1051/0004-6361/202142337](https://doi.org/10.1051/0004-6361/202142337)
- [105] Feng, F., Shectman, S. A., Clement, M. S., et al. 2020, Search for Nearby Earth Analogs .III. Detection of 10 New Planets, 3 Planet Candidates, and Confirmation of 3 Planets around 11 Nearby M Dwarfs, *ApJ Supplements*, 250, 29, doi: [10.3847/1538-4365/abb139](https://doi.org/10.3847/1538-4365/abb139)
- [106] Foreman-Mackey, D., Agol, E., Angus, R., & Ambikasaran, S. 2017, Fast and scalable Gaussian process modeling with applications to astronomical time series, *ArXiv*. <https://arxiv.org/abs/1703.09710>
- [107] Fouqué, P., Moutou, C., Malo, L., et al. 2018, SPIRou Input Catalogue: global properties of 440 M dwarfs observed with ESPaDOnS at CFHT, *Monthly Notices of the Royal Astronomical Society*, 475, 1960, doi: [10.1093/mnras/stx3246](https://doi.org/10.1093/mnras/stx3246)
- [108] Fouqué, P., Martioli, E., Donati, J. F., et al. 2023, The SPIRou legacy survey. Rotation period of quiet M dwarfs from circular polarization in near-infrared spectral lines: The SPIRou APERO analysis, *Astronomy and Astrophysics*, 672, A52, doi: [10.1051/0004-6361/202345839](https://doi.org/10.1051/0004-6361/202345839)
- [109] Fulton, B. J., Petigura, E. A., Howard, A. W., et al. 2017, The California-Kepler Survey. III. A Gap in the Radius Distribution of Small Planets, *The Astronomical Journal*, 154, 109, doi: [10.3847/1538-3881/aa80eb](https://doi.org/10.3847/1538-3881/aa80eb)
- [110] Gaia Collaboration. 2020, VizieR Online Data Catalog: Gaia EDR3 (Gaia Collaboration, 2020), *VizieR Online Data Catalog*, I/350
- [111] Gaidos, E., Mann, A. W., Kraus, A. L., & Ireland, M. 2016, They are small worlds after all: revised properties of Kepler M dwarf stars and their planets, *Monthly Notices of the Royal Astronomical Society*, 457, 2877, doi: [10.1093/mnras/stw097](https://doi.org/10.1093/mnras/stw097)
- [112] Garcia-Sage, K., Glocer, A., Drake, J. J., Gronoff, G., & Cohen, O. 2017, On the Magnetic Protection of the Atmosphere of Proxima Centauri b, *ApJ Letters*, 844, L13, doi: [10.3847/2041-8213/aa7eca](https://doi.org/10.3847/2041-8213/aa7eca)
- [113] Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, The James Webb Space Telescope, *Space Science Reviews*, 123, 485, doi: [10.1007/s11214-006-8315-7](https://doi.org/10.1007/s11214-006-8315-7)
- [114] Gatewood, G., & Eichhorn, H. 1973, An unsuccessful search for a planetary companion of Barnard's star BD +4 3561., *The Astronomical Journal*, 78, 769, doi: [10.1086/111480](https://doi.org/10.1086/111480)
- [115] Gillon, M., Jehin, E., Delrez, L., et al. 2013, in *Protostars and Planets VI Posters*
- [116] Gillon, M., Demory, B. O., Barman, T., et al. 2007, Accurate Spitzer infrared radius measurement for the hot Neptune GJ 436b, *Astronomy and Astrophysics*, 471, L51, doi: [10.1051/0004-6361:20078283](https://doi.org/10.1051/0004-6361:20078283)
- [117] Gillon, M., Triaud, A. H. M. J., Demory, B.-O., et al. 2017, Seven temperate terrestrial planets around the nearby ultracool dwarf star TRAPPIST-1, *Nature*, 542, 456, doi: [10.1038/nature21360](https://doi.org/10.1038/nature21360)

- [118] Glazier, A. L., Howard, W. S., Corbett, H., et al. 2020, Evryscope and K2 Constraints on TRAPPIST-1 Superflare Occurrence and Planetary Habitability, *The Astrophysical Journal*, 900, 27, doi: [10.3847/1538-4357/aba4a6](https://doi.org/10.3847/1538-4357/aba4a6)
- [119] Gomes da Silva, J., Santos, N. C., Bonfils, X., et al. 2012, VizieR Online Data Catalog: M dwarfs activity and radial velocity (Gomes da Silva+, 2012), *VizieR Online Data Catalog*, J/A+A/541/A9
- [120] Gomes da Silva, J., Santos, N. C., Adibekyan, V., et al. 2021, Stellar chromospheric activity of 1674 FGK stars from the AMBRE-HARPS sample. I. A catalogue of homogeneous chromospheric activity, *Astronomy and Astrophysics*, 646, A77, doi: [10.1051/0004-6361/202039765](https://doi.org/10.1051/0004-6361/202039765)
- [121] Gorrini, P., Astudillo-Defru, N., Dreizler, S., et al. 2022, Detailed stellar activity analysis and modelling of GJ 832, *Astronomy & Astrophysics*, 664, A64, doi: [10.1051/0004-6361/202243063](https://doi.org/10.1051/0004-6361/202243063)
- [122] Guinan, E. F., & Engle, S. G. 2013, in *American Astronomical Society Meeting Abstracts*, Vol. 221, *American Astronomical Society Meeting Abstracts #221*, 333.02
- [123] Gullikson, K. 2014, TelFit: Fitting the telluric absorption spectrum, *Astrophysics Source Code Library*, record ascl:1405.002. <http://ascl.net/1405.002>
- [124] Günther, M. N., Zhan, Z., Seager, S., et al. 2020, Stellar Flares from the First TESS Data Release: Exploring a New Sample of M Dwarfs, *The Astronomical Journal*, 159, 60, doi: [10.3847/1538-3881/ab5d3a](https://doi.org/10.3847/1538-3881/ab5d3a)
- [125] Hale, G. E. 1908, On the Probable Existence of a Magnetic Field in Sun-Spots, *The Astrophysical Journal*, 28, 315, doi: [10.1086/141602](https://doi.org/10.1086/141602)
- [126] Halverson, S., Roy, A., Mahadevan, S., & Schwab, C. 2015, “Modal Noise” in Single-mode Fibers: A Cautionary Note for High Precision Radial Velocity Instruments, *ApJ Letters*, 814, L22, doi: [10.1088/2041-8205/814/2/L22](https://doi.org/10.1088/2041-8205/814/2/L22)
- [127] Haywood, R. D., Collier Cameron, A., Queloz, D., et al. 2014, Disentangling planetary orbits from stellar activity in radial-velocity surveys, *International Journal of Astrobiology*, 13, 155, doi: [10.1017/S147355041300044X](https://doi.org/10.1017/S147355041300044X)
- [128] Haywood, R. D., Collier Cameron, A., Dumusque, X., et al. 2016, in *19th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun (CS19)*, *Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun*, 47, doi: [10.5281/zenodo.55693](https://doi.org/10.5281/zenodo.55693)
- [129] Hébrard, É. M., Donati, J. F., Delfosse, X., et al. 2016, Modelling the RV jitter of early-M dwarfs using tomographic imaging, *Monthly Notices of the Royal Astronomical Society*, 461, 1465, doi: [10.1093/mnras/stw1346](https://doi.org/10.1093/mnras/stw1346)
- [130] Heintz, W. D. 1978, Reexamination of suspected unresolved binaries., *The Astrophysical Journal*, 220, 931, doi: [10.1086/155982](https://doi.org/10.1086/155982)
- [131] —. 1987, The red-dwarf binary sigma 2398., , 99, 1084, doi: [10.1086/132082](https://doi.org/10.1086/132082)
- [132] —. 1988, The binary star 70 Ophiuchi revisited., , 82, 140

## BIBLIOGRAPHY

---

- [133] Henry, T. J., Jao, W.-C., Subasavage, J. P., et al. 2006, The Solar Neighborhood. XVII. Parallax Results from the CTIOPI 0.9 m Program: 20 New Members of the RECONS 10 Parsec Sample, *The Astronomical Journal*, 132, 2360, doi: [10.1086/508233](https://doi.org/10.1086/508233)
- [134] Hershey, J. L. 1973, Astrometric analysis of the field of AC +65 6955 from plates taken with the Sproul 24-inch refractor., *The Astronomical Journal*, 78, 421, doi: [10.1086/111436](https://doi.org/10.1086/111436)
- [135] Hirano, T., Kuzuhara, M., Kotani, T., et al. 2020, Precision radial velocity measurements by the forward-modeling technique in the near-infrared, , 72, 93, doi: [10.1093/pasj/psaa085](https://doi.org/10.1093/pasj/psaa085)
- [136] Hobson, M. J., Bouchy, F., Cook, N. J., et al. 2021, The SPIRou wavelength calibration for precise radial velocities in the near infrared, *Astronomy and Astrophysics*, 648, A48, doi: [10.1051/0004-6361/202038413](https://doi.org/10.1051/0004-6361/202038413)
- [137] Hodapp, K. W., Jensen, J. B., Irwin, E. M., et al. 2003, The Gemini Near-Infrared Imager (NIRI), , 115, 1388, doi: [10.1086/379669](https://doi.org/10.1086/379669)
- [138] Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, Planet Occurrence within 0.25 AU of Solar-type Stars from Kepler, *ApJ Supplements*, 201, 15, doi: [10.1088/0067-0049/201/2/15](https://doi.org/10.1088/0067-0049/201/2/15)
- [139] Howell, S. B., Sobeck, C., Haas, M., et al. 2014, The K2 Mission: Characterization and Early Results, , 126, 398, doi: [10.1086/676406](https://doi.org/10.1086/676406)
- [140] Hsu, D. C., Ford, E. B., & Terrien, R. 2020, Occurrence rates of planets orbiting M Stars: applying ABC to Kepler DR25, Gaia DR2, and 2MASS data, *Monthly Notices of the Royal Astronomical Society*, 498, 2249, doi: [10.1093/mnras/staa2391](https://doi.org/10.1093/mnras/staa2391)
- [141] Ilin, E., Poppenhäger, K., Schmidt, S. J., et al. 2021, in *Plato Mission Conference 2021. Presentations and posters of the online PLATO Mission Conference 2021*, 54, doi: [10.5281/zenodo.5559296](https://doi.org/10.5281/zenodo.5559296)
- [142] Jacob, W. S. 1855, On certain Anomalies presented by the Binary Star 70 Ophiuchi, *Monthly Notices of the Royal Astronomical Society*, 15, 228, doi: [10.1093/mnras/15.9.228](https://doi.org/10.1093/mnras/15.9.228)
- [143] Jarvis, M., & Jain, B. 2004, Principal Component Analysis of PSF Variation in Weak Lensing Surveys. <https://arxiv.org/abs/astro-ph/0412234>
- [144] Johnson, S. A., Penny, M., Gaudi, B. S., et al. 2020, Predictions of the Nancy Grace Roman Space Telescope Galactic Exoplanet Survey. II. Free-floating Planet Detection Rates, *The Astronomical Journal*, 160, 123, doi: [10.3847/1538-3881/aba75b](https://doi.org/10.3847/1538-3881/aba75b)
- [145] Jolliffe, I. T. 1986, *Principal component analysis*
- [146] Joshi, M. M., Haberle, R. M., & Reynolds, R. T. 1997, Simulations of the Atmospheres of Synchronously Rotating Terrestrial Planets Orbiting M Dwarfs: Conditions for Atmospheric Collapse and the Implications for Habitability, *Icarus*, 129, 450, doi: [10.1006/icar.1997.5793](https://doi.org/10.1006/icar.1997.5793)

- [147] Jurgenson, C., Fischer, D., McCracken, T., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, ed. C. J. Evans, L. Simard, & H. Takami, 99086T, doi: [10.1117/12.2233002](https://doi.org/10.1117/12.2233002)
- [148] Kane, S. R., Barclay, T., & Gelino, D. M. 2013, A Potential Super-Venus in the Kepler-69 System, *ApJ Letters*, 770, L20, doi: [10.1088/2041-8205/770/2/L20](https://doi.org/10.1088/2041-8205/770/2/L20)
- [149] Kasting, J. F., Whitmire, D. P., & Reynolds, R. T. 1993, Habitable Zones around Main Sequence Stars, *Icarus*, 101, 108, doi: [10.1006/icar.1993.1010](https://doi.org/10.1006/icar.1993.1010)
- [150] Khodachenko, M. L., Ribas, I., Lammer, H., et al. 2007, Coronal Mass Ejection (CME) Activity of Low Mass M Stars as An Important Factor for The Habitability of Terrestrial Exoplanets. I. CME Impact on Expected Magnetospheres of Earth-Like Exoplanets in Close-In Habitable Zones, *Astrobiology*, 7, 167, doi: [10.1089/ast.2006.0127](https://doi.org/10.1089/ast.2006.0127)
- [151] Kiefer, F., Hébrard, G., Lecavelier des Etangs, A., et al. 2021, Determining the true mass of radial-velocity exoplanets with Gaia. Nine planet candidates in the brown dwarf or stellar regime and 27 confirmed planets, *Astronomy and Astrophysics*, 645, A7, doi: [10.1051/0004-6361/202039168](https://doi.org/10.1051/0004-6361/202039168)
- [152] Kiefer, F., Hébrard, G., Martioli, E., et al. 2023, A sub-Neptune planet around TOI-1695 discovered and characterized with SPIRou and TESS, *Astronomy and Astrophysics*, 670, A136, doi: [10.1051/0004-6361/202245129](https://doi.org/10.1051/0004-6361/202245129)
- [153] Kipping, D. M. 2013, Parametrizing the exoplanet eccentricity distribution with the beta distribution., *Monthly Notices of the Royal Astronomical Society*, 434, L51, doi: [10.1093/mnrasl/slt075](https://doi.org/10.1093/mnrasl/slt075)
- [154] Kirkpatrick, J. D., Henry, T. J., & McCarthy, Donald W., J. 1991, A Standard Stellar Spectral Sequence in the Red/Near-Infrared: Classes K5 to M9, *ApJ Supplements*, 77, 417, doi: [10.1086/191611](https://doi.org/10.1086/191611)
- [155] Klein, B., Debras, F., Donati, J.-F., et al. 2023, ATMOSPHERIX: I- An open source high resolution transmission spectroscopy pipeline for exoplanets atmospheres with SPIRou, *Monthly Notices of the Royal Astronomical Society*, doi: [10.1093/mnras/stad2607](https://doi.org/10.1093/mnras/stad2607)
- [156] Kopparapu, R. K., Ramirez, R. M., SchottelKotte, J., et al. 2014, Habitable Zones around Main-sequence Stars: Dependence on Planetary Mass, *ApJ Letters*, 787, L29, doi: [10.1088/2041-8205/787/2/L29](https://doi.org/10.1088/2041-8205/787/2/L29)
- [157] Kossakowski, D., Kürster, M., Trifonov, T., et al. 2023, The CARMENES search for exoplanets around M dwarfs. Wolf 1069 b: Earth-mass planet in the habitable zone of a nearby, very low-mass star, *Astronomy and Astrophysics*, 670, A84, doi: [10.1051/0004-6361/202245322](https://doi.org/10.1051/0004-6361/202245322)
- [158] Kotani, T., Tamura, M., Suto, H., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V, ed. S. K. Ramsay, I. S. McLean, & H. Takami, 914714, doi: [10.1117/12.2055075](https://doi.org/10.1117/12.2055075)

## BIBLIOGRAPHY

---

- [159] Kullback, S., & Leibler, R. A. 1951, On Information and Sufficiency, *Ann. Math. Statist.*, 22, 79
- [160] 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 and Astrophysics*, 636, A36, doi: [10.1051/0004-6361/201937222](https://doi.org/10.1051/0004-6361/201937222)
- [161] 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, doi: [10.1038/339038a0](https://doi.org/10.1038/339038a0)
- [162] Latouf, N., Wang, S. X., Cale, B., & Plavchan, P. 2022, Characterizing and Mitigating Telluric Absorption in Precise Radial Velocities. II. A Study of an M2-type Star, *The Astronomical Journal*, 164, 212, doi: [10.3847/1538-3881/ac947b](https://doi.org/10.3847/1538-3881/ac947b)
- [163] Leighton, R. B., Noyes, R. W., & Simon, G. W. 1962, Velocity Fields in the Solar Atmosphere. I. Preliminary Report., *The Astrophysical Journal*, 135, 474, doi: [10.1086/147285](https://doi.org/10.1086/147285)
- [164] Lépine, S., Hilton, E. J., Mann, A. W., et al. 2013, A Spectroscopic Catalog of the Brightest ( $J < 9$ ) M Dwarfs in the Northern Sky, *The Astronomical Journal*, 145, 102, doi: [10.1088/0004-6256/145/4/102](https://doi.org/10.1088/0004-6256/145/4/102)
- [165] Lillo-Box, J., Millholland, S., & Laughlin, G. 2021, Follow-up of non-transiting planets detected by Kepler. Confirmation of three hot-Jupiters and validation of three other planets, *Astronomy and Astrophysics*, 654, A9, doi: [10.1051/0004-6361/202140746](https://doi.org/10.1051/0004-6361/202140746)
- [166] Lopez, E. D., & Fortney, J. J. 2013, The Role of Core Mass in Controlling Evaporation: The Kepler Radius Distribution and the Kepler-36 Density Dichotomy, *The Astrophysical Journal*, 776, 2, doi: [10.1088/0004-637X/776/1/2](https://doi.org/10.1088/0004-637X/776/1/2)
- [167] Lovis, C., Dumusque, X., Santos, N. C., Udry, S., & Mayor, M. 2011, in *AAS/Division for Extreme Solar Systems Abstracts, Vol. 2, AAS/Division for Extreme Solar Systems Abstracts, 2.02*
- [168] Lovis, C., & Fischer, D. 2010, in *Exoplanets*, ed. S. Seager, 27–53
- [169] Lovis, C., Snellen, I., Mouillet, D., et al. 2017, Atmospheric characterization of Proxima b by coupling the SPHERE high-contrast imager to the ESPRESSO spectrograph, *Astronomy and Astrophysics*, 599, A16, doi: [10.1051/0004-6361/201629682](https://doi.org/10.1051/0004-6361/201629682)
- [170] Lubin, J., & HPF Science. 2022, in *Bulletin of the American Astronomical Society, Vol. 54, 403.02*
- [171] Lueftinger, T., Tinetti, G., Eccleston, P., & Salvignol, J.-C. 2022, in *Bulletin of the American Astronomical Society, Vol. 54, 502.03*
- [172] Madhusudhan, N., Sarkar, S., Constantinou, S., et al. 2023, Carbon-bearing Molecules in a Possible Hycean Atmosphere. <https://arxiv.org/abs/2309.05566>
- [173] Mahadevan, S., Ramsey, L. W., Terrien, R., et al. 2014, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9147, Ground-based and*

- Airborne Instrumentation for Astronomy V, ed. S. K. Ramsay, I. S. McLean, & H. Takami, 91471G, doi: [10.1117/12.2056417](https://doi.org/10.1117/12.2056417)
- [174] Mahmud, N. I., Crockett, C. J., Johns-Krull, C. M., et al. 2011, Starspot-induced Optical and Infrared Radial Velocity Variability in T Tauri Star Hubble I 4, *The Astrophysical Journal*, 736, 123, doi: [10.1088/0004-637X/736/2/123](https://doi.org/10.1088/0004-637X/736/2/123)
- [175] Marconi, A., Abreu, M., Adibekyan, V., et al. 2022, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 12184, *Ground-based and Airborne Instrumentation for Astronomy IX*, ed. C. J. Evans, J. J. Bryant, & K. Motohara, 1218424, doi: [10.1117/12.2628689](https://doi.org/10.1117/12.2628689)
- [176] Marcy, G. W., & Butler, R. P. 1995, in *American Astronomical Society Meeting Abstracts*, Vol. 187, *American Astronomical Society Meeting Abstracts*, 70.04
- [177] Marcy, G. W., & Butler, R. P. 1996, A Planetary Companion to 70 Virginis, *ApJ Letters*, 464, L147, doi: [10.1086/310096](https://doi.org/10.1086/310096)
- [178] Marois, C., Macintosh, B., Barman, T., et al. 2008, Direct Imaging of Multiple Planets Orbiting the Star HR 8799, *Science*, 322, 1348, doi: [10.1126/science.1166585](https://doi.org/10.1126/science.1166585)
- [179] Marois, C., Zuckerman, B., Konopacky, Q. M., Macintosh, B., & Barman, T. 2010, Images of a fourth planet orbiting HR 8799, *Nature*, 468, 1080, doi: [10.1038/nature09684](https://doi.org/10.1038/nature09684)
- [180] Martioli, E., Hébrard, G., Moutou, C., et al. 2020, Spin-orbit alignment and magnetic activity in the young planetary system AU Mic, *Astronomy and Astrophysics*, 641, L1, doi: [10.1051/0004-6361/202038695](https://doi.org/10.1051/0004-6361/202038695)
- [181] Martioli, E., Hébrard, G., Fouqué, P., et al. 2022, TOI-1759 b: A transiting sub-Neptune around a low mass star characterized with SPIRou and TESS, *Astronomy and Astrophysics*, 660, A86, doi: [10.1051/0004-6361/202142540](https://doi.org/10.1051/0004-6361/202142540)
- [182] Mayor, M., & Queloz, D. 1995, A Jupiter-mass companion to a solar-type star, *Nature*, 378, 355, doi: [10.1038/378355a0](https://doi.org/10.1038/378355a0)
- [183] Mayor, M., Pepe, F., Queloz, D., et al. 2003, Setting New Standards with HARPS, *The Messenger*, 114, 20
- [184] Mayor, M., Marmier, M., Lovis, C., et al. 2011, The HARPS search for southern extra-solar planets XXXIV. Occurrence, mass distribution and orbital properties of super-Earths and Neptune-mass planets, *arXiv e-prints*, arXiv:1109.2497, doi: [10.48550/arXiv.1109.2497](https://doi.org/10.48550/arXiv.1109.2497)
- [185] Mazeh, T., Holczer, T., & Faigler, S. 2016, Dearth of short-period Neptunian exoplanets: A desert in period-mass and period-radius planes, *Astronomy and Astrophysics*, 589, A75, doi: [10.1051/0004-6361/201528065](https://doi.org/10.1051/0004-6361/201528065)
- [186] Meadows, V. S., Arney, G. N., Schwieterman, E. W., et al. 2018, The Habitability of Proxima Centauri b: Environmental States and Observational Discriminants, *Astrobiology*, 18, 133, doi: [10.1089/ast.2016.1589](https://doi.org/10.1089/ast.2016.1589)

## BIBLIOGRAPHY

---

- [187] Mennesson, B., Gaudi, S., Seager, S., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9904, Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave, ed. H. A. MacEwen, G. G. Fazio, M. Lystrup, N. Batalha, N. Siegler, & E. C. Tong, 99040L, doi: [10.1117/12.2240457](https://doi.org/10.1117/12.2240457)
- [188] Merlis, T. M., & Schneider, T. 2010, Atmospheric dynamics of Earth-like tidally locked aquaplanets, *Journal of Advances in Modeling Earth Systems*, 2, 13, doi: [10.3894/JAMES.2010.2.13](https://doi.org/10.3894/JAMES.2010.2.13)
- [189] Metcalf, A. J., Anderson, T. B., Bender, C. F., et al. 2019, Stellar spectroscopy in the near-infrared with a laser frequency comb, *Optica*
- [190] Meunier, N., Lagrange, A. M., Mbemba Kabuiku, L., et al. 2017, Variability of stellar granulation and convective blueshift with spectral type and magnetic activity. I. K and G main sequence stars, *Astronomy and Astrophysics*, 597, A52, doi: [10.1051/0004-6361/201629052](https://doi.org/10.1051/0004-6361/201629052)
- [191] Micheau, Y., Kouach, D., Donati, J.-F., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10702, Ground-based and Airborne Instrumentation for Astronomy VII, ed. C. J. Evans, L. Simard, & H. Takami, 107025R, doi: [10.1117/12.2305937](https://doi.org/10.1117/12.2305937)
- [192] Mignon, L., Meunier, N., Delfosse, X., et al. 2023, Characterisation of stellar activity of M dwarfs. I. Long-timescale variability in a large sample and detection of new cycles, arXiv e-prints, arXiv:2303.03998, doi: [10.48550/arXiv.2303.03998](https://doi.org/10.48550/arXiv.2303.03998)
- [193] Mishra, L., Alibert, Y., Udry, S., & Mordasini, C. 2023, Framework for the architecture of exoplanetary systems. I. Four classes of planetary system architecture, *Astronomy and Astrophysics*, 670, A68, doi: [10.1051/0004-6361/202243751](https://doi.org/10.1051/0004-6361/202243751)
- [194] Mondrik, N., Newton, E., Charbonneau, D., & Irwin, J. 2019, An Increased Rate of Large Flares at Intermediate Rotation Periods for Mid-to-late M Dwarfs, *The Astrophysical Journal*, 870, 10, doi: [10.3847/1538-4357/aabee64](https://doi.org/10.3847/1538-4357/aabee64)
- [195] Morbidelli, A. 2014, Scenarios of giant planet formation and evolution and their impact on the formation of habitable terrestrial planets, *Philosophical Transactions of the Royal Society of London Series A*, 372, 20130072, doi: [10.1098/rsta.2013.0072](https://doi.org/10.1098/rsta.2013.0072)
- [196] Mordasini, C. 2018, in 42nd COSPAR Scientific Assembly, Vol. 42, E4.2-1-18
- [197] Mordasini, C. 2020, Planetary evolution with atmospheric photoevaporation. I. Analytical derivation and numerical study of the evaporation valley and transition from super-Earths to sub-Neptunes, *Astronomy and Astrophysics*, 638, A52, doi: [10.1051/0004-6361/201935541](https://doi.org/10.1051/0004-6361/201935541)
- [198] Morin, J., Donati, J. F., Petit, P., et al. 2010, Large-scale magnetic topologies of late M dwarfs\*, *Monthly Notices of the Royal Astronomical Society*, 407, 2269, doi: [10.1111/j.1365-2966.2010.17101.x](https://doi.org/10.1111/j.1365-2966.2010.17101.x)
- [199] —. 2008, Large-scale magnetic topologies of mid M dwarfs, *Monthly Notices of the Royal Astronomical Society*, 390, 567, doi: [10.1111/j.1365-2966.2008.13809.x](https://doi.org/10.1111/j.1365-2966.2008.13809.x)

- [200] Morley, C. V., Kreidberg, L., Rustamkulov, Z., Robinson, T., & Fortney, J. J. 2017, Observing the Atmospheres of Known Temperate Earth-sized Planets with JWST, *The Astrophysical Journal*, 850, 121, doi: [10.3847/1538-4357/aa927b](https://doi.org/10.3847/1538-4357/aa927b)
- [201] Morton, T. D., Bryson, S. T., Coughlin, J. L., et al. 2016, FALSE POSITIVE PROBABILITIES FOR ALLiKEPLER/iOBJECTS OF INTEREST: 1284 NEWLY VALIDATED PLANETS AND 428 LIKELY FALSE POSITIVES, *The Astrophysical Journal*, 822, 86, doi: [10.3847/0004-637x/822/2/86](https://doi.org/10.3847/0004-637x/822/2/86)
- [202] Moulton, F. R. 1899, The limits of temporary stability of satellite motion, with an application to the question of the existence of an unseen body in the binary system 70 Ophiuchi, *The Astronomical Journal*, 20, 33, doi: [10.1086/103096](https://doi.org/10.1086/103096)
- [203] Moutou, C., Hébrard, E. M., Morin, J., et al. 2017, SPIRou input catalogue: activity, rotation and magnetic field of cool dwarfs, *Monthly Notices of the Royal Astronomical Society*, 472, 4563, doi: [10.1093/mnras/stx2306](https://doi.org/10.1093/mnras/stx2306)
- [204] Moutou, C., Delfosse, X., Petit, A. C., et al. 2023, Characterizing planetary systems with SPIRou: M-dwarf planet-search survey and the multiplanet systems GJ 876 and GJ 1148, arXiv e-prints, arXiv:2307.11569, doi: [10.48550/arXiv.2307.11569](https://doi.org/10.48550/arXiv.2307.11569)
- [205] Murphy, S. J., Bedding, T. R., & Shibahashi, H. 2016, A Planet in an 840 Day Orbit around a Kepler Main-sequence A Star Found from Phase Modulation of Its Pulsations, *ApJ Letters*, 827, L17, doi: [10.3847/2041-8205/827/1/L17](https://doi.org/10.3847/2041-8205/827/1/L17)
- [206] Naponiello, L., Mancini, L., Sozzetti, A., et al. 2023, A super-massive Neptune-sized planet, *Nature*, 1–6, doi: [10.1038/s41586-023-06499-2](https://doi.org/10.1038/s41586-023-06499-2)
- [207] Nutzman, P., & Charbonneau, D. 2008, Design Considerations for a Ground-Based Transit Search for Habitable Planets Orbiting M Dwarfs, , 120, 317, doi: [10.1086/533420](https://doi.org/10.1086/533420)
- [208] Ould-Elhkim, M., Moutou, C., Donati, J. F., et al. 2023, Wapiti: A data-driven approach to correct for systematics in RV data. Application to SPIRou data of the planet-hosting M dwarf GJ 251, *Astronomy and Astrophysics*, 675, A187, doi: [10.1051/0004-6361/202346472](https://doi.org/10.1051/0004-6361/202346472)
- [209] Owen, J. E., & Lai, D. 2018, Photoevaporation and high-eccentricity migration created the sub-Jovian desert, *Monthly Notices of the Royal Astronomical Society*, 479, 5012, doi: [10.1093/mnras/sty1760](https://doi.org/10.1093/mnras/sty1760)
- [210] Owen, J. E., & Wu, Y. 2013, Kepler Planets: A Tale of Evaporation, *The Astrophysical Journal*, 775, 105, doi: [10.1088/0004-637X/775/2/105](https://doi.org/10.1088/0004-637X/775/2/105)
- [211] Parès, L., Donati, J. F., Dupieux, M., et al. 2012, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 8446, *Ground-based and Airborne Instrumentation for Astronomy IV*, ed. I. S. McLean, S. K. Ramsay, & H. Takami, 84462E, doi: [10.1117/12.925410](https://doi.org/10.1117/12.925410)
- [212] Pâris, I., Petitjean, P., Aubourg, É., et al. 2014, The Sloan Digital Sky Survey quasar catalog: tenth data release, *Astronomy and Astrophysics*, 563, A54, doi: [10.1051/0004-6361/201322691](https://doi.org/10.1051/0004-6361/201322691)

## BIBLIOGRAPHY

---

- [213] Pearson, K. 1915, On the Partial Correlation Ratio, Proceedings of the Royal Society of London Series A, 91, 492, doi: [10.1098/rspa.1915.0041](https://doi.org/10.1098/rspa.1915.0041)
- [214] Penny, M. T., Gaudi, B. S., Kerins, E., et al. 2019, Predictions of the WFIRST Microlensing Survey. I. Bound Planet Detection Rates, ApJ Supplements, 241, 3, doi: [10.3847/1538-4365/aafb69](https://doi.org/10.3847/1538-4365/aafb69)
- [215] Pepe, F., & Lovis, C. 2008, in 2007 ESO Instrument Calibration Workshop, ed. A. Kaufer & F. Kerber, 375, doi: [10.1007/978-3-540-76963-7\\_50](https://doi.org/10.1007/978-3-540-76963-7_50)
- [216] Pepe, F., Mayor, M., Rupprecht, G., et al. 2002, HARPS: ESO's coming planet searcher. Chasing exoplanets with the La Silla 3.6-m telescope, The Messenger, 110, 9
- [217] Pepe, F., Cameron, A. C., Latham, D. W., et al. 2013, An Earth-sized planet with an Earth-like density, Nature, 503, 377, doi: [10.1038/nature12768](https://doi.org/10.1038/nature12768)
- [218] Pepe, F., Cristiani, S., Rebolo, R., et al. 2021, ESPRESSO at VLT. On-sky performance and first results, Astronomy and Astrophysics, 645, A96, doi: [10.1051/0004-6361/202038306](https://doi.org/10.1051/0004-6361/202038306)
- [219] Pepper, J., Pogge, R. W., DePoy, D. L., et al. 2007, The Kilodegree Extremely Little Telescope (KELT): A Small Robotic Telescope for Large-Area Synoptic Surveys, , 119, 923, doi: [10.1086/521836](https://doi.org/10.1086/521836)
- [220] Perruchot, S., Kohler, D., Bouchy, F., et al. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II, ed. I. S. McLean & M. M. Casali, 70140J, doi: [10.1117/12.787379](https://doi.org/10.1117/12.787379)
- [221] Perryman, M. 2011, The Exoplanet Handbook
- [222] Phillips, D. F., Glenday, A. G., Li, C.-H., et al. 2012, Calibration of an astrophysical spectrograph below 1 m/s using a laser frequency comb, Optics Express, 20, 13711, doi: [10.1364/OE.20.013711](https://doi.org/10.1364/OE.20.013711)
- [223] Pollacco, D. L., Skillen, I., Collier Cameron, A., et al. 2006, The WASP Project and the SuperWASP Cameras, , 118, 1407, doi: [10.1086/508556](https://doi.org/10.1086/508556)
- [224] Potter, S. B., Romero-Colmenero, E., Ramsay, G., et al. 2011, Possible detection of two giant extrasolar planets orbiting the eclipsing polar UZ Fornacis, Monthly Notices of the Royal Astronomical Society, 416, 2202, doi: [10.1111/j.1365-2966.2011.19198.x](https://doi.org/10.1111/j.1365-2966.2011.19198.x)
- [225] Qian, S. B., Liao, W. P., Zhu, L. Y., & Dai, Z. B. 2010, Detection of a Giant Extrasolar Planet Orbiting the Eclipsing Polar DP Leo, ApJ Letters, 708, L66, doi: [10.1088/2041-8205/708/1/L66](https://doi.org/10.1088/2041-8205/708/1/L66)
- [226] Qian, S. B., Liu, L., Liao, W. P., et al. 2011, Detection of a planetary system orbiting the eclipsing polar HU Aqr, Monthly Notices of the Royal Astronomical Society, 414, L16, doi: [10.1111/j.1745-3933.2011.01045.x](https://doi.org/10.1111/j.1745-3933.2011.01045.x)
- [227] Queloz, D., Henry, G. W., Sivan, J. P., et al. 2001, No planet for HD 166435, Astronomy and Astrophysics, 379, 279, doi: [10.1051/0004-6361:20011308](https://doi.org/10.1051/0004-6361:20011308)

- [228] 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, doi: [10.1117/12.2313689](https://doi.org/10.1117/12.2313689)
- [229] 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, doi: [10.1093/mnras/stv1428](https://doi.org/10.1093/mnras/stv1428)
- [230] Ranalli, P., Hobbs, D., & Lindegren, L. 2018, Astrometry and exoplanets in the Gaia era: a Bayesian approach to detection and parameter recovery, *Astronomy and Astrophysics*, 614, A30, doi: [10.1051/0004-6361/201730921](https://doi.org/10.1051/0004-6361/201730921)
- [231] Rasmussen, C. E., & Williams, C. K. I. 2006, *Gaussian processes for machine learning.*, Adaptive computation and machine learning (MIT Press), I–XVIII, 1–248
- [232] Rauer, H., Catala, C., Aerts, C., et al. 2014, The PLATO 2.0 mission, *Experimental Astronomy*, 38, 249, doi: [10.1007/s10686-014-9383-4](https://doi.org/10.1007/s10686-014-9383-4)
- [233] Rayner, J., Tokunaga, A., Jaffe, D., et al. 2022, iSHELL: a 1-5 micron R = 80,000 Immersion Grating Spectrograph for the NASA Infrared Telescope Facility, , 134, 015002, doi: [10.1088/1538-3873/ac3cb4](https://doi.org/10.1088/1538-3873/ac3cb4)
- [234] Rein, H., & Liu, S. F. 2012, REBOUND: an open-source multi-purpose N-body code for collisional dynamics, *Astronomy and Astrophysics*, 537, A128, doi: [10.1051/0004-6361/201118085](https://doi.org/10.1051/0004-6361/201118085)
- [235] Reiners, A. 2009, Activity-induced radial velocity jitter in a flaring M dwarf, *Astronomy and Astrophysics*, 498, 853, doi: [10.1051/0004-6361/200810257](https://doi.org/10.1051/0004-6361/200810257)
- [236] Reiners, A., Bean, J. L., Huber, K. F., et al. 2010, Detecting Planets Around Very Low Mass Stars with the Radial Velocity Method, *The Astrophysical Journal*, 710, 432, doi: [10.1088/0004-637X/710/1/432](https://doi.org/10.1088/0004-637X/710/1/432)
- [237] Reiners, A., & Zechmeister, M. 2020, Radial Velocity Photon Limits for the Dwarf Stars of Spectral Classes F–M, *ApJ Supplements*, 247, 11, doi: [10.3847/1538-4365/ab609f](https://doi.org/10.3847/1538-4365/ab609f)
- [238] Reiners, A., Zechmeister, M., Caballero, J. A., et al. 2018, The CARMENES search for exoplanets around M dwarfs. High-resolution optical and near-infrared spectroscopy of 324 survey stars, *Astronomy and Astrophysics*, 612, A49, doi: [10.1051/0004-6361/201732054](https://doi.org/10.1051/0004-6361/201732054)
- [239] Reuyl, D., & Holmberg, E. 1943, On the Existence of a Third Component in the System 70 Ophiuchi., *The Astrophysical Journal*, 97, 41, doi: [10.1086/144489](https://doi.org/10.1086/144489)
- [240] Reylé, C. 2021, in *SF2A-2021: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*. Eds.: A. Siebert, ed. A. Siebert, K. Baillié, E. Lagarde, N. Lagarde, J. Malzac, J. B. Marquette, M. N’Diaye, J. Richard, & O. Venot, 451–454
- [241] Ribas, I., Tuomi, M., Reiners, A., et al. 2018, A candidate super-Earth planet orbiting near the snow line of Barnard’s star, *Nature*, 563, 365, doi: [10.1038/s41586-018-0677-y](https://doi.org/10.1038/s41586-018-0677-y)

- [242] Ribas, I., Reiners, A., Zechmeister, M., et al. 2023, The CARMENES search for exoplanets around M dwarfs, *Astronomy and Astrophysics*, 670, A139, doi: [10.1051/0004-6361/202244879](https://doi.org/10.1051/0004-6361/202244879)
- [243] Ricker, G. R., & TESS Team. 2015, in *American Astronomical Society Meeting Abstracts*, Vol. 225, American Astronomical Society Meeting Abstracts #225, 202.01
- [244] Ridgway, R. J., Zamyatina, M., Mayne, N. J., et al. 2023, 3D modelling of the impact of stellar activity on tidally locked terrestrial exoplanets: atmospheric composition and habitability, *Monthly Notices of the Royal Astronomical Society*, 518, 2472, doi: [10.1093/mnras/stac3105](https://doi.org/10.1093/mnras/stac3105)
- [245] Robertson, P., & Mahadevan, S. 2014, Disentangling Planets and Stellar Activity for Gliese 667C, *ApJ Letters*, 793, L24, doi: [10.1088/2041-8205/793/2/L24](https://doi.org/10.1088/2041-8205/793/2/L24)
- [246] Robertson, P., Mahadevan, S., Endl, M., & Roy, A. 2014, Stellar activity masquerading as planets in the habitable zone of the M dwarf Gliese 581, *Science*, 345, 440, doi: [10.1126/science.1253253](https://doi.org/10.1126/science.1253253)
- [247] Rodriguez-Lopez, C., Gizis, J. E., MacDonald, J., & Amado, P. J. 2015, in *Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun*, Vol. 18, 18th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, 253–256
- [248] Route, M. 2016, The Discovery of Solar-like Activity Cycles Beyond the End of the Main Sequence?, *ApJ Letters*, 830, L27, doi: [10.3847/2041-8205/830/2/L27](https://doi.org/10.3847/2041-8205/830/2/L27)
- [249] Ryu, Y. H., Yee, J. C., Udalski, A., et al. 2018, OGLE-2016-BLG-1190Lb: The First Spitzer Bulge Planet Lies Near the Planet/Brown-dwarf Boundary, *The Astronomical Journal*, 155, 40, doi: [10.3847/1538-3881/aa9be4](https://doi.org/10.3847/1538-3881/aa9be4)
- [250] Saar, S. H., & Donahue, R. A. 1997, Activity-Related Radial Velocity Variation in Cool Stars, *The Astrophysical Journal*, 485, 319, doi: [10.1086/304392](https://doi.org/10.1086/304392)
- [251] Sabotta, S., Schlecker, M., Chaturvedi, P., et al. 2021, The CARMENES search for exoplanets around M dwarfs. Planet occurrence rates from a subsample of 71 stars, *Astronomy and Astrophysics*, 653, A114, doi: [10.1051/0004-6361/202140968](https://doi.org/10.1051/0004-6361/202140968)
- [252] Sahlmann, J., Lazorenko, P. F., Ségransan, D., et al. 2013, Astrometric orbit of a low-mass companion to an ultracool dwarf, *Astronomy and Astrophysics*, 556, A133, doi: [10.1051/0004-6361/201321871](https://doi.org/10.1051/0004-6361/201321871)
- [253] Sameshima, H., Matsunaga, N., Kobayashi, N., et al. 2018, Correction of Near-infrared High-resolution Spectra for Telluric Absorption at 0.90-1.35  $\mu\text{m}$ , *Astronomy and Astrophysics*, 613, A130, doi: [10.1088/1538-3873/aac1b4](https://doi.org/10.1088/1538-3873/aac1b4)
- [254] Santerne, A., Moutou, C., Tsantaki, M., et al. 2016, SOPHIE velocimetry of Kepler transit candidates. XVII. The physical properties of giant exoplanets within 400 days of period, *Astronomy and Astrophysics*, 587, A64, doi: [10.1051/0004-6361/201527329](https://doi.org/10.1051/0004-6361/201527329)
- [255] 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 and Astrophysics*, 623, A44, doi: [10.1051/0004-6361/201834114](https://doi.org/10.1051/0004-6361/201834114)

