Contributions to the study of exoplanets : improvement of the stars radial velocity measurements, transit spectroscopy modeling

Anastasiia Ivanova

université PARIS-SACLAY

Contributions to the study of exoplanets : improvement of the stars radial velocity measurements,transit spectroscopy modeling

Contributions à l'étude des exoplanètes : amélioration des mesures de vitesse radiale des étoiles, modélisation de la spectroscopie de transit

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 127 : Astronomie et Astrophysique d'Île de France (AAIF)
Spécialité de doctorat : Astronomie et Astrophysique
Graduate School : Physique. Référent : Université de Versailles-Saint-Quentin-en-Yvelines

Thèse préparée dans l'unité de recherche **LATMOS(Université Paris-Saclay, CNRS)**, sous la direction de **Emmanuel MARCQ**, Maitre de conferences, et la co-direction de **Rosine LALLEMENT**, DRCE CNRS émérite

Thèse soutenue à Guyancourt, le 20 Décembre 2023, par

Anastasiia IVANOVA

Composition du jury

Membres du jury avec voix délibérative

Cyril SZOPA Professeur des universités, UVSQ/Paris Saclay, LATMOS	Président
François BOUCHY Professeur, HDR, Université de Genève, Observatoire de Genève	Rapporteur & Examinateur
Nadège MEUNIER Astronome, HDR, Université Grenoble Alpes	Rapporteur & Examinateur
Claire MOUTOU Directrice de Recherche, Institut de recherche en astrophysique et planétologie, IRAP	Rapporteur & Examinateur
Magali DELEUIL Professeur, Université Aix-MArseille, LAM	Examinatrice

THESE DE DOCTORAT
NNT : 2023UPASP147

université PARIS-SACLAY

ÉCOLE DOCTORALE
Astronomie
et Astrophysique
d'Île-de-France (AAIF)

Titre : Contributions à l'étude des exoplanètes : amélioration des mesures de vitesse radiale des étoiles, modélisation de la spectroscopie de transit.

Mots clés : exoplanetes, vitesse radiale, absorption tellurique, atmosphere

Résumé : La méthode des vitesses radiales a été la première méthode de détection d'une exoplanète. Depuis 1995, elle a permis de découvrir plus de 2000 exoplanètes et la méthode des RV a ses avantages. Outre la découverte de nouvelles exoplanètes et la détermination de leur masse apparente, elle est utilisée pour confirmer les exoplanètes découvertes par d'autres méthodes et pour mesurer la masse des exoplanètes en transit. Cependant, la méthode RV est limitée par les capacités techniques : la taille et la disponibilité du télescope, la stabilité et la précision de l'étalonnage de la longueur d'onde du spectrographe. Les spectrographes de nouvelle génération sont extrêmement précis d'un point de vue instrumental, ils sont stables et parfaitement calibrés, mais leur précision n'est pas encore suffisante pour détecter une exoplanète jumelle de la Terre. Ainsi, toute amélioration du traitement associé à l'extraction de la RV devrait se traduire par une amélioration de la précision, ou par une réduction du temps de télescope nécessaire pour atteindre une précision définie. Les spectres utilisés pour la méthode des RV étant obtenus à partir de la surface de la Terre, ils contiennent le spectre de l'étoile et le spectre de l'atmosphère terrestre (absorption tellurique). Jusqu'à récemment, les régions spectrales contaminées par l'absorption tellurique étaient exclues du traitement. Après correction, ces régions deviennent disponibles pour être utilisées pour la récupération de RV, améliorant ainsi la précision sur RV. Thèse est consacrée au développement de l'algorithme de correction (basé sur le service web TAPAS), et à l'estimation de l'augmentation de la précision des mesures de RV due à l'augmentation de la gamme de longueurs d'onde disponibles. La méthode de correction a été développée pour l'absorption tellurique causée par O_2 et H_2O dans le domaine des longueurs d'onde visibles et testée sur les données ouvertes d'ESPRESSO. Pour notre étude de cas de l'étoile HD40307, l'erreur formelle sur la RV est réduite de 0.77 à 0.64 ms^{-1}, ce qui correspond à une économie significative de 45% de temps de télescope. En plus du travail sur la correction de l'absorption tellurique, la thèse se concentre également sur le calcul direct de la RV. 2 méthodes principales existent et sont largement utilisées : la méthode de la fonction de corrélation croisée (CCF) et la méthode du template matching. Il existe également la méthode de Pierre Connes (décrite dans Absolute Astronomical Accelerometry), qui a ses limites en raison de son incapacité à s'appliquer à de grandes variations de vitesse. Dans la thèse, nous proposons et montrons l'application d'une nouvelle méthode, qui est une combinaison de la méthode classique de la CCF et de la méthode de Pierre Connes (PC). Nous proposons d'utiliser la méthode PC pour calculer le changement de RV directement sur la CCF, avec une correction préalable pour le BERV, au lieu de l'ajustement gaussien de la CCF. La méthode a été testée pour les données d'une étoile avec un système planétaire et sans système planétaire, les résultats montrent que la précision du calcul de la RV en utilisant la méthode PC est plus élevée que celle de l'ajustement gaussien. La thèse montre également les débuts d'une correction de l'activité stellaire, basée sur la méthode PC. La deuxième partie de la thèse est consacrée au développement d'un modèle de spectroscopie en transmission pour les exoplanètes telluriques en début de formation. Le développement d'un modèle est particulièrement pertinent après le lancement du JWST et en prévision du lancement d'Ariel, car la plupart des modèles actuels se concentrent davantage sur les géantes gazeuses et les naines brunes. Un modèle d'absorption raie par raie à haute résolution a été développé pour une atmosphère composée de vapeur d'eau H_2O et de CO_2.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Astronomie
et Astrophysique
d'Île-de-France (AAIF)

Title : Contributions to the study of exoplanets : improvement of the stars radial velocity measurements, transit spectroscopy modeling.

Keywords : exoplanets, radial velocity, telluric absorption, atmosphere

Abstract : The radial velocity method was the first method to detect an exoplanet. Since 1995 it has discovered more than 2000 more exoplanets and, although the transit photometry method is currently the most productive, the radial velocity method has its advantages. In addition to being able to discover new exoplanets and determine their apparent mass msini it is used to confirm exoplanets found by other methods and measure the mass of transit exoplanets. However, the RV method is limited by technical capabilities : the size and availability of the telescope, the stability and accuracy of the wavelength calibration of the spectrograph. New generation spectrographs are extremely accurate from an instrumental point of view, they are stable and perfectly calibrated, but their accuracy is still not enough to detect Earth's exoplanet twin. If it is not possible to improve the instrument the data processing needs to be refined, so any improvement of the processing associated to the radial velocity retrieval should result in an improvement of precision, or in a shorter telescope time required to achieve a defined-goal precision. Because the spectra for the radial velocity method are obtained from the Earth's surface, they contain the spectrum of the star and the spectrum of the Earth's atmosphere (telluric absorption). Until recently, spectral regions contaminated by telluric absorption were excluded from consideration. After telluric correction, those regions become available to be used for RV retrieval, improving the precision on RV. This thesis is devoted to the development of telluric correction algorithm (based on TAPAS web service), describing the telluric correction method and estimating an increase in the precision of radial velocity measurements due to increased available wavelength range. The correction method was developed for telluric absorption caused by O_2 and H_2O in the visible wavelength range and probed on ESPRESSO open data. For our case study of star HD40307, the formal error on RV is reduced from 0.77 to 0.64 ms^{-1}, corresponding to a significant saving of 45% telescope time. In addition to work on the telluric absorption correction, the thesis also focuses on the direct calculation of the radial velocity. At the moment, there are 2 main methods that exist and are widely used : the cross-correlation function (CCF) method and the template matching method. There is also the Pierre Connes method (described at Absolute Astronomical Accelerometry), which has its limitations due to its inability to be applied to large velocity variations. In this thesis we propose and show the application of a new method, which is a combination of the classical method of cross-correlation function and the method of Pierre Connes (PC). We propose to use the PC method of calculating the change of radial velocity directly on the CCF, with prior correction for BERV, instead of the classical Gaussian fitting of the CCF.The method has been tested for data of a star with a planetary system as well as without a planetary system, the results show that the precision of the radial velocity calculation using the PC method is higher than that of the Gaussian fit. The thesis also shows the beginnings of a stellar activity correction, based on the PC method. As the CCF can be considered an averaged stellar line, counting the change in radial velocity for the left and right wings separately, stellar activity can be analysed. The second part of the thesis is dedicated to the development of a transmission spectroscopy model for telluric exoplanets at the beginning of formation. The development of such a model is particularly relevant after the JWST launch and in anticipation of the Ariel launch, as most current models focus more on gas giants and brown dwarfs. A high-resolution line-by-line model has been developed for an atmosphere consisting of H_2O vapour and CO_2.

Résumé long en français

La détection et la caractérisation de systèmes planétaires autour d'autres étoiles que le soleil est un des sujets majeurs de l'astrophysique. Deux aspects principaux sont prédominants : tout d'abord, leurs propriétés, leur diversité ainsi que les liens avec l'étoile hôte apportent des contraintes sur la formation planétaire en général et celle de notre propre système solaire. Ensuite, l'étude des atmosphères et de l'habitabilité de chacune est motivée par la recherche de la présence possible de formes de vie extra-terrestres. La première exoplanète a été découverte à l'Observatoire de Haute-Provence en 1995 (Mayor and Queloz, 1995) et, depuis lors, le nombre d'exoplanètes découvertes augmente chaque année. L'une des méthodes les plus productives pour détecter et étudier les exoplanètes est la méthode des vitesses radiales (RV). Celle-ci est basée sur l'effet Doppler « réflexe ». Si l'étoile possède un système planétaire, étoile et planètes se déplacent autour du centre de masse du système étoile-planètes et, lorsqu'il observe le spectre stellaire avec des spectrographes terrestres, l'observateur constate un déplacement des raies spectrales stellaires vers les régions rouges et bleues du spectre avec des périodicités qui reflètent le nombre et l'orbite des planètes. Malgré les progrès considérables déjà obtenus, il existe encore un certain nombre de limitations qui empêchent la détection de planètes de type terrestre dans des étoiles de type solaire à l'heure actuelle (il faudrait une précision de 9 cms^{-1}). La thèse vise à étudier ces limitations et à tester des méthodes pour améliorer la précision des mesures. En outre, la thèse comprend des travaux sur la modélisation de l'atmosphère d'une planète semblable à la Terre au stade de l'océan de lave en fusion (magma), des travaux sur les statistiques des exoplanètes et des travaux sur la calibration et l'étude des exoplanètes à l'aide d'un nouveau spectrographe.

Correction de l'absorption tellurique

L'une des limitations rencontrées dans la méthode des vitesses radiales est liée au fait que la lumière des étoiles traverse l'atmosphère terrestre et est absorbée par les molécules de celle-ci, créant ainsi des régions d'absorption dans le spectre. Les raies stellaires d'une part, et les raies ou bandes (systèmes de raies) d'absorption par l'atmosphère terrestre dites raies telluriques, d'autre part, sont deux systèmes différents, attachés à des référentiels différents (centre de masse étoile-planètes pour le premier, référentiel de l'observatoire terrestre pour le second) et leurs déplacements sont donc indépendants. Les raies telluriques sont donc contaminantes et brouillent le signal planétaire. Classiquement, les régions du spectre sujettes à l'absorption tellurique sont exclues du domaine spectral utilisé pour déterminer RV. Au contraire, dans la thèse on considère que l'on peut corriger le spectre de l'absorption tellurique et ne plus exclure autant de régions spectrales. Nous avons donc étudié l'effet d'une correction de l'absorption tellurique sur la précision des mesures de vitesse radiale RV. Une méthode de correction de l'absorption tellurique a été développée, et les résultats de la correction ont été analysés pour les longueurs d'onde du visible (la correction est faite pour les molécules H_2O et O_2) et du proche infrarouge (dans ce domaine la correction est faite pour H_2O, O_2, CO_2 et CH_4), pour les étoiles de la classe FGK. Nous montrons une amélioration de la précision des mesures de vitesse radiale, pour notre exemple: pour l'étoile HD40307 (archive de données ouverte ESPRESSO, correction O_2 et H_2O), l'erreur formelle sur la vitesse radiale est réduite de 0,77 à 0,64 ms^{-1}, ce qui correspond à une économie significative de 45% du temps de télescope.

Détection de l'activité stellaire

Parmi les facteurs qui aujourd'hui limitent la précision des mesures, l'activité stellaire est le principal. Dans ce terme d'activité on englobe granulation, super-granulation, effets des plages et taches, cycles d'activité. L'activité d'une étoile mère peut parfois masquer le signal d'une exoplanète (si l'amplitude de la variation de vitesse radiale due à l'activité est plus forte que

l'amplitude de la variation de vitesse radiale due à la planète), elle peut également imiter la présence d'une planète ou introduire une erreur dans la mesure des paramètres de la planète. Il existe un certain nombre de techniques pour améliorer la précision des vitesses radiales. La thèse propose une autre méthode de mesure de changement de RV basée sur la méthode de Pierre Connes (EPiCA). Cette nouvelle méthode EPiCA est appliquée à la fonction de corrélation croisée CCF présente dans les données archivées du spectromètre ESPRESSO de l'ESO au Very Large Telescope. La méthode a été testée sur les données d'une étoile possédant un système planétaire et sur les données d'une étoile sans système planétaire. Sur ces exemples, les résultats montrent que la précision du calcul de la vitesse radiale en utilisant la méthode EPiCA est plus élevée que celle de l'ajustement d'une Gaussienne à la CCF, méthode standard dont le résultat est archivé avec les spectres stellaires.
Par ailleurs, comme la CCF peut être considérée comme une image moyenne des raies stellaires, l'algorithme EPiCA peut être utilisé pour détecter des changements de forme dans les raies stellaires et donc pour détecter la présence de l'activité stellaire qui perturbe la mesure de RV. La méthode EPiCA semble moins sensible à ces perturbations que la méthode standard d'ajustement Gaussien de la CCF.

Parties supplémentaires

La deuxième partie de la thèse est consacrée au développement d'un modèle de spectroscopie en transmission pour les exoplanètes telluriques en début de formation. Le développement d'un tel modèle est particulièrement pertinent après le lancement du James Webb Sapce Telescope (JWST) et en prévision du lancement d'Ariel, car la plupart des modèles actuels se concentrent davantage sur les géantes gazeuses et les naines brunes. Un modèle raie par raie à haute résolution a été développé pour une atmosphère composée de vapeur d'eau H_2O et de CO_2. La thèse résume aussi quelques résultats sur les statistiques des exoplanètes, ainsi que les premiers résultats du nouveau spectrographe FFOREST de l'observatoire SAO à Zelenchuk, dans le Caucase.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my gratitude towards my amazing thesis supervisors: Rosine Lallement, Emmanuel Marcq. I am extremely grateful to you for sharing your experience with me, explaining complex concepts to me and guiding me in the world of science for all 3 years.

I am grateful to Jean-Loup Bertaux, who, not being my supervisor according to the documents, was one of the key figures in my development as a scientist. Without his personal participation this PhD would not have happened. I have learned a lot since the beginning of our work in 2019. I am exceptionally lucky to have this experience.

I would also like to thank Jean-Philippe Beaulieu and Pierre Drossart for accepting me to ARIEL summer school in 2021 and thus included me in a community of excellent and talented doctoral students, post-docs and professors. It really brought a lot of joy and interest to my scientific life and life in general, especially after all COVID restrictions.

I am grateful to Piercarlo Bonifacio for sharing his experience in stellar modeling. This knowledge helped a lot during my thesis and I am sure will be useful in the future.

Thanks to all the great people who have met on my way over these three years, with whom I could talk about work and life events. I am happy to say that this PhD brought me not only knowledge, but friends as well. Thank you Natalia, Emilie and Yassin. I would also like to thank the LATMOS and Observatoire de Paris (Meudon) PhD students and post-docs and all the ARIEL summer school students.

I am happy to thank my colleagues. Especially Daria for being my guide in the world of life in France. Thank you Sacha for all the help, discussions and jokes.

I have to thank Thierry Fouchet and Laurent Verstraete for all their help and support from the ecole doctorale.

Along with gratitude to Thierry and Laurent, I want to thank French Embassy in Moscow for giving me this opportunity, specially Olivier Dubert, who helped me a lot with the documents and administrative

things.

I cannot express my full gratitude to my family. Without their constant support and love this could never happened. Thank you to my dear friend Maria, who have brightened my sad days. And thank you Andrey for additional motivation to write the thesis text.

There is a proverb: “It takes a village to raise a child”. I can rephrase it: “It takes a village to make a PhD thesis” and I am grateful and lucky to have all the amazing people around me. Thank you all.

Introduction

Almost 30 years have passed since the discovery of the first exoplanet (Mayor and Queloz, 1995). Since then, several thousand exoplanets have been discovered. These planets differ in radii, masses, periods, for some of them we know only the main parameters, for others, scientists were able to detect the atmosphere and disassemble it into components. Despite the fact that many exoplanets have been discovered, the search for the Earth's twin planet is still relevant. In order to detect such a planet, classical methods of detection and observation are not enough, their improved versions are needed.

The first and still one of the most productive methods is the Radial Velocity method. Since the discovery of the first exoplanet, it has undergone some changes. Improvements were made both in the instrumental part and in the data processing part. In the instrumental field, the most accurate spectrographs have been created in recent years, the wavelength calibration and temperature-pressure stability have been brought to the ideal. After the instrumental possibilities for improvement have been exhausted, it is necessary to turn to additional data processing. One of the possible improvements in data processing is the correction of telluric absorption, which is the main subject of this thesis. Along with correction of telluric absorption a new method of RV calculation is proposed. Short studies of exoplanetary atmospheres and statistics of exoplanets are presented as well.

The main subject of the thesis is telluric absorption correction and how its application can improve the precision of radial velocity measurements. Chapter 1 is dedicated to this study, we first explain the method and then demonstrate how its application improves the precision of radial velocity measurements. In Chapter 2 we describe the methods that can be used to calculate the radial velocity of a star. A new approach of RV calculation is also proposed using the Pierre Connes algorithm (Section 2.3.3): the improvement in RV precision applying this new method (Chapter 4) and the possibility of using a new method to study the impact of activity of stars on retrieved RV are discussed (Chapter 3, 5). Analysis of sources of various bias have been studied and explained in Chapter 6. Most of the studies made use of an exceptional series of spectra publicly available, recorded with the ESPRESSO spectrograph.

Additionally work on observations and stability evaluation of SAO/RAS spectrograph was done and is presented in Chapter 7. Some results in the field of exoplanetary atmospheres is shown in Chapter 8, followed by work on exoplanet statistics showed in Chapter 9.

Chapter 1

Telluric correction to improve the precision of Radial Velocities retrieval

Telluric absorption is absorption produced by the Earth's atmosphere. The ground-based astronomical spectroscopy is affected by the telluric absorption because the light from a star is crossing the Earth's atmosphere and is partially or fully absorbed by it. Even a minimization of atmospheric influence by choosing a high-altitude observatory location cannot help to avoid this effect. As defined originally, the main subject of this PhD work was to correct stellar spectra used for the detection of exo-planets from the telluric absorption, to make available new spectral domains otherwise discarded form Radial Velocity retrievals, because of severe telluric contamination (see figure 1.8). The telluric absorption is imprinted on the spectrum of the target star as a system of spectral lines, it is called transmission (or transmittance) spectrum of the Earth's atmosphere. For each molecule the vertical distribution of its abundance, the atmospheric conditions (temperature and pressure vertical profiles) affect shape, depth and position in wavelengths of telluric spectral lines. Even wind can produce a shift of telluric lines as it is shown in Figueira et al. (2012). Resulting transmission spectrum strongly depends on the site and weather conditions. For the ground-based spectroscopy in the optical domain, the two main molecules are H_2O and O_2, for the near-infrared domain additionally CO_2, CH_4 and N_2O. While O_2 is quite stable, represented by 3 large bands (in optical domain, γ, B and A band), with strong, sharp doublet-lines, which are easy to identify, the H_2O is widely spread over all wavelength ranges and shows strong hourly and seasonal variations.

In this chapter the application of the method of telluric correction based on TAPAS online web service is described. In section 1.3 we describe the application of the developed method to correct spectra of

hot stars in order to perform a 3D kinetic tomography of Taurus clouds (published in Ivanova et al. (2021a)), which can be considered as a first step towards developing the key method of the thesis. In section 1.4 we describe the new method of telluric correction for cooler stars and in section 1.5 we demonstrate the gain in precision caused by the telluric correction (published in Ivanova et al. (2023)).

1.1 Overview of existing methods

There are several approaches to correct the spectra of star from telluric absorption. A first technique is to divide the spectrum of the target star by a "telluric standard" spectrum recorded as close as possible to the target star in time and direction (Vacca et al., 2003). A "telluric standard" spectrum is a high signal-to-noise spectrum of a hot, rapidly rotating star, because for such objects stellar lines are broad and their observed spectrum is simply the spectrum of the atmospheric transmittance superimposed onto a smooth continuum. A second type of method is based on the accumulation of observational data. For example Artigau et al. (2014) have constructed libraries of empirical telluric absorption spectra based on series of observations of hot fast-rotating stars. These spectra are used to feed a Principal Component Analysis (PCA) of any new dataset. Leet et al. (2019) used series of spectra of fast-rotating stars observed in various atmospheric conditions to extract the whole series of water vapour lines. Bedell et al. (2019) proposed a data-driven method based on a very large series of observations covering a variety of stellar-telluric Doppler shifts, allowing to disentangle stellar and telluric contributions without prior knowledge of spectral features for the star or for the Earth's atmosphere. Cretignier et al. (2021) developed a similar method and a PCA on residuals to eliminate telluric contamination and also instrumental systematics. A third group of correction methods is based on extremely precise atmospheric transmission spectra. The most detailed models of time and location dependent atmospheric profiles are the Global Data Assimilation System (GDAS) and the European Centre for Medium-Range Weather Forecasts (ECMWF). Both facilities provide the altitude distribution of pressure, temperature and H_2O concentration. Several tools are using this profiles and HITRAN molecular database to compute transmission spectra (Rothman, 2021; Gordon et al., 2017). The MolecFit (Smette et al., 2015) and the Telfit (Gullikson et al., 2014) are using the GDAS profiles, while TAPAS (Bertaux et al., 2014) is using ECMWF profiles. Both MolecFit and TAPAS are using Line-by-line Radiative Transfer Model (LBLRTM) code and HITRAN database to compute the atmospheric transmission spectrum. The MolecFit includes a software for telluric

correction, it retrieves automatically the atmospheric profile at the time and location of observations from ESO observatory, while TAPAS only provides the atmospheric profile at the time and location without a tool for correction. The advantage of the third method is that it does not require much additional data.

Recently, several kinds of combinations of the above methods have been published. For example, Allart et al. (2022) showed how it is possible to correct a stellar spectrum based on a single-layer atmosphere. The authors used the HITRAN database to compute the transmittance of H_2O and O_2, but, unlike the third type of technique mentioned above based on a detailed atmospheric profiles, they used a unique pressure and a unique temperature. Cook et al. (2022) presented a different approach which is used in the SPIRou pipeline. It starts with TAPAS absorption spectra and use them to perform a first correction. In a second step, they perform an empirical correction of the residuals using a PCA whose vectors are taken from a library of accumulated corrections of hot stars spectra. In what follows, we study the results of the application of a unique, adjusted correction based on TAPAS.

1.2 The TAPAS on-line service

The web-based service, TAPAS (in French: "Transmissions Atmosphériques Personnaliséees pour l'Astronomie", translated to "Transmissions of the AtmosPhere for AStromomical data", https://tapas.aeris-data.fr/en/home/) has been developed in collaboration between IPSL, LATMOS, GEPI and ACRI-ST for researchers using spectra of astronomical targets acquired from ground-based instruments or spectra of sunlight reflected by Earth acquired from orbital platforms. Both types of data are affected by the atmospheric transmission and TAPAS aims at providing a model of the atmospheric transmittance adapted to the observations. This service, freely available, is developed and maintained within the atmospheric CNES and INSU/CNRS ESPRI center operated by IPSL. TAPAS computes the atmospheric transmission in the line-of-sight to the target indicated by the user. The user files a request indicating the time, ground location, and either the equatorial coordinates of the target or the Zenith Angle of the line-of sight (LOS). The actual atmospheric profile (temperature, pressure, humidity, ozone content) at that time and place is retrieved from the ESPRI atmospheric data base. The database includes measurements acquired by satellite, balloons and aircraft or from other sources to make them available to the entire scientific community for exploitation by modeling or assimilation. ESPRI stores

atmospheric profiles computed by using the ECMWF (European Center for Medium-range Weather Forecast) meteorological field and the code ARLETTY developed for ESPRI by ACRI-ST from the algorithms of Alain Hauchecorne at LATMOS. For atmospheric levels which are above the highest ECMWF field, Arletty is computing a profile by using the MSISE-90 algorithm (Hedin, 1991). The full atmospheric model is made of 127 layers for which temperature, pressure and constituents are computed and the model is refreshed every six hours.

The atmospheric transmission is computed from LBLRTM software (Clough et al., 1992, 2005) and HITRAN data base (Rothman, 2021) for a number of gases including O_2, H_2O, O_3, CO_2, N_2O and CH_4 as well as the Rayleigh extinction. Recently TAPAS was updated to include NO_2 and the wavelength domain has been extended in both near UV and near-IR, covering now 300 to 3500 nm.

The first purpose of TAPAS output is to allow identification of observed spectral features as being from atmospheric origin. But the TAPAS transmittance spectra can be used to perform corrections of astronomical spectra by removing those features. The user input parameters are:

- the UTC time of the observation
- the location of the observer in longitude, latitude, and altitude (a list of selected observatories is available in a roll menu)
- the RA and DEC of the astronomical target (RA, Right Ascension, DEC, Declination, in the J2000 system). From RA and DEC J2000, TAPAS computes the zenith angle, as an input parameter to the atmospheric transmission. Since an observer may wish to keep confidential its target (defined by RA and DEC), another available option is to give simply the zenith angle and the air mass will be computed by TAPAS.

In addition, the user must select output parameters:

- wavelength reference system: wave number (unit cm^{-1}, which has the advantage of being independent of the index of refraction of air, or wavelength in vacuum (unit, nm, nanometer, also independent of the index of refraction of air), or wavelength (in nm) at 15^oC and 760 mm Hg pressure (it is a commonly used laboratory reference wavelength standard; in particular, lines of Thorium Argon lamps which are frequently used to calibrate the wavelength scale of spectrometers are given in this particular wavelength reference system.)
- definition of spectral interval for the computation

- selection of atmospheric constituents
- choice between independent transmittance for each species (plus Rayleigh) or combined transmittance
- optional choice of application of the so-called BERV (Barycentric Earth Radial Velocity) correction to compute a Doppler stretch and recover the exact spectrum which would be seen if the observatory were, not on Earth, but rather at the barycenter of the solar system.

Currently TAPAS is operating on the basis of HITRAN 2016 (Gordon et al., 2017).

1.3 Telluric correction of hot stars for 3D kinetic tomography of Taurus interstellar clouds.

As a training exercise for telluric correction of stellar spectra and manipulation of TAPAS outputs, I took part to an analysis dedicated to the motions of interstellar clouds, localised in 3D space based on the starlight extinction they produce. This study has been published in Ivanova et al. (2021a) and is summarized below.

Efforts are currently undertaken to achieve an automated 3D kinetic tomography of Milky Way interstellar clouds, based on the combination of gas emission, gas and dust absorption of stellar light and newly available Gaia parallaxes. In this context, preliminary tomographic results allowing to test the new techniques are precious. Useful preliminary results can be obtained based on 3D dust maps on the one hand, and large series of interstellar absorption lines measured in spectra of targets star distributed in 3D space on the other hand. We performed such a preliminary study in the region of the well-known Taurus-Perseus-California nearby clouds (hereafter called Taurus clouds), using interstellar neutral potassium (K I) absorption and recent 3D dust maps. This study involved telluric transmittance and telluric correction, therefore it has been a preparation for the subsequent work at the core of the thesis for me.

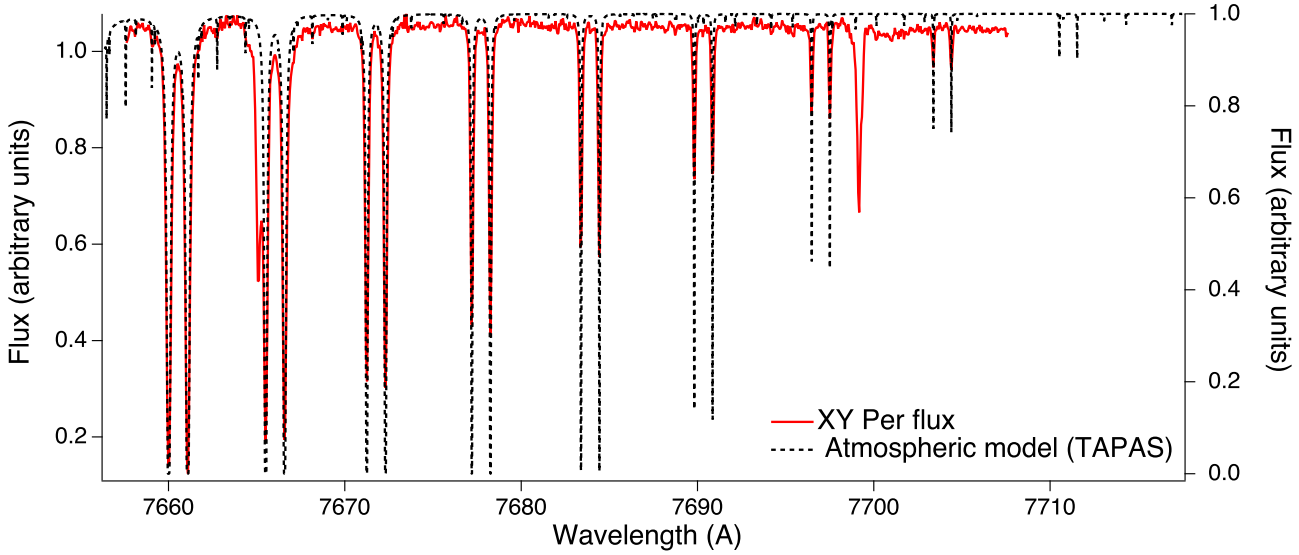
We assume that interstellar dust and gas concentrations are spatially co-located, which is universally verified in the case of dense atomic and molecular clouds such like the Taurus clouds. 3D distributions of the dust clouds in the Sun vicinity are now publicly available. They are obtained by inversion of massive amounts of individual extinction measurements for target stars distributed in space. In order to assign a radial velocity to the clouds appearing in the maps, we use stars located in front of, within

and beyond the clouds and measure the interstellar absorption lines in their spectra. The combination of all results allows to distinguish which lines are generated by which clouds and to measure their individual velocities.

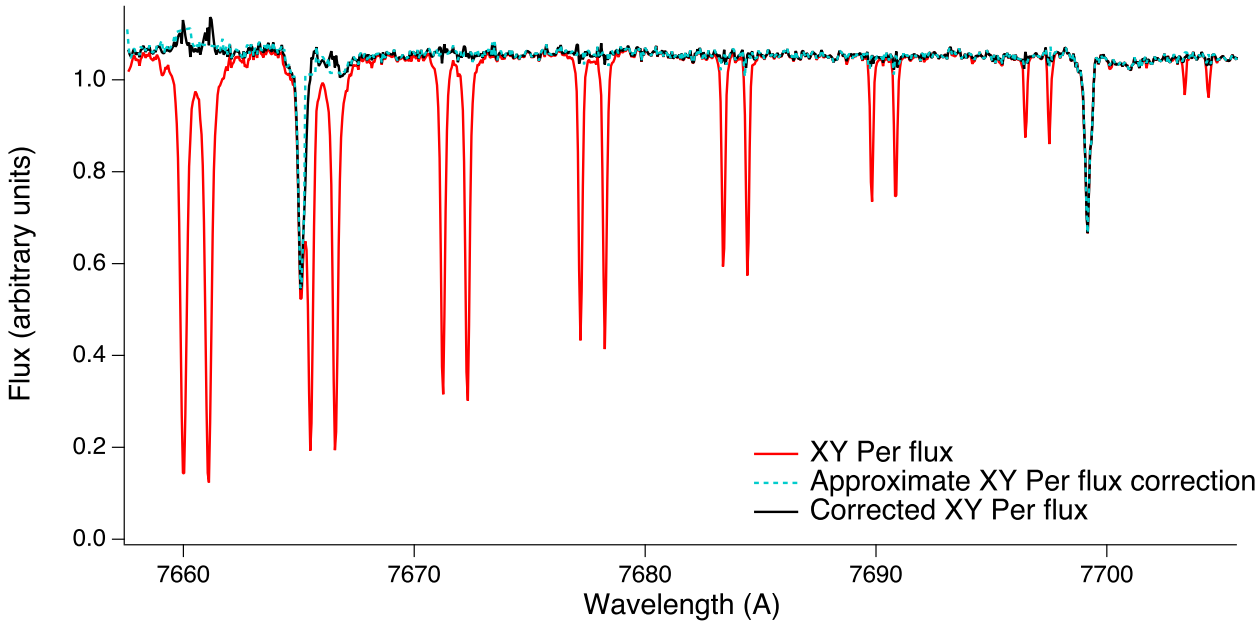
To perform this tomography, we used the 3D dust maps from Lallement et al. (2019) on the one hand, and stellar spectra recorded with the Echelle spectro-polarimeter NARVAL at the Pic du Midi Bernard Lyot Telescope (TBL) during two dedicated programs on the other hand (L152N07 and L162N04, P.I. Lallement). We complemented the latter by archival data from the Polarbase facility, recorded with Narval or the ESPaDOnS spectro-polarimeter at CFHT (Donati et al., 1997; Petit et al., 2014) and by a few other archival data. Because it generates narrow lines ideally suited for radial velocity measurements, we used interstellar neutral potassium (K I).

The first, stronger transition of the (K I) $\lambda\lambda$7665, 7699 Å resonance doublet is located in the central part of the A-band of telluric molecular oxygen. This is why the second, weaker transition is used alone. This is unfortunate, since the combined use of the two transitions provides much better determinations of the cloud structure and individual densities and temperatures of the absorbing clouds. This is the reason why we devised a novel method to extract a maximum of information from the two transitions simultaneously. The new method is a forward modeling of both telluric and interstellar lines and their best fit adjustment to spectra. The reason for that choice is what follows. In principle, telluric-corrected (telluric-free) spectra can be obtained following the adjustment of a telluric model (such like TAPAS) and subsequent division of the data by the adjusted model. However, such corrected spectra are often characterized by strong residuals in the regions of the deepest telluric lines that make the determination of the continuum and the subsequent profile-fitting difficult. This is the case for the $\lambda\lambda$7665 Å transition region, where atmospheric oxygen lines are (before convolution by the instrumental PSF) saturated (see Fig 1.1a). To avoid this, we used two distinct, consecutive adjustments. First, we used the atmospheric transmittance models from the TAPAS online facility (Bertaux et al., 2014), and selected those that are suitable for the respective observing sites. We used these models and their adjustments to the data to (i) refine the wavelength calibration, (ii) determine the instrumental function (or line spread function, hereafter LSF) as a function of wavelength along the echelle order, and (iii) derive the atmospheric transmittance that prevailed at each observation, prior to entry into the instrument. To do so, we exclude the spectral regions containing the interstellar lines. In a second step, this adjusted transmittance and the spectral resolution measurements were

used in a global forward model in combination with a multi-cloud model of the two IS K I 7665 and 7699 Å transitions.



(a)



(b)

Figure 1.1: Illustration of the preliminary telluric correction, prior to the dual interstellar-telluric profile-fitting. The TAPAS model was selected for the Pic du Midi observatory.
(a) Atmospheric profile (black color, right axis) after adjustment to the observation, superimposed on the initial stellar spectrum (here the star XY Per, red color, left axis). We note that this is the atmospheric transmittance before convolution by the instrumental profile, which is used at the next step of profile-fitting.
(b) Corrected stellar spectrum (black line) obtained by division of the raw data by the above atmospheric profile, after its convolution by the instrumental profile. It is superimposed on the initial stellar spectrum (red line). There are some spiky residuals in the strong absorption areas which are due to the division of weak quantities for both the data and models. Also shown is the quasi-continuum obtained from the corrected spectrum after some interpolation at these regions (turquoise dashed line). Here, there is only a very small difference between the two curves. We note that only the adjusted pre-instrumental (i.e. non convolved) profile shown at the top will be used at the next step.

1.3.1 Data

The Narval spectra cover a wide wavelength range from 370 to 1058 nm. Target stars were selected in direction and distance to sample the volumes of Taurus, Perseus, and California clouds, and they were chosen preferentially among the earliest types. The observations were done in the "star only" (pure spectroscopic) mode that provides a resolving power on the order of 75,000. Most of the observations were performed at low airmass. Spectra of good and excellent quality were obtained for a total of 57 new targets. We used the fully reduced and wavelength-calibrated spectra from the Narval pipeline. The pipeline provides the spectrum in separate echelle orders, and here we use the order that covers the $\simeq$ 764.5-795.0 nm interval and contains the two K I transitions at 7664.911 and 7698.974 Å. The first transition is also included in an adjacent order; however, we used a unique order and non resampled data for the K I extraction.

We also performed an online search in the POLARBASE database and found ten Narval and 15 ESPaDOnS spectra of target stars that have been observed as part of other programs and are suitable for our study. Their spectral resolution R is comprised between $\simeq$ 55000 and $\simeq$ 85000, with the lower value for the polarimetric mode.

We also searched for useful spectra in the ESO archive facility and found spectra for seven UVES, two XSHOOTER, and one FEROS targets of interest (R= $\simeq$70000, 18000, and 80000 respectively). We also used six spectra from the Chicago database recorded with the ARCES spectrograph at R=$\simeq$ 31000 (Wang et al., 2003; Fan et al., 2019). Finally, one spectrum recorded with the OHP 1.93m Elodie spectrograph R$\simeq$48000 has been added. Elodie does not cover the K I doublet; however, we derived the Doppler velocity of the main absorbing cloud from neutral sodium lines (5890 Å doublet). Finally, in addition to these archival data that entered our telluric correction and profile-fitting, we complemented the observation list with Chaffee and White (1982) Doppler velocities of Taurus clouds for eight targets, also based on neutral potassium lines and high resolution observations. One additional result by Welty et al. (1994) based on sodium lines has also been included. Tables describing the data can be found in the attached PDF of the published article.

1.3.2 Spectral analysis

The analysis has two phases. The first phase consists in fitting a telluric model to the data and determining the wavelength-dependent LSF. During this step, the spectral regions around the K I

lines are excluded. The second phase is the dual profile-fitting itself. The previously derived telluric model is multiplied by as many Voigt profiles as necessary to represent the interstellar K I absorption lines and the velocity distribution of intervening clouds, and the total product is convolved by the wavelength-dependent LSF. The number of clouds and the interstellar parameters are optimized to fit this convolved product to the data.

First phase: preliminary telluric correction and determination of the wavelength-dependent instrumental width

For the telluric correction, we used the TAPAS online facility: TAPAS uses vertical atmospheric profiles of pressure, temperature, humidity, and most atmospheric species, interpolated in the meteorological field of the European Centre for Medium-range Weather Forecasts (ECMWF), the HIgh-Resolution TRANsmission molecular absorption database (HITRAN), and the Line-by-Line Radiative Transfer Model (LBLRTM) software to compute the atmospheric transmission at ultra-high resolution (wavelength step $\simeq 4\ 10^{-6}$ Å). See Bertaux et al. (2014) for details about TAPAS products and all references. Here we use TAPAS calculations for the observatory from which each spectrum under analysis was obtained. At first we created a Gaussian LSF for an estimated preliminary resolution, adapted to the central wavelength of the correction interval. Via this LSF, we then convolved the initial TAPAS transmission model for one airmass unit, and through cross-correlation we obtained a preliminary Doppler shift between the observed spectrum (data) and this model (convolved TAPAS). To do so, we excluded the potassium lines regions from the Doppler shift estimate, using a mask with null value from 7663 to 7667 and from 7697 to 7701 Å for both the data and model. We then used a comprehensive list of unblended or weakly blended telluric O_2 lines located in the spectral interval containing the K I doublet to obtain a first estimate of the airmass at observation as well as a more refined estimate of the spectral resolution. To do so, Gaussian profiles were fitted in both the stellar and earth transmission spectra at the locations of all these potentially useful lines and the results were compared one-by-one. Undetected lines that are too weak and/or lines with profiles that are too noisy in the data were removed from the list. The strongest lines were selected and all corresponding ratios between the data and model equivalent widths were computed. The average ratio provides a preliminary airmass (or relative optical thickness with respect to the initial model), and the line-width

increase between the model and data provides an estimate of the mean resolution.

After all these preparatory steps, the main procedure starts. It can be broken down into the following three main steps.

1. The first step involves fitting the telluric model to the data by means of the *rope-length method* , masking the central parts of the strong telluric lines (where the flux is close or equal to zero). This step assumes a unique Doppler and a unique Gaussian LSF. The free parameters are the airmass factor, the Doppler shift, and the LSF width. The convergence criterion is obtaining the minimal length of the spectrum obtained after division of the data by the convolved transmission spectrum. Here the length is simply the sum of distances between consecutive data points. Such a technique assumes that the minimum length corresponds to the smoothest curve and therefore to minimum residuals after the data to model ratio, that is to say a good agreement on all the line shapes. It is important to note that this technique is sufficient for weak to moderate lines, but it is applied here only as an intermediate solution.

2. The second step consists of the computation of a decontaminated spectrum by division of the data by the adjusted transmission model and subsequent replacement of the ratio by an interpolated polynomial at the locations of the strong telluric lines. This provides a quasi-continuum without strong residuals. By quasi-continuum, we mean the telluric-free spectrum, that is the stellar spectrum and its interstellar features.

3. The third step involves a series of fits of the convolved product of the quasi-continuum from step 2 and a TAPAS transmission to the data, gradually removing constraints on the width of the instrumental profile and on the Doppler shift, and allowing their variability within the fitted spectral interval, along with the airmass factor variability. The third step is adapted to strong lines and corresponds to a forward modeling. We assume that the data, after division by the already well determined atmospheric transmission model, provide the telluric-free continuum outside the strong lines' centers, and that the interpolated polynomials may represent at first order the telluric-free data elsewhere. We then performed a series of fits to the data of convolved products of this adopted continuum by a transmission model, masking the interstellar K I regions. We let the Doppler shift free to vary along the spectral interval, then, in a final stage, we allowed for a variation of the LSF width. We found that this last stage is very important since the two lines are at very different locations along the echelle order, the first one being very close to

the short wavelength limit of the order, and this results in a significant difference reaching a 30% relative variation of the LSF width. We tested the whole procedure for a unique echelle order and for adjacent orders merging. We found that the rebinning and the interpolations required during the order merging as well as the resulting non-monotonous LSF wavelength variation introduce strong residuals at the locations of the deep lines and make the correction more difficult. For this reason, we have chosen to restrict the analysis to one order. After final convergence, we stored the wavelength-dependent LSF width and, importantly, the adjusted transmission model before its convolution by the LSF.

This method was applied to all stars. A different, dedicated TAPAS model was used for each observatory. An example of telluric model adjustment is represented in Figures 1.1a and 1.1b. The fitted TAPAS transmission before entrance in the instrument is displayed on top of the data in Figure 1.1a. At the end of the process, if we convolve this transmission by the LSF and divide data by the resulting convolved model, we obtain a partially corrected spectrum, represented in Figure 1.1b .

There are still narrow spikes at the locations of the strongest lines (see e.g., the first pair of O_2 lines in Figure 1.1b). This is unavoidable in the case of deep lines since the flux in the data and model is close to zero. However, because we did not use the result of the division, but only kept the best telluric model as an ingredient to enter the K I profile-fitting, we are not impacted by such spikes, which is the main advantage of the method. There is an additional advantage: Dividing data by a convolved telluric transmission profile is not a fully mathematically correct solution because the convolution by a LSF of the product of two functions is not equivalent to the product of the two functions separately convolved. The latter computation gives an approximate solution, however, only if all features in the initial spectrum are wider than the LSF. This is not the case in the presence of sharp stellar or interstellar features, and this may have an impact on the accuracy of the results. We note that this type of method can be used in exo-planetary research for the removal of telluric contamination of the spectra obtained from ground-based facilities.

1.3.3 Interstellar K I doublet extraction

As a result of the telluric correction procedure, two files were produced for each target, the first one being the fitted TAPAS model, before convolution by the LSF. The second contains the LSF width as

a function of wavelength. We used the second to extract the two separate LSFs appropriate for each K I line (7665 and 7700 Å). After this preparation, we then proceeded to the K I doublet profile-fitting in a rather classical manner (see, e.g., Puspitarini and Lallement, 2012), except in two ways: (i) the inclusion of the telluric profile, now combined with Voigt profiles before convolution by the LSF; and (ii) the use of two distinct LSFs.

The profile-fitting procedure starts with continuum fitting for the two transitions, followed by fitting of normalized data to the convolved products of Voigt profiles and the telluric model by means of a classical Levenberg-Marquardt convergence algorithm in both K I line areas. The number of clouds is increased, if it appears necessary. Here, we have imposed a maximum Doppler broadening (i.e., combination of thermal broadening and internal motions) of 2.5 km s^{-1} for each cloud. This arbitrary limit was guided by the spectral resolution of the data, the spatial resolution of the present 3D maps of dust, and our limited objectives in the context of this preliminary study (see below). As is well known, the advantage of fitting a doublet instead of a single line is that more precise cloud disentangling and velocity measurements can be made since there must be agreement between shifts, line depths, and widths for the two transitions and therefore tighter constraints are obtained. An advantage of our technique is obtaining maximal information from the doublet regardless of the configuration of telluric and K I lines. An example of profile-fitting results is presented in figure 1.2. All results are summarized in the annexed article.

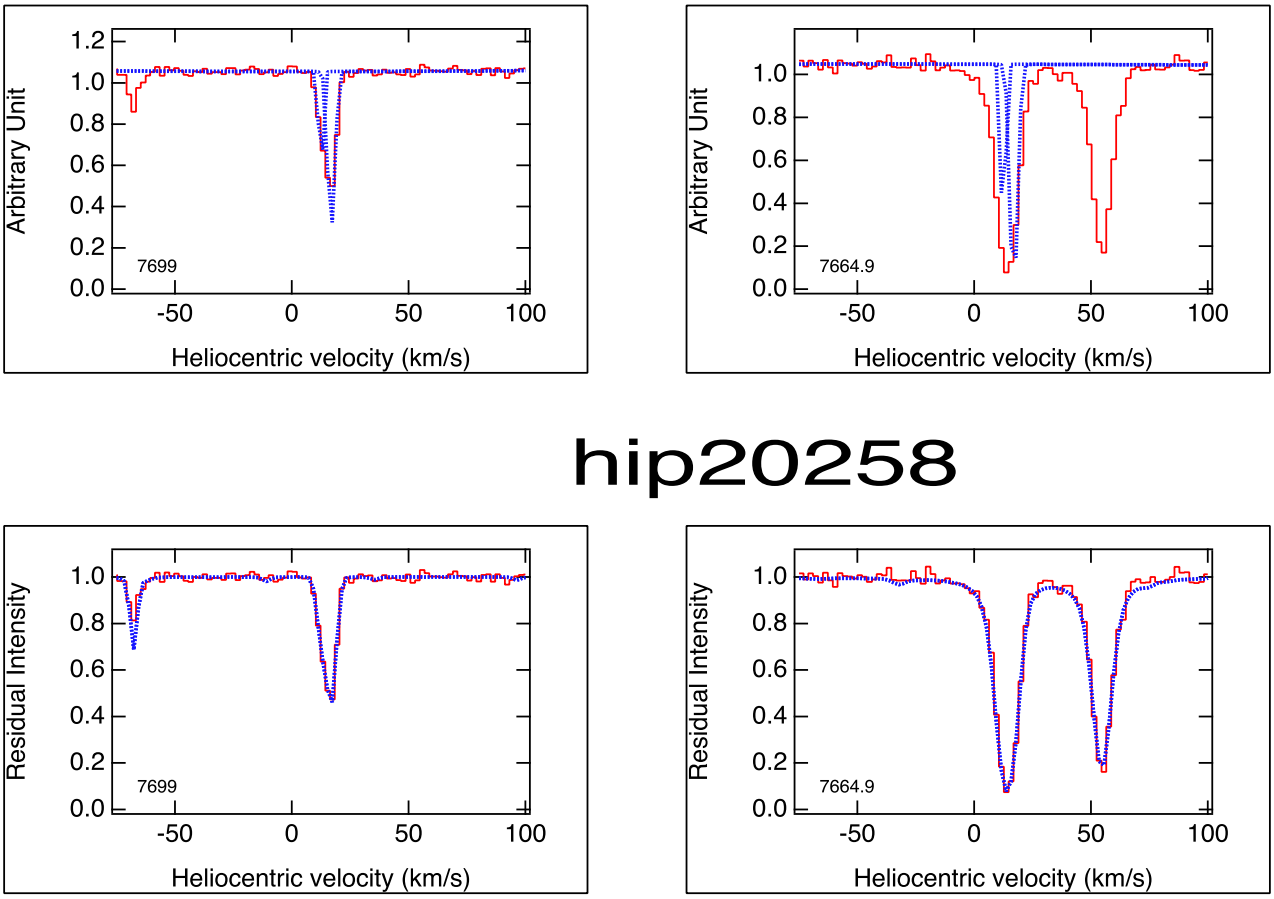


Figure 1.2: Example of K I doublet dual interstellar-telluric profile-fitting: star HIP 20258. The Pic du Midi-NARVAL spectra are shown around the 7699 (resp. 7665) Å line on the left (resp. on the right). The wavelength scale has been converted into radial velocity in the frame of each transition. The top graphs show the fitted interstellar model before convolution by the instrumental profile (blue line), superimposed on the raw data. The bottom graphs show the adjusted interstellar-telluric model (blue line) superimposed on the normalized data.

1.3.4 Illustration of 3D kinetic study

We show in Figure 1.3 an image of the dust clouds in a vertical planes containing the Sun and oriented toward Galactic longitudes l=157.5° . It is a planar cut within the 3D dust distribution. It is important to keep in mind that the mapped structures have a minimum size of 10 pc, which prevents detecting smaller clumps. We selected the target stars whose longitudes are less than 1.3 degrees from the longitude of the plane and we plotted the projection of their line of sight on the plane. We then compared their observed K I absorption velocities as well as their associated approximate columns (in units of 10^{10} cm^{-2}) and the dense structures encountered along their trajectories. As we have already pointed out, this exercise would take too much time in the case of more numerous data and the results are partially arbitrary, but we want to illustrate here what can be done based only on a preliminary

visual method. Work is underway to develop automated techniques.

To help visualize which star corresponds to which velocity assignment, we have numbered them and the numbers correspond to the objects listed in the text included in each figure. The velocity assignments to the dense clouds are indicated by arrows pointing outward (inward) for positive (negative) radial velocities. The length of the arrow is proportional to the velocity modulus. Two or more velocities at very close locations in the map indicate that information on the same cloud is provided by different stars. In some cases the arrows are slightly displaced to avoid the superimposition of different measurements based on stars at very close latitudes. In all cases, differences between measured velocities on the order of 1 kms^{-1} and sometimes up to 2 kms^{-1} may be due to profile-fitting uncertainties and do not necessarily imply a different clump.

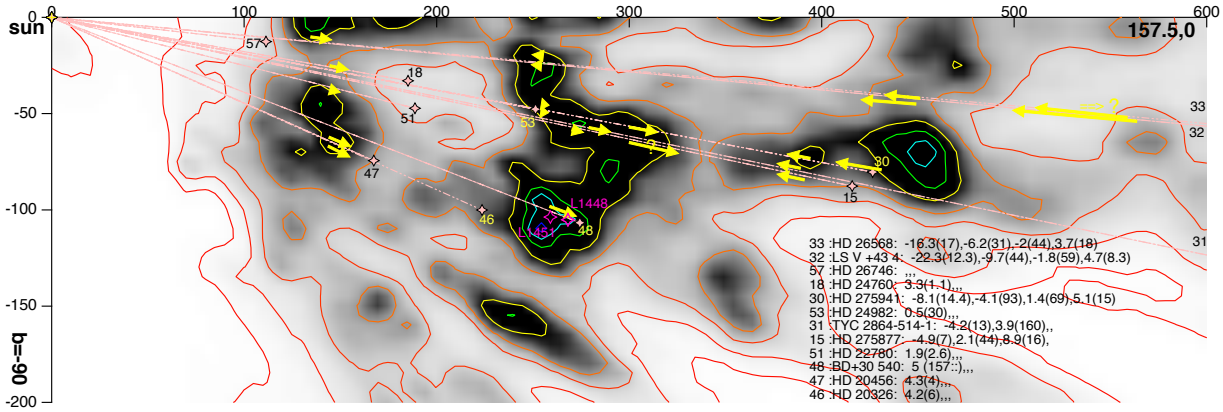


Figure 1.3: Image of the inverted interstellar dust distribution in a vertical plane containing the Sun (located at 0,0) and oriented along the Galactic longitude l=157.5° . The image is extracted from the 3D distribution of Lallement et al. (2019). Distances are in parsecs. The quantity represented in black and white is the differential extinction. Iso-contours are shown for 0.0002, 0.00045, 0.001, 0.002, 0.005, 0.01 , 0.015, and 0.02 mag per parsec. The paths to the target stars whose longitudes are within 1.3 degrees from the longitude of the represented plane are superimposed. The stars are numbered as in the text and drawn at the bottom of the figure; additionally, the Doppler velocities and approximate columns of absorbing K I in 10^{10} cm^{-2} units are listed for each target (in parentheses after velocities). Stars' numbers are indicated in yellow or black. The list of targets is given by decreasing latitude. For distant targets falling outside the image, their numbers are indicated along their paths at the boundary of the figure. The preliminary velocity assignments to the dense clouds are indicated by arrows pointing outward (inward) for positive (negative) radial velocities. The length of the arrow is proportional to the velocity modulus.

1.4 Telluric correction of cooler stars (spectral class from F to M).

Exoplanets are searched for around cooler stars than those for which a correction was made in Section 1.3.Therefore, it is necessary to develop an algorithm that can perform the correction in the presence of a large number of stellar spectral lines. The method which was developed and will be described can be classified as Type 3 following order in 1.1. We are using a TAPAS model of the atmospheric transmission and ESPRESSO data. A short description of ESPRESSO instrument is in subsection 1.4.1 and the description of the dataset can be found in the subsection 1.4.2.

1.4.1 Instrument ESPRESSO/VLT

ESPRESSO (Echelle SPectrograph for Rocky Exoplanets and Stable Spectroscopic Observations) is a fiber-fed, cross-dispersed, high-resolution echelle spectrograph installed at the VLT in Paranal, Chile. ESPRESSOs optical scheme is a "white pupil" (more about the "white pupil scheme" is in Section 7.1) but with some modifications which are explained in Pepe et al. (2021). ESPRESSO has a pupil slicer, which leads to the reduction of echelle grating size, without loose of spectral resolution. It is important to mention that ESPRESSO has two different camera arms a red and a blue. There are two individual CCDs to record the blue and red spectra separately, they are temperature stabilized separately in their individual cryostats. Two fibers illuminate the spectrograph: one is dedicated to the target and the other one carries the light from a calibration source (either for flat-fielding or wavelength calibration or drift measurements), or the light from the sky background (to be subtracted from the stellar spectra, for weak stars, or moonlit nights). The wavelength calibration is done by a combination of Th-Ar lamp and white-light illuminated Fabry-Perot. There is a plan for replacement of Th-Ar plus Fabry-Perot by a LFC (Laser Frequency Comb).

ESPRESSO is placed in a vacuum vessel which is located inside a thermally controlled room. Such a configuration should minimize any temperature and pressure fluctuations: the index of refraction is 1, and constant, so there should be almost no drift associated to changes of the index of refraction .

The overall optical scheme is sketched in Figure 1.4 (from the ESPRESSO User manual P102 at https://www.eso.org/sci/facilities/paranal/instruments/espresso/).

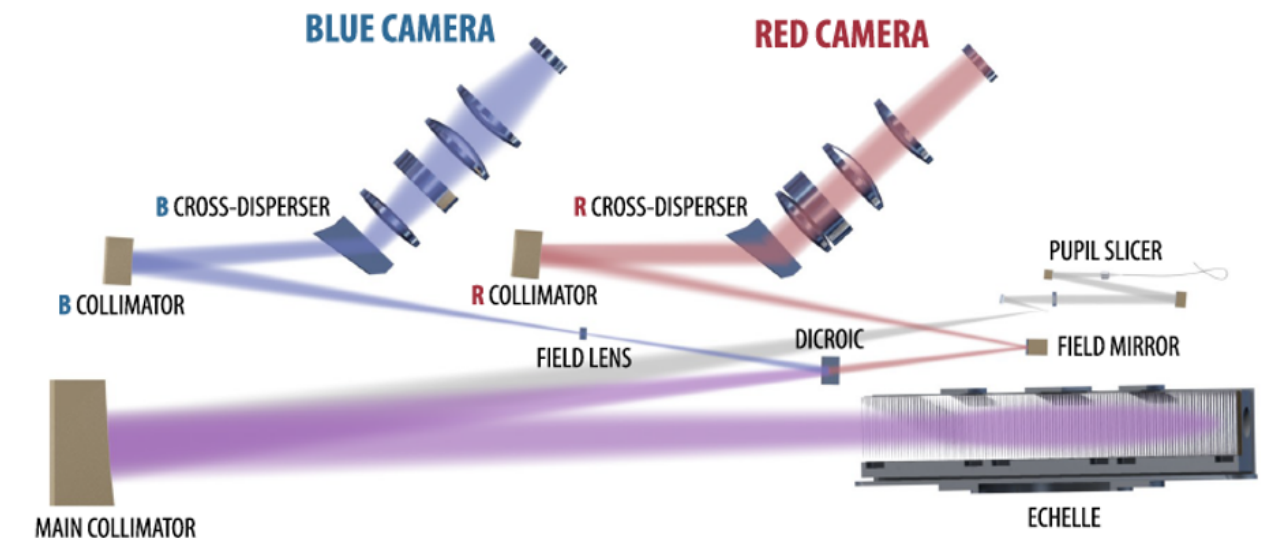


Figure 1.4: Layout of the ESPRESSO spectrograph and its optical elements (from ESPRESSO User Manual version 1.0, Molaro et al, 2018).

The operating wavelength range of ESPRESSO is from 3782 to 7887 Å. The resolving power in the HR (high resolution) mode is around 140 000, and more than 190 000 in the UHR (ultra high resolution). We have used data only from the HR mode. The resolving power is variable across the detector. From the color-code figure 11 of Pepe et al. (2021) we have built a numerical function giving the resolving power as a function of the ESPRESSO order and pixel number, and Figure 1.5 is a color-coded image of this function. On this figure 1.5, we see that the blue arm orders (from 0 to 89) contains 6540 pixels , while the red arm (orders 90 to 169) contain 9111 pixels (actually , "spectels", but we will use "pixels" instead most of time in this document.) Note that what is called one order here is NOT the grating interference order, but rather a reference number. The fiber image-slicer gives two images of the same spectral region; one even reference order number n and the following order n+1 have the same wavelength coverage. The values of central wavelength of order as a function of order number is shown on Figure 1.6. The transition in wavelength coverage between blue and red arms of ESPRESSO is shown on Figure 1.7.

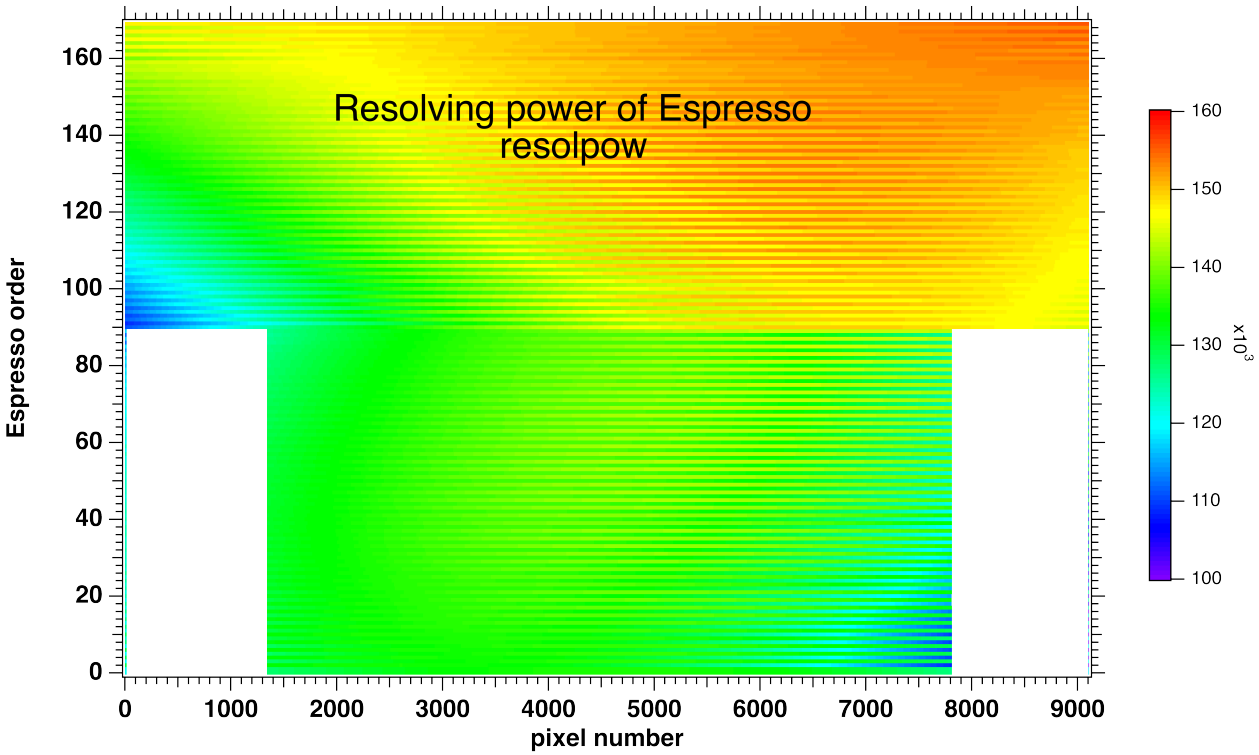


Figure 1.5: Color-coded map of the resolving power of HR mode of ESPRESSO. The map was obtained by digitizing the figure 11 of Pepe et al. (2021) (both image and color scale). We used this matrix of resolving power to compute accurately the spectrum of the atmospheric transmittance as seen with ESPRESSO, by the (pseudo) convolution of the TAPAS returned transmittance with this map of the PSF (or ILSF, Instrument Line Spread Function).

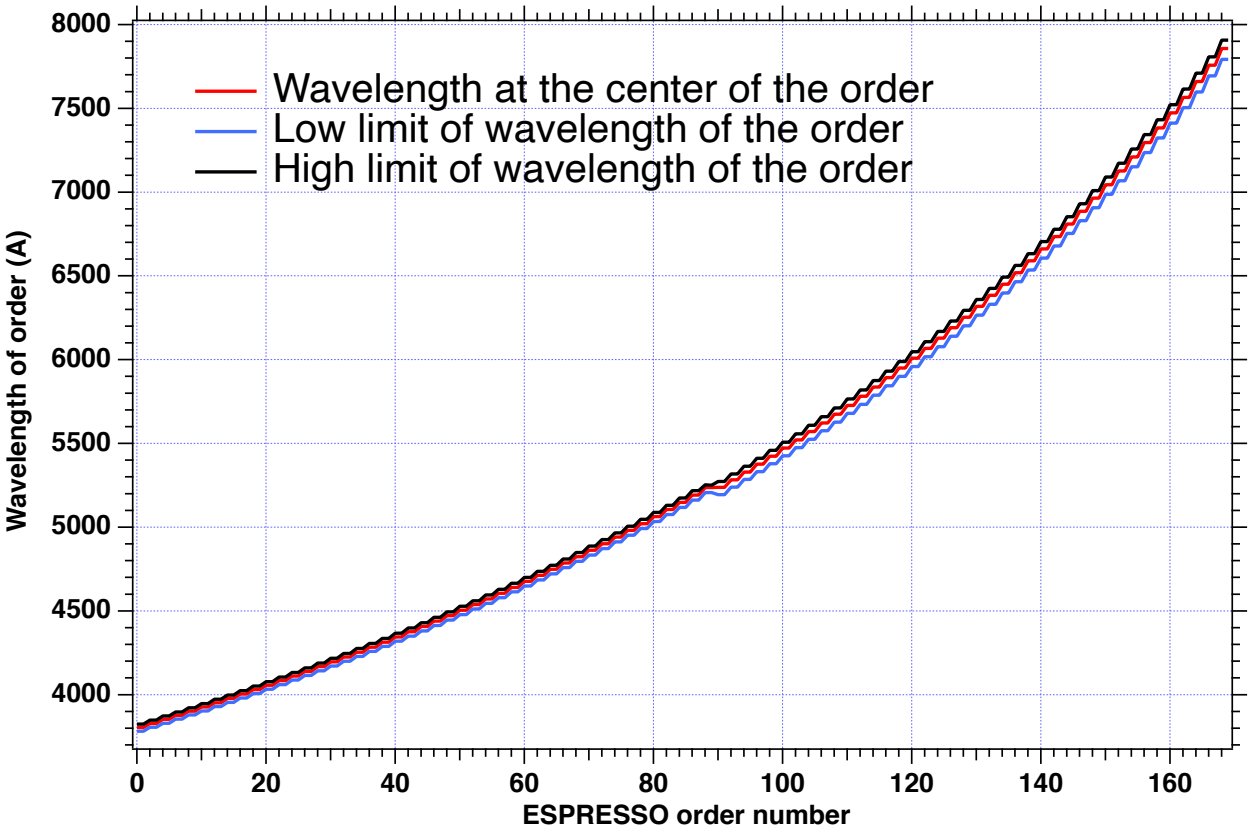


Figure 1.6: The red curve is the central wavelength (in vacuum) of each ESPRESSO order as a function of the order number. The black and blue curve are (respectively) the higher and lower wavelength limits of each order.

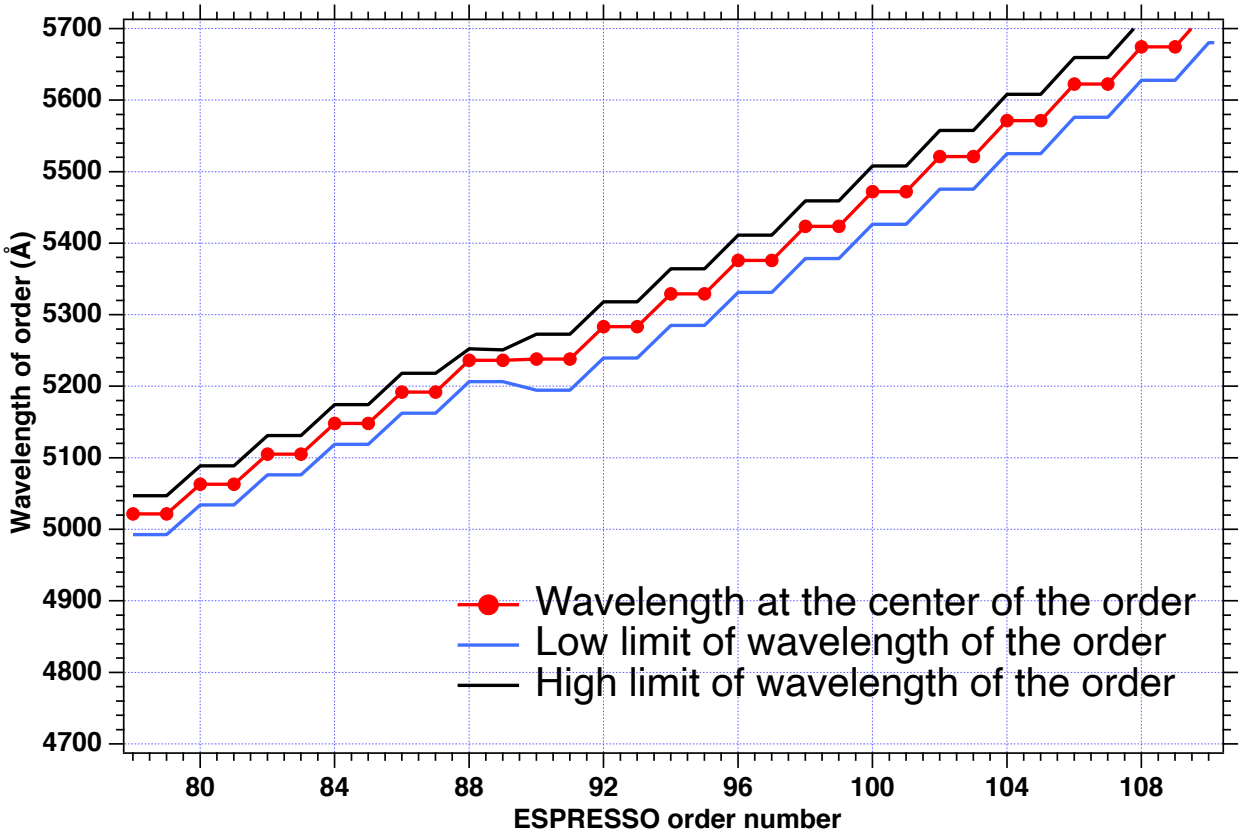


Figure 1.7: Same as figure 1.6. Zoom on the region around the transition between blue arm detector (ending at orders 88-89) and red arm detector (starting at orders 90-91. Two consecutive orders, having an even and an odd number, cover the same wavelength interval, thanks to the image slicer. There is usually some wavelength overlap between consecutive (even) orders.

1.4.2 Data

HD 40307 is a K2.5V type star with visual magnitude 7.147. It was observed during 5 nights from 22 December to 28 December 2018 by ESPRESSO. Several series of exposures to get spectra were obtained in HR mode which resolution varies between 100 000 and 160 000 across the spectrum. In December 2018 HD 40307 was a part of an asteroseismological campaign[1], 1151 measurements were obtained, the exposure time was set to 30 seconds, and the sampling period was about 78 seconds.

From exoplanet.eu we know that star HD 40307 is a multi-planetary system. It was reported by Mayor et al. (2009) that HD 40307 hosts 3 super-Earth-type planets, with orbital periods P_b=4.311 days, P_c=9.6 days and P_d=20.5 days. Later, Tuomi et al. (2013) independently analyzed HARPS publicly available data and concluded to the presence of 3 additional exoplanets with periods: P_e=34.62 days, P_f=51.76 and P_g=197.8. Díaz et al. (2016) confirmed the 3 planets found by Mayor et al. (2009) and planet f from Tuomi et al. (2013), but doubted the presence of planets e and g. The most recent

[1]Prog.ID:0102.D-0346(A); PI: Bouchy

analysis made by Coffinet et al. (2019) confirm the findings of Díaz et al. (2016): they found the 3 exoplanets from Mayor et al. (2009) and also the additional planet f from Tuomi et al. (2013), but as well as Díaz et al. (2016) they doubt the presence of planets e and g. After subtraction of a long-term signal, which probably represents a magnetic cycle of the star, the signal around the period of 200 days has a very low significance and the signal around 34.6 days is absent. We used data from Coffinet et al. (2019) because they corrected the raw data from a HARPS wavelength calibration problem, linked to the presence of a small gap between two adjacent sub-blocks into several blocks of the CCD detector (stitching). Each time a stellar line goes over this gap, due to Earth velocity excursion, it induces a large (spurious) shift of RV retrieval for this line. Coffinet et al. (2019) corrected HARPS data from this effect and re-determined the planetary parameters of seven stars, including HD 40307. Therefore, the Bayesian analysis of Díaz et al. (2016) was done on HARPS data uncorrected from this effect, and the Coffinet et al. (2019) results on HD 40307 are at present the best estimates of the planetary parameters.

1.4.3 Developed method

We are using a TAPAS model of the atmospheric transmission for H_2O and O_2 (due to the ESPRESSO wavelength range these are the two main Earth's atmosphere gases) for the date of 25 December 00:24, location – Paranal observatory (an example of TAPAS transmission is shown on figures 1.8 and 1.9). For the preliminary adjustment of H_2O and O_2 absorption prior to telluric correction, we choose 3 main orders: one which contains numerous H_2O lines of various strengths and 2 which contain O_2 bands (one order contains the γ band and the other contains the B band).

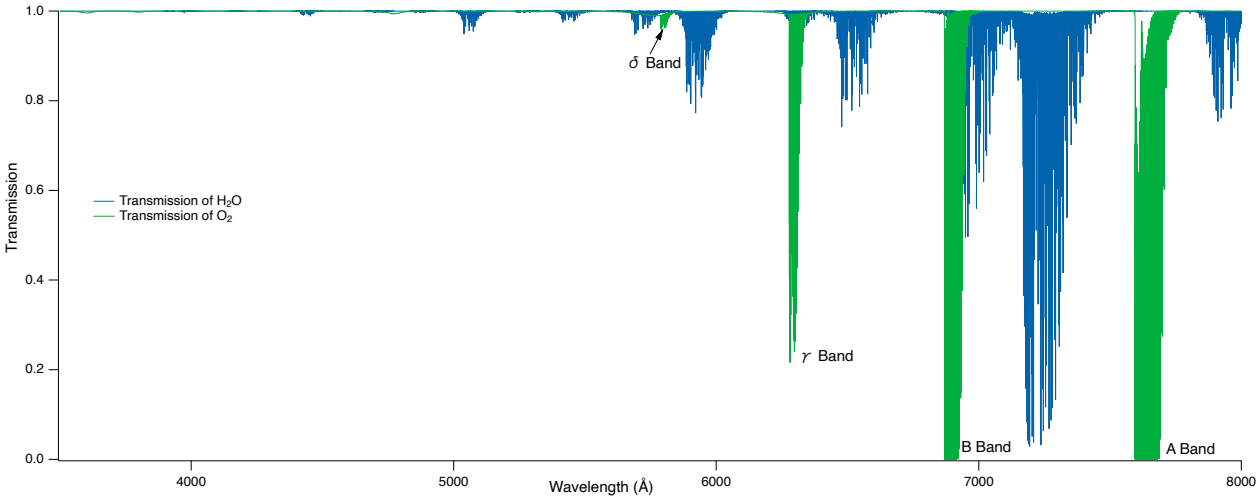


Figure 1.8: Transmission spectra provided by TAPAS of H_2O (blue) and O_2 (green) at the wavelength range of ESPRESSO (380 – 788 nm).

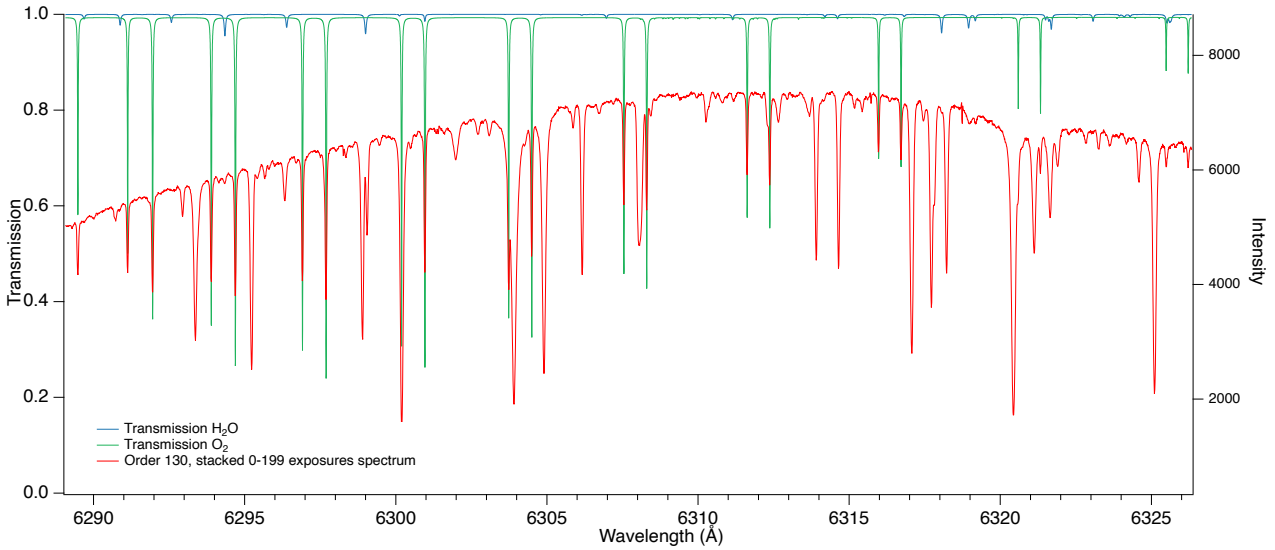


Figure 1.9: Transmission spectra provided by TAPAS of H_2O (blue) and O_2 (green) and observed data (red). Here order 130 is shown, which contains the O_2 γ band.

The correction method can be separated into 3 main steps, which are the followings:

- Preparation of the stellar mask. The objective is to select spectral regions in which there is only telluric absorption and to mask all parts in which stellar lines exist alone or are mixed with telluric lines. As a matter of fact, it is essential to eliminate the influence of stellar lines to obtain the actual quantity of gases in the atmosphere.

- Determination of the atmospheric H_2O and O_2 columns for each exposure. This procedure consists in fitting data in unmasked regions (mask from step 1) by a TAPAS model for the two species consecutively.

- Computation of the product of H_2O and O_2 transmission spectra for each exposure, not only for the orders used in steps 1 and 2, but for all orders contaminated by telluric absorption, using the results from step 2. The transmission spectra are convolved by the point spread function of ESPRESSO adapted to the order and to the wavelength in each order.

- Division of data by the result of step 3.

We now describe each point in more detail.

Stellar mask preparation

Telluric absorption is a problem for radial velocities techniques due to the fact that telluric features remain fixed in the instrument frame and there is a varying Doppler between them and the stellar lines. If tellurics contaminate stellar lines, they prevent a correct determination of the star radial velocity. But, on the other hand, stellar lines cause the same problem if one wants to model the telluric absorption – stellar lines can be mixed with telluric lines and this will prevent the accurate estimation of gaseous quantities.

In the case of hot stars, it is rather simple to correct spectra from telluric absorption, since this kind of stars almost have no stellar line or very broad and shallow ones. So, it is enough to have an Earth atmosphere transmittance model and the PSF of spectrograph. But the cooler is the star, the higher is the number of stellar lines, also deeper and narrower. In the case of HD 40307, the temperature of star is around 4980 K, and we are in the second case. However, thanks to the high resolution of ESPRESSO and to the low metallicity of star (-0.3) it is possible to find spectral domains in which a number of telluric lines are separated from stellar lines, allowing to create the stellar mask.

The mask is created on the basis of 3 spectra: a high SNR stellar spectrum obtained by stacking 10 consecutive exposures; a TAPAS spectral transmittance of atmospheric H_2O and O_2 computed for the date of the observations and the Paranal site; a stellar synthetic spectrum computed for the temperature, metallicity, gravity and rotation of HD 40307, using an ATLAS-SYNTHE environment provided by Piercarlo Bonifacio at GEPI (Sbordone et al., 2004) The visual inspection of the set of 3 spectra allowed us to construct the mask. It is done for each of 3 orders that we choose. This step could be made automated in the case of numerous target stars, but we started this way. An example of mask is shown on Figure 1.10. We can see how important it is to use a precise synthetic stellar spectrum, because of lines such as the one near 7178 Å, which is produced by a blend of telluric lines

and stellar lines. The mask is equal to 1 when there is a telluric absorption only and 0 if there is a stellar line, or a blend.

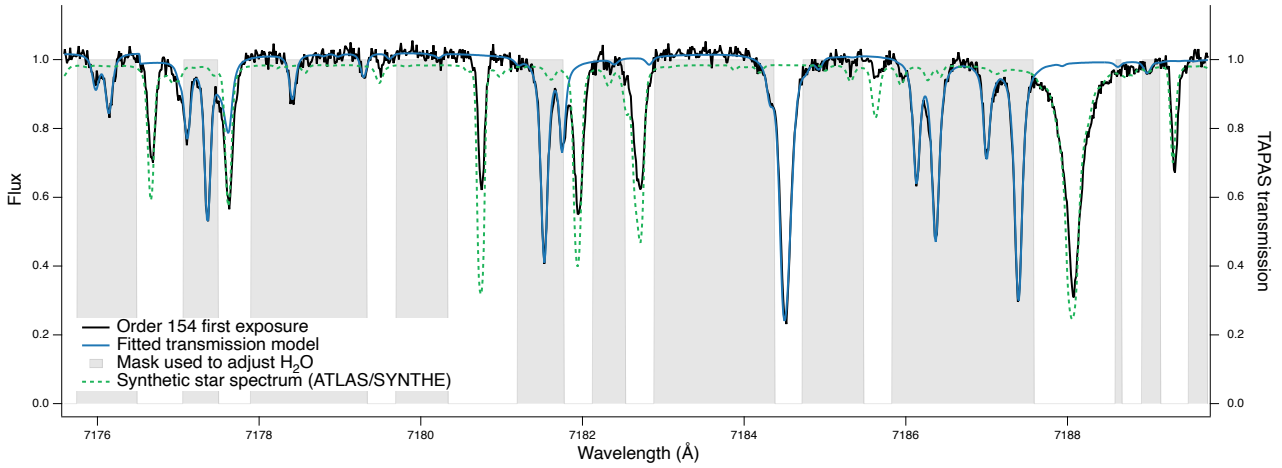


Figure 1.10: Stellar mask for ESPRESSO order 154, contaminated by water vapor. Black – original data; blue – TAPAS transmission model; green dashed line – synthetic star spectrum made by ATLAS/SYNTHE with a Castelli line list; the stellar mask is shown by grey bars.

Determination of H_2O and O_2 columns

Due to the target airmass evolution during the night and the temporal variation of atmospheric parameters and constituents, telluric lines may vary during the observation time. The second effect is particularly noticeable in the case of water vapor, which can vary on timescales of hours and even minutes, as it is shown on Figure 1.11. Therefore, even for a short period observation as the one we are using, it is important to measure temporal changes of H_2O and O_2 columns.

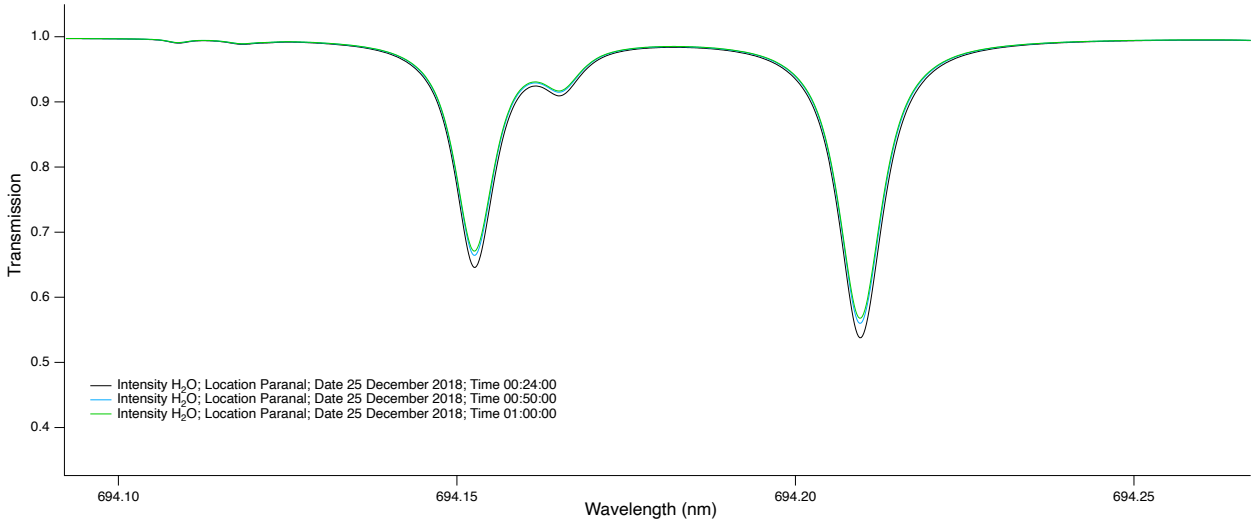


Figure 1.11: Demonstration of water vapor intensity variation with short time scale. All three spectra are TAPAS models for the same night, same observatory and same zenital angle, but for different times. The variation of intensity between the first model (black) and the last model (green) for a time difference of 36 minutes is 2%.

For the three chosen orders, each exposure was divided into chunks of 2 nm and each chunk was fitted in the unmasked regions by the product of the telluric transmittance, convolved by a Gaussian PSF, and a quadratic polynomial representing the stellar continuum. During the fit the free parameters are the columns of H_2O or O_2 (depending on the order) and the coefficients of the polynomial representing the stellar continuum. According to Bertaux et al. (2014), the telluric transmittance varies as a power law $T^X(H_2O)$ or $T^X(O_2)$ where the X factor is proportional to the column of corresponding absorbing gas, T is the TAPAS transmission spectrum. During this step the resolving power was kept fixed and independent of the wavelength, its value was taken for each order from the Figure 11 from Pepe et al. (2021) (which was transformed into a color-map shown on 1.5, on Figure 1.12 spectral resolution as a function of pixel is shown for orders 130, 146 and 154). We used a Gaussian PSF since we did not find any evidence for a PSF profile departing from Gaussian, this will be discussed later. The result obtained for all chunks were averaged. This procedure is illustrated on Figures 1.15, 1.16 and 1.17.

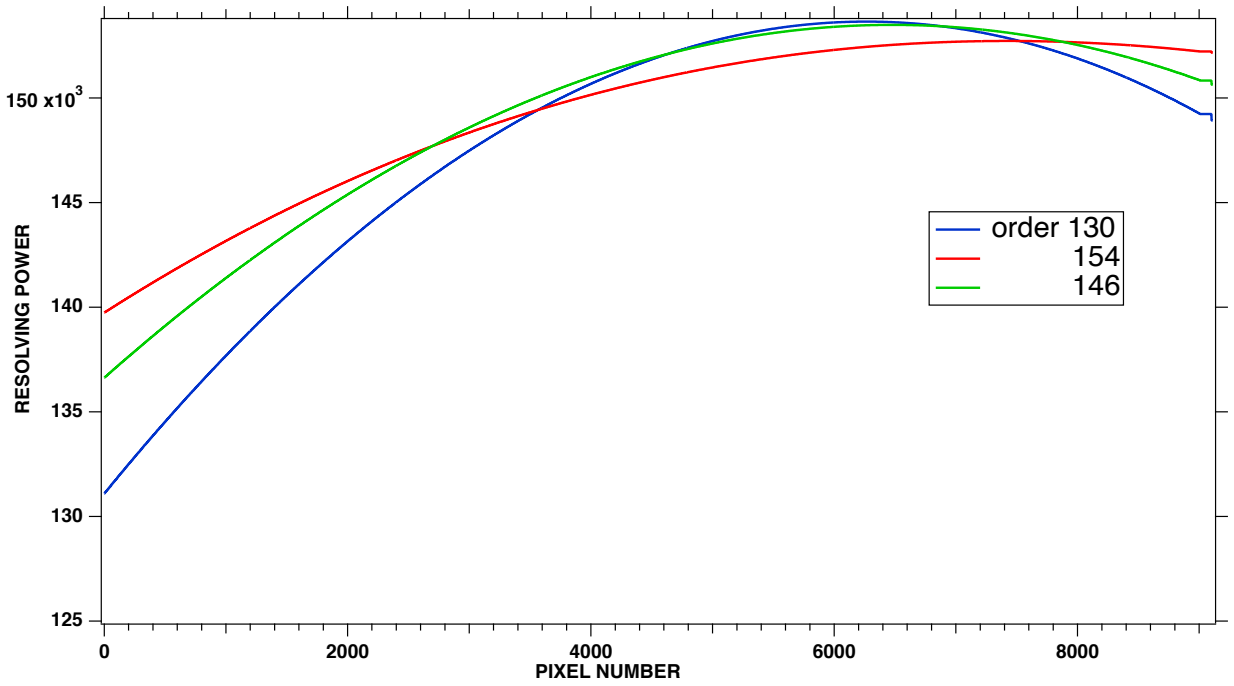


Figure 1.12: Spectral resolution as a function of pixel number, based on the digitization of Figure 11 of Pepe et al. (2021) Shown are three examples, for three orders of the red arm.

Evolution of the H_2O quantity We started with the TAPAS model requested for the first exposure of the night, 25 December 2018 00:24:00 GMT, Paranal observatory. The water vapor vertical column predicted by TAPAS does not correspond to the column deduced from the data, nor with the column estimated locally by the infrared radiometer installed at Paranal, which monitors continuously

the night sky. This radiometer-based column is included in the header of each observation and expressed in mm liquid equivalent. The difference between TAPAS and both the spectrum and the local measurement can be explained by the fact which we already mentioned – water vapor can vary a lot in a short time, and the ECMWF meteorological models used in TAPAS computations is refreshed only every 6 hours, so variations can be missed. In order to adjust the initial TAPAS model to the Paranal conditions we elevated the transmittance to the power of the ratio of the two values – i.e. the vertical column of H_2O measured at Paranal for the first exposure divided by vertical column of H_2O predicted by TAPAS, and set the X value equal to 1 for a reference spectrum. We then fitted separately all exposures using X as free parameter.

The evolution of the fitted H_2O column (i.e. the product of the reference column estimated on site at the time of the first exposure and the fitted coefficient X) is shown on Figure 1.13 for the 200 exposures of the 25 December night. It is shown along with the estimated column deduced locally from the Paranal infrared radiometer. Both estimations have a similar amplitude of variation of about 15% and a similar time at which the column is at its minimum. The absolute levels of both estimations are also similar. During some time intervals, the values from the spectral adjustment differ slightly from the onsite measurements. This can be explained by the sensitivity of the radiometric data to other parameters than the molecular column. Because the spectral measurements are more direct, we used for the following steps of correction for H_2O lines the fitted values.

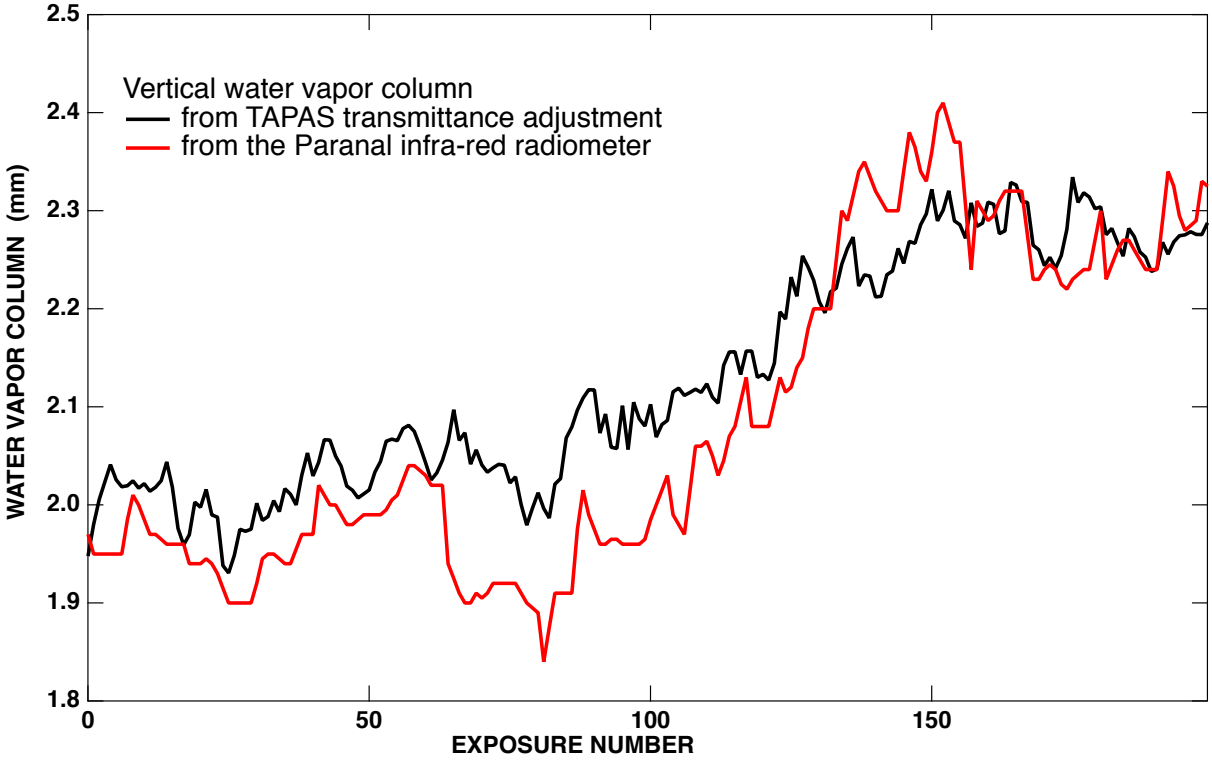


Figure 1.13: Comparison between the vertical column of water vapor adjusted through spectral fitting (black curve) and the column estimated from the infrared monitor at Paranal (in red) for the 200 consecutive exposures.

Evolution of the O_2 quantity The reference O_2 transmission spectrum for the first exposure is the TAPAS model, and a reference value X=1 was assigned to it for the fitting procedure. The retrieved O_2 column in the line of sight for the 200 exposures is shown on Figure 1.14. The results of two fits are displayed, one for the order 130 for containing the O_2 γ band near 630 nm, and one for the order 146 for containing the O_2 B band near 688 nm; the evolution of the airmass for all exposures is also shown. One can notice that both fitted values follow very closely the evolution of the airmass. The determination for γ band is more noisy than the one for B band, but the level of determination for γ band is equal to the expected level (X=1 for the first exposure, which means that the fitted column is exactly the TAPAS reference value for the first column of the exposure), while the determination based on the for B band has a small relative difference of about 3 percent. The reason for such small discrepancy is not obvious. It may be due to the lack of precision in the modeled PSF, or small deviations from the Gaussian approximation. Another possible reason could be the existence of actual differences between the modeled and actual vertical pressure/temperature atmospheric profiles. Finally, the last (least likely reason) could be some inaccuracy in the cross-sections in the HITRAN database. TAPAS is currently based on HITRAN 2016 database, but forseen to be upgraded to the

HITRAN 2020 (Gordon et al., 2022). It will be interesting to check if the same discrepancy between B and γ band of O_2 will persist. We decided to use the B band result for the absolute value of X for telluric-line removal, due to the fact that B band lines are stronger. For the temporal evolution we simply used the airmass as a multiplicative factor for the O_2 vertical column.

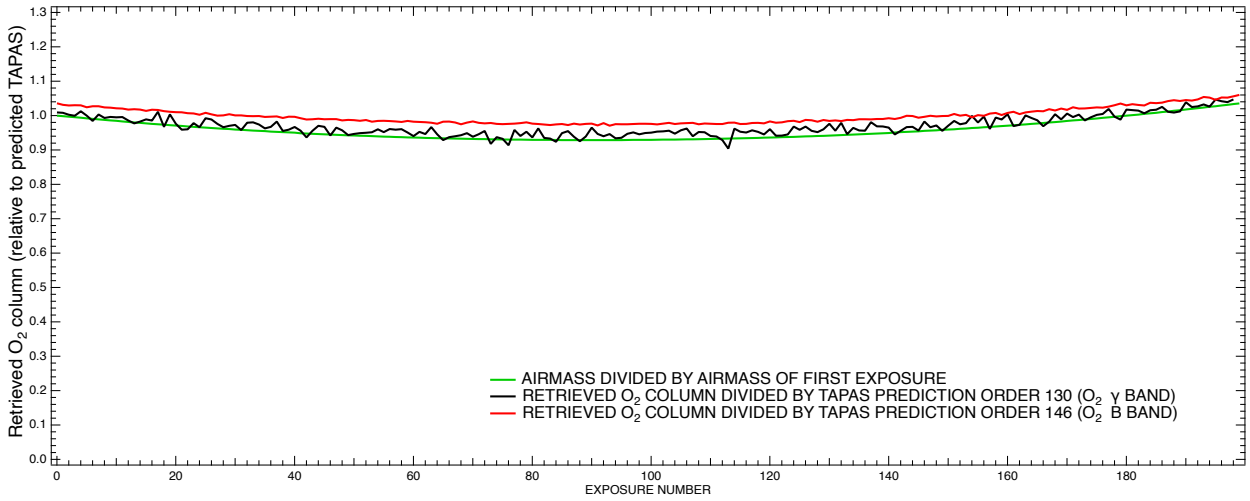


Figure 1.14: Comparison between the column of O_2 adjusted through spectral fitting of the B band (black curve) and the γ band (red curve) and the air mass (green curve). The retrieved columns of O_2 follow the air mass relatively well, which means that, unlike the H_2O variations, there is no significant variation of the O_2 vertical column during the time interval of the 200 exposures.

1.4.4 Correction of all spectra for telluric lines based on H_2O and O_2 adjusted columns

The final step of the method is the division of the whole wavelength range of each exposure data by a synthetic transmission computed on the basis of TAPAS, using for each exposure the adjusted H_2O and O_2 quantities.

To compute the most realistic synthetic model of the transmission we need to adapt it to the ESPRESSO resolution, which implies using an as accurate as possible PSF as a function of order and wavelength. To do so we used the Figure 11 extracted from the Pepe et al. (2021) paper displaying an image of the resolving power. We digitized the image and the color-scale, identified the series of blue and red orders in the image and attributed a resolution to each pixel of each order (shown on Figure 1.12). The point spread function (PSF) of the instrument was assumed to be Gaussian with a width corresponding to the resolving power determined for each wavelength and order from the digitized image. We assumed that the PSF has a Gaussian shape for two reasons: first, we found no mention of the opposite in the ESPRESSO manuals and descriptions; second, if it was not the case, i.e. the

PSF shape was different from a Gaussian, then we would find a repetitive pattern of residuals at the locations of the wings of all "corrected" lines, with a particularly prominent pattern for deep lines, and, indeed, we did not find such patterns.