- [256] Schrijver, C. J., De Rosa, M. L., & Title, A. M. 2002, in AGU Fall Meeting Abstracts, Vol. 2002, SH52A–0436
- [257] Schwarz, G. 1978, Estimating the Dimension of a Model, *Annals of Statistics*, 6, 461
- [258] Schwieterman, E., Olson, S. L., Reinhard, C. T., et al. 2018, in AGU Fall Meeting Abstracts, Vol. 2018, P41E–3779
- [259] Seabold, S., & Perktold, J. 2010, in 9th Python in Science Conference
- [260] See, T. J. J. 1896, Researches on the orbit of 70 Ophiuchi, and on a periodic perturbation in the motion of the system arising from the action of an unseen body, *The Astronomical Journal*, 16, 17, doi: [10.1086/102368](https://doi.org/10.1086/102368)
- [261] See, V., Jardine, M., Vidotto, A. A., et al. 2014, The effects of stellar winds on the magnetospheres and potential habitability of exoplanets, *Astronomy and Astrophysics*, 570, A99, doi: [10.1051/0004-6361/201424323](https://doi.org/10.1051/0004-6361/201424323)
- [262] Silva, A. M., Faria, J. P., Santos, N. C., et al. 2022, A novel framework for semi-Bayesian radial velocities through template matching, *Astronomy and Astrophysics*, 663, A143, doi: [10.1051/0004-6361/202142262](https://doi.org/10.1051/0004-6361/202142262)
- [263] Silvotti, R., Schuh, S., Janulis, R., et al. 2007, A giant planet orbiting the ‘extreme horizontal branch’ star V391 Pegasi, *Nature*, 449, 189, doi: [10.1038/nature06143](https://doi.org/10.1038/nature06143)
- [264] Silvotti, R., Charpinet, S., Green, E., et al. 2014, Kepler detection of a new extreme planetary system orbiting the subdwarf-B pulsator KIC 10001893, *Astronomy and Astrophysics*, 570, A130, doi: [10.1051/0004-6361/201424509](https://doi.org/10.1051/0004-6361/201424509)
- [265] 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 and Astrophysics*, 576, A77, doi: [10.1051/0004-6361/201423932](https://doi.org/10.1051/0004-6361/201423932)
- [266] Snellen, I., de Kok, R., Birkby, J. L., et al. 2015, Combining high-dispersion spectroscopy with high contrast imaging: Probing rocky planets around our nearest neighbors, *Astronomy and Astrophysics*, 576, A59, doi: [10.1051/0004-6361/201425018](https://doi.org/10.1051/0004-6361/201425018)
- [267] Snellen, I., Albrecht, S., Anglada-Escude, G., et al. 2019, ESA Voyage 2050 White Paper: Detecting life outside our solar system with a large high-contrast-imaging mission, arXiv e-prints, arXiv:1908.01803, doi: [10.48550/arXiv.1908.01803](https://doi.org/10.48550/arXiv.1908.01803)
- [268] Spina, L., Nordlander, T., Casey, A. R., et al. 2020, How Magnetic Activity Alters What We Learn from Stellar Spectra, *The Astrophysical Journal*, 895, 52, doi: [10.3847/1538-4357/ab8bd7](https://doi.org/10.3847/1538-4357/ab8bd7)
- [269] Stock, S., Nagel, E., Kemmer, J., et al. 2020, The CARMENES search for exoplanets around M dwarfs. Three temperate-to-warm super-Earths, *Astronomy and Astrophysics*, 643, A112, doi: [10.1051/0004-6361/202038820](https://doi.org/10.1051/0004-6361/202038820)
- [270] Strand, K. A. 1942, The orbital motion off the triple system zeta Aquarii, *The Astronomical Journal*, 49, 165, doi: [10.1086/105681](https://doi.org/10.1086/105681)
- [271] —. 1943a, The Triple System  $\mu$  Draconis, , 55, 26, doi: [10.1086/125482](https://doi.org/10.1086/125482)

## BIBLIOGRAPHY

---

- [272] —. 1943b, The Triple System  $\xi$  Boötis, , 55, 28, doi: [10.1086/125483](https://doi.org/10.1086/125483)
- [273] —. 1943c, 61 Cygni as a Triple System, , 55, 29, doi: [10.1086/125484](https://doi.org/10.1086/125484)
- [274] —. 1957, The orbital motion of 61 Cygni., *The Astronomical Journal*, 62, 35, doi: [10.1086/107588](https://doi.org/10.1086/107588)
- [275] Strassmeier, K. G. 2009, Starspots, , 17, 251, doi: [10.1007/s00159-009-0020-6](https://doi.org/10.1007/s00159-009-0020-6)
- [276] Struve, O. 1952, Proposal for a project of high-precision stellar radial velocity work, *The Observatory*, 72, 199
- [277] Suárez Mascareño, A., González-Álvarez, E., Zapatero Osorio, M. R., et al. 2023, Two temperate Earth-mass planets orbiting the nearby star GJ 1002, *Astronomy and Astrophysics*, 670, A5, doi: [10.1051/0004-6361/202244991](https://doi.org/10.1051/0004-6361/202244991)
- [278] Szekely, G. J., & Rizzo, M. L. 2013, Partial Distance Correlation with Methods for Dissimilarities, arXiv e-prints, arXiv:1310.2926, doi: [10.48550/arXiv.1310.2926](https://doi.org/10.48550/arXiv.1310.2926)
- [279] Tamai, R., Cirasuolo, M., González, J. C., Koehler, B., & Tuti, M. 2016, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 9906, *Ground-based and Airborne Telescopes VI*, ed. H. J. Hall, R. Gilmozzi, & H. K. Marshall, 99060W, doi: [10.1117/12.2232690](https://doi.org/10.1117/12.2232690)
- [280] Tarter, J. C., Backus, P. R., Mancinelli, R. L., et al. 2007, A Reappraisal of The Habitability of Planets around M Dwarf Stars, *Astrobiology*, 7, 30, doi: [10.1089/ast.2006.0124](https://doi.org/10.1089/ast.2006.0124)
- [281] Terrien, R. C., Keen, A., Oda, K., et al. 2022, Rotational Modulation of Spectroscopic Zeeman Signatures in Low-mass Stars, *ApJ Letters*, 927, L11, doi: [10.3847/2041-8213/ac4fc8](https://doi.org/10.3847/2041-8213/ac4fc8)
- [282] Thatte, N. A., Clarke, F., Bryson, I., et al. 2016, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 9908, *Ground-based and Airborne Instrumentation for Astronomy VI*, ed. C. J. Evans, L. Simard, & H. Takami, 99081X, doi: [10.1117/12.2230629](https://doi.org/10.1117/12.2230629)
- [283] The Astropy Collaboration. 2018, astropy v3.1: a core python package for astronomy, 3.1, Zenodo, Zenodo, doi: [10.5281/zenodo.4080996](https://doi.org/10.5281/zenodo.4080996)
- [284] Thibault, S., Brousseau, D., Poulin-Girard, A.-S., et al. 2022, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 12184, *Ground-based and Airborne Instrumentation for Astronomy IX*, ed. C. J. Evans, J. J. Bryant, & K. Motohara, 1218454, doi: [10.1117/12.2630640](https://doi.org/10.1117/12.2630640)
- [285] Thompson, M. A., Krissansen-Totton, J., Wogan, N., Telus, M., & Fortney, J. J. 2022, The case and context for atmospheric methane as an exoplanet biosignature, *Proceedings of the National Academy of Science*, 119, e2117933119, doi: [10.1073/pnas.2117933119](https://doi.org/10.1073/pnas.2117933119)
- [286] Trifonov, T., Tal-Or, L., Zechmeister, M., et al. 2020, Public HARPS radial velocity database corrected for systematic errors, *Astronomy and Astrophysics*, 636, A74, doi: [10.1051/0004-6361/201936686](https://doi.org/10.1051/0004-6361/201936686)

- [287] Udalski, A., Zebrun, K., Szymanski, M., et al. 2000, The Optical Gravitational Lensing Experiment. Catalog of Microlensing Events in the Galactic Bulge, , 50, 1, doi: [10.48550/arXiv.astro-ph/0002418](https://doi.org/10.48550/arXiv.astro-ph/0002418)
- [288] 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 and Astrophysics*, 621, A79, doi: [10.1051/0004-6361/201833282](https://doi.org/10.1051/0004-6361/201833282)
- [289] Vacca, W. D., Cushing, M. C., & Rayner, J. T. 2003, A Method of Correcting Near-Infrared Spectra for Telluric Absorption, , 115, 389, doi: [10.1086/346193](https://doi.org/10.1086/346193)
- [290] van de Kamp, P. 1969, Alternate dynamical analysis of Barnard’s star., *The Astronomical Journal*, 74, 757, doi: [10.1086/110852](https://doi.org/10.1086/110852)
- [291] —. 1982, The planetary system of Barnard’s star, *Vistas in Astronomy*, 26, 141, doi: [10.1016/0083-6656\(82\)90004-6](https://doi.org/10.1016/0083-6656(82)90004-6)
- [292] Vida, K., Leitzinger, M., Kriskovics, L., et al. 2019, The quest for stellar coronal mass ejections in late-type stars. I. Investigating Balmer-line asymmetries of single stars in Virtual Observatory data, *Astronomy and Astrophysics*, 623, A49, doi: [10.1051/0004-6361/201834264](https://doi.org/10.1051/0004-6361/201834264)
- [293] Vieira, V. 2012, Permutation tests to estimate significances on Principal Components Analysis, *Computational Ecology and Software*, 2, 103
- [294] Vogt, S. S., Allen, S. L., Bigelow, B. C., et al. 1994, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 2198, *Instrumentation in Astronomy VIII*, ed. D. L. Crawford & E. R. Craine, 362, doi: [10.1117/12.176725](https://doi.org/10.1117/12.176725)
- [295] Wang, J., Fischer, D. A., Horch, E. P., & Huang, X. 2015, On the Occurrence Rate of Hot Jupiters in Different Stellar Environments, *The Astrophysical Journal*, 799, 229, doi: [10.1088/0004-637X/799/2/229](https://doi.org/10.1088/0004-637X/799/2/229)
- [296] Wang, S. X., Latouf, N., Plavchan, P., et al. 2022, Characterizing and Mitigating the Impact of Telluric Absorption in Precise Radial Velocities, *The Astronomical Journal*, 164, 211, doi: [10.3847/1538-3881/ac947a](https://doi.org/10.3847/1538-3881/ac947a)
- [297] Wikimedia. 2007, Diagram illustrating and explaining various terms in relation to Orbits of Celestial bodies. <https://commons.wikimedia.org/wiki/File:Orbit1.svg>
- [298] Winn, J. N. 2009, in *Transiting Planets*, ed. F. Pont, D. Sasselov, & M. J. Holman, Vol. 253, 99–109, doi: [10.1017/S174392130802629X](https://doi.org/10.1017/S174392130802629X)
- [299] Winters, J. G., Henry, T. J., Jao, W. C., et al. 2019, VizieR Online Data Catalog: Stellar multiplicity rate of M dwarfs within 25 pc (Winters+, 2019), *VizieR Online Data Catalog*, J/AJ/157/216
- [300] Wolszczan, A., & Frail, D. A. 1992, A planetary system around the millisecond pulsar PSR1257 + 12, *Nature*, 355, 145, doi: [10.1038/355145a0](https://doi.org/10.1038/355145a0)
- [301] Worth, M. D., & Heintz, W. D. 1974, Parallax, orbit, and mass of the binary star 70 Ophiuchi., *The Astrophysical Journal*, 193, 647, doi: [10.1086/153202](https://doi.org/10.1086/153202)

## BIBLIOGRAPHY

---

- [302] Xuan, W. J., Mawet, D., Ngo, H., et al. 2018, Characterizing the Performance of the NIRC2 Vortex Coronagraph at W. M. Keck Observatory, *The Astronomical Journal*, 156, 156, doi: [10.3847/1538-3881/aadae6](https://doi.org/10.3847/1538-3881/aadae6)
- [303] Yana Galarza, J., Meléndez, J., Lorenzo-Oliveira, D., et al. 2019, The effect of stellar activity on the spectroscopic stellar parameters of the young solar twin HIP 36515, *Monthly Notices of the Royal Astronomical Society*, 490, L86, doi: [10.1093/mnrasl/slz153](https://doi.org/10.1093/mnrasl/slz153)
- [304] Yang, J., Cowan, N. B., & Abbot, D. S. 2013, Stabilizing Cloud Feedback Dramatically Expands the Habitable Zone of Tidally Locked Planets, *ApJ Letters*, 771, L45, doi: [10.1088/2041-8205/771/2/L45](https://doi.org/10.1088/2041-8205/771/2/L45)
- [305] Zechmeister, M., & Kürster, M. 2009, The generalised Lomb-Scargle periodogram, *Astronomy and Astrophysics*, 496, 577, doi: [10.1051/0004-6361:200811296](https://doi.org/10.1051/0004-6361:200811296)
- [306] 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 and Astrophysics*, 609, A12, doi: [10.1051/0004-6361/201731483](https://doi.org/10.1051/0004-6361/201731483)
- [307] Zhang, Z., Liu, M. C., Claytor, Z. R., et al. 2021, The Second Discovery from the COCONUTS Program: A Cold Wide-orbit Exoplanet around a Young Field M Dwarf at 10.9 pc, *ApJ Letters*, 916, L11, doi: [10.3847/2041-8213/ac1123](https://doi.org/10.3847/2041-8213/ac1123)
- [308] Zieba, S., Kreidberg, L., Ducrot, E., et al. 2023, No thick carbon dioxide atmosphere on the rocky exoplanet TRAPPIST-1 c, arXiv e-prints, arXiv:2306.10150, doi: [10.48550/arXiv.2306.10150](https://doi.org/10.48550/arXiv.2306.10150)
- [309] Zucker, S. 2018, Detection of periodicity based on independence tests - III. Phase distance correlation periodogram, *Monthly Notices of the Royal Astronomical Society*, 474, L86, doi: [10.1093/mnrasl/slx198](https://doi.org/10.1093/mnrasl/slx198)
- [310] Zuluaga, J. I., Cuartas, P. A., & Hoyos, J. H. 2012, Evolution of magnetic protection in potentially habitable terrestrial planets, arXiv e-prints, arXiv:1204.0275, doi: [10.48550/arXiv.1204.0275](https://doi.org/10.48550/arXiv.1204.0275)



# Appendix A

## Other detection methods for exoplanets

### A.1 Imaging

Directly imaging exoplanets faces two significant challenges: the star's brightness compared to that of a planet in the optical spectrum, and the fact that, when observed from Earth, a planet is very close to its star. For example, a separation of 1 au from a star located 10 parsecs away corresponds to an angular separation of only 0.1 arcseconds. This is theoretically feasible, but atmospheric turbulence causes the spreading of images, making detection more difficult.

The first challenge is typically addressed using coronagraphy, a technique pioneered by Bernard Lyot, which allows for the blocking of light originating from the host star. The second challenge can be overcome either by using space telescopes such as the JWST<sup>1</sup> (Carter et al., 2023) or by employing adaptive optics for ground-based observations. Adaptive optics involves deforming a mirror to correct for atmospheric wavefront distortions.

Although these mitigation techniques work in the optical spectrum, in practice it is often more advantageous to observe in the infrared. This is because coronagraphs and adaptive optics perform better with longer wavelengths, and the contrast between the star and the planet is more favorable due to the planet's thermal emission. If the planet is young and in its contraction phase, it may emit its own infrared radiation. Otherwise, it can only be detected by reflecting or scattering the light from its host star, but this requires the planet to be relatively close to its star.

The first successful direct imaging of exoplanets occurred in 2008, employing the Very Large Telescope (VLT) in Chile and the NACO<sup>2</sup> instrument. They detected a multiplanetary system orbiting HR 8799 (Marois et al., 2008) which can be seen in figure A.1. Subsequently, this technique has led to the discovery of 69 exoplanets. Among them, COCONUTS-2b (Zhang et al., 2021) gained attention due to its exceptionally large orbital separation of 7,506 au from its host star.

---

<sup>1</sup>James Webb Space Telescope

<sup>2</sup>short for NAOS-CONICA

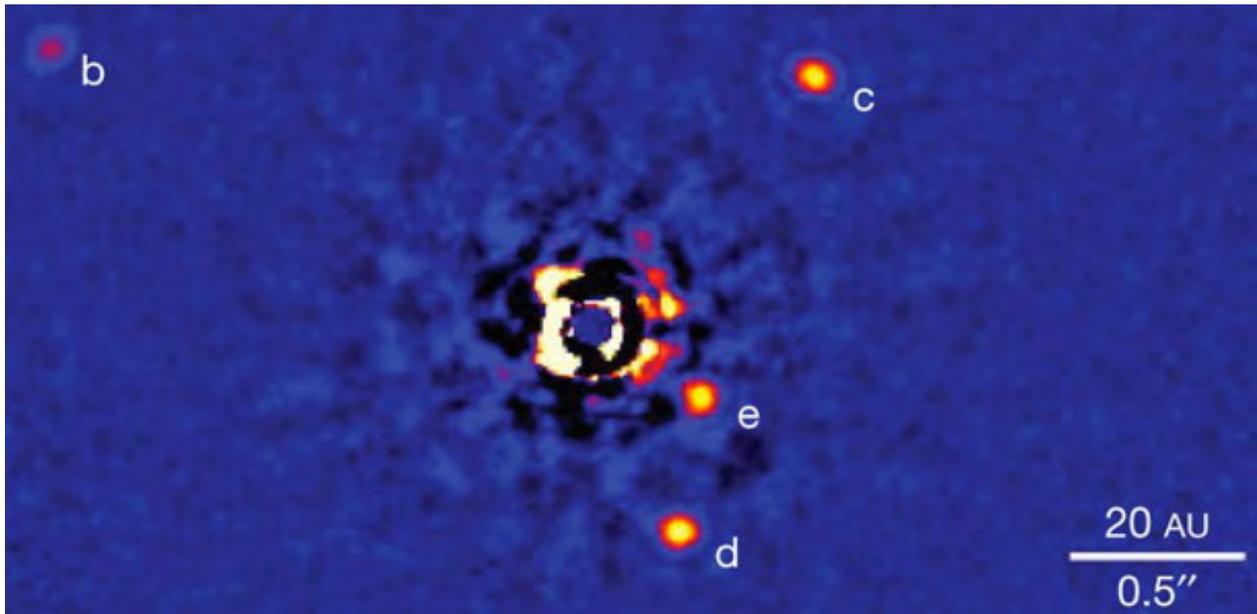


Figure A.1: HR8799 direct imaging planet detections. Credit: [Marois et al. \(2010\)](#)

A dedicated instrument, SPHERE<sup>3</sup> ([Beuzit et al., 2019](#)), was constructed for the VLT. Other similar instruments include NIRC2<sup>4</sup> ([Xuan et al., 2018](#)) and NIRC2<sup>5</sup> ([Hodapp et al., 2003](#)). Additional projects aim to install new instruments operating in the near-infrared (nIR) at the E-ELT, expected to be operational by 2027 ([Tamai et al., 2016](#)). These projects include METIS<sup>6</sup> ([Brandl et al., 2012](#)), HARMONI<sup>7</sup> ([Thatte et al., 2016](#)), and MICADO<sup>8</sup> ([Davies et al., 2016](#)).

## A.2 Astrometry

As mentioned earlier, astrometry involves detecting exoplanets by observing the position of their host stars. It was the initial method attempted for exoplanet discovery ([Jacob, 1855](#)), but it faced challenges and produced false positives, including instances such as [Jacob \(1855\)](#); [See \(1896\)](#); [Berman \(1932\)](#); [Strand \(1957\)](#); [van de Kamp \(1969\)](#). Nevertheless, astrometry remains relevant today because it uniquely offers a direct means to determine the complete 3D layout of planetary systems.

Furthermore, while other techniques such as velocimetry perform better when planets orbit close to their stars, astrometry proves more effective for more distant planets. This is because the star's displacement due to the planet's presence increases as the planet orbits farther away. However, it is essential to note that the effectiveness of astrometry also depends on the star's distance since we measure an angle, making this method more sensitive when the

<sup>3</sup>Spectro Polarimetric High contrast Exoplanet REsearch

<sup>4</sup>Near-InfraRed Camera

<sup>5</sup>Near-InfraRed Imager

<sup>6</sup>Mid-infrared ELT Imager and Spectrograph

<sup>7</sup>High Angular Resolution Monolithic Optical and Near-infrared Integral field spectrograph

<sup>8</sup>Multi-AO Imaging Camera for Deep Observations

star is closer. In any case, there is also a bias in favor of detecting short-period planets since they require fewer extended observations.

As of the present moment, only two exoplanets have been detected using the astrometry method. The first of these is a planet with a mass roughly 30 times that of Jupiter, orbiting an L dwarf star with a period of 246 days (Sahlmann et al., 2013). The second exoplanet orbits GJ 896A and is estimated to have a mass of 2.3 Jupiter masses, with an orbital period of 284.4 days (Curiel et al. (2022)), as shown in Figure A.2). The Gaia mission has made notable progress, and it is anticipated to play a substantial role in the detection of exoplanets with astrometry by identifying numerous signals in the future (Ranalli et al., 2018).

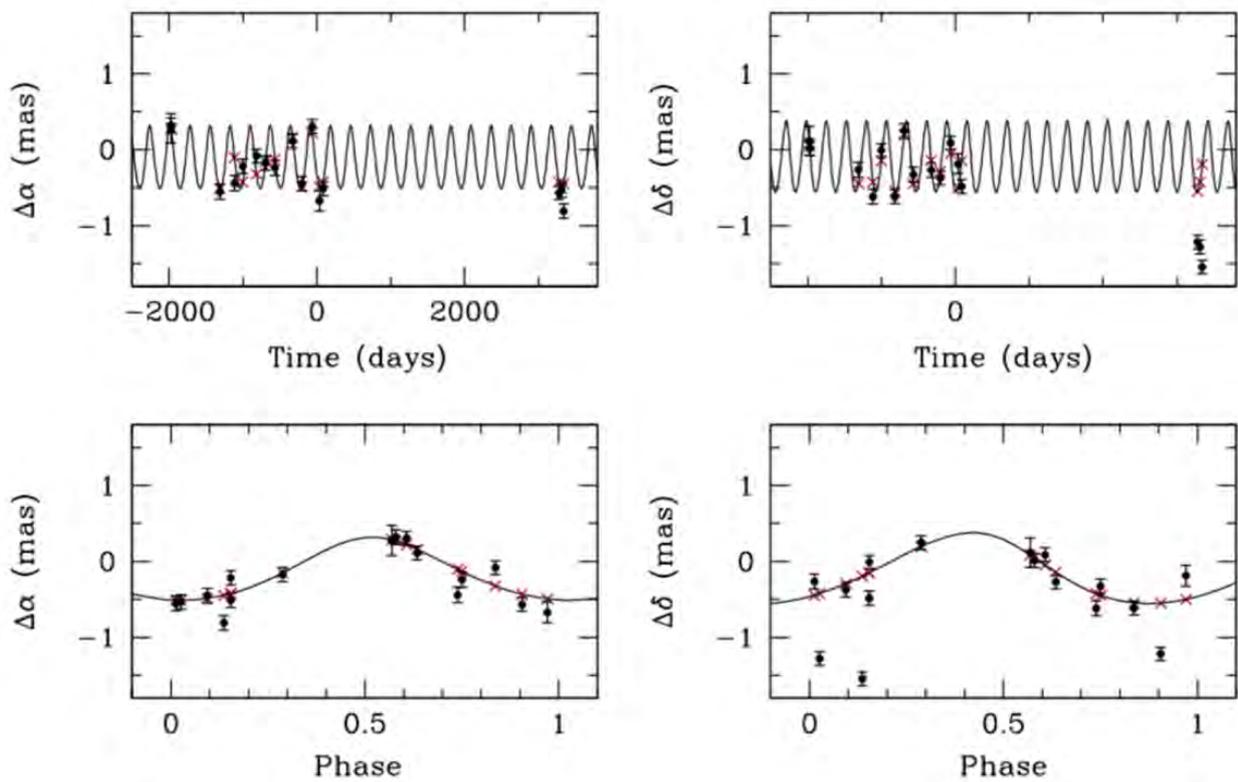


Figure A.2: Best-fitted solution including the astrometric signal due to the planetary companion GJ 896Ab, obtained from the full combined astrometric fit. The upper panels show the astrometric signal of the main star GJ 896A due to the planetary companion as a function of time, with the parallax and proper motion subtracted, and taking into account the orbital motion of the stellar companion GJ 896B around the main star GJ 896A. The lower panels also show the astrometric signal of the main star GJ 896A due to the planetary companion, but in this case the signal is folded in phase. The black dots correspond to the absolute astrometric observations of the main star GJ 896A. The red crosses indicate the fitted position on the orbit for each observed epoch. The bars correspond to the observational astrometric errors. Credit: Curiel et al. (2022)

### A.3 Transit

The transit method is quite straightforward. It involves observing the phenomenon wherein the light emitted by a star appears dimmer when a planet passes in front of it (see figure A.3). A key characteristic of this method is the need for continuous and highly precise monitoring of the star and its flux, denoted as  $F$ . This continuous observation is crucial for measuring the reduction in flux, which is proportional to  $\Delta F = \left(\frac{R_p}{R_s}\right)^2$ , where  $R_s$  represents the radius of the star, and  $R_p$  is the radius of the transiting planet.

The probability of detecting a transit scales as  $\frac{R_s}{a}$ , where  $a$  is the semi-major axis of the star-planet system. Consequently, this method is particularly favorable for detecting planets that orbit in close proximity to their host stars.

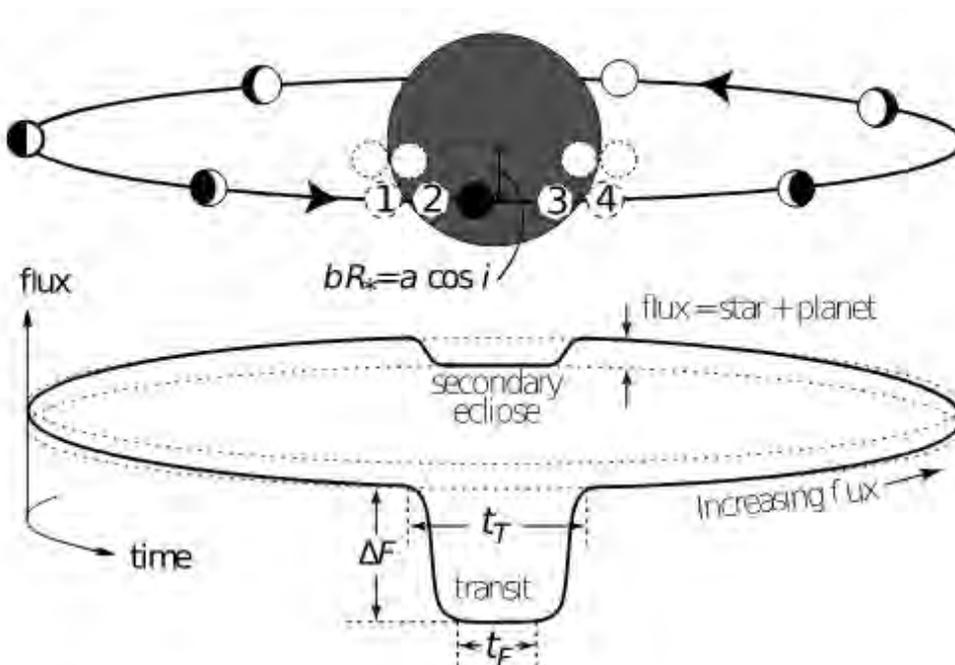


Figure A.3: Illustration of the orbit of a transiting planet (top) and the corresponding light curve (bottom). Credit: (Winn, 2009; Perryman, 2011).

Assuming that the stellar properties are accurately determined, which is usually the case, transit curves provide information about the planet's radius and orbital inclination (depending on the shape of the transit curve). Additionally, the recurrence of the primary transit can give access to the planet's orbital period. However, it is important to note that the planet's mass cannot be directly obtained from the transit curve alone.

As the planet continues along its orbit after the primary transit, it progressively reflects more of the star's light (as depicted in Figure A.3) until it is eventually hidden by the star, a phenomenon known as the secondary eclipse.

Since the first detection of a transiting planet (Charbonneau et al., 2000), research intensified

and ground-based surveys like HAT<sup>9</sup> (Bakos et al., 2004), WASP<sup>10</sup> (Pollacco et al., 2006) or KELT<sup>11</sup> (Pepper et al., 2007) made it possible to detect increasingly smaller planets, from warm Neptunes (Gillon et al., 2007) to super Earths (Charbonneau et al., 2009).

However, photometric observations from Earth are inevitably constrained by atmospheric limitations. To overcome these constraints, space-based missions were undertaken. The first of these missions was CoRoT<sup>12</sup> (Auvergne et al., 2009), succeeded by Kepler (Borucki et al., 2010) and its follow-up mission, K2 (Howell et al., 2014). These programs have been instrumental in detecting thousands of exoplanets, marking a significant achievement in the field. Kepler confirmed 2,780 planets and has 1,984 more awaiting confirmation, while K2 confirmed 548 planets with 977 awaiting confirmation.

Collaborative efforts with high-precision ground-based surveys like SPECULOOS<sup>13</sup> (Gillon et al., 2013) and MEarth (Nutzman & Charbonneau, 2008) resulted in the discovery of the first Earth-sized planet (Pepe et al., 2013). The TESS<sup>14</sup> mission (Ricker & TESS Team, 2015) has also made significant contributions, with 383 confirmed TESS candidates to date and 4,471 awaiting confirmation. The most recent discovery being a super-massive Neptune-sized planet (Naponiello et al., 2023). Looking ahead, future missions such as PLATO 2.0 (Rauer et al., 2014) and ARIEL<sup>15</sup> (Lueftinger et al., 2022) hold promising prospects for the future of transit detections

## A.4 Orbital brightness modulation

This is another technique that relies on the analysis of photometric curves, even in the case of a non-transiting planet. During its orbit, the planet reflects starlight toward the observer, which combines with the light directly received from the star. Consequently, the observer observes a brightness modulation with the planet’s orbital period since it is viewed at various illumination phases. This method has resulted in the confirmed detection of 9 exoplanets (Charpinet et al., 2011; Silvotti et al., 2014; Esteves et al., 2015; Lillo-Box et al., 2021).

## A.5 Timing Variations

In this section, we will delve into a form of indirect detection that capitalizes on the fact that unseen planets can reveal their presence through the gravitational forces they exert on other planets and stars. This influence leads to variations in the timing of predictable events.

### A.5.1 Transit timing variations

The transit method relies on the periodic dimming of a star’s brightness to detect exoplanets. This dimming is expected to follow an extremely regular pattern as the planet orbits its

---

<sup>9</sup>Hungarian Automated Telescope

<sup>10</sup>Wide Angle Search for Planets

<sup>11</sup>Kilodegree Extremely Little Telescope

<sup>12</sup>Convection, Rotation and Planetary Transits

<sup>13</sup>Search for Habitable Planets Eclipsing Ultra-Cool Stars

<sup>14</sup>Transiting Exoplanet Survey Satellite

<sup>15</sup>Atmospheric Remote-sensing Infrared Exoplanet Large-survey

star. However, there may be instances where the transit occurs either earlier or later than predicted, hinting at the presence of another planet in the system. The first detection using this method was conducted by [Ballard et al. \(2011\)](#), where the authors discovered the planet Kepler-19c. To date, 25 exoplanets have been detected using this technique.

### A.5.2 Eclipse timing variations

Eclipsing binary stars consist of pairs of stars that orbit each other, periodically passing in front of one another as observed from Earth. The combined brightness of the system fluctuates when one star blocks the light from the other. Analogous to a planetary transit, a planet orbiting around an eclipsing binary can perturb the timing of the binary eclipses, causing them to occur earlier or later than expected. This method has resulted in 17 confirmed detections thus far (e.g., [Qian et al. \(2010, 2011\)](#); [Potter et al. \(2011\)](#)).

### A.5.3 Pulsar timing variations

We have previously encountered this method when discussing the detection of the first exoplanets around a pulsar ([Wolszczan & Frail, 1992](#)). This approach involves identifying irregularities in the periodicity of radio pulses emitted by the pulsar. It enables the detection of planets that were either captured from elsewhere or formed from the debris left behind by a supernova. However, pulsars are relatively rare, and only seven exoplanets have been discovered using this method.

### A.5.4 Pulsation timing variations

Certain stars exhibit pulsations, characterized by regular fluctuations in brightness as their outermost envelope of hot gas expands and contracts over periods ranging from hours to years. These stars are well-known variable stars. When such stars have planetary companions, these planets can exert gravitational forces on the stars, disrupting the timing of those pulses.

Planets orbiting such stars would pose challenges for detection using alternative methods. For example, their pulsating fluctuations in brightness would hinder transit detections. However, these detections themselves are not straightforward and can only be accomplished when the star's pulsations are extremely regular, which is relatively rare. Consequently, only two confirmed exoplanets have been discovered using this method ([Silvotti et al., 2007](#); [Murphy et al., 2016](#)).

## A.6 Microlensing

Microlensing is an astronomical phenomenon predicted by Einstein's General Theory of Relativity. According to this theory, when light emitted by a star passes in close proximity to another star while traveling toward an observer on Earth, the gravitational pull of the intervening star will slightly deflect the light rays from the source star. This deflection causes the two stars to appear more distant from each other than they would under normal circumstances.

This phenomenon can be used for exoplanet detection. If the star acting as the gravitational lens has an exoplanet, the system behaves like a double lens, causing a significant increase in the light curve of the source star. The first planet detected using this method occurred in 2004, and it was a planet with a mass 1.5 times that of Jupiter, situated 3 au away from its host star OGLE 2003-BLG-235. Notably, this star is positioned at a distance of 5,000 parsecs from Earth (Bond et al., 2004).

In fact, one of the advantages of this method is its capability to detect exoplanets at extremely large distances from Earth. The most distant exoplanet detected using this method is OGLE-2016-BLG-1190Lb, situated a staggering 22,000 light-years away (Ryu et al., 2018). Additionally, microlensing is highly sensitive to planets occupying moderate to large orbits around their stars. This sensitivity complements the radial velocity and transit detection methods, both of which are most effective when identifying planets in close proximity to their stars. Moreover, microlensing events can unveil the presence of rogue planets, which are planets that do not orbit stars.

As of now, 201 objects have been detected through microlensing, with the majority of these detections coming from the OGLE<sup>16</sup> program (Udalski et al., 2000). The recently launched Euclid space telescope is expected to further contribute to exoplanet detection via microlensing (Beaulieu et al., 2010). Additionally, the upcoming Nancy Grace Roman Space Telescope (formerly WFIRST<sup>17</sup>) has a planned launch in 2027 and is poised to significantly enhance our ability to detect exoplanets using microlensing (Penny et al., 2019; Johnson et al., 2020).

In addition to the detection of free-floating planets, the types of exoplanets predicted to be discovered through this method should open up the possibility of identifying complementary exoplanet types that are not detectable through the transit or radial velocity methods, as illustrated in figure A.4.

However, the primary drawback of this method is that microlensing events are unique occurrences, and any planets detected through this method will not be observed again. This limitation is exacerbated by the rarity and randomness of such events, making the method challenging and unpredictable. These are the fundamental constraints of the technique.

Another limitation is that the distance of the detected planet can only be determined through rough approximations. If these planets are located thousands of light-years away, this could result in errors of thousands of light-years. Nevertheless, joining the Roman-Euclid survey has the potential to overcome this microlens mass-distance-velocity degeneracy (Bachelet et al., 2022).

---

<sup>16</sup>Optical Gravitational Lensing Experiment

<sup>17</sup>Wide-Field Infrared Survey Telescope

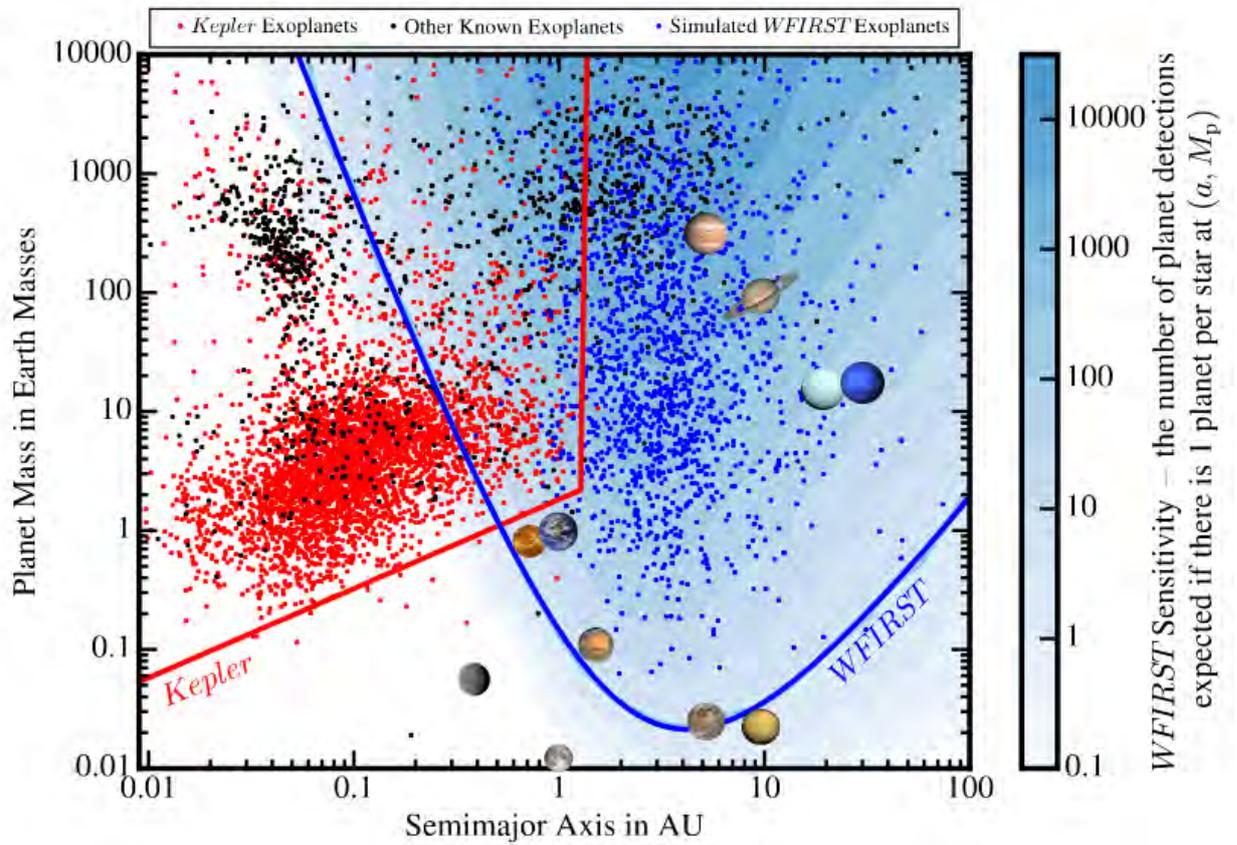


Figure A.4: Simulations of exoplanets that the Roman Space Telescope should be capable of detecting. Credit: Penny et al. (2019)

# Appendix B

## BERV-Invariant periodogram

### B.1 Overview

The objective of this appendix is to introduce a novel type of periodogram designed to disregard signals linked to the Barycentric Earth Radial Velocity (BERV). The rationale behind this approach is rooted in our observations from chapters 2 and 3, where we noted that telluric lines tend to generate systematic effects that are closely connected to the BERV.

The `Wapiti` method relies on radial velocities (RVs) computed using the line-by-line (LBL) framework (Dumusque, 2018; Artigau et al., 2022). While it is increasingly applied to various instruments beyond SPIRou, such as NIRPS (Thibault et al., 2022) or HARPS (Mayor et al., 2003), other instruments may still employ alternative methods like the cross-correlation function (CCF) method (Mayor & Queloz, 1995; Baranne et al., 1996; Pepe et al., 2002; Lafarga et al., 2020) or template matching techniques such as `harps-terra` (Anglada-Escudé & Butler, 2012), `naira` (Astudillo-Defru et al., 2017b), `wobble` (Bedell et al., 2019a), or `serval` (Zechmeister et al., 2018).

For these reasons, our objective was to develop a periodogram that remains invariant to BERV variations so as to distinguish between signals of telluric origins and those from other sources.

This periodogram was influenced by the research of Binnenfeld et al. (2022), which in turn was inspired by the notion of semi-partial distance correlation. In their study, this novel approach enabled the construction of periodograms capable of effectively discriminating between Doppler shifts and spectral shape variations in astronomical spectra.

### B.2 Partial and semi-partial correlation

Partial correlation (Pearson, 1915), quantifies the extent of association between two variables while accounting for and eliminating the influence of other variables. Semi-partial correlation, also known as part correlation, serves a similar purpose, but it only mitigates the impact of nuisance parameters on one of the two variables.

As a straightforward example, let us examine three time series denoted as  $(x_i, y_i, z_i)$ , where  $i$

represents the measurement epoch. To standardize these series, we subtract their respective means, resulting in:

$$\mathbf{x} = \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} z_1 - \bar{z} \\ z_2 - \bar{z} \\ \vdots \\ z_n - \bar{z} \end{bmatrix}. \quad (\text{B.1})$$

We presume that the observed values are sampled from a joint distribution involving three random variables, namely,  $X$ ,  $Y$ , and  $Z$ . These variables cannot be assumed to be independent. The covariance between these random variables can be represented using the standard Euclidean inner product, which is given by:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_i (x_i - \bar{x})(y_i - \bar{y}) = n \text{Cov}(X, Y). \quad (\text{B.2})$$

To mitigate the influence of the variable  $Z$  on the relationship between  $X$  and  $Y$ , one approach is to independently model two linear relationships: one between  $X$  and  $Z$ , and the other between  $Y$  and  $Z$ . The residuals from these regression models, denoted as  $\mathbf{e}_x$  and  $\mathbf{e}_y$ , are assumed to be less affected by  $Z$ . Once these residuals are obtained, the correlation coefficient between  $\mathbf{e}_x$  and  $\mathbf{e}_y$  can be computed, providing a measure of correlation while accounting for the impact of the nuisance parameter  $Z$ . This approach is commonly referred to as "partial correlation" (Pearson, 1915). The residual vectors with respect to  $Z$  are calculated as follows:

$$\mathbf{e}_x = \mathbf{x} - \frac{\langle \mathbf{x}, \mathbf{z} \rangle}{\langle \mathbf{z}, \mathbf{z} \rangle} \mathbf{z}, \quad \mathbf{e}_y = \mathbf{y} - \frac{\langle \mathbf{y}, \mathbf{z} \rangle}{\langle \mathbf{z}, \mathbf{z} \rangle} \mathbf{z}. \quad (\text{B.3})$$

The partial correlation is defined as the correlation coefficient of the residuals:

$$R_{xy} = \frac{\langle \mathbf{e}_x, \mathbf{e}_y \rangle}{\sqrt{\langle \mathbf{e}_x, \mathbf{e}_x \rangle \langle \mathbf{e}_y, \mathbf{e}_y \rangle}}. \quad (\text{B.4})$$

In certain situations, there might be a specific interest in managing the influence of  $Z$  on just one of the two variables, for instance,  $X$ . This process is known as "semi-partial correlation," and it is computed by:

$$R_{xy} = \frac{\langle \mathbf{e}_x, \mathbf{y} \rangle}{\sqrt{\langle \mathbf{e}_x, \mathbf{e}_x \rangle \langle \mathbf{y}, \mathbf{y} \rangle}}. \quad (\text{B.5})$$

### B.3 Semi-partial distance correlation

The semi-partial distance correlation is an extension of the semi-partial correlation. In the subsequent subsection, we introduce the Hilbert space of  $\mathcal{U}$ -centered matrices, which serves as the foundation for defining the semi-partial distance correlation (Szekely & Rizzo, 2013).

Now, let us consider three variables: X, Y, and Z, each with a size of  $n$ . The distance matrix of X, computed using the metric function  $d_X$ , can be expressed as follows:

$$x_{ij} = d_X(x_i, x_j), \quad (\text{B.6})$$

where  $i$  and  $j$  represent indices corresponding to measurements.  $\mathcal{U}$ -centering (Szekely & Rizzo, 2013) is then applied to this distance matrix:

$$\bar{x}_{ij} = x_{ij} - \bar{x}_{ij}, \quad (\text{B.7})$$

where

$$\bar{x}_{ij} = \begin{cases} \frac{1}{n-2} \left[ \sum_k (x_{ik} + x_{kj}) - \frac{1}{n-1} \sum_{kl} x_{kl} \right] & \text{if } i \neq j \\ x_{ij} & \text{otherwise.} \end{cases} \quad (\text{B.8})$$

The inner product of these  $\mathcal{U}$ -centered matrices is defined as:

$$\langle \bar{\mathbf{x}}, \bar{\mathbf{y}} \rangle_{\mathcal{U}} = \frac{1}{n(n-3)} \sum_{i \neq j} \bar{x}_{ij} \bar{y}_{ij}. \quad (\text{B.9})$$

As demonstrated in Szekely & Rizzo (2013), this expression serves as an unbiased estimator of the distance covariance. Now, focusing on Z as the nuisance parameter, we can calculate the residual matrix with respect to Z as follows:

$$\bar{\mathbf{e}}_w = \bar{\mathbf{w}} - \frac{\langle \bar{\mathbf{w}}, \bar{\mathbf{z}} \rangle_{\mathcal{U}}}{\langle \bar{\mathbf{z}}, \bar{\mathbf{z}} \rangle_{\mathcal{U}}} \bar{\mathbf{z}} \quad (\text{B.10})$$

Here, the variable  $w$  represents either the  $\mathcal{U}$ -centered matrices X or the Y variable. Ultimately, we can formulate the semi-partial distance correlation between variables X and Y, while taking into account the influence of the nuisance variable Z on X, as follows:

$$D_{xy} = \frac{\langle \bar{\mathbf{e}}_x, \bar{\mathbf{y}} \rangle}{\sqrt{\langle \bar{\mathbf{e}}_x, \bar{\mathbf{e}}_x \rangle \langle \bar{\mathbf{y}}, \bar{\mathbf{y}} \rangle}} \quad (\text{B.11})$$

## B.4 BERV-I periodogram

Now, we can straightforwardly define the BERV-Invariant (BERV-I) periodogram. To begin, we select a period  $P$ . For each measurement epoch  $t_i$ , we assign a phase  $\phi_i = t_i \bmod P$ . As detailed in Zucker (2018), we calculate a distance matrix  $\bar{\phi}$  using the phase metric  $d_\phi$ :

$$d_\phi = \phi_{ij} \cdot (P - \phi_{ij}) \quad (\text{B.12})$$

where  $\phi_{ij} = t_i - t_j \bmod P$ . Next, we compute the distance matrix  $\bar{v}$  for each epoch of the RV using the RV metric (Binnenfeld et al., 2020, 2022):

$$d_v = |v_i - v_j|. \quad (\text{B.13})$$

Here, the phase and RV construct the first two random variables in the problem. The third variable is the BERV, for which we construct the distance matrix for the BERV using the BERV metric:

$$d_{BERV} = |BERV_i - BERV_j|. \quad (\text{B.14})$$

The BERV-I periodogram is constructed by considering the BERV as the nuisance parameter in Equation B.11 and its effect on the phase.

To assess the effectiveness of the BERV-I periodogram, we employed the simulations conducted for a star with a narrow BERV coverage and telluric contamination levels of 3% from chapter 3, with an injected signal at an orbital period  $P = 20$  d and semi-amplitude  $K = 2, 5$  or  $10 \text{ m.s}^{-1}$ .

As done in [Binnenfeld et al. \(2020, 2022\)](#), the peak significance in these periodograms is assessed using the bootstrap method. Specifically,  $N_{bootstrap} = 1,000,000$  random time series are created using white noise, and the peak value of the maximum peak is recalculated for these time series. This information is used to obtain a threshold value that matches a desired level of False Alarm Probability (FAP).

The results are presented in Figure B.1. It can be observed that in the BERV-I periodograms, the intended penalty for telluric harmonics is observed, while the injected 20-day signal starts to become detectable, even at a  $5 \text{ m.s}^{-1}$  semi-amplitude level while it was not discernible at a statistically significant level in the GLS periodogram. This demonstrates that the BERV-I periodogram successfully differentiates the nature of the signals in the data, allowing us to detect the injected planet signal without having to correct for the systematics.

This encouraging result led us to test the BERV-I periodogram in the case of real data with Gl 725 B (trend corrected). However, its application to this target in figure B.2 shows that it did not lead to the detection of the two planets that were previously identified using the *Wapiti* method. Additionally, the peaks associated with telluric systematics were not entirely corrected, indicating the need for further improvement. One potential avenue for improving this tool could involve incorporating more nuisance parameters than just the BERV. Determining the most suitable additional parameters will be a subject of future research.

## B.4. BERV-I PERIDOGRAM

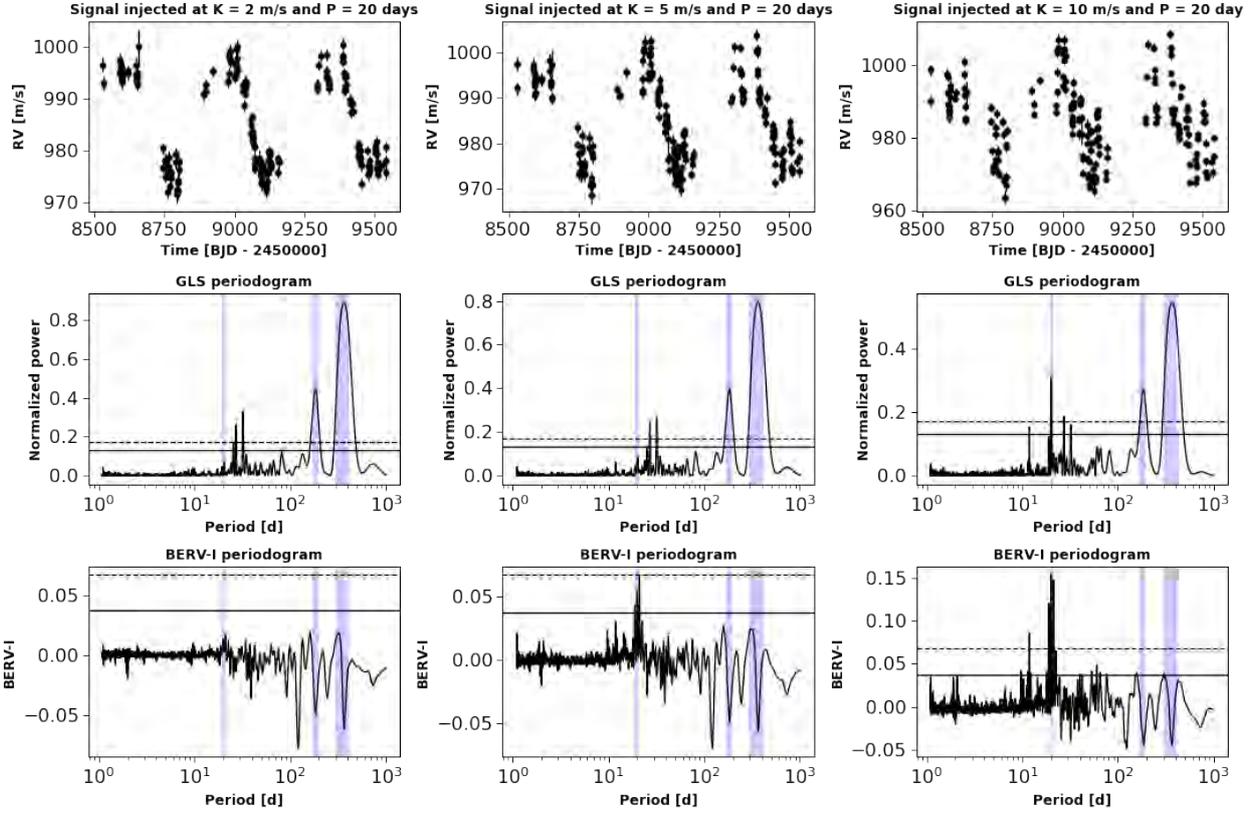


Figure B.1: The BERV-I periodogram applied on simulations with a narrow BERV coverage and a planetary signal of 2, 5 and 10  $\text{m.s}^{-1}$  with a 20 d orbital period. Top panels show the simulated RV time series, middle panels the associated GLS periodograms and bottom panels the associated BERV-I periodograms.

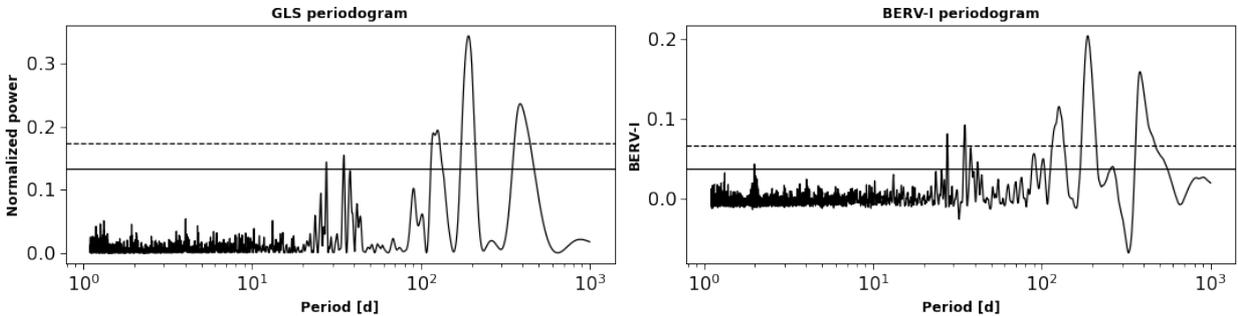


Figure B.2: BERV-I periodogram applied on G1725 B RV timeseries. Left panel shows the GLS periodogram and right panel the BERV-I periodogram.



# Appendix C

## Publications

### C.1 List of publications

- **Ould-Elhkim, M.**, “Confirmation of an exoplanet around Gl 480 and detection of a candidate exoplanet orbiting Gl 382”, submitted to A&A.
- Donati, J.-F., “Magnetic fields and rotation periods of M dwarfs from SPIRou spectra”, *Monthly Notices of the Royal Astronomical Society*, vol. 525, no. 2, pp. 2015–2039, 2023. doi:10.1093/mnras/stad2301. (**I am author 7**).
- Moutou, C., “Characterizing planetary systems with SPIRou: M-dwarf planet-search survey and the multiplanet systems GJ 876 and GJ 1148”, arXiv e-prints, 2023. doi:10.48550/arXiv.2307.11569. (**I am author 8**).
- **Ould-Elhkim, M.**, “Wapiti: A data-driven approach to correct for systematics in RV data. Application to SPIRou data of the planet-hosting M dwarf GJ 251”, *Astronomy and Astrophysics*, vol. 675, 2023. doi:10.1051/0004-6361/202346472.
- Carmona, A., “Near-IR and optical radial velocities of the active M dwarf star Gl 388 (AD Leo) with SPIRou at CFHT and SOPHIE at OHP. A 2.23 day rotation period and no evidence for a corotating planet”, *Astronomy and Astrophysics*, vol. 674, 2023. doi:10.1051/0004-6361/202245660. (**I am author 5**).
- Cadieux, C., “TOI-1452 b: SPIRou and TESS Reveal a Super-Earth in a Temperate Orbit Transiting an M4 Dwarf”, *The Astronomical Journal*, vol. 164, no. 3, 2022. doi:10.3847/1538-3881/ac7cea. (**I am author 38**).
- Martioli, E., “TOI-1759 b: A transiting sub-Neptune around a low mass star characterized with SPIRou and TESS”, *Astronomy and Astrophysics*, vol. 660, 2022. doi:10.1051/0004-6361/202142540. (**I am author 20**).

### C.2 First author publications

#### C.2.1 Wapiti: A data-driven approach to correct for systematics in RV data

# Wapiti: A data-driven approach to correct for systematics in RV data

## Application to SPIRou data of the planet-hosting M dwarf GJ 251

M. Ould-Elhkim<sup>1,\*</sup>, C. Moutou<sup>1</sup>, J-F. Donati<sup>1</sup>, É. Artigau<sup>2,3</sup>, P. Fouqué<sup>1</sup>, N.J. Cook<sup>2</sup>, A. Carmona<sup>4</sup>, P. I. Cristofari<sup>1,5</sup>, E. Martioli<sup>6,7</sup>, F. Debras<sup>1</sup>, X. Dumusque<sup>8</sup>, J. H. C. Martins<sup>10</sup>, G. Hébrard<sup>6</sup>, C. Cadieux<sup>2</sup>, X. Delfosse<sup>4</sup>, R. Doyon<sup>2</sup>, B. Klein<sup>9</sup>, J. Gomes da Silva<sup>10</sup>, T. Forveille<sup>4</sup>, T. Hood<sup>1</sup>, and P. Charpentier<sup>1</sup>

<sup>1</sup> Université de Toulouse, UPS-OMP, IRAP, 14 avenue E. Belin, Toulouse, F-31400, France, France

e-mail: merwan.ould-elhkim@irap.omp.eu

<sup>2</sup> Trottier Institute for Research on Exoplanets, Université de Montréal, Département de Physique, C.P. 6128 Succ. Centre-ville, Montréal, QC H3C 3J7, Canada

<sup>3</sup> Observatoire du Mont-Mégantic, Université de Montréal, Département de Physique, C.P. 6128 Succ. Centre-ville, Montréal, QC H3C 3J7, Canada

<sup>4</sup> Univ. Grenoble Alpes, CNRS, IPAG, 38000 Grenoble, France

<sup>5</sup> Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA

<sup>6</sup> Institut d'Astrophysique de Paris, CNRS, UMR 7095, Sorbonne Université, 98 bis bd Arago, 75014 Paris, France

<sup>7</sup> Laboratório Nacional de Astrofísica, Rua Estados Unidos 154, 37504-364, Itajubá - MG, Brazil

<sup>8</sup> Département d'astronomie, Université de Genève, Chemin des Maillettes 51, CH-1290 Versoix, Switzerland

<sup>9</sup> Department of Physics, University of Oxford, Oxford OX1 3RH, UK

<sup>10</sup> Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, 4150-762 Porto, Portugal

July 11, 2023

### ABSTRACT

**Context.** Recent advances in the development of precise radial velocity (RV) instruments in the near-infrared (near-IR) domain, such as SPIRou, have facilitated the study of M-type stars to more effectively characterize planetary systems. However, the near-IR presents unique challenges in exoplanet detection due to various sources of planet-independent signals which can result in systematic errors in the RV data.

**Aims.** In order to address the challenges posed by the detection of exoplanetary systems around M-type stars using near-IR observations, we introduced a new data-driven approach for correcting systematic errors in RV data. The effectiveness of this method is demonstrated through its application to the star GJ 251.

**Methods.** Our proposed method, Weighted principal component reconstruction (referred to as Wapiti), used a dataset of per-line RV time series generated by the line-by-line (LBL) algorithm and employed a weighted Principal Component Analysis (wPCA) to reconstruct the original RV time series. A multistep process was employed to determine the appropriate number of components, with the ultimate goal of subtracting the wPCA reconstruction of the per-line RV time series from the original data in order to correct systematic errors.

**Results.** The application of Wapiti to GJ 251 successfully eliminated spurious signals from the RV time series and enabled the first detection in the near-IR of GJ 251b, a known temperate super-Earth with an orbital period of 14.2 days. This demonstrates that, even when systematics in SPIRou data are unidentified, it is still possible to effectively address them and fully realize the instrument's capability for exoplanet detection. Additionally, in contrast to the use of optical RVs, this detection did not require us to filter stellar activity, highlighting a key advantage of near-IR RV measurements.

**Key words.** Techniques: radial velocities; Techniques: spectroscopic; Methods: data analysis; Stars: individual: GJ 251; Stars: planetary systems; Stars: low-mass

## 1. Introduction

Over the past few years, the development of precise radial velocity (RV) instruments in the near-infrared (near-IR) domain, such as HPF (Mahadevan et al. 2012), IRD (Kotani et al. 2014), CARMENES-NIR Becerril et al. (2017), NIRPS (Bouchy & Doyon 2018) and SPIRou (Donati et al. 2020), has facilitated

the study of exoplanets. SPIRou is an acronym for SPectropolarimétrie InfraROUGE (infrared spectropolarimeter).

\* Based on observations obtained at the Canada-France-Hawaii Telescope (CFHT) which is operated from the summit of Maunakea by the National Research Council of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique of France, and the University of Hawaii. The observations at the Canada-France-Hawaii Telescope were performed with care and respect from the summit of Maunakea which is a significant cultural and historic site. Based on observations obtained with SPIRou, an international project led by Institut de Recherche en Astrophysique et Plané-

the detection and characterization of new exoplanetary systems (Cadieux et al. 2022; Martioli et al. 2022). These instruments have enabled the expansion of research on M-type stars to the *YJHK* spectral bandpasses (Quirrenbach et al. 2018; Reiners & Zechmeister 2020). Those stars are of particular interest because they constitute the majority of main sequence stars in the solar neighborhood (Henry et al. 2006). Moreover, thanks to their low masses, M dwarfs are particularly well suited for the detection of planet RV signatures of a few  $\text{m s}^{-1}$  (Anglada-Escudé et al. 2016; Bonfils et al. 2013; Faria et al. 2022; Mascareño et al. 2023) and they are known to harbor more rocky planets than stars of other types (Bonfils, X. et al. 2013; Dressing & Charbonneau 2015; Gaidos et al. 2016; Hsu et al. 2020; Sabotta et al. 2021).

However, M dwarfs pose challenges for exoplanet detection due to their high levels of magnetic activity, which produce spurious RV signals (Bonfils et al. 2007; da Silva et al. 2012; Newton et al. 2016; Hébrard et al. 2016). Additionally, similar to other instruments operating in the optical range, some systematics may originate from the instrument itself, such as an average intra-night drift and a nightly zero-point variation that can be corrected (Trifonov et al. 2020; Tal-Or et al. 2019), or instrumental defects (Dumusque et al. 2015). Furthermore, the near-IR range required to analyze the red spectral energy distribution of M dwarfs is prone to strong telluric absorption lines which can create an additional source of uncertainty in the measurement of RV (Artigau et al. 2014). The near-IR poses other challenges, including persistence in infrared arrays. Most notably, this phenomenon is characterized by the fact that not only is the preceding frame of significance, but the entire illumination history spanning the last few hours, is as well. Specifically, bright targets observed at the outset of a night can impact the observation of fainter targets later on in the same night (Artigau et al. 2018).

Despite efforts to correct these systematics, it is possible that spurious signals may still remain after applying existing techniques, potentially due to incorrect assumptions, technical limitations, or unidentified sources. To address this issue, we present a new data-driven method for correcting systematics in RV data. We focus our work on data collected with the SPIRou instrument, a high-precision RV spectrograph in the near-IR domain with two science topics that are at the forefront: the quest for Earth-like planets in the habitable zones of very-low-mass stars, and the study of low-mass stars and planet formation in the presence of magnetic fields (Donati et al. 2020).

To achieve these science goals, SPIRou has a large program (LP) called the SPIRou Legacy Survey (SLS)<sup>1</sup> which was allocated 310 nights over seven semesters (2019a to 2022a) on the Canada-France-Hawaii Telescope. The SLS-Planet Search program, which is a part of this LP, aims to detect and characterize planetary systems around  $\sim 50$  nearby M dwarfs (Moutou et al. in prep).

The SPIRou data are prone to the previously mentioned systematics and it is necessary to develop methods to correct them. Our method, Weighted-principal component analysis reconstruction, referred to as *Wapiti*<sup>2</sup>, is agnostic to the origin of the systematics and reconstructs the per-line RV time series. This method takes advantage of the new line-by-line (LBL) algorithm (Artigau et al. 2022) used to compute RV data with SPIRou, which provides a RV time series for each line of the spectrum. The application of principal component analysis (PCA) to mitigate systematics in RV data is a well-established technique in the

exoplanet community. However, existing methods are primarily aimed at specifically reducing the impact of stellar activity, such as the methods presented in Collier Cameron et al. (2021) and Cretignier et al. (2022). PCA based methods were also explored to correct for telluric or instrumental effects in prior studies (Artigau et al. 2014; Cretignier et al. 2021). Our method, on the other hand, is unique in that it does not focus solely on reducing the impact of stellar activity or telluric effects but instead is a versatile and flexible approach that can be applied to any target, as long as LBL data are available. We demonstrate the effectiveness of *Wapiti* by applying it to the star GJ 251, a target with a known planet in the SLS-Planet Search program.

This paper is structured as follows. Section 2 provides an overview of the observations, data reduction, and LBL algorithm used for the study. Section 3 presents the *Wapiti* method in detail, including how it determines the optimal number of principal vectors. Section 4 demonstrates the efficiency of the *Wapiti* method by eliminating systematic peaks and confirming the detection of a planet with a 14.2-day period. Finally, section 5 summarizes the main results and discusses potential future work.

## 2. Observations

### 2.1. GJ251

The star GJ 251 (HD 265866) is a red dwarf with a spectral type of M3V (Kirkpatrick et al. 1991) located  $5.5846 \pm 0.0009$  pc away (Gaia Collaboration 2020) from Earth. From the analysis of SPIRou spectra, Cristofari et al. (2022) derived the stellar parameters listed in Table 1 that we used in our study.

Parameters	Value
$T_{\text{eff}}$ (K)	$3420 \pm 31$
$[M/H]$	$-0.01 \pm 0.10$
$[\alpha/\text{Fe}]$	$-0.01 \pm 0.04$
$\log g$	$4.71 \pm 0.06$
Radius ( $R_{\odot}$ )	$0.365 \pm 0.007$
Mass ( $M_{\odot}$ )	$0.35 \pm 0.02$
$\log (L/L_{\odot})$	$-1.786 \pm 0.003$

**Table 1.** Retrieved stellar parameters of GJ 251 (Cristofari et al. 2022)

In Butler et al. (2017), the detection of two planets with orbital periods of 1.74 and 601.32 days was claimed to be made using the HIRES spectrograph at the Keck observatory. However, in Stock et al. (2020), no such evidence was found in the CARMENES visible data. Instead, a planet with an orbital period of  $14.238 \pm 0.002$  days was identified in the data, with a semi-amplitude  $K = 2.11 \pm 0.20 \text{ m s}^{-1}$ , which falls within the category of temperate super-Earths, with a calculated mass of  $M_p \sin i = 4.0 \pm 0.4 M_{\oplus}$  and an equilibrium temperature of  $T_{\text{eq}} = 351 \pm 1.4 \text{ K}$  (Stock et al. 2020). The small amplitude of this planet makes GJ 251 an ideal candidate to demonstrate that SPIRou is also a suitable instrument for detecting Earth-like planets in the habitable zone of M dwarfs

The observations were collected as series of circular polarization sequences of 4 individual sub exposures per night. The data collection for GJ,251 spans from December 12, 2018 to April 14, 2022, comprising a total of 769 subexpositions, with each visit comprising four subexposures. Out of the available data, three visits are incomplete, leaving us with a total of 191

<sup>1</sup> <https://spirou-legacy.irap.omp.eu/doku.php>

<sup>2</sup> <https://github.com/HkmMerwan/wapiti>

complete visits. The airmass during these observations varied from a minimum of 1.028 to a maximum of 2.003, with a median value of 1.083. The signal-to-noise ratio per  $2.28\text{-km s}^{-1}$  pixel bin in the middle of the *H* band, as measured with 61-second exposures, has a median value of 137.0.

## 2.2. The SPIRou spectropolarimeter

SPIRou is a high-resolution near-infrared spectropolarimeter installed in 2018 on the Canada-France-Hawaii Telescope (CFHT). It has a spectral range from  $0.98$  to  $2.35\ \mu\text{m}$  at a spectral resolving power of about 70 000 (Donati et al. 2020). The instrument operates in vacuum at 73K, regulated to sub-mK level, with a temperature-controlled environment (Challita et al. 2018). It has precise light injection devices (Parès et al. 2012) and splits light into two orthogonal polarization states using ZnSe quarter-wave rhombs and a Wollaston prism located in its Cassegrain unit. It uses two science fibers (A and B) and one calibration fiber (C) (Micheau et al. 2018). The detector is a 15-micron science-grade HAWAII 4RG<sup>TM</sup> (H4RG) from Teledyne Systems used in its up-the-ramp readout mode (Artigau et al. 2018).

## 2.3. SPIRou data reduction with APERO

The spectra of GJ 251 were reduced with the 0.7.275 version of the SPIRou data reduction system (DRS) called APERO (A PipelinE to Reduce Observations) which is detailed in Cook et al. (2022). The APERO pipeline processes science frames to produce 2D and 1D spectra from the two science channels and the calibration channel. The wavelength calibration uses a combination of exposures from a UNe hollow cathode lamp and a Fabry-Pérot etalon (Hobson et al. 2021). Barycentric Earth radial velocity (BERV) is calculated to shift the wavelengths to the barycentric frame of the solar system using *barycorrpy* (Kandia & Wright 2018). Finally, the pipeline incorporates a correction for telluric absorptions and night-sky emission, which will be detailed in an upcoming publication (Artigau in prep.).

## 2.4. LBL algorithm

Radial velocities of GJ 251 were obtained from the telluric-corrected spectra using the line-by-line (LBL) method described in the paper by Artigau et al. (2022). This method, which is based on the framework developed in Bouchy et al. (2001), has been proposed by Dumusque (2018) and further explored by Cretignier et al. (2020) in the context of optical RV observations with HARPS. The LBL calculates Doppler shifts for individual spectral lines. To achieve this, a template with low noise is required, as the RVs are inferred from the comparison between the residuals of the observed spectrum and the template, as well as the derivative of the template. In practice, a high signal-to-noise ratio (S/N) combined spectrum is used as a template for a given target, to ensure that any remaining noise is small compared to an individual spectrum. The LBL method has been compared to other methods such as the cross-correlation function and template matching in Martioli et al. (2022) and Artigau et al. (2022).

For each spectrum, the LBL algorithm combines the velocities of individual lines (27702 lines in the case of GJ 251) into a single RV measurement. This is achieved using a mixture model, which assumes that the velocities of individual lines are distributed according to a Gaussian distribution centered on the mean velocity, or another distribution of high-sigma outliers with multiple origins (e.g., bad pixels, cosmic rays, telluric resid-

uals). The RV measurements are grouped by night by calculating the inverse-variance weighted mean of all observations for each night, resulting in a reduction from 769 to 177 data points. To address the variation in velocities of the day-to-day drift, a correction is applied using simultaneous Fabry-Pérot measurements. Furthermore, to take into account any long-term variation, a Gaussian Process (GP) regression was computed using the most observed stars in the SLS in order to estimate the zero point in the time domain. This procedure is similar to the one conducted for SPIRou data in Cadieux et al. (2022) or for SOPHIE (Courcol et al. 2015), HARPS (Trifonov et al. 2020) and HIRES (Tal-Or et al. 2019) spectrographs. In our case, we employed the *celerite* package (Foreman-Mackey et al. 2017) to fit a GP model using a Matérn 3/2 kernel. The estimation of the hyper-parameters was achieved by minimizing the likelihood function. Studies on common trends in SPIRou data are being conducted in parallel, and will address the differences between GPs and more commonly used tools such as a running median or nightly offsets in forthcoming work by the SPIRou team.

The final outcome of the data reduction is presented in Figure 1, wherein the red data points represent the results obtained after applying zero point correction and the black data points represent the results without such correction. To identify periodicities in the time series, a generalized Lomb-Scargle (GLS) periodogram was used (Zechmeister & Kürster 2009), with a grid of 100,000 points ranging from 1.1 days to 1,000 days, using the *astropy* package (The Astropy Collaboration 2018). The false alarm probability (FAP) of the maximum peak was calculated using the analytical formula in Baluev (2008).

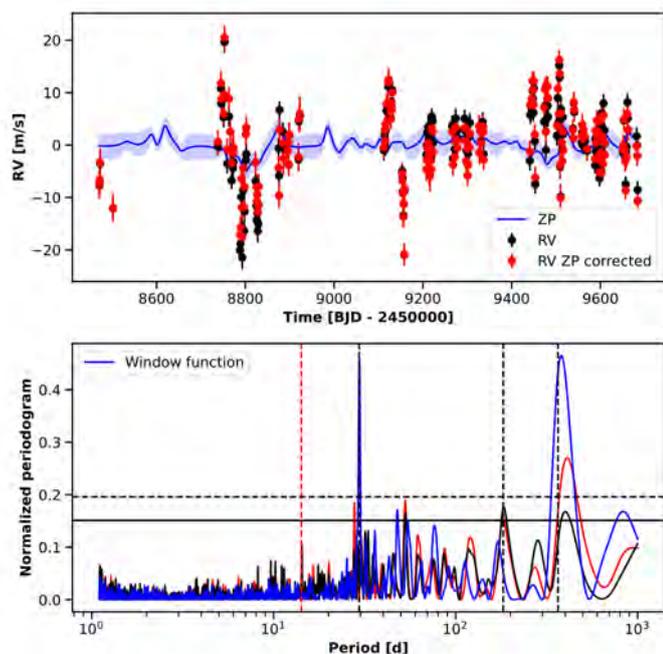
As is illustrated, even with the careful processing steps in APERO, the LBL algorithm produced time series with spurious peaks at a period of 6 months and 1 year. However, the planet as reported in Stock et al. (2020) was not detected. Notably, the zero-point correction did not alleviate these systematic effects but rather amplified them. As this 180-day peak is not present in the window function, it can be inferred that it did not arise from the observation strategy but from an unclear systematic effect. These signals may be attributed to a defective telluric correction; however, the underlying cause is poorly understood and a practical, physics-based correction is currently unavailable. Therefore, GJ 251 serves as an appropriate example of how our method can eliminate spurious signals from RV data.

## 3. The Wapiti method

Our approach applies a PCA to the per-line RV time series produced by the LBL algorithm. Since these datasets may be noisy and contain missing data, we use the weighted principal component analysis (wPCA) method described in Delchambre (2015), implemented in the *wPCA* module<sup>3</sup> and each observation is weighted by the inverse of its uncertainty squared. The objective of the Wapiti method is to eliminate the sources of systematic error. This is achieved by performing a wPCA reconstruction of the original per-line RV time series and subtracting its reconstructions from it. The rationale behind this approach is that, in a time series dominated by spurious signals of unknown origin, selecting appropriate components can serve as a proxy for correcting the effects responsible for the resulting systematic errors.

To determine the appropriate number of principal vectors, we suggest the following approach. First, we discarded all individual line RV time series that have more than 50% of missing

<sup>3</sup> <https://github.com/jakevdp/wPCA>



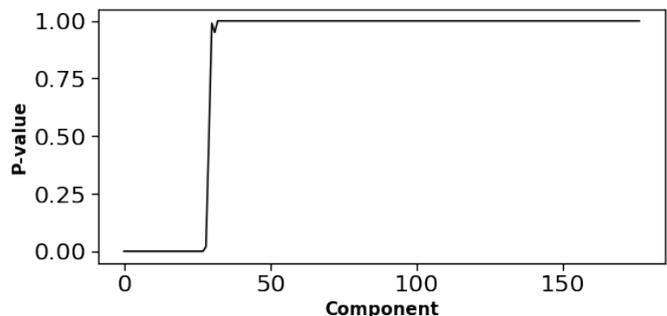
**Fig. 1.** Comparison of LBL RVs data for GJ 251 with and without zero point Correction. The top panel shows the data with and without zero point correction, indicated by black and red points respectively. The zero point offset is shown in blue. The bottom panel shows the corresponding periodograms for the time series in the same color scheme, and the blue line represents the window function. The significance levels of  $FAP = 10^{-3}$  and  $FAP = 10^{-5}$  are indicated by the solid and dashed horizontal lines, respectively. The 1-year period, the 180 days period, and the synodic (29.5 days) period of the Moon are represented by vertical black dashed lines, and the period of the claimed planet in Stock et al. (2020) is in red.

data, reducing the number of lines from 27 201 to 17 923. Such missing data may occur for a line when the RV value deviates by more than five sigma from the standard deviation in relation to the line’s RMS (Artigau et al. 2022). Subsequently, we binned each individual per-line RV time series by night, by computing the inverse-variance weighted mean of all observations for each night. Afterward, we night-binned in a similar fashion the Fabry-Pérot drift also computed with the LBL and removed it from each individual per-lines RV time series. Finally, we standardized the time series using their inverse-variance weighted mean and variance.

### 3.1. Reducing the number of components through a permutation test

The selection of the appropriate number of components for a wPCA is a critical aspect. Too few components may fail to effectively remove undesired contamination, while an excessive number of components may lead to over-fitting and an increase in the levels of random noise. To address this issue, a method is needed to determine which wPCA components are relevant and necessary for the analysis. In our work, we first used a permutation test (Vieira 2012) which involves independently shuffling the columns of the dataset, and then conducting a wPCA on the resulting dataset. We repeated this process 100 times and compared the explained variance of the original dataset with the permuted versions. If a principal component is relevant, we expect that the permuted explained variances of that component will not

be greater than the original explained variance. We can identify the relevant components by examining the p-values and selecting those with low p-values ( $\leq 10^{-5}$ ). The result was that only the first 26 components can not be explained by noise as it can be seen in Figure 2.



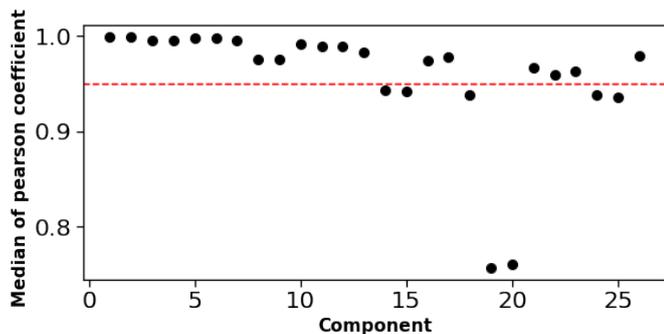
**Fig. 2.** p-values of each component as determined by the permutation test. Only the components with a p-value below a certain threshold ( $\leq 10^{-5}$ ) are considered to have a significant contribution, rather than being just random noise.

### 3.2. Leave-p-out cross-validation

To evaluate the significance of the number of components retained after the permutation test, we conducted a leave-p-out cross-validation. This approach has been used by Cretignier et al. (2022) to select relevant components, albeit with variations in the methodology. In this study, we generated 100 subsets of the per-lines RV dataset, each containing 80% of the lines randomly selected. The wPCA was then recalculated for each of these N datasets to assess the robustness of the 26 remaining components.

The underlying idea involves evaluating the robustness of the eigenvector basis, also known as principal vectors, by examining their stability under the removal of a few lines. Should a principal vector exhibit excessive sensitivity to the lines in use, it is deemed superfluous and may be discarded. We employed the Pearson coefficient of correlation as the evaluation metric, for the reasons stated in Cretignier et al. (2022) being that it is computationally efficient, bounded, and insensitive to the amplitudes of vectors. Specifically, we computed the Pearson coefficient between the  $n^{\text{th}}$  principal vector obtained from the wPCA using the full dataset and the  $n^{\text{th}}$  principal vector obtained from each of the N subsets. The absolute value of the coefficient was taken, as the wPCA vectors may change sign. The median Pearson coefficient value was calculated for each component number, and only those with a value above an arbitrary threshold of 95% were retained, as in Cretignier et al. (2022).

As depicted in Figure 3, it was observed that the value first goes below the threshold after the 13<sup>th</sup> component. This indicates that the principal vectors after the 3<sup>th</sup> one are not correlated with their corresponding vectors in the original dataset, implying that these eigenvectors are reliant on the lines used. As a result, they were disregarded, leaving us with 13 components for the remainder of the *Wapiti* method. In future research, efforts will be conducted to study how the number of components obtained through the application of the two methods, namely the permutation test and the leave-p-out cross-validation, varies with the target being studied.



**Fig. 3.** Mean Pearson coefficients as a function of the components. Those with a value below 95% are discarded, leaving the number of components to 13.

### 3.3. Removing outliers

Per-line RVs have the potential to be affected by short-term fluctuations that are not correlated with the most significant principal vectors. These outliers can adversely affect the ability to identify systematics and reduce the accuracy of the reconstruction. Collier Cameron et al. (2021) applied a singular value decomposition to the shape of the CCF and used a one-dimensional rejection mask in the time domain to remove outliers. For each principal component, they classified observations as "bad" if their absolute deviations are farther from the median value than a specified number  $k_{mad}$  of the median absolute deviations (MAD), and "good" otherwise. If even one epoch in a principal component is an extreme outlier, it is likely that the entire observation is contaminated. In this study the authors determined that a  $k_{mad}$  value of 6 yielded satisfactory results at the cost of rejecting 33 data points out of 886.

In our work, we considered the per-line RV time series and applied a similar method to the principal vectors. However, selecting an appropriate value for  $k_{mad}$  is not a straightforward task. Although one could potentially choose a value such as  $k_{mad} = 15$  to correspond to a  $10\sigma$  rejection, this would seem arbitrary. Furthermore, such an approach would not align with the aim of our method, which is to be able to operate independently on stars without requiring an intervention to select the appropriate value of  $k_{mad}$  for each target. For this reason, we have chosen to employ a different strategy. We computed the ratio of the absolute deviation ( $AD$ ) of the  $n^{\text{th}}$  principal vector  $V_n(t)$  to its  $MAD$ , and assigned the maximum value of this ratio to each epoch across all components (we recall that for a time-series  $X$  and naming its median  $\bar{X}$  we have  $AD(X) = |X - \bar{X}|$  and  $MAD = \text{median}(AD(X))$ ). We use the letter  $D$  to denote this value, which measures the degree of anomaly in an observation:

$$D(t) = \max_n \frac{AD(V_n(t))}{MAD(V_n(t))}. \quad (1)$$

Consequently, we reorganized the epochs in a descending order based on the value of  $D$ , resulting in a list that ranks the degree of anomaly of each observation. We chose to discard the first  $N$  elements from this list. To identify the optimal value of  $N$ , we gradually removed them one by one and determined the number that maximizes the significance of the signal in the corrected RV once the Wapiti method had been applied. We found that removing the first seven outliers produced the best outcomes, reducing the time series from 178 to 171 data points. It is acknowledged that readers may perceive this method to be

specific to individual stars. However, we contend that this is a justifiable approach for our selected target GJ 251, as the signal remains consistent regardless of the number of rejected outliers. Moreover, the main benefit of this approach is that it provides a framework to assess the level of anomaly for any target at a given epoch. For further information about this process, please refer to Appendix A.

The top-left panel of Figure 4 displays the RV time series with flagged outliers highlighted in red. The periodogram of the RV time series after the removal of outliers is displayed on the right. Furthermore, the first two principal vectors are presented on the left, with the previous principal vectors using all available data points depicted in red, and the new principal vectors obtained from applying the wPCA on the remaining data points shown in black. The periodograms of the new principal vectors are displayed on the right. From the periodograms, it can be observed that the presence of a 1-year period and a 180-day period is evident in the two vectors, indicating that the wPCA method identified signals within the per-line RV dataset that may be contributing to the resulting systematic errors. We furthermore observed that the principal vectors did not exhibit any signs of the known 14.2 d planet, indicating that our approach did not inadvertently capture the Keplerian signal in our data. It can also be noted that a prominent signal with a period of 120 days appears in the first vector, while not present in the RV time series. This period could correspond to a harmonic of the 1 year signal, however we also note that it is close to the reported star's rotation period of  $P_{rot} = 124 \pm 5$  days in Stock et al. (2020). We attributed this signal to stellar activity for reasons that will be more detailed in subsequent sections.