We note that a way to retrieve the exact shape of the PSF would be a method of deconvolution of telluric lines using a hot star spectrum and a telluric model. Some results of the correction are shown on Figures 1.15 and 1.16 for H_2O and on Figure 1.17 and 1.18 for O_2, along with the telluric model and the mask.

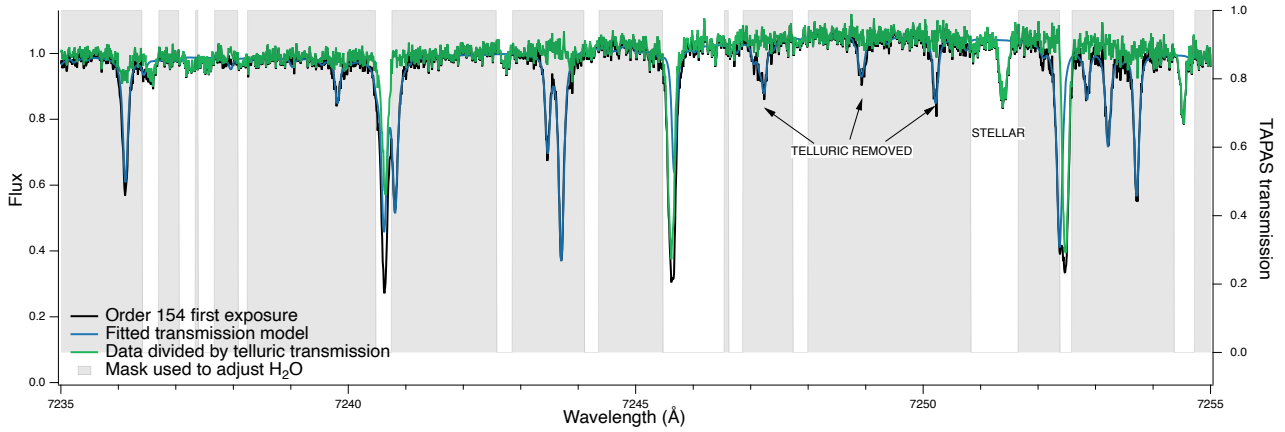


Figure 1.15: Illustration of the procedure used to determine the H_2O column for each exposure (here exposure 0 and order 154) and the telluric H_2O removal. The mask of stellar lines is shown as grey bars. The data and adjusted TAPAS transmittance are shown in black and light blue, respectively. The division of the data by the fit is shown in green. The correction works for isolated telluric lines and also for blended telluric and stellar lines (e.g., at 7240 Å; see figure 1.16)

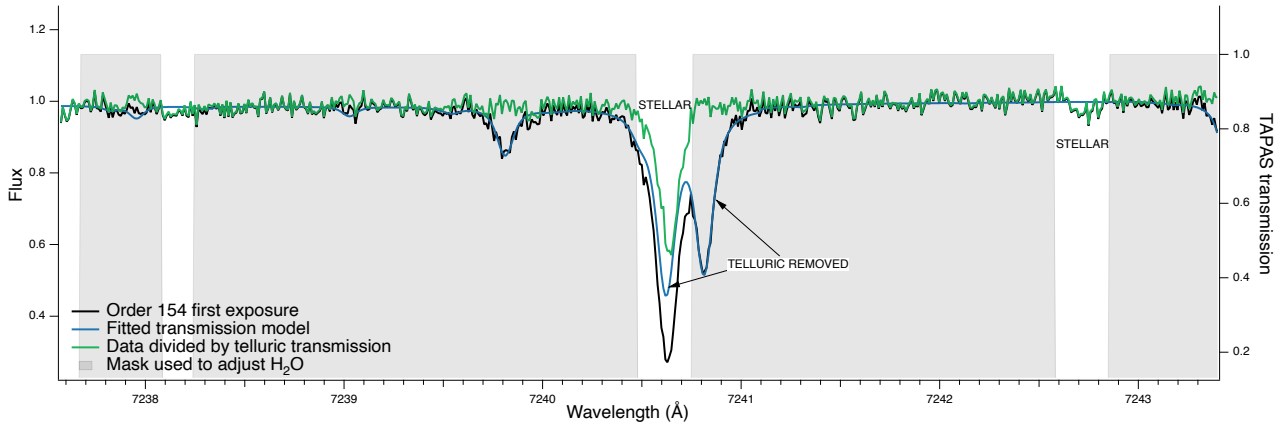


Figure 1.16: Zoom onto Figure 1.15 Even stellar lines heavily contaminated by H_2O lines may be kept for RV retrieval after correction from an adjusted telluric transmission. The mask of stellar lines is shown as grey bars.

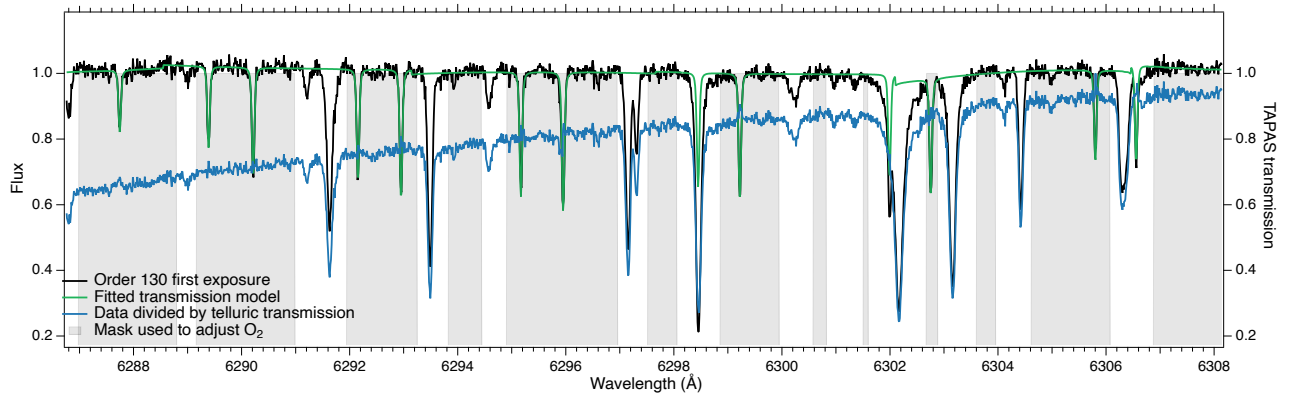


Figure 1.17: Illustration of the procedure used to determine the O_2 column for each exposure (here exposure 0 and order 130) and the telluric O_2 removal. The mask of stellar lines is shown as grey bars. The data and adjusted TAPAS transmittance are shown in black and green, respectively. The division of the data by the fit is shown in blue.

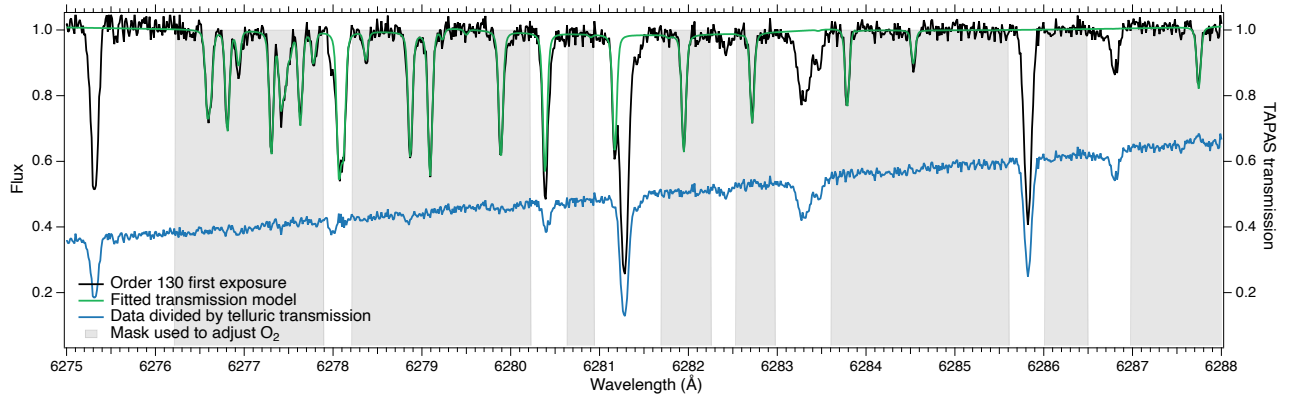


Figure 1.18: Illustration of the procedure used to determine the O_2 column for each exposure (here exposure 0 and order 130) and the telluric O_2 removal. The wavelength region 6277-6278 Å is an example of difficult to correct region, it is a beginning of γ band of O_2.The mask of stellar lines is shown as grey bars. The data and adjusted TAPAS transmittance are shown in black and green, respectively. The division of the data by the fit is shown in blue.

It should be noted that the correction is not perfect in every case, in particular there are residual peaks at the center of the strongest lines. After division of the observed spectra by the TAPAS transmission convolved by PSF there are residuals peaks at the center of the strongest O_2 lines. In such regions TAPAS lines before the convolution are saturated (transmittance = 0) and it is impossible to recover the actual stellar signal in this kind of spectral regions. The new binary mask, which is created for CCF procedure after telluric correction, should avoid lines which are near the center of such strong lines. An example of residuals is shown in figure 1.19.

We should point out that almost in all cases, these residuals peaks are well centered, and this indicates an excellent match between the HITRAN database wavelengths of lines and their actual measurements with the (precisely) calibrated ESPRESSO spectra. However, there are several cases of discrepancy

between the HITRAN prediction and the observed spectrum. The most notable one is at 7185.5 Å(Figure 1.20). It is most likely linked to an incorrect wavelength assignment of H_2O transitions. We note that it is not a single line, but a mixture of two or several lines. The same problem with correction of this line can be found in Allart et al. (2022), the authors mentioned that this line is excluded from the subset of lines due to the residual problem. After discussion with TAPAS team we found that TAPAS is currently using HITRAN 2016 database (Gordon et al., 2017), in Allart et al. (2022) authors used HITRAN 2016 as well, therefore in future such a problem should be eliminated by using HITRAN 2020 (Gordon et al., 2022).

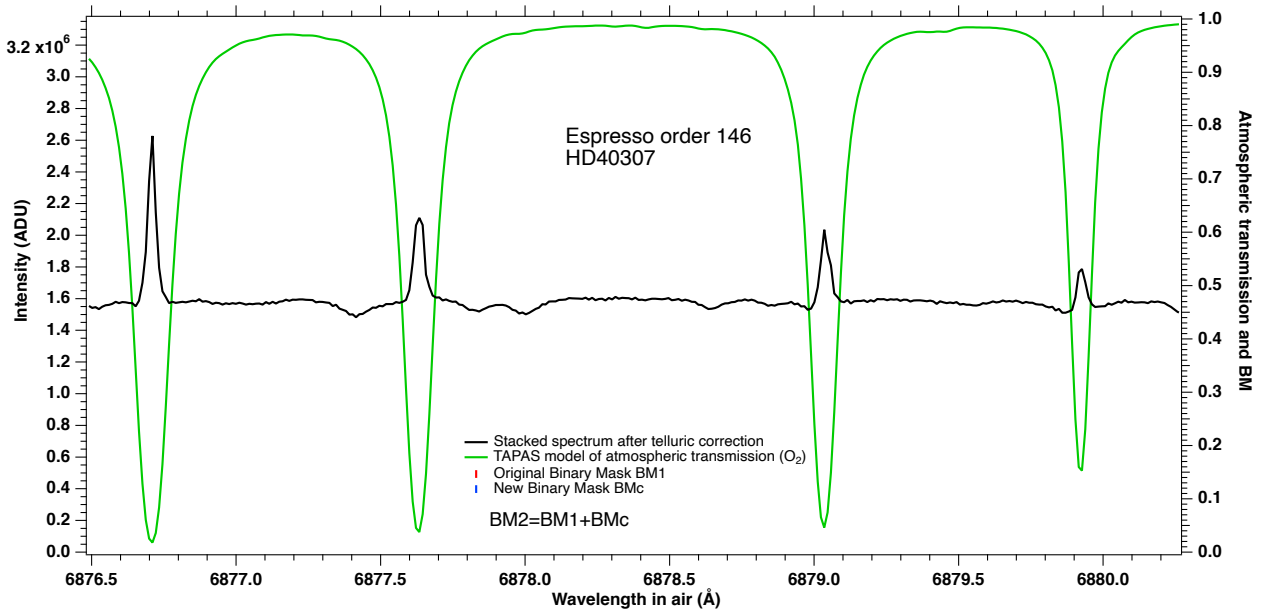


Figure 1.19: Example of residuals at the center of lines after division of data by TAPAS (convolved with PSF). The new binary mask for CCF avoids the proximity of such lines.

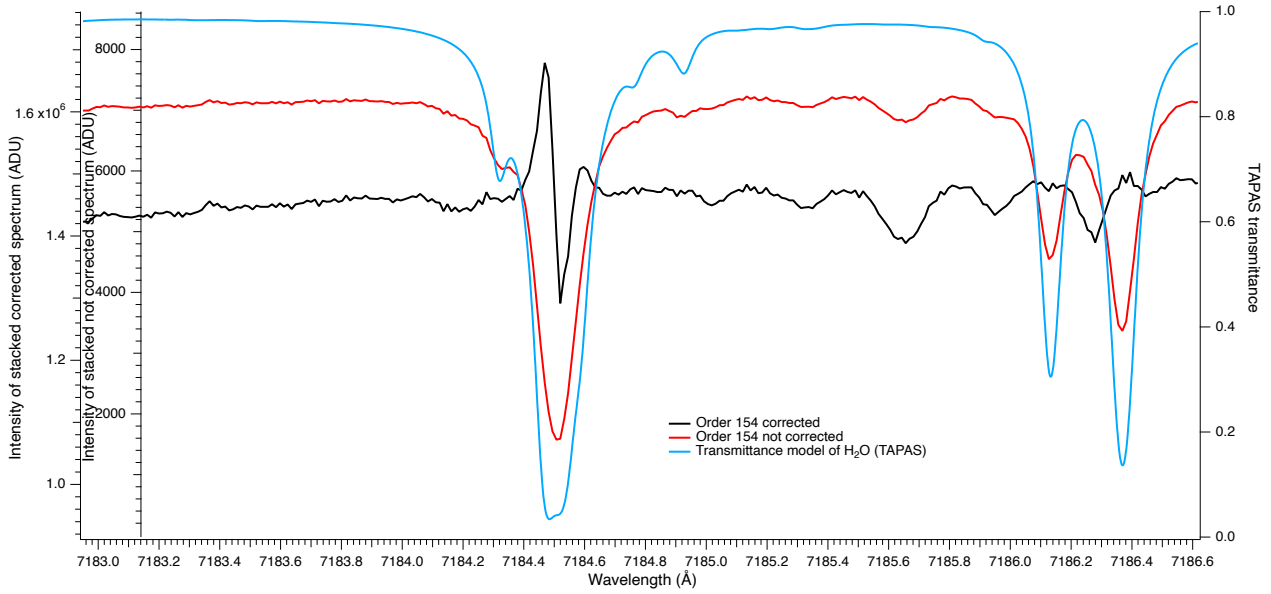


Figure 1.20: Example of the residuals (after telluric correction), which is caused by a potentially wrong wavelength assignment (or temperature coefficient) in HITRAN database. Red: the measured spectrum in order 154. Blue: Transmittance computed by TAPAS (right scale). Black: the spectrum after correction.

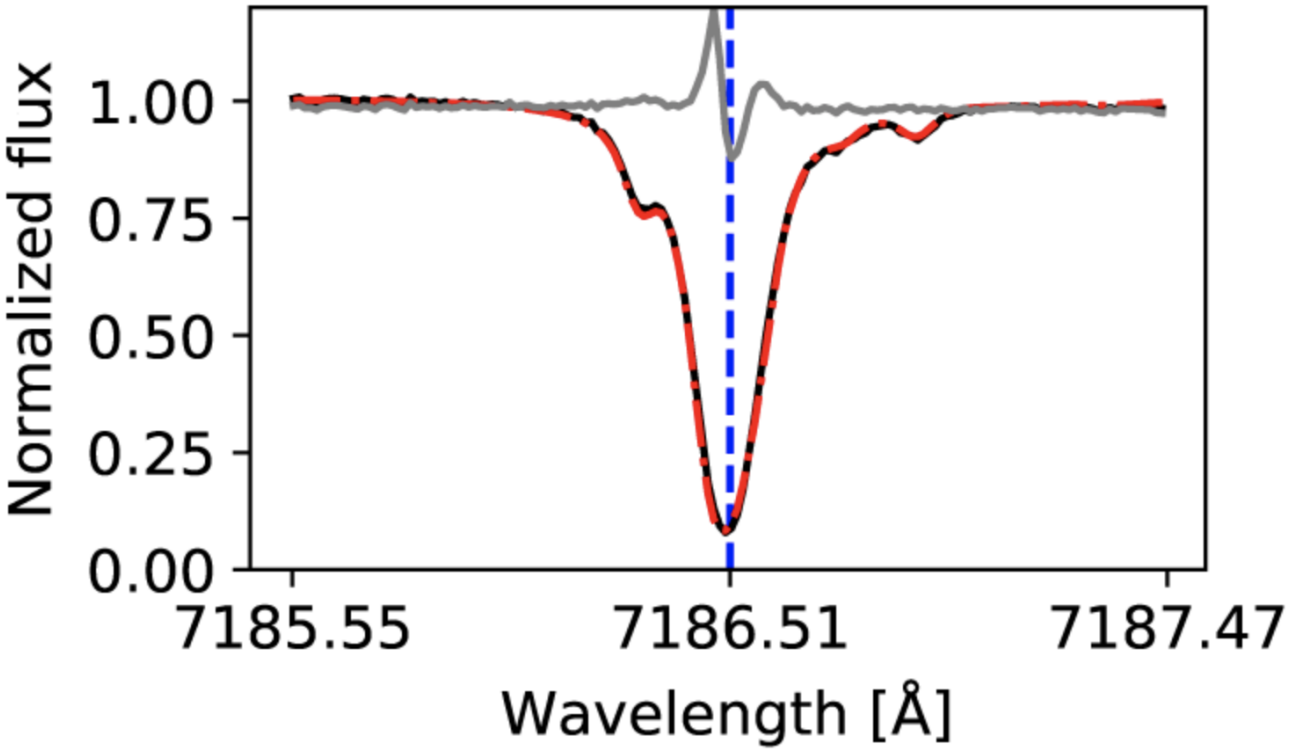


Figure 1.21: Same line as at 1.20, figure is taken from Allart et al. (2022) paper.

1.5 Radial velocity algorithm used for testing the telluric correction

We first performed the telluric corrections as previously described for all the 200 exposures of night 25 ($\sim$ over 4 hours). Then, we made our own RV retrieval with a classical CCF method, applied to both time series, original (no telluric correction, with the standard binary mask BM1 avoiding telluric contaminated spectral regions), and after telluric correction with an augmented mask including telluric corrected regions. Apparently the RV average signal over this lapse of time shows no global time trend: the perturbations due to planets are negligible. Still, fluctuations of RV may exist, due to granulation and oscillations. The oscillations seem undetected on this time series, see Appendix A).
Let us call δRV the error on one measurement of RV. Let us call F_{gbs} the amplitude of the apparent Doppler shift fluctuations, due to stochastic variations of the field of granulation. Let us call Sdev the standard deviation of a time series of RV measurements, which is a measure of the dispersion (or spread) over the 200 exposures. In the context of random Gaussian errors, it is likely that there is a quadratic combination of the independent phenomena (GBS fluctuations and shot noise/instrument errors on measurement), yielding the relationship:

$$S^2dev = \delta RV^2 + F_{gbs}^2 \tag{1.1}$$

The improvement on δ RV brought by telluric correction is therefore reflected on a decrease of Sdev. We note that the model of Meunier and Lagrange (2020) indicates a value of $F_{gbs} \sim 0.2$ ms^{-1} for a star of K4 type (HD40307 is K2.5), significantly smaller than Sdev.

Globally, the Sdev is significantly reduced for the RV time series of telluric corrected spectra. The results have been published in Ivanova et al. (2023), and this paper is attached to this manuscript.
We have implemented our own code for the classical CCF method, which is described in more details in Section 2.3.1. As a binary mask we have used the official ESPRESSO pipeline mask for K2 star (for HD 40307 a K2V star), which contains a list of laboratory wavelengths of spectral transitions, together with the so-called "contrast", which is an estimate of the expected relative depth of the corresponding star line.
When we are computing CCF by piling-up all CCFs for each line of one order, we implemented a selection of deepest mask lines by a contrast threshold. We choose only lines for which a contrast is larger than 0.2 (c>0.2). It is interesting to note how a CCF changes in correspondence to the contrast

threshold. Figure 1.22 illustrates how threshold conditions affect the number of lines which are used in CCF approach. For example, for order 104 the number of lines with c>0.5 is 8; c>0.2 is 30; c>0.1 is 40. Figure 1.23 illustrates the calculated CCFs for each threshold value. It is shown that the more lines are used - the higher is the intensity of CCF, the CCF based on more lines is narrower and the wings of CCF function are less noisy. On our side we decided finally to compute CCFs without any weights. At first we used the contrast as a weight, but after we switched to the no weight approach. We could not use the weights considering the width of the lines, since they were not available in the official ESPRESSO mask: only the contrast was included in the mask (since February 2023, the width of the line is also included in the official mask for ESPRESSO).

Also after the study of ESPRESSO official CCFs downloaded from ESO archive (http://archive.eso.org/cms.html),

we found that the pipeline produces CCF for each order and all single CCFs are summed up into one, which is actually used for RV determination by Gaussian fit. We did not follow such an approach, we fit each single CCF by Gaussian function and after RV_{order} values are summed up taking into account uncertainties, such approach allows us to take into account orders with more RV information (higher number of lines, higher SNR) and avoid orders which can spoil resulting RV (affected by instrument instability, low SNR).

RV values extracted from fits headers for each exposure of night 25 December of HD 40307 are shown on Figure 1.24. These values are produced by the official ESPRESSO pipeline, we have subtracted 31.3718 kms^{-1} from each value as a mean.

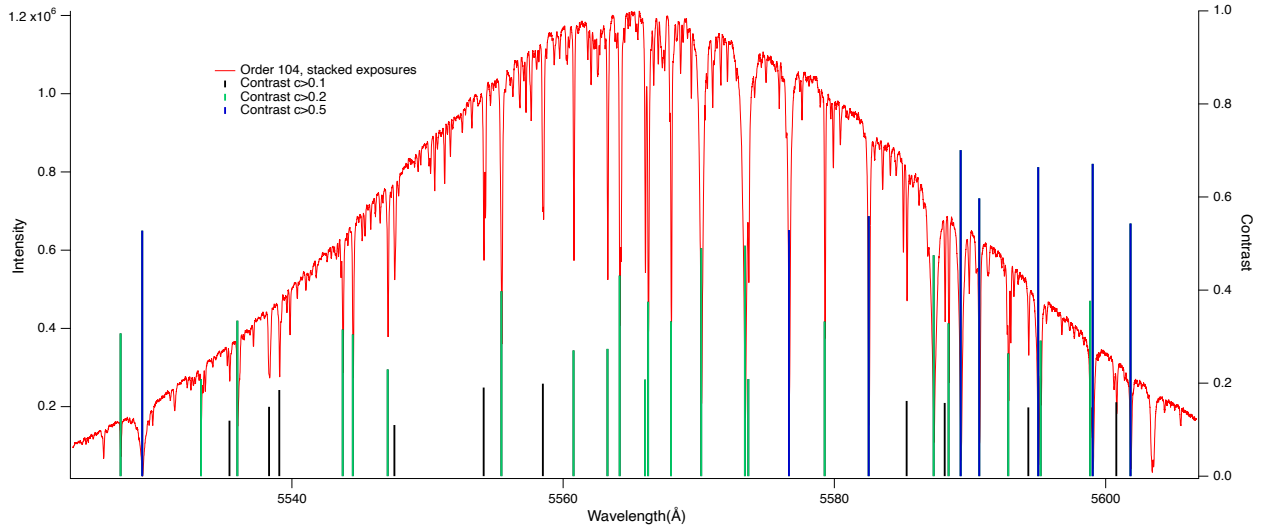


Figure 1.22: Example of mask change as response to the change of contrast threshold. Order 104 is not affected by telluric contamination (left scale); black – mask lines with contrast >0.1; green - mask lines with contrast > 0.2; blue – mask lines with contrast > 0.5.

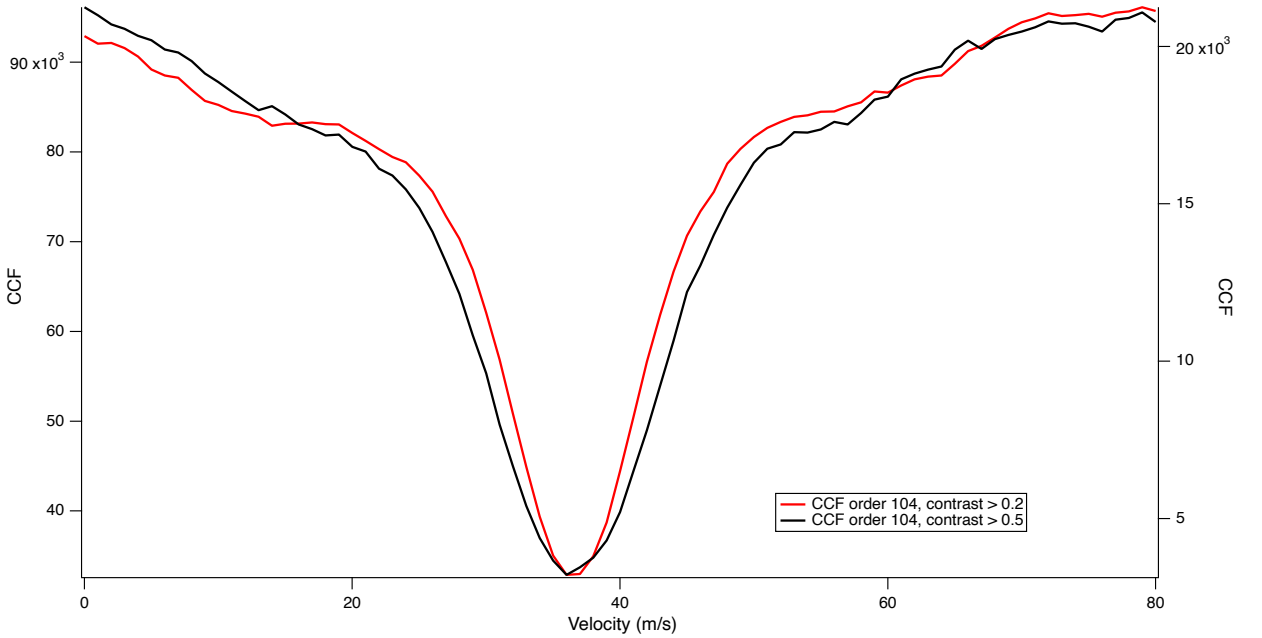


Figure 1.23: Example of CCF change in response to a change in contrast threshold. Red – CCF for order 104 with contrast threshold > 0.2 (left scale); black – CCF for order 104 with contrast threshold > 0.5 (right scale). See text for comments.

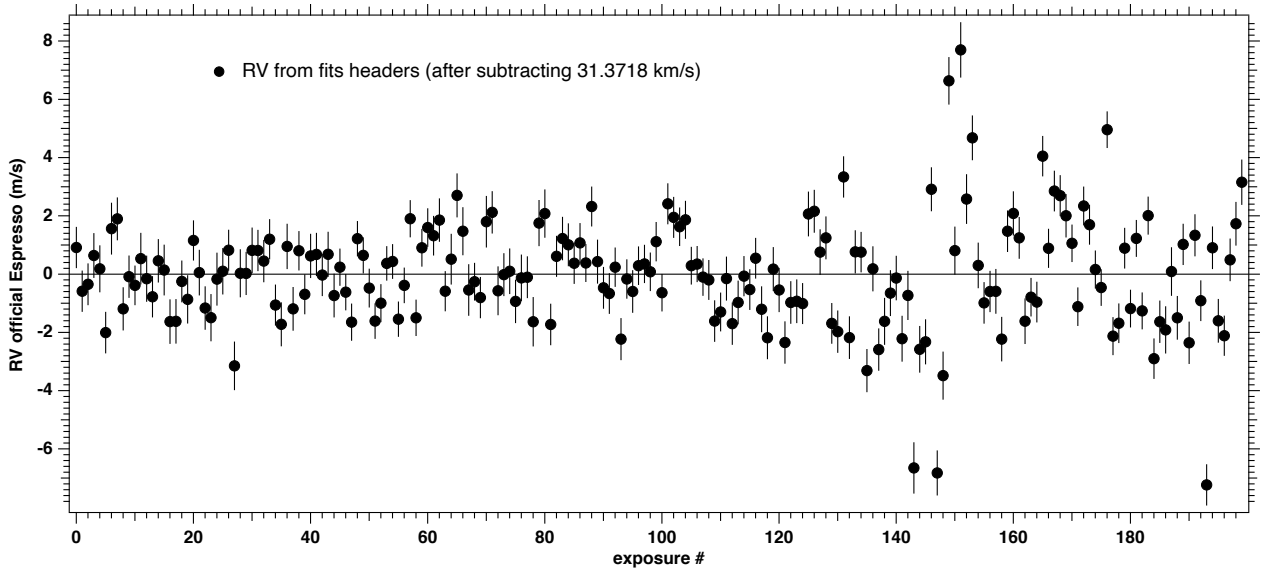


Figure 1.24: The official retrieved RV from ESO/ESPRESSO pipeline contained in the header of each exposure of night 25 December 2018, target HD 40307 with its error bar, in ms^{-1} and after the subtraction of the average of 31.3718 kms^{-1}. We note a substantial increase of the fluctuations after exposure $\sim$100, which is an instrumental effect (an instability of the cooling system of the blue arm detector).

The no-weight approach has the advantage that the content of each element of the CCF contains N_{e^-} independent electrons, and therefore the error (mainly photon shot-noise) is well known: $\sqrt{(N_{e^-})}$. At variance, as soon as a different weight is attached to each individual CCF, it is difficult to track the associated error due to shot-noise. At any rate, in order to estimate the gain in accuracy (or precision) when more spectral regions, corrected from tellurics, are used in the RV retrieval, it is enough to estimate the gain with a particular RV retrieval method. It is not needed that the method would be the absolute best. The gain is entirely due to the additional photons which become available when spectral regions, otherwise discarded because of telluric contamination, are used for RV retrieval.

1.5.1 Binary mask for new wavelength range.

After we made the telluric correction (see Section 1.4), a new range of wavelength starts to be accessible, for which no mask of stellar lines did exist yet. So, the first step is to produce a new mask for these new regions. To create the new mask we stacked 200 exposures of star HD 40307 corrected from telluric absorption, wavelength was assigned as for exposure number 100. For a set of stellar lines we used a Gaussian fit to determine the wavelength position of center (minimum) of line and depth (contrast), only lines with contrast > 0.2 were kept. Since all lines in mask have to be in the same reference system we have to shift them back to the rest frame. It is done by the back-shifting

lines by the RV_0=31381.572 ms^{-1}, this value is determined with order from 92 to 115 of the official mask (BM1). All new lines from telluric corrected stacked spectrum have been back-shifted by the same stretch factor. Example of back-shifting procedure is shown on Figure 1.25, it is done for the line which is in the official binary mask (Binary Mask 1, BM1), and found in the stacked spectrum as well. The black vertical line corresponds to the line in BM1, the blue vertical line corresponds to the position of line in stacked spectrum and the red vertical line corresponds to the position of line after back-shift. Positions of line in BM1 and after back-shifting match.

The number of lines in mask as a function of order is shown on Figure 1.26, a total of 724 new lines were added after telluric correction. New mask is called Binary Mask 2 or BM2 and is shown on Figure 1.27.

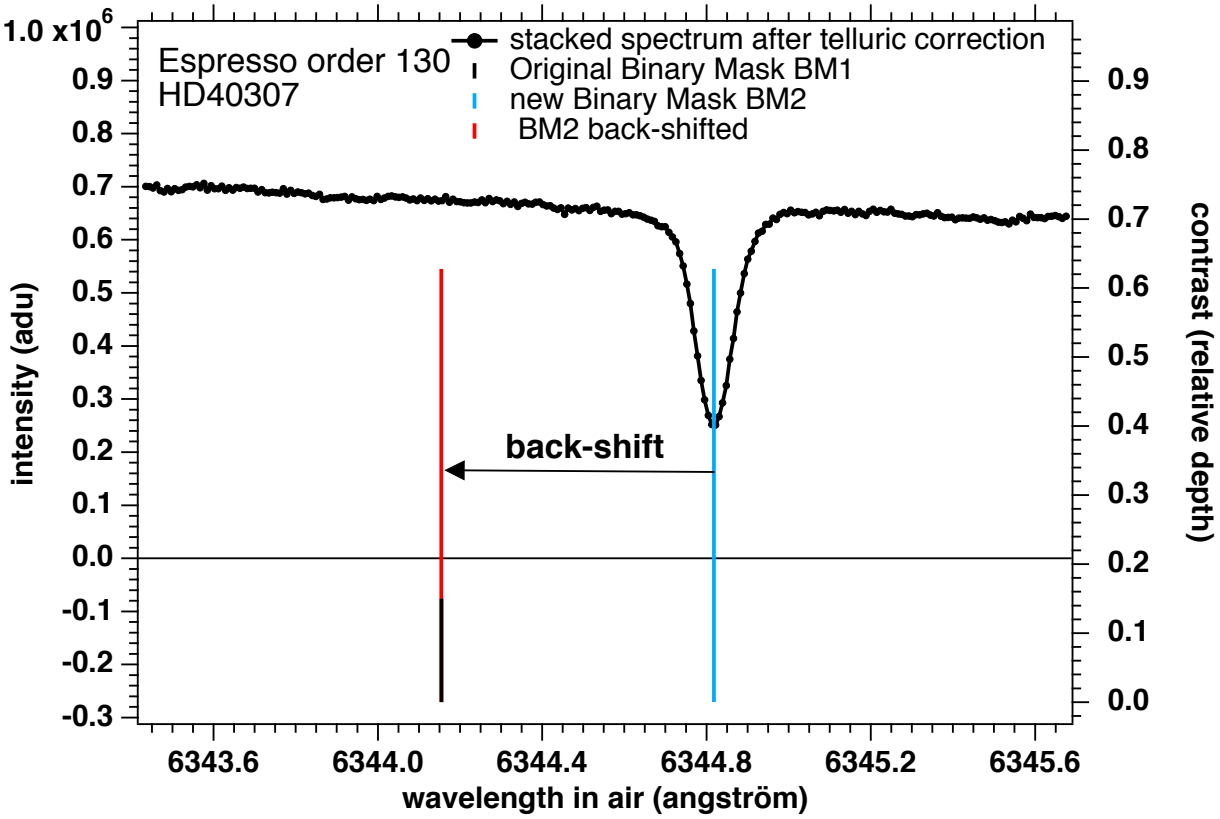


Figure 1.25: Adjustment of the two masks. The vertical blue line gives the wavelength position of the center of one stellar line of the observed spectrum, after telluric correction and stacking of all exposures. Its height is the contrast of the line (relative depth, 0.64). The red line represents the same line, back-shifted to account for the radial velocity RV_0 of the star. The velocity RV_0 was determined with the original mask BM1 and data from orders 92 to 115. As a result, the position of the red line coincides well (but not perfectly) with the corresponding line of the original BM1 mask (vertical bar, bottom).

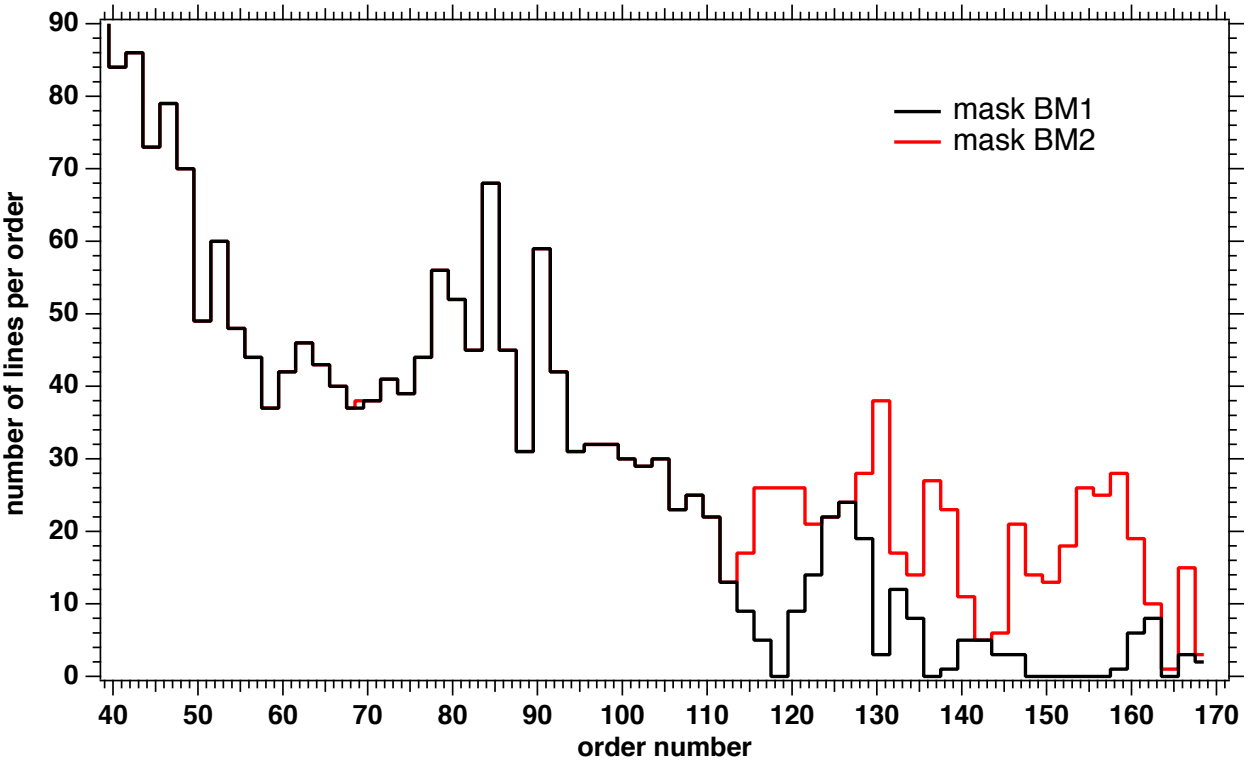


Figure 1.26: Number of stellar lines per order from the standard BM1 (black) and the augmented binary mask BM2=BM1+BMc (red) , restricted to lines with a contrast >0.2 and used in our study. For the red arm, the new mask contains 724 more lines than the 1060 lines of the original mask.

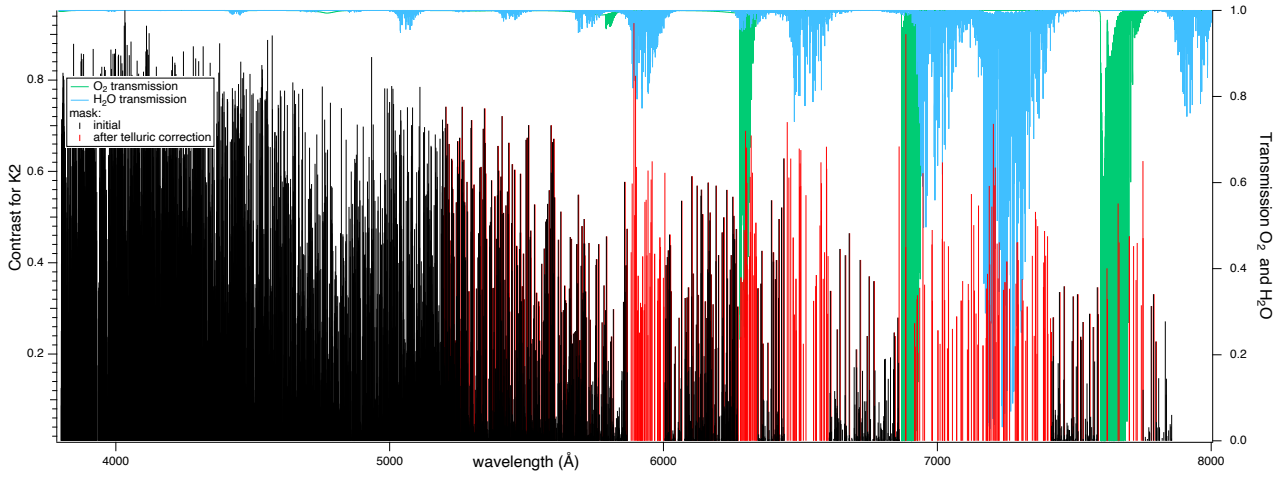


Figure 1.27: Old and new mask for K2 star. Black – original mask used by ESPRESSO pipeline; red – new mask with addition of lines (overlap with black in no telluric case); green – O_2 transmission (TAPAS model); blue – H_2O transmission (TAPAS model)

Order 130 is a good example to illustrate the effect of telluric correction, this order is partially contaminated by O_2 (figure 1.28) and in original mask only 3 lines for a whole wavelength range (6264-6358 Å) is used, while after correction it is possible to add 23 lines to existing mask (figure 1.29).

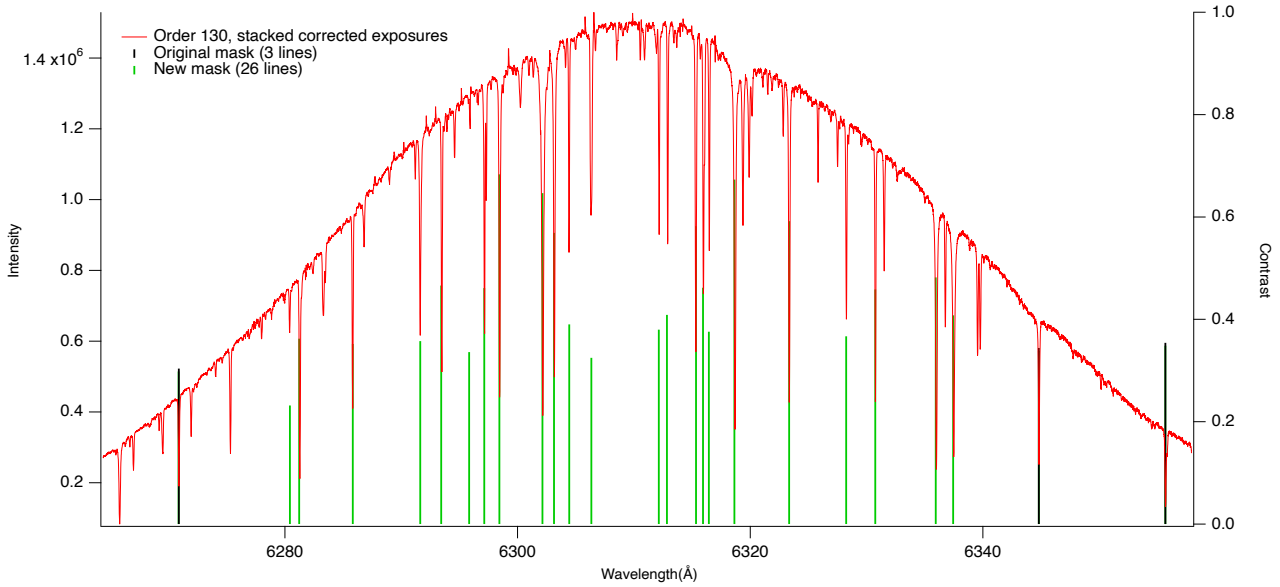


Figure 1.28: Example of order 130 partially contaminated by O_2. Order is shown after telluric correction (right scale). Black - original mask used by ESPRESSO pipeline; green – new mask with addition of lines (overlap with black in no telluric case).

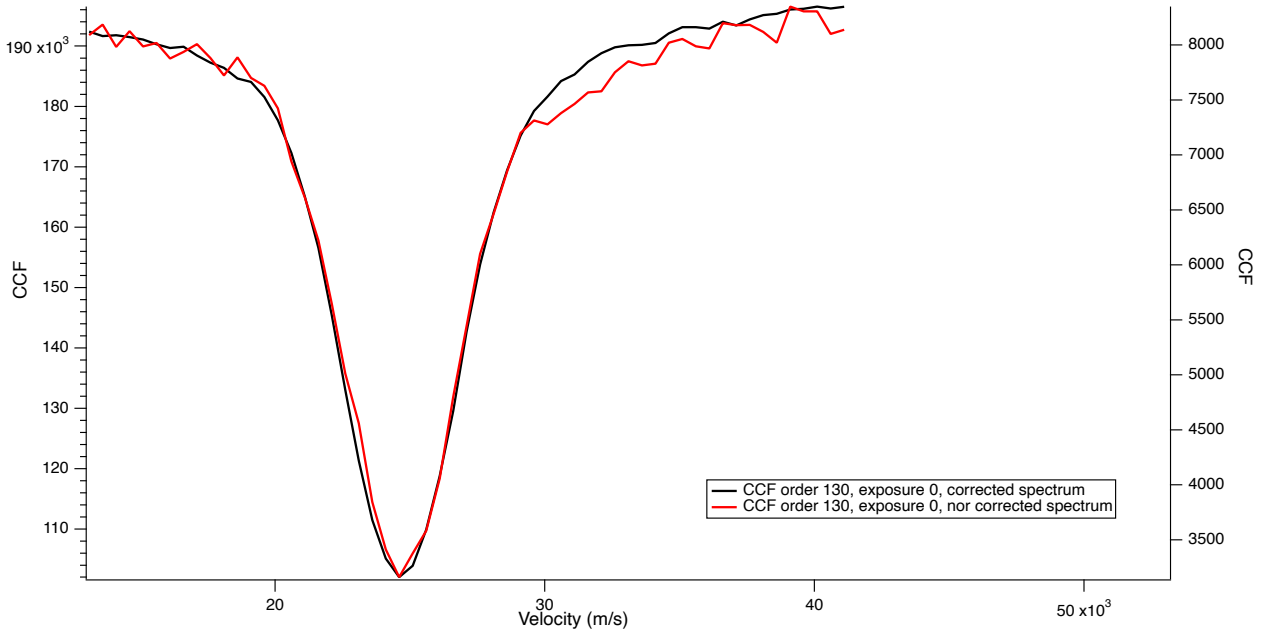


Figure 1.29: Example of CCF for order 130 (6264-6358 Å) which is partially contaminated by O_2. The original mask contains 3 lines; mask after correction contains 26 lines. Red - result of CCF computation with original mask (right scale); Black - result of CCF computation with new mask (left scale).

1.5.2 The result of telluric correction in terms of RV precision.

The results of computations are the following:

Adding spectral regions corrected from tellurics improves the precision of radial velocity retrieval

(Ivanova et al., 2023). To evaluate the improvement in precision after telluric correction we decided (as it was already mentioned) to use CCF method without weighting. In the case of uncorrected data the CCF is computed with BM1 (original mask), in the case of data after correction the CCF is computed with BM2 new mask = BM1+BM). Comparison of RV obtained for 3 orders are shown on Figures 1.31, 1.30 and 1.32. On Figure 1.30 order 116 is shown, this order is contaminated by H_2O and is easy to correct, in red the result of RV computation for not corrected data and BM1 is shown, in black - result of RV computation for corrected data and BM2. On Figure 1.31 order 130 is shown, this order is contaminated by γ band of O_2 and is moderately easy to correct, in red the result of RV computation for not corrected data and BM1 is shown, in black - result of RV computation for corrected data and BM2. On Figure 1.32 order 145 is shown, this order is contaminated by B band of O_2 and is more difficult to correct, in red the result of RV computation for not corrected data and BM1 is shown, in black - result of RV computation for corrected data and BM2.

For all 3 cases the average (over time) of RV has been subtracted from the whole series. The main criterion of precision improvement that we consider is the dispersion. Since all 200 exposures are obtained during $\sim$ 4 hours we should find a constant value. Both the error bars (error returned by Gaussian fit of CCF) and the spread of measurements (dispersion, S_{dev}) are decreased with the telluric corrected data. For order 116 (H_2O) the spread of measurements is reduced from 20.5 to 9.6 ms^{-1}, the average formal error is reduced from 12.8 to 7.1 ms^{-1}; for order 130 (γ band of O_2) the spread of measurement is reduced from 35.4 to 7.0 ms^{-1}, the average formal error reduced from 22.5 to 5.3 ms^{-1} and for order 146 (B band of O_2) the spread decreased from 47 to 32 ms^{-1}, the formal error increased from 21 to 25 ms^{-1}. The average error bar of the ensemble of orders in the red arm is reduced from 1.04 to 0.78 ms^{-1} with the new mask (error returned by Gaussian fit of CCF for all orders and their combination). If we use the photon-noise estimation on the CCF, the error bar decreases from 0.89 to 0.72 ms^{-1}, a reduction factor of 1.24. The spread of RV measurements S_{dev} is reduced from 2.83 to 2.37 ms^{-1}. The average error bar of combination of blue and red arms is reduced from 0.77 to 0.64 ms^{-1} while the spread of RV measurements S_{dev} is reduced from 3.53 to 2.59 ms^{-1}. These results are listed in Table 1.1.