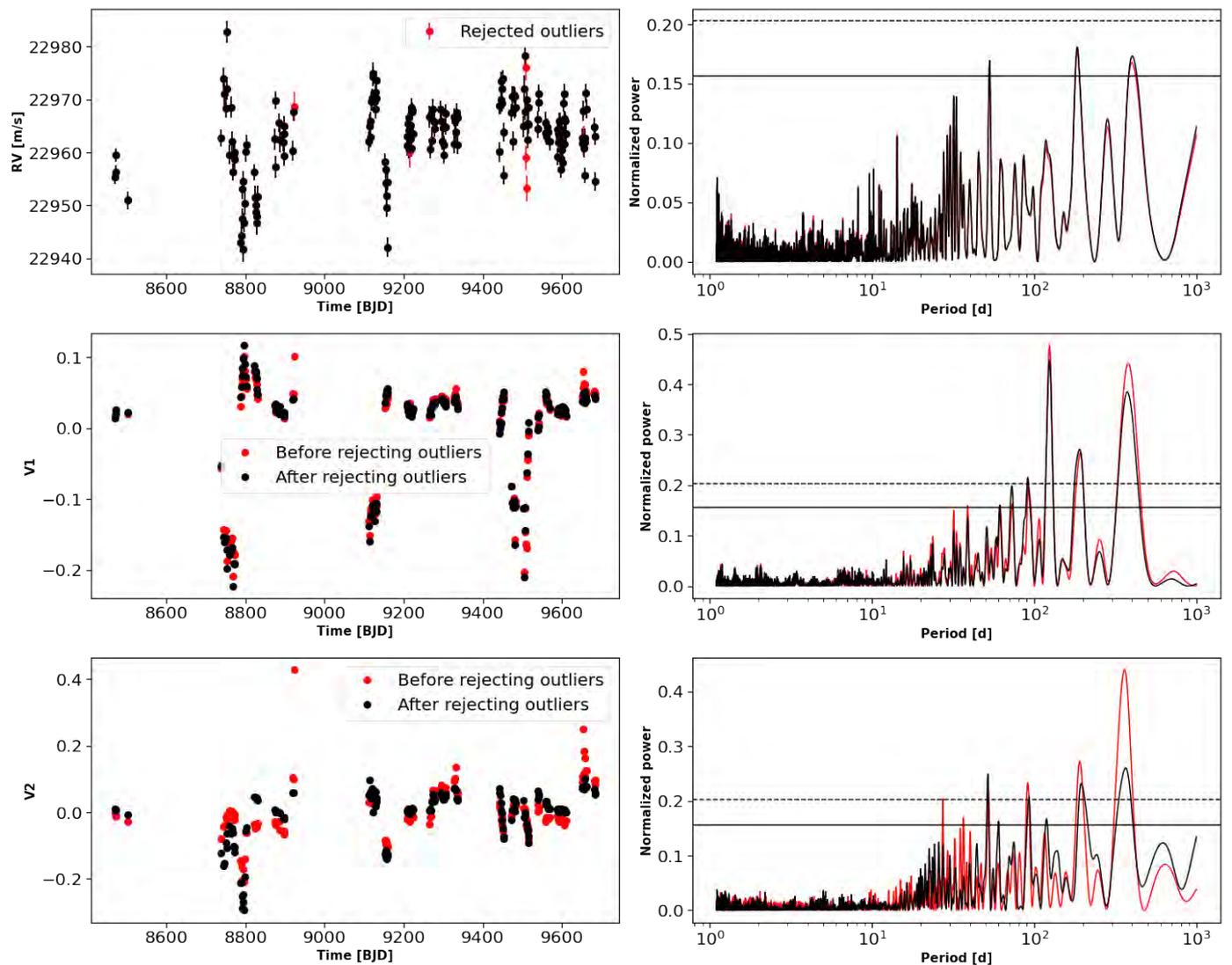
## 4. Results

### 4.1. The Wapiti corrected time series and its analysis

The method outlined in the previous section has allowed us to determine the most appropriate number of components for our dataset on GJ 251. We subsequently subtracted the reconstruction of the original time series to itself to generate a new time series, which we refer to as the Wapiti corrected time series. We evaluated the effectiveness of the correction by analyzing the resulting periodogram as illustrated in Figure 5. In this Figure, the time series resulting from the correction is depicted in blue and the original data are represented in black. The same color scheme is used when plotting the periodograms below the time series.

As desired, the Wapiti correction effectively eliminated the 180 d and 365 d peaks that we attributed to systematics present in our data since those periods are no longer statistically significant and the planet previously reported by Stock et al. (2020) is now discernible in our data, making it the first detection of GJ 251b with near-IR velocities. More precisely, the most prominent peak went from a 184d period peak with a FAP level of  $9.15 \times 10^{-5}$  to a 14.2d peak with a FAP level of  $4.01 \times 10^{-9}$  after correction. In Appendix B, we explore the impact of the number of components used for the wPCA on this detection.

To analyze the corrected RV time series obtained from the Wapiti algorithm, a model  $\mu$  consisting of a Keplerian orbit with a period of 14.2 days and an offset was employed, using the radvel package (Fulton et al. 2018) on the corrected data. However, caution must be exercised when fitting this model, as the reconstruction process may degrade it. The reason being that although there were no signs of the Keplerian signal found in the principal vectors, it is important to note that they may not be or-



**Fig. 4.** Radial velocity time series and principal vectors. Left panels show the RV time series and the first two principal vectors. Their respective periodograms are on the right. The significance levels of  $FAP = 10^{-3}$  and  $FAP = 10^{-5}$  are indicated by the solid and dashed horizontal lines, respectively.

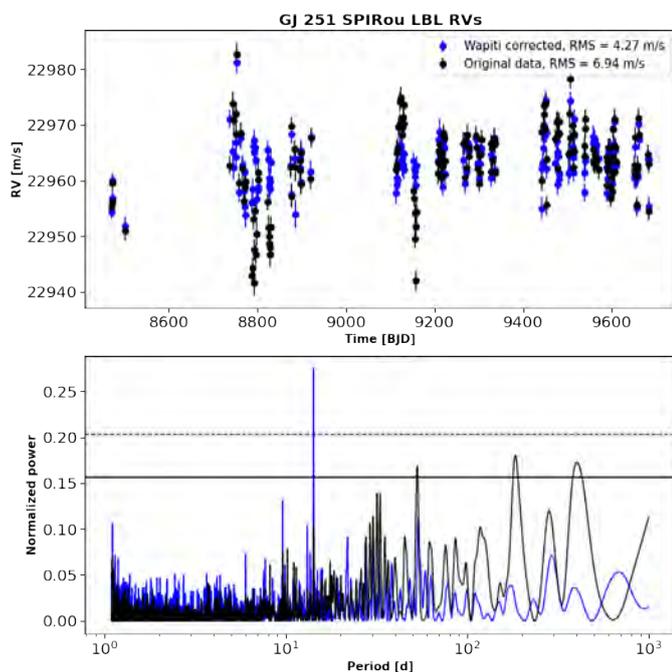
thogonal to it. Therefore, it is possible that when subtracting the wPCA projection from the time series, the projection of the Keplerian signal could still have a nonzero value, posing a potential issue when trying to correctly characterize the signal. Therefore, it could be necessary to take into account the filtering effect of the wPCA reconstruction when fitting the model. This can be achieved with minimal computational cost by using the wPCA to reconstruct the desired model and then fitting the difference between the original model  $\mu$  and its wPCA reconstruction  $\hat{\mu}$  to the corrected RV time series obtained from Wapiti. Concretely, denoting  $R$  the wPCA reconstruction we have  $\hat{\mu} = R(\mu)$  and we suggest to fit  $\mu - \hat{\mu}$  to the corrected data instead of simply  $\mu$ . This process bears resemblance to that described in Gibson et al. (2020), wherein they used this method to take into account the filtering impact of the SysRem algorithm (Tamuz et al. 2005) on their model. The periodogram of the residuals in 6 does not show any signs of one of the planets reported in Butler et al. (2017).

To obtain estimates of the uncertainties on the free parameters, we employed a Bayesian Markov Chain Monte Carlo (MCMC) framework to sample the posterior distribution, with 100 000 samples, 10 000 burn-in samples, and 100 walkers us-

ing the emcee module (Foreman-Mackey et al. 2013). The orbital and derived planetary parameters are shown in Table 2 and compared to those obtained in Stock et al. (2020). In order to calculate the mass of the planet,  $M_p \sin i$ , we used the mass of the star and its uncertainty obtained from Cristofari et al. (2022) and listed in Table 1. One may note that the eccentricity is poorly constrained from our RV data, prompting the question of whether a circular fit would yield a better result. To address this, we conducted a comparison of the maximum likelihood estimation (MLE) for both Keplerian and circular models, allowing us to compute their respective Bayesian Information Criterion (BIC; see Schwarz (1978)). By taking the logarithm of the Bayes factor ( $\log BF = \frac{BIC_{circ.} - BIC_{kep.}}{2}$ ), we were able to compare the two models. Our analysis strongly favored the Keplerian model ( $\log BF = 15$ ), and as such, we decided not to perform a MCMC estimation of the orbital parameters using a circular model. The phase-folding of the best-fitted model to the data is displayed in Figure 6 which also present the periodogram of the residuals with no significant signals within it.

	Wapiti	Priors	Stock et al. (2020)
<b>Orbital Parameters</b>			
Period (days)	$14.24^{+0.01}_{-0.01}$	$\mathcal{U}(14, 15)$	$14.238^{+0.002}_{-0.002}$
$t_0 - 2450000$ (BJD)	$8628^{+3}_{-3}$	$\mathcal{U}(8622.01, 8632.01)$	$8626.69^{+0.34}_{-0.35}$
$K$ (m s <sup>-1</sup> )	$3.25^{+0.41}_{-0.40}$	$\mathcal{U}(0, 4)$	$2.11^{+0.21}_{-0.20}$
$\sqrt{e} \cos \omega$	$-0.12^{+0.32}_{-0.23}$	$\mathcal{U}(-1, 1)$	$0.06^{+0.21}_{-0.22}$
$\sqrt{e} \sin \omega$	$0.10^{+0.22}_{-0.22}$	$\mathcal{U}(-1, 1)$	$0.20^{+0.16}_{-0.22}$
<b>Derived Parameters</b>			
$M_p \sin i$ (M <sub>⊕</sub> )	$6.02^{+0.80}_{-0.78}$	-	$4.00^{+0.40}_{-0.40}$
$\omega$ (degrees)	$60^{+86}_{-209}$	-	$78.8^{+47.6}_{-44.7}$
$e$	$0.11^{+0.11}_{-0.07}$	-	$0.10^{+0.09}_{-0.07}$

**Table 2.** Orbital and derived planetary parameters from the analysis of GJ 251. The parameters derived from the Wapiti corrected RV data from SPIRou are compared to those from Stock et al. (2020) using CARMENES and HIRES spectrographs.



**Fig. 5.** The effectiveness of the Wapiti correction method on the RV time series. The top panel of the Figure shows the original data in black and the time series from the Wapiti correction in blue. The bottom panel shows the corresponding periodograms using the same color code, it can be seen that the correction has successfully eliminated the 180d and 365d systematic peaks. The reported 14.2d period planet is now detected at high significance in the blue periodogram with a FAP level of  $4.01 \times 10^{-9}$ .

#### 4.2. Injection recovery test

In order to assess the effectiveness of the Wapiti method in retrieving correct planetary parameters and to verify our claim that it is necessary to take into account the wPCA reconstruction of our model, we performed an injection recovery test. We first removed the Wapiti corrected RV time series from each per-line RV time series. We then injected a signal consisting of a Keplerian with the parameters corresponding to the best-fit param-

eters of GJ 251b from our analysis, except for the semi-amplitude  $K_{injected}$  that we made vary from 2 to 10 m s<sup>-1</sup>.

We chose a range of semi-amplitude values that explore the low and intermediate regime where the planetary signal is either of lower or of comparable magnitude with the systematics present in our data. In our case, the systematics consist of a 180 d signal of  $\sim 6$  m s<sup>-1</sup> amplitude. This range of semi-amplitude allowed us to assess the ability of the Wapiti method to recover the injected planetary signal.

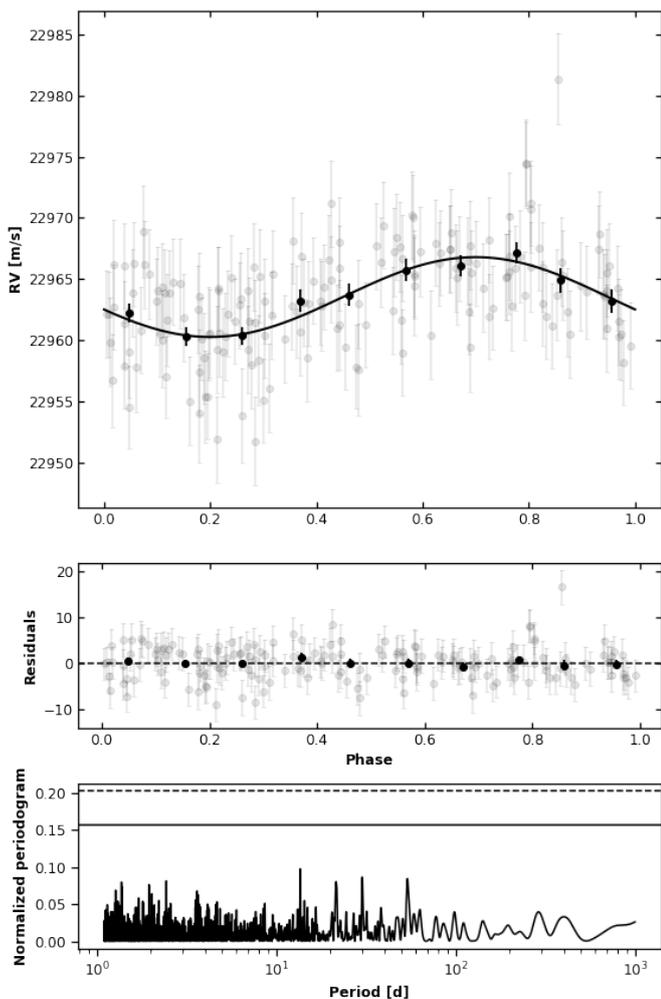
For each injected signal, we employed a simple model comprising of only the injected Keplerian with varying semi-amplitude and an offset. Our aim was to quantitatively assess the effect of the Wapiti method on the estimated semi-amplitude  $K_{computed}$ . To accomplish this, we used a MCMC approach, which consisted of 100,000 steps, 10,000 burn-in samples, and 100 walkers. We set a uniform prior distribution between -50 and 50 m s<sup>-1</sup> over the semi-amplitude.

In order to evaluate the importance of subtracting the wPCA reconstruction  $\hat{\mu}$  from the model  $\mu$ , we conducted a comparative analysis by fitting both  $\mu$  and  $\mu - \hat{\mu}$  to the data and examining their impact on the estimated semi-amplitude. Our findings displayed in figure 7 demonstrate that the injection recovery test is generally successful and the Wapiti method accurately recovers the correct semi-amplitude. However, as the semi-amplitude increases, the exclusion of the reconstruction  $\hat{\mu}$  becomes increasingly significant. Fitting only  $\mu$  results in an underestimation of the correct semi-amplitude.

#### 4.3. Stellar activity in GJ 251

It is noteworthy that contrary to the method employed by Stock et al. (2020), we did not use a quasi-periodic GP to account for stellar activity in our analysis. In their paper, they reported a rotation period of  $124 \pm 5$  days while Fouqué et al. (2023) detected a periodicity of  $98.7^{+11.5}_{-4.8}$  days by analyzing the time series of the longitudinal magnetic field ( $B_\ell$ ), yet we did not observe such signals in our RV periodogram. Nonetheless, for completeness, we fit two models consisting of one planet and one planet plus a quasi-periodic GP of the form

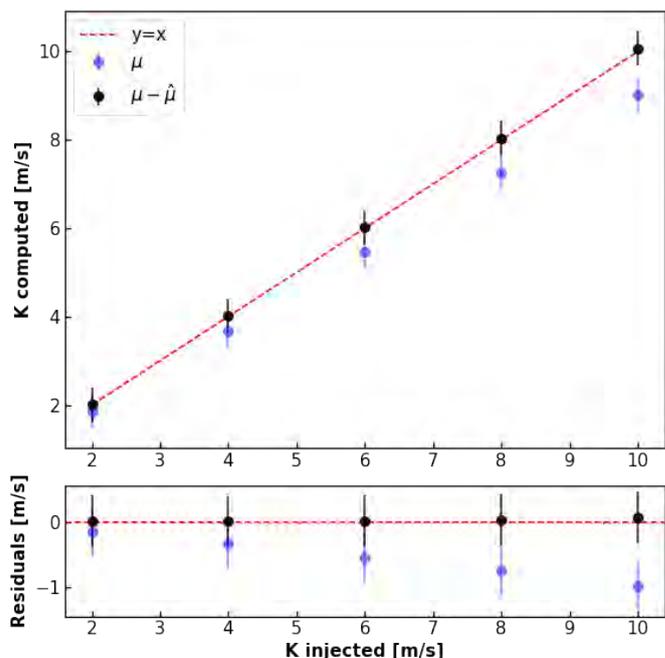
$$k(\tau) = A \exp\left(-\frac{\tau^2}{2l^2} - \Gamma \sin^2\left(\frac{\pi}{P}\tau\right)\right). \quad (2)$$



**Fig. 6.** Results of the RV analysis using the *Wapiti* corrected time series. The top panel illustrates the phase-folding of the planet with the best-fit 14.2-day period, represented by a solid black curve, and the black circles depict the binned RV measurements. The middle panel displays the residuals after the model has been fit, and the bottom panel presents the resulting periodogram, which indicates the absence of any remaining significant signals within the data.

The GP was computed using *george* (Ambikasaran et al. 2015). We calculated the Bayes factor to compare the two models and it was found that adding the quasi-periodic GP did not improve the fit ( $\log BF = -5.5$ ), reinforcing our conclusion that stellar activity does not impact our RV data. Ultimately, we determined that it was not necessary to account for the influence of stellar activity when characterizing the planet orbiting GJ 251.

However, when performing a wPCA on the differential line width (dLW; introduced in Zechmeister et al. (2018)), which is another data product obtained from the LBL analysis that measures variations in the average width of line profiles relative to a template (Artigau et al. 2022), Cadieux et al. (in prep.) found that the first principal vectors can be sensitive to the presence of stellar activity. In our case, we observed that even though no periodicity can be detected in dLW, the first principal vector ( $W_1$ ) presents a signal that is similar to the one reported in Stock et al. (2020) as displayed in Figure 8. We fit a quasi-periodic GP in order to estimate the value of the periodicity in  $W_1$  fitting a quasi-periodic kernel of the same form as in Equation 2 as well as a jitter term and a constant. We computed an MCMC using 10,000

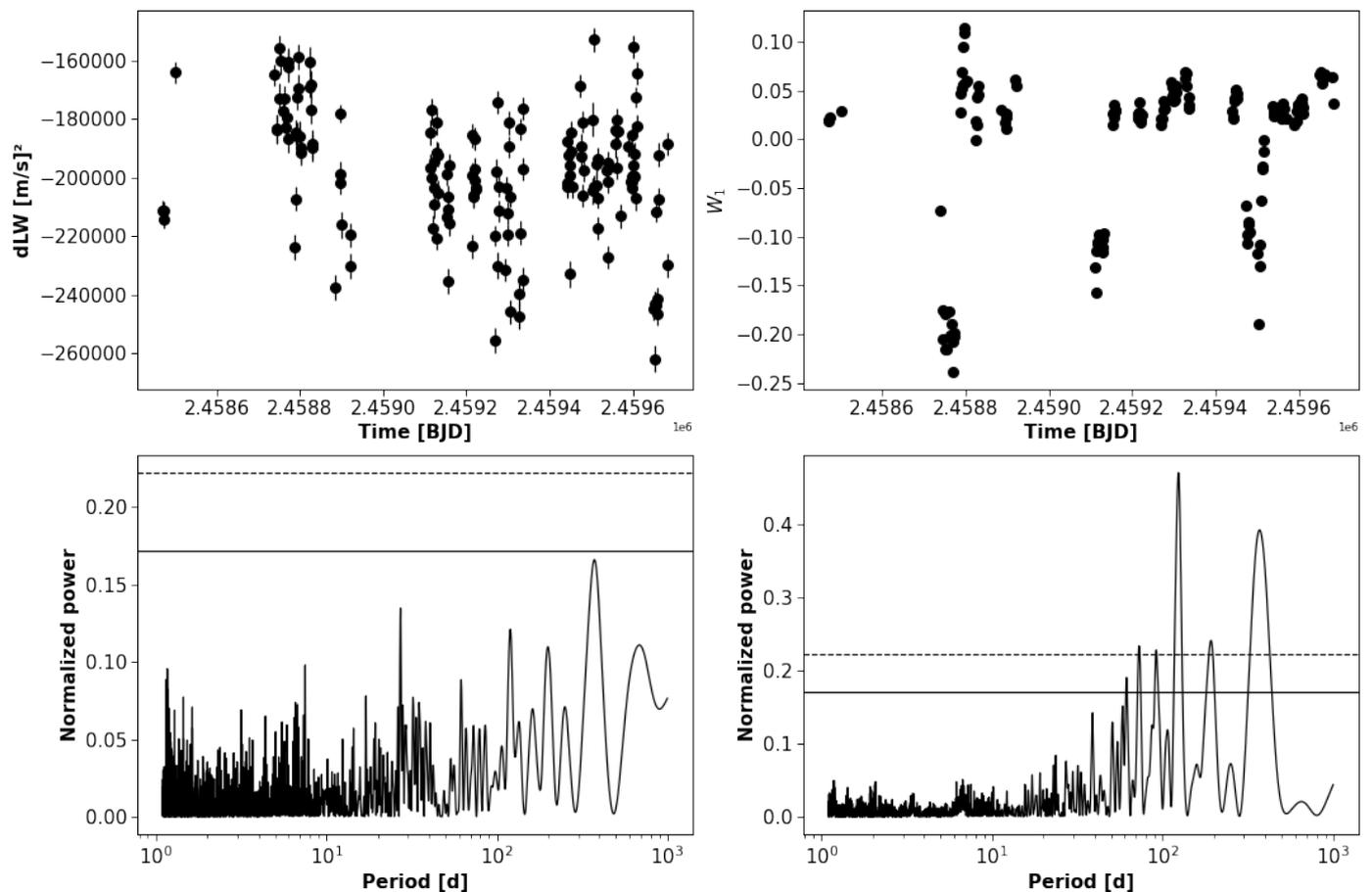


**Fig. 7.** Comparison of retrieved and injected semi-amplitudes. Top panel displays the retrieved semi-amplitude  $K_{computed}$  versus the injected semi-amplitude  $K_{injected}$ . The bottom panel illustrates the residuals, demonstrating the successful recovery of the injected signal using the *Wapiti* method, with the inclusion of the wPCA reconstruction in the fit model.

steps, 1,000 burn-in samples, and 100 walkers which revealed a periodicity of  $137^{+37}_{-16}$  days comparable within uncertainties to the one obtained in Stock et al. (2020). This led us to conclude that  $W_1$  could serve as a good proxy for stellar activity and the specifics of this analysis can be found in Appendix D. In addition, in Donati et al. (2023) it was found when studying AUMic that the first time derivative of  $W_1$  correlated well with the RV which is not the case in our study (Pearson coefficient of -0.11). That said, a noteworthy result is the correlation between the first principal component of the wPCA applied to the RVs and  $W_1$  with a Pearson coefficient of 0.96, suggesting that the secondary peak of the first principal vector at 120d may be related to stellar activity and not be a 1 year harmonic. This highlights that despite the lack of any clear signs of activity in both RV and dLW time series, the presence of such activity can still be detected in their principal vectors. The results from both our data and Stock et al. (2020) demonstrate that the impact of stellar activity on RV data can manifest with a  $\sim 2\sigma$  difference to the periodicity reported in Fouqué et al. (2023) when studying  $B_\ell$ . This indicates that the effect of stellar activity on RV data may not always occur at the same period, for instance due to differential rotation, and must be evaluated with available indicators on a case-by-case basis.

## 5. Discussion, summary, and prospects

The *Wapiti* method successfully eliminated the spurious signals in the GJ 251 data and detected the planetary signal. Although our study demonstrated the effectiveness of the method in accurately estimating planetary parameters, it is important to carefully consider the characteristics of the data. For instance, in situations where a high semi-amplitude planetary signal or trend is present, it would be recommended to first fit these signals and then apply the method to the residuals.



**Fig. 8.** Analysis of dLW and W1 time series. The left panels display, at the top, the dLW time series and its corresponding periodogram at the bottom, which shows no indication of periodicity in the data. On the right, the first principal component of the per-line dLW time series obtained from the wPCA is displayed, which exhibits a pronounced 124-day period peak.

Furthermore, due to its data-driven nature, this method does not require a model for the origins of the corrected systematic errors. Nevertheless, the *Wapiti* method does not aim to replace more physically motivated approaches, such as the zero point correction (which in the instance of GJ 251 is not sufficient), but rather to be a tool that works in conjunction with them. In fact, this method is quite helpful and convenient when investigating the sources of systematics, as outlined in the following discussion.

### 5.1. A BERV effect in SPIRou LBL RVs of GJ 251

From the outset of our study, it was apparent that the spurious signals at 180 and 365 days in our dataset could be attributed to a residual telluric effect, as verified through visual examination of the data in the BERV space in Figure 9. This analysis revealed the presence of a BERV effect, with significant fluctuations occurring in the original data at BERV values of approximately 10 and 25 km s<sup>-1</sup>. It is worth noting that the largest fluctuation observed in our data corresponds to a velocity of the star that is close to zero in the observer’s frame of reference due to the star’s velocity of  $22.98^{+0.22}_{-0.22}$  km s<sup>-1</sup>.

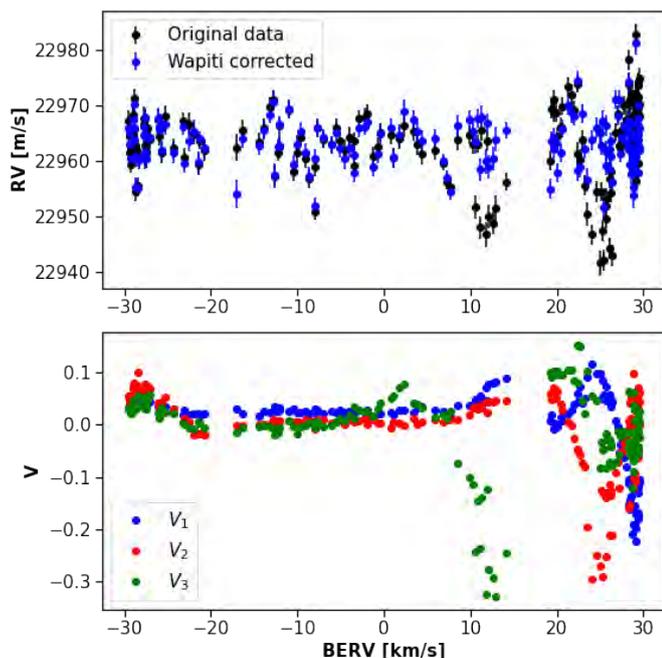
The principal vectors successfully captured the BERV dependency, as evidenced by an evident structure in the BERV space. For the sake of readability we only displayed the first three vectors in Figure 9. Therefore, it can be concluded that the *Wapiti* method was able to correct for systematics in the

original data by removing the BERV dependency through subtracting the wPCA reconstruction. Although the origin of this BERV effect in the reduced data remains somewhat unclear, the next subsection describes how the *Wapiti* products can be used to investigate the source of systematics.

### 5.2. Investigating the sources of systematics

An approach to search for the underlying physical causes of systematic variations in our data would be to examine the scores obtained from the wPCA. These scores represent the transformed data in the new coordinate system established by the principal vectors. To quantify the influence of a specific principal component on a given line, we examined its score values  $U_n$ . Our method involved computing the value  $\tilde{U}_n = \frac{U_n - \bar{U}_n}{\sigma(U_n)}$ , where  $\bar{U}_n$  and  $\sigma(U_n)$  denote the weighted mean and uncertainties of  $U_n$ , respectively. For each line  $\lambda$ , we determined the maximum absolute Z-score value of  $\tilde{U}_n(\lambda)$  over all the components. Recall that the Z-score of a value  $x$  in a sample with mean  $\mu$  and standard deviation  $\sigma$  is given by  $Z(x) = \frac{x - \mu}{\sigma}$ . Hence, the metric we employed to quantify the extent to which a line is affected by the systematics is determined for each line  $\lambda$  as follows:

$$z(\lambda) = \max_n |Z(\tilde{U}_n(\lambda))|. \quad (3)$$



**Fig. 9.** Variability of RV data and principal vectors with respect to BERV. The top panel displays the original data’s variability with respect to the BERV and how our method removed this dependency in the Wapiti corrected RV. The bottom panel demonstrates the behavior of the first three principal vectors with the BERV.

The higher the value of  $z$ , the more a line can be considered to be affected by at least one of the 13 components. In Figure 10, we have constructed a plot that displays the lines with a  $z$  value above an arbitrary threshold of  $z_0 = 10$ . This selection criterion resulted in a total of 63 lines being displayed on the plot. Additionally, we have overplotted in blue a synthetic telluric spectrum obtained using TAPAS (Bertaux et al. 2014) available on the APERO GitHub repository<sup>4</sup> as well as the template of GJ 251 that was used to perform the LBL algorithm.

Figure 10 reveals that the lines most affected by the systematics are predominantly situated within the  $H$  band (1500–1800 nm), as well as in the immediate vicinity of the telluric regions surrounding this band. However, we note that 7 of them are located outside this shared region. To try to establish a connection between affected lines and stellar lines, we computed atomic and molecular lines with VALD (Pakhomov et al. 2017) for a star with a temperature of 3500 K. The atomic mask consists of 2 932 lines from 24 elements (H, Na, Mg, Al, Si, S, K, Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Ga, Rb, Sr, Y, Zr, Ba, Lu). As for the molecular mask, it includes 4 666 lines from 3 elements (OH, CO, CN). In Figure 10, we employed a color-coding scheme to highlight any line from the two masks that met two criteria: first, it fell within the start and end points of an affected line, and second, its depth was above 0.7% to avoid being considered as noise due to the S/N of our observations. The depth of the line being determined as the contrast of the line’s background in comparison to the mean of the boundaries before and after it. Based on our analysis, we observed that 14 atomic lines (in blue), 8 OH lines (in red) and 9 CO lines (in green) could correspond to the affected lines, which accounts for 50% of the total affected lines. Furthermore, we observed that 27 of those lines that have a po-

<sup>4</sup> <https://github.com/njcuk9999/apero-drs/tree/main/apero/data/spirou/telluric>

tential association with the star are located in the  $H$  band. For the remaining lines located in the telluric region, our findings suggest a possible complex interplay between telluric and stellar lines. Specifically, the overlapping of molecular or atomic lines with telluric lines may explain why both a 1-year signal and the star’s rotational period can be distinguished in  $V_1$ .

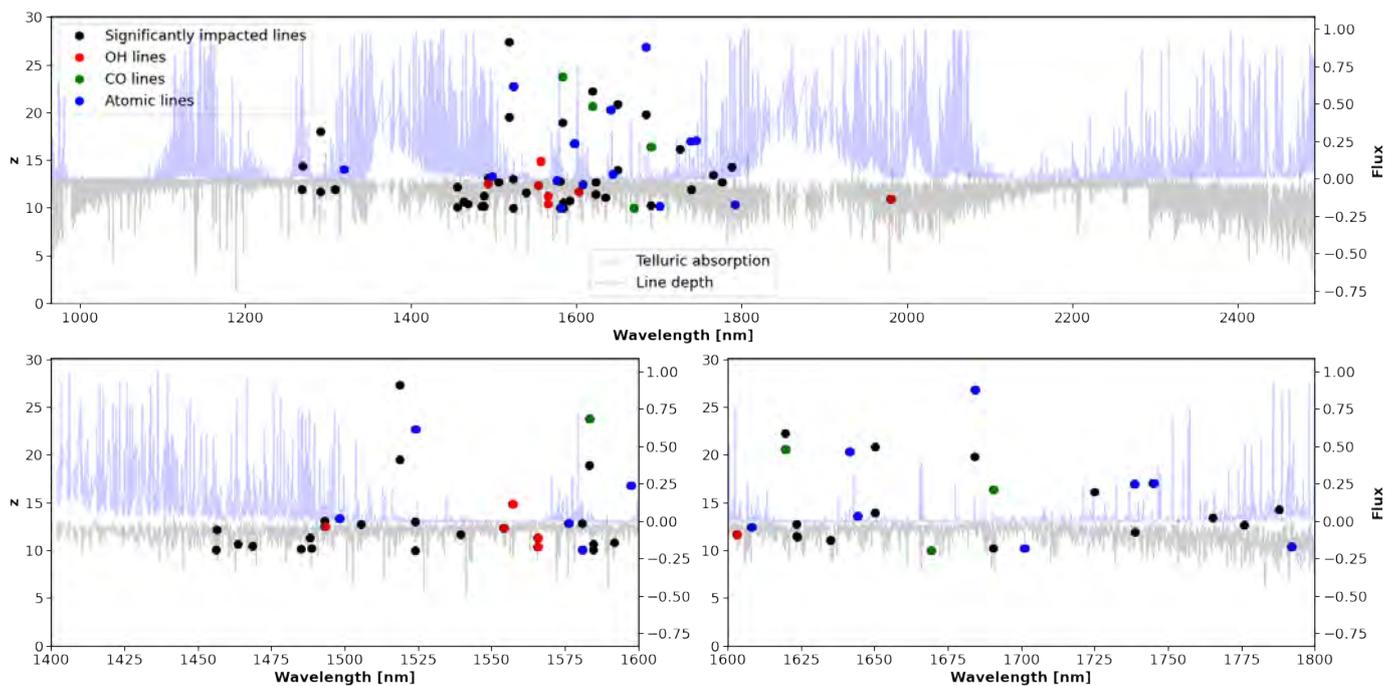
However, the origin of the 32 remaining lines with a  $z$  value above 10 is uncertain, some of them could be molecular lines as of yet unidentified but also of a completely different origin. While telluric or stellar activity could potentially be contributing factors, they may not fully explain the observed impact, suggesting the presence of other sources of spurious signals in the data. The complex interplay between different sources of systematics makes it challenging for the wPCA to distinguish them clearly, particularly since the wPCA is a linear method that can not differentiate between nonlinear relationships between features in the data. Therefore, should the interplay between overlapping stellar lines and telluric lines be nonlinear, we may encounter such an issue. Ultimately, the purpose of this discussion is not to conclusively explain the origin of the systematics, but rather to demonstrate how the results obtained from the Wapiti method can lead to a better understanding of them. In order to identify the exact cause of the systematics, it would be necessary to conduct an investigation of the scores associated with the principal vectors across multiple targets, which is beyond the scope of this study.

We also wanted to examine if those systematics only impact a few lines and what would happen if we were to simply remove those high impacted lines from the calculation when computing the final RV time series. We tested this idea by removing lines for various threshold values of  $z$ . We computed the periodogram of the resulting time series and we see in Figure 11 that even a very strict criterion of removing lines with  $z \geq 3$  only very weakly enhanced the 14.2 d period signal ( $FAP = 4.9 \times 10^{-5}$ ) while not effectively removing the systematics from the RV data, at the cost of removing 1 939 lines from the computation. These results suggest that simply removing the most impacted lines does not eliminate systematics from the RV time series. Instead, the wPCA reconstruction performed by the Wapiti method is necessary to successfully eliminate the systematics

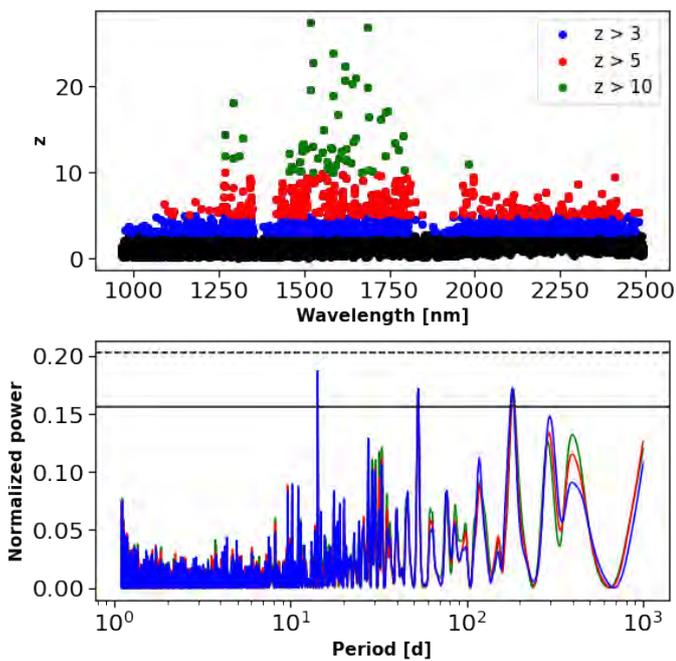
Further research could explore the potential of variational autoencoders (VAE, Kingma & Welling (2013)) to reconstruct the per-line RV time series, which may be able to capture nonlinear relationships between the systematic errors. Previous research has explored the use of neural networks in the examination of RV time series by incorporating both RV and stellar activity indicators as input (Perger et al. 2023; Camacho et al. 2023). However, the present study concentrates on the performance of the Wapiti method and demonstrates that a wPCA, more straightforward, easier to follow, and more efficient computationally, already produces promising outcomes. Additional investigation would be necessary to assess the advantages of applying a VAE to per-line RV time series.

### 5.3. Conclusion and future work

In this paper, we have presented Wapiti, a new data-driven algorithm for correcting systematic errors in RV data. The algorithm used a dataset of per-line RV time series generated by the LBL algorithm and employed a weighted principal component analysis to subtract the wPCA reconstruction from the original data. This process created a new RV time series free of systematic errors. To determine the appropriate principal vectors, the algorithm incorporated a permutation test followed by leave-p-out



**Fig. 10.** The scores of the most affected lines using all components. In the top panel, the lines that are significantly affected by the component are displayed in black. If these lines could be attributed to a OH line and a CO line, they are color-coded in red and green respectively, while if they correspond to an atomic line, they are colored in blue. The telluric absorption computed with TAPAS is also displayed, as well as the template of GJ 251 used by the LBL algorithm. The two lower panels provide zoomed-in views from 1400 to 1800 nm.



**Fig. 11.** The impact of removing affected lines on the RVs. Top panel shows the  $z$  metric of all lines. Bottom panel displays the periodogram of the resulting time series in green, red and blue when respectively rejecting lines with a  $z$  value above 10, 5 and 3.

cross-validation algorithm. Once the relevant number of components have been identified an outlier rejection step followed and the wPCA was applied to the remaining data points. The wPCA reconstruction was consequently removed from original RV time

series, allowing for the derivation of a new Wapiti corrected RV time series.

The application of Wapiti to GJ 251 allowed to successfully free the RV data from spurious signals at 6 months and 1 year while allowing the detection of a known temperate super-Earth with a period of 14.2 days. An examination of the sources of corrected systematics, as revealed by the  $z$  metric, suggested that the method was able to address various sources of spurious signals in the data, including but not limited to stellar activity, telluric lines, and other unidentified sources. This highlights the potential for SPIRou to detect exoplanets, even in the presence of known or unknown sources of systematics. Furthermore, the algorithm has the potential to enhance the precision and accuracy of RV data from future exoplanet detection surveys, thereby advancing our understanding of the diversity and frequency of exoplanetary systems. Finally, this work provided further evidence to support the idea that the first principal vector  $W_1$  derived from the wPCA of the dLW time series can be employed as a reliable activity indicator to detect the stellar rotation period in our data.

While the Wapiti algorithm has shown encouraging results, there are still some remaining questions and challenges that need to be addressed. Future research should aim to evaluate the algorithm on a broader range of RV data and exoplanetary systems and better understand the sensitivity of planetary parameters to the method. A specific direction for future work will be to apply the Wapiti algorithm to targets of the SLS-Planet Search program that exhibit systematic signals of unknown origin at similar periods.

*Acknowledgements.* This project received funding from the European Research Council under the H2020 research & innovation program (grant #740651 NewWorlds). E.M. acknowledges funding from FAPEMIG under project number APQ-02493-22 and research productivity grant number 309829/2022-4 awarded by the CNPq, Brazil. J.H.C.M. is supported in the form of a work contract funded by Fundação para a Ciência e Tecnologia (FCT) with the reference DL 57/2016/CP1364/CT0007;

and also supported from FCT through national funds and by FEDER-Fundo Europeu de Desenvolvimento Regional through COMPETE2020-Programa Operacional Competitividade e Internacionalização for these grants 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, PTDC/FIS-AST/29942/2017. This work has been carried out within the framework of the NCCR PlanetS supported by the Swiss National Science Foundation under grants 51NF40\_182901 and 51NF40\_205606. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement SCORE No 851555) We acknowledge funding from Agence Nationale pour la Recherche (ANR, project ANR-18-CE31-0019 SPLASH) X.D. and A.C. acknowledge funding by the French National Research Agency in the framework of the Investissements d'Avenir program (ANR-15-IDEX-02), through the funding of the "Origin of Life" project of the Université Grenoble Alpes. BK acknowledges funding from the European Research Council (ERC) under the H2020 research & innovation programme (grant agreements #865624 GPRV). This research made use of the following software tools: NumPy; a fundamental package for scientific computing with Python. pandas; a library providing easy-to-use data structures and data analysis tools. Astropy; a community-developed core Python package for Astronomy. SciPy; a library for scientific computing and technical computing. RadVel; a python package for modeling radial velocity data. matplotlib; a plotting library for the Python programming language. tqdm; a library for creating progress bars in the command line. seaborn; a data visualization library based on matplotlib. wpc; a python package for weighted principal component analysis. emcee; a python package for MCMC sampling. PyAstronomy; a collection of astronomical tools for Python. celerite: a fast Gaussian process library in Python.

## References

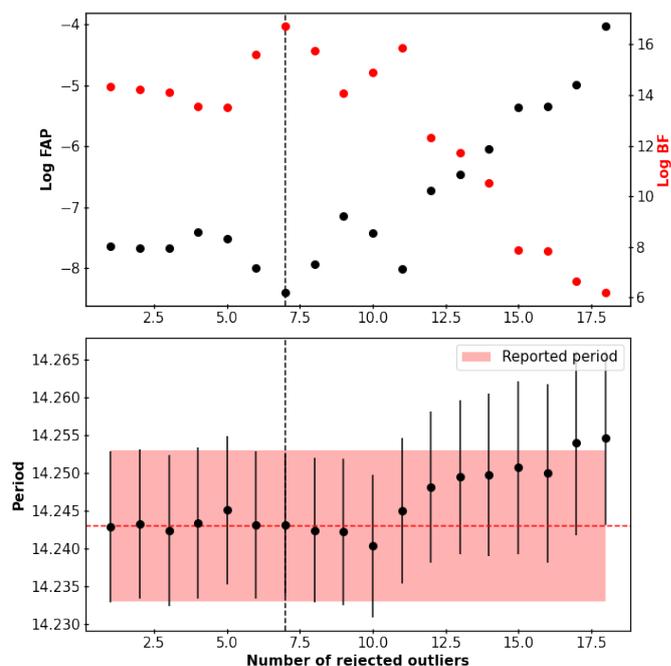
- Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O'Neil, M. 2015, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38, 252
- Anglada-Escudé, G., Amado, P. J., Barnes, J., et al. 2016, *Nature*, 536, 437
- Artigau, É., Astudillo-Defru, N., Delfosse, X., et al. 2014, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 9149, *Observatory Operations: Strategies, Processes, and Systems V*, ed. A. B. Peck, C. R. Benn, & R. L. Seaman, 914905
- Artigau, É., Cadieux, C., Cook, N. J., et al. 2022, *AJ*, 164, 84
- Artigau, É., Saint-Antoine, J., Lévesque, P.-L., et al. 2018, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 10709, *High Energy, Optical, and Infrared Detectors for Astronomy VIII*, ed. A. D. Holland & J. Beletic, 107091P
- Baluev, R. V. 2008, *MNRAS*, 385, 1279
- Becerril, S., Mirabet, E., Lizon, J. L., et al. 2017, in *Materials Science and Engineering Conference Series*, Vol. 278, *Materials Science and Engineering Conference Series*, 012191
- Bertaux, J.-L., Lallement, R., Ferron, S., & Boonne, C. 2014, in *13th International HITRAN Conference*, 8
- Bonfils, X., Delfosse, X., Udry, S., et al. 2013, *A&A*, 549, A109
- Bonfils, X., Mayor, M., Delfosse, X., et al. 2007, *A&A*, 474, 293
- Bonfils, X., Delfosse, X., Udry, S., et al. 2013, *A&A*, 549, A109
- Bouchy, F. & Doyon, R. 2018, in *European Planetary Science Congress, EPSC2018-1147*
- Bouchy, F., Pepe, F., & Queloz, D. 2001, *A&A*, 374, 733
- Butler, R. P., Vogt, S. S., Laughlin, G., et al. 2017, *AJ*, 153, 208
- Cadieux, C., Doyon, R., Plotnikov, M., et al. 2022, *AJ*, 164, 96
- Camacho, J. D., Faria, J. P., & Viana, P. T. P. 2023, *MNRAS*, 519, 5439
- Challita, Z., Reshetov, V., Baratchart, S., et al. 2018, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 10702, *Ground-based and Airborne Instrumentation for Astronomy VII*, ed. C. J. Evans, L. Simard, & H. Takami, 1070262
- Collier Cameron, A., Ford, E. B., Shahaf, S., et al. 2021, *MNRAS*, 505, 1699
- Cook, N., Hobson, M., Bouchy, F., & Martioli, E. 2022, *APER0: A Pipeline to Reduce Observations*, *Astrophysics Source Code Library*, record ascl:2211.019
- Courcol, B., Bouchy, F., Pepe, F., et al. 2015, *A&A*, 581, A38
- Cretignier, M., Dumusque, X., Allart, R., Pepe, F., & Lovis, C. 2020, *Astronomy and Astrophysics*, 633, A76
- Cretignier, M., Dumusque, X., Hara, N. C., & Pepe, F. 2021, *A&A*, 653, A43
- Cretignier, M., Dumusque, X., & Pepe, F. 2022, *A&A*, 659, A68
- Cristofari, P. I., Donati, J. F., Masseron, T., et al. 2022, *MNRAS*, 516, 3802
- da Silva, J. G., Santos, N. C., Bonfils, X., et al. 2012, *Astronomy and Astrophysics*, 541, A9
- Delchambre, L. 2015, *MNRAS*, 446, 3545
- Donati, J. F., Cristofari, P. I., Finocciety, B., et al. 2023, *MNRAS*[arXiv:2304.09642]
- Donati, J. F., Kouach, D., Moutou, C., et al. 2020, *MNRAS*, 498, 5684
- Dressing, C. D. & Charbonneau, D. 2015, *ApJ*, 807, 45
- Dumusque, X. 2018, *Astronomy and Astrophysics*, 620, A47
- Dumusque, X., Pepe, F., Lovis, C., & Latham, D. W. 2015, *ApJ*, 808, 171
- Faria, J. P., Mascareño, A. S., Figueira, P., et al. 2022, *Astronomy and Astrophysics*, 658, A115
- Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, *celerite: Scalable 1D Gaussian Processes in C++, Python, and Julia*, *Astrophysics Source Code Library*, record ascl:1709.008
- Foreman-Mackey, D., Conley, A., Meierjürgen Farr, W., et al. 2013, *emcee: The MCMC Hammer*, *Astrophysics Source Code Library*, record ascl:1303.002
- Fouqué, P., Martioli, E., Donati, J. F., et al. 2023, *A&A*, 672, A52
- Fulton, B. J., Petigura, E. A., Blunt, S., & Sinukoff, E. 2018, *PASP*, 130, 044504
- Gaia Collaboration. 2020, *VizieR Online Data Catalog*, I/350
- Gaidos, E., Mann, A. W., Kraus, A. L., & Ireland, M. 2016, *MNRAS*, 457, 2877
- Gibson, N. P., Merritt, S., Nugroho, S. K., et al. 2020, *MNRAS*, 493, 2215
- Hébrard, É. M., Donati, J. F., Delfosse, X., et al. 2016, *MNRAS*, 461, 1465
- Henry, T. J., Jao, W.-C., Subasavage, J. P., et al. 2006, *AJ*, 132, 2360
- Hobson, M. J., Bouchy, F., Cook, N. J., et al. 2021, *A&A*, 648, A48
- Hsu, D. C., Ford, E. B., & Terrien, R. 2020, *MNRAS*, 498, 2249
- Kanodia, S. & Wright, J. 2018, *Research Notes of the American Astronomical Society*, 2, 4
- Kingma, D. P. & Welling, M. 2013, arXiv e-prints, arXiv:1312.6114
- Kirkpatrick, J. D., Henry, T. J., & McCarthy, Donald W., J. 1991, *ApJS*, 77, 417
- Kotani, T., Tamura, M., Suto, H., et al. 2014, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 9147, *Ground-based and Airborne Instrumentation for Astronomy V*, ed. S. K. Ramsay, I. S. McLean, & H. Takami, 914714
- Mahadevan, S., Ramsey, L., Bender, C., et al. 2012, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 8446, *Ground-based and Airborne Instrumentation for Astronomy IV*, ed. I. S. McLean, S. K. Ramsay, & H. Takami, 84461S
- Martioli, E., Hébrard, G., Fouqué, P., et al. 2022, *A&A*, 660, A86
- Mascareño, A. S., González-Álvarez, E., Osorio, M. R. Z., et al. 2023, *Astronomy and Astrophysics*, 670, A5
- Micheau, Y., Kouach, D., Donati, J.-F., et al. 2018, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 10702, *Ground-based and Airborne Instrumentation for Astronomy VII*, ed. C. J. Evans, L. Simard, & H. Takami, 107025R
- Newton, E. R., Irwin, J., Charbonneau, D., Berta-Thompson, Z. K., & Dittmann, J. A. 2016, *ApJ*, 821, L19
- Pakhomov, Y., Piskunov, N., & Ryabchikova, T. 2017, in *Astronomical Society of the Pacific Conference Series*, Vol. 510, *Stars: From Collapse to Collapse*, ed. Y. Y. Balega, D. O. Kudryavtsev, I. I. Romanyuk, & I. A. Yakunin, 518
- Parès, L., Donati, J. F., Dupieux, M., et al. 2012, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 8446, *Ground-based and Airborne Instrumentation for Astronomy IV*, ed. I. S. McLean, S. K. Ramsay, & H. Takami, 84462E
- Perger, M., Anglada-Escudé, G., Baroch, D., et al. 2023, *A&A*, 672, A118
- 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
- Reiners, A. & Zechmeister, M. 2020, *ApJS*, 247, 11
- Sabotta, S., Schlecker, M., Chaturvedi, P., et al. 2021, *A&A*, 653, A114
- Schwarz, G. 1978, *Annals of Statistics*, 6, 461
- Stock, S., Nagel, E., Kemmer, J., et al. 2020, *A&A*, 643, A112
- Tal-Or, L., Trifonov, T., Zucker, S., Mazeh, T., & Zechmeister, M. 2019, *MNRAS*, 484, L8
- Tamaz, O., Mazeh, T., & Zucker, S. 2005, *MNRAS*, 356, 1466
- The Astropy Collaboration. 2018, *astropy v3.1: a core python package for astronomy*, Zenodo
- Trifonov, T., Tal-Or, L., Zechmeister, M., et al. 2020, *A&A*, 636, A74
- Veira, V. 2012, *Computational Ecology and Software*, 2, 103
- Zechmeister, M. & Kürster, M. 2009, *Astronomy and Astrophysics*, 496, 577
- Zechmeister, M., Reiners, A., Amado, P. J., et al. 2018, *A&A*, 609, A12

## Appendix A: Supplementary information regarding the outliers rejection

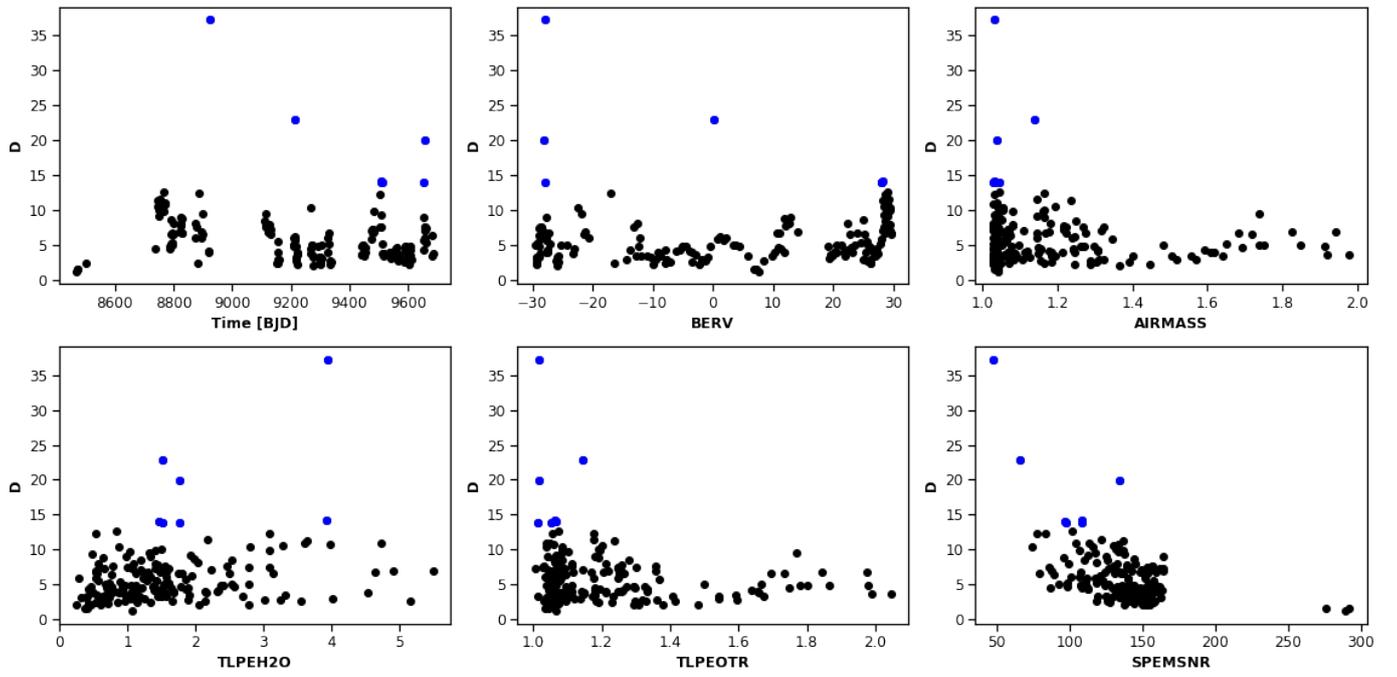
The process of identifying outliers for rejection involved using the  $D$  metric which quantifies the degree of anomaly of an epoch. They were subsequently ranked in decreasing order of  $D$  and were removed one by one, and we stop when  $D \leq 10$ . The Wapiti method was applied to the resulting time series and the one with the maximum statistical significance of the highest peak in the resulting periodogram was chosen, which in our case resulted in the rejection of 7 outliers. This significance was determined using the log false alarm probability (FAP), but we also verified that the same conclusion was reached using the log Bayes factor.

Such a criterion to remove outliers is justified for our data set from the observation that the Wapiti method yields a prominent peak at the same period independently of the number of rejected outliers as shown in Figure A.1. From the definition of  $D$  it comes that those rejected epochs are outliers in at least one principal component, yet they do not visually appear in the RV time series and one may be interested in what could be their origin.

This is the reason why we tried to visualize their physical origin in Figure A.2 using the most-straightforward parameters available from the APERO reduction, namely the BERV, the airmass (AIRMASS), the SPIRou Exposure Meter S/N Estimate (SPEMSNR), and the telluric water exponent (TLPEH2O) and other telluric molecules (TLPEOTR) calculated with TAPAS during the precleaning of the telluric correction. Regrettably, our investigation did not provide conclusive evidence to determine the source of the rejected outliers. These outliers exhibit a wide range of BERV values and are present at low airmasses ( $\leq 1.2$ ). Furthermore, some of the rejected outliers have a SPEMSNR that is not necessarily lower than the average, even though we do notice that the first two outliers are the data points with the lowest SPEMSNR. Ultimately, it is likely that multiple factors contribute to the anomalous behavior of these observations, and a single parameter is insufficient to explain this anomaly. We advise that when applying this outlier removal method to other targets, it should be done with caution and the results should be carefully verified, as shown in Figure A.1.



**Fig. A.1.** Effect of removing outliers on the signal detection. The top panel displays the evolution of the log FAP of the signal as the number of outliers removed increases. The red dots overlaid on the plot represent the evolution of the log BF, which follows a similar pattern and reaches its maximum value at the same number of rejected outliers. The bottom panel of the figure shows the detected period in the resulting time series, which remains consistent within error-bars with the reported planetary period.



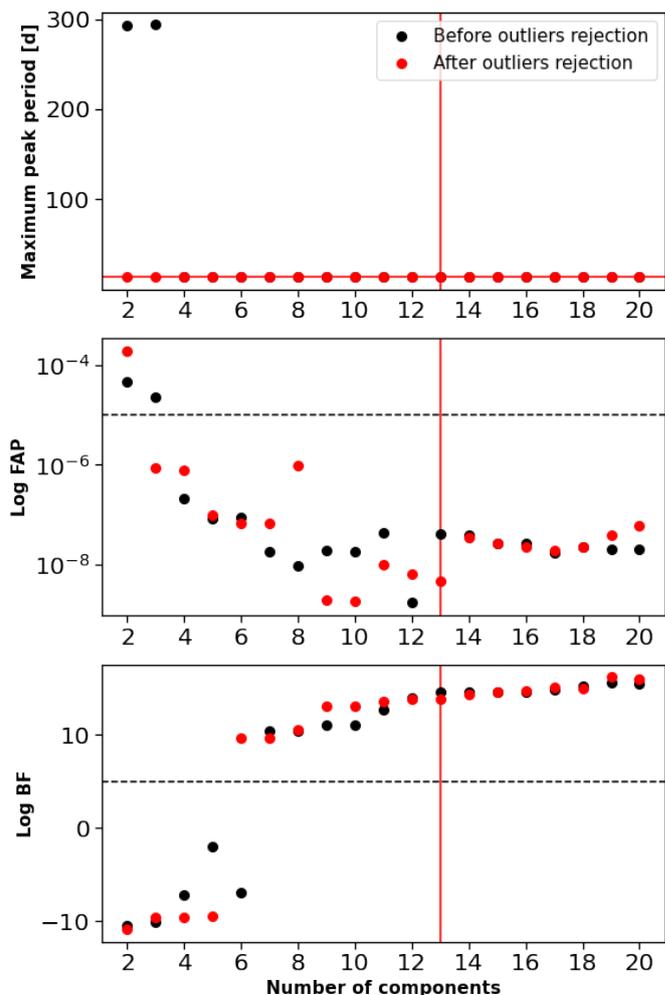
**Fig. A.2.** Analysis of anomaly degree and rejected observations. All panels display the degree of anomaly  $D$  of the observations with respect to various parameters. The data points that appear in blue represent the rejected observations that were excluded to maximize the significance of the planetary signal in our data.

## Appendix B: Impact of the components on the signal

After determining that 13 components were required to accurately reconstruct our data using wPCA, one may wonder how the number of components affects the signals reported in our data. To address this question, we conducted a straightforward test by applying the method to time series with an increasing number of components ranging from 2 to 20. For each time series, we computed its periodogram and the FAP of its maximum peak period. Additionally, we performed a computation of the  $\log BF$  to compare a Keplerian model with a 14.2-day orbital period and an offset with a model solely consisting of an offset. The outcomes are exhibited in Figure B.1. As our rejection of outliers is expected to favor the use of 13 components, we conducted this examination on both RV time series, both before and after outliers rejection.

Contrary to expectations, the use of 13 components did not yield the highest level of statistical significance for the 14.2-day signal, but instead required 12 components without outlier rejections and 10 components with such rejections. It is noteworthy that the planetary signal can be directly extracted using only 2 components in the time series data that underwent Wapiti correction and outlier rejection, compared to 3 components required without such rejection. Notably, in the case of no outliers rejection, the number of components had minimal impact on the results after reaching 5 components, with the signal remaining above a  $\log BF$  of 5 and a FAP value below  $10^{-5}$ .

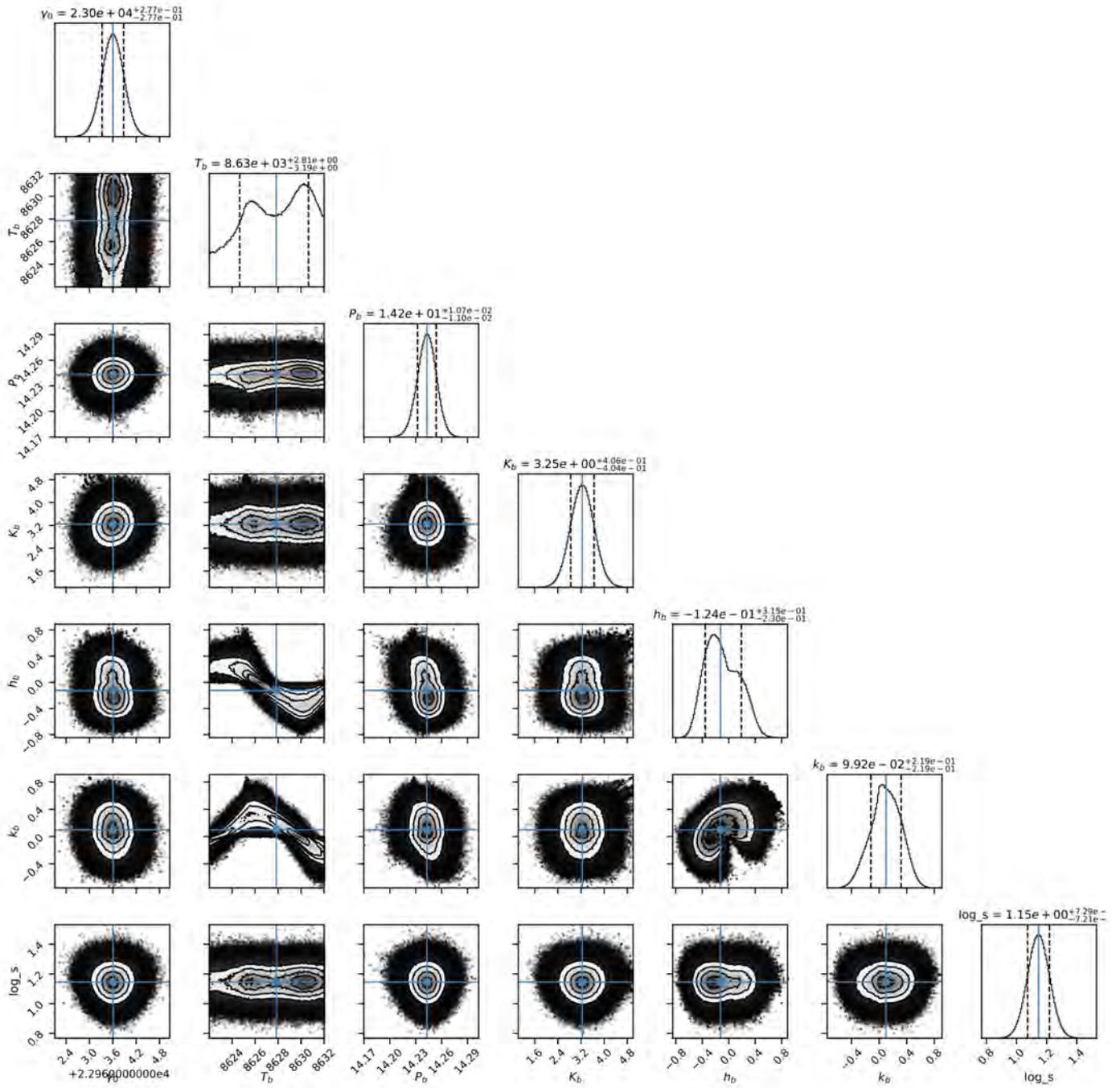
To sum up, the results indicate that the number of components employed does not significantly impact the outcomes beyond a certain threshold. However, given that the objective is to use the Wapiti method on unidentified planetary systems where the orbital period is not a priori known, an unbiased and robust approach is necessary to determine the appropriate number of components. The identification of 13 as the optimal number of components through our independent preprocessing methods provides robust evidence of the effectiveness of our methodology.



**Fig. B.1.** Impact of the choice of number of components on the signal detection. The top panel displays the peak associated with the maximum period in the periodogram when the wPCA reconstruction is eliminated from the data using varying numbers of components. In the middle panel, the logarithm of the FAP linked to this signal is presented, while the bottom panel depicts the  $\log BF$  evaluating the significance of a 14.2 d Keplerian signal compared to an offset. The black and red data points correspond, respectively, to the outcomes obtained from the pre- and post-outlier rejection RV time series.

### **Appendix C: MCMC results from the RV analysis**

Orbital parameters computed using an MCMC over the *Wapiti* corrected RV data from *SPIRou* with 100,000 samples, 10,000 burn-in samples, and 100 walkers.



**Fig. C.1.** Corner plot displaying the results of MCMC analysis on the Wapiti corrected RV time series is presented. This analysis involves fitting a model that consists of an offset  $\gamma_0$ , an RV jitter  $s$  (in log), and a Keplerian with parameters including the time of periastron  $T_b$ , the orbital period  $P_b$ , the semi-amplitude  $K_b$ , and the eccentricity and argument of periastron represented by  $\sqrt{e} \cos \omega$  ( $h_b$ ) and  $\sqrt{e} \sin \omega$  ( $k_b$ ), respectively.

## Appendix D: MCMC results from the W1 analysis

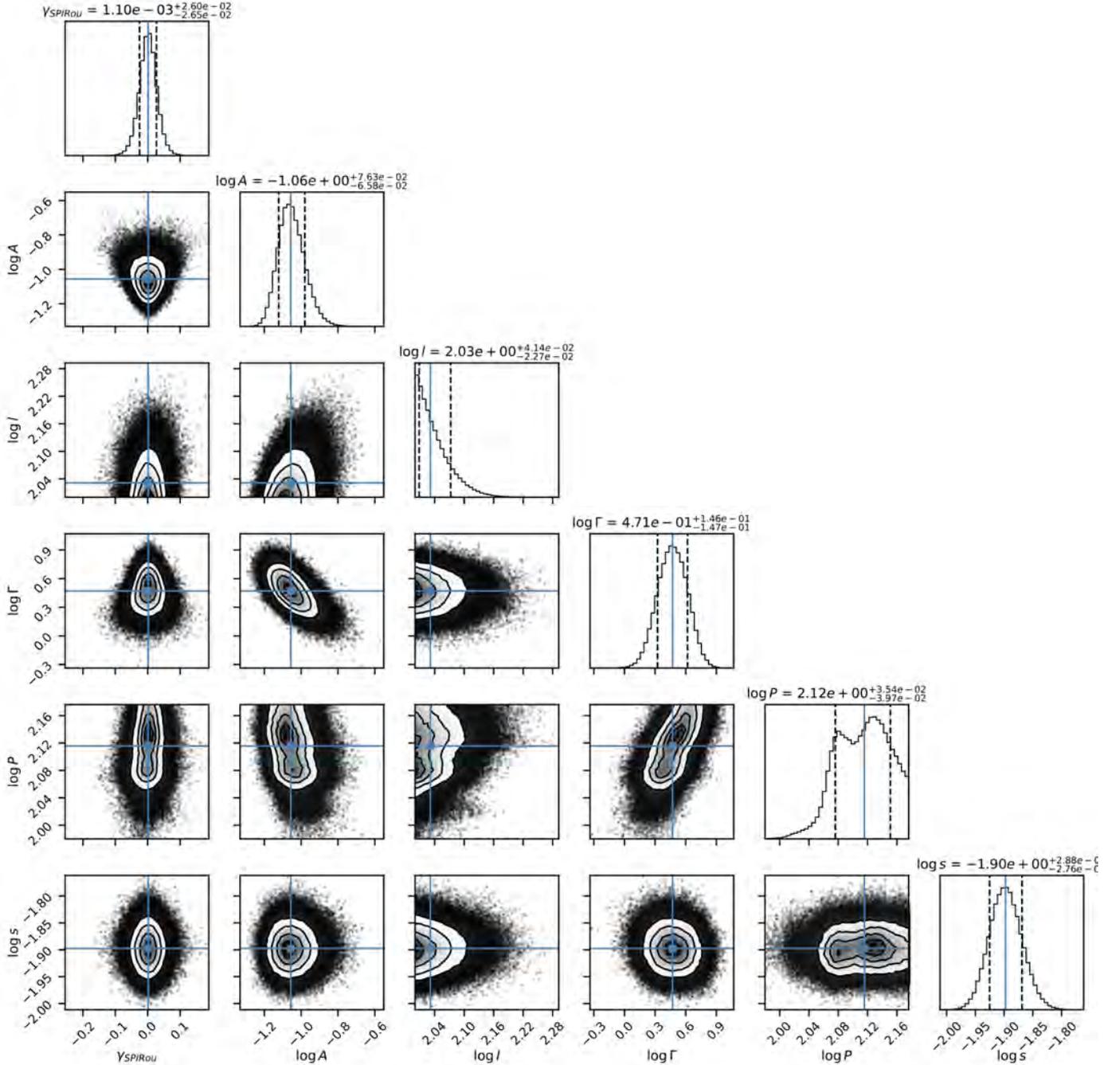
We fit a quasi-periodic GP in order to estimate the value of the periodicity in  $W_1$  fitting a quasi-periodic kernel of the form

$$k(\tau) = A \exp\left(-\frac{\tau^2}{2l^2} - \Gamma \sin^2\left(\frac{\pi}{P}\tau\right)\right) \quad (\text{D.1})$$

as well as a jitter term  $s$  and a constant  $\gamma$ . We computed an MCMC using 10,000 steps, 1,000 burn-in samples, and 100 walkers. The result is displayed in Table D.1 and the corner plot is shown in Figure D.1.

**Table D.1.** Prior and posterior distributions of the quasi-periodic GP model of the SPIRou  $W_1$  time series.

Parameter	Prior distribution	Posterior distribution	
		In units	linear units
<b>GP <math>W_1</math></b>			
Amplitude	$\log A \quad \mathcal{U}(-3, 3)$	$-1.07^{+0.08}_{-0.07}$	$0.09^{+0.02}_{-0.01}$
Decay time [days]	$\log l \quad \mathcal{U}(2, 3)$	$2.03^{+0.04}_{-0.02}$	$107^{+11}_{-5}$
Smoothing parameter	$\log \Gamma \quad \mathcal{U}(-5, 5)$	$0.55^{+0.25}_{-0.19}$	$3.5^{+2.7}_{-1.3}$
Cycle length (period) [days]	$\log P \quad \mathcal{U}(\ln(80), \ln(200))$	$2.14^{+0.10}_{-0.05}$	$137^{+37}_{-16}$
Uncorr. Noise	$\log \sigma \quad \mathcal{U}(-5, 0)$	$-1.90^{+0.03}_{-0.03}$	$0.0127^{+0.0009}_{-0.0008}$


**Fig. D.1.** Corner plot displaying the results of MCMC analysis on the  $W_1$  time series. This analysis involves fitting a quasi-periodic GP over the data and an offset.



**C.2.2 (Submitted to A&A) Confirmation of an exoplanet around  
Gl 480 and detection of a candidate exoplanet orbiting Gl 382**

# The SPIRou Legacy Survey

## Confirmation of an exoplanet around Gl 480 and detection of a candidate exoplanet orbiting Gl 382

M. Ould-Elhkim<sup>1,\*</sup> , C. Moutou<sup>1</sup> , J-F. Donati<sup>1</sup> , É. Artigau<sup>2,3</sup>, C. Cadieux<sup>2</sup>, E. Martioli<sup>4,5</sup> , T. Forveille<sup>6</sup>, Ryan Cloutier<sup>7,8</sup> , A. Carmona<sup>6</sup>, P. Fouqué<sup>1</sup>, P. Charpentier<sup>1</sup>, P. Larue<sup>6</sup>, N.J. Cook<sup>2</sup>, X. Delfosse<sup>6</sup>, and R. Doyon<sup>2</sup>

<sup>1</sup> Université de Toulouse, UPS-OMP, IRAP, 14 avenue E. Belin, Toulouse, F-31400, France, France  
e-mail: merwan.ould-elhkim@irap.omp.eu

<sup>2</sup> Trottier Institute for Research on Exoplanets, Université de Montréal, Département de Physique, C.P. 6128 Succ. Centre-ville, Montréal, QC H3C 3J7, Canada

<sup>3</sup> Observatoire du Mont-Mégantic, Université de Montréal, Département de Physique, C.P. 6128 Succ. Centre-ville, Montréal, QC H3C 3J7, Canada

<sup>4</sup> Laboratório Nacional de Astrofísica, Rua Estados Unidos 154, 37504-364, Itajubá - MG, Brazil

<sup>5</sup> Institut d'Astrophysique de Paris, CNRS, UMR 7095, Sorbonne Université, 98 bis bd Arago, 75014 Paris, France

<sup>6</sup> Univ. Grenoble Alpes, CNRS, IPAG, 38000 Grenoble, France

<sup>7</sup> Center for Astrophysics | Harvard Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA

<sup>8</sup> Department of Physics & Astronomy, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4L8, Canada

August 28, 2023

### ABSTRACT

**Context.** Advancements in the field of exoplanetary research have extended radial velocity (RV) observations from the optical to the near-infrared (nIR) domain. M dwarf stars, characterized by their lower masses and higher prevalence of rocky planets, have become a focal point of investigation. This study uses data from the SPIRou Legacy Survey Planet Search program (SLS-PS) in the nIR and from the CARMENES spectrograph in the optical to analyze RVs of two nearby M dwarfs, Gl 480 and Gl 382.

**Aims.** The primary objective of this study within the SLS is to apply recently developed data analysis techniques and methodologies that can efficiently mitigate the impact of systematics on RV measurements. By applying these techniques, the study aims at improving the accuracy and reliability of RV data, in order to enable the detection and characterization of exoplanets with higher precision. Furthermore, this work seeks to investigate some more the potential of the line-by-line (LBL) framework applied to SPIRou data, in conjunction with optical RV measurements from other instruments, to facilitate a joint analysis of RV measurements in both the nIR and optical domains.

**Methods.** The study employs a comprehensive approach that combines the LBL framework, the weighted Principal Component Analysis (wPCA) technique, and the innovative Wapiti algorithm. Through an extensive analysis of available stellar activity indicators and data products from Wapiti such as the principal vectors derived from the wPCA, we gain a deeper understanding of the systematics present in our RV data.

**Results.** We demonstrate the efficacy of the Wapiti method in addressing telluric-induced systematics. Specifically, we establish that the per-line radial velocity's first principal vector can be a useful tool for understanding the remaining correlation between the star template's first derivative and the residuals from the telluric correction. Additionally, we show that telluric interference can also affect stellar indicators provided by the LBL framework. By applying the Wapiti correction technique, we achieve the first detection of Gl 480 b in the nIR. By performing a combined analysis of both nIR and optical RV data, we obtain a more precise determination of its orbital parameters, leading to an estimated minimum planetary mass of  $9.7^{+0.9}_{-0.9} M_{\oplus}$  and an orbital period of  $9.5545^{+0.0007}_{-0.0007}$  days. Furthermore, our analysis has suggested the presence of a strong planetary candidate orbiting around Gl 382. Upon characterization, this planet has been identified as sharing similar characteristics to Gl 480 b with a minimum planetary mass of  $8.71^{+0.85}_{-0.83} M_{\oplus}$  and an orbital period of  $10.3843^{+0.0008}_{-0.0007}$  days.

**Key words.** Techniques: radial velocities; Techniques: spectroscopic; Methods: data analysis; Stars: individual: GJ 251; Stars: planetary systems; Stars: low-mass

## 1. Introduction

The first detection of an exoplanet around 51 Pegasi was made possible by the radial velocity (RV) method (Mayor & Queloz 1995) using the optical spectrograph ELODIE at the Haute-

tional project led by Institut de Recherche en Astrophysique et Planétologie, Toulouse, France. SPIRou is an acronym for SPectropolarimetre InfraROuge (infrared spectropolarimeter).

\* Based on observations obtained at the Canada-France-Hawaii Telescope (CFHT) which is operated from the summit of Maunakea by the National Research Council of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique of France, and the University of Hawaii. The observations at the Canada-France-Hawaii Telescope were performed with care and respect from the summit of Maunakea which is a significant cultural and historic site. Based on observations obtained with SPIRou, an interna-

Provence Observatory in France Baranne et al. (1996). Since then, advancements in optical instrumentation with instruments such as HARPS (Mayor et al. 2003) and ESPRESSO (Mégévand et al. 2014) has resulted in the identification of numerous exoplanetary systems (Faria et al. (2022); Mascareño et al. (2023); to quote just a few of the latest results). Despite having historically started in the optical domain, RV observations have recently extended to the near-infrared (nIR) domain with instruments such as GIANO (Carleo et al. 2016), HPF (Metcalf et al. 2019), CARMENES-NIR (Bauer et al. 2020), IRD (Hirano et al. 2020), NIRPS (Thibault et al. 2022) and SPIRou (Donati et al. 2020). The nIR does not benefit from the extensive research that has been conducted in the optical domain with respect to instrumental and data-reduction techniques. Most notably, the nIR infrared comes with new challenges such as telluric correction, instrument stability at low temperatures, wavelength calibration, as well as persistence and other yet to investigate effects of infrared arrays (Artigau et al. 2018; Cook et al. 2022). However, this spectral range provides a unique opportunity to explore the *YJHK* spectral bandpasses (Quirrenbach et al. 2018; Reiners & Zechmeister 2020), which are especially rich in molecular bands such as CO, OH, and H<sub>2</sub>O, in addition to being the domain in which cool stars are the most luminous.

The nIR is thus of interest to better detect and characterize exoplanetary systems around M dwarfs (Cadieux et al. 2022; Martioli et al. 2022; Almenara et al. 2023). Those stars have attracted a great deal of interest due to their abundance in the solar neighborhood (Henry et al. 2006). They are of particular significance as they have lower masses, making them excellent candidates for detecting planet RV signatures that are as low as a few  $\text{m.s}^{-1}$  (Anglada-Escudé et al. 2016; Bonfils et al. 2013; Faria et al. 2022; Mascareño et al. 2023). Additionally, these stars are known to host a larger number of rocky planets than other star types (Bonfils et al. 2013; Dressing & Charbonneau 2015; Gaidos et al. 2016; Hsu et al. 2020; Sabotta et al. 2021).

On the other hand, M dwarfs are known to exhibit a high level of magnetic activity, which can lead to spurious RV signals (Bonfils et al. 2007; da Silva et al. 2012; Newton et al. 2016; Hébrard et al. 2016). However, this activity is wavelength-dependent (Reiners et al. 2013; Baroch et al. 2020) and less important in the nIR. Consequently, an example of the potential synergy between both optical and nIR is to conduct RV measurements in both domains so as to confirm or rule out the planetary nature of ambiguous signals (Cortes-Zuleta et al. 2023; Carmona et al. 2023).

Within this context, the SPIRou Legacy Survey Planet Search program (SLS-PS<sup>1</sup>) is a large program (LP) that aims at detecting and characterizing planetary systems around approximately  $\sim 50$  nearby M dwarfs (Moutou et al. 2023). The program was allocated 153 nights of observation time at the Canada-France-Hawaii Telescope (CFHT) from February 2019 to June 2022. In this paper, we present the analysis of the nearby M dwarfs G1480 and G1382 using SPIRou data obtained as part of this program and other publicly available datasets.

This paper is organized as follows. In Section 2, we describe our targets and the RV data used in our analysis. In Section 3, we provide an overview of the Wapiti correction method to mitigate systematics in the SPIRou data of our two targets. Section 4 presents a separate analysis for each target, discussing their characteristics and the determination of orbital parameters. Finally, Section 5 summarizes the main findings and discusses potential avenues for future study.

## 2. Observations

### 2.1. Targets

G1480 (Wolf 433) is a red dwarf star with spectral type M3.5V (Lépine et al. 2013) that is located  $14.2625 \pm 0.0060$  pc away from Earth (Gaia Collaboration 2020). In Cristofari et al. (2022), its physical properties were determined from the analysis of SPIRou spectra and are summarized in Table 1. Additionally, time-series of the longitudinal magnetic field ( $B_\ell$ ) of this star were obtained from circularly polarized spectropolarimetric sequences. However, no clear periodic variation was detected in the timeseries (Fouqué et al. 2023). In the study by Donati et al. (2023b), which is a parallel analysis of the  $B_\ell$  timeseries using different data reduction and analysis of the spectra, the authors identified a rotational period  $P_{rot}$  of  $25.00 \pm 0.36$  days. However, a prior study by Feng et al. (2020) reported an activity-related signal occurring at a period of  $49.3 \pm 0.2$  days. It is plausible that the 25-day signal observed in the  $B_\ell$  data might actually represent the first harmonic of the true rotational period. Finally, a planet orbiting at a  $9.567 \pm 0.005$  d orbital period was claimed to orbit G1480 from the analysis of HIRES and HARPS data (Feng et al. 2020). The signal has a semi-amplitude  $K = 6.80 \pm 0.87 \text{ m.s}^{-1}$  with an estimated  $M_p \sin i$  of  $13.2 \pm 1.7 M_\oplus$ .

G1382 (V\* AN Sex) is a red dwarf of spectral type M2.0V (Kirkpatrick et al. 1991) located  $7.707 \pm 0.0015$  pc away from Earth (Gaia Collaboration 2020). Its physical properties are also listed in Table 1. Time-series of  $B_\ell$  of this star were also obtained from Stokes V spectropolarimetric sequences and led to the detection and characterization of a signal at  $21.32^{+0.04}_{-0.03}$  days (Fouqué et al. 2023) ( $21.91 \pm 0.16$  d in Donati et al. (2023b)). In Sabotta et al. (2021), a signal with a period of 10.65 days was detected in CARMENES optical RV data. However, the signal was automatically flagged as spurious due to this period being close to the first harmonic of the rotational period. This star was also part of the activity study conducted by Mignon et al. (2023) from HARPS spectra, and they did not find any signs of an activity cycle in the chromospheric activity indices.

### 2.2. The SPIRou spectropolarimeter, the APERO reduction and the LBL algorithm

SPIRou is a high-resolution near-infrared spectropolarimeter installed in 2018 at the Cassegrain focus of CFHT. It has a spectral range going from  $0.98$  to  $2.35 \mu\text{m}$  at a resolving power of 70,000, and is stabilized in a vacuum chamber with milli-K temperature stability (Challita et al. 2018). A detailed presentation of the instrumental design and early measurement performances is provided in Donati et al. (2020).

All SPIRou spectra are reduced using the 0.7.275 version of the SPIRou data reduction system (DRS), called APERO (A PipelinE to Reduce Observations) detailed in Cook et al. (2022). APERO processes science frames to produce 2D and 1D spectra from the two science channels and the RV reference channel, and it uses a combination of exposures from a UNe hollow cathode lamp and a Fabry-Pérot etalon for wavelength calibration (Hobson et al. 2021). Barycentric Earth radial velocity (BERV) is calculated using barycorrpy (Kanodia & Wright 2018) to shift the spectra to the barycentric frame of the solar system. APERO also performs optimal extraction of both science channels and the Fabry-Pérot RV reference channel, flat-fielding, a correction of thermal background, and a correction of the residual leak of the Fabry-Pérot RV reference spectrum onto the science channel.