During our study we found out that RV fluctuations, which were mentioned in Pepe et al. (2021) as the p-modes of star appear only in analysis of the blue arm, while the red arm did not show such fluctuations. We found sudden fluctuations to appear after exposure number 100, so we divided

exposures into 2 groups: before and after exposure number 100 (to get both samples equal in size). For the blue arm S_{dev} increased from 4.13 to 6.71 ms^{-1}, for the red arm S_{dev} decreased from 2.50 to 2.33 ms^{-1}. In the case of a stellar-nature cause of fluctuations, they should be detectable in both arms, since p-modes implied Doppler shifts of the stellar atmosphere would affect the blue and red arm equally. The referee of the paper Ivanova et al. (2023) mentioned that ESPRESSO had a blue cryostat thermal stability problem (Figueira et al., 2021). Therefore this type of fluctuations is certainly caused mainly by the thermal instability of the blue detector.

Table 1.1: CCF statistical results. Units are ms^{-1}.

Used exposures	0-199 no correction	0-199 with correction	0-99 with correction	100-199 with correction
Formal error err Red Arm	1.04	0.78	0.79	0.78
Formal error err Blue Arm	1.15	1.15	1.21	1.10
Formal error err Red + Blue Arms	0.77	0.64	0.65	0.63
Standard deviation S_{dev} Red Arm	2.83	2.37	2.50	2.23
Standard deviation S_{dev} Blue Arm	5.78	5.56	4.13	6.71
Standard deviation S_{dev} Red + Blue Arms	3.53	2.59	3.03	3.97

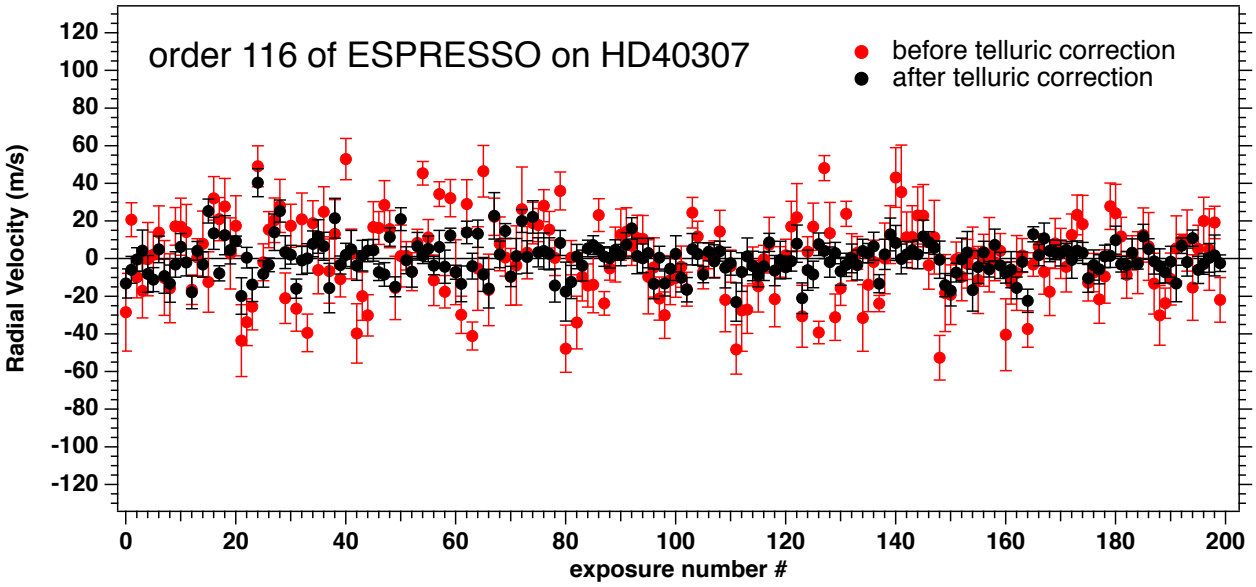


Figure 1.30: Radial velocity of star HD 40307 for order 116, which is contaminated by H_2O. In red - result with the standard binary mask BM1 before telluric correction; in black - result with our new binary mask BM2 after telluric correction. The average radial velocities of 31386.4 and 31388 ms^{-1} were subtracted respectively from the series before and after correction. The dispersion of black points is smaller than those of uncorrected RV measurements.

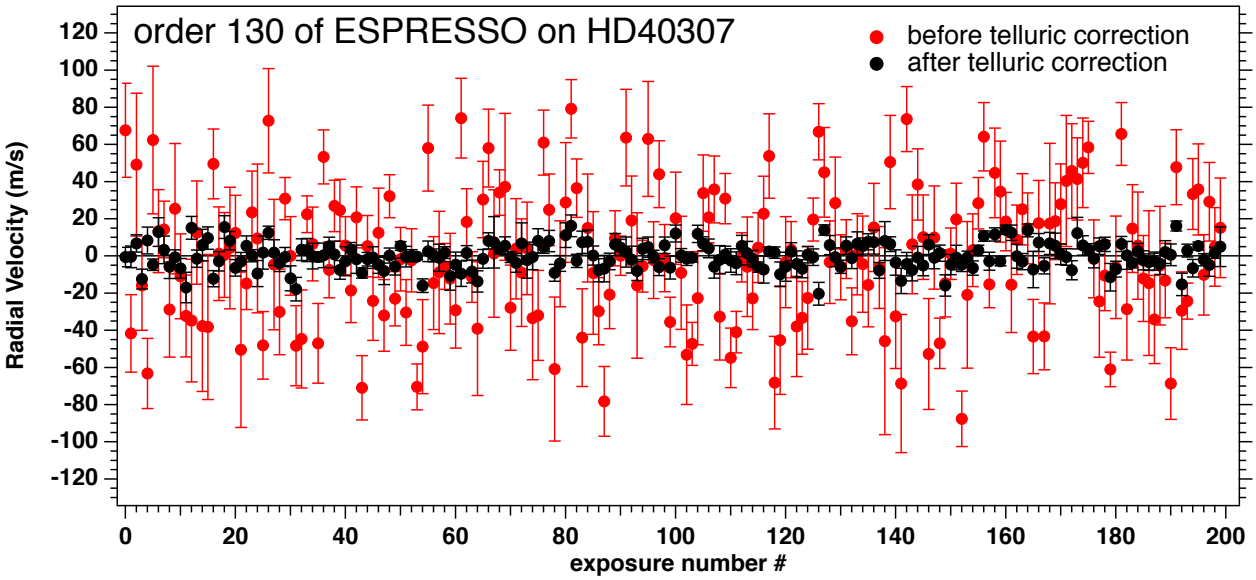


Figure 1.31: Radial velocity of star HD 40307 for order 130, which is contaminated by O_2. In red - result with the standard binary mask BM1 before telluric correction; in black - result with our new binary mask BM2 after telluric correction. The average radial velocities of 31372.4 and 31384.2 ms^{-1} were subtracted respectively from the series before and after correction. The dispersion of RV points is drastically reduced after correction.

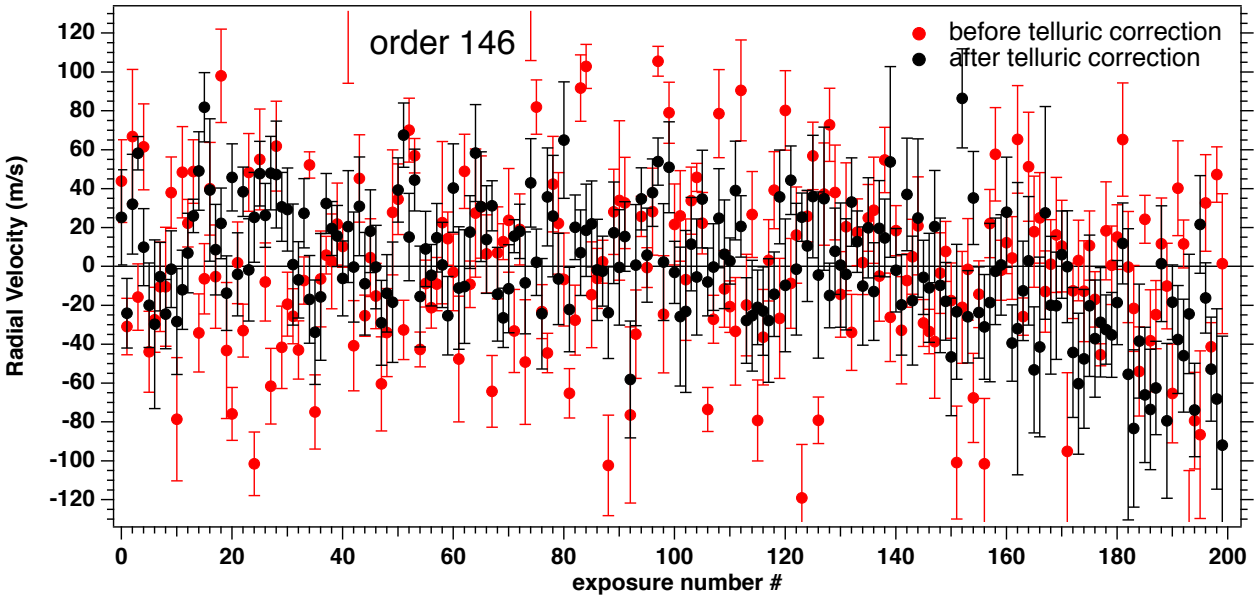


Figure 1.32: Radial velocity of star HD 40307 for the extreme case of order 146, which is heavily contaminated by O_2. In red - result with the standard binary mask BM1 before telluric correction; in black - result with our new binary mask BM2 after telluric correction. The average radial velocities of 31386.4 and 31388 ms^{-1} were subtracted respectively from the series before and after correction.

To evaluate the uncertainty of retrieved radial velocities it is possible to use the evaluation returned by Gaussian fit or, as it is described in Boisse et al. (2010) and ESPRESSO Pipeline User Manual (ESO-331895 Issue 2.3.3 p.67), the uncertainty computed with 2^{nd} Formula of Connes (more about it can be found at Section 2.3.3). During our work on the paper (Ivanova et al., 2023) we used both

the evaluation returned by a Gaussian fit of CCF and the evaluation by the 2^{nd} formula of Connes, after a remark of the referee. We followed the approach described in Boisse et al. (2010). We applied this computation of the uncertainty in two cases: to the CCF computed on not-corrected spectra with the official ESPRESSO mask on the one hand, and to the CCF computed on corrected spectra with custom made mask on the other hand.

For each exposure we made the CCF and computed δV_Q(k) for each order k, according to the second formula of Connes where N_{e^-} is the total number of electrons in the CCF(k) and Q(k) the quality factor of the CCF(k) (over a defined number of points):

$$\delta V_Q(k) = \frac{c}{Q(k)\sqrt{N_{e^-}}} \tag{1.2}$$

then $\overline{\delta V_Q}$ for one exposure is made by the combination of δV_Q(k) calculated for each order according to the following equation:

$$\overline{\delta V_Q} = \frac{1}{\sqrt{\sum(\frac{1}{\delta V_Q(k)})^2}} \tag{1.3}$$

Uncertainties computed both from our Gaussian fit of the CCF δV_G and from the application of the above quality factor to the CCF δV_{QCCF} are shown on figures 1.33 and 1.34. In both cases estimated uncertainties decrease after telluric correction. We have verified that the fluctuations of δV_{QCCF} (Figure 1.34) are almost entirely due to intensity fluctuations reflected on the number of electrons in the CCF. This explains the parallelism of both curves, before and after telluric correction. On the last figure 1.35 we show a comparison between our estimates based on a Gaussian fit δV_G and those based on the quality factor δV_{QCCF} for all orders (blue and red) simultaneously. We can see that generally the uncertainty estimation obtained from the quality factor δV_{QCCF} is smaller than the one obtained by Gaussian fit and also displays smaller fluctuations. This is because Boisse et al. (2010) used the full CCF to perform their estimate while we used only 10 points of the CCF for our Gaussian fit (we recall that in this way we attempt to avoid RV biases appearing when fitting a whole spectral line as described by González Hernández et al. (2020).

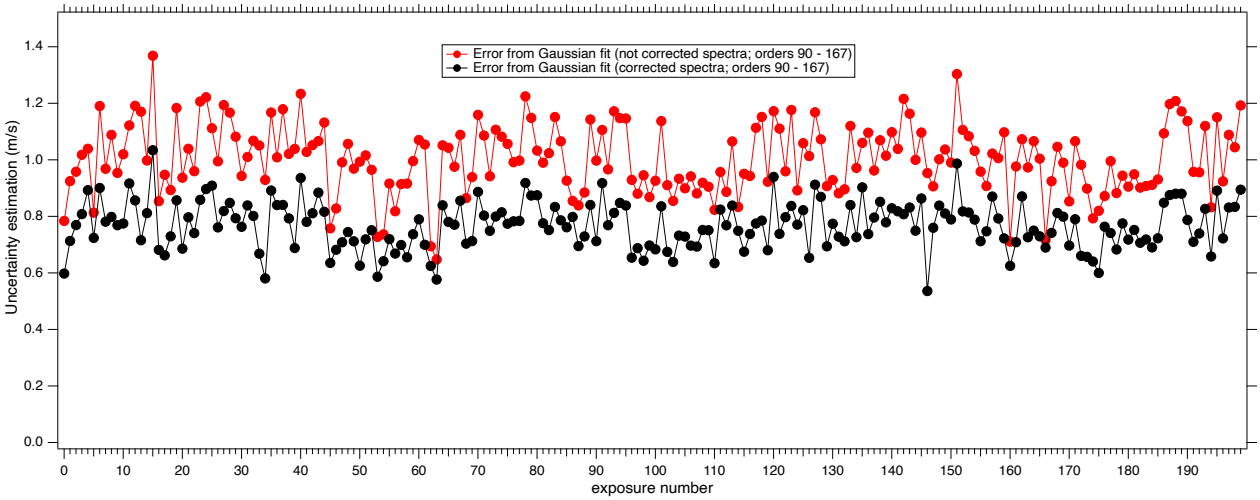


Figure 1.33: Estimated uncertainty δV_G based on the Gaussian fit for the red arm. Red curve - before telluric correction; black curve - after telluric correction. The estimated uncertainty is substantially decreased after telluric correction.

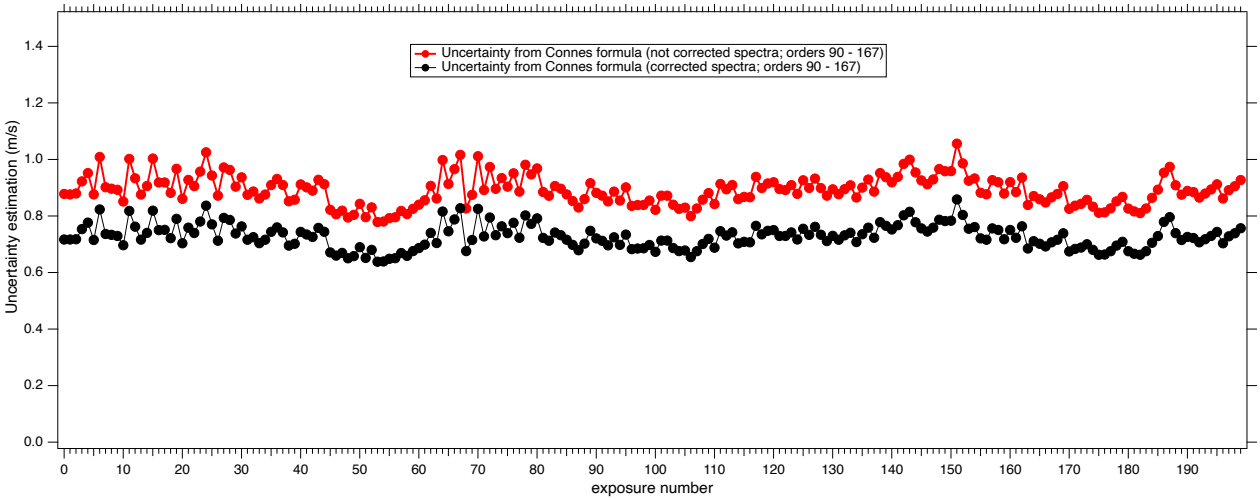


Figure 1.34: Estimated uncertainty δV_{QCCF} based on the quality factor for CCF from Boisse et al. (2010) for the red arm. Red curve - before telluric correction; black curve - after telluric correction. Fluctuations are connected to the fluctuations of spectral intensity among various exposures. Again, the estimated uncertainty is decreased after telluric correction.

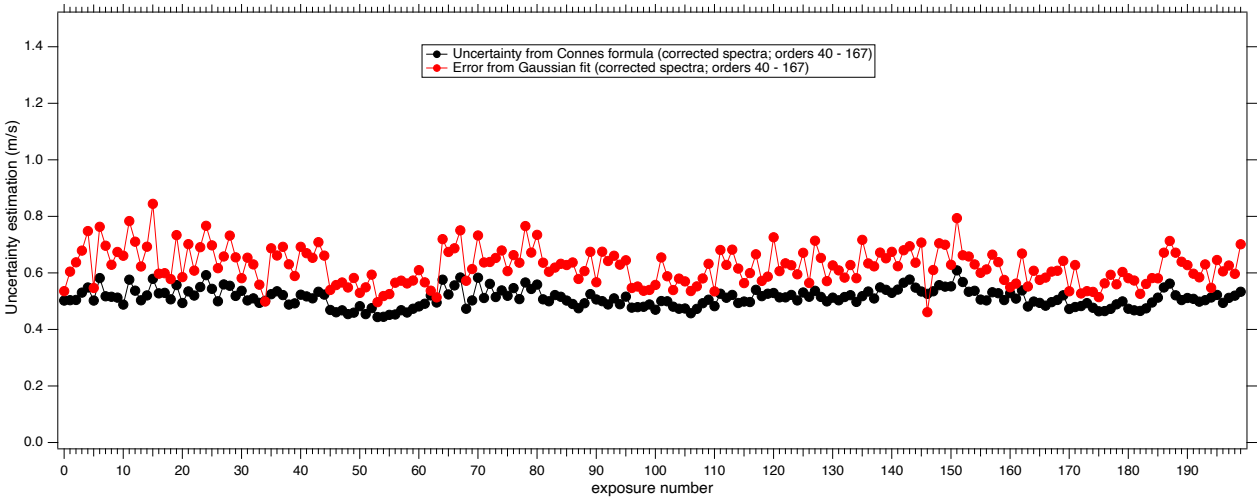


Figure 1.35: Comparison of estimated uncertainties: based on Gaussian fit of the CCF (δV_G in red) and on the quality factor for the CCF from Boisse et al. (2010) (δV_{QCCF} in black), for red and blue arms combined. For the blue arm, orders 0 to 39 are excluded because of too low signal.

1.5.3 Results of telluric correction in terms of RV precision for the 24 December night.

As it was already mentioned in Section 1.4.2 the star HD 40307 has been observed for several consecutive nights. In particular, during the night 24 December 2018 it has been observed 406 times and we extracted 401 exposures. It was decided to correct the full series of measurements obtained on the night of December 24th to confirm the previously obtained improvements in the precision of the radial velocity measurements. The result of the correction in terms of radial velocities is shown on figure 1.36. The S_{dev} after correction is decreased from 3.41 ms^{-1} to 2.74 ms^{-1}.

In order to take into account the small evolution of RV due to the planets we fitted the two series of RVs to a second order polynomial and subtracted each series to get the corresponding series of RV residuals. The histograms of the two RV residuals are displayed in Figure 1.37. Gaussian functions fitted to the histograms are also displayed. The width of the Gaussian fit to the histogram is decreased from 3.39 ms^{-1} to 2.73 ms^{-1} after telluric correction, consistent with the decrease of S_{dev} noted above. Since the S_{dev} is a measure of the mean error, it means that the precision has increased by 20% when adding telluric corrected portions of the spectrum in the RV retrieval algorithm. At the same time, there is a small variation of $\simeq$ -0.4 ms^{-1} of the average RV which is due to the fact that we have imposed a RV for the computation of the new lines of the BM. t has no incidence on exoplanet searches.

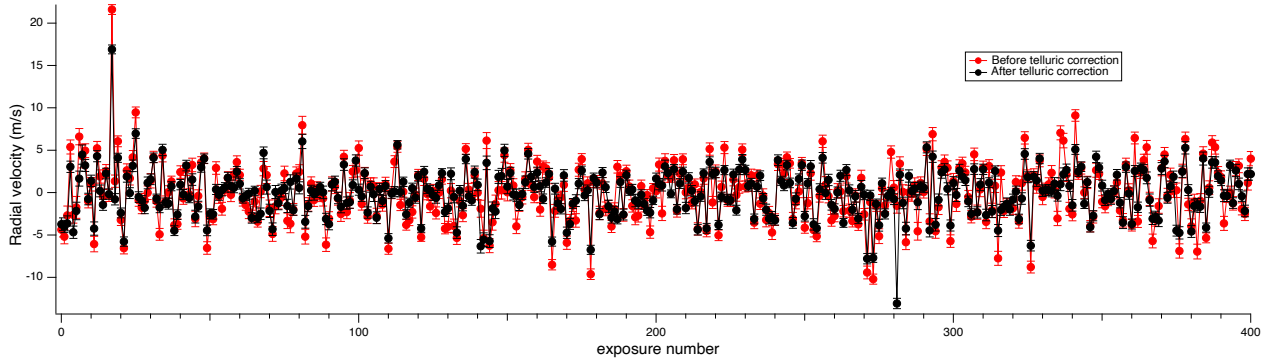


Figure 1.36: Radial velocity of star HD 40307 for nigth December 24th. In red - result with the standard binary mask BM1 before telluric correction; in black - result with our new binary mask BM2 after telluric correction. The average radial velocities of 31383.5 and 31383.4 ms^{-1} were subtracted respectively from the series before and after correction. The S_{dev} after correction is decreased from 3.41 ms^{-1} to 2.74 ms^{-1}.

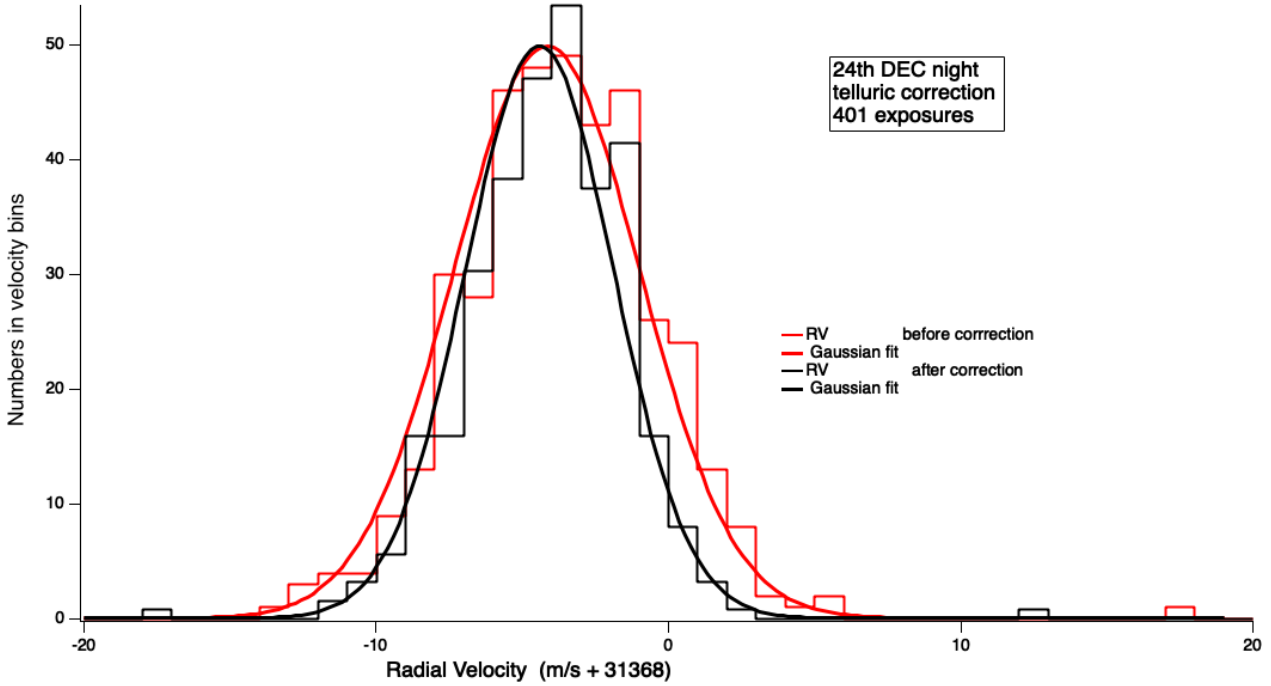


Figure 1.37: Histograms of the two series of radial velocities of HD 40307 for the December 24th night. In red - result with the standard binary mask BM1 before telluric correction; in black - result with our new binary mask BM2 after telluric correction. The width of the Gaussian is significatly reduced after telluric corrections, from 3.39 ms^{-1} to 2.73 ms^{-1}.

1.6 Testing the telluric correction for other spectral types and in the near-infrared

We tested the ability of the code to correct different types of data. Firstly, we tested an extension to stars of earlier spectral types, namely F and G types. Secondly, we also tested a correction of CH_4

and CO_2 lines in the near-infrared.

1.6.1 Data

WASP-76

WASP-76 is a F7 star with visual magnitude of 9.52. WASP-76 has one transiting hot Jupiter (West et al., 2016). It was observed by ESPRESSO in 2018 and 2019 with 3 observational programs:

- Prog.ID:1102.C-0744(C); PI/Col: PEPE/ ADIBEKYAN/ ALIBERT/ ALLENDE PRIETO/ BOUCHY/ CRISTIANI/ D'ODORICO/ EHRENREICH/ FIGUEIRA/ GÉNOVA SANTOS/ GONZÁLEZ HERNÁNDEZ/ LEITE/ LOVIS/ MARTINS/ MICELA/ MOLARO/ NUNES/ SOZZETTI/ PORETTI/ REBOLO/ SANTOS/ SOUSA/ UDRY/ ZAPATERO OSORIO

- Prog.ID:1102.C-0744(D); PI/Col: PEPE/ ADIBEKYAN/ ALIBERT/ ALLENDE PRIETO/ BOUCHY/ CRISTIANI/ D'ODORICO/ EHRENREICH/ FIGUEIRA/ GÉNOVA SANTOS/ GONZÁLEZ HERNÁNDEZ/ LEITE/ LOVIS/ MARTINS/ MICELA/ MOLARO/ NUNES/ SOZZETTI/ PORETTI/ REBOLO/ SANTOS/ SOUSA/ UDRY/ ZAPATERO OSORIO

- Prog.ID:0104.C-0642(A); PI/Col: GIBSON/ DE MOOIJ/ MERRITT/ NUGROHO/ WATSON

The main purpose of programs 1102.C-0744(C) and 1102.C-0744(D) was to detect and characterise Earth-like planets with ESPRESSO. The last program 0104.C-0642(A) was focused on study of exoplanets atmospheres with ESPRESSO.

To test our correction algorithm we used the dataset obtained during Prog.ID:1102.C-0744(C). All data was retrieved from the ESO archive with open access.

WASP-127

WASP-127 is a G5 star with visual magnitude of 10.172. WASP-127 has one transiting hot Jupiter (Lam et al., 2017). It was observed by ESPRESSO in 2021 during 1 observational program Prog.ID:106.21R1.001 PI/Col:GIBSON/ NUGROHO/ MERRITT/ DE MOOIJ/ WATSON. The program is dedicated to the study of exoplanets atmospheres with ESPRESSO. All data was retrieved from the ESO archive with open access.

HD 189733

HD 189733 is a K2V star with visible magnitude of 7.648. HD 199733 has one transiting hot Jupiter (Bouchy et al., 2005). It was observed by SPIROU during several observational programs since 2018. We used the dataset obtained in July 2023. All data was retrieved from the "Centre Canadien de Données Astronomiques" archive with open access.

1.6.2 Results

WASP-76 and WASP-127 were chosen to demonstrate that the algorithm can work for other spectral classes than one of HD 40307. The correction procedure was the same as described above. Different stellar masks were done using observations and synthetic spectra of star taken from THE SYNTHETIC STELLAR SPECTRA DATABASE POLLUX

(https://pollux.oreme.org/collections/)

Examples of WASP-76 corrections Examples of correction for WASP-76 are shown below on figures 1.38, 1.39, 1.40:

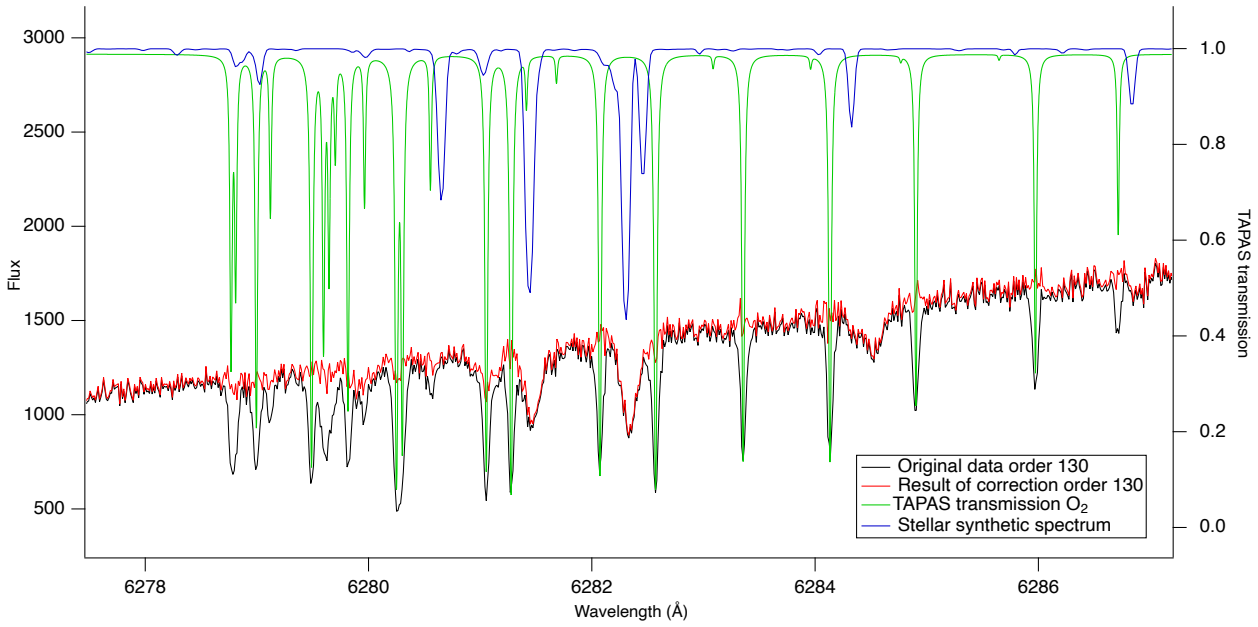


Figure 1.38: Illustration of the correction for WASP-76 order 130 contaminated by O_2

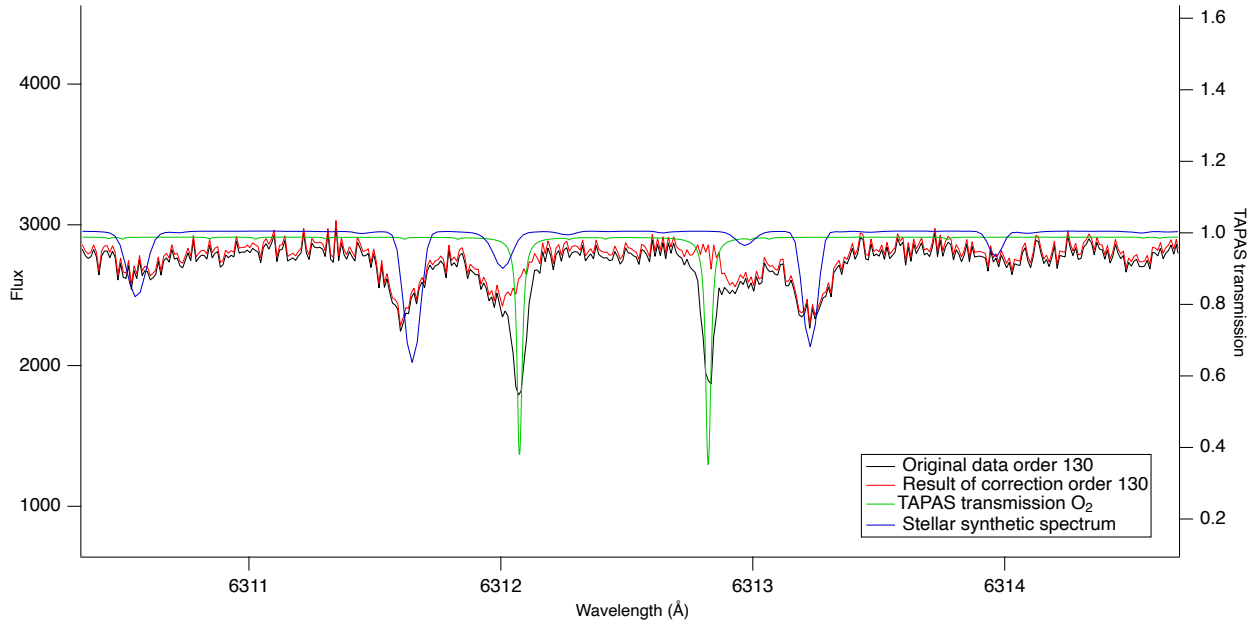


Figure 1.39: Illustration of the correction for WASP-76 order 130 contaminated by O_2. Zoom into region around 6312 Å

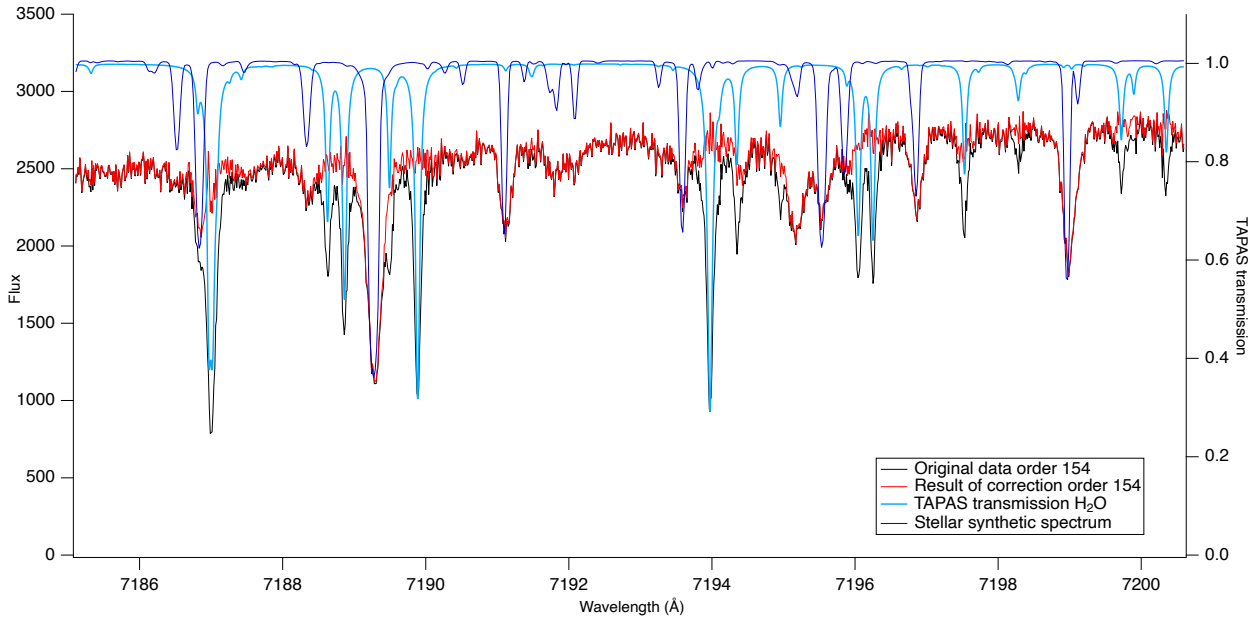


Figure 1.40: Illustration of the correction for WASP-76 order 154 contaminated by H_2O.

Examples of WASP-127 corrections Examples of correction for WASP-127 are shown below on figures 1.41, 1.42, 1.43

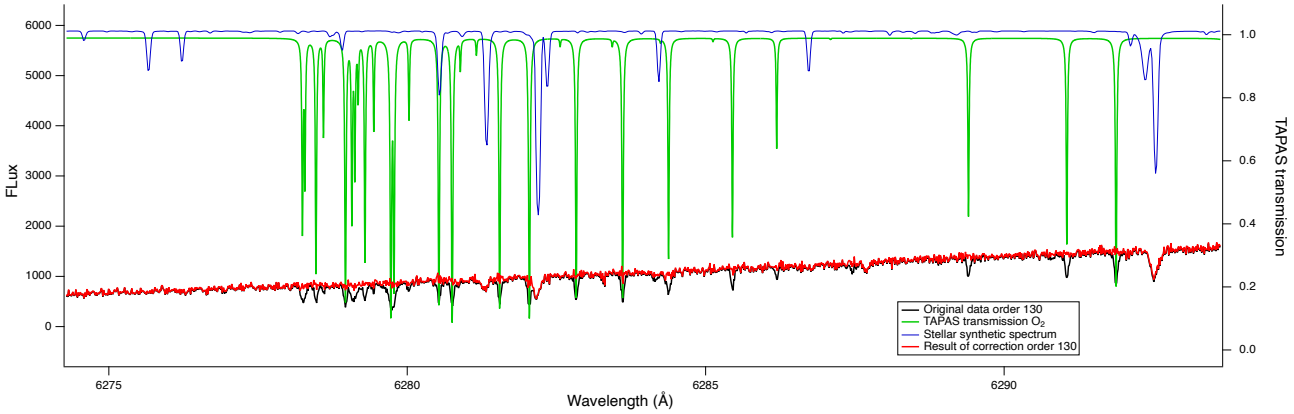


Figure 1.41: Illustration of the correction for WASP-127 order 130 contaminated by O_2

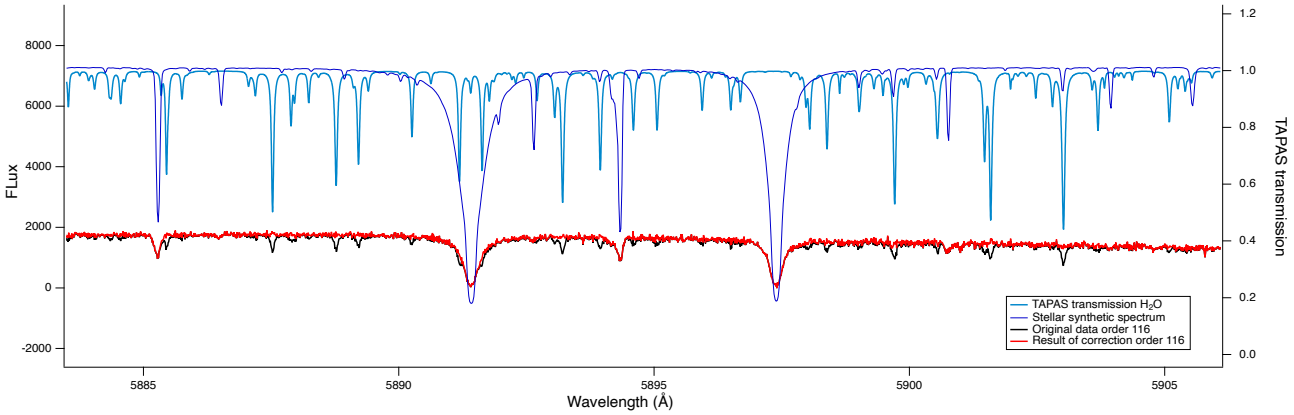


Figure 1.42: Illustration of the correction for WASP-127 order 116 contaminated by H_2O.

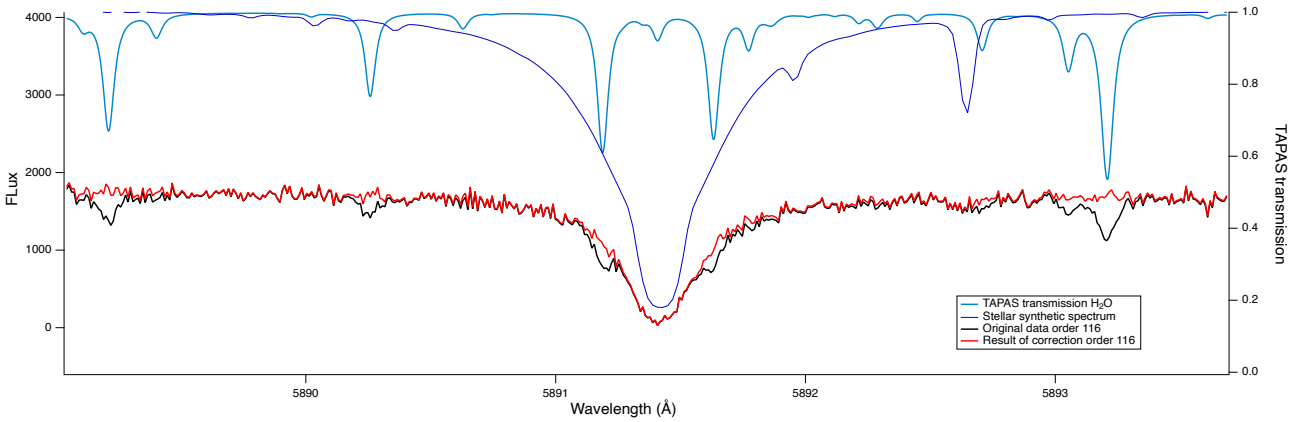


Figure 1.43: Illustration of the correction for WASP-127 order 116 contaminated by H_2O. Zoom into region around NaI line.

Examples of HD 189733 corrections In the case of SPIROU data, the correction procedure was slightly different; instead of using a resolution map (as it was done for ESPRESSO data) one single resolution value was imposed, 70000 as mentioned in Moutou et al. (2020). Another difference is that in addition to the already discussed H_2O and O_2, the operating range of SPIROU also includes

CH_4 and CO_2. We will focus on correcting CH_4 and CO_2 with our algorithm. We started with TAPAS downloaded CH4 and CO2 transmittance spectra for the CFHT location and altitude and the date of observations. But the abundances of the two species were free parameters during the adjustment.

As it was already mentioned in Section 1.1, it is the APERO - pipeline which is used for SPIROU, and as stated in Cook et al. (2022) the results of telluric corrections are archived, so it is possible to see how the results of correction by our new algorithm differ from the results of correction produced by APERO.

Examples of corrections are shown on figures 1.44 (for CO_2), 1.45 (for CH_4). It can be seen that in some places our result coincides with the APERO result, but in some places and especially at the location of strong lines they differ and the APERO result is closer to the continuum level than the result of our correction. This is not surprising for two reasons: first, we used a unique resolution, and the correction of strong lines is very sensitive to its exact value. More work is needed to perform a correction with adjusted resolving power. Second, as described earlier in section 1.1, the APERO pipeline starts the correction with TAPAS and improves the accuracy of the correction with accumulated data, correcting for the slight discrepancies between the modeled transmittance and data.

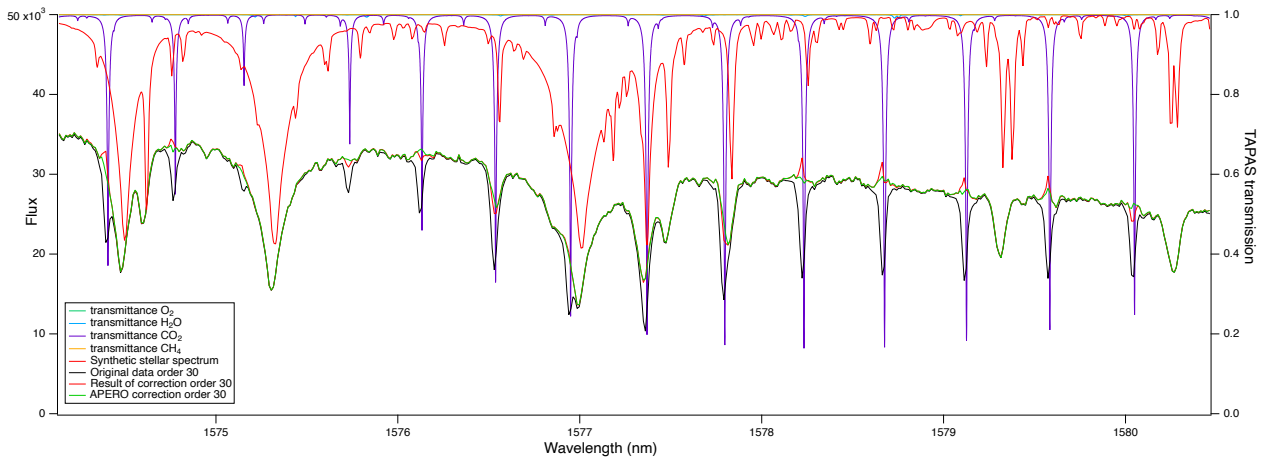


Figure 1.44: Illustration of the correction for HD189733 order 30 contaminated by CO_2.

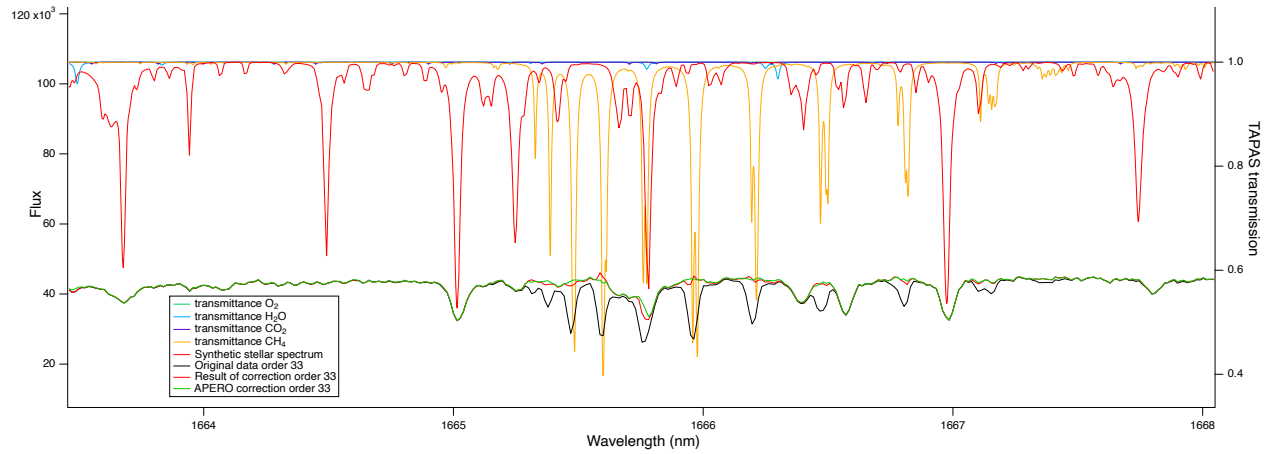


Figure 1.45: Illustration of the correction for HD189733 order 33 contaminated by CH_4.

1.7 Future work: treatment of residuals and detection of potential annual signature of telluric absorption

As shown by Cook et al. (2022), a perfect telluric line removal is not achievable if one uses the molecular data and LBLRTM results as they are currently produced, despite their already extraordinary quality, since there are small wavelength shifts or oscillator strengths inaccuracies at the origin of post-correction remaining residuals, i.e., after a telluric correction, the observed corrected spectrum may display sometimes a small telluric residual at the place of the original telluric line. Sophisticated methods using corrections of hot star spectra as done by Cook et al. (2022) in the near-infrared are necessary. Another technique could be an additional spectrum of a hot star recorded before or after the target star, its correction and the use of the residuals measured in the hot star corrected spectrum to apply to the target spectrum.

The existence of residuals is potentially affecting the RV measurements. When observing a star all along the year, the system of spectral telluric absorptions is moving w.r.t. the system of stellar lines, because of the variable Barycentric Earth Radial Velocity (BERV) [to the star] along its orbit, which may reach $\pm$ 30 kms^{-1} for a star in the ecliptic plane. If not corrected, or improperly corrected, the retrieved RV of a star may display a yearly signature, detrimental to the monitoring of a planetary system. A good example of such an effect is shown in figure 3 of Cretignier et al. (2020) reproduced below (figure 1.46), which examines the yearly signature of one single star line (Cr I line at 5038.90 Å) which is swept across by a (not corrected) telluric line of H_2O at 5038.85 Å inducing a spurious RV variation of $\pm$ 300 ms^{-1}. Such an up-and-down signature is due to one single sweep of the telluric

line across the star line around a particular value of the BERV, affecting the blue side and the red side of the star line in turn.