<sup>1</sup> <https://spirou-legacy.irap.omp.eu/doku.php>

Parameters	Gl 480	Gl 382	Ref.
<b>Stellar parameters</b>			
$T_{\text{eff}}$ (K)	$3509 \pm 31$	$3644 \pm 31$	1
$[M/H]$	$0.26 \pm 0.10$	$0.15 \pm 0.10$	1
$[\alpha/\text{Fe}]$	$-0.01 \pm 0.04$	$-0.02 \pm 0.04$	1
$\log g$	$4.88 \pm 0.06$	$4.75 \pm 0.05$	1
Radius ( $R_{\odot}$ )	$0.449 \pm 0.008$	$0.511 \pm 0.009$	1
Mass ( $M_{\odot}$ )	$0.45 \pm 0.02$	$0.51 \pm 0.02$	1
$\log (L/L_{\odot})$	$-1.562 \pm 0.002$	$-1.384 \pm 0.001$	1
mean $B_{\ell}$ (G)	$6.6^{+3.5}_{-3.3}$	$-0.7^{+4.2}_{-4.2}$	2
$P_{\text{rot}}$ (d)	-	$21.32^{+0.04}_{-0.03}$	2
$P_{\text{rot}}$ (d)	$25.00 \pm 0.24$	$21.91 \pm 0.16$	3
$P_{\text{rot}}$ (d)	$49.3 \pm 0.2$	-	4
$P_{\text{rot}}$ (d)	$25.00 \pm 0.24$	$21.91 \pm 0.16$	3
<b>Astrometry</b>			
Right ascension (J2000)	$12^{\text{h}}38^{\text{m}}52.44^{\text{s}}$	$10^{\text{h}}12^{\text{m}}17.67^{\text{s}}$	5
Declination (J2000)	$+11^{\circ}41'46.14''$	$-03^{\circ}44'44.39''$	5
Parallax (mas)	$70.1139 \pm 0.0297$	$129.7544 \pm 0.0252$	5
Distance (pc)	$14.2625 \pm 0.0060$	$7.707 \pm 0.0015$	5
Proper motion RA (mas/year)	$-1155.756 \pm 0.034$	$-152.760 \pm 0.028$	5
Proper motion D (mas/year)	$-250.862 \pm 0.038$	$-243.693 \pm 0.028$	5
<b>Photometry</b>			
G [mag]	$10.335732 \pm 0.002806$	$8.331761 \pm 0.002821$	5

**Table 1.** Stellar parameters of Gl 480 and Gl 382. **1:** Cristofari et al. (2022), **2:** Fouqué et al. (2023), **3:** Donati et al. (2023b), **4:** Feng et al. (2020), **5:** Gaia Collaboration (2020).

Based on the TAPAS model spectrum of the Earth atmosphere (Bertaux et al. 2014) and a library of hot-star spectra obtained with SPIRou in various conditions of airmass and humidity, AP-ERO corrects the science spectra from telluric spectrum using a three-step telluric correction approach which will be detailed in an upcoming publication (Artigau et al. (2023), in prep.). The final product is a wavelength calibrated telluric corrected 2D spectrum

RVs are subsequently computed from the telluric-corrected spectra using the Line-By-Line (LBL<sup>2</sup>) method. This method is based on the framework developed in Bouchy et al. (2001) and was first proposed by Dumusque (2018) then further explored by Cretignier et al. (2020) for optical RV observations with HARPS and more recently revisited by Artigau et al. (2022) in the context of nIR observations. The LBL algorithm requires a high-signal-to-noise ratio (S/N) template spectrum of the observed star, simply called template in the following, and derived from the median of all collected spectra. RVs are then computed by comparing each spectrum to the template and its first few derivatives. The LBL algorithm is applied to both the stellar extracted spectrum and the Fabry-Pérot spectrum in the RV reference channel, with the latter providing a precise measurement of the instrumental drift at the exact time of our observations. The final RV time series is corrected from this drift. The LBL method has been compared to other methods such as the cross-correlation function and template matching in Martioli et al. (2022) and Artigau

et al. (2022). It provides a mean RV value of the stellar spectrum and per-line RV time-series on which it is possible to apply a weighted Principal Component Analysis (wPCA<sup>3</sup>; Delchambre (2015)). The Wapiti<sup>4</sup> method (Ould-Elhkim et al. 2023) uses this approach and applies a wPCA to the LBL per-line RV time-series in order to clean the final RV time-series from systematics of stellar, tellurics, instrumental or other unknown origins.

### 2.3. RVs data sets

#### 2.3.1. Gl 480<sup>5</sup>

SPIRou observations were collected as series of circular polarization sequences of 4 individual sub exposures per night. The data collection for Gl 480 spans from November 13, 2019 to June 13, 2022, comprising a total of 446 sub-exposures, with each visit comprising four sub-exposures. The airmass during these observations has a median value of 1.3 and the signal-to-noise ratio per  $2.28 \text{ km s}^{-1}$  pixel bin in the middle of the  $H$  band has a median value of 110.7. Data are night-binned using the inverse-variance weighted mean of all observations for each night leading to 111 night-binned data points.

We also retrieved RV time-series collected with HIRES, HARPS and CARMENES (optical data). We found 20 RV data points from HIRES in Feng et al. (2020) ( $RMS = 5.5 \text{ m.s}^{-1}$ ) in

<sup>3</sup> <https://github.com/jakevdp/wpca>

<sup>4</sup> <https://github.com/HkmMerwan/wapiti>

<sup>5</sup> CFHT program: P45. From semester 19A to 22A included.

<sup>2</sup> <https://lbl.exoplanets.ca/>

addition to 37 HARPS ( $RMS = 6.1 \text{ m.s}^{-1}$ ) and 7 CARMENES ( $RMS = 3.3 \text{ m.s}^{-1}$ ) RV measurements, respectively from Trifonov et al. (2020) and Ribas et al. (2023), which were reduced using the SERVAL pipeline (Zechmeister et al. 2018a).

### 2.3.2. Gl 382<sup>6</sup>

SPIRou observations were again conducted using circular polarization sequences. The observations for Gl 382 were carried out from February 02, 2019 to January 09, 2023, totaling 542 sub-exposures with each visit comprising four sub-exposures. The median airmass during these observations was 1.3, and the median signal-to-noise ratio per  $2.28 \text{ km s}^{-1}$  pixel bin in the middle of the  $H$  band was 154.5. Binning the data resulted in 133 night-binned data points.

In addition to the SPIRou data, RV time-series were retrieved from HIRES, HARPS, and CARMENES (optical data). We used 95 RV measurements from HIRES as published in Tal-Or et al. (2019) ( $RMS = 6.1 \text{ m.s}^{-1}$ ). Furthermore, we retrieved 32 HARPS ( $RMS = 6.7 \text{ m.s}^{-1}$ ) and 77 CARMENES ( $RMS = 5.5 \text{ m.s}^{-1}$ ) RV measurements from Trifonov et al. (2020) and Ribas et al. (2023), respectively. These measurements were reduced using the SERVAL pipeline (Zechmeister et al. 2018a).

## 3. Method: Wapiti correction

For the SPIRou time-series, we applied the Wapiti algorithm to correct the RV (drift corrected) for systematics using the 2D (lines  $\times$  epochs) LBL per-line RV time-series. The algorithm is described in detail in Ould-Elhkim et al. (2023). In summary, it uses a wPCA on the per-line RV time-series and automatically selects an appropriate number of components. Once the wPCA is computed, the reconstruction is subtracted from the original RV time series. More details of this correction on each target are provided in Appendix A and the results are displayed in Figure 1.

We analyzed the time-series using a generalized Lomb-Scargle (GLS) periodogram (Zechmeister & Kürster 2009), with a grid of 100,000 points ranging from 1.1 days to 1,000 days, using the astropy package (The Astropy Collaboration 2018). The false alarm probability (FAP) of the maximum peak was calculated using the analytical formula in Baluev (2008). In Figure 1, we displayed both the original and Wapiti-corrected RV time-series, as well as their respective periodograms. We also included the window function of the time-series and the periodogram of the subtracted wPCA reconstruction for better insight.

In the case of Gl 480, two clear outliers were excluded from the data set, along with an anomalous epoch characterized by a low signal-to-noise ratio ( $S/N = 30$ ), prior to employing the Wapiti method, hence three rejected epochs in total. Seven wPCA components were most appropriate for this target and the application of the Wapiti method resulted in a reduction in the RMS and an enhancement of the statistical significance associated with the 9.6 day signal. The original data already indicated the presence of the known RV signal of planetary origin at an orbital period of 9.6 days, with a FAP of  $1.1 \times 10^{-5}$ . However, the application of the Wapiti method further improved the significance of the signal, resulting in a reduced FAP of  $2.2 \times 10^{-6}$ , accompanied by a decrease in the RMS of the corrected time series from  $6.0 \text{ m.s}^{-1}$  to  $4.9 \text{ m.s}^{-1}$ . The periodogram analysis of

the subtracted wPCA reconstruction revealed that this improvement was attributed to the elimination of a 1-year signal from the data, most likely from the yearly modulation of telluric absorption lines.

Regarding Gl 382, no evident outliers were detected in the RV time series. However, prior to employing the Wapiti algorithm, epochs flagged as outliers were still excluded using the methodology outlined in Ould-Elhkim et al. (2023). Specifically, anomalous observations were identified by identifying outliers in the principal vectors, and further details can be found in Appendix A. Consequently, 16 epochs were removed from the analysis and, using eleven components, the statistical significance of a 10.4 day signal increased from an initial FAP of  $1.2 \times 10^{-4}$  in the original RV data to  $6.7 \times 10^{-11}$  in the corrected RV data, while the RMS decreased from  $5.7 \text{ m.s}^{-1}$  to  $4.5 \text{ m.s}^{-1}$ . It is noteworthy that the observed peak corresponds to the signal previously identified in the CARMENES optical data (Sabotta et al. 2021), which was attributed to the rotation of the star. However, in Cortes-Zuleta et al. (2023) and Carmona et al. (2023), it was reported that peaks originating from stellar activity in optical RV data tend to diminish or even vanish in the nIR. Thus the persistence of this peak in the SPIRou data raises doubts regarding its origin from stellar activity.

In the following section, a comprehensive analysis of Gl 480 and Gl 382 will be presented, using the Wapiti corrected RV time series along with other available datasets. The analysis will be conducted separately for each target. Firstly, we will begin by characterizing the stellar activity of the observed targets. Subsequently, we propose a novel approach that leverages the correlations between principal vectors derived with wPCA applied on per-line RVs and telluric absorption lines. This approach enables us to gain a deeper comprehension of the physical meaning of the first few principal vectors. Finally, we will perform an in-depth RV analysis of the targets and accurately determine the orbital parameters associated with their planetary companion.

## 4. Analysis and results

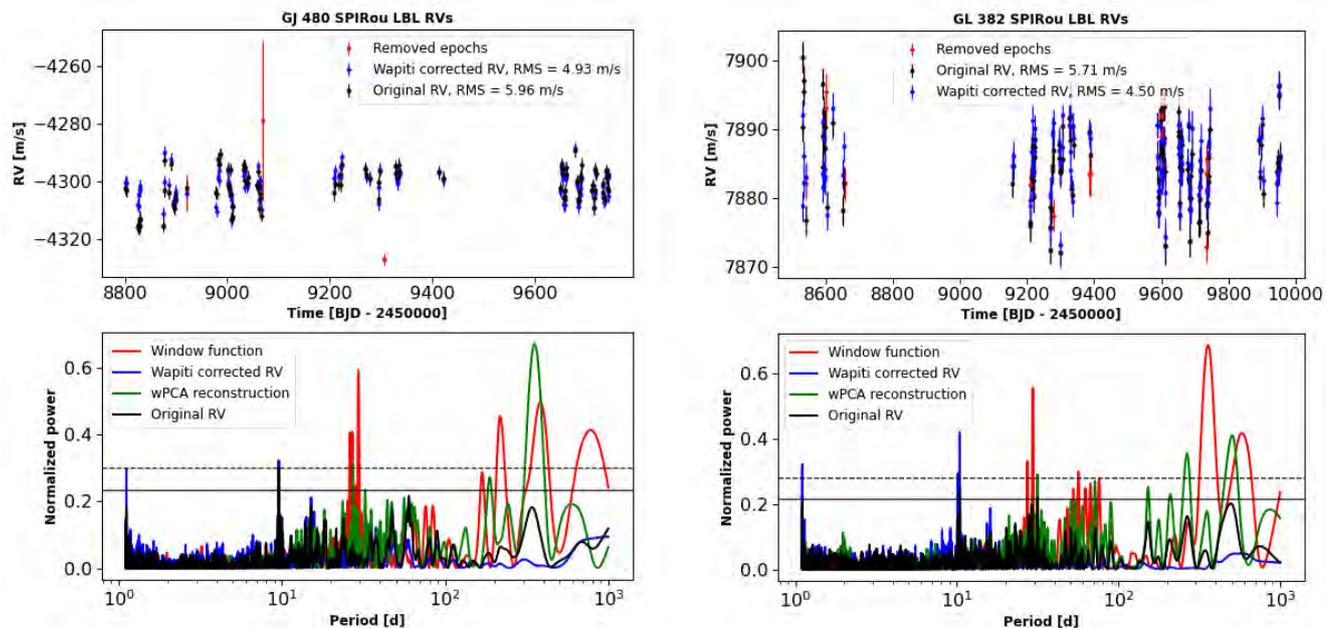
### 4.1. Gl 480

#### 4.1.1. Analysis of the stellar activity signatures

In their study, Feng et al. (2020) performed an activity analysis that detected a signal with a period of  $49.3 \pm 0.2$  days in two stellar indicators,  $H\alpha$  and NaD2, in the HARPS data. We thus examined those indicators as well as the differential line width (dLW), which is another product obtained from SERVAL (Zechmeister et al. 2018b), in the HARPS data. Our results are presented in the panels of Figure 2, where the 9.6 day signal is marked in blue. We find no evidence of activity at the reported period in dLW (panels 1 and 4) or NaD2 (panels 2 and 5), while we did observe a signal in  $H\alpha$  (panels 3 and 6).

Regarding SPIRou, we used three stellar activity indicators:  $V_1$  (the first principal vector obtained through wPCA applied to the per-line RVs), dLW, and  $W_1$  (the first principal vector from wPCA applied to the per-line dLWs). The latter has been demonstrated as a good tracer for stellar rotation in previous studies (Donati et al. 2023a; Ould-Elhkim et al. 2023). Our analysis indicates that dLW does not display a period associated with activity; instead, it exhibits a periodicity of  $\sim 1$  and 0.5 yr (panels 7 and 10). However, both  $W_1$  (panels 8 and 11) and  $V_1$  (panels 9 and 12) exhibit significant signals which could correspond to the reported activity period, with  $W_1$  showing a prominent peak at 46 d. It is important to note that they also demonstrate other various significant periods, such as 120 and 365 days.

<sup>6</sup> CFHT program: Q57. From semester 19A to 22A included.



**Fig. 1.** Wapiti correction of the SLS targets. Left panels GJ 480 and right panels GL 382. The top panels display the RV time series in black, the corrected RV time series in blue and the epochs removed from the computation in red. The bottom panels show the periodogram of the original RV time series in black, of the corrected RV time series in blue, of the wPCA reconstruction subtracted to the original RVs in green and the window function in red. The significance levels of  $FAP = 10^{-3}$  and  $FAP = 10^{-5}$  are indicated by the solid and dashed horizontal lines, respectively.

Additionally, we investigated the behavior of  $W_1$  and  $V_1$  in the " $V_{\text{tot}}$ " space (figure 2), where " $V_{\text{tot}}$ " is defined as the difference between the star's velocity and its Barycentric Earth Radial Velocity (BERV). In doing so, we observed a distinct dependence of both indicators on this parameter, suggesting that they are likely capturing telluric systematics rather than reflecting stellar activity. Additionally, we found a strong correlation between  $W_1$  and  $V_1$  (with a Pearson coefficient of  $r = -0.97$ ), similar to what was observed in a previous study involving GJ 251 (Ould-Elhkim et al. 2023) and indicative of a telluric contamination of this indicator.

It is important to mention that in the case of the young active M dwarf AU Mic (Donati et al. 2023a) a significant correlation between the first time derivative of  $W_1$  and the RV was attributed to stellar activity. However, in our investigation, we did not uncover a similar correlation, as evident in panel 15 ( $r = 0.05$ ). This disparity could stem from differing levels of activity between the two targets. Notably, the stellar activity in AU Mic might be strong enough to overcome the telluric contamination. Based on our analysis, we interpret these findings as indicating that the  $W_1$  indicator is contaminated by tellurics in the case of GJ 480.

#### 4.1.2. Telluric contamination of stellar lines

The observed relationship with  $V_{\text{tot}}$  indicates that the presence of telluric lines has an impact on our radial velocity (RV) data. However, this effect had already been noted in a previous study (Ould-Elhkim et al. 2023). In this section, we aim to further investigate the specific information captured by  $V_1$  and  $W_1$ .

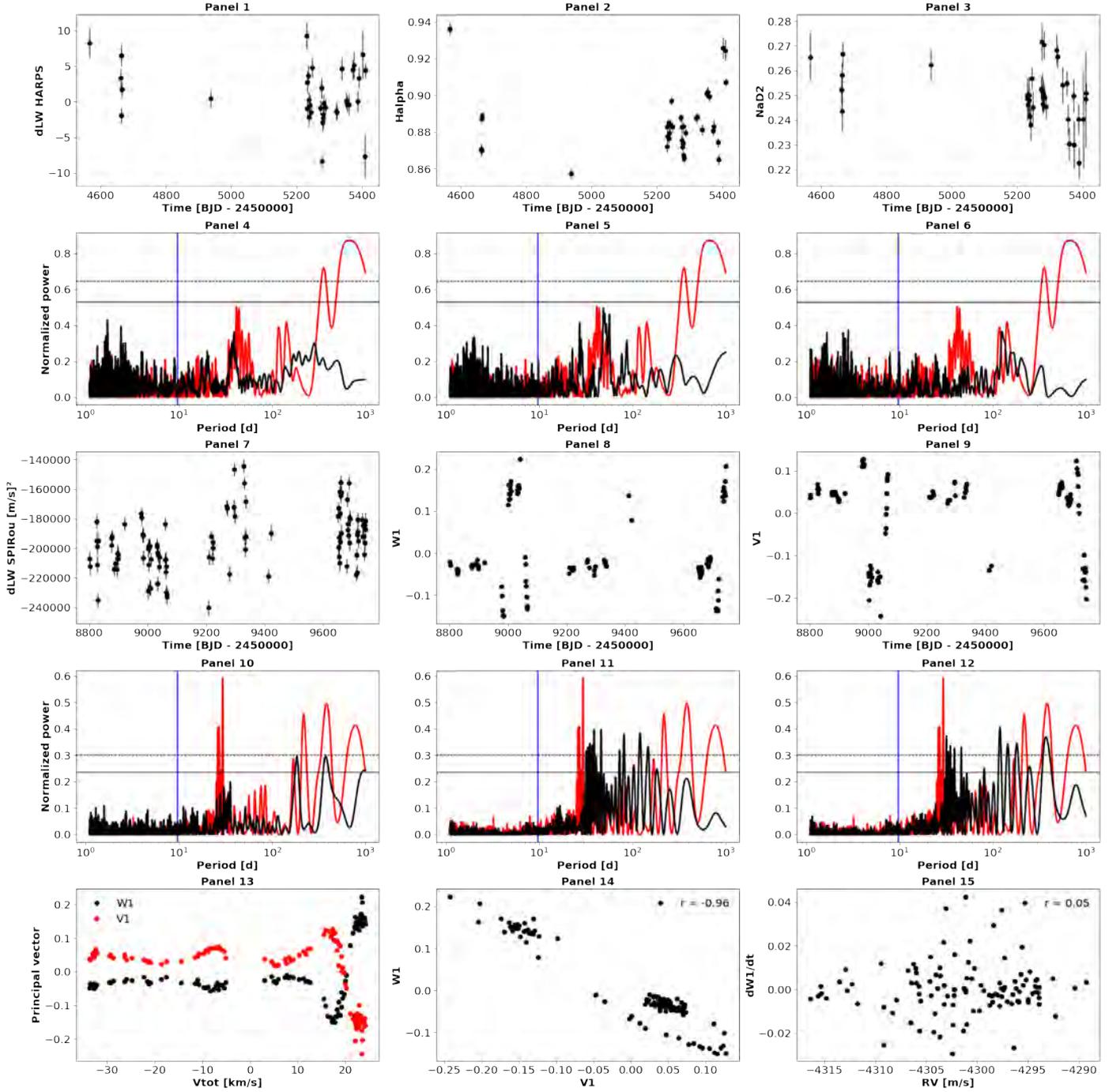
To assess the influence of telluric lines on our data, we calculate the Pearson correlation between the telluric 2D spectra and the first derivative of the star template for each individual line. As a reminder, in the LBL algorithm, the RV is determined by subtracting the star template from the observed spectrum and then projecting the residuals onto the first derivative of the star

template (Artigau et al. 2022). Therefore, the use of Pearson correlation is a natural choice since it allows us to investigate the linear correlation of this derivative with the telluric lines.

An initial approach to address the impact of telluric effects would involve removing the lines that exhibit high correlation with tellurics (Pearson coefficient  $\geq 0.8$ ) from the RV computation at each epoch. However, upon testing this idea, we discovered that, unlike the Wapiti method, it did not effectively reduce systematic errors or improve the significance of the signal, even when setting the threshold to 0.5. It appears that simply rejecting lines is not the key to mitigating telluric effects. Instead, a more successful approach involves reconstructing the telluric signatures and subtracting them from the data, similar to what the Wapiti method achieves through wPCA reconstruction. This conclusion aligns with the finding of Ould-Elhkim et al. (2023), where the authors attempted to reject lines based on their scores (the transformed values of the data along a specific principal component) but were unable to effectively eliminate systematics.

Nevertheless it would still be valuable understand what  $V_1$  represents in the data. By combining the computed Pearson correlation of the derivative of the template with tellurics and the previous work of Ould-Elhkim et al. (2023) we found that it is possible to exhibit a clear correspondence between  $V_1$  and the linear correlation of lines with tellurics.

In order to do so, we first followed a similar approach as described in Ould-Elhkim et al. (2023) to identify the lines most influencing the principal vectors. This involved determining the score  $U_n$  and its uncertainty  $dU_n$  for each line with respect to the  $n^{\text{th}}$  principal vector  $V_n$ . Next, we calculated the Z-score  $z_n$  relative to  $U_n$  for each line, using  $z_n = \frac{U_n - \mu}{dU_n}$ , where  $\mu$  is the weighted mean of  $U_n$ . To quantify the extent to which a line affects the systematics is determined for each line  $\lambda$  we use the z metric:



**Fig. 2.** Various activity indicators of Gl 480 from HARPS and SPIRou. The blue band represents the reported planetary period. The red periodogram represents the window function. The significance levels of  $FAP = 10^{-3}$  and  $FAP = 10^{-5}$  are indicated by the solid and dashed horizontal lines, respectively. **Panel 1 and 4:** dLW time series from HARPS and its periodogram. **Panel 2 and 5:**  $H\alpha$  index from HARPS and its periodogram. **Panel 3 and 6:** NaD2 index from HARPS and its respective periodogram. **Panel 7 and 10:** dLW from SPIRou and its respective periodogram. **Panel 8 and 11:**  $W_1$  and its periodogram. **Panel 9 and 12:**  $V_1$  and its respective periodogram. **Panel 13:**  $V_1$  and  $W_1$  in the  $V_{tot}$  space. **Panel 14:** Correlation plot between  $V_1$  and  $W_1$ . **Panel 15:** Correlation plot between the first time derivative of  $W_1$  and the RV.

$$z(\lambda) = \max_n |z_n(\lambda)|. \quad (1)$$

The higher the value of  $z$ , the more a line can be considered to affect at least one of the 7 PCA components previously identified by Wapiti. In Figure 3, we presented the  $z$  metric values for all lines. Lines with a  $z$  value above 5 are highlighted in red. Additionally, we overlaid the telluric template ( $T_{\oplus}$ ) computed

by taking the median of the 2D telluric spectrum derived from the observations of Gl 480.

We can also observe that the mean Pearson correlation between the tellurics and the derivative of the template ( $r_{\oplus}$ ) considering only the lines with a  $z$  value above 5 (4,827 lines, 27.4% of the total number of lines), has a strong correlation with  $V_1$  ( $r = 0.89$ ) and  $W_1$  ( $r = 0.88$ ).

This result suggests that one possible interpretation of  $V_1$  is that it provides a direct method to investigate  $r_\oplus$ . Moreover, the strong correlation between  $W_1$ , the first principal vector of the per-line dLWs, and  $r_\oplus$  may indicate a flaw in the way the LBL algorithm computes it. According to Artigau et al. (2022), it is computed independently of the per-line RVs, but with the difference that the residuals of the difference between the observed spectrum and the star template are projected onto the second derivative of the template.

The presence of telluric contamination could potentially impact not only the Doppler shift but also the width of a spectral line. This can lead to contaminated dLW values, which could explain the ineffectiveness of using it as a reliable stellar activity indicator. The advantage of our interpretation of  $V_1$  lies in its potential use as a proxy in future research to identify and separate telluric contamination on each line before computing its dLW value.

We conducted a similar analysis to the one for  $V_1$  with the other principal vectors but did not find any correlations, even when attempting to use the Pearson correlation between the tellurics and the second derivative of the template. This shows that while it is likely the remaining principal vectors are linked to more advanced effects of telluric contamination, their interpretation remains unclear.

Moreover, in our dataset, there might be other sources of systematic effects that could complicate the investigation of the relationship between the principal vectors and telluric contamination. This is evident from the observation that the values of the  $z$  metric exceeding 5 are not confined solely to regions of pronounced telluric absorption lines but are distributed across the entire spectrum. To address this issue and better understand the specific impact of telluric contamination, we are planning to conduct controlled simulations in an upcoming study (Ould-Elhkim et al., 2023c, in preparation). These simulations will allow us to precisely regulate the injection of telluric residuals, isolating their effects on the RV data. By doing so, we will gain deeper insights into what the principal vectors capture and how they are linked to telluric contamination.

In the upcoming section, we will proceed with the RV analysis of Gl 480, using the *Wapiti*-corrected RV data in combination with other available time series.

#### 4.1.3. Confirmation of Gl 480 b

Firstly, we validated our ability to recover the orbital parameters described in Feng et al. (2020) through the use of our HARPS and HIRES data. A comparison with SPIRou of the periodograms generated from the data of the two instruments is presented in Figure 4. In their investigation, Feng et al. (2020) employed a Keplerian model to characterize the signal produced by the planet, following their combination of the HIRES and HARPS data. Additionally, several noise models, such as a white noise model, a moving average (MA) model, and an autoregressive (AR) model, were considered to account for possible red noise present in the data. They determined that the signal remained robust under all noise models. In contrast, we will employ solely a white noise model in our investigation, but we will consider both circular and eccentric models to describe the planetary orbit. Specifically, we will examine the circular model  $\mu_C$  and the Keplerian (eccentric) model  $\mu_K$

$$\mu_C = \gamma + K \sin(2\pi(t - T)/P) \quad (2)$$

$$\mu_K = \gamma + K [\sin(\omega + \nu(t)) + e \cos \omega] \quad (3)$$

where the parameters of both models include the offset of the radial velocity ( $\gamma$ ), the semi-amplitude of the signal ( $K$ ), and the orbital period ( $P$ ). In the circular model, the time offset is represented by  $T$ , while in the Keplerian model, the eccentricity, the argument of periastron, and the true anomaly are represented respectively by  $e$ ,  $\omega$ , and  $\nu$ . It is noteworthy that the offset of the radial velocity is fitted separately for each instrument. The *radvel* package (Fulton et al. 2018) was used to model the Keplerian.

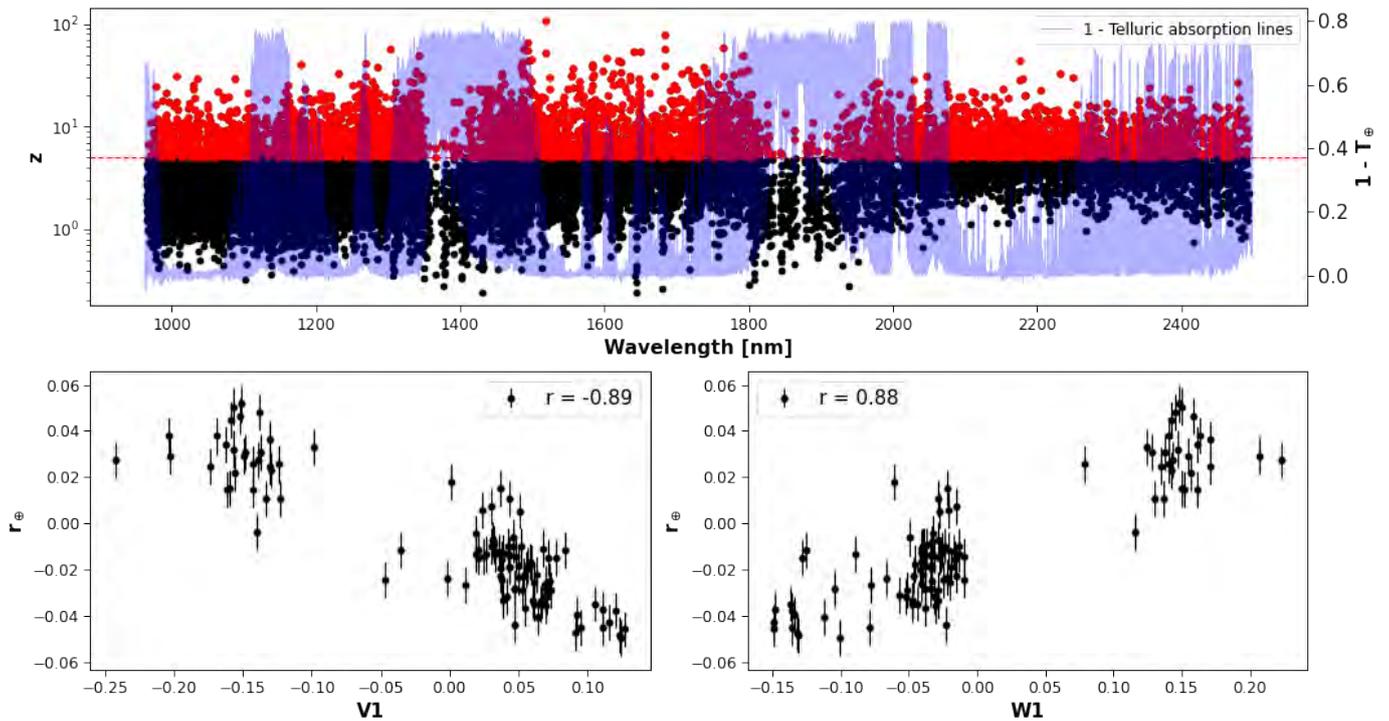
We conducted the analysis of three datasets consisting of one combined HARPS and HIRES dataset, another containing only SPIRou data, and finally one using all data ("ALL") including CARMENES ones. For the three datasets, we applied the circular and Keplerian models and used maximum likelihood estimation (MLE) to compute their respective Bayesian Information Criterion (BIC), as described in Schwarz (1978). It is important to note that the model used may not be orthogonal to the principal vectors used in the wPCA projection when applying the *Wapiti* method. This can lead to imprecise characterization of the orbital parameters and must be taken into account. Thus, denoting  $R$  as the wPCA reconstruction, the two models fitted to the SPIRou data are actually  $\hat{\mu}_C = \mu_C - R(\mu_C)$  and  $\hat{\mu}_K = \mu_K - R(\mu_K)$ . An injection-recovery test, as detailed in Ould-Elhkim et al. (2023), has confirmed that this approach is sufficient to obtain accurate estimations of the semi-amplitude of injected signals within the data.

The circular model demonstrated the lowest BIC value for all data sets (with a difference ranging from 8 to 10, see Table 2), consequently we used this model to characterize Gl 480 b in all of the data sets. To estimate the uncertainties of the model parameters, we employed a MCMC approach using the *emcee* module (Foreman-Mackey et al. 2013). Specifically, we generated 10,000 samples with a burn-in of 1,000 samples and 100 walkers. The resulting orbital and derived planetary parameters are summarized in Table 2, and are compared to those obtained in Feng et al. (2020). We estimated the planet's mass,  $M_p \sin i$ , by using the star's mass and its uncertainty obtained from Cristofari et al. (2022), which is listed in Table 1. For estimating the equilibrium temperature ( $T_{eq}$ ), we relied on the star's radius and effective temperature also derived from the same study. The corner plot from the MCMC computed over the data sets are available in Appendix B.1.

In our independent analysis, we obtained a semi-amplitude of  $6.78_{-0.81}^{+0.80}$  m s<sup>-1</sup> for the HARPS and HIRES data set, which is consistent with the value reported in Feng et al. (2020) within two sigmas. The final fit on the dataset using all instruments is presented in Figure 5, we provide the phase-folded representation of the data and the residuals. No other statistically significant signals were found in the periodogram of the residuals once the model was fitted.

#### 4.1.4. Detection sensitivity of the *Wapiti* correction

In this section, our focus is on evaluating the detection sensitivity of the *Wapiti* correction when applied to the SPIRou LBL RV data of Gl 480. Our aim is to determine what signals can be reliably detected using this method. To achieve this, we followed a specific procedure. First, we subtracted the Gl 480 b signal from each per-line RV time series. Next, we introduced a circular planetary signal into these RVs, characterized by a given semi-amplitude ( $K$ ) and orbital period ( $P$ ). Finally, we applied the



**Fig. 3.** The top panel displays the  $z$  values for all wavelengths of GJ 480 using 7 components, with the lines having values above 5 marked in red, defining them as "affected lines." Moving to the bottom left panel, it illustrates the correlation between  $V_1$  and  $r_{\oplus}$ , represented as the mean Pearson correlation between the tellurics and the derivative of the template for all affected lines. Finally, the bottom right panel presents the correlation plot between  $W_1$  and  $r_{\oplus}$ .

Wapiti correction with 7 principal components to these modified time series to assess if we could successfully recover the injected signal.

To carry out this analysis, we constructed a grid of orbital periods ranging from 2 to 80 days, with intervals of 0.25 days, and a semi-amplitude grid spanning from 1 to 5  $\text{m.s}^{-1}$ . At fixed values of  $P$  and  $K$ , we introduced a simulated signal, repeating this process 10 times with random phase. To consider a signal as detected, it had to meet specific criteria: it should be the most prominent peak in the periodogram, and its FAP should be below  $10^{-3}$ . We then calculated the detection rates before and after applying the Wapiti correction, as illustrated in Figure 6. To provide a more comprehensive understanding of the results, we also computed the percentage of successfully recovered periods at a given semi-amplitude. We applied two criteria for recovery: a less conservative one, where the detection rate was above 50%, and a more conservative one, where the detection rate was above 80%. The resulting figures are also presented in Figure 6.

In summary, the application of the Wapiti correction significantly enhances the recovery of signals with low semi-amplitudes compared to cases without correction. Specifically, when considering a conservative criterion, it becomes apparent that signals with semi-amplitudes as low as  $3.5 \text{ m.s}^{-1}$  can be efficiently recovered thanks to the Wapiti correction. The non-conservative criterion shows even more favorable results, with recovery rates reaching up to 95% for signals with semi-amplitudes exceeding  $4.5 \text{ m.s}^{-1}$ . However, it's important to note that, based on the current dataset and correction method, signals below  $3.5 \text{ m.s}^{-1}$  remain undetectable.

## 4.2. GJ382

### 4.2.1. Analysis of the stellar activity signatures

For GJ 382, we employed the same analysis approach used for GJ 480. Initially, we examined various indicators of stellar activity available for both HARPS and SPIRou, just like we did for GJ 480. Additionally, we had access to a larger dataset from CARMENES, comprising 77 data points, which allowed us to use multiple stellar activity indicators. These indicators included the dLW, the  $H\alpha$  and NaD2 indices, as well as the Ca II IRT1 (849.802,nm), Ca II IRT2 (854.209,nm), and Ca II IRT3 (866.214,nm) indices.

The time series data along with their respective periodograms are presented in Figure 7 for CARMENES and Figure 8 for HARPS and SPIRou. In these figures, we have highlighted the 10.4 day signal detected in the SPIRou and CARMENES RV data with a blue line, and the window functions are shown in red in all periodograms.

Upon an examination of these indicators, we found that both CARMENES and HARPS data exhibited power in their  $H\alpha$  periodograms near the known 21 day rotational period (panel 5, Figure 7 and 8), although it was not statistically significant in the CARMENES data. However, for the CARMENES data, the Ca II IRT 1, 2, and 3 indices (panels 10, 11, and 12) did show a significant period corresponding to the rotational period. On the other hand, the NaD2 index from both instruments did not provide retrieval of the rotational period.

As for the SPIRou indicators such as  $W_1$  and  $V_1$ , they are primarily modulated by a 1 year signal, indicating yet again the presence of telluric contamination in this target, which is even more apparent when examining the principal vectors in the  $V_{\text{tot}}$  space (panel 13). Once more, we observe a similar pattern where

	SPIRou	HARPS + HIRES	ALL	Priors	Feng et al. (2020)
<b>Model comparison</b>					
$BIC_{circular}$	630	342	1014	-	-
$BIC_{Keplerian}$	639	350	1024	-	-
<b>Instrumental Parameters</b>					
$\gamma_{SPIRou}$ (m s <sup>-1</sup> )	$-4301.6^{+0.4}_{-0.4}$	-	$-4301.6^{+0.4}_{-0.4}$	$\mathcal{U}(-4311.6, -4291.6)$	-
$\gamma_{CARMENES}$ (m s <sup>-1</sup> )	-	-	$0.29^{+0.51}_{-0.51}$	$\mathcal{U}(-9.71, 10.29)$	-
$\gamma_{HARPS}$ (m s <sup>-1</sup> )	-	$-0.25^{+0.66}_{-0.66}$	$0.04^{+0.72}_{-0.71}$	$\mathcal{U}(-9.96, 10.03)$	-
$\gamma_{HIRES}$ (m s <sup>-1</sup> )	-	$3.73^{+1.01}_{-1.00}$	$3.24^{+1.02}_{-1.01}$	$\mathcal{U}(-6.69, 13.3)$	-
$\sigma_{SPIRou}$ (m s <sup>-1</sup> )	$3.5^{+0.3}_{-0.3}$	-	$3.5^{+0.3}_{-0.3}$	$\mathcal{U}(10^{-5}, 10^5)$	-
$\sigma_{CARMENES}$ (m s <sup>-1</sup> )	-	-	$0.1^{+0.7}_{-0.1}$	$\mathcal{U}(10^{-5}, 10^5)$	-
$\sigma_{HARPS}$ (m s <sup>-1</sup> )	-	$3.6^{+0.6}_{-0.5}$	$4.0^{+0.9}_{-0.7}$	$\mathcal{U}(10^{-5}, 10^5)$	-
$\sigma_{HIRES}$ (m s <sup>-1</sup> )	-	$3.5^{+0.9}_{-0.7}$	$4.0^{+0.6}_{-0.5}$	$\mathcal{U}(10^{-5}, 10^5)$	-
<b>Orbital Parameters</b>					
Period (days)	$9.552^{+0.006}_{-0.006}$	$9.567^{+0.004}_{-0.004}$	$9.5545^{+0.0007}_{-0.0007}$	$\mathcal{U}(9, 11)$	$9.567 \pm 0.005$
$K$ (m s <sup>-1</sup> )	$4.29^{+0.60}_{-0.60}$	$6.76^{+0.81}_{-0.80}$	$4.98^{+0.43}_{-0.43}$	$\mathcal{U}(0, 20)$	$6.80 \pm 0.87$
$T_{SPIRou}$ (BJD-245000)	$9220.4^{+0.2}_{-0.2}$	-	-	$\mathcal{U}(9215, 9230)$	-
$T_{HARPS+HIRES}$ (BJD-245000)	-	$5284.0^{+0.2}_{-0.2}$	-	$\mathcal{U}(5285, 5295)$	-
$T_{ALL}$ (BJD-245000)	-	-	$8981.6^{+0.1}_{-0.1}$	$\mathcal{U}(8975, 8990)$	-
$e$	-	-	-	-	$0.10 \pm 0.7$
$\omega$ (deg)	-	-	-	-	$151 \pm 92$
<b>Derived Parameters</b>					
$M_p \sin i$ (M <sub>⊕</sub> )	$8.4^{+1.2}_{-1.2}$	$13.2^{+1.6}_{-1.6}$	$9.7^{+0.9}_{-0.9}$	-	$13.2 \pm 1.7$
$a$ (AU)	$0.068^{+0.001}_{-0.001}$	$0.068^{+0.001}_{-0.001}$	$0.068^{+0.001}_{-0.001}$	-	$0.068 \pm 0.001$
$T_{eq}$ (K)	$406^{+6}_{-6}$	$406^{+6}_{-6}$	$406^{+6}_{-6}$	-	-

**Table 2.** Orbital and derived planetary parameters from the analysis of Gl 480. The parameters derived from the Wapiti corrected RV data from SPIRou are compared to those from Feng et al. (2020) using HARPS and HIRES spectrographs. The parameters derived by combining all data sets including the few CARMENES data correspond to the ALL column. We also show for completeness the eccentricity  $e$  and argument of periastron  $\omega$  from the Keplerian fit conducted in Feng et al. (2020).

$W_1$  and  $V_1$  are anticorrelated ( $r = -0.88$ ), while there is no correlation between the time derivative of  $W_1$  and the RV ( $r = 0.04$ ). This suggests that for this target  $W_1$  also fails to be a proxy for stellar activity as in Donati et al. (2023a) due to telluric contamination.