We note that, since the BERV is changing as a sine function of Earth's longitude or day-of-the year, in principle there should be two such crossings per year. Here we see only one, because all dates of the year cannot be covered by observations, when the star is angularly too near by the Sun. There could also exist only one signature, when the sweep across is obtained at one maximum or minimum of the BERV, and is incomplete. In practice, there are many telluric lines which can perturb the RV retrieved for many different stellar lines, and with phases of the year which are different for each couple of star line/telluric line. Therefore, the retrieved RV, an average of all RV for each line, will display a complicated pattern, in addition blurred by the fact that one single signature extends over $\sim$ 0.15 phase, or about 2 months.

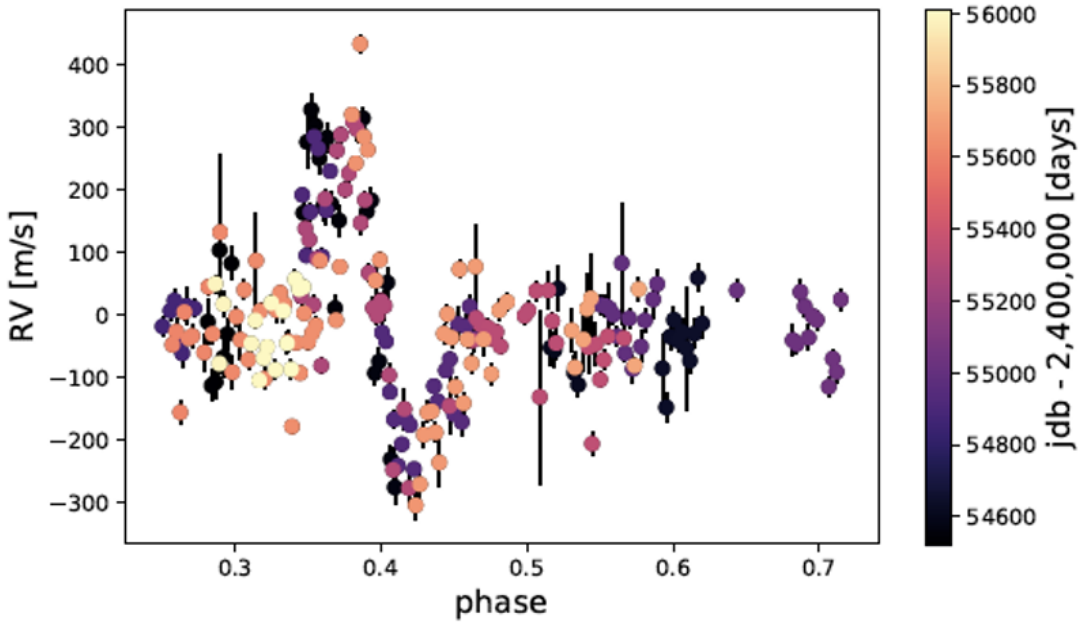


Fig. 3. RV_i time series of the shallow Cr I line at 5038.90 Å, from 2008 to 2012, phase-folded with the period of a year. This stellar line is contaminated by the 5038.85 Å telluric line, which has a depth ratio of 5%. A peak-to-peak RV variation of 600 m s^{-1} is observed when the spectral line moves through the telluric line over the year.

Figure 1.46: Figure 3 from Cretignier et al. (2020)

In the above example, the telluric line has a depth of about half the depth of the stellar line, i.e. ,

it corresponds to a rather strong case. In our correction, we selected, as additional lines of the binary mask, only those lines whose depth is at least 20%. This implies that, to get an effect as strong as in the quoted example, the depth (or relative excess) of the post-correction residual of the telluric line should be at least 10%. Such residuals would be clearly seen, and we can safely exclude them. However, we can not exclude smaller residuals not visible by eye and producing a signal perturbing the RV measurement.

A strong difficulty is that, in order to detect a repetitive yearly pattern, a lot of observations are needed over several years, and for a star without planet preferentially. We did not find ESPRESSO data with enough adapted observations to perform tests of our correction, for a star without planet. One way to evaluate the improvement of one particular method of telluric correction would be to simulate the behaviour of one single stellar line perturbed by one single telluric line, like the example of Figure 1.46, and compare the RV retrieved pattern, with and without telluric correction. As a rule-of-thumb, we may estimate that the up-and-down yearly signature in RV of one single line will decrease by the same factor by which the original telluric has been reduced to its remnant, after telluric correction. The up-and-down signature should be very significantly reduced for the whole set of telluric corrections to be valid. Several thresholds could be added and the lines of the BM selected accordingly.

Chapter 2

Description of some RV retrieval algorithms and RV changes retrieval algorithms

In this section dedicated to radial velocity retrieval methods we first describe the absolute radial velocity of star in Section 2.1. Then we discuss the difference between absolute RV and variation of RV in Section 2.2. In Section 2.3 we describe different approaches for RV and RV changes determination (Cross-correlation function, Pierre Connes algorithms, template matching) and show how they have been used in our study.

2.1 The absolute radial velocity of a star and the influence of stellar activity

Formally speaking, the radial velocity of a star in the Solar barycentric system is dR/dt, the time derivative of R, distance of the center of the star to the barycenter of the Solar System. We have no direct access to this quantity. What is measured is the spectrum of the star integrated over the whole disc, from which may be determined the Doppler effect Vd of the light emitting surface of the star. This Doppler effect is therefore a proxy of the dynamical velocity. To first order, the Doppler effect Vd is measured by the difference between the observed wavelength λ_{obs} of a certain spectral line from a given element, (say, Fe I), with the laboratory measurement λ_{lab} of the transition of this element. However, a number of second order effects must be considered when accuracy/precision of 1 to 10 ms^{-1} or better are targeted for. We distinguish two groups of effects: the effects that affect the

absolute radial velocity and those that affect the variation of radial velocity.

The effects affecting the absolute radial velocity are:

- No values of λ_{lab} are given in the NIST data base with a relative accuracy of 3e-9 necessary to reach an absolute accuracy of 1 ms^{-1} for dR/dt. According to González Hernández et al. (2020) (GH 2020 in the following), the best quality laboratory measurements (Nave et al., 1994) have an accuracy in the range 0.28 to 1.25 mÅ, corresponding to 16 to 75 ms^{-1}.

- The Gravitational Red Shift (GRS) : the photon lose energy to escape the gravity g_{surf} of the star and mimics a Doppler effect which depends of g_{surf}. The theoretical value is 636.31 ms^{-1} for the Sun, which has been actually measured several times (e.g., GH 2020).

- The convection in the photosphere of the star, manifested by granulation. We quote from GH 2020 :

  "In the Sun, the gravitational redshift is competing with the blue shift of the lines due to the convective motions. Convection produces granulation phenomena, which induce line shifts which are opposite to the gravitational redshift. The granulation pattern is formed by upward flows with a velocity of 1–2 kms^{-1} in the inner parts of the granules and downward flows in the intergranular lanes, which are faster by a factor of between two and three."...

  "The upward flows dominate, inducing a blue shift." "This complex granulation pattern produces asymmetric disk integrated photospheric lines with observed Doppler blue shifts down to $\sim$ – 500 ms^{-1} (Allende Prieto and Garcia Lopez 1998b; Reiners et al. 2016) after subtracting the theoretical value of the solar gravitational redshift." We may consider the convective blue shift as a constant effect in average, with some stochastic fluctuations (when it is not attenuated by stellar activity, see below).

The effects affecting radial velocity variability (following Meunier (2021)).

- Random changes in the small scale convection (granulation) at the surface of the star. When integrated over the whole stellar disc, there remain stochastic fluctuations around the average.

- Supergranulation

- Convective blueshift inhibition in spots and plages

- Active regions, which decrease the flux on one side of the star (receding or advancing half-disk), change the integrated flux, with a period equal to the rotation of the star (spots and plages).
- Solar oscillations with periods around 5 minutes (for the Sun) giving an amplitude of up to 4 ms^{-1}.

The granulation is the irregular cellular pattern at the surface of stars that arise because stars have a convective envelope in their photospheres where the hotter bubbles of gas rise (blueshift) and the cooler bubbles sink (redshift). Illustration of the granulation is shown on figure 2.1, which is taken from paper Dalal et al. (2023). According to Meunier (2021) a typical size of granule is of the order of 1 Mm and lifetime of 10 minutes. The imbalance between contribution of hot granules and cool intergranular lanes leads to the blueshift of RV or convective blueshift (Dravins et al., 1981).

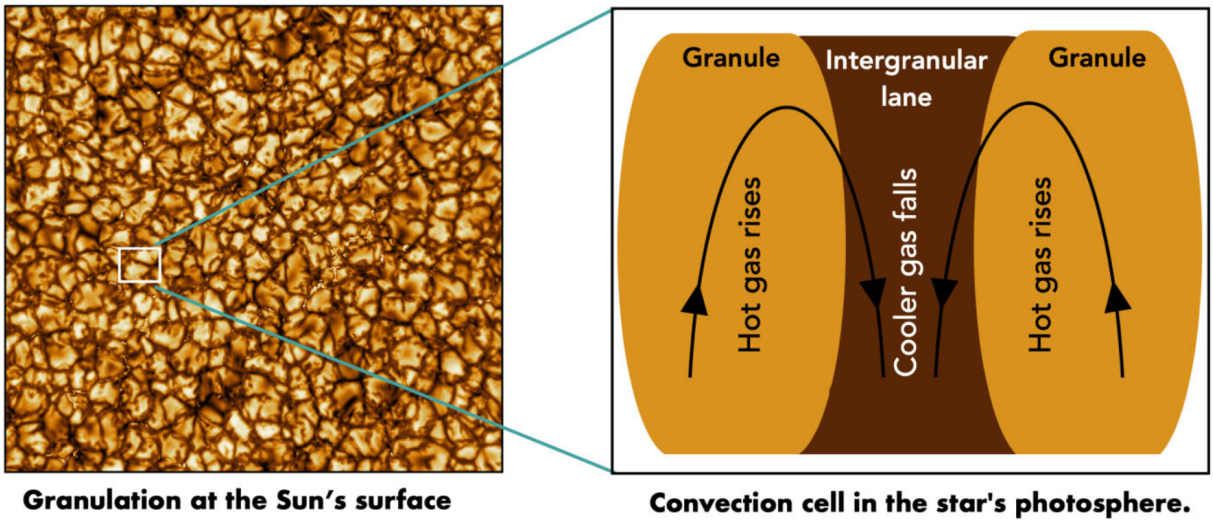


Figure 2.1: Figure 1 from Dalal et al. (2023). Solar granulation: Left-hand panel: Our Sun's surface taken at 789 nm using the Daniel K. Inouye Solar Telescope (by NSO/AURA/NSF, for illustrative purposes only), which covers an area of 36 500 km × 36 500 km (51 arcsec × 51 arcsec). The image shows the granulation pattern, where the cell-like structures or granular cells (with dimensions of 30–1500 km) are the convection cells covering the entire surface of the Sun. Hot plasma rises in the bright cells (granules), cools down, and then falls into the dark (intergranular) lanes. Right-hand panel: Illustration of the cross-section of a granular cell, spanning approximately 30–1500 km in length, where hot gas rises in the granules (orange region) and falls in the intergranular lane (dark brown region).

Concerning the pressure-mode stellar oscillations, they can be mitigated by adjusting the exposure time to an integer number of period.

There is a process of stochastic fluctuations of granulation and supergranulation convective blue shift. The life time of one granule is of the order of a few minutes, and there are about 10^4 on the visible disc of the star; the average RV will not be rigorously constant, but rather will display stochastic

fluctuations. The magnitude of these fluctuations has been estimated from model simulation by Meunier and Lagrange (2020) for stars from F6 to K4 spectral classes.

We show below their simulation case for a K4 star, nearby our star HD40307 (K2.5): there are peak-to-peak fluctuations of the order of ± 0.5 ms^{-1}, even down to time scales of hours.

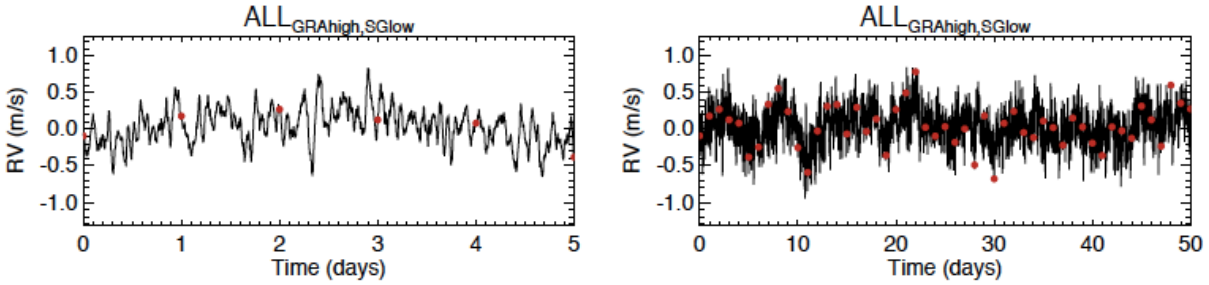


Figure 2.2: Simulations of the variation of RV due to granulation and supergranulation changes for a K4 star, in the case of a high effect of granulation and low effect of supergranulation. Extracted from figure A.4 from Meunier and Lagrange (2020). The span of times is extended on the right panel, and red points are sampled at one per day.

Also attenuation of granulation blue shift takes part in. It has been studied from solar observations that the convective granulation blueshift is inhibited or attenuated by magnetic activity, which manifests itself on the Sun by solar spots, faculae and plages (Meunier, 2021). From analysis of solar Dopplergrams and magnetograms obtained with the MDI instrument of board of SOHO Meunier et al. (2010) have measured along a whole solar cycle (11 years) that the average RV of the integrated solar disc is increasing with solar activity (due to the attenuation of the granulation blueshift) from 0 (quiet Sun as a reference) up to 8 ms^{-1}, almost 100 times larger than the reflex motion induced by Earth around a G star.

Therefore, it appears that the granulation and supergranulation and their time variations seem to be the limiting factors for the detection of Earth size planets via radial velocity methods (Meunier et al., 2015).

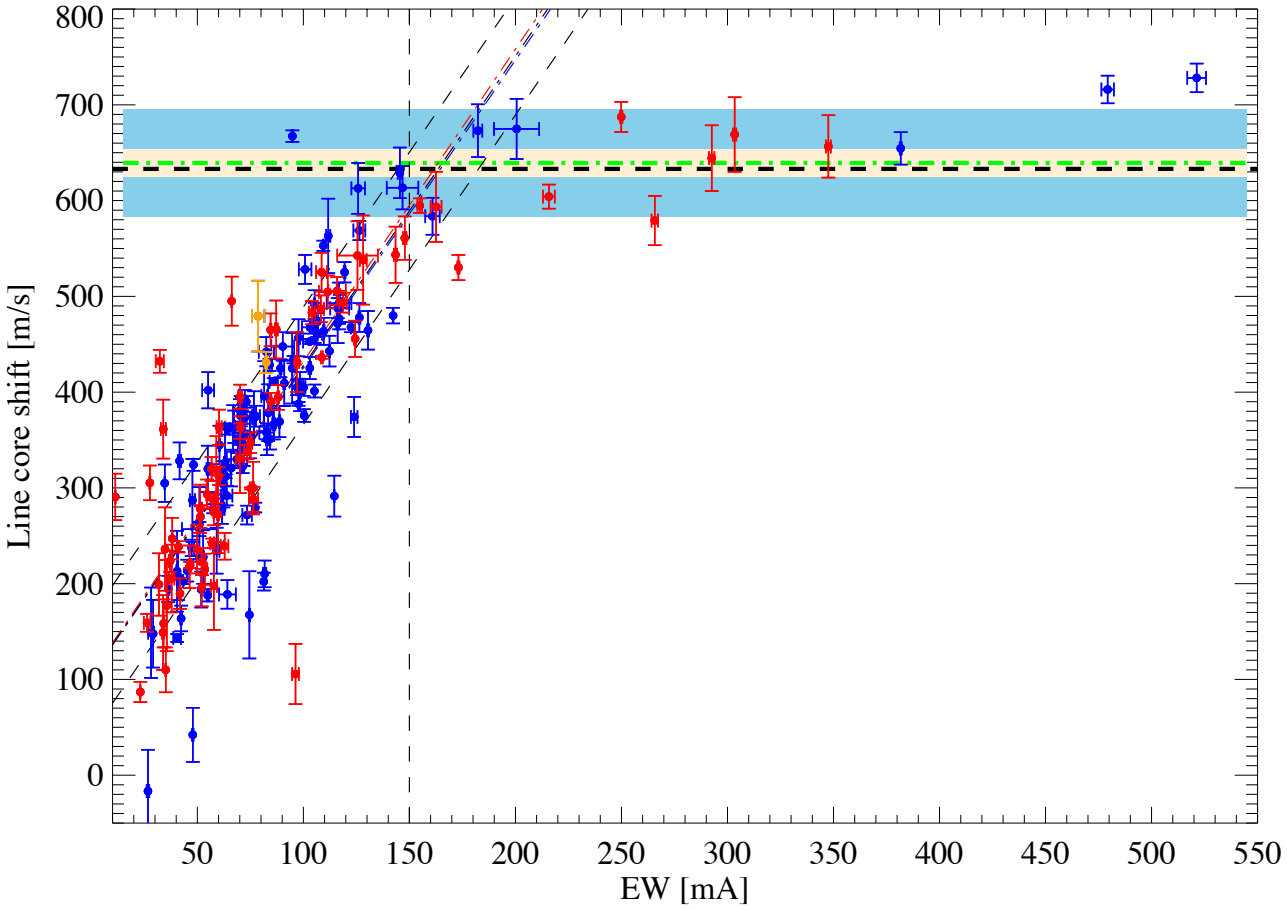


Figure 2.3: Figure 5 from González Hernández et al. (2020). Line core shift of FeI spectral lines, estimated using the recalibrated wavelengths λ_{nist} as reference laboratory wavelengths, measured on HARPS-LFC Moon spectra in the blue chip (blue dots) and in the red chip (red dots). Line core shifts of the two FeII lines are displayed as orange symbols. Blue, red, and black dashed-dotted lines show the linear fits of the blue, red, and all dots corresponding to the lines with EWs smaller then 170 mÅ. The horizontal black dashed line shows the theoretical solar gravitational redshift (GRS) measured on Earth, $v_{GRS,theo}$=633.10 ms^{-1}. The green dashed-dotted line shows the mean line core shift, $v_{GRS,obs}$, of the N=15 lines at EWs greater than 150 mÅ. The light blue and light yellow regions show the standard deviation, $\sigma_{GRS,obs}$, and the $\sigma_{GRS,obs}/\sqrt{N}$.

The very interesting study of González Hernández et al. (2020) reports measurements of the Moon reflected solar spectrum with HARPS, and laser-comb wavelength calibration system. The analysis of many observed Fe I lines shows a distinct pattern of the measured shift (difference λ_{obs} - λ_{lab} expressed in ms^{-1}) as a function of the EW of the line, showing the competition between the gravitational redshift (GRS) and the convective/granulation effect (their figure 5, reproduced on Figure 2.3). For strong lines (EW>150 mÅ), which form high in the photosphere, the convective effect is negligible, and the observed shift is due entirely to the GRS. For weaker line (EW from 0 to 150 mÅ), which form deeper and deeper in the photosphere when EW is smaller, the convective effect produces a negative shift down to $\sim-$ 500 m/s, almost compensating the GRS.

In this study, they find also that the value of observed wavelength λ_{obs} for a given spectral line depends on the extent of the wavelength region on which is determined λ_{obs}: either the core of the line around the minimum (5-7 pixels covering $\sim$ 3.2-4.8 kms^{-1}) or the whole spectral line up to the continuum fitted by a Gaussian (global line fit, after normalisation to a flat continuum). The difference may reach several tens of ms^{-1}, an effect linked to the asymmetry of the line, characterized for instance by the shape of the bisector of the line. In fact Dravins et al. (1981) studying high resolution solar spectra, have already pointed that "the bottoms of strong absorption lines should show smaller shifts since they form high up in the atmosphere where the granulation is no longer distinctly visible" and "the lines in different wavelength regions with different granular/intergranular contrast or different atmospheric opacity should show different amounts of line asymmetries and shifts".

2.2 The absolute radial velocity or time variations of the radial velocity of a star.

Now we must distinguish the question that we wish to address: either we wish to determine, at a certain time, the absolute value of the star's radial velocity dR/dt, or we are interested only by the time variations of this absolute radial velocity. Indeed, for the search of exoplanets, we are only interested by the motion of the star around the barycenter of the stellar system, induced by the presence of exoplanets in the system.

We discuss this question in the frame of the use of a Binary Mask (BM), composed of a series of spectral transitions defined by their wavelengths, and the Lafarga-type of algorithm (described at Lafarga et al. (2020)), in which the minimum of a CCF (between one line of the BM and an observed line) is determined. Actually, Lafarga et al. (2020) followed many previous authors.

2.2.1 Determination of the Absolute radial velocity dR/dt.

This is useful, for instance, for the study of the dynamics of stars or groups of stars. As a result of the description above about the absolute radial velocity, we may establish a set of general rules and rules for the constitution of the BM that would be built for the purpose of estimating the absolute radial velocity dR/dt:

- The lines of the BM must be taken as the best possible laboratory measurements.

- The GRS correction: the spectral type of the star should be determined, in order to get an accurate estimate of the surface gravity and compute the GRS of the star from the gravity potential at the surface of the star with radius R_*, GM_*/R_* then divided by the speed of light c. The surface gravity is g=GM_*/ R_*^2 and the stellar parameter is usually log g.

- The effect of granulation. If it is not accounted for, then, each individual line will give a different value for the RV of the star, affected by a "personal" bias, which may amount to several hundreds ms^{-1}; as described above, taking only strong Fe I lines (with large EW) will give the same (true) answer before GRS correction (and presumably strong lines of other elements): minimal effect of granulation, minimal bias. If one want also to use weaker lines, then a model of the star photosphere must be used, as in González Hernández et al. (2020), in order to determine and correct for the effect of granulation. The lines of the BM may be kept as the best possible laboratory measurements, and the granulation correction made afterwards, different for each line of the BM. Still, as we shall see later, the convective granulation blueshift is changing with time, up to $\sim$ 10 ms^{-1} (Meunier et al., 2010).

- The stellar oscillations may be averaged out, either by having an exposure time which is the same as the oscillation period (if it is known), or by having an exposure with a much longer time than the anticipated oscillation period, which depends on the spectral type.

- The stellar rotation : the quantity V_{rot}sin i can be measured from the line width of the lines, or determined from quasi-periodic variations of apparent RV

2.2.2 Determination of the time variations of the radial velocity of a star.

For the search of exoplanets, we are not so much interested by the absolute value of dR/dt, but only by the time variations of RV, revealing (or not) the presence of exoplanets around the given star. Therefore, we may completely ignore the GRS correction which is (almost) constant (if the eccentricity of Earth's orbit is ignored).

As shown on figure 2.3, for the convective Granulation Blue Shift (GBS), each line has a different bias, and this bias is anti-correlated with the EW of each line: weak lines have a strong bias, while strong lines have a small bias. If this bias were constant in time, it would not be so important to

keep each line with its own bias. However, as shown by several authors (e.g. Meunier and Lagrange (2020)) the bias may be changing rather quickly, producing a "jitter" in the RV. Moreover, some line shapes were observed to change also, complicating the concept of variable Doppler shift. We may consider that the variations of the convective GBS is presently the most important limiting factor for the accuracy of RV changes (DRV) determination. We address further this convective GBS in a later section 5.

Then comes the way to make an average of all RVs recorded for one exposure:

- Simple average over all lines, weighted by their individual uncertainty.
- or average inside each order of the spectrometer, with or without weights (Lafarga-type algorithm uses the weight=contrast/width of line, where the contrast is the relative depth of the line and is given for each line of the official ESPRESSO BM). This step is then followed by a combination of all RV_{order} determined for each order with appropriate weights connected with the uncertainties for each RV_{order}.

Whatever is the method to combine the various RVs, a few principles must be applied when one want to examine a series of exposures:
- the BM should be identical for all exposures.
- the weighting method should also be identical for all exposures.
- the details of the fitting process (to determine the minimum of the CCF) must also be kept identical for all exposures. In particular, to avoid the difference of shifts due to the core-of-line/ full line, the number of points kept to fit a CCF by a Gaussian should be kept constant, and preferably the same for all lines of the BM. This is what was adopted in the present study.

We conclude also that when only time variations need to be determined, the exact wavelength position of each line of the BM is not critical, since changing its absolute wavelength position will only change its bias by a constant value, and will not change at all the time variation of this line.

2.3 Overview of the radial velocity retrieval methods

After discussion about absolute RV and variation of RV we are coming back to the Radial velocity retrieval method. The radial velocity retrieval method (RV) is based on Doppler effect. One can measure the radial velocity of a star by measuring the wavelength shift of several known spectral

lines from high-resolution spectrum with respect to (w.r.t.) laboratory measurements and combine the shifts of the various lines to get an estimate of RV. If the observed star has a planet or a planetary system, it will move with respect to the center of mass of the star—planet system. This movement will produce a periodical shift towards the red part of spectrum (if star is moving away from the observer) and towards the blue part (if star moves towards the observer). If one computes the radial velocity of the star for a long time which covers the periods of a planet – it is possible to see changes of radial velocity which will follow a sinus-like curve (in the simple case of a circular orbit). This is a so-called indirect method of detection of an extra-solar planet (exoplanet).

The method of radial velocity was historically the first method used to detect exoplanets around a main sequence star with success: Pegasi 51 b was discovered in 1995 by the RV method (Mayor and Queloz, 1995). The difficulty of the method is due to the fact that planets of small masses, m or planets with a strong inclination of the orbit to the angle of view i cause only small periodic fluctuations of the radial velocity: low values of $m \sin i$ are extremely difficult to detect. Presently, the precision of radial velocity measurements is as low as 1 ms^{-1} for some of the best high-resolution spectrometers. Although the radial velocity method gave way to transit photometry with the main Kepler mission to observe exoplanet transits, in terms of numbers of exoplanet discoveries (1033 exoplanets detected by RV, 3751 exoplanets detected by transit), the RV method is still important for exoplanet research and continues to evolve. For example, the radial velocity method allows us to measure the projective mass $m \sin i$ of an exoplanet, which, in a transit configuration, is the real mass of the exoplanet ($\sin i$=1) ; the transit photometry method does not allow us to measure the mass of the planet. Therefore, the application of the two methods simultaneously allows us to obtain the most complete information about the exoplanet.

Up to date several approaches of Doppler shift RV measurements exist. We will discuss several of them which we have explored and used during our study.

- Cross-correlation function (CCF): allows to retrieve one estimate of the absolute radial velocity RV for each exposure. Then a time series of RV is examined to detect changes.
- Template matching: allows to determine the difference DRV of RV between one exposure and a template spectrum used as a reference.
- Pierre Connes algorithm: allows to determine the change DRV of RV between two exposures (at two epochs)

- Combination of template matching and Pierre Connes algorithm: allows to determine the change DRV of RV between one exposure and a template (e.g. Dumusque (2018); Cretignier et al. (2020); Artigau et al. (2022)

One thing which is common for all 3 approaches: the spectral regions affected by telluric contamination (which are affected due to the Earth's year motion) are taken out of consideration. Binary masks (an ensemble of spectral lines used to measure the RV of the star) which are used in the CCF algorithm have no lines in such regions. The two other methods must also avoid such telluric contaminated spectral regions.

2.3.1 Cross-correlation function (CCF) with a binary mask

The radial velocity RV of a star can be found by measuring the minimum of a cross-correlation function. The correlation is made between the observed stellar spectrum and a suitable ensemble of spectral lines (a binary mask, BM). Actually, a CCF may be obtained for each spectral line of the BM. The mathematical correlation for one line is made by changing the Doppler shift of the line, and one element of the CCF is the product of a boxcar attached to the Doppler shifted BM line and the stellar spectral line. One estimate of RV of the star is obtained at the minimum of this individual CCF. Combining all estimates for all lines (with their individual error) gives a more precise estimate of RV. The minimum of the CCF is usually determined by a Gaussian fit to the CCF.
The method has proved to be robust and stable, as well as simple and fairly fast.Actually, the cross-correlation method was first used mechanically with the CORAVEL spectrograph installed by the team of Michel Mayor (at Observatoire de Haute Provence, France (OHP). The CORAVEL combines all lines together, optically and mechanically (Baranne et al., 1979). A mechanical mask was implemented in the CORAVEL spectrograph focal plane, which is a chromium-coated glass plate with holes drilled in places corresponding to the stellar lines of the spectrum. There is a cross-disperser, so the mask in 2D, with one line per grating order. The mask is fixed in front of a photoelectric detector, focusing on the transverse scattered spectrum of the star. The light that passes through all the holes is refocused on the photoelectric detector, which counts the photons registered. The mechanism then optically shifts the spectrum of the star relative to the mask, by varying the angle of incidence of the light on the grating. When the spectral lines coincide with the holes in the mask, the number of photons detected is minimal and the radial velocity is calculated from the position of the mechanism.

It is worth noting that this approach has several disadvantages:

- All spectral orders are shifted simultaneously
- The spectrum is shifted, not stretched. In an ideal system, the spectrum is stretched because the Doppler shift is actually a stretch $\beta = \sqrt{\frac{1+\frac{v}{c}}{1-\frac{v}{c}}}$

It is also important to keep in mind that, depending on the spectral class of the star, the number and quality (relative depth) of spectral lines changes. Therefore, to improve the precision of the cross-correlation calculation it is necessary to create masks for each spectral class of the observed stars. One binary mask has been produced for CORAVEL, for the star Arcturus (spectral class K2 3), and the practical measurements have shown that with such a mask it is possible to measure the cross-correlation for stars of spectral classes later than F5 V. But it is also worth noting that there is a limitation in precision when using a slit spectrograph with a physical mask. Practice has shown that the instrumental precision is limited to about 200 ms^{-1} and this greatly limits the ability to search for giant planets or brown dwarfs orbiting stars. This important limitation is mainly due to the errors that are caused by the variable luminosity of the slit. In addition to the problem of variable slit illuminations, there is the problem of a fixed geometry of the pattern-mask, as it has been calculated for a well-defined index of refraction and this leads to a wavelength shift which cannot be fully compensated by a zero-point shift defined by the calibration when atmospheric pressure is changing. A precision of 200 ms^{-1} is a good result for a first development, but insufficient for detecting exoplanets.

The work of Baranne et al. (1979) describing the CORAVEL instrument was continued and a development of the method was published in Baranne et al. (1996). Like the previous work, the new paper is inextricably linked to the development of the spectrograph. The new spectrograph is named ELODIE and is based on the already working CORAVEL. This spectrograph will take its place in history as the discoverer, because it was the ELODIE spectrograph data that discovered the first exoplanet near a main sequence star in 1995 (Mayor and Queloz, 1995). A major improvement was the abolition of the physical mask. New large CCD arrays with low readout noise opened up the possibility to replace the optical correlation between the spectrum and the physical template with a numerical correlation between the CCD spectrum and the "software" template. The difficulties previously described associated with fixed mask geometry are completely eliminated by numerical correlation. Also, the advantage of numerical cross-correlation is the possibility to construct different templates

for different classes of stars or measurement targets. And the problems associated with input slit illumination are solved by replacing the input slit with a fiber optic system, including an optical "scrambler", which minimizes the fluctuations in position of the star spectrum on the CCD detector. The algorithm for calculating the cross-correlation function is very similar to the optical cross-correlation process. The spectral pattern consists of rectangular emission lines of equal amplitude and acts as a real physical spatial filter for the CORAVEL spectrograph. The cross-correlation function is calculated from a calibrated 2-D spectrum, with associated wavelengths. Based on the wavelength bounds of each spectral order and velocity range, the algorithm selects, for each order, the lines that will always be in the wavelength range during the scan. The cross-correlation function is calculated step by step for each velocity value V, without joining orders or splitting them into new areas.
The formula describing the calculation of the new cross-correlation function is as follows

$$C(V) = \sum_{l} \sum_{x,o} p_{l,x,o}(v) f_{x,o} \tag{2.1}$$

Where $f_{x,o}$ – is the value of the 2-D spectrum at order o and pixel x, $p_{l,x,o}$ is the fraction of the l line of the pattern that hits pixel (x,o) at speed velocity V. The resulting function C(V) can be fitted by a Gaussian. This is an elegant way to combine the CCF of all the lines together, rather than combining the results that could be obtained for each line.
Let us refer to the paper by Mayor and Queloz (1995), which describes the discovery of the first exoplanet near a main sequence star. At that time, spectrographs aimed at studying the effects of the Doppler effect reached an accuracy of 15 ms^{-1}. However, the Sun response to Jupiter's gravitational influence is 13 ms^{-1} over a (long) period of 12 years, so all the investigations carried out at the time were restricted to a minimum mass of an object equal to the mass of Jupiter. Mayor's group was able to achieve an accuracy of 13 ms^{-1} for about 100 stars, with 18 months of measurements of radial velocity variation, and a small number of stars showing significant variability in radial velocity. One such star was 51 Pegasi, which is a sun-like star hosting a planet with minimum mass of 0.5 Jupiter masses and a maximum mass of 2 Jupiter masses; the distance from the star to the planet is only 0.05 astronomical units (1 a.u. is the distance from the Sun to Earth). Such a small distance between the star and its companion (planet) was unexpected and not predicted by the then existing models of giant planet formation. And for a complete revolution around its parent star, planet 51 Peg b needs only 4.23 days. After the discovery of the first exoplanet, the number of candidate exoplanets (detected

in the data, but waiting to be confirmed as actually existing and not as an error of observation or interpretation) began to grow rapidly. And at the time of writing (Pepe et al., 2002) the number of candidate exoplanets was 80. Pepe et al. (2002) describe the discovery of two exoplanets with masses approximately equal to Saturn, with the minimum mass of the planet HD168746b being only 0.77 Saturn masses, one of the 4 known sub-saturn planets at the time. The discovery of exoplanets with a mass less than that of Jupiter was made possible in large part by improvements in the cross-correlation method. The cross-correlation method described earlier is not optimal for obtaining radial velocity information. For example, deep and sharp lines have no greater influence, although they contain more radial velocity information than do broad and weak lines. Lines have different relative depths, which are averaged into the resulting cross-correlation function. An improvement is to create a weighted cross-correlation function which takes into account the relative importance in RV retrieval of each spectral line contained in the mask. The essence of introducing weights can be explained in the following way: Let us imagine that there is line l with relative depth c_1=0.5 and continuum level S_1=I_0, and there is line m with relative depth c_m=1 and continuum level S_m=I_0/2, their contribution to the global cross-correlation function is the same, the signal is proportional to the product of relative depth and continuum level, and equal to I_0/2 in both cases. And the noise of each point function is proportional to the square root of the real spectral signal S_i and it is $\sqrt{2}$ times larger in the cross-correlation function of line l. If we approximate the cross-correlation functions of each line independently, we would get $\sqrt{2}$ times the accuracy for line m, so the weight of line l should be higher.

For a long time, researchers have looked for exoplanets in stars similar to the Sun or, to be more precise, in stars of the FGK spectral classes. Stars of these types show in their spectra a number of spectrally resolved and unmixed absorption lines in the visible wavelength range and are well suited to representation as a weighted binary mask. But in recent years there has been increasing interest in M-class stars because, due to the lower star-to-planet mass ratio and the closer location of the habitable zone, the gravitational pull of a planet orbiting in the habitable zone of the parent star is stronger for M dwarfs than for hotter and more massive FGK-class stars. This gravitational influence creates a larger amplitude of radial velocities, and the shorter orbital periods allow for shorter observational durations. All these features of M dwarfs make it possible to detect small, rocky exoplanets orbiting stars of this type, and M type stars are also the most common stars in the vicinity of the Sun, so they can be

considered the most common potential hosts of exoplanets, especially those most interesting orbiting in the habitable zone of the star. The main problem of using the radial velocity method to study M dwarfs is the high level of stellar activity. In the spectra, the activity appears as distortions in the spectral line profiles used to calculate the radial velocities. In addition, due to the lower temperatures of M-class stars, the number of their spectral lines is much higher than that of FGK-class stars, and most of the spectral lines get mixed, which makes it impossible to compose a sufficiently accurate mask.

In Lafarga et al. (2020), the authors focused on the cross-correlation method for M dwarfs in the visible and near-infrared regions of the spectrum. The masking approach differs from those previously discussed, although it still builds on the basics. The authors created the masks based on a combination of multiple observations of stars of the same type. After this, there is one template for several stars of the same type; the next step to create a quality mask is to select suitable lines. Each line is characterized by 3 main parameters: half-width, depth and contrast (depth relative to the adjacent continuum). The contrast in this case is 2 values, being the difference between the light around the line and the light at the line minimum, the difference of the obtained 2 contrast values can be used to study the asymmetry of the lines. After a practical investigation of different masks, the authors concluded that it is best to select lines with half-widths within a certain range, while taking care that most of the lines selected have the same width, avoiding the narrowest lines, which tend to correspond to weak noisy lines not previously selected. As mentioned earlier, weights can be added to each line to improve cross-correlation accuracy. In this paper the weight is calculated as the contrast divided by the line half-width, so that the lines with the steepest profiles (i.e. those that contain the greatest amount of radial velocity information) contribute the most to the cross-correlation function. We should mention that since 03 February 2023 the official ESPRESSO pipeline mask use FWHM and contrast as a weight in mask (before that only contrast was used), as stated at the ESO website (https://www.eso.org/sci/software/pipelines/espresso/espresso-pipe-recipes.html). In wavelength spectral regions which are/or may be affected by telluric absorption, there are no lines in classical masks which are used in spectrographs pipelines. When a telluric correction is performed in a spectral region, one has to create a new mask of stellar lines, as have been discussed in 1.5.1.

2.3.2 Template matching

The approach of template matching was first described by Chelli (2000), who proposed the method of correlating, in Fourier space, each spectrum of the target star and a reference spectrum specific to that star - template. The method had been applied to solar-type stars. In Galland et al. (2005) authors applied this method to the earlier type stars, from A to F spectral class. One of the recent papers about template matching technique is Anglada-Escudé and Butler (2012). The algorithm is based on the minimization of the difference between an observed spectrum and a template. The idea is that the template should differ from data only by the Doppler shift (actually, stretch) due to the Earth's motion and/or companion presence. To account for some instrumental effects authors propose to empirically re-evaluate the uncertainties for each grating order. The approach mainly consists of minimizing χ^2 , which is a weighted sum over all the observed wavelengths of differences R between the template and the observed flux at each wavelength. This approach was slightly changed in Zechmeister et al. (2018); the authors have included a polynomial in the model to the template, while Anglada-Escudé and Butler (2012) have applied it to the data. Anglada-Escudé and Butler (2012) did not re-adjust errors the way it is done at Zechmeister et al. (2018) and the polynomial function variations should be a few percent, so both approaches are feasible.

With a model of Earth's atmosphere absorption one can set weights equal to 0 to the wavelengths which correspond to the telluric lines. The procedure minimizing the χ^2 will not account for lines with 0 weights, so the telluric contaminated regions will not produce any bias and any RV information either.

During the course of this work, our first idea as an algorithm to retrieve the RV of the star was to apply a template matching method of the observed spectrum to a model of the stellar spectrum, with one parameter of the fit being the star RV. However, we had to discriminate the telluric absorption lines imprinted on the spectrum from the stellar lines. For this purpose, we used the TAPAS model (Bertaux et al., 2014): more details about TAPAS are written in section 1.2.

It is important to recall here that the wavelength assigned to each pixel of one ESPRESSO spectrum by the calibration pipeline is not in the frame of the laboratory, but rather is corrected for the known motion of the Earth (orbit around the Earth and diurnal rotation). The wavelength of each pixel is given in the so-called Solar system Barycentric system. The radial velocity of the Earth in this system along the line of sight is the so-called BERV (Barycentric Earth Radial Velocity, BERV). The BERV

velocity correction is found in the header of each exposure. Therefore, the position of a telluric absorption returned by TAPAS may be put also in the BERV system. Doing this, we found out that telluric lines predicted by TAPAS are at the right place, in good correspondence with the observed data (most often, better than $\sim$ 1 ms^{-1}) , which demonstrates the exquisite wavelength calibration of ESPRESSO spectrometer as well as the quality of the laboratory wavelength used in TAPAS.

Building of a star spectrum model (synthetic).

According to the chosen strategy it requires 2 main steps to obtain a stellar synthetic spectrum:

- The first step is to create a stellar model. It is done using ATLAS9 code (Kurucz, 2017a) with parameters corresponding to the observed star. Such parameters are: temperature, metallicity and gravity. Scripts that were used were provided by Piercarlo Bonifacio along with a grid of precomputed models.

- The second step is to take a stellar model from step 1 and create a synthetic stellar spectrum on its basis. It is done by using SYNTHE code (Kurucz, 2017a). Scripts to run correctly SYNTHE were provided to me by Piercarlo Bonifacio.

It is important to indicate which line list should be used for a synthetic spectrum creation. During the current study two line lists were tested: Castelli line list (Castelli and Kurucz, 2010) and Gaia-ESO line list (Heiter et al., 2021). The main difference between two mentioned line lists is the source of the lines parameters. The Castelli line list is based on the Kurucz line list with addition of new lines or refinement of already included lines. Parameters of lines are calculated semi-empirically, energy levels, wavelengths, and transition probabilities are calculated theoretically Kurucz (2011, 2017b, 2018) with additional check of some lines with experimental data (Peterson and Kurucz, 2022).
The Gaia ESO line list is constructed as follows: the highest priority is given to the laboratory measurements; in case such laboratory data are not available, the data is taken from the semi-empirical approach of Kurucz. Astrophysical gf-values were avoided (Heiter et al., 2021).
Two synthetic stellar spectra with equal star atmosphere model parameters but with different line lists were produced in order to check which line list is better for the current study. The comparison of synthetic and ESPRESSO observed spectra was done order by order, and chunk by chunk (a chunk is a small piece of one order, which may encompasses several lines) for orders which are not affected

by telluric contamination, to avoid any bias caused by it. Each chunk of observed spectrum was fitted by a "fitting" model. This fitting model is obtained from the stellar synthetic model as follows:

- a preliminary Doppler stretch 31 kms^{-1} is applied, which is the approximate RV of star HD40307
- then it is multiplied by a polynomial (which coefficients are determined in the best fit process), to fit the level of signal in ADU, and to account for "local" variations of the relative spectral sensitivity of ESPRESSO
- shifted by an additional ΔRV

In theory both fit by the two synthetic models should return the same, or almost the same ΔRV values, but sometimes the best fit retrieved values of ΔRV differ greatly.

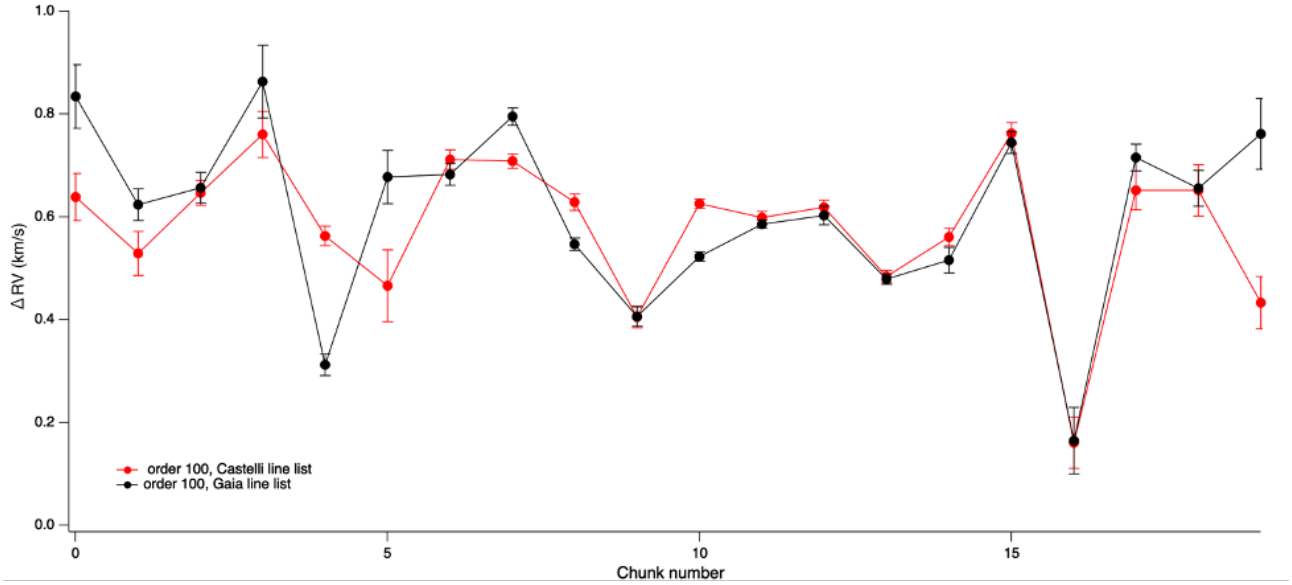


Figure 2.4: Example of chunk-by-chunk ΔRV extraction by model template matching; Order 100 (not affected by tellurics), star HD40307; Result of fit by the synthetic spectrum based on Castelli line list is in red; by the synthetic spectrum based of Gaia-ESO line list is in black. The error bar on ΔRV is the one returned from the best-fit process.

On figure 2.4 one of the examples is shown. The chunk 4 from Figure 2.4 is shown on Figure 2.5 for further examination. It shows that the line around pixel 2268 in Gaia-ESO line list is shifted w.r.t. the observed line, while there is a Castelli line at the right place.

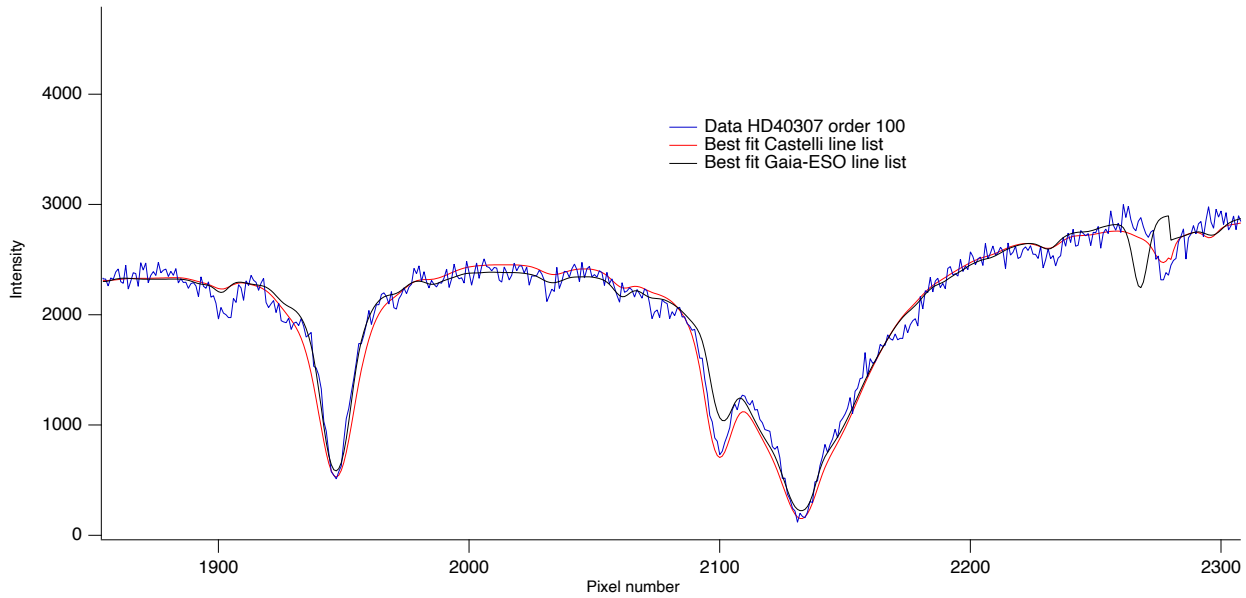


Figure 2.5: Chunk 4 from figure 2.4. Blue – original data; Red – best fit by synthetic spectrum based on Castelli line list; Black - best fit by synthetic spectrum based on Gaia-ESO line list. Note that the line of Gaia-ESO line list, near pixel 2268, is misplaced versus the observed line and Castelli line.

The overall fit of both models to data is rather good, but seems better with the Castelli line list. Another important example is shown on Figures 2.6 and 2.7, it is a case when the observed line is absent in the Gaia-ESO line list.