#### 4.2.2. Telluric contamination of stellar lines

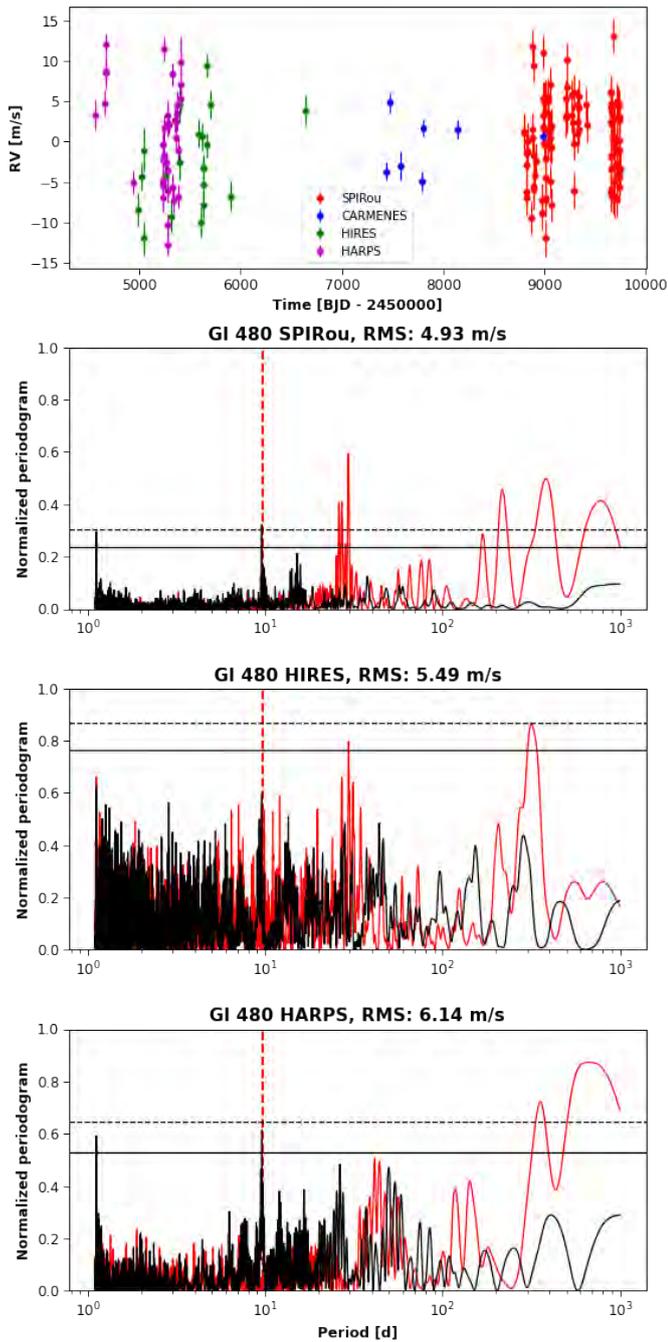
To ensure that our interpretation of the results is correct, we performed a similar analysis as we carried out for Gl 480. This involved calculating the  $z$  metric for Gl 382 using all 11 components used for the Wapiti correction which is displayed in Figure 9. Additionally, we computed the Pearson correlation between the tellurics and the derivative of the template for all lines. Finally, we computed  $r_{\oplus}$ , which represents the average Pearson correlation of all lines most influencing the principal vectors (with a  $z$  value above 5, ). This led to selecting 4,789 lines which constitutes 28.1% of the total number of lines.

As depicted in Figure 9, the correlation between  $r_{\oplus}$  and  $V_1$  for this target remains significant, although lower compared to

Gl 480, with a Pearson coefficient of  $-0.70$ . The correlation with  $W_1$  is weaker but still notable, with a value of 0.55. It is worth noting that the correlation plot indicates a deviation from a simple linear relationship, suggesting that the Pearson coefficient might not be the most appropriate measure as it assesses linear correlations between variables.

Nevertheless, the main conclusion remains that  $V_1$  can serve as a useful proxy for evaluating  $r_{\oplus}$ . The weaker correlation value could indicate that for this target, the tellurics are better corrected, resulting in reduced telluric contamination of the per-line dLWs which would explain the weaker correlation with  $W_1$ .

In summary, we believe that the increased statistical significance of the 10.4 d signal in our RV data could primarily be due to the effectiveness of the Wapiti method in correcting for contamination from telluric residuals. Initially, this signal was automatically attributed to stellar activity in (Sabotta et al. 2021). However, even after applying the Wapiti method, the persistence of this signal in the nIR led us to consider the possibility that it might be of planetary origin. Therefore, while acknowledging the need for caution when detecting a signal at the first

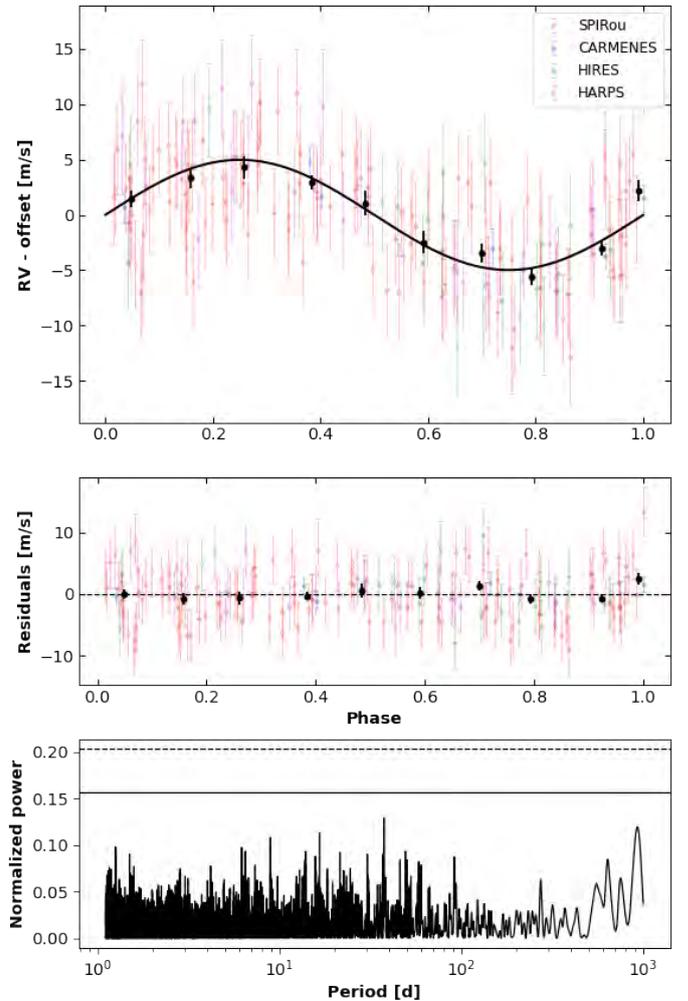


**Fig. 4.** RV time series from the available instruments and their respective periodogram. The red periodogram represents the window function. The significance levels of  $FAP = 10^{-3}$  and  $FAP = 10^{-5}$  are indicated by the solid and dashed horizontal lines, respectively.

harmonic of the star’s rotational period, we proceed with a characterization of this signal assuming a potential planetary origin.

#### 4.2.3. Planetary candidate orbiting GI382

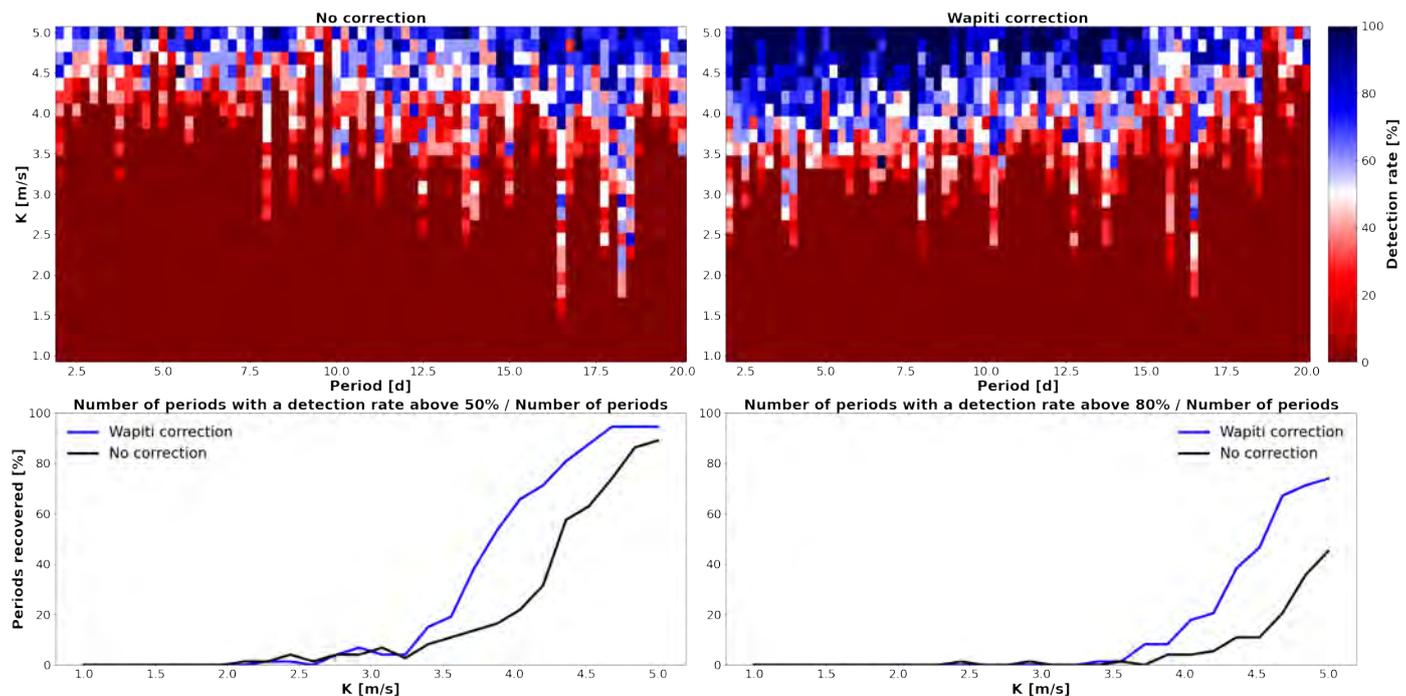
Initially, we conducted an analysis of the signals present in the additional datasets, and the periodograms corresponding to each instrument are presented in Figure 10. The 10.4 day signal is detected in the CARMENES data, in agreement with the results presented in Sabotta et al. (2021), where it was automatically identified as a harmonic of the rotational period of the star.



**Fig. 5.** Top panel: Phase-folding of the model on the RVs of GI480. Middle panel: Phase-folding of the residuals. Bottom panel: Periodogram of the residuals. The significance levels of  $FAP = 10^{-3}$  and  $FAP = 10^{-5}$  are indicated by the solid and dashed horizontal lines, respectively

Importantly, the 10.4 day signal, which was detected in the SPIRou and CARMENES optical datasets, was notably absent in both the HARPS and HIRES datasets. To investigate whether this absence was expected given the data, a straightforward test was conducted. The test involved injecting a circular signal, as defined in Equation 2, into the RV data by randomly drawing  $K$ ,  $T$ , and  $P$  from normal distributions which we chose to be  $\mathcal{N}(4, 1)$ ,  $\mathcal{N}(0, 2\pi)$ , and  $\mathcal{N}(10.4, 0.1)$ , respectively. This injection process was repeated 1,000 times, resulting in a periodogram for each injection. These periodograms were then averaged to create a mean periodogram, shown in Figure 11.

Remarkably, the mean periodogram revealed that the injected signal was not detected on average. Finally, to define the criterion for signal detection in the simulations, the injected signal was considered detected if a peak appeared in its periodogram within the range of 10.3 to 10.5 days with a FAP below  $10^{-3}$ . Based on this criterion, the success rate in detecting the injected signal for the two instruments was found to be 0% with HARPS and 18% with HIRES. If a less cautious approach were taken by increasing the FAP threshold to  $10^{-4}$ , the detection rate rises to 0.4% with HARPS and 24% with HIRES.



**Fig. 6.** Top panels show the detection rate of the injected signals without (left) and with Wapiti correction (right) using the SPIRou LBL RVs of Gl 480. The bottom panels show the number of periods recovered, meaning the number of periods with a detection rate above 50% (left) or 80% (right) with and without Wapiti correction.

In conclusion, the lack of detection of the 10.4 day signal in the HARPS and HIRES RV data appears to be consistent with their sampling and/or dispersion properties and we thus decided to proceed with the characterization of this signal.

To characterize the signal, assuming it originates from a planet, we employed a circular and Keplerian model, similar to the approach used in the previous section for Gl 480. These models were compared using the BIC obtained from likelihood maximization. Initially, we examined subsets of data from SPIRou and CARMENES individually to ensure compatibility of the signal between the instruments. Subsequently, we used the datasets consisting of SPIRou and CARMENES data as well as the one using all data to achieve the final characterization of the signal.

Due to the potential stellar activity origin of the 10.4 day signal, we adopted a more conventional approach to account for stellar activity involving the use of a quasi-periodic Gaussian process (GP) with the following form:

$$k(\tau) = A \exp\left(-\frac{\tau^2}{2l^2} - \Gamma \sin^2\left(\frac{\pi}{P}\tau\right)\right). \quad (4)$$

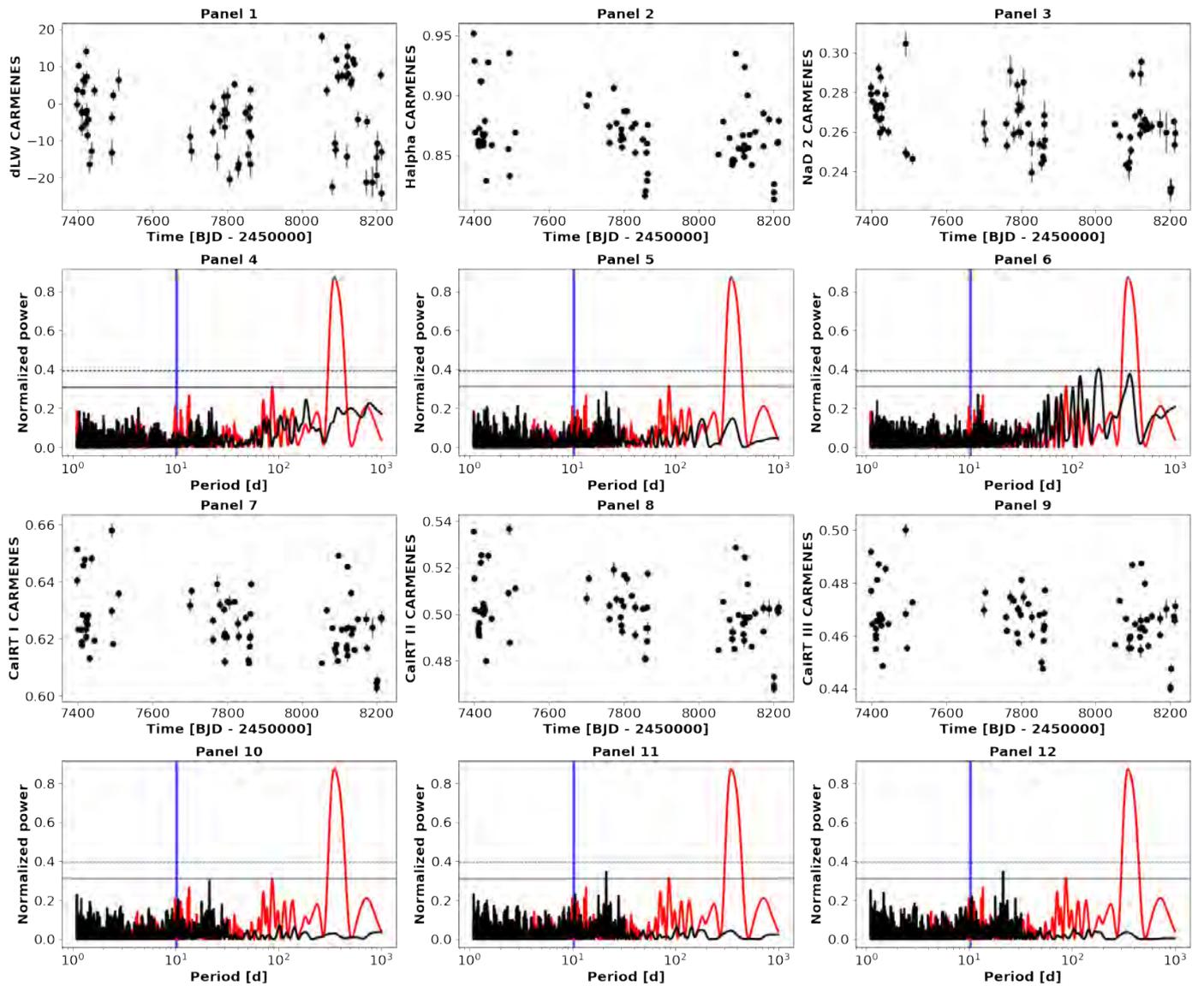
The GP was computed using the george package (Ambikasaran et al. 2015), and the GP was initialized using the known 21 day rotational period (Fouqué et al. 2023).

We considered a circular orbit planet plus a quasi-periodic GP model and a GP-only model in addition to the circular and Keplerian model for the SPIRou and CARMENES optical datasets. The model comparison results are shown in Table 3. For the SPIRou dataset and similarly to Gl 480, the circular model (BIC = 632) is strongly favored over the Keplerian model (BIC = 642). Moreover, the inclusion of a GP (named  $GP_{nIR}$ ) did not lead to an improved fit with (BIC = 650) or without (BIC = 662) the circular planet. This implies that our data in the nIRA is minimally affected by stellar activity, as anticipated, or

at least that the 10.4 days signal present in this target has a higher amplitude compared to activity in the nIR.

On the other hand, for the CARMENES dataset, the circular + GP model is preferred (BIC = 450), indicating that for optical data, the influence of stellar activity needs to be taken into account through the use of a GP (named  $GP_{optical}$ ). Notably, the GP-only model (BIC = 457) is not favored either, implying that even in the CARMENES data alone, the potential planetary origin of the 10.4 day signal was already distinguishable.

From these results, we performed a joint nIR and optical RV analysis by using a model consisting of a circular planet and a quasi-periodic GP fitted on optical data only. This model is indeed favored compared to a simple circular model, as outlined in Table 3. The resulting orbital and derived planetary parameters were retrieved using 10,000 samples with a burn-in of 1,000 samples and 100 walkers. We estimated the planet's minimum mass  $M_p \sin i$  by using the star's mass and its uncertainty obtained from Cristofari et al. (2022). We can see that the planetary mass between SPIRou and CARMENES optical data alone are compatible within  $1\sigma$ , even if somewhat smaller in the CARMENES data. This could be attributed to the fact that we used a GP in the CARMENES data, which can tend to underestimate the retrieved semi-amplitude and consequently the derived planetary mass. The compatibility of the orbital parameters using two different datasets in different wavelength domains adds to the evidence that this signal could be of planetary origin instead of stellar origin. All the more so as if the signal was of stellar origin we would expect it to have a much larger semi-amplitude in the optical compared to the nIR instead of being of similar magnitude. Thus, we interpreted this signal as a planet and will refer to it as Gl 382 b, and the joint RV fit using all instruments leads to a final minimum planetary mass of  $M_p \sin i = 8.71^{+0.85}_{-0.83} M_{\oplus}$ . The corner plot from the MCMC computed over all datasets is available in Appendix B.2.



**Fig. 7.** Various activity indicators of Gl 382 from CARMENES. The red periodogram represents the window function, the blue band highlights the 10.4 d period. The significance levels of  $FAP = 10^{-3}$  and  $FAP = 10^{-5}$  are indicated by the solid and dashed horizontal lines, respectively. **Panel 1 and 4:** dLW time series from CARMENES and its periodogram. **Panel 2 and 5:**  $H\alpha$  index time series from CARMENES and its periodogram. **Panel 3 and 6:** NaD II index time series from CARMENES and its periodogram. **Panel 7 and 10:** CaIRT I index time series from CARMENES and its periodogram. **Panel 8 and 11:** CaIRT II index time series from CARMENES and its periodogram. **Panel 9 and 12:** CaIRT III index time series from CARMENES and its periodogram.

The final fit on the dataset using all instruments is presented in Figure 12, we provide the phase-folded representation of the data and the residuals. No other statistically significant signals were found in the periodogram of the residuals once the model was fitted.

#### 4.2.4. Detection sensitivity of the Wapiti correction

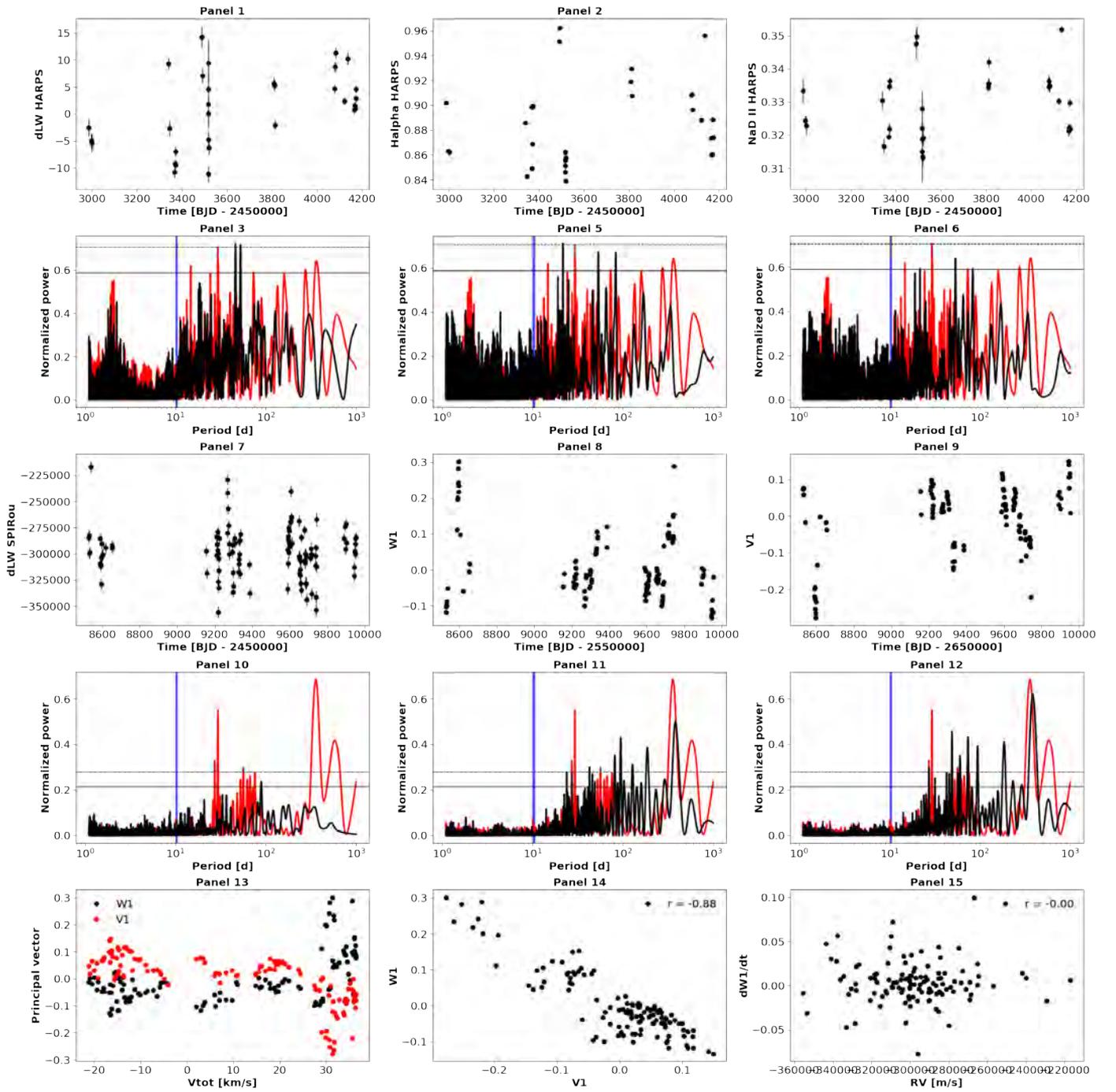
To assess the detection sensitivity of the Wapiti correction when applied to the RV data of Gl 382, we employed the same injection recovery method as in the case of Gl 480 and the results are presented in Figure 13.

The results for Gl 382 are even more favorable compared to Gl 480. When applying the Wapiti correction, we can detect signals with even lower semi-amplitudes. Specifically, under the conservative criterion, we start detecting 3 m.s<sup>-1</sup> semi-

amplitude signals, with a detection rate exceeding 50% at 3.56 m.s<sup>-1</sup>. In contrast, without correction, only 3% of periods are recovered. When applying the non-conservative criterion, the Wapiti correction leads to nearly 100% recovery of the periods, while it remains below 50% in the case of no correction. These findings underscore the promising capability of the Wapiti method, although it is important to note that, with the current data, signals below 3 m.s<sup>-1</sup> remain undetectable.

## 5. Summary, discussion and prospects

The work presented in this paper can be summarized as follows: Firstly, we conducted an analysis of all available stellar activity indicators from various instruments for each target. This step is crucial to understand the activity signals present in the RV data of our target. This understanding is particularly important when

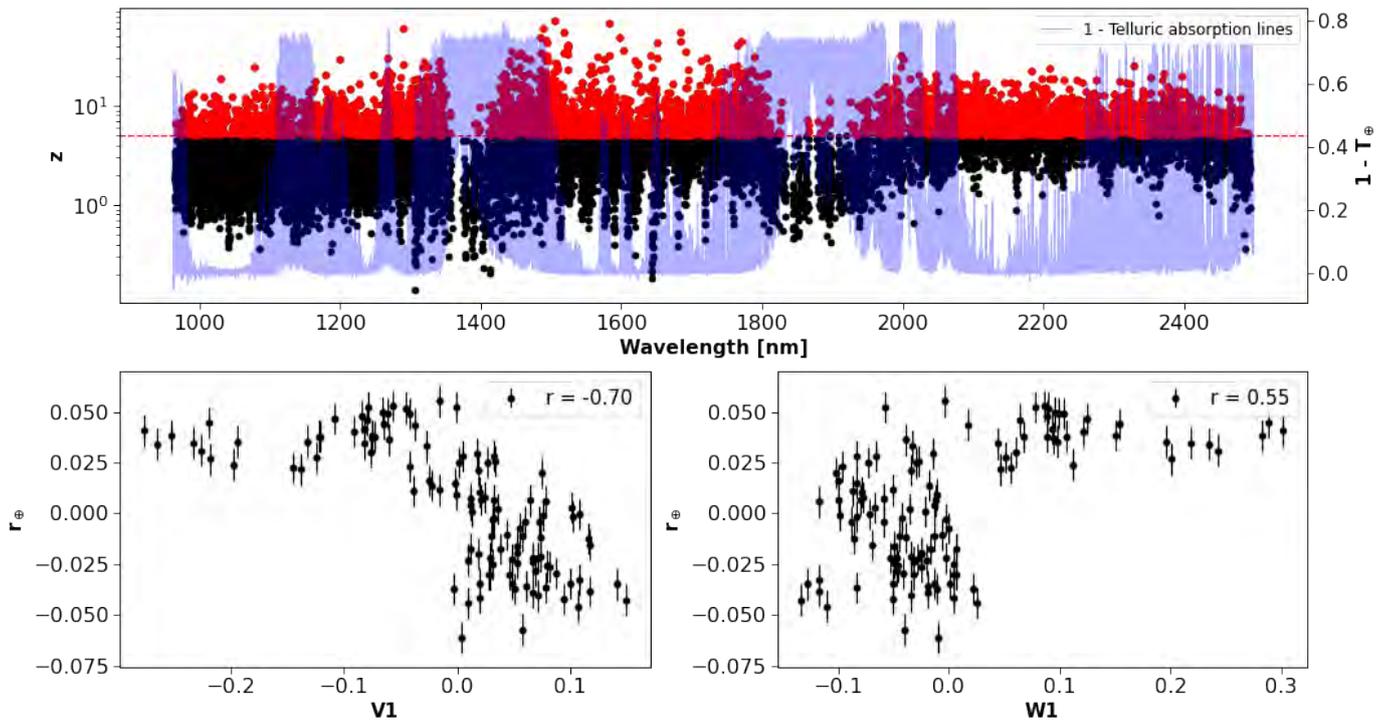


**Fig. 8.** Various activity indicators of G1382 from HARPS and SPIRou. The red periodogram represents the window function, the blue band highlights the 10.4 d period. The significance levels of  $\text{FAP} = 10^{-3}$  and  $\text{FAP} = 10^{-5}$  are indicated by the solid and dashed horizontal lines, respectively. **Panel 1 and 4:** dLW time series from HARPS and its periodogram. **Panel 2 and 5:** H $\alpha$  index time series from HARPS and its periodogram. **Panel 3 and 6:** NaD II index time series from HARPS and its periodogram. **Panel 7 and 10:** dLW time series from SPIRou and its periodogram. **Panel 8 and 11:** W1 and its periodogram. **Panel 9 and 12:** V1 and its respective periodogram. The green bands represent epochs when  $V_{\text{tot}} \geq 20 \text{ km.s}^{-1}$ . **Panel 13:** V1 and W1 in the  $V_{\text{tot}}$  space. **Panel 14:** Correlation plot between V1 and W1. **Panel 15:** Correlation plot between the first time derivative of W1 and the RV.

applying the *Wapiti* correction, as it provides a clearer comprehension of our data before and after correction.

The LBL framework provided valuable insights into the SPIRou data analysis. We observed that by applying the wPCA decomposition within this framework, we could use additional data products such as the principal vectors, and first principal vector of the per-line dLWs,  $W_1$ . Although the principal vectors may not be directly correlated with stellar activity (as opposed

to the findings on the young active M dwarf AU Mic; Donati et al. (2023a), which probably comes from lower level of stellar activity, they can still offer valuable information regarding the systematics present in the data.



**Fig. 9.** The top panel displays the  $z$  values for all wavelengths of Gl 382 using 11 components, with the lines having values above 5 marked in red, defining them as "affected lines." Moving to the bottom left panel, it illustrates the correlation between  $V_1$  and  $r_{\oplus}$ , represented as the mean Pearson correlation between the tellurics and the derivative of the template for all affected lines. Finally, the bottom right panel presents the correlation plot between  $W_1$  and  $r_{\oplus}$ .

### 5.1. Using the principal vectors to investigate the impact of tellurics on LBL data

As demonstrated in our study, a significant fraction of the behavior exhibited by the principal vectors  $V_1$  and  $W_1$  obtained from the wPCA decomposition of per-line RVs and dLWs can be directly attributed to the influence of telluric effects. This connection arises from the linear correlation analysis between telluric lines and the derivative of the star template. The computation of the Pearson coefficient of correlation can be readily performed using the available data products provided by the APERO DRS, including the 2D telluric spectra and the template used for the computation of LBL RVs.

The mean Pearson coefficient  $r_{\oplus}$  was computed by selecting lines with a  $z$  value above 5, revealing a significant correlation with  $V_1$  for both Gl 480 and Gl 382. This indicates that the behavior of  $V_1$  is predominantly influenced by the telluric contamination in those specific lines. As a result,  $V_1$  can serve as a proxy for estimating the quality of the telluric correction for these stars.

Additionally, we could find a comparable correlation with  $W_1$  indicating that the calculation of per-line dLWs is also affected by the presence of telluric residuals. This contamination introduces inaccuracies in the computation, consequently limiting its utility as a reliable proxy for stellar activity. To address this issue, further research is warranted to develop correction methods for mitigating these telluric-induced effects, with the possibility of using  $V_1$  as a potential solution.

### 5.2. Joint nIR and optical RV analysis in the Wapiti framework

Based on our comprehensive analysis, we concluded that in the Wapiti corrected RV time series of our targets, the stellar activ-

ity could be neglected. However, owing to a higher activity level, we still needed to take activity into account for data obtained from other optical instruments in the case of Gl 382. To address this, we opted to employ a quasi-periodic GP specifically for those optical data to account for the stellar activity. The choice of using a single GP for all the optical data could be a subject of discussion. Conceptually, it might be more appropriate to employ a different GP amplitude for each instrument, considering their distinct wavelength domains. However, in our analysis, the chosen approach proved to be sufficient, and the retrieved hyperparameters remained consistent across various combinations of data sets.

It is important to emphasize that the Wapiti correction method involves a trade-off. Although it effectively corrects for systematic errors in the RV data, it can introduce a challenge in precisely determining the desired signal's properties. This challenge may lead to an underestimation of the semi-amplitude. However, this problem can be resolved by using the reconstructed model of the signal that is being fitted as demonstrated in Cortes-Zuleta et al. (2023) through an injection-recovery test.

Moreover, it could be beneficial to consider applying the LBL algorithm to other instruments. This approach might not only enhance the accuracy of RV measurements but also for investigating how well the Wapiti algorithm performs in mitigating optical stellar activity. Given that quasi-periodic GPs can occasionally provide inaccurate estimates of the semi-amplitude, and there exists a current lack of tools designed to rectify this concern, the compensation offered by the Wapiti method could play a crucial role in minimizing the impact of stellar activity while still achieving precise characterization of planetary systems.

	SPIRou	CARMENES	SPIRou + CARMENES	ALL	Priors
<b>Model comparison</b>					
$BIC_{circular}$	632	471	1106	1936	-
$BIC_{Keplerian}$	642	480	1115	1945	-
$BIC_{GP_{nIR}}$	662	-	-	-	-
$BIC_{circular+GP_{nIR}}$	650	-	-	-	-
$BIC_{GP_{optical}}$	-	457	-	-	-
$BIC_{circular+GP_{optical}}$	-	450	1085	1829	-
<b>Instrumental Parameters</b>					
$\gamma_{SPIRou}$ (m s <sup>-1</sup> )	$7884.8^{+0.3}_{-0.3}$	-	$7884.8^{+0.3}_{-0.3}$	$7884.8^{+0.3}_{-0.3}$	$\mathcal{U}(7874.7, 7885.7)$
$\gamma_{CARMENES}$ (m s <sup>-1</sup> )	-	$-0.02^{+1.03}_{-0.97}$	$-0.02^{+0.98}_{-0.93}$	$0.05^{+1.2}_{-1.2}$	$\mathcal{U}(-9.88, 10.12)$
$\gamma_{HIRES}$ (m s <sup>-1</sup> )	-	-	-	$-1.38^{+0.81}_{-0.82}$	$\mathcal{U}(-9.98, 10.03)$
$\gamma_{HARPS}$ (m s <sup>-1</sup> )	-	-	-	$-0.64^{+1.38}_{-1.39}$	$\mathcal{U}(-9.71, 10.29)$
$\sigma_{SPIRou}$ (m s <sup>-1</sup> )	$2.6^{+0.3}_{-0.3}$	-	$2.7^{+0.3}_{-0.3}$	$2.7^{+0.3}_{-0.3}$	$\mathcal{U}(10^{-4}, 10^4)$
$\sigma_{CARMENES}$ (m s <sup>-1</sup> )	-	$0.5^{+1.2}_{-0.4}$	$0.5^{+1.1}_{-0.4}$	$0.3^{+1.0}_{-0.3}$	$\mathcal{U}(10^{-4}, 10^4)$
$\sigma_{HARPS}$ (m s <sup>-1</sup> )	-	-	-	$0.2^{+0.5}_{-0.2}$	$\mathcal{U}(10^{-4}, 10^4)$
$\sigma_{HIRES}$ (m s <sup>-1</sup> )	-	-	-	$0.8^{+1.2}_{-0.7}$	$\mathcal{U}(10^{-4}, 10^4)$
<b>GP Parameters</b>					
$A$ (m s <sup>-1</sup> )	-	$4.3^{+0.7}_{-0.6}$	$4.3^{+0.7}_{-0.5}$	$5.3^{+0.4}_{-0.4}$	$\mathcal{U}(10^{-4}, 10^4)$
$l$ (days)	-	$87^{+36}_{-26}$	$80^{+27}_{-24}$	$111^{+26}_{-23}$	$\mathcal{U}(10, 10^4)$
$\Gamma$	-	$14^{+8}_{-8}$	$15^{+7}_{-7}$	$13^{+4}_{-4}$	$\mathcal{U}(10^{-4}, 10^4)$
$P_{rot}$ (days)	-	$21.6^{+0.2}_{-0.1}$	$21.6^{+0.2}_{-0.1}$	$21.6^{+0.2}_{-0.1}$	$\mathcal{U}(18, 25)$
<b>Orbital Parameters</b>					
Period (days)	$10.393^{+0.004}_{-0.004}$	$10.38^{+0.02}_{-0.02}$	$10.383^{+0.002}_{-0.002}$	$10.3843^{+0.0008}_{-0.0007}$	$\mathcal{U}(10, 12)$
$K$ (m s <sup>-1</sup> )	$4.70^{+0.48}_{-0.48}$	$3.7^{+0.8}_{-0.8}$	$4.35^{+0.42}_{-0.42}$	$4.00^{+0.37}_{-0.36}$	$\mathcal{U}(0, 20)$
$T_{SPIRou}$ (BJD-245000)	$9594.8^{+0.2}_{-0.2}$	-	-	-	$\mathcal{U}(9590, 9600)$
$T_{CARMENES}$ (BJD-245000)	-	$7829.8^{+0.4}_{-0.4}$	-	-	$\mathcal{U}(7825, 7835)$
$T_{SPIRou+CARMENES}$ (BJD-245000)	-	-	$9210.5^{+0.1}_{-0.2}$	-	$\mathcal{U}(9205, 9215)$
$T_{ALL}$ (BJD-245000)	-	-	-	$7798.2^{+0.2}_{-0.1}$	$\mathcal{U}(7795, 7805)$
<b>Derived Parameters</b>					
$M_p \sin i$ ( $M_{\oplus}$ )	$10.2^{+1.1}_{-1.1}$	$8.0^{+1.8}_{-1.8}$	$9.47^{+0.95}_{-0.94}$	$8.71^{+0.85}_{-0.83}$	-
$a$ (AU)	$0.074^{+0.001}_{-0.001}$	$0.074^{+0.001}_{-0.001}$	$0.074^{+0.001}_{-0.001}$	$0.074^{+0.001}_{-0.001}$	-
$T_{eq}$ (K)	$428^{+6}_{-6}$	$428^{+6}_{-6}$	$428^{+6}_{-6}$	$428^{+6}_{-6}$	-

**Table 3.** Orbital and derived planetary parameters from the analysis of Gl 382. The models fit to the data are the circular model for SPIRou and circular + GP<sub>optical</sub> model for other combination of datasets.

In summary, our joint nIR and optical RV analysis has significantly improved the characterization of the orbital parameters for Gl 480, revealing a minimum planetary mass of  $9.7^{+0.9}_{-0.9} M_{\oplus}$  and an orbital period of  $9.5545^{+0.0007}_{-0.0007}$  d. Additionally, our analysis strongly supported the existence of an exoplanet in the Gl 382 system, which shares similar characteristics with Gl 480 b with a minimum mass of  $8.71^{+0.85}_{-0.83} M_{\oplus}$  and an orbital period of  $10.3843^{+0.0008}_{-0.0007}$  d. These two planets likely belong to the category of warm sub-Neptunes, with characteristics making them

relatively common in this category. However, what makes them particularly interesting is their proximity to bright and nearby M dwarf stars, which makes them ideal candidates for follow-up observations.

### 5.3. Conclusion and future work

By employing the SPIRou LBL RV measurements and incorporating RV data from other available instruments within the community, we conducted a joint analysis of the RV in both the nIR

and optical domain for two targets: Gl 480 and Gl 382, as part of the SLS large program. This analysis resulted in the first detection of Gl 480 b in the nIR, allowing for a more precise determination of its orbital parameters. Concerning Gl 382, our study led us to consider the possibility that the previously identified 10.4 day signal, as reported in Sabotta et al. (2021), may originate from a planet rather than stellar activity, despite its proximity to the first harmonic of the star's rotational period. Notably, the nIR data from SPIRou reveal that the influence of stellar activity is not prominent in the RV and dLW time series. Furthermore, the amplitude of the activity GP signal is lower in comparison to the CARMENES dataset. Additionally, the obtained semi-amplitude remains consistent between the optical and nIR data. As a result, within this paper, we concluded that we had the first evidence indicating the planetary origin of this signal and we conducted a characterization of its orbital parameters.

This research also facilitated further use of the *Wapiti* algorithm and demonstrated that it is a valuable tool in the LBL framework. As previously announced in Ould-Elhkim et al. (2023), we conducted a more detailed analysis of the data products generated by *Wapiti*, namely the principal vectors and scores obtained from the wPCA decomposition of the LBL RVs. The scores allowed for a rapid assessment of the spectral lines that exerted the most significant influence on the principal vectors. Through our investigation, a clear correlation between  $V_1$  and the telluric contamination in those lines was established. It was observed that this contamination also extended to the second derivative of the template, resulting in a computation of dLW contaminated by tellurics, thereby hindering its potential as a reliable stellar activity indicator for relatively quiet stars.

Nonetheless, the behavior of the principal vectors is probably influenced by systematics originating from other sources. We have previously raised this point in our earlier work, highlighting that wPCA, being a linear method, is unable to differentiate between nonlinear relationships among data features. Consequently, unraveling the non-linear interplay between systematics would be challenging using wPCA alone. Alternatively, techniques such as autoencoders (Kingma & Welling (2013)), which are renowned for capturing non-linear relationships compared to PCA, could be employed in future investigations. However, incorporating autoencoders would introduce complex considerations, including the selection of an appropriate neural network, handling uncertainties, and managing NaN values.

Future endeavors will involve the careful application of the *Wapiti* algorithm in simulations aimed at isolating and mitigating the influence of telluric absorption lines on LBL RVs data. This approach will enable a thorough exploration of the corrected data, allowing us to gain valuable insights into the content captured by the principal vectors and their connection to telluric contamination.