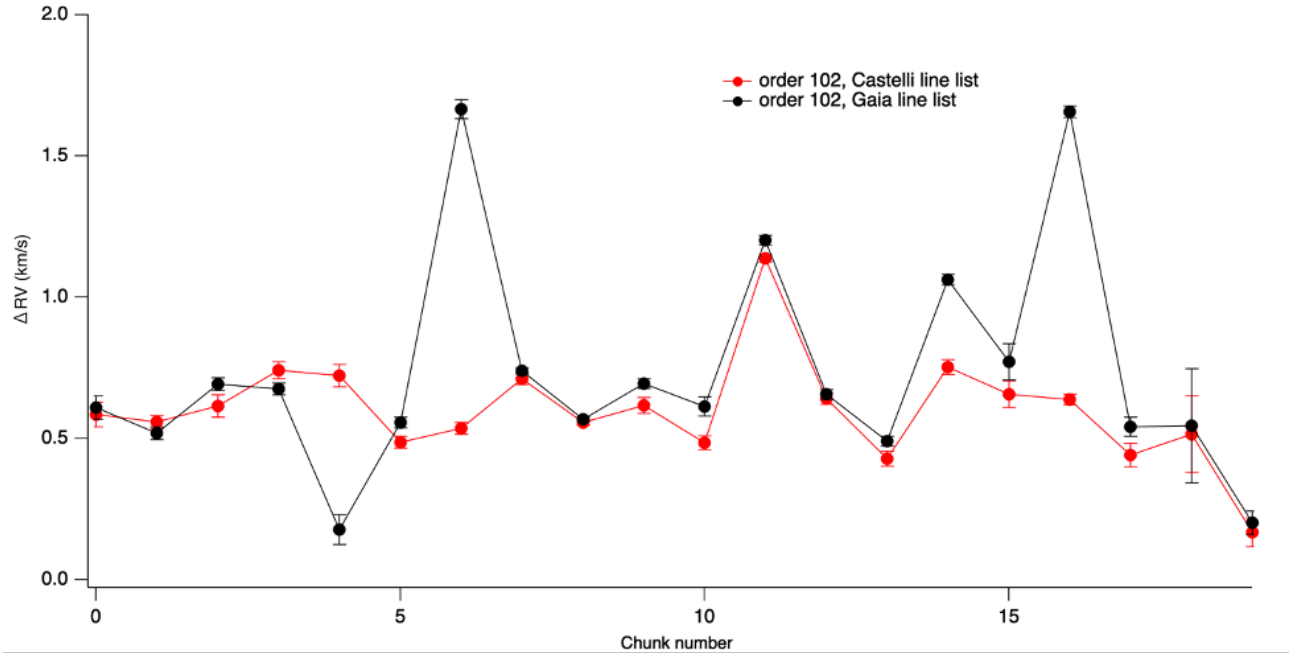


Figure 2.6: Example of chunk-by-chunk ΔRV extraction by template matching; Order 102 (not affected by tellurics), star HD40307; Result of fit by synthetic spectrum based on Castelli line list is in red; by synthetic spectrum based of Gaia-ESO line list is in black

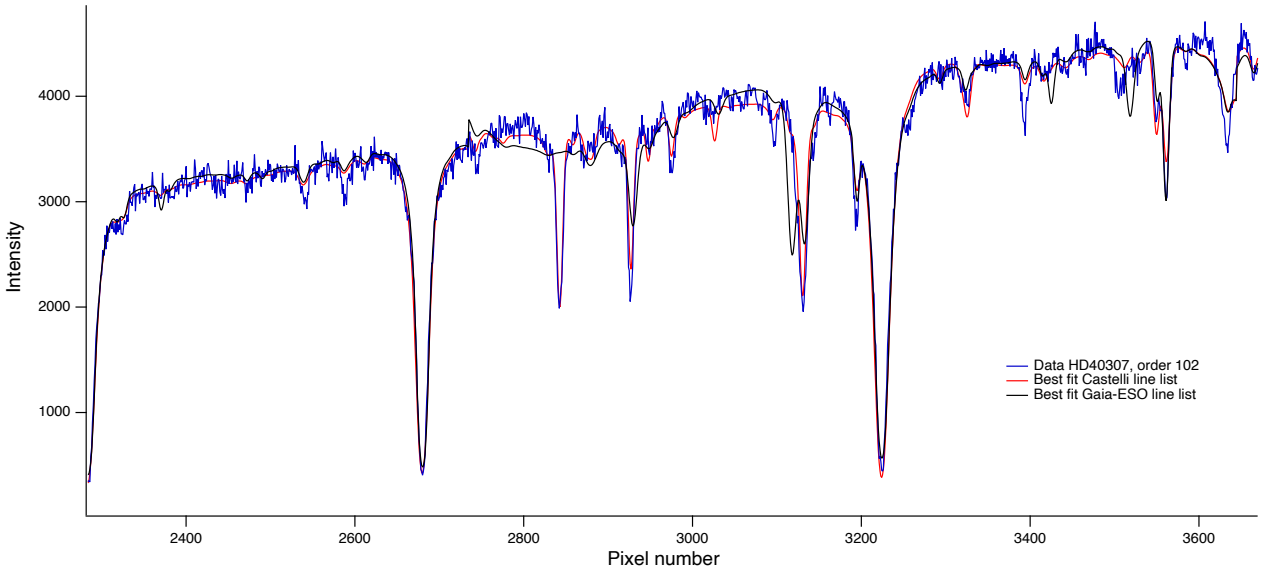


Figure 2.7: Comparison of a piece of observed spectrum with both best fit models. Example of case when the Gaia-ESO line (black curve) list is missing a stellar line, near pixel 2850. Also, this model displays a line around pixel 3100 which does not exists either in the data or in the Castelli model.

After examination of several orders, it was decided to use the Castelli line list as it was more suited to the observations.

However, we were concerned by the fact that, even with the stellar model based on the Castelli list, the variation of the retrieved ΔRV from chunk to chunk (Fig. 2.4 and 2.6) was very large, several hundreds of ms^{-1}, even $\sim$700 ms^{-1} for chunk 11 of Figure 2.6. And this variation was much larger than the error bars. It casted some doubt in our mind about the applicability of the stellar model to our problem.

At that time, we were not fully aware of the work of Gray (2009); Reiners et al. (2016); Liebing et al. (2021); González Hernández et al. (2020) on the position of solar lines, showing that most of spectral features are affected by the granulation blue-shift (GBS), with a magnitude of several hundreds of ms^{-1} w.r.t. the laboratory measured wavelengths. This GBS is different for all lines, and can be accounted for in stellar atmosphere models. The models that we used did not do a perfect simulation of the GBS.

Therefore, to avoid such problems with models, we explore another type of algorithm, still doing template matching, but replacing the star atmosphere synthetic template by a high SNR based on the data themselves, as described in the next section. We will see that this course of action failed, because of an unexpected instrumental problem: the presence of minute fringes of interference (produced by the optics of ESPRESSO). This is why we came back to the more classical algorithm (for RV

retrieval) of using the CCF, piled up for each order, and the use of a standard Binary Mask of spectral lines (Figure 1.27). Actually, we figured out that a substantial number of grating orders gave retrieved RV some hundreds of ms^{-1} from the average over all orders. The reason is again the dependence of GBS on a line-by-line basis, since the wavelengths of the BM lines are based also on laboratory measurements of transitions. The GBS effect is quite detrimental when one is interested by the absolute value of RV. However, we are only interested by *time variations* of RV. If the GBS were systematic and stable, it should not be dramatic for the retrieval of such time variations of RV, and in retrospect we now think that the method of star model template matching could be used, as well as the CCF method that we finally used to evaluate the gain in RV retrieval when new parts of the spectrum are made available after correction of the telluric absorption. However, the GBS is not stable, and the template matching method will encounter the same problem as the CCF method. Still, if the SNR is high enough, the template matching method could allow to detect, on a line-by-line basis, RV changes due to changes in the GBS.

Data template matching after correction of telluric absorption.

Everything was done for the one night, the data set of 200 exposures taken on 25 December 2018 of HD 40307. As a first try only orders with water vapour (and corrected from) were taken into account. The new idea was to correct spectra from telluric absorption, and stack many exposures to create a high-SNR spectrum to be used as a template reference, against which may be determined the variation of RV of each individual exposure. Each exposure was corrected from telluric absorption; after that all 200 exposures were stacked pixel by pixel into a "dry superspectrum", which was used as the reference a template afterwards. A classical fitting with Levenberg-Marquardt was used (within IGOR software). The radial velocity (w.r.t. the reference template) was a parameter of the fit. As previously each spectrum was divided on a number of chunks, each chunk was fitted separately, after the results were averaged for each order and for each exposure. The result of RV measurements is shown on figure 2.8 (it should be noted that since stacked spectrum is used as a reference template, the obtained RV is not an actual RV (which should be around 31 kms^{-1}) but the RV with respect to the actual RV of the reference template (kind of DRV, not to be mistaken with the ΔRV of figures 2.4 and 2.6. DRV=0 for the reference template spectrum).

To our great surprise and disappointment, we found that over the course of the $\sim$ 4 hours of

measurements, DRV changed from $\sim$ + 15 ms^{-1} to $\sim$ -20 ms^{-1}. Even for a star with planets, such accelerations are just impossible. On figure 2.8 is also plotted the BERV as a function of exposure number. It varies from -500 ms^{-1} to -700 ms^{-1}. One may note that the change of radial velocity follows a linear slope, as well as the variation of BERV (Barycentric Earth Radial Velocity) change during the night. Therefore there is a hint that the variation of the BERV has something to do with the spurious change of RV that we found.

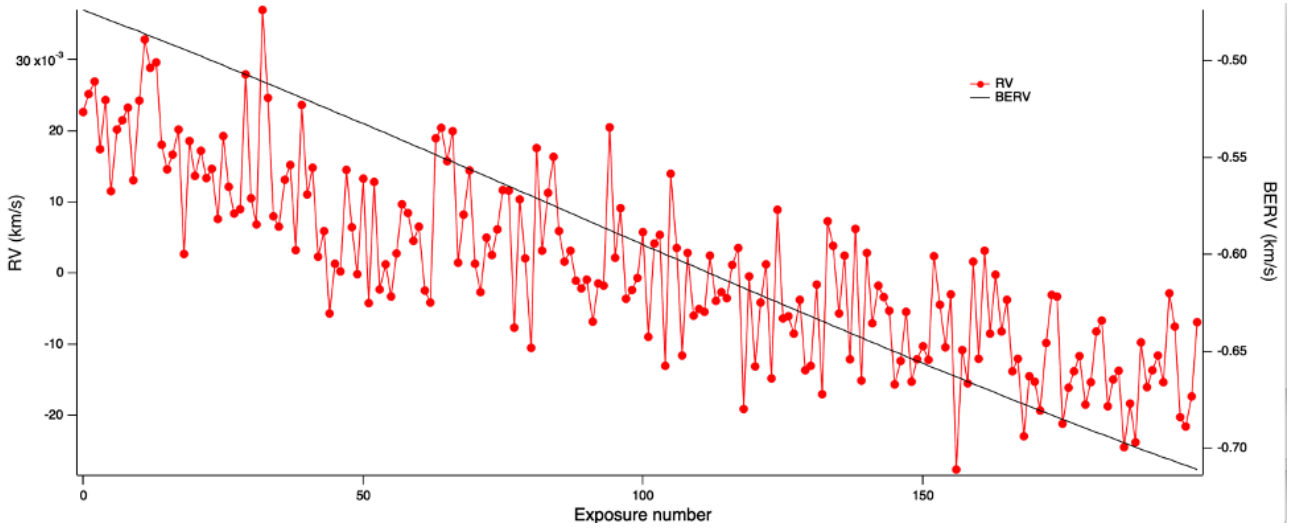


Figure 2.8: Result of fit with a "dry superspectrum". Red – radial velocity DRV of star w.r.t. the reference of stacked spectrum (left scale); black – Barycentric Earth Radial Velocity (right scale). The DRV changes from + 20 ms^{-1} to -15 ms^{-1}, while it should stay constant. The BERV is changing from -470 ms^{-1} to -710 ms^{-1}

In order to understand what is happening, it is useful to remember what happens during the night on the CCD detectors of ESPRESSO. Because the spectrometer is extremely stable, the telluric absorption are staying exactly on the same position in pixels, while the spectrum of the star is slowly moving by about half-a-pixel over the $\sim$ 4 hours of measurements because of BERV changing. Of course, the wavelength assignment (of every pixel) for each exposure is changing also, since the laboratory wavelength calibration is modified by the (very well) known BERV, and attached to the data. To the "dry super spectrum" template we attached the wavelength scale of exposure 100, just in the middle of the series. Since we are doing the template matching of each single exposure with the "dry superspectrum", where all atmospheric features have disappeared by atmospheric correction, we should retrieve only the RV change due to planets, well below 1ms^{-1} for $\sim$ 4 hours of measurements. On the contrary, if all exposures (and the superspectrum) are not perfectly well corrected from tellurics, then we may understand that the remaining, not well corrected, system of tellurics, will force the template matching process away from 0, and towards a BERV motion, depending on the

relative strength of the telluric system and the stellar system of spectral lines.

To illustrate this effect, we have also made a template spectrum by stacking all spectra, without any correction of tellurics, and examined the evolution of the retrieved RV by the data template matching method, but we kept the retrieved RV for each order separately. The result is illustrated on figure 2.9 for the 1st and last exposures (exposure number 0 and 199). For heavily telluric contaminated regions after order 146 (by H_2O and O_2) the retrieved RV changes from $\sim$ + 100 ms^{-1} to $\sim$ -100 ms^{-1}, as expected from the full BERV excursion (from -470 ms^{-1} to -710 ms^{-1}) for spectral regions dominated by telluric absorption. What was fully surprising was that a much smaller, but still well present, change of retrieved RV was also seen in all orders, including the ones that were not supposed to be affected by tellurics.

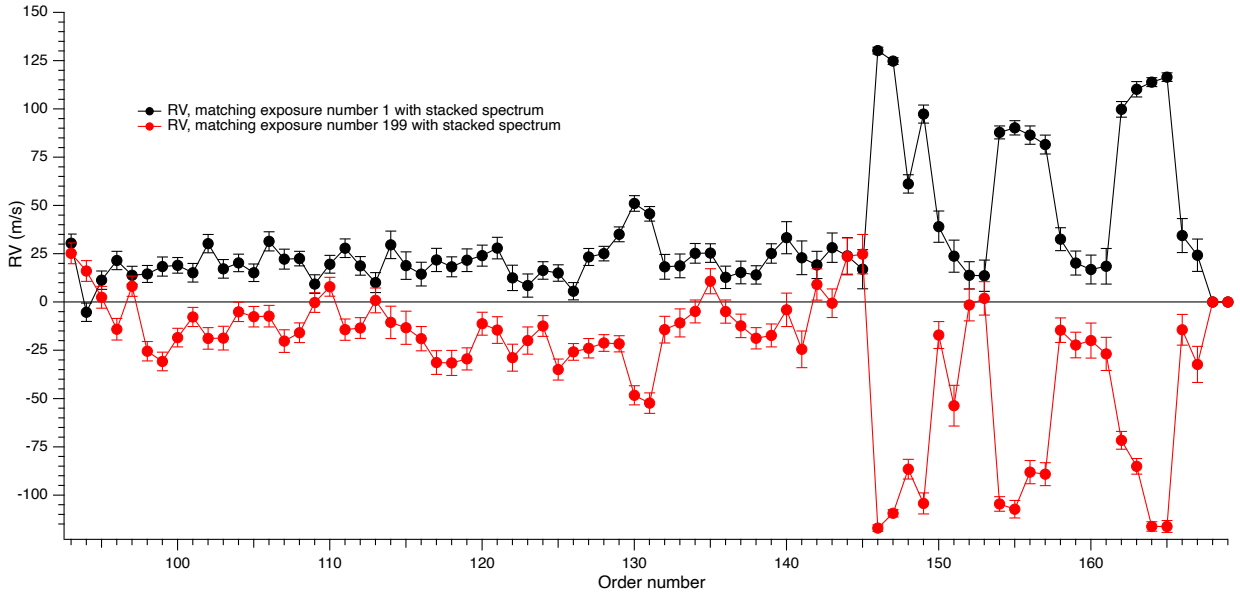


Figure 2.9: The retrieved RV value from the template matching method is displayed as a function of order, both for the 1st exposure (number 0, black) and the last exposure (number 199, red)

For instance, we studied in more details the order 124 which is void of telluric contamination, by dividing it in 20 “chunks” of about 450 pixel each. A negative slope of RV with time was found for almost all chunks, but quite variable from chunk to chunk (figure 2.10):

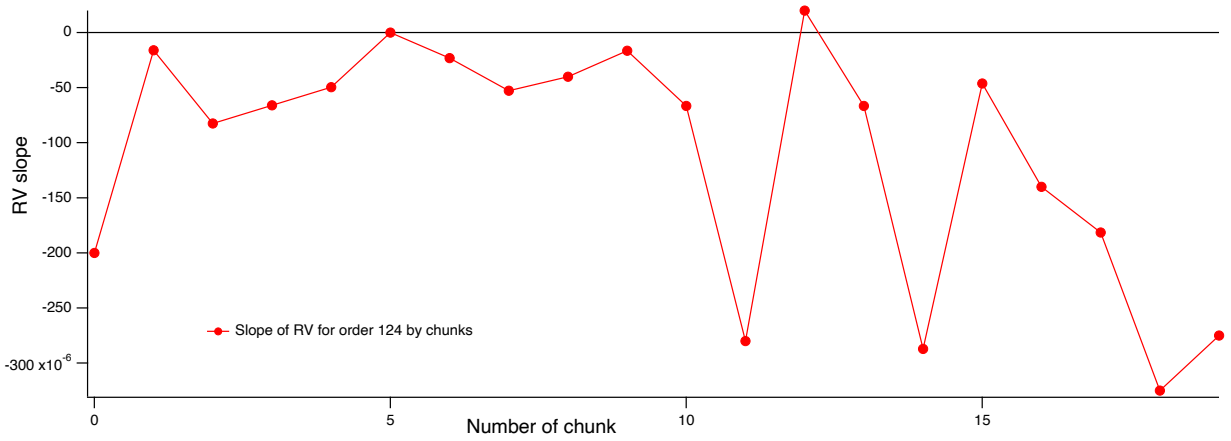


Figure 2.10: The trend of time variation of retrieved RV is characterized by a slope, which depends on the particular chunk of order 124. A slope of -100 10^{-6} translates into a total variation of – 20 ms^{-1} over the 200 points.

Figure 2.11 displays the star spectrum of order 124, and also the negative slope found in figure 2.10, at the position of the 20 chunks. When looking at the order, strong negative slopes correspond to the absence (or small number) of strong stellar lines. Slopes close to zero are in general with deeper or more numerous stellar lines. This finding illuminates the mental representation of the two systems of lines described above: the stellar one (that we want to track) and another system attached to the pixels. It could not be telluric, or even micro-tellurics, because we corrected from them using TAPAS.

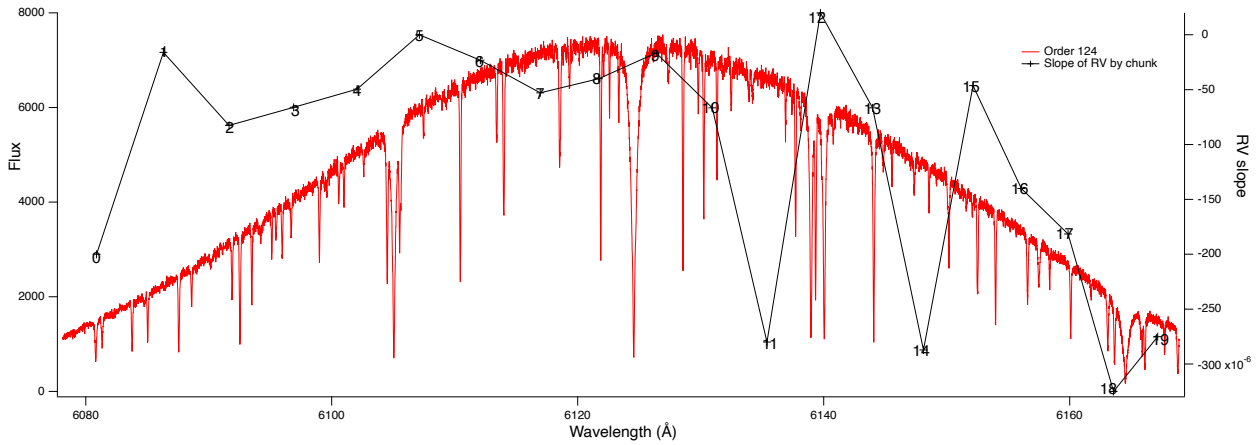


Figure 2.11: In red (left scale): Intensity of star stacked spectrum of order 124 (in principle, not affected by tellurics). In black (right scale): slope of RV with time over the 200 exposures for each chunk. The chunks with a large slope (11, 14) correspond to chunks where stellar lines are weak.

It took a while to understand that it could be due, not to tellurics, but to small interference fringes produced by the instrument optics. After some investigation and literature check it was confirmed that there are indeed in the ESPRESSO spectra so-called “wiggles”, which may be found in ESPRESSO continuum and caused by the optical interference patterns of the Coudé train (Allart et al., 2020;

Tabernero et al., 2021), ESO manual. Actually, we may also foresee that any PRNU (Pixel-to-Pixel Non Uniformity) remaining after Flat Field correction will have a similar effect, and both fringes and uncorrected PRNU, which are imprinted always in the same pixels, as are the tellurics, are a limiting factor of the precision of the data template matching method. Also, the fringe pattern of ESPRESSO, being associated with the Coudé train, is somewhat variable with the variation of the LOS to the star along the night, and may move somewhat across the pixels, increasing its power of disturbance of the method. Therefore, we could not continue on this line of RV retrieval method and we had to come back to the more classical method of CCF fit by Gaussians, in order to estimate the gain attached to the correction of tellurics.

2.3.3 Pierre Connes algorithm

In 1985 Pierre Connes Connes (1985) presented his Absolute Astronomical Accelerometry (AAA). The concept proposed by Pierre Connes (PC) completely differs from the CCF approach. While in classical CCF we work with stellar lines, taking them into account in a mask, weighting them accordingly to their contrast or FWHM (trying to account for their content of RV information), in PC-algorithm he stresses the point that his method is not line-based: "we refuse to pose the problem in terms of lines". Instead of line-approach, the task is done with the full spectrum as a continuous function. If the star is stable, without any motion, we have to register the same spectrum every time we observe the star. "Same spectrum" means that at each spectrum wavenumber we detect a same intensity, but if star has some motion and shifts during observations – we will see, that intensity at the wavenumber has changed.

Following the notation from Connes (1985) let A(σ) is the spectrum we observe at the beginning of observations (epoch 0); and $A_n(\sigma)$ is the spectrum we observe at epoch n, when spectrum stretched, σ is the wavenumber. $\delta V_n = V_n - V_0$ is the velocity change; wavenumber shift is $\delta\sigma_n = \sigma\delta V_n$ (Connes used wavenumber, which is $10^4/\lambda(\mu\text{m})$).

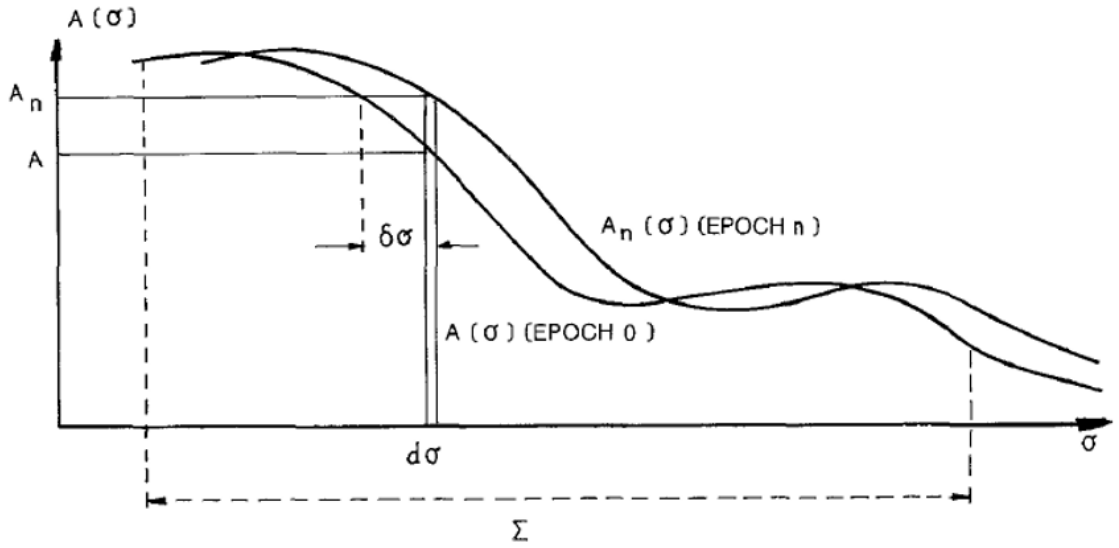


Fig. 4. An arbitrary spectral profile at epoch 0 and epoch n; the intensity change is measured within a given $d\sigma$ slice.

Figure 2.12: Original graph from Connes (1985). Illustrates the approach.

The observable intensity change at wavenumber σ is given by

$$A_n(\sigma) - A(\sigma) \simeq \frac{\partial A}{\partial \sigma}\delta\sigma_n = \frac{\delta V_n}{c}\sigma\frac{\partial A}{\partial \sigma} \tag{2.2}$$

We can rewrite this equation in terms of pixels and displacement δsp:

$$dA(i) = A_n(i) - A(i) = -\frac{\partial A(i)}{\partial \sigma}\delta sp(i) \tag{2.3}$$

[actually we note that in the Connes paper of 1985, the sign – is lacking by mistake in his formula (2). On his figure 2.12 the displacement δsp(i) from epoch 0 to epoch n is positive, the derivative is negative, and the intensity is increasing from $A_0 = A$ to A_n].

So it is possible to extract for each pixel i the mechanical (or physical) shift δsp(i). We call it the first formula of Connes:

$$\delta sp(i) = -\frac{A_n(i) - A(i)}{\partial A(i)/\partial sp} \tag{2.4}$$

Equation 2.4 can be transformed into a Doppler shift and the corresponding change ΔRV(i) of the radial velocity of the star between two epochs. We define sp(i)=$\frac{\lambda}{\Delta\lambda}=\frac{c}{\Delta V}$ (some kind of "velocity

resolving power" of one pixel). We have then :

$$\frac{\delta V(i)}{c} = \frac{\delta sp(i)}{sp} = -\frac{A_n(i) - A(i)}{sp(i)\partial A(i)/\partial sp} \tag{2.5}$$

Then the ΔRV(i) have to be combined accounting for their individual errors σ(i) (not to be confused with the wavenumber of Connes, that we will no longer use). In the case of Gaussian errors, it is known that the optimal combination is to use weights = 1/ σ^2(i), while the error σ(RV) on the combined retrieved RV is such that

$$\frac{1}{\sigma^2} = \sum_i \frac{1}{\sigma_i^2} \tag{2.6}$$

Variance σ(i) of $\delta V(i)$ can be calculated by the relation Var(aX+bY)=a^{2}Var(X)+b^{2}Var(Y)

$$\sigma^2(i) = \frac{1}{sp(i)\partial A(i)/\partial sp(i)}^2 Var(A(i)) + \frac{1}{sp(i)\partial A(i)/\partial sp(i)}^2 Var(A_n(i)) \tag{2.7}$$

Variance of a number of photoelectrons A is equal to A, considering that A and A_n have the same intensity level it is possible to set A_n=A. Hence the equation 2.7 will be:

$$\sigma^2(i) = 2(\frac{1}{sp(i)\partial A(i)/\partial sp(i)})^2 A_n(i) \tag{2.8}$$

If A is a perfect noise-free spectrum, then equation 2.8 will be the same, but without factor of 2. Following the notations of Bouchy et al. (2001) who revisited the Connes approach, the next step is to introduce the weight function W(i).

$$W(i) = \frac{1}{\sigma(i)^2} = \frac{(sp(i)\partial A(i)/\partial sp(i))^2}{2A_n(i)} \tag{2.9}$$

The velocity change then is:

$$\frac{\delta V}{c} = \frac{\sum \frac{\delta V(i)}{c} W(i)}{\sum W(i)} = \frac{\sum (A_n(i) - A(i)(\frac{W(i)}{2A_n})^{1/2}}{\sum W(i)} \tag{2.10}$$

Also Connes (1985) has shown that the uncertainty on a RV measurements can be calculated *a priori* on the basis of photon noise. Even if there is no velocity change in between epochs due to the motion, we can find small velocity change caused by the noise perturbation of spectrum at epoch n. There are two cases: pure photon noise and pure detector noise. In the following we consider the

photon noise, which corresponds to the highest precision. The approach is based on the quality factor Q which can be computed for any star, is independent of the flux (if we neglect the detector noise contribution), and in fact is the photon noise limit which can be achieved to determine the absolute wavelength position of a piece of an observed spectrum.

$$\delta V_{min} = \frac{c}{Q\sqrt{N_{e^-}}} \tag{2.11}$$

where c is the speed of light, and N_{e^-} is the total number of photoelectrons counted over the whole spectral range considered, and Q is the quality factor equal to

$$Q = \frac{\sqrt{\sum W(i)}}{\sqrt{\sum A(i)}} \tag{2.12}$$

We call it the second formula of Connes. It should be noted that the observed spectra are given in ADU units. In order to convert into a number of electrons, one must multiply by the so-called "gain" of the CCF read-out system, the number of electrons per ADU, which is included in the headers of ESPRESSO spectra (0.9 electrons per ADU).

$$\delta V_{min} = \frac{c}{Q\sqrt{N_{e^-}}} = \frac{c}{Q\sqrt{0.9 * N_{ADU}}} \tag{2.13}$$

The quality factor Q is best estimated with a noise-free spectrum, which is possible only if computations are done on the synthetic spectra. Alternately it can be measured with a high SNR observed spectrum . As said above, the method of Connes was also revisited in Bouchy et al. (2001), where the authors also show the computation of quality factor for several star types using synthetic spectra and estimations of the RV uncertainty for some spectrographs.

A new usage of Connes approach for uncertainty estimation was presented in Boisse et al. (2010). While the original approach applies to the whole spectrum or part of it, Boisse et al. (2010) proposed to apply it to the cross-correlation function instead. In the case of quality factor Q computation on a noisy spectrum, the retrieved value for Q will be overestimated due to the high frequency structures and the uncertainty of RV will be underestimated. Authors investigated that the computation of quality factor is correct for spectra with relatively high SNR (higher than 200 per pixel). Such a high SNR may not be present for one individual spectral line, but will easily be achieved when the CCF is computed for all the spectral lines in one grating order, by piling up (with appropriate weights) all the

CCFs of all the lines of the order, as is common practice. We should mention that this approach for RV uncertainty estimation is adopted in several spectrograph pipelines, such as ESPRESSO at VLT and SOPHIE at OHP. It is specified in user manuals.

The Pierre Connes approach is quite different from the CCF approach, but none of them are working with telluric contaminated regions. To deal with such wavelength ranges Pierre Connes proposed to use a mask function, which is basically the product of wavenumber derivative of intensity and weight function, which can be used to just eliminates telluric contaminated regions. by putting a weighting function F(σ)=0 in such spectral regions.

There is a well known limitation to the use of the Connes approach, which is illustrated on Figure 2.13. We take a spectral line observed at Epoch 0 as a reference (black curve) ; on the red curve, shifted by 2 kms^{-1}, the decrease of the signal dA corresponds to the local slope and the wavelength shift dλ. However, for a larger shift of 15 kms^{-1} (quite possible with the $\pm$ 30 kms^{-1} Earth orbital velocity), the value of dA would be opposite. Therefore, the use of 1st formula of Connes 2.10 must be limited to small shifts only.

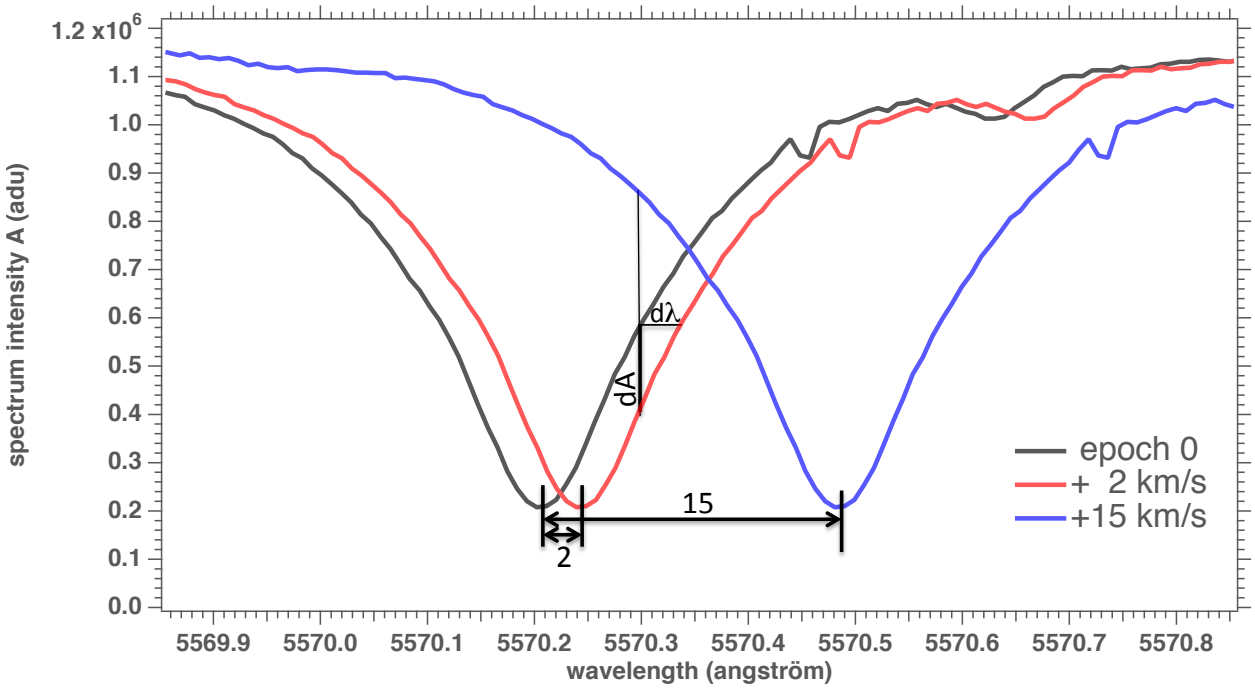


Figure 2.13: Example of a spectral line at 3 epochs. Epoch 0: reference (black), Epoch 1: shifted by +2 kms^{-1} (red), Epoch 2: shifted by +15 kms^{-1}. The intensity change is measured within the dλ slice. This figure is similar to the figure 4 from Connes (1985), and shows in addition that when the shift is too large (+15 kms^{-1}), the 1st formula of Connes (equation 2.4) does not hold any longer.

However, we figured out that this limitation may be circumvented by the use of a wavelength scale computed in the system attached to the barycenter of the solar system.

We describe in Section 2.3.3 a "shift finding" mathematical algorithm of any function y(x) , which is exactly the same as the approach of Connes. We then apply it to the total CCF of a whole spectrum (y), which is a function of the radial velocity (x) applied to the lines of the Binary Mask when building the CCF.

Application of the 1st and 2nd formulas of Connes to CCF

While Boisse et al. (2010) have used the 2nd formula of Connes to estimate the error on RV content of one particular CCF, in the frame of this PhD work, we have explored also the application of the 1st formula of Connes to CCF instead of a spectrum, in order to determine a radial velocity change between two epochs. To the best of our knowledge, this is a new approach, which makes sense,

because a CCF represents an average of many lines of a spectrum, and its use is robust to several artefacts, mitigating by averaging them out (for instance, the effect of optical interference fringes). We first establish the mathematics, then we perform a simulation to validate the concept, and apply it to a time series of exposures and their CCF.

Generalization: mathematical problem of determining the displacement (shift) Δx of a curve y(x).

Here we generalize the approach of Pierre Connes to any function y(x). Let us consider a function y_0(x), not known analytically, but known from its discretization at a certain number of sampling abscissa x(i). Let us consider that the curve is displaced (shifted) by an unknown amount Δx along the x axis, but we have also its discretization at sampling sites x(i) (figure 2.14). It becomes y_1(x):

$$y_1(x) = y_0(x - \Delta x) \tag{2.14}$$

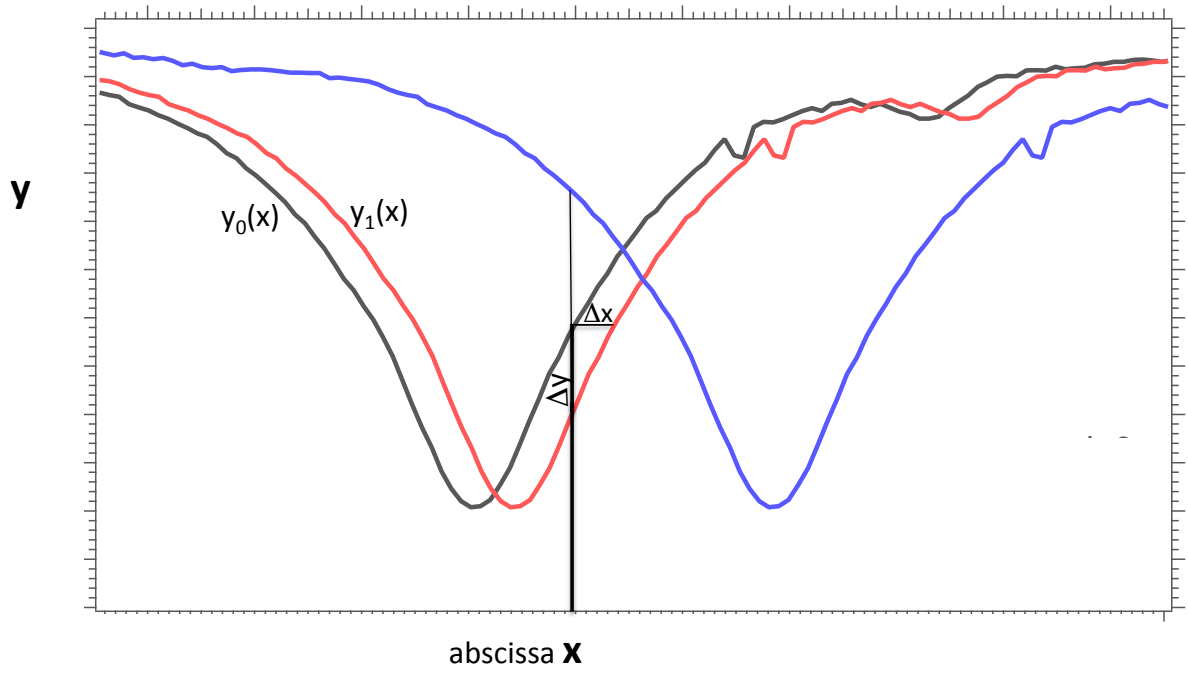


Figure 2.14: The black curve $y_0(x)$ is displaced to the position of the red curve. A small displacement (shift) Δx induces a change Δy directly linked to the local slope dy/dx of the curve. If the displacement of the curve is too large, there is no more link to the slope.

The mathematical problem that we address here is: having the discretization of both $y_0(x)$ and $y_1(x)$, what is the algorithm which allows to find the most likely estimate of Δx, and with which uncertainty? (a shift finding algorithm).

While the curve is displaced by Δx along the x axis, the ordinate at sampling site x(i) changes by $\Delta y = y_1(x(i)) - y_0(x(i))$.

We relate Δx, Δy and the derivative y'_0 by:

$$\Delta x = -\frac{\Delta y}{y'_0} \tag{2.15}$$

In the following we note in short $\Delta x(i)$ for $\Delta x(xi)$, $y_0(i)$ for $y_0(xi)$, $y_1(i)$ for $y_1(xi)$, $y'_0(i)$ for $y'_0(xi)$, i is

the index of sampling abscissae x(i). Equation 2.15 becomes:

$$\Delta x(i) = -\frac{\Delta y(i)}{y_0'(i)} = \frac{y_0(i) - y_1(i)}{y_0'(i)} \tag{2.16}$$

Therefore, for each sampling point x(i), we have an estimate of the displacement Δx.

Now, we estimate the error (or uncertainty) σ(i) on this estimate Δx(i) associated to the errors on y_0(i) and y_1(i), noted respectively $\sigma(y_0(i))$ and $\sigma(y_1(i))$. As did Pierre Connes, we assume that there is no error on the derivative y_0'(i).

The variance of a random variable X is Var(X)= σ^2(X). We use the formula of error propagation, or the formula of variances of two independent random variables X and Y, with constant factors a and b, applied to 2.16:

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) \tag{2.17}$$

Here we use:

a=$\frac{1}{y_0'(i)}$, b= -$\frac{1}{y_0'(i)}$, X=y_0(i), Y=y_1(i)

It comes:

$$Var(\Delta x(i)) = \frac{1}{y_0'^2}(Var(y_0(i)) + Var(y_1(i))) \tag{2.18}$$

$$\sigma^2(\Delta x(i)) = Var(\Delta x(i)) = \frac{1}{y_0'^2}(\sigma^2(y_0(i)) + \sigma^2(y_1(i))) \tag{2.19}$$

Now we consider two extreme cases:

1. The curve y_0(x) has no noise, and only y_1(x) has noise.

2. Both curves y_0(x) and y_1(x) have an equal noise.

In the first case, the uncertainty $\sigma_i \equiv \sigma(\Delta x_i)$ becomes:

$$\sigma_i \equiv \sigma(\Delta x(i)) = \frac{\sigma(y_1(i))}{(abs(y_0'))} \tag{2.20}$$

In the second case, the uncertainty from equation 2.20 must be multiplied by $\sqrt{2}$.

Therefore, for each sampling point of the curves x(i), we have an estimate Δ x(i) of the displacement between the two curves (equation 2.16), and an estimate of the error (or uncertainty) on this estimate given by equation 2.20.

We know, from the theory of Gaussian errors, that the optimal estimate Δ x_{opt} of the true, but unknown displacement Δx, is obtained by a weighted average of all determinations Δx(i), where the weights

are $1/\sigma_i^2$:

$$\Delta x_{opt} = \frac{\sum_i \frac{1}{\sigma_i^2} \Delta x_i}{\sum_i \frac{1}{\sigma_i^2}} \tag{2.21}$$

and the uncertainty (error) σ_{opt} on this estimate is given by the equation 2.22:

$$\frac{1}{\sigma_{opt}^2} = \sum_i \frac{1}{\sigma_i^2} \tag{2.22}$$

In the case of Poisson statistics where y(x) is a number of discrete events, like electrons created by photons in a pixel of a CCD, we know that the variance of y is just y, a feature used by Pierre Connes in his formulation of the shift of a spectrum.

***A posteriori* check of shape constancy hypothesis.**

This shift finding algorithm relies on the strict hypothesis that the shape and amplitude of $y_0(x)$ and $y_1(x)$ are rigorously identical, the two functions differ only by a single shift. In real life, a change of amplitude might happen in addition to a shift. This is why the two functions must be normalized to each other (the same integral), before applying the algorithm, as was recognized by Connes.

It is possible to check *a posteriori* if the hypothesis of shape constancy (after normalization) is fulfilled, or not. Indeed, one estimate $\Delta x(i)$ of the shift is provided on each of the discretization point x_i, with its error bar σ_i. If all the values $\Delta x(i)$ are equal, within their error bar, then the shape constancy hypothesis is not contradicted. On the contrary, if the values $\Delta x(i)$ are departing from a constant (within the error bars), it is a proof that the shape of the curve is not constant. Therefore, the algorithm provides a powerful tool to study potential changes of shapes, which may be interpreted in terms of physical phenomenons, like changes due to convective granulation blue shifts. In this respect, it should be remarked that the higher is the signal y, the smaller is the error bar and the easier to detect a departure from constancy. This is why in the following we apply the shift finding algorithm to CCF, which are a composite result of many spectral lines, increasing the SNR on each single point.

Application to the CCF of a spectrum.

We now develop the application of the above shift finding algorithm to the case of the total CCF of a spectrum, CCF_{tot}. This CCF_{tot} is computed on a fixed grid of radial velocities RV of the observed star, which is the x axis of the previous section, while the values of CCF_{tot} are the y(x) values: $CCF_{tot}(RV)$

is the curve that is shifted due to the existence of exoplanets around the star.

The CCF_{tot} is just the sum of all CCF per echelle order of ESPRESSO spectrum, and one CCF for one order is obtained by piling-up all CCF of all lines of the Binary Mask which are contained in that order.

Since the wavelength assignment for each pixel of the spectrum is computed in the system Barycentric Solar System, the CCF are also computed in this system. The reflex motions induced by planets on the star are at most a few hundred ms^{-1}, a shift much smaller than the typical width of spectral lines. The condition (a small shift) required for the applicability of the shift finding algorithm is therefore fulfilled, and equation 2.21 can be applied to curve $CCF_{tot}(RV)$ to find a planetary-induced shift.

Simulation

At first we decided to test this approach on simulated data. We have used the series of 200 exposures taken on star HD 40307 (Prog.ID:0102.D-0346(A); PI: Bouchy) taken during the Christmas night from 24 to 25 December, 2018.

- We have averaged all 200 spectra together (to get a high SNR spectrum), spectel to spectel, for order 104 (spectrum S1), and computed the CCF1 on a given grid Vrad of radial velocities (2.15). The wavelength scale WS1 was the spectrometer wavelength scale for exposure 100. Order 104 was selected because this order is not affected by tellurics, and the signal is rather high with a single exposure (covering wavelength range 5523.8 – 5606.7 Å).

- Then, a synthetic spectrum S2 was derived from the stacked spectrum S1 just by inflating the wavelength scale, as if the star had changed its velocity by +100 ms^{-1}: same intensities but on an inflated scale WS2.

- In order to mimic what would be the values of spectrum S2 sampled on the grid WS1, we interpolate the spectrum S2 vs WS2 on all the points of the grid WS1.

- The CCF2 of the spectrum S2 sampled on WS1 is done on the same grid Vrad of radial velocities, yielding the green CCF curve of Figure 2.15.

Also displayed on Figure 2.15 (in red) is the weight attached to each point of the Vrad grid, according to the photon noise limit of Pierre Connes first formula, here applied to the CCF1. The double-peak

shape of the weight curve (in red) just shows that the relative contribution of each point of the CCF to determine a shift, which is depending on the square of the derivative (w.r.t. the Vrad grid).

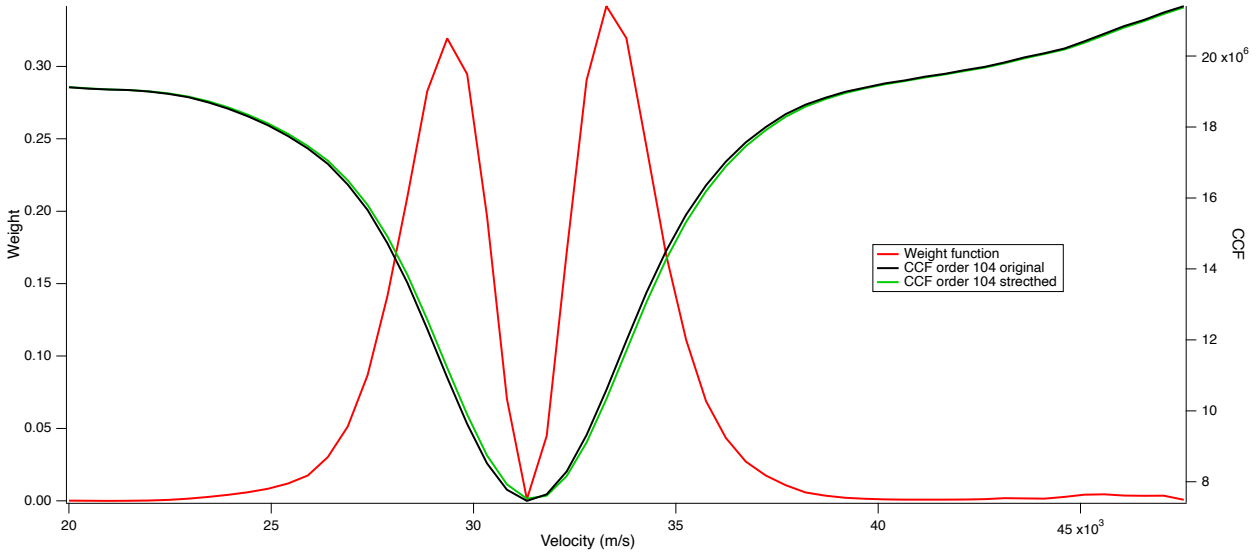


Figure 2.15: Example with a simulated Doppler shift. Black – CCF1 of original data (right scale); green – CCF2 of simulated shifted from original data (left scale); red – weight function (left scale).

The CCF were computed on 2 grids, one with a number of points of 57 (start point 20000 ms^{-1}, step 492 ms^{-1}) and other one with 81 points (start point 13100 ms^{-1}, step 500 ms^{-1}), the one with 81 points is the grid which is produced by ESPRESSO pipeline in automated mode. When the first formula of Connes is applied to the CCF1 and CCF2, a radial velocity of 99.613 ms^{-1} (in case of V_{grid} 81 points – 99.9703 ms^{-1}) is returned, while the second formula of Connes giving the uncertainty applied to CCF1 (or CCF2) is 0.36628 ms^{-1} (in case of V_{grid} 81 points - 0.367348 ms^{-1}). Therefore, this simulation exercise validates this new algorithm, and the uncertainty associated to it.