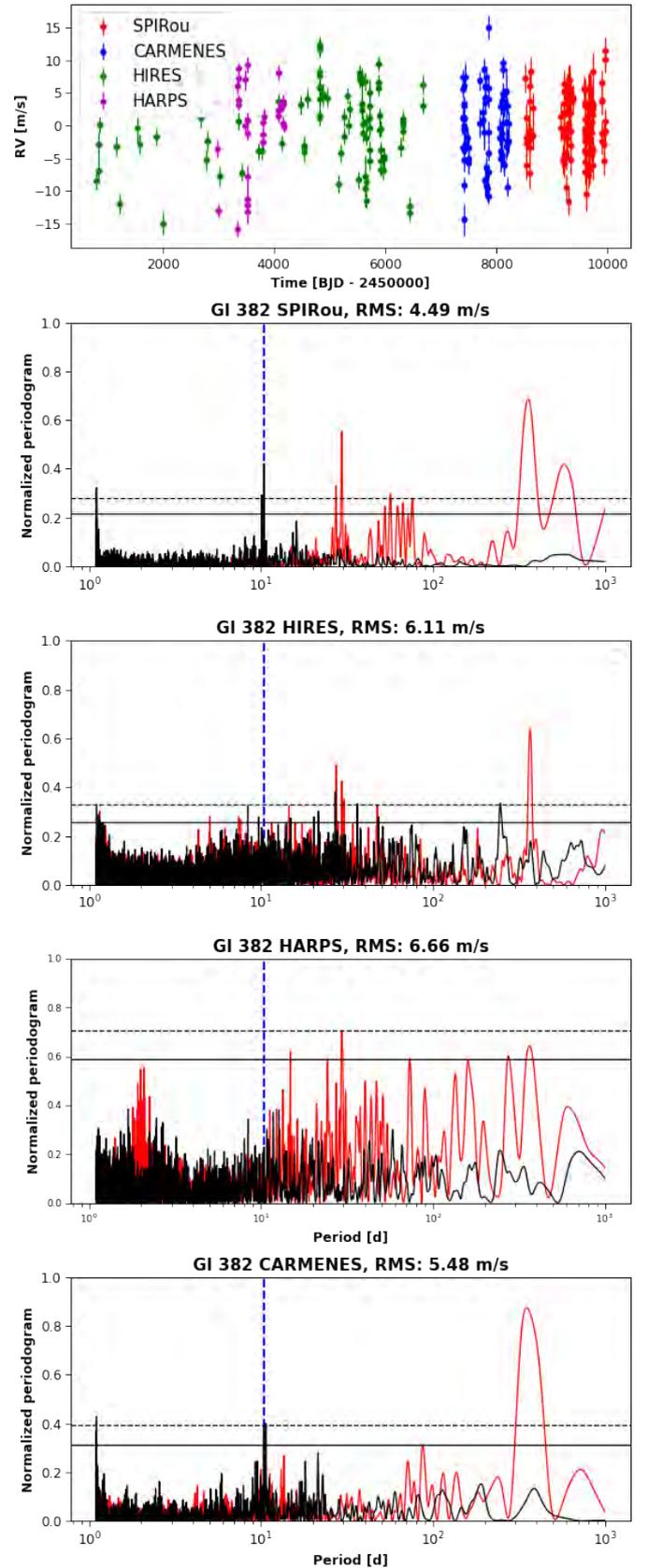
*Acknowledgements.* JFD and CM acknowledge funding from the European Research Council (ERC) under the H2020 research & innovation program (grant agreements 740651 NewWorlds). X.D. and A.C. acknowledge funding by the French National Research Agency in the framework of the Investissements d'Avenir program (ANR-15-IDEX-02), through the funding of the "Origin of Life" project of the Université Grenoble Alpes. E.M. acknowledges funding from FAPEMIG under project number APQ-02493-22 and research productivity grant number 309829/2022-4 awarded by the CNPq, Brazil. This paper makes use of data from the first public release of the WASP data (Butters et al. 2010) as provided by the WASP consortium and services at the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program. This research made use of the following software tools: NumPy; a fundamental package for scientific computing with Python. pandas; a library providing easy-to-use data structures and data analysis tools. Astropy; a community-developed core Python package for Astronomy. SciPy; a library

for scientific computing and technical computing. RadVel; a python package for modeling radial velocity data. matplotlib; a plotting library for the Python programming language. tqdm; a library for creating progress bars in the command line. seaborn; a data visualization library based on matplotlib. wPCA; a python package for weighted principal component analysis. emcee; a python package for MCMC sampling. PyAstronomy; a collection of astronomical tools for Python. celerite; a fast Gaussian process library in Python. george; a module for Gaussian process regression in Python.

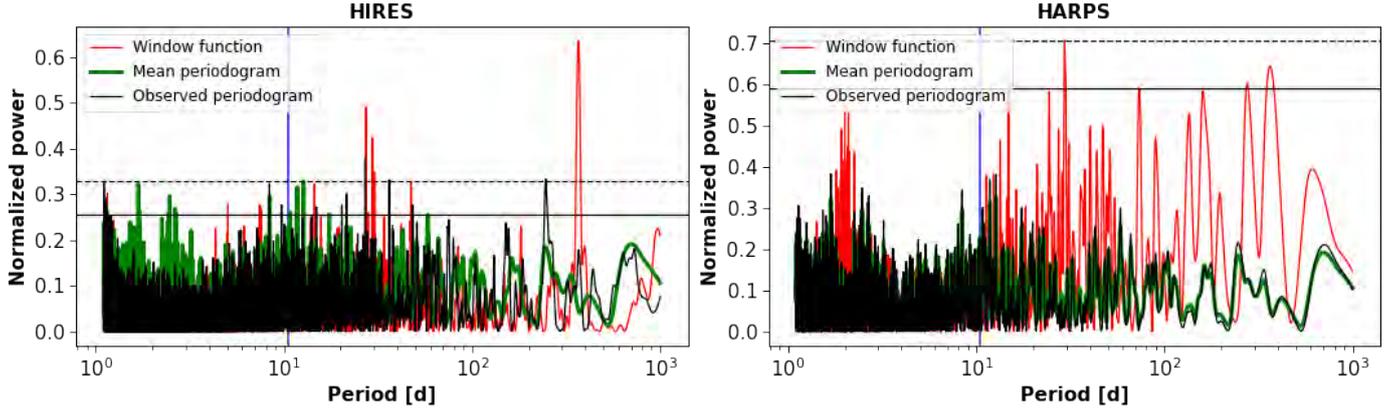
## References

- Almenara, J. M., Bonfils, X., Bryant, E. M., et al. 2023, arXiv e-prints, arXiv:2308.01454
- Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O'Neil, M. 2015, IEEE Transactions on Pattern Analysis and Machine Intelligence, 38, 252
- Anglada-Escudé, G., Amado, P. J., Barnes, J., et al. 2016, Nature, 536, 437
- Artigau, É., Cadieux, C., Cook, N. J., et al. 2022, AJ, 164, 84
- Artigau, É., Saint-Antoine, J., Lévesque, P.-L., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10709, High Energy, Optical, and Infrared Detectors for Astronomy VIII, ed. A. D. Holland & J. Beletic, 107091P
- Bluev, R. V. 2008, MNRAS, 385, 1279
- Baranne, A., Queloz, D., Mayor, M., et al. 1996, A&AS, 119, 373
- Baroch, D., Morales, J. C., Ribas, I., et al. 2020, A&A, 641, A69
- Bauer, F. F., Zechmeister, M., Kaminski, A., et al. 2020, A&A, 640, A50
- Bertaux, J.-L., Lallement, R., Ferron, S., & Boonne, C. 2014, in 13th International HITRAN Conference, 8
- Bonfils, X., Delfosse, X., Udry, S., et al. 2013, Astronomy and Astrophysics, 549, A109
- Bonfils, X., Delfosse, X., Udry, S., et al. 2013, A&A, 549, A109
- Bonfils, X., Mayor, M., Delfosse, X., et al. 2007, A&A, 474, 293
- Bouchy, F., Pepe, F., & Queloz, D. 2001, A&A, 374, 733
- Cadieux, C., Doyon, R., Plotnykov, M., et al. 2022, AJ, 164, 96
- Carleo, I., Sanna, N., Gratton, R., et al. 2016, Experimental Astronomy, 41, 351
- Carmona, A., Delfosse, X., Bellotti, S., et al. 2023, Near-IR and optical radial velocities of the active M dwarf star Gl 388 (AD Leo) with SPIRou at CFHT and SOPHIE at OHP: A 2.23 day rotation period and no evidence for a co-rotating planet
- Challita, Z., Reshetov, V., Baratchart, S., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10702, Ground-based and Airborne Instrumentation for Astronomy VII, ed. C. J. Evans, L. Simard, & H. Takami, 1070262
- Cook, N., Hobson, M., Bouchy, F., & Martioli, E. 2022, APERO: A Pipeline to Reduce Observations, Astrophysics Source Code Library, record ascl:2211.019
- Cortes-Zuleta, P., Boisse, I., Klein, B., et al. 2023, arXiv e-prints, arXiv:2301.10614
- Cretignier, M., Dumusque, X., Allart, R., Pepe, F., & Lovis, C. 2020, Astronomy and Astrophysics, 633, A76
- Cristofari, P. I., Donati, J. F., Masseron, T., et al. 2022, MNRAS, 516, 3802
- da Silva, J. G., Santos, N. C., Bonfils, X., et al. 2012, Astronomy and Astrophysics, 541, A9
- Delchambre, L. 2015, MNRAS, 446, 3545
- Donati, J. F., Cristofari, P. I., Finocci, B., et al. 2023a, MNRAS[arXiv:2304.09642]
- Donati, J. F., Kouach, D., Moutou, C., et al. 2020, MNRAS, 498, 5684
- Donati, J. F., Lehmann, L. T., Cristofari, P. I., et al. 2023b, arXiv e-prints, arXiv:2307.14190
- Dressing, C. D. & Charbonneau, D. 2015, ApJ, 807, 45
- Dumusque, X. 2018, Astronomy and Astrophysics, 620, A47
- Faria, J. P., Mascareño, A. S., Figueira, P., et al. 2022, Astronomy and Astrophysics, 658, A115
- Feng, F., Shectman, S. A., Clement, M. S., et al. 2020, ApJS, 250, 29
- Foreman-Mackey, D., Conley, A., Meierjürgen Farr, W., et al. 2013, emcee: The MCMC Hammer, Astrophysics Source Code Library, record ascl:1303.002
- Fouqué, P., Martioli, E., Donati, J. F., et al. 2023, A&A, 672, A52
- Fulton, B. J., Petigura, E. A., Blunt, S., & Sinukoff, E. 2018, PASP, 130, 044504 Gaia Collaboration. 2020, VizieR Online Data Catalog, I/350
- Gaidos, E., Mann, A. W., Kraus, A. L., & Ireland, M. 2016, MNRAS, 457, 2877
- Hébrard, É. M., Donati, J. F., Delfosse, X., et al. 2016, MNRAS, 461, 1465
- Henry, T. J., Jao, W.-C., Subasavage, J. P., et al. 2006, AJ, 132, 2360
- Hirano, T., Kuzuhara, M., Kotani, T., et al. 2020, PASJ, 72, 93
- Hobson, M. J., Bouchy, F., Cook, N. J., et al. 2021, A&A, 648, A48
- Hsu, D. C., Ford, E. B., & Terrien, R. 2020, MNRAS, 498, 2249
- Kanodia, S. & Wright, J. 2018, Research Notes of the American Astronomical Society, 2, 4

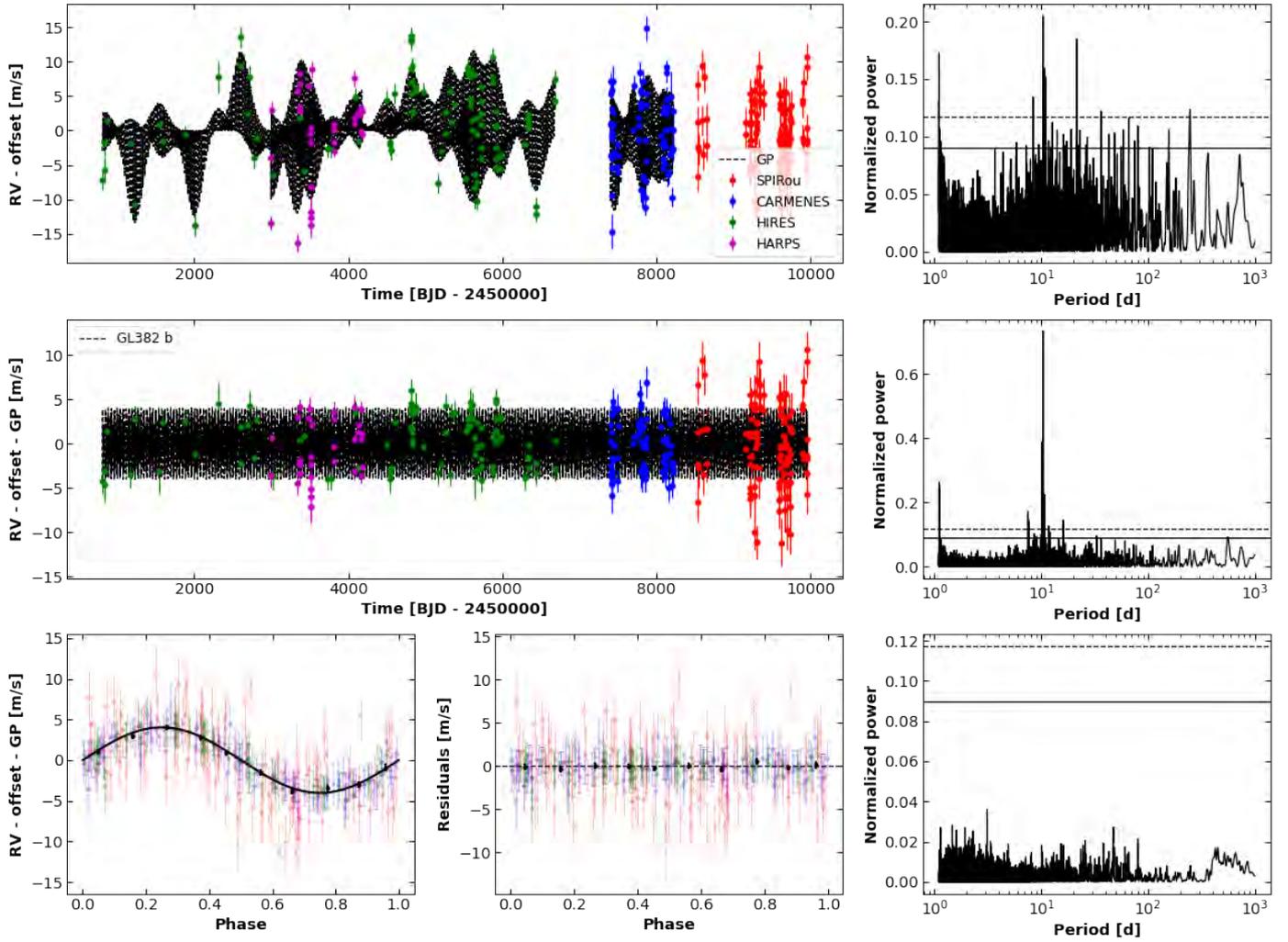
- Kingma, D. P. & Welling, M. 2013, arXiv e-prints, arXiv:1312.6114
- Kirkpatrick, J. D., Henry, T. J., & McCarthy, Donald W., J. 1991, *ApJS*, 77, 417
- Lépine, S., Hilton, E. J., Mann, A. W., et al. 2013, *AJ*, 145, 102
- Martioli, E., Hébrard, G., Fouqué, P., et al. 2022, *A&A*, 660, A86
- Mascareño, A. S., González-Álvarez, E., Osorio, M. R. Z., et al. 2023, *Astronomy and Astrophysics*, 670, A5
- Mayor, M., Pepe, F., Queloz, D., et al. 2003, *The Messenger*, 114, 20
- Mayor, M. & Queloz, D. 1995, *Nature*, 378, 355
- Mégevand, D., Zerbi, F. M., Di Marcantonio, P., et al. 2014, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 9147, *Ground-based and Airborne Instrumentation for Astronomy V*, ed. S. K. Ramsay, I. S. McLean, & H. Takami, 91471H
- Metcalfe, A. J., Anderson, T. B., Bender, C. F., et al. 2019, *Optica*
- Mignon, L., Meunier, N., Delfosse, X., et al. 2023, arXiv e-prints, arXiv:2303.03998
- Moutou, C., Delfosse, X., Petit, A. C., et al. 2023, arXiv e-prints, arXiv:2307.11569
- Newton, E. R., Irwin, J., Charbonneau, D., Berta-Thompson, Z. K., & Dittmann, J. A. 2016, *ApJ*, 821, L19
- Ould-Elhkim, M., Moutou, C., Donati, J.-F., et al. 2023, arXiv e-prints, arXiv:2305.02123
- 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
- Reiners, A., Shulyak, D., Anglada-Escudé, G., et al. 2013, *A&A*, 552, A103
- Reiners, A. & Zechmeister, M. 2020, *ApJS*, 247, 11
- Ribas, I., Reiners, A., Zechmeister, M., et al. 2023, *Astronomy and Astrophysics*, 670, A139
- Sabotta, S., Schlecker, M., Chaturvedi, P., et al. 2021, *A&A*, 653, A114
- Schwarz, G. 1978, *Annals of Statistics*, 6, 461
- Tal-Or, L., Trifonov, T., Zucker, S., Mazeh, T., & Zechmeister, M. 2019, *MNRAS*, 484, L8
- The Astropy Collaboration. 2018, *astropy v3.1: a core python package for astronomy*, Zenodo
- Thibault, S., Brousseau, D., Poulin-Girard, A.-S., et al. 2022, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 12184, *Ground-based and Airborne Instrumentation for Astronomy IX*, ed. C. J. Evans, J. J. Bryant, & K. Motohara, 1218454
- Trifonov, T., Tal-Or, L., Zechmeister, M., et al. 2020, *A&A*, 636, A74
- Vieira, V. 2012, *Computational Ecology and Software*, 2, 103
- Zechmeister, M. & Kürster, M. 2009, *Astronomy and Astrophysics*, 496, 577
- Zechmeister, M., Reiners, A., Amado, P. J., et al. 2018a, *A&A*, 609, A12
- Zechmeister, M., Reiners, A., Amado, P. J., et al. 2018b, *A&A*, 609, A12



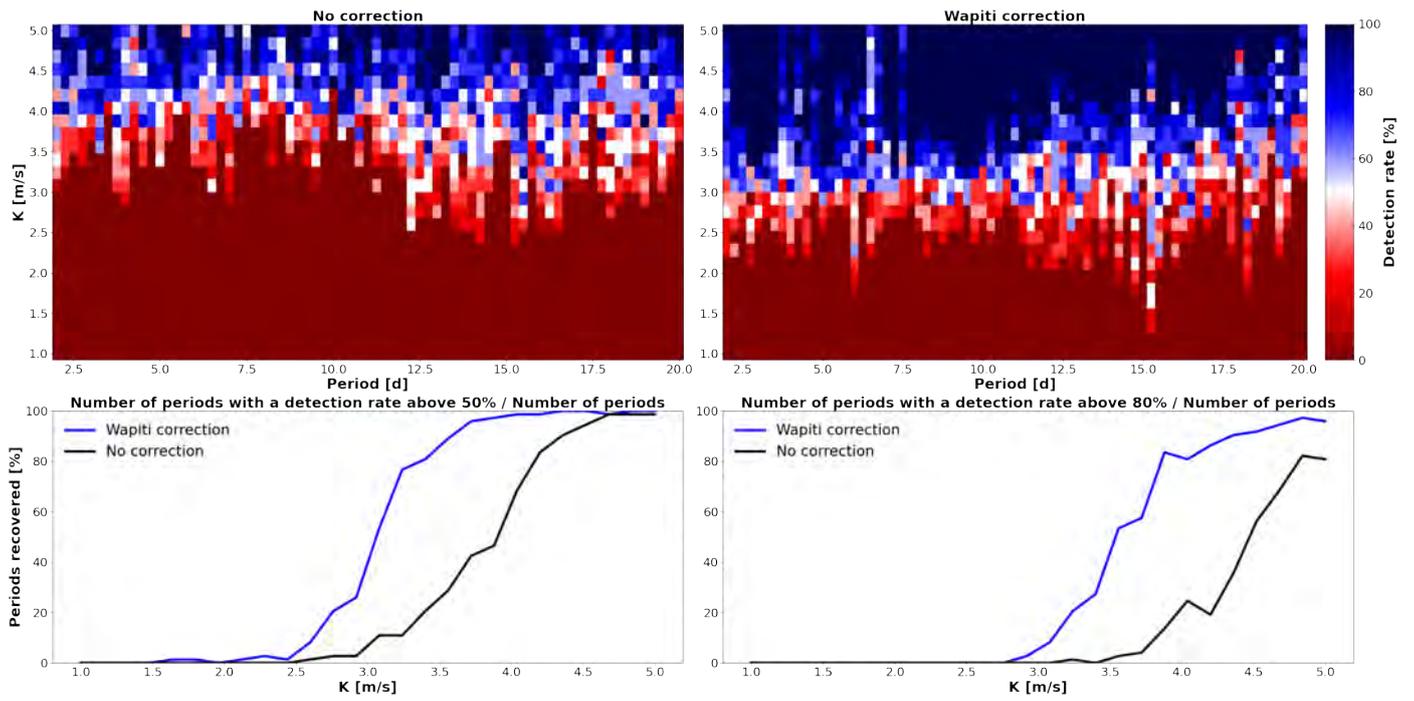
**Fig. 10.** RV time series from the available instruments and their respective periodogram. The red periodogram represents the window function. The significance levels of  $FAP = 10^{-3}$  and  $FAP = 10^{-5}$  are indicated by the solid and dashed horizontal lines, respectively

**Periodograms when injecting a circular signal**


**Fig. 11.** Results of the 1,000 simulations when injecting a 10.4 day signal in HIRES and HARPS RV data. The mean periodogram is in green, the periodogram from the observations in black and the window function in red. We highlighted in blue the periods ranging from 10.3 to 10.5 days. The significance levels of  $FAP = 10^{-3}$  and  $FAP = 10^{-5}$  are indicated by the solid and dashed horizontal lines, respectively



**Fig. 12.** Top left panel: RV time series of Gl 382 from all instruments, overlaid with the best-fitted GP in the optical from the GP + circular model is shown in black. Top right panel: Periodogram of the combined RV time series. Middle left panel: RV time series obtained when the optical data are corrected for activity using the GP, with the best-fitted circular model overlaid in black. Middle right panel: Periodogram of the combined RV time series corrected for activity. Bottom left panel: Phase-folding of the best-fitted model to the RVs after correction for activity. Bottom middle panel: Phase-folding of the residuals. Bottom right panel: Periodogram of the residuals. The significance levels of  $FAP = 10^{-3}$  and  $FAP = 10^{-5}$  are indicated by the solid and dashed horizontal lines, respectively



**Fig. 13.** Top panels show the detection rate of the injected signals without (left) and with *Wapiti* correction (right) using the SPIRou LBL RVs of Gl 382. The bottom panels show the number of periods recovered, meaning the number of periods with a detection rate above 50% (left) or 80% (right) with and without *Wapiti* correction.

## Appendix A: Supplementary material regarding the Wapiti correction

All these steps are explained in greater detail in Ould-Elhkim et al. (2023). They are designed to robustly determine the appropriate number of components for each target.

### Appendix A.1: Gl480 Wapiti correction

We employ a permutation test (Vieira 2012) for relevant component determination in wPCA analysis. This involves shuffling the dataset columns independently and performing wPCA on the shuffled data. This process is repeated 100 times to compare the explained variance between the original dataset and the permuted versions. We then identify the relevant components based on the p-values obtained from the permutation test, selecting those with p-values  $\leq 10^{-5}$ . As a result, Figure A.1 illustrates that only the first 21 components exhibit variance beyond noise for Gl480.

To assess the significance of the retained components after the permutation test, we employ leave-p-out cross-validation. We create 100 subsets of the RV dataset, each containing 80% of the lines chosen randomly. For each of these subsets, we recalculate the wPCA, resulting in N datasets, to examine the robustness of the remaining components.

We calculate the Pearson coefficient between the  $n^{\text{th}}$  principal vector obtained from the whole dataset using wPCA and the  $n^{\text{th}}$  principal vector obtained from each of the N subsets. As the wPCA vectors may change sign, we consider the absolute value of the coefficient. For selection purposes, we calculate the median Pearson coefficient value for each component number and retain only those components with a value above a predefined threshold of 95%. Consequently, Figure A.2 displays that the Wapiti method retains 7 components after this process.

### Appendix A.2: Gl382 Wapiti correction

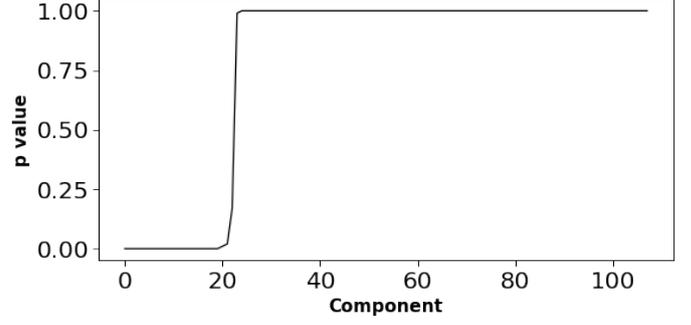
The same pre-processing steps are applied to Gl382. The permutation test (Figure A.3) leaves us with 33 components while the subsequent permutation test retains 11 components (Figure A.4).

For this target, we followed the outlier rejection step from Ould-Elhkim et al. (2023) to improve the statistical significance of the signal. Per-line RVs can be influenced by short-term fluctuations unrelated to the main patterns. Removing these outliers helps enhance the detection of systematic patterns and improves the accuracy of the reconstruction process.

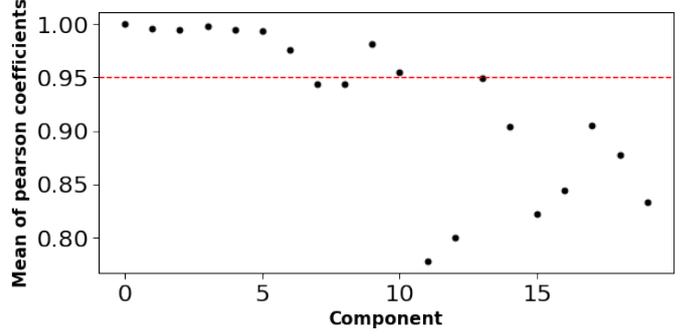
To address this issue, we calculate the anomaly degree in an observation by computing the ratio of the absolute deviation (AD) of the  $n^{\text{th}}$  principal vector  $V_n(t)$  to its median absolute deviation (MAD). To obtain the MAD, we calculate the median of the absolute deviations  $AD(X) = |X - \bar{X}|$  of a timeseries  $X$  with  $\bar{X}$  its median. We then determine the maximum value of this ratio across all components and epochs, denoted as  $D$ . This value serves as a measure of anomaly in the observation.

$$D(t) = \max_n \frac{AD(V_n(t))}{MAD(V_n(t))}. \quad (\text{A.1})$$

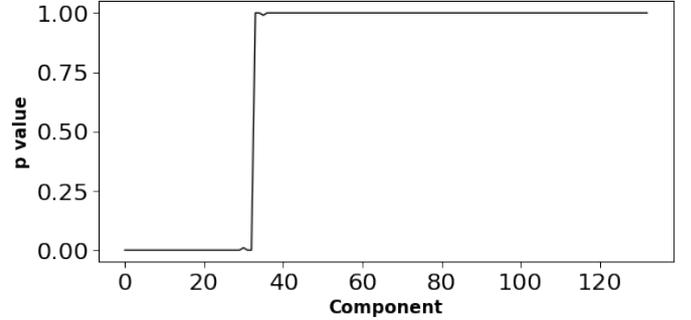
Consequently, we sort the epochs in descending order based on the value of  $D$ , creating a ranked list that indicates the degree of anomaly for each observation. We then decide to discard the initial  $N$  elements from this list. To determine the optimal value



**Fig. A.1.** p-values of each component of Gl480 as determined by the permutation test. Only the components with a p-value below a certain threshold ( $\leq 10^{-5}$ ) are considered to have a significant contribution, rather than being just random noise.



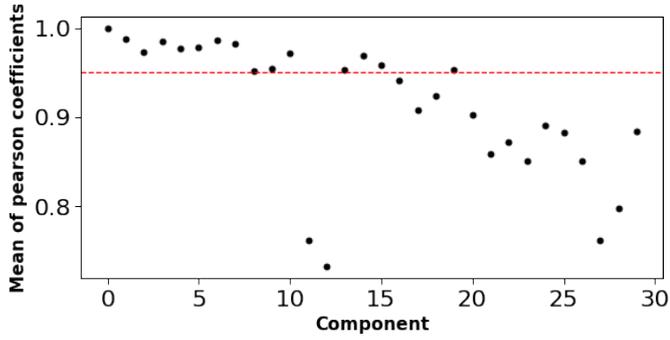
**Fig. A.2.** The median Pearson coefficients for each component of Gl480, components with values below the 95% threshold were eliminated from consideration.



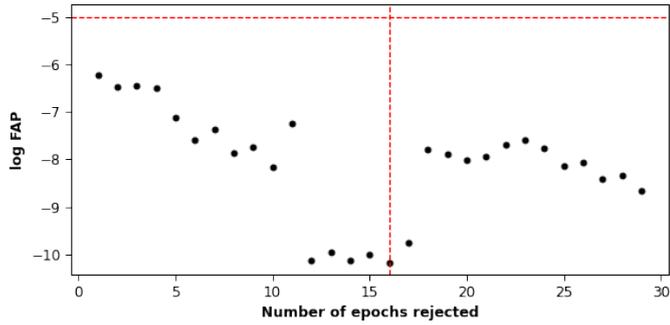
**Fig. A.3.** p-values of each component of Gl382 as determined by the permutation test. Only the components with a p-value below a certain threshold ( $\leq 10^{-5}$ ) are considered to have a significant contribution, rather than being just random noise.

of  $N$ , we systematically remove one element at a time and evaluate the resulting significance of the corrected RV's signal using the Wapiti method.

The analysis presented in Figure A.5 shows that by removing the first twelve outliers, there is a significant increase in the statistical significance of the signal. Initially, the FAP value is in the range of  $10^{-6}$ , but it improves to  $10^{-10}$  after this removal. We finalize the process by rejecting 16 outliers, which minimizes the FAP and reduces the time series from 133 to 117 data points.



**Fig. A.4.** The median Pearson coefficients for each component of Gl 382, components with values below the 95% threshold were eliminated from consideration.



**Fig. A.5.** The statistical significance of the 10.4 day period signal in Gl 382 is shown as a function of the number of epochs rejected. The epochs are rejected in descending order according to the value of  $D$ .

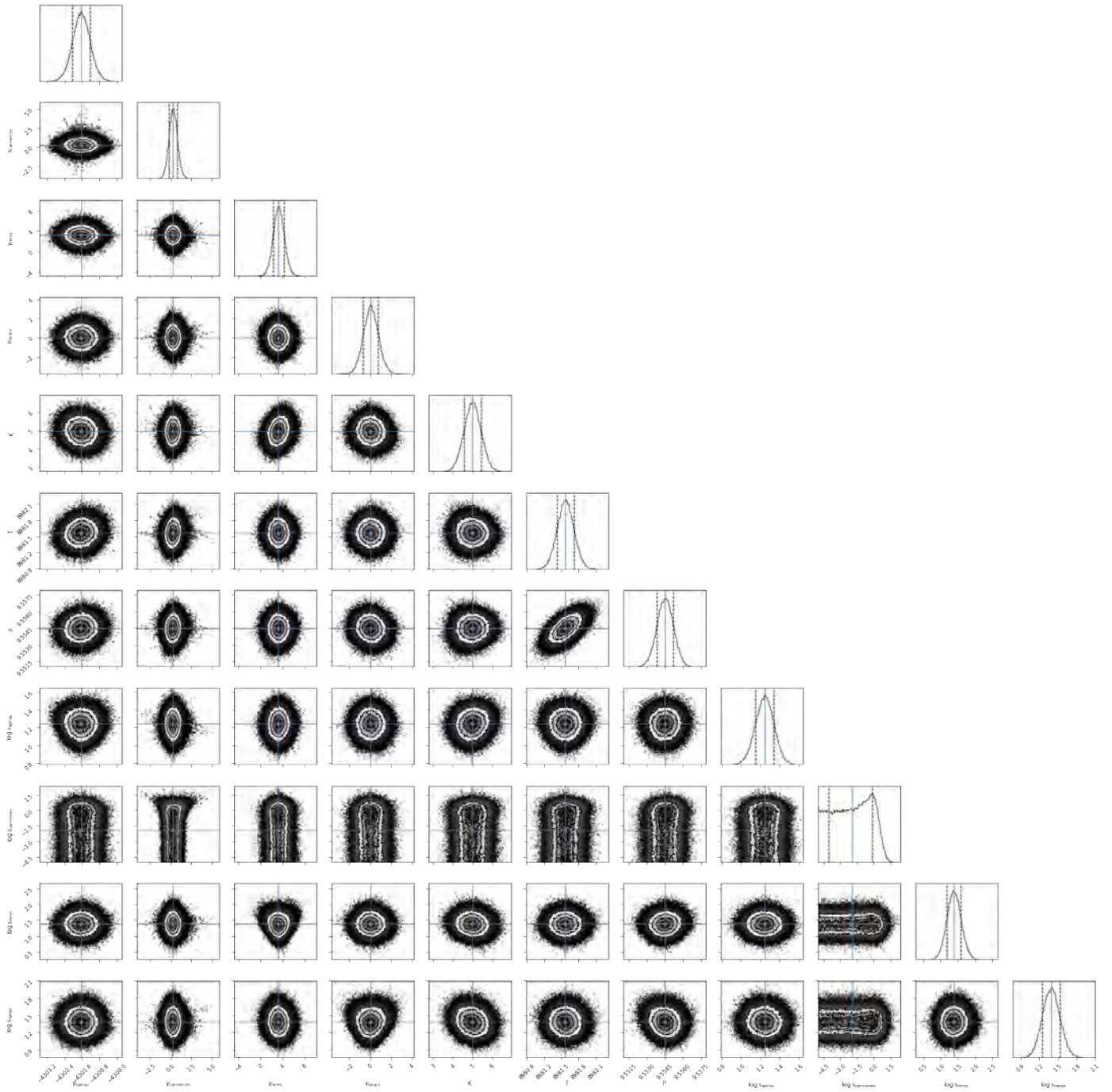
## Appendix B: Supplementary material regarding the MCMC analysis

### Appendix B.1: Gl 480 RV analysis

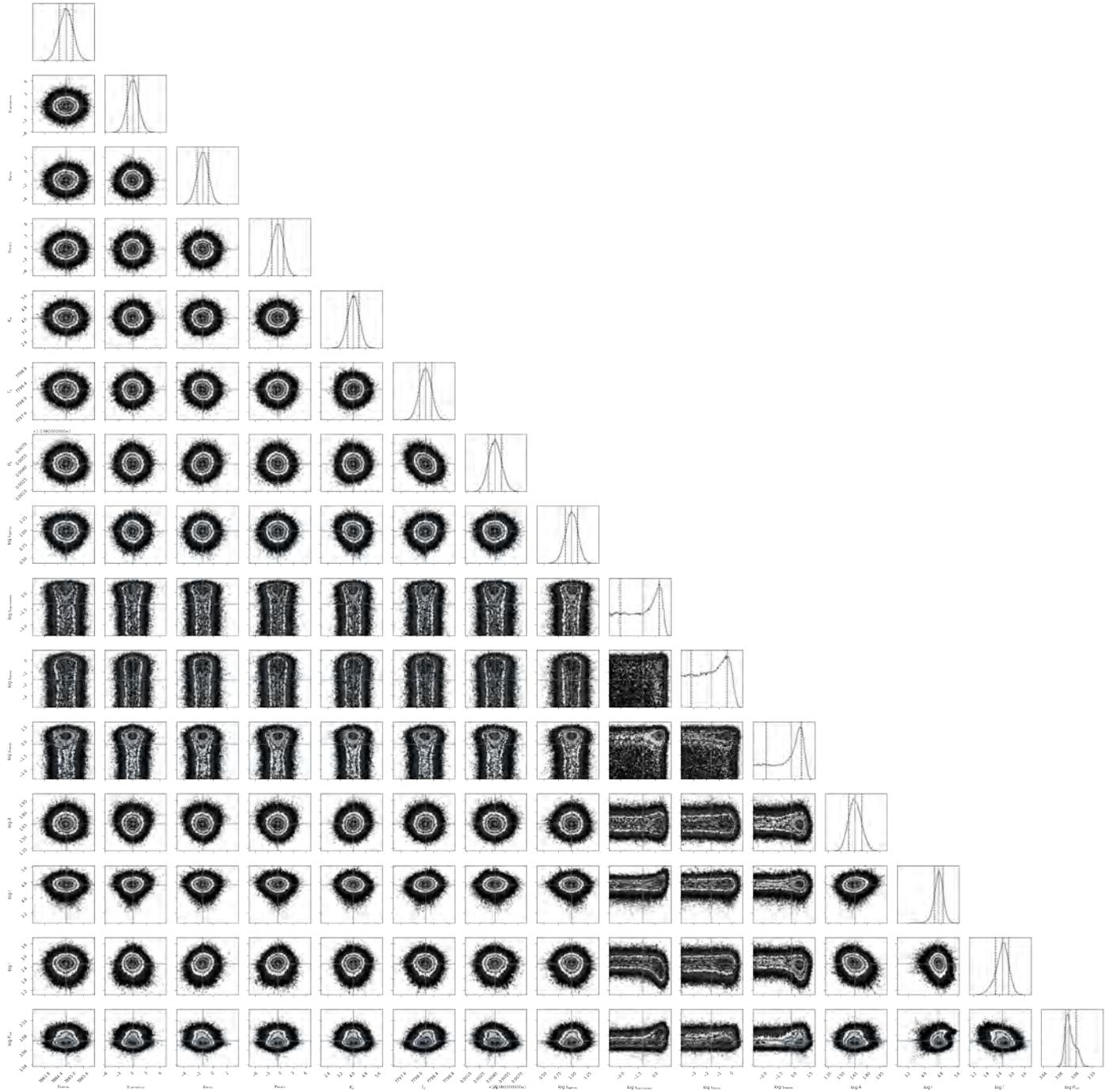
We present the corner plot obtained from the MCMC analysis performed on all instruments using a circular model in Figure B.1. The corresponding parameters can be found in Table 2.

### Appendix B.2: Gl 382 RV analysis

In Figure B.2, we present the corner plot resulting from the MCMC analysis conducted on all instruments. The analysis uses a circular model for the planetary signal, combined with a quasi-periodic GP for the instruments operating in the optical (CARMENES, HARPS and HIRES). The corresponding parameters can be found in Table 3.



**Fig. B.1.** Corner plot displaying the results of MCMC analysis on the RV time series of Gl 480



**Fig. B.2.** Corner plot displaying the results of MCMC analysis on the RV time series of Gl 382



**Titre :** Systèmes planétaires autour d'étoiles de faible masse avec SPIRou

**Mots clés :** Exoplanètes, SPIRou, Vitesses radiales, Spectroscopie, Naines M

**Résumé :** Les progrès de la recherche sur les exoplanètes dépendent de la précision des mesures de la vitesse radiale (VR). Les étoiles de type M ou naines rouges, en raison de leur abondance et de leur aptitude à accueillir des planètes rocheuses, font l'objet d'un intérêt grandissant dans cette recherche. Le domaine spectral du proche infrarouge offre une voie prometteuse, car il contient plus d'informations sur la vitesse radiale pour ces types d'étoiles tout en permettant une possible réduction de l'activité stellaire. Cependant, la présence de raies telluriques représente un défi important, compliquant les efforts de détection des exoplanètes. En s'appuyant sur le système de réduction de données SPIRou et sur l'approche line-by-line (LBL) pour extraire les VR, cette thèse présente des simulations qui reproduisent les effets systématiques induits par les raies d'absorption telluriques. Ces simulations sont utilisées pour démontrer l'efficacité de la technique d'analyse en composantes principales pondérées (wPCA) dans l'atténuation et la caractérisation de ces effets systématiques induits par les telluriques dans les données LBL. Ensuite, la méthode Wapiti (Weighted-principal component analysis reconstruction) est présentée comme un exemple d'une telle technique basée sur la wPCA, montrant comment elle facilite la détection de nouveaux signaux dans les données SPIRou qui étaient auparavant difficiles à démêler avec les telluriques et autres systématiques.

**Title:** Planetary systems around low mass stars with SPIRou

**Key words:** Exoplanets, SPIRou, Radial velocity, Spectroscopy, M dwarf

**Abstract:** The advancement of exoplanetary research hinges on precise radial velocity (RV) measurements. M dwarf stars, due to their abundance and suitability for hosting rocky planets, are a focal point in this pursuit. The near-infrared spectral domain offers a promising avenue, boasting increased RV information content and a potential reduction in stellar activity. However, the presence of telluric lines presents a substantial challenge, complicating exoplanet detection efforts. Leveraging the SPIRou data reduction system and the line-by-line (LBL) approach to extract RVs, this thesis displays simulations that replicate telluric-induced systematic effects arising from telluric absorption lines. Those simulations are used to demonstrate the effectiveness of the weighted-Principal Component Analysis (wPCA) technique in successfully mitigating and characterizing these telluric-induced systematics in LBL data. Following this, the Wapiti (Weighted-principal component analysis reconstruction) method is introduced as an example of such a wPCA-based technique, showcasing how it facilitates the detection of new signals in SPIRou data that were previously challenging to disentangle with tellurics and other systematics.