After the confirmation that this approach is able to work on CCF we decided to test it on real data, and compare the results with the classical approach: the star RV included in the header of each ESPRESSO spectrum, derived from a Gaussian fit to the CCF (at order level, then merging all orders with appropriate weights). Two criteria of quality of the new approach will be, for a star with no planets, the constancy of a time series, and the dispersion of RV around the mean value.

Our new approach was tested on datasets of two stars : HD 40307 (described in subsection 1.4.2) and HD 67458 (described in subsection 3.1)

In the case of HD 40307 where is a planetary system, the changes of RV includes the reflex motion of the star; the RV signal is changing with time, and furthermore a comparison of dispersion (same as standard deviation) is not a good criterion for efficiency evaluation. In such a case, the comparison

criteria between the two methods may be the χ^2 characterizing the difference between data and a model of RV with 3 planets, or, for shorter time slots (some hours), the χ^2 from a linear model (linear change of RV with time). In the case of HD 67458 where is no planet at all, the constancy and the dispersion are good criteria.

When we are using the PC method we have to pay attention to the quality of CCF. The narrower CCF shape is, the higher precision of velocity is. For each order, the general shape of CCF will be the same from one exposure to the next, but the intensity, and its noise will vary. It depends on the quality of original data and on the number of lines in mask. Even if the SNR of spectrum is high, but in CCF computation there are only 3 lines to be used for the whole order – the quality of CCF will be worse, than if the SNR of spectrum is lower, but for CCF computation 30 lines are used. At SNR spectrum constant, the SNR of the CCF will scales approximately as the square root of number of lines.

2.3.4 Line-by-line method

A relatively new approach which is based on the Connes formula is the line-by-line method, as first described in Dumusque (2018) with further development in Cretignier et al. (2020); Artigau et al. (2022). The concept of line-by-line method is actually the same method as described in Bouchy et al. (2001) but the analysis is performed in small fractions of spectrum, so-called "lines". After the velocity change for each "line" is computed it is necessary to check for consistency of velocities of lines and average them with uncertainty propagation to obtain one velocity for one spectrum. Such an approach allows to find and track stellar activity. When operating two loops in succession it is possible to exclude from consideration lines which are polluted or show high activity. The main benefit of such an approach is that if one divide the spectrum into N segments with equal ΔRV content, the ΔRV accuracy of each domain will be reduced by factor $\sqrt{N}$ w.r.t the ΔRV combined for all domains. At the same time the velocity error caused by one single glitch on one single segment will increase by a factor of N w.r.t the error caused by this single glitch on the full spectrum, because of average over N segments. Therefore it will be much easier to find affected segments and exclude them from further consideration.

Chapter 3

Study of ESPRESSO series of short time exposures

We used ESPRESSO series of short time exposures to perform several investigations. Their aim was to test various changes in the construction and/or the analysis of the CCFs, their effects and their potential ability to improve the precision of RV measurements. In short, the studies can be divided into the following groups:

- Comparison of ways to merge order-by-order CCFs
- Comparison of ways to compute the CCFs for individual orders
- Comparison of RV results based on CCF Gaussian fits and on the application to the CCFs of the first Formula of Connes
- Application of the first Formula of Connes to test differences between detectors
- Application of CCF technique to detect stellar lines temporal distortions

The first three studies made use of HD40307 data. The study of the application of the first Formula of Connes to tests of the differences between detectors was performed for a dataset of the star HD 67458 (described below in 3.1). Similarly, the search for the effect of stellar lines temporal distortions was performed for the HD 67458 dataset and is explained in 3.5.3.

3.1 Data

Most studies were performed for the dataset of HD 40307, this dataset has been described in Section 1.4.2. HD 67458 is a G0V type star with visual magnitude 6.79. It was observed during 3 full nights from 17th to 21st of January, 2021 by ESPRESSO. Data were obtained in HR mode with resolution between 100 000 and 160 000. During this January 2021 run HD 67458 was a part of a coordinated spectroscopic and photometric campaign on signals of stellar granulation [1]: 587 measurements were obtained, the exposure time was set to 60 seconds. For this star there is no evidence for the existence of a planetary system (there are no references in the literature indicating the presence of exoplanets). Therefore, the radial velocity of this star (in the Solar system barycentric frame) should have no modulations caused by planets. In SIMBAD the radial velocity of HD 67458 is indicated at -15.672 kms^{-1}. Such kind of star is an appropriate target for studies of the physical phenomena that influence the observed Doppler shift. In particular, Sulis et al. (2023) studied the stochastic signal due to the granulation for this star.

3.2 Tests on Gaussian fits of CCFs

3.2.1 Comparison of ways to merge pipeline CCFs provided for each order

If CCFs have been Gaussian fitted separately for each order of a single exposure, the question arises of their merging to obtain unique RV. The ESPRESSO pipeline provides CCFs for each order which are ready to simply sum, i.e. their respective weights have been pre-computed along with the CCFs themselves (we believe the weights are based on depths and widths of all lines of the BM in the order). In the following study, we have investigated other ways to combine RVs from all orders, using the HD 40307 time series. More precisely, we performed Gaussian fits to CCFs obtained for individual orders, both pipeline CCFs and CCFs computed by us (next section), and combined them in two different ways.

The first, classical way is to use the uncertainties σ(ord, exp) associated with each individual CCF fit as the sources of the relative weights W, i.e. W is proportional to $1/(\sigma(\text{ord, exp}))^2$. We also tested a second way, here permitted by the existence of the time series and subsequent availability of statistics based on the series. For each order, we used a long series of fitted RVs obtained from consecutive

[1] Prog.ID:0106.D-0445(A), 0106.D-0445(B),0106.D-0445(C); PI: Sulis

exposures, performed a linear fit to the series and derived the RV dispersion associated with the order. These dispersions were then used as σ(ord). The tests were made based on the first 100 exposures of each night, and the measured σ(ord) were subsequently applied during the order merging for all exposures of the night. The reason for such a test is the following: thanks to the statistics on 100 exposures we obtain an empirical, data-based evaluation of the uncertainty associated to each order. This estimate is more precise than individual uncertainties associated to a single exposure, largely governed by the noise. However, doing so we implicitly assume that the spectral shapes of the lines do not change during the 100 exposures and during the night, and that subsequently the distribution of order weights does not change, which may not be the case (as we will see in next sections).

We used the ESPRESSO CCFs provided for each order and performed a Gaussian fit over the entirety of the official velocity grid of the pipeline (81 points). During this computation we stored the RVs and the uncertainties σ(ord, exp) associated with each fit, i.e. for each individual order and for each of all exposures.

Classical on-the-fly method:

We used all the uncertainties σ(ord,exp) derived from the fits, i.e. for each order and each exposure, and for each exposure we combined the order-based RVs according to weights inversely proportional to the square of σ(ord, exp).

Time-series based method:

For each order, we performed a linear fit as a function of exposure number over the series of RVs fitted for this order for the 100 first exposures of a night and extracted the dispersion S_{dev}(order) around the fitted linear relationship (from the χ^2 returned by the fit). We used a linear fit and not simply the global dispersion w.r.t. the mean to account for a radial velocity shift during that time period. As explained above, this quantity served as a new estimator of the uncertainty associated with RV for this order for the entire night, i.e. we no longer used all individual σ(ord, exp). As a consequence, the individual weights for order merging are independent of the exposure (or time) for this night. The first night has only 10 exposures and could not be used this way. The distribution of the weights is very similar for the 5 nights. A combination of RVs for all orders was then done using these new weights, i.e.replacing σ(ord, exp) by S_{dev}(order), a quantity no longer depending on the exposure.

The difference between the two types of weights is illustrated in Figures 3.1 and 3.2 . In the figures are superimposed the weights attributed to each order for all exposures of the Dec 24th night, deduced from the observed dispersion (second method), and the on-the-fly weights derived from the individual fit for one single selected exposure (here the last exposure of the night). The sum of the weights is one in both cases. Surprisingly, significant differences do appear, which means that the existence of large uncertainties on the individual order fits for one particular exposure does not necessarily correspond to a large dispersion of the RVs of an order in the second method. We will discuss this point below.

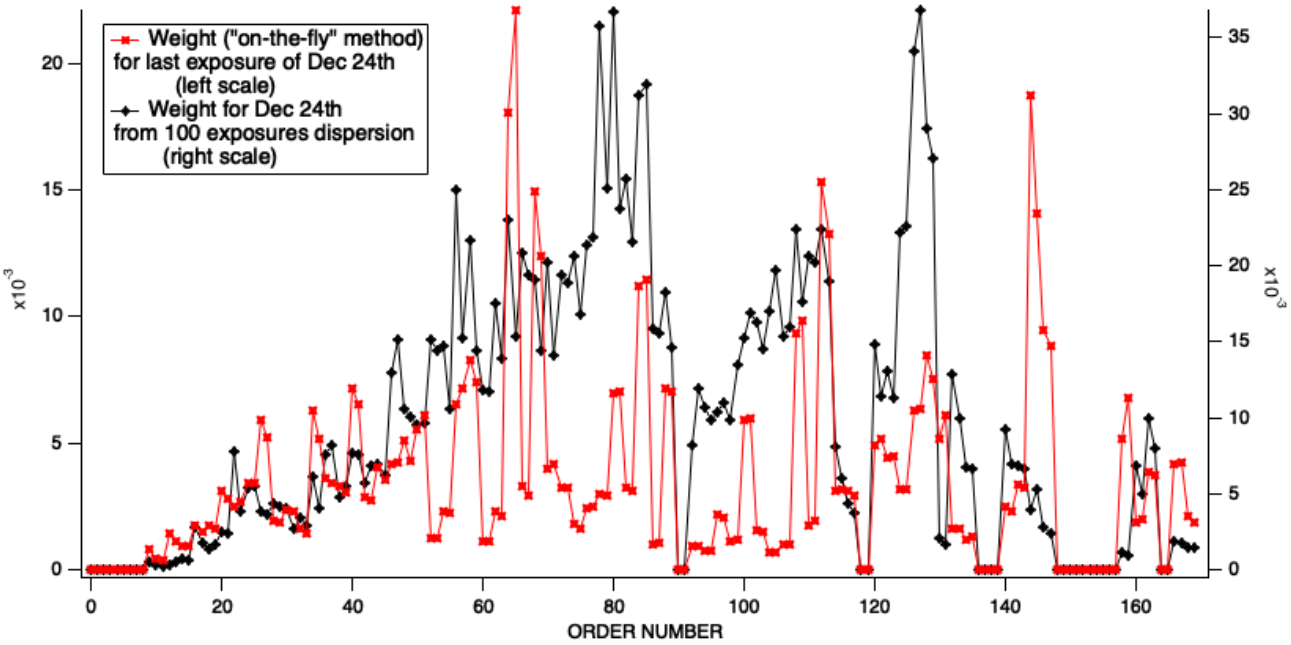


Figure 3.1: Weights for each order obtained with the two methods, as a function of order number. Orders for which binary mask lines are absent have a null weight. Weights based on uncertainties on the CCF Gaussian fits are shown in red for the last exposure of the Dec 24th night. Weights deduced from the dispersion of the RVs around a linear fit to the first 100 exposures of 24 Dec night are shown in black. Significant differences can be seen.

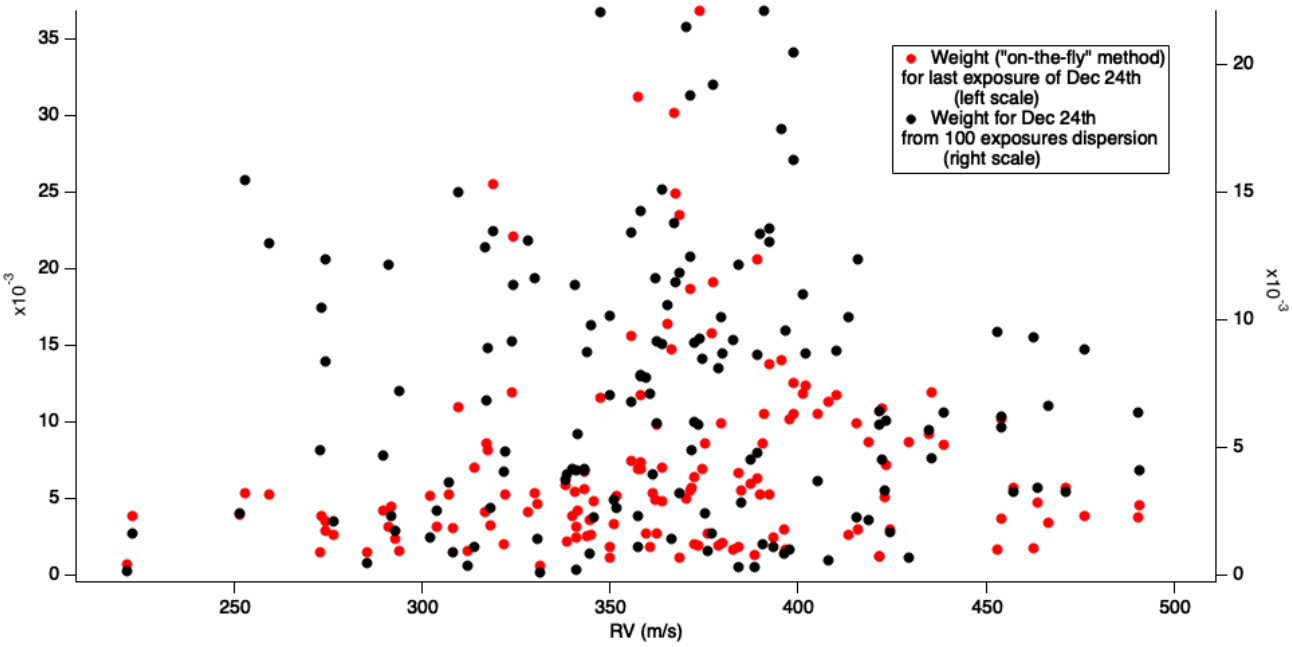


Figure 3.2: Weights for each order obtained with the two methods, as a function of RV computed for the order (a constant 31 kms^{-1} has been subtracted from RV). Orders for which binary mask lines are absent have a null weight. Weights based on uncertainties on the CCF Gaussian fits are shown in red for the last exposure of the Dec 24th night. Weights deduced from the dispersion of the RVs around a linear fit to the first 100 exposures of 24 Dec night are shown in black. Note the large and small values of RVs for some orders. Each pair of one black and one red point for the same order have the same RV, but their weights may be quite different, changing the overall resulting RV after order merging. See for instance the pair at 252 ms^{-1}.

The two different resulting RV series are shown superimposed in Figure 3.3. As can be seen, the dispersion on RVs **within a night** is significantly reduced for the second method. This is true for all nights, and may appear as a strong advantage; however, simultaneously all RVs are shifted down by about 3 ms^{-1}. The global shift comes from the fact that every change in the weights of individual orders has a very strong effect. Figure 3.2 illustrates this point. The figure displays the series of δ(RV) values, one for each order, to be summed during order combination to obtain RV for a typical exposure. The individual values are extremely different from one order to the order, and reflect the strong offsets between model and measured stellar line centers. This explains why changing the distribution of weights may drastically change the average value of RV. We will come back to this point in the next section.

The influence of the weights is so strong that we may detect differences in average RV values if, instead of computing and using the first 100 exposures of each night, we use for all nights the same distribution of weights, namely the one found for Dec 24th. This is illustrated on Figure 3.4. The figure shows that RV for the Dec 25 night is slightly different depending on the choice of weights.

This casts shadows on the improvements in precision obtained from second method. The precision is indeed increased, but may be at the cost of a variable bias. It suggests that different, more adapted BMs must be used, with much smaller offsets of line centers, which can be produced from high SNR spectra of the star. This is discussed in the perspective section of this work.

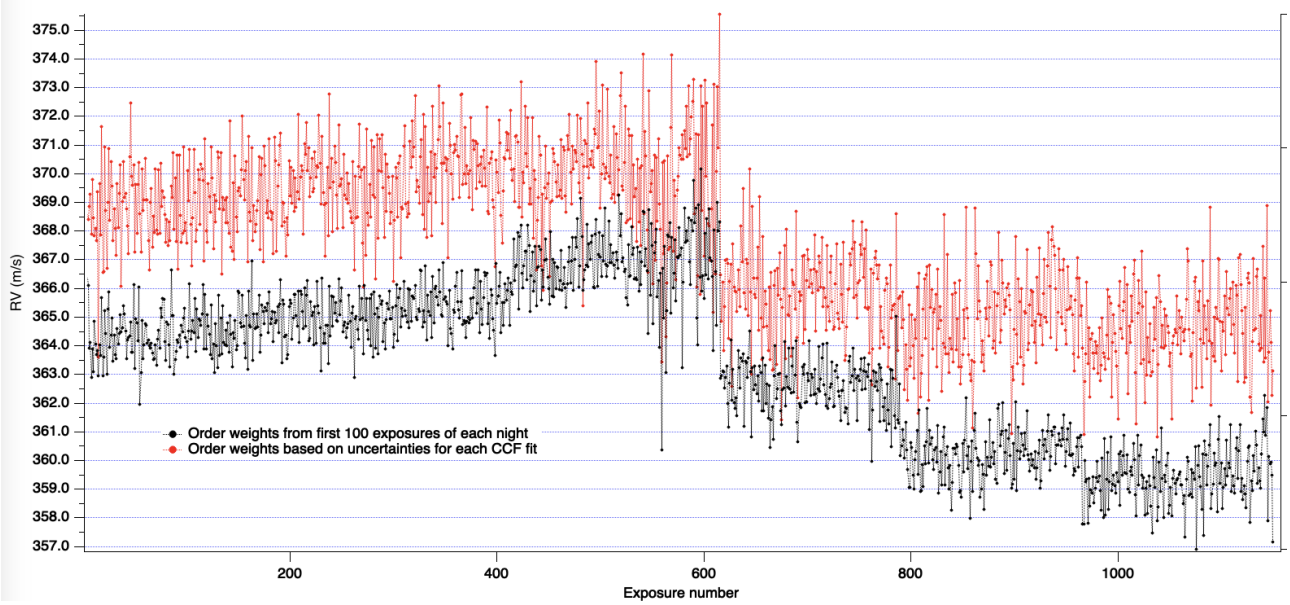


Figure 3.3: RVs obtained using order weights based on the dispersion among the 100 first exposures of each night (black line) and RVs computed classically using order weights from each CCF fit (red lines). Note the smaller dispersion of RVs for the first method, and the differences of average levels between the two determinations.

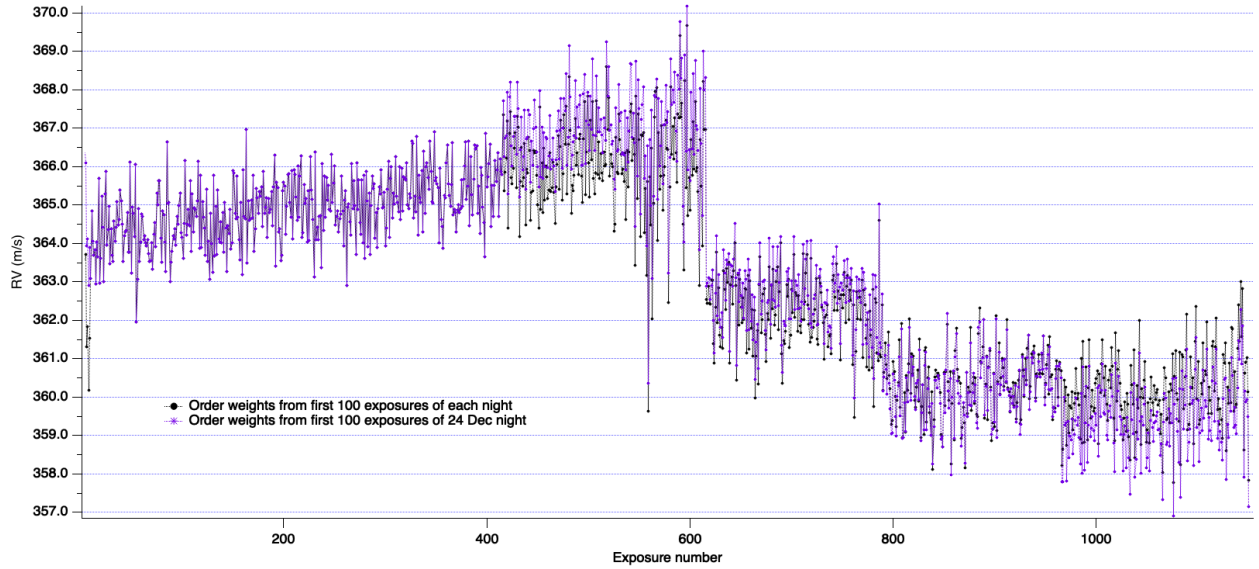


Figure 3.4: RVs obtained using order weights based on the dispersion among the 100 first exposures of each night (violet line, same as black line of Figure 3.3), and RVs obtained using for all nights the same, unique distribution of weights, namely those computed from the first 100 exposures of the Dec 24th night (black line). Note the differences between the average RVs from night to night according to the chosen weights.

3.2.2 Comparison of ways to merge CCFs built by us

Building CCFs

During this study we have also computed our own CCFs in a classical way as well. First, we selected only the lines of the BM with a contrast (relative depth) >0.2. A RV velocity grid with 57 points and a grid step of 492 ms^{-1} (equivalent to about one pixel in the observed spectrum). For each step of the RV grid, a boxcar integration was performed over the observed stellar line. The width of the boxcar is a parameter. When it is equal to one pixel of the observed spectrum, the resulting CCF looks very much alike the original spectrum for obvious reasons. We call this the J1 case, for which all electrons of the spectrum are used, and they are used only once. With a boxcar of 3 pixels width, case J3, each electron is used 3 times, and one must be careful about the error attached to each point of the CCF. Taking more than one pixel smooths the spectrum and decreases the noise on each point of the CCF. Then, for each order separately, all the CCF for each line of the BM are piled up together, without assigning any weight to the individual lines. This is not like the Lafarga et al. (2020) scheme, where a weight=contrast/width is selected. As a result, the signal in our CCFs are much larger than if a weight is assigned, as is the case of the CCF of the official pipeline. During the Gaussian fit of each CCF per order, the number of points over which the fit is performed is also a parameter. González Hernández et al. (2020) have shown that, depending if the fit is done around the core only, or over the whole line, the returned best fit RV is different.

Comparison of ways to merge our CCFs order by order

We made Gaussian fits of all our computed CCFs and combined the results obtained for each order in the same two ways as for the pipeline CCFs. The difference between the two methods is illustrated on Figure 3.5 for one of our CCF computation mode, namely with boxcar of type J1 and Gaussian fit over 39 points of the CCF. We can see that the RV dispersion using uncertainties from the fits are quite large for this particular CCF computation, significantly above the one obtained for the pipeline CCFs (the pipeline values can be seen in Figure as displayed on Figure 3.7). On the other hand, the dispersion is considerably reduced for the second weighting method, with weights estimated form the dispersion in the series of results for each order, permitted by the existence of long time series, and is now reaching the level of the pipeline results.

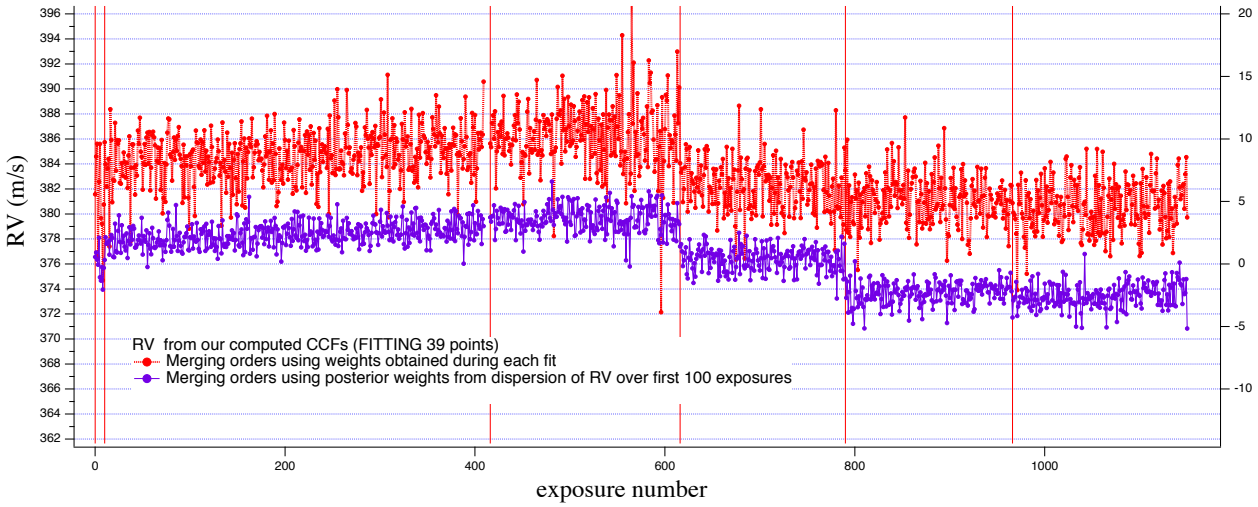


Figure 3.5: RVs obtained from our computed CCFs using two different types of order weighting. Red points show the results obtained using uncertainties for each CCF fit. Violet points use weights deduced from the dispersion around a linear fit of the first 100 measurements, obtained during a preliminary computation with the first method. The boxcar of type J1 and Gaussian fit made over 39 points of the CCF

Influence of boxcar size and fit window, optimization

Based on the above results we searched for the optimal way to compute our CCFs. Namely, we used several boxcar widths and CCF Gaussian fit windows, and we used the same weights for order combination as those obtained above based on the individual dispersion among each order for the 100 exposures of the Dec 24 night. For each case we performed a linear fit to the 100 exposures of the Dec, 25th night and used the dispersion (measured by χ^2) as a diagnostic of the precision of the RV measurement. The results are displayed on Figure 3.6. We also performed several determinations using the pipeline CCFs and varying the Gaussian fit window width. As can be seen the optimal solution for our own CCF computations is found for a boxcar of 11 points and a fit window of 17 points total, i.e. 8 points on each side around the center.

The results also show that using the RV dispersion per order as weights for the order combination allows to reach an even lower uncertainty compared to the pipeline CCF Gaussian fitting (providing the smallest dispersion), as shown on Figure 3.7. The gain in RV precision is visible by eye for all nights.

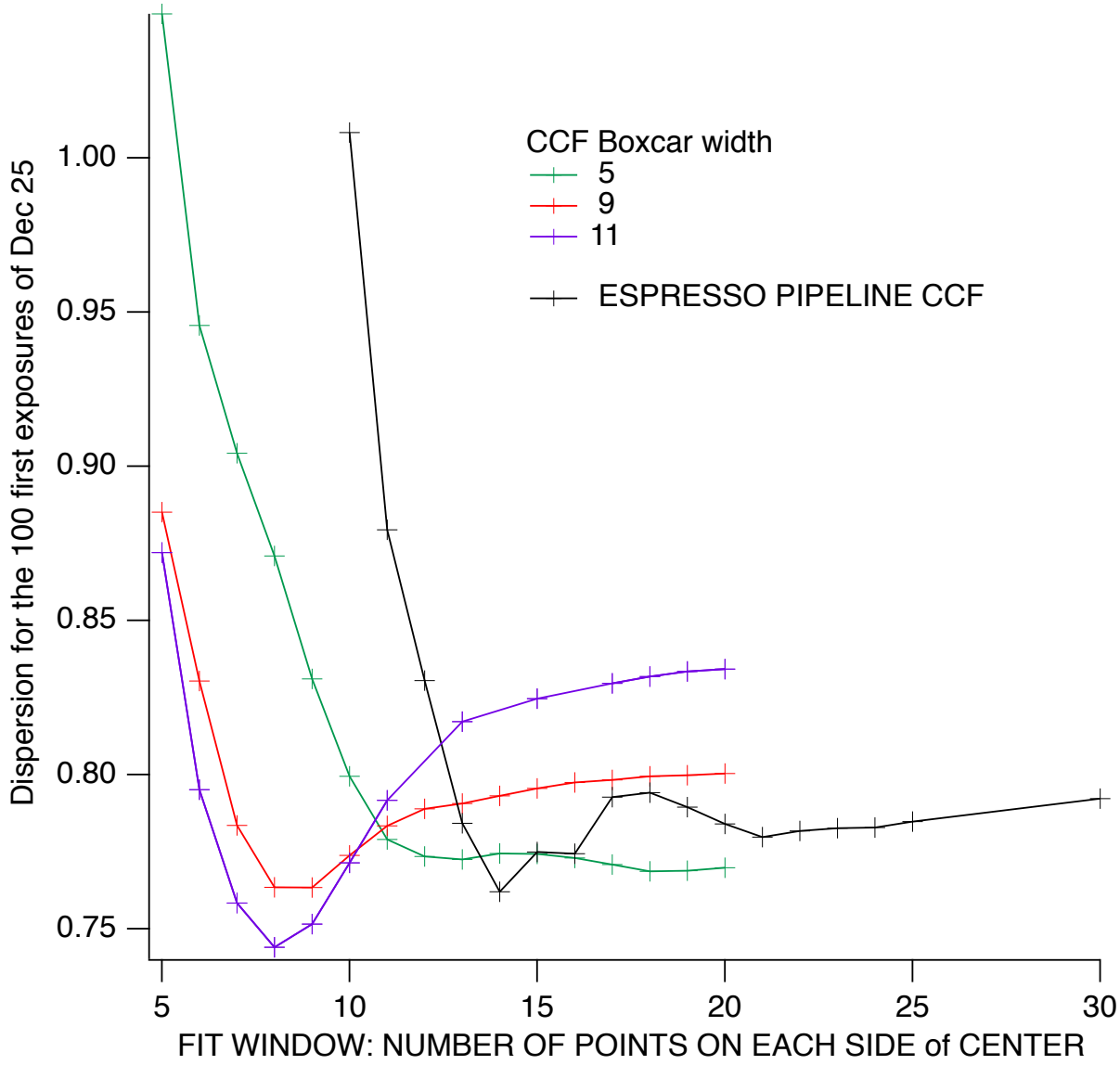


Figure 3.6: Dispersion on RVs around a linear fit for the first 100 exposures of Dec, 25th. Shown are the results obtained from the Gaussian fits to our computed CCFs and to the ESPRESSO pipeline CCFs as a function of the fit window, here expressed by the number of points on each side of the CCF center. Our CCFs are computed with three different boxcar widths (see text)

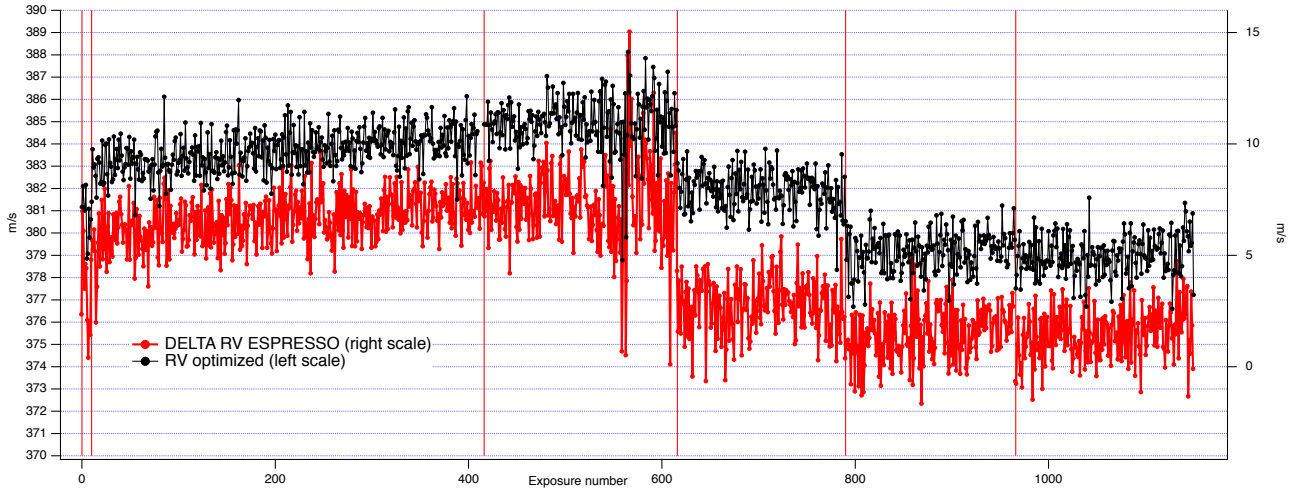


Figure 3.7: RVs obtained from our CCFs using the optimized window and CCF boxcar width (black points). The RV time series from the ESPRESSO pipeline (after subtraction of 31 kms^{-1}) is shown for comparison (red points, right scale). The left and right scales are adjusted to correspond to the same height per ms^{-1}).

3.3 Comparison of Gaussian fits and application of the First Formula of Connes

Initially P. Connes devised a computation of Doppler shift between two spectra, a computation we have named the first Pierre Connes formula (or CF1), and he demonstrated that this is the optimal way to measure a shift, in other terms it takes maximum information from data. There is a condition for the applicability to two spectra: the Doppler shift must be very small with respect to line widths. For measurements taken at different epochs and for large Doppler shifts, it is no longer valid. However, new generation, high resolution and highly stable spectrographs as ESPRESSO and, also, the possibility to compute the Earth velocity vector in the solar system barycentric frame with a precision on the order of a few cms^{-1} allows to assign to each pixel of the measured spectrum a wavelength scale in the barycentric frame (the intensity of the pixel is not modified).

Because Doppler shifts due to planets between spectra that have been computed back in the barycentric frame are all very small, the application of the CF1 remains an option. The CF1 method is applicable, not on a pixel basis as was imagined by Pierre Connes, but on the wavelength scale in the barycentric system. In order to minimize the effect of fringes and detector pixel-to-pixel response non uniformity (PRNU), we have investigated the applicability of the CF1, not to the spectra as was used by Pierre Connes, but instead to the CCFs computed order by order, and also to their sum. It is expected that it

should provide a better information than a Gaussian fit.

3.3.1 Application of the CF to CCFs of individual orders: "on-the-fly" derivatives and order merging

The CF1 method is measuring the Doppler shift between two CCFs, called CCF1 and CCF2. CCF1, taken as a reference, comes from one particular exposure (say, the first one of a series). CCF2 is then one of the CCF of a time series of exposures. The application of the CF1 method will result in an estimate of dRV for each point of the velocity grid for which is computed the CCF. Then a weight is assigned to each point, with the so-called Mask, containing the derivative of the CCF (vs RV grid scale), and all dRV are combined with the Mask of weight. A global error for this order is derived from the second formula of Connes, containing the quality factor Q.

The results of computation is shown on Figure 3.8. Dispersion for each night reduced with the CF method, results are shown in Table 3.1.

Night	S_{dev} pipeline	S_{dev} CF1
24	1.08954	1.08857
25	1.86163	1.5203
26	1.25311	1.04921
27	1.18426	0.910429
28	1.13666	1.04968

Table 3.1: Dispersion of each night (without night 22)

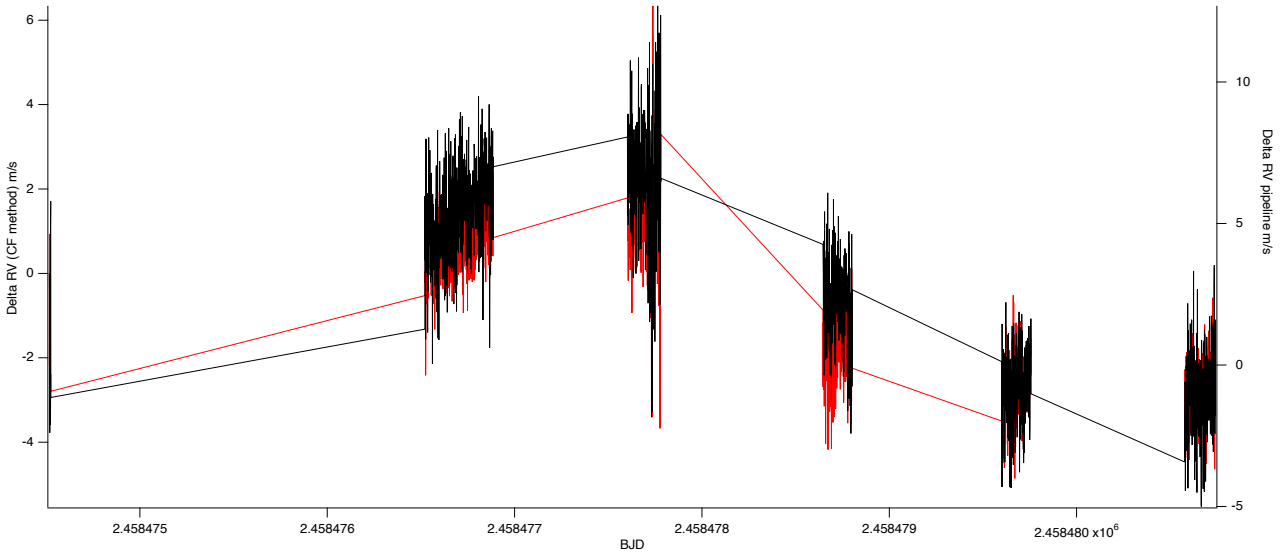


Figure 3.8: RVs obtained from ESPRESSO CCFs for individual orders. The RV time series based on CF1 "on-the-fly" is shown in black and compared with the pipeline series (in red). The pipeline is obtained with a Gaussian fit. The left and right scales are adjusted in such a way the full amplitudes are identical. The dispersion is significantly reduced for CF1 results (see also table 3.1).

3.3.2 Application of the CF to CCFs of individual orders: using pre-computed averaged parameters

In 3.3.1 we used an "on-fly-method" of computation of CCF derivatives and of weights used for order merging. We also explored another way to use the CF1 method. Considering that the derivative of the CCF is quite sensitive to the noise, we have computed the derivative, called the Mask, and the quality factor Q based on the average of the CCFs of 50 exposures. The average is much less noisy, and is used for all exposures. Of course we keep the values of CCF2 to compute the dRV, and the sum of electrons to compute an estimate of the error (on RV) per order.

Then, when combining the orders together, the order merging was based on the dispersion of returned RV around a linear fit to the 100 first exposures of each night (a two-step technique similar to what we did for Gaussian fits). This provided the best (i.e. minimum) dispersion of RVs. However, the comparison with results with "on-the-fly" mask, Qccf and fit-based order weights again shows some offset discrepancies, similar to what we found for Gaussian fits. Figure 3.9 shows the results of this method applied to ESPRESSO CCFs and compares with the pipeline. It can be seen how the dispersion with each night is very significantly reduced. E.g., for the 406 exposures of Dec 24, the dispersion around a linear fit to the series is 0.63 ms^{-1} vs 1.0 ms^{-1} for the pipeline. It also reveals the offset discrepancies.

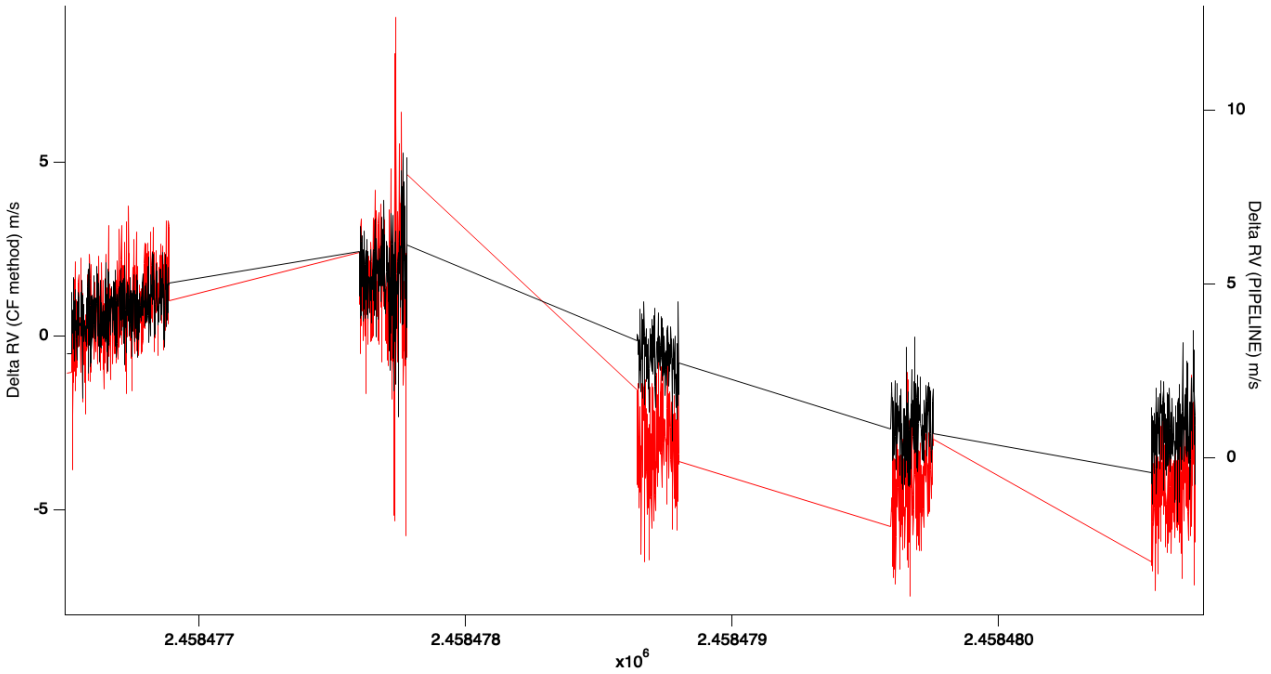


Figure 3.9: RVs obtained from ESPRESSO CCFs for individual orders. The RV time series based on CF1 with the Mask is shown in black and compared with the pipeline series (in red). The pipeline is obtained with a Gaussian fit. The left and right scales are adjusted in such a way the full amplitudes are identical. The dispersion is significantly reduced for CF1 results.

3.3.3 Application of CF to order-merged pipeline CCFs and comparison with the Gaussian fit

In CCF fits files taken from ESO archive the CCF matrix has 171 columns, while only 170 orders exists. The last column of CCF matrix is a order-merged CCF produced by the pipeline itself. We have verified that this CCF is computed by the summing up all CCF per order for the exposure.

We applied the CF technique to the order-merged CCFs of the pipeline (column 170, since numeration starts from 0 up to 170). All computations used "on-the-fly" values for mask and and quality factor. The CCF for the first exposure of Dec 24th was chosen as the unique reference for all nights. On Figure 3.10 we compare the CF results with the results from Gaussian fits to the order-merged CCFs (similar to RVs from the pipeline). Although not easily visible by eye, there is a gain (i.e. a reduction) in RV dispersion in the case of the CF method. A careful look at all results shows that in clean regions the dispersion is clearly reduced, while, on the contrary, in regions characterized by strong variability (like the second half of the Dec 25th night characterized by blue detector instabilities) the dispersion is significantly augmented, showing that the CF method is highly sensitive.

In order to quantify the gain we made a linear fit to the 406 values of RVs of night Dec 24th for both

types of results and removed the trend (essentially due to planets) to keep only the dispersion around the average value. We made two histograms of the RV values with the same bins and compared them. The result is shown on Figure 3.11. The figure clearly shows the gain in precision, confirmed by the half-widths at 1/e of Gaussian fits to the two histograms. The half-width = $\sigma\sqrt{2}$ is reduced from σ=1.46 to 1.17 ms^{-1} and the error σ is reduced by 20% from 1.03 to 0.83 ms^{-1} for this series of exposures taken during night 24 december 2018.

The result is not very surprising, since P. Connes demonstrated that his first formula is the optimal way to measure Doppler shifts. On the other hand, as we will discuss below, the hypothesis underlying the CF formulation is an absence of variability of the shape of the spectra (or here CCFs) . Only pure Doppler shifts are allowed. As we will see below, this is not the case here, due to stellar variability. Figure 3.10 shows that, along with dispersion changes, the average RV value does not vary exactly in the same way from night-to-night for the two methods. We suspect that this behavior is associated with stellar line variable shapes.

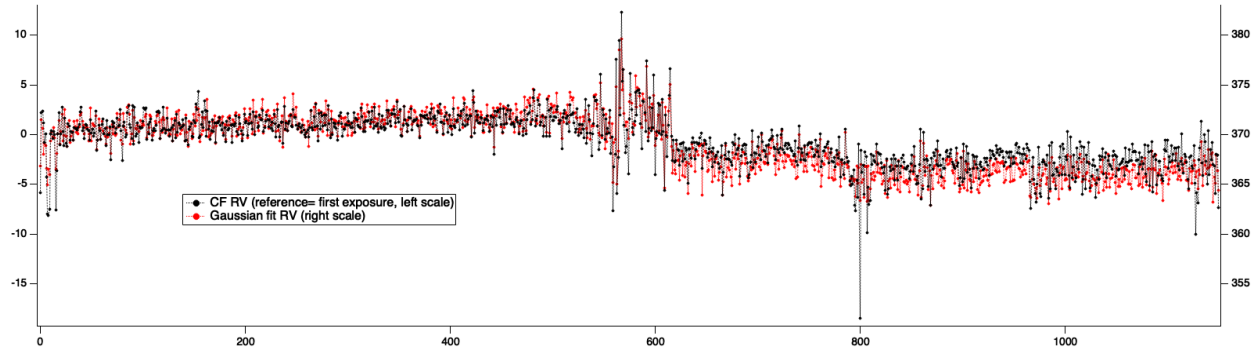


Figure 3.10: Comparison between CF and Gaussian fit methods applied to the order-merged pipeline CCFs. Concatenated RVs derived from the two methods are displayed as a function of exposure number: Gaussian fit (red) and CF method (black).

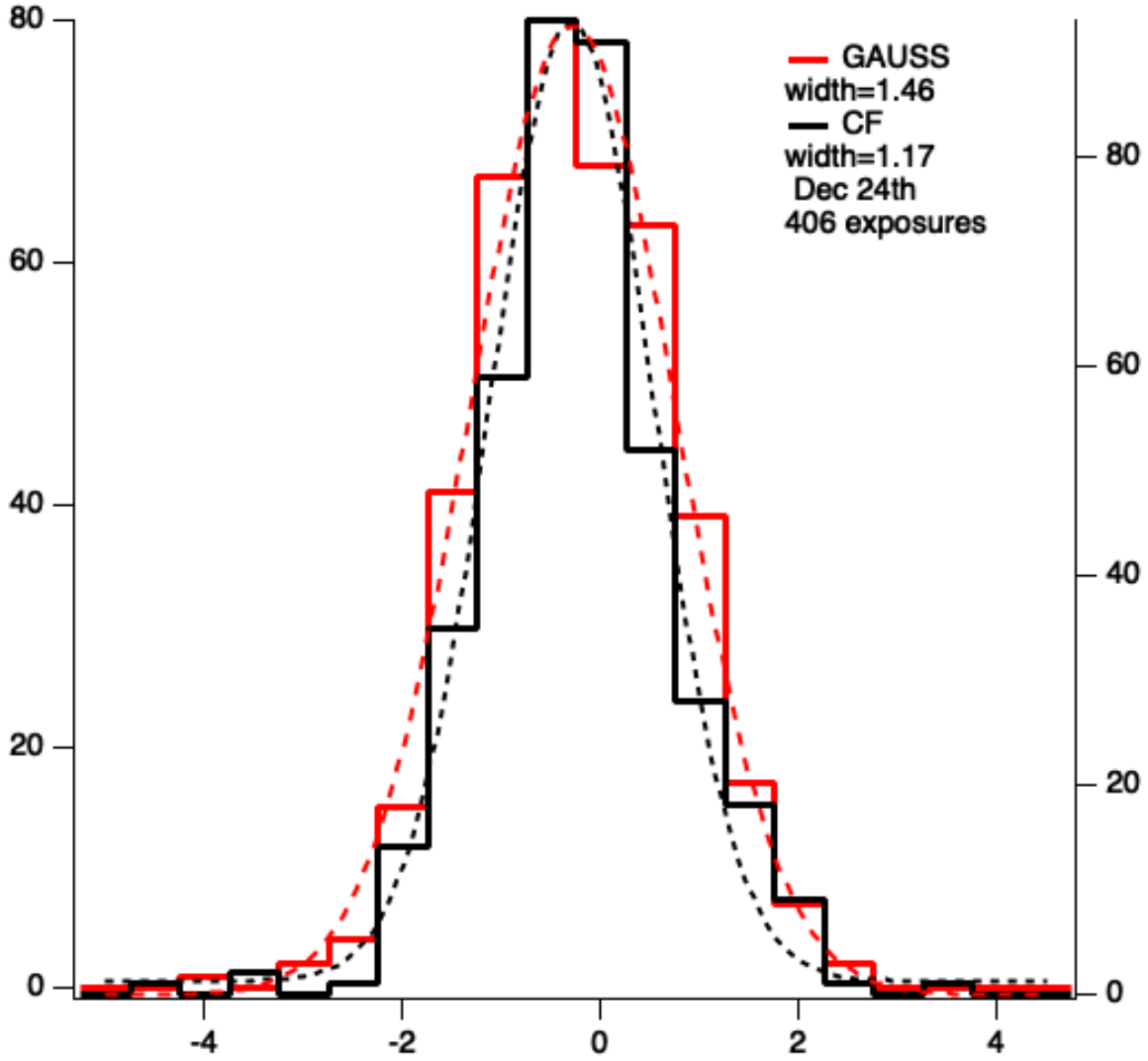


Figure 3.11: Histograms of RV obtained from Gaussian fit (red) and CF (black) for the 406 exposures of nigh December 24 2018, after removal of linear fits to each series. Gaussian widths of histograms are indicated for each method.

3.4 Application of First Formula of Connes to test differences between detectors

The star HD67458 was observed with ESPRESSO in the nights of January 17, 18, 21, 2021, in the frame of a co-ordinated program of observations with CHEOPS, ESPRESSO and TESS, as reported in Sulis et al. (2023). As stated by the authors, the time fluctuations of observed RV for the star HD 67458 (with no planet, and relatively low magnetic activity) "results from the contribution of three phenomena: instrumental noise, stellar oscillations, and stellar granulation."

The time scale of photon shot noise is of course the sampling period; the spurious instrumental noise

may also be as short as this sampling period, and may be sporadic. For the intrinsic stellar fluctuations, we quote Sulis et al. (2023) : "The stellar oscillations (or p -modes) evolve on short timescales for both stars, typically over 10 min or less. The stellar granulation, which we seek to identify here, evolves on longer timescales."

It should be noted that both datasets (HD 40307 and HD 67458) were taken before the ESPRESSO upgrade, so the blue detector thermal stability was affected and this can introduce some spurious fluctuations into the measured RV time series and increases the dispersion. In order to take this into account one can limit the number of working orders only to the red arm (orders from 90 to 170). Or we can select any combination of orders, since we have access to the CCF for each order, and apply to the CCF the 1^{st} formula of Connes, something impossible to do *a posteriori* for the RV pipeline values.

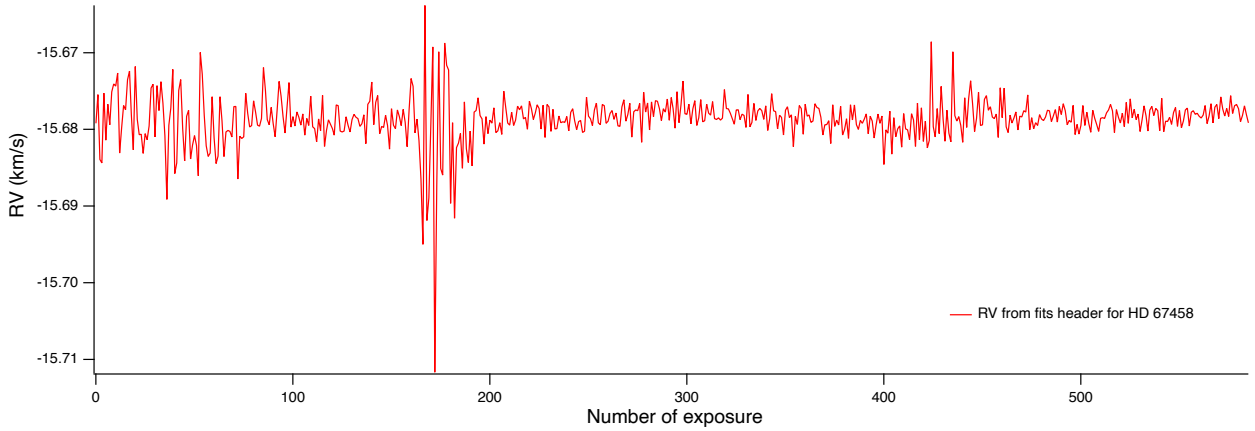


Figure 3.12: Star HD 67458: Dataset of 585 RV values from fits headers of ESPRESSO CCFs. Here we have concatenated all observations obtained during nights of 17,18, 21 January 21. We find indeed strictly the same pattern as displayed on Figure 1 of Sulis et al. (2023) which is normal, since we use the same data.

Let us show how the retrieved RV values changes if we change the ensemble of orders which we take into consideration for the computation of RV retrieved from the combination of orders. Figure 3.12 illustrate the original dataset of 585 RV values for star HD67458; it can be noted, that at exposure number 165 the RV values fluctuate quite strongly and fast. This set of exposures corresponds to the beginning of the second night of observation. One can suspect that this fluctuation has an instrumental nature. It may be possible to check this by the PC algorithm.

On figure 3.13 are displayed two time series of RV changes. In red, this is the ΔRV, difference of the RV pipeline contained in the data header with the first one. We computed with our PC algorithm on the basis of all orders with the 1st formula of Connes, the reference exposure being the first exposure

of the first night. We obtain the black curve, right scale. We see also the strong fluctuations at the beginning of the second day. Roughly speaking, the two time series overlap each other.

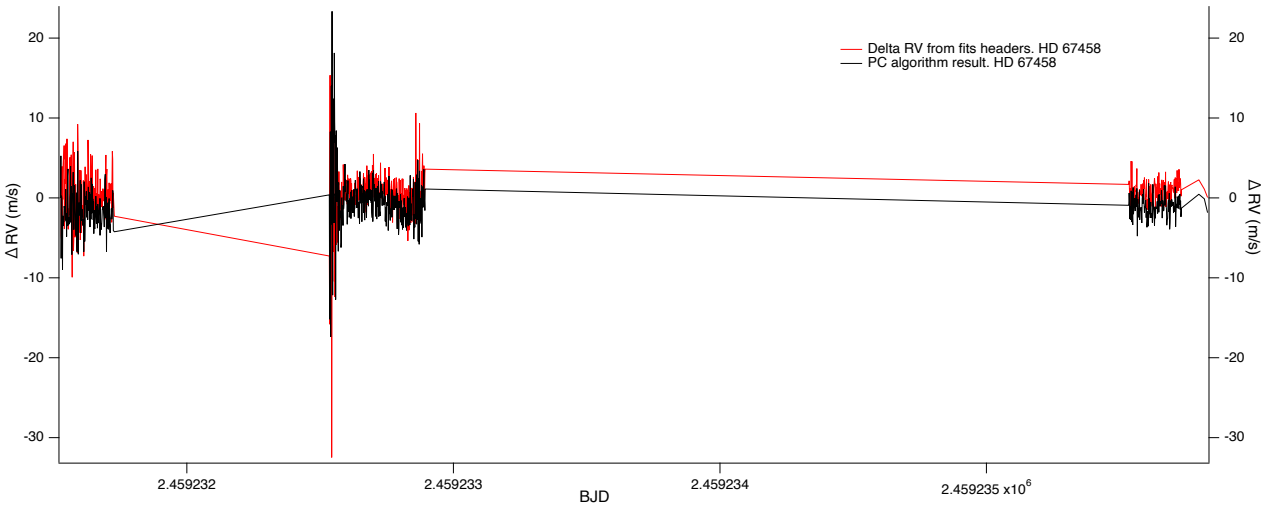


Figure 3.13: Red - ΔRV from fits headers (left scale). Black - ΔRV from PC (right scale).

Figure 3.14 is a zoom on the time series, it is possible to see that PC algorithm results and official pipeline results of ESPRESSO are in good agreement. Both methods display during the first half of the night relatively large fluctuations, greatly reduced during the second half of the night.

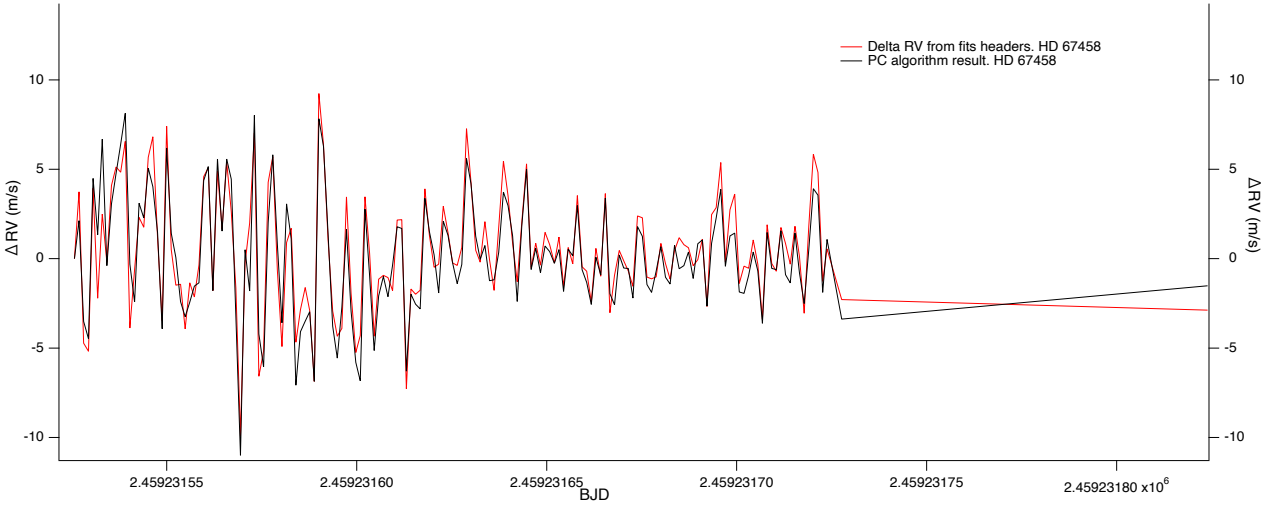


Figure 3.14: Red - ΔRV from fits headers (left scale). Black - ΔRV from PC (right scale).

It is possible to make an assumption that such strong fluctuations for the first half of the night have an instrumental cause and it is possible to explore which orders cause it. To start with, as it was already described, one can divide orders on blue (0-89 orders) and red arm (90-169 orders). Computed ΔRV for only the blue arm and only the red arm are shown on Figure 3.15 and 3.16 respectively.

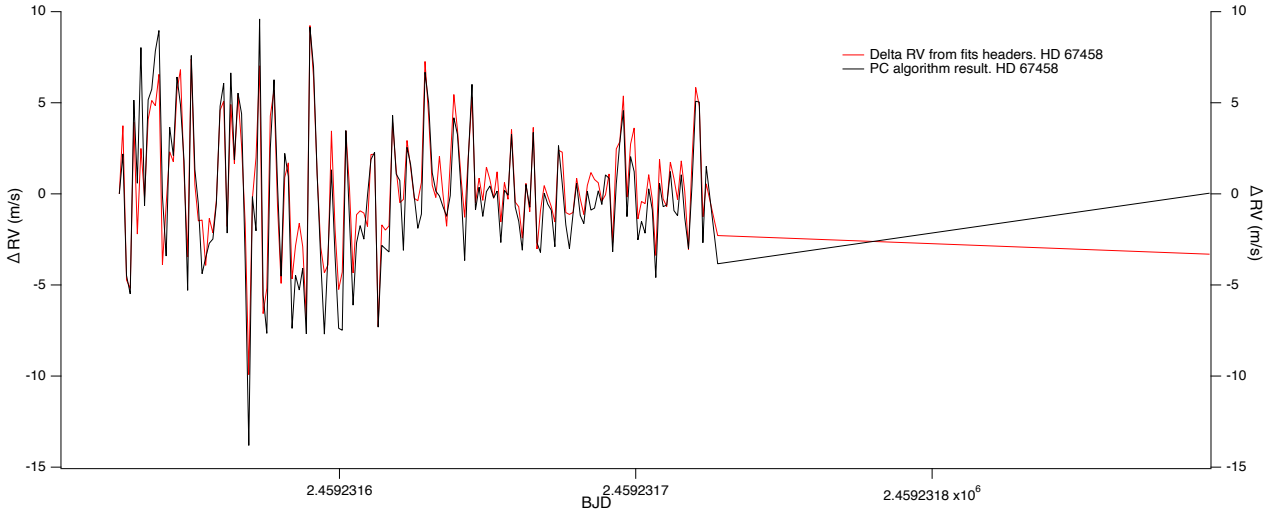


Figure 3.15: PC algorithm only for the blue arm (0-89 orders). Red - ΔRV from fits headers (left scale). Black - ΔRV from PC (right scale).

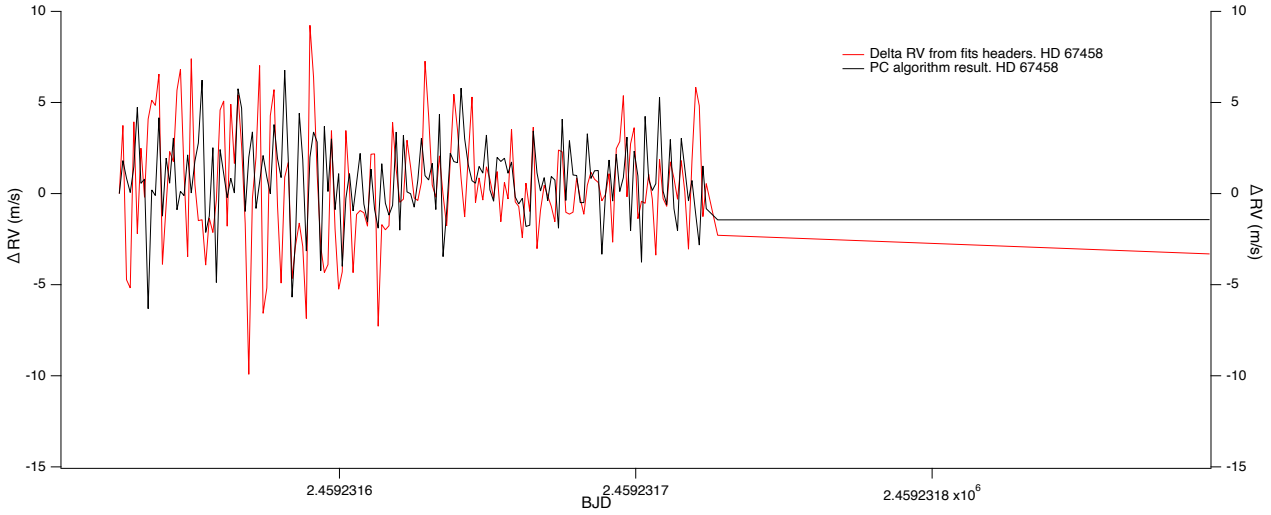


Figure 3.16: PC algorithm only for the red arm (90-169 orders).Red - ΔRV from fits headers (left scale). Black - ΔRV from PC (right scale).

It is shown that the fluctuations during the 1^{st} half of the night are much smaller for the red arm only than for the blue arm only, and smaller than the fluctuations of the RV pipeline. Quantitatively, the standard deviation of RV measurements is 3.787 ms^{-1} for the blue arm, and 2.201 ms^{-1} for the red arm. This wavelength dependence excludes the p-modes as an explanation of these fluctuations, since Doppler effect would manifest at all wavelength simultaneously. The role of granulation is also excluded because the stochastic fluctuations should be present simultaneously both in the blue and in the red arms. Therefore, these fast fluctuations much larger in the blue arm are most likely the indication that the blue arm detector instability is the cause of this difference. And since the weight of blue orders is much larger than the weight of red orders (partly because telluric regions discarded

from the analysis affect mainly the red arm orders), it is not so surprising that both algorithms give similar results when all orders are combined together, as shown on Figure 3.14.

It is possible to check exposures order by order, as it is shown on Figure 3.17.

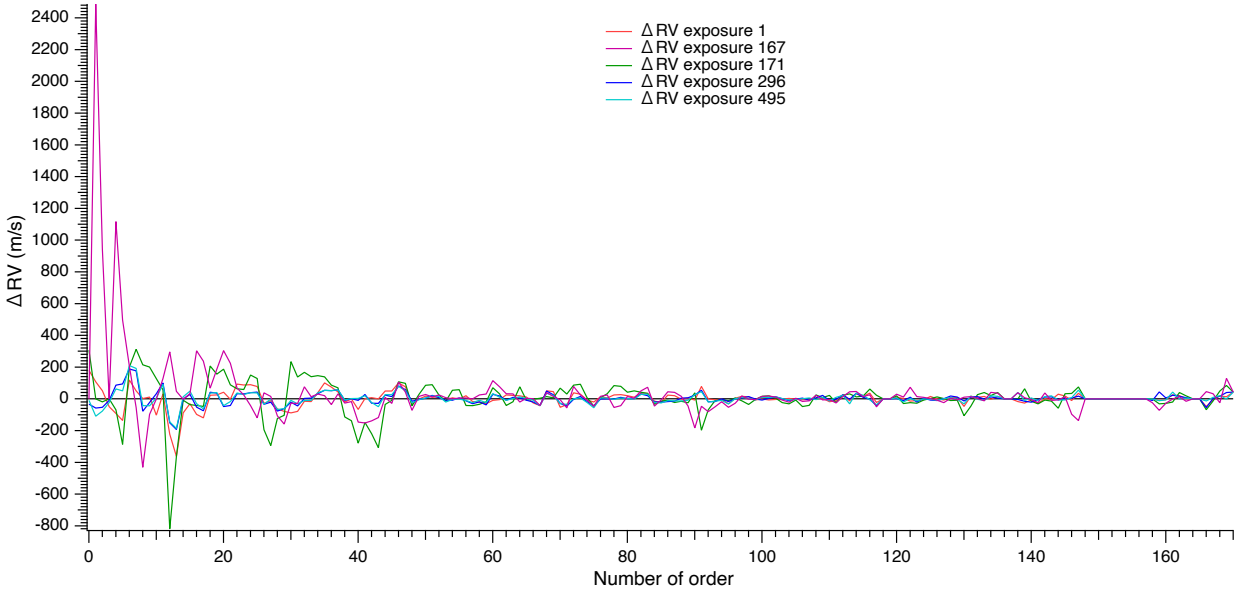


Figure 3.17: ΔRV as a result of PC algorithm by order for several exposures. Exposure 1 - second exposure of the 1^{st} night. Exposure 167 and 171 - exposures of the first half of 2^{nd} night. Exposure 296 - exposure of the second half of 2^{nd} night. Exposure 495 - exposure of the 3^{d} night.

In fact, RV for this star should be stable and ΔRV should be around 0 (with some fluctuations due to granulation stochastic fluctuations and random error). On figure 3.17 it is shown that some orders, such as blue orders, are more biased.

On Figure 3.18 are plotted for several exposures of this time series, the value of Q_{CCF} returned by the 2^{nd} formula of Connes, allowing to give an estimate of the error on RV derived from a particular CCF for a given order. There are no values for orders contaminated by tellurics and having no lines in the original binary mask. Clearly the average value of Q_{CCF} is larger for blue orders than for red orders, and this is mainly due to a larger number of lines in the blue. The dotted blue line is the value of Q_{CCF} computed from a spectrum stacked over 100 exposures therefore with a very high SNR. It is clear from the plot that this Q_{CCF100} is the smallest of all values of CCF. This is because some noise in the CCF is interpreted as a real spectral feature, in which case the derivative from one point to the next in the CCF is overestimated, and the value of Q will be overstimated as well. We can see that the values of Q_{CCF} are in general quite similar for all exposures, except with two main spectral regions: - the first orders of the blue arm, below order 20, and orders 90 and 91. Indeed the signal of the star is very weak in the first blue orders, as well as in orders 90 and 91 which are the first orders of the red

arm. The orders for which there are important variations of the Q_{CCF} should be discarded from the combination of orders to retrieve RV from the 1^{st} formula of Connes.

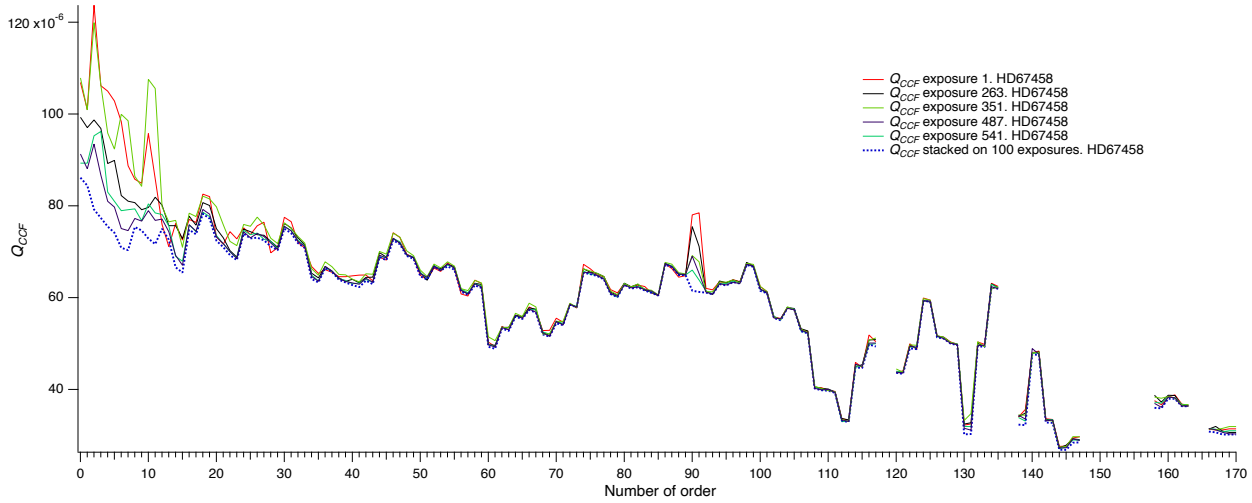


Figure 3.18: Example of Q_{CCF} values. Color solid lines – Q_{CCF} per order of exposures; Dashed blue line - Q_{CCF} per order based on stacked CCF of 100 exposures;

3.5 CCF technique for stellar lines temporal distortions detection

This section describes several computations using the blue and red parts of the CCFs that can be seen as an introduction to the more extended studies of the CCFs point by point in Chapter 5.

An important advantage of the CF method is its applicability to parts of the spectra (or here CCFs). In parallel with the above CF application to the entire 81 points of the CCFs, we applied CF separately to the blue and red wings of the CCFs. More precisely, since the central value is at pixel number 38, we used the intervals 0-38 for the blue part, and 38-80 for the red part. The resulting RVs for the Dec 24th night are shown on Figure 3.19. The figure reveals significant changes along the night of the two RVs, and, interestingly, essentially symmetrical variations, a sign for line broadening. The same changes are obtained if we apply the CF method to individual orders and combine the orders using our method discussed above. This is shown on Figure 3.20.

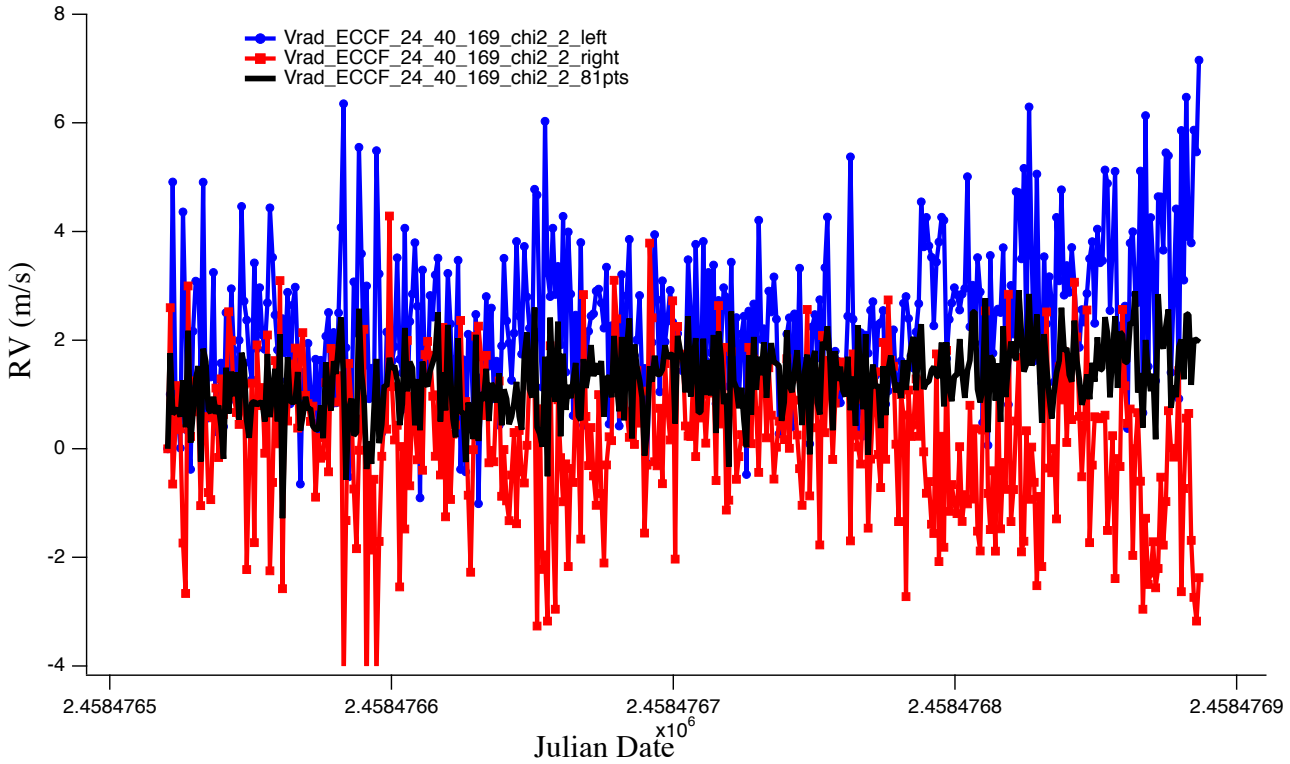


Figure 3.19: RVs obtained from pipeline CCFs using CF formula for left and right parts of CCFs. In blue, RV from left side of the CCF (blue shift). In red, RV from right side of the CCF (red shift). In black, RV from the full CCF. All curves use the left scale in ms^{-1}. The right scale is not used in this graph.

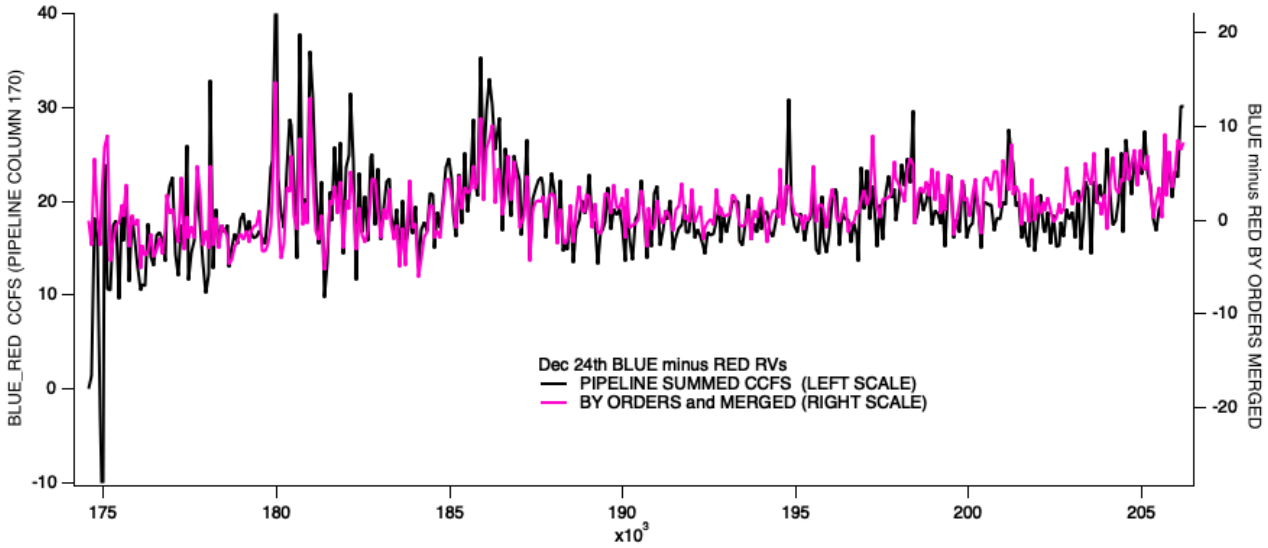


Figure 3.20: Temporal evolution of the BLUE-RED RV difference based on CF during the Dec 24th night. Deduced from the pipeline order-merged CCFs (black) and computed independently for each order and merged (pink line). The same features are detected.

We checked this effect (potential change of shape of spectral lines) by normalizing the CCFS of the first and last exposures and subtracting the first from the last. The result is shown on Figure 3.21. The shape of the difference is compatible with a broadening of the last CCF w.r.t. the first,

and also reveals some asymmetry. As discussed below, more detailed applications of the CF to the CCFs should help disentangling the broadening (which is a distortion) from a simple global shift. Figure 3.22 displays the blue and red RVs for all six nights. The first three nights have an average RV (Red)-RV(Blue) difference larger than the last three nights, although strong variations are also visible within a single night.

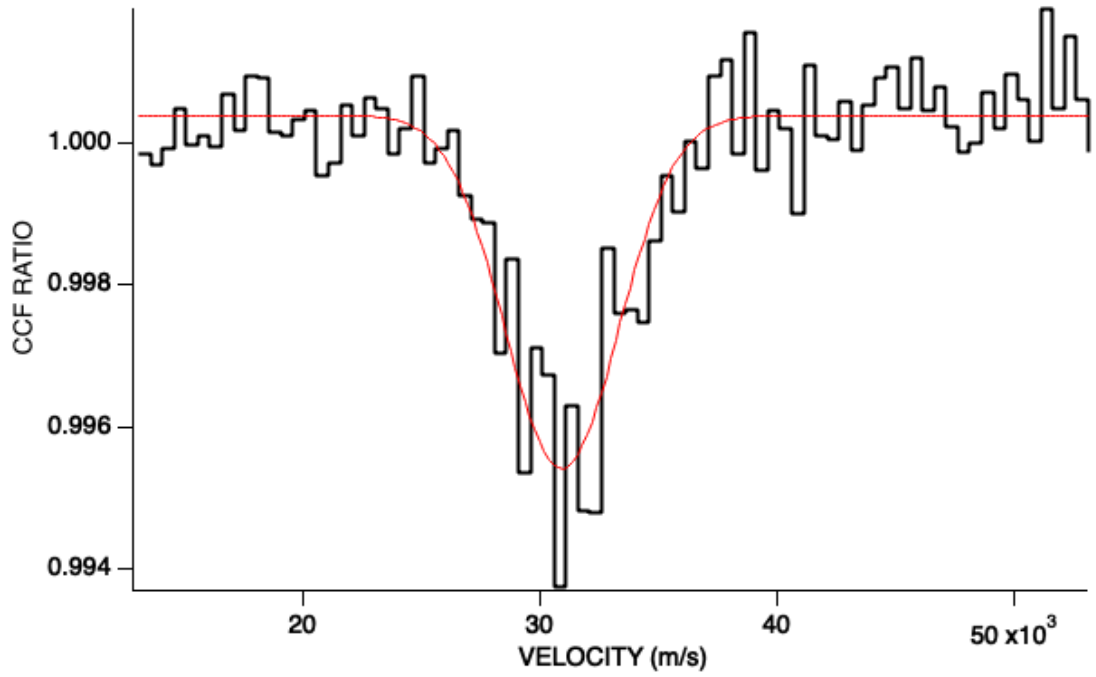


Figure 3.21: Ratio between the summed twelve last and summed twelve first order-merged pipeline CCFs of Dec 24th night. Before the ratio computation the summed CCFs were normalized to the same area. The pattern shows that the end-of-night CCFs were relatively deeper by about 0.5%. This is a sign of variation of shape.

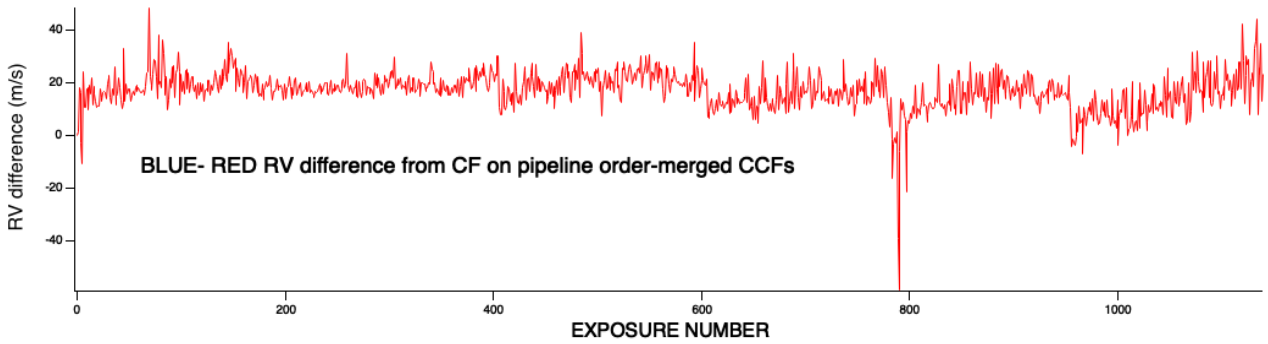


Figure 3.22: Temporal evolution of the BLUE-RED RV difference based on CF during all the nights deduced from the pipeline order-merged CCFs.

3.5.1 An attempt to link line width and discrepancy between Gaussian fit and CF method

The 1st formula of Connes, when applied to a piece of spectrum or a CCF, is proven to get the best precision (optimal method) on its displacement (change of RV, dRV) between two epochs of sampling. Indeed, the dispersion of a time series of dRV is found smaller than the use of the Gaussian fit to the same time series. It is more precise. However, CF1 assumes that the piece of spectrum (or CCF) does not change at all in shape : only a shift, and no distortion at all. When applying CF1 in presence of a distortion, it will introduce a bias (difference with the true value of dRV), which will depend on the amount and type of the distortion, therefore will depend on sampling epoch. Actually, this distortion may be studied with CF1, since it gives an information on each of all the points of the CCF. Therefore, the question arises : can we find a method, based on CF1, to disentangle the distortion from the global displacement when comparing two CCF at two epochs?

We wished to study this problem with an empirical approach, rather crude to start with. We built 4 time series from the application of CF1 to the total CCF (column 170 in the official product):

- S_g - obtained with RV retrieved from a Gaussian fit to the CCF (81 points) and subtraction of one particular value RVref determined for one exposure serving as a reference.
- S_{tot} - CF1 applied to the 81 points.
- S_{blue} - CF1 applied the blue part of the CCF (points 1 to 38)
- S_{red} - CF1 applied the red part of the CCF (points 38 to 81)

In reality, when applied separately to the blue and red half of the CCF, we do not find the same displacement result. This is a proof that the shape of the CCF is changing. In order to characterize crudely the distortion, we built the difference $S_{b-r} = S_{blue}$-S_{red}, which corresponds to the width of the total CCF. It is indeed changing along the time series. Next, in order to characterize the bias associated to the CF1 application to a distorting shape, we built also the difference S_b=S_g- S_{tot} as another time series.On one hand, we see that there is a noticeable trend for this time series of differences. Both S_g and S_{tot} are supposed to follow the RV variations of HD 40307 induced by the 3 planets, and their difference should remain 0 or constant. Therefore, since the difference is changing with time, at least one of the series S_g and S_{tot}, contains a variable bias, and may be both series. We assume for a while

that the Gaussian fit to retrieve RV does not contain a variable bias. When plotting S_b=S_g- S_{tot} as a function of S_{b-r} = S_{blue}-S_{red} we note that the scatter plot shows a linear trend. We may inverse this linear trend, and compute one value S_{bc} for each measured value of S_{b-r}: S_{bc}=a+bS_{b-r}. Then, similarly to the relation S_g= S_{tot}+S_b, we compute a new estimate of a time series S_{gc} = S_{tot}+S_{bc}, which is entirely based on CF1 applied to CCF, and not at all on a Gaussian fit. So doing, we expect that the new corrected series S_{gc} will have a similar bias as the series S_g (therefore no variable bias, according to our assumption), and a smaller dispersion than the Gaussian fit.

3.5.2 Actual preliminary application to HD 40307

Figure 3.23 displays the difference in RV between the CF and Gaussian fit methods (i.e. S_{tot}-S_g) as a function of the difference between the blue and red (S_{blue}-S_{red}). Colors refer to the selected night . We omitted Dec 22d due to the low number of exposures. We can see a trend and mainly distinguish two groups of points, correspond to the differences between the first and last three nights mentioned above. We show in red a line connecting the two main groups. This line corresponds to a relationship that we have inverted and we used the inverted relationship to correct the bias due to the stellar line shape variability. The comparison between RV from the Gaussian fit and the corrected RV from the CF method is displayed on Figure 3.24. A large fraction of the bias visible in the case of the initial CF1 results (Figure 3.10) from Dec 26th has disappeared. However, because we used the blue-red differences to compute the correction, the dispersion now is strongly increased. This shows that more work is needed to keep the main advantage of the CF method, i.e. a high precision and reduced dispersion, while simultaneously disentangling what part of the bias is due to CCF broadening, from what part is due to distortion.

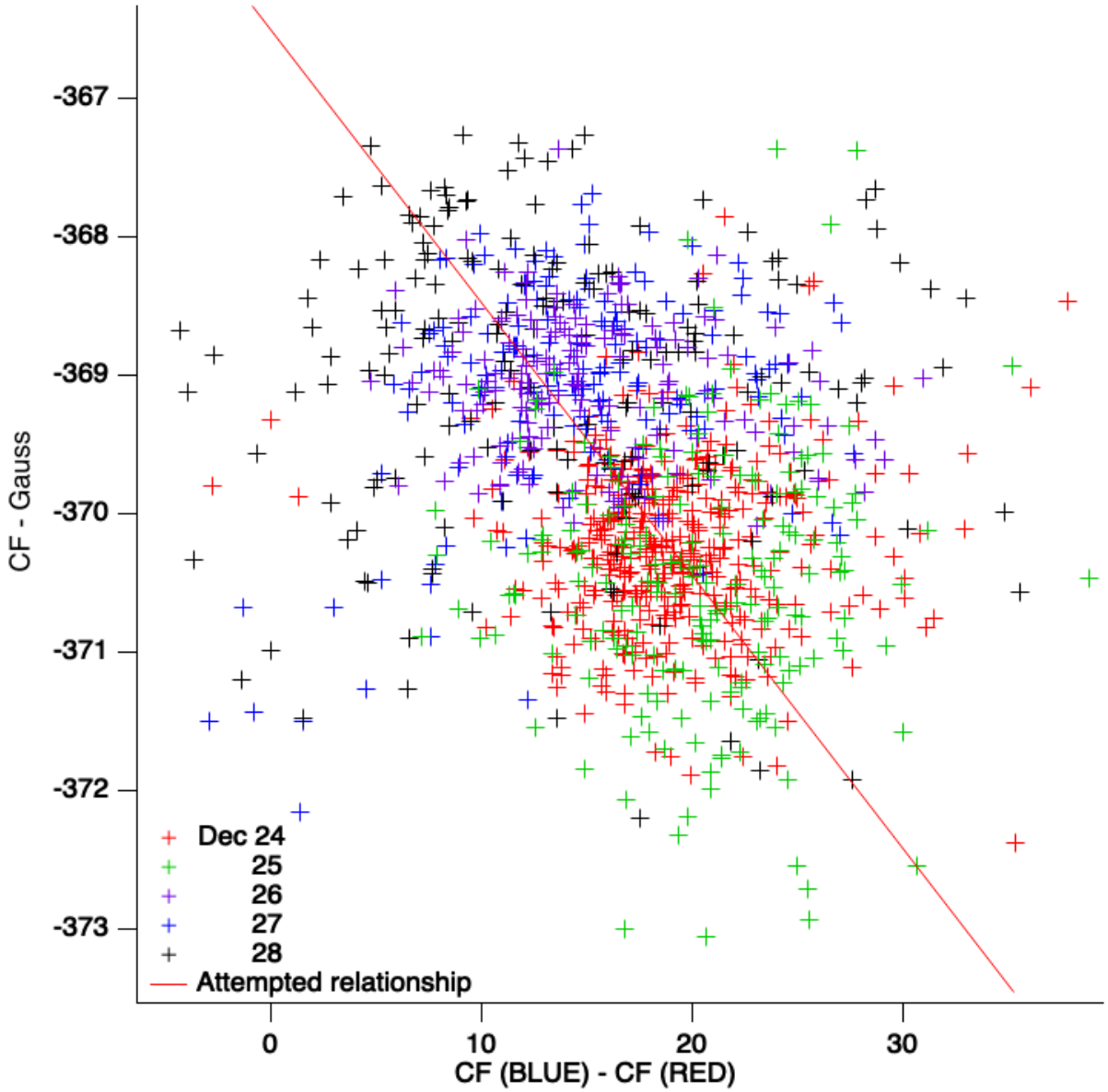


Figure 3.23: RV offset between the CF and Gauss methods as a function of the BLUE-RED RV difference based on CF1.

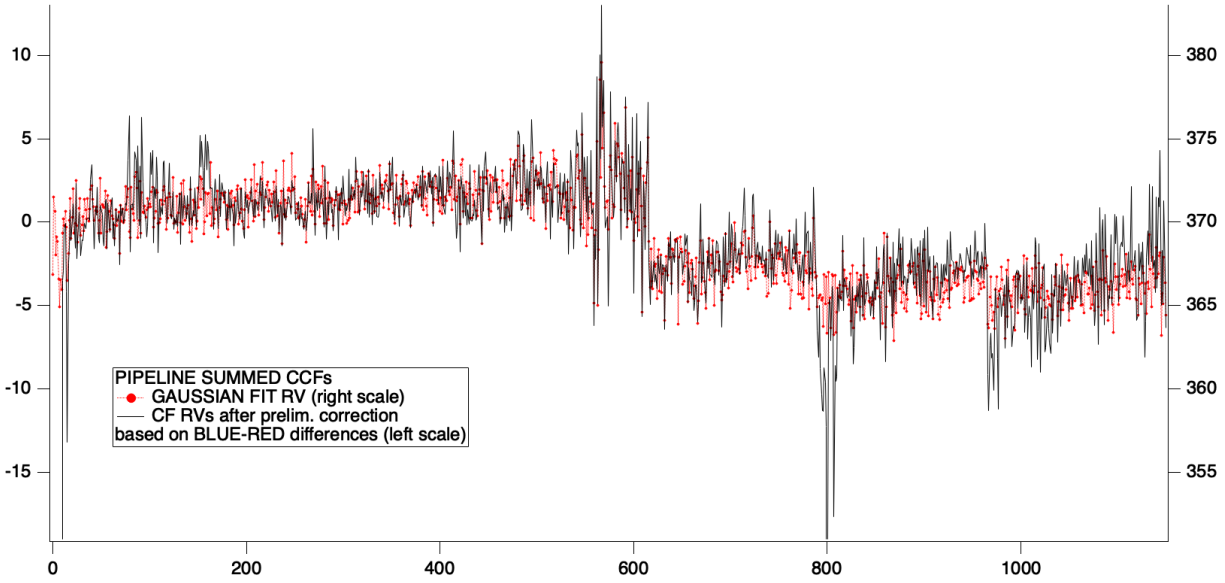


Figure 3.24: Attempted correction of CF1 RV level based on Figure 3.23. The average levels are in better agreements with the RVs from the Gaussian fits. However the correction has introduced a significant additional dispersion, to be treated separately.

3.5.3 Preliminary application to HD 67458

It is interesting to examine the relationship between the blue-red difference and the RV value from the CF method for the no-planet star HD 67458. If there are variations in the CCF shape they should affect both the RV derived from the full CCF and the blue-red difference. If a strong relationship were found, it could be used as a way to build a RV correction and mitigate the effects of shape variability. The dataset of HD 67458 is also spoiled by the instrumental instabilities (especially blue orders), so it is not possible to use whole time series. We built by-eye a mask on the RV values to exclude all points of high variability regions, also we used a median filter and smoothing sliding window over 10 points to transform the blue-red difference.

Figure 3.25 shows the computed RV values, obtained by CF to the CCFs from order 20 to order 170 (black); along with the computed RV values, obtained by CF used only to the blue (resp. red) wings of CCFs in blue color (red color). The chosen mask is shown by the absence of markers. Figure 3.26 shows the difference between RV calculated on blue and red wings as a function of exposure number, exposures which were removed by the filtering procedure are shown by the red line with markers, while the exposures, which passed the filtering procedure are shown in plain red line with over-plotted blue line (mask).

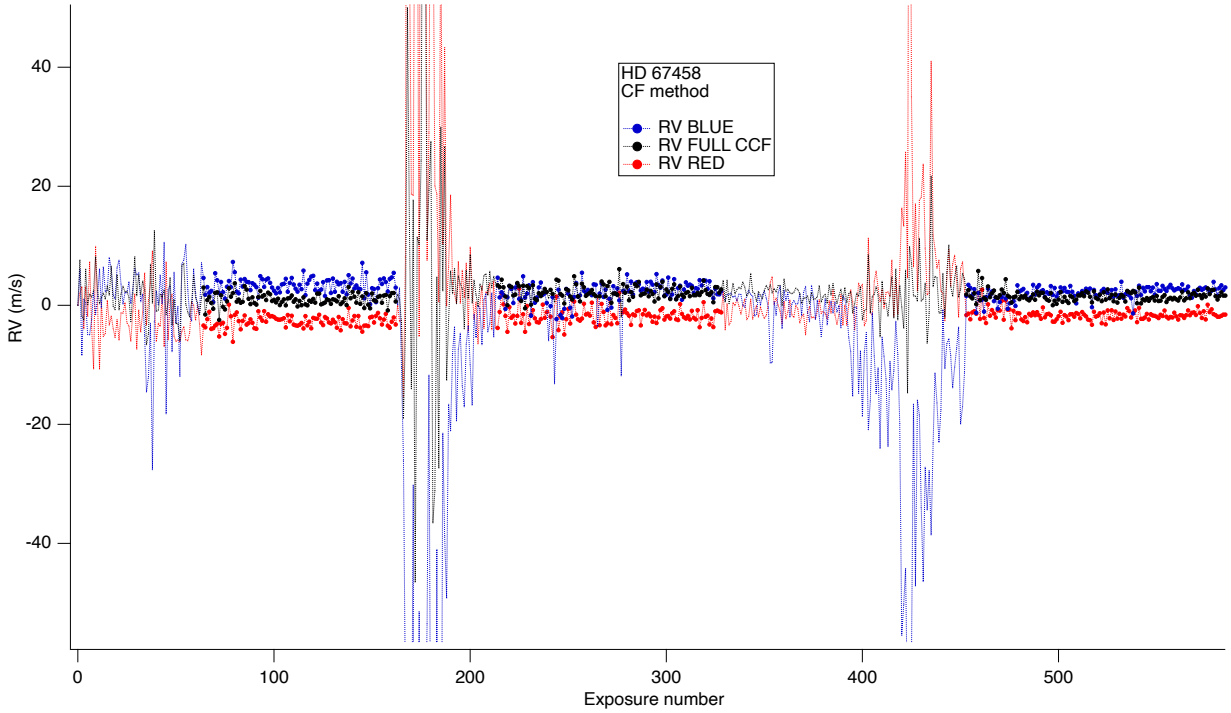


Figure 3.25: RV from CF based on orders 20 to 170 (black line and markers). Superimposed are RVs from CF applied to the same CCFs, using only the blue part (blue line and markers) or only the red part (red line and markers).

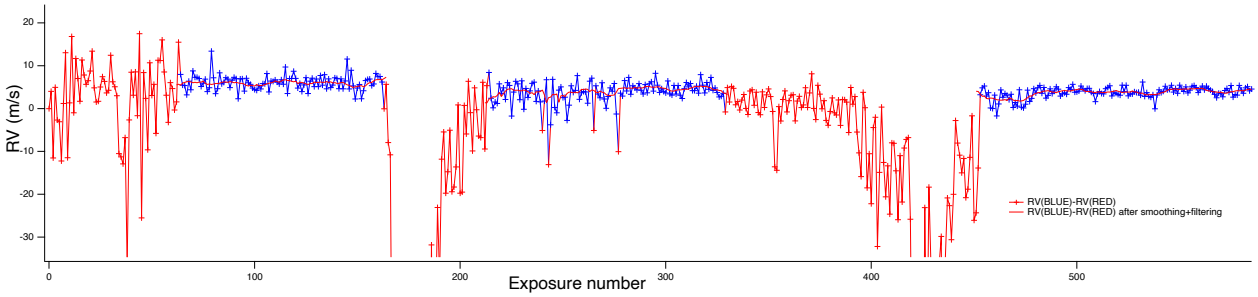


Figure 3.26: RV (BLUE) - RV (RED) after smoothing and filtering.

Fig. 3.27 displays the BLUE-RED difference as a function of the RV from CF applied to the full CCF. A linear slope confirms a relationship between the two quantities, namely : RV (full CCF) = 5.05$\pm$ 0.10 - 0.35 $\pm$ 0.05.

Applying a correction to RV following the above gradient, we find new RV values that are more constant along the three nights, as shown on Figure 3.28. Such correction is a preliminary example, further research and development of the method will help improve the accuracy of the correction.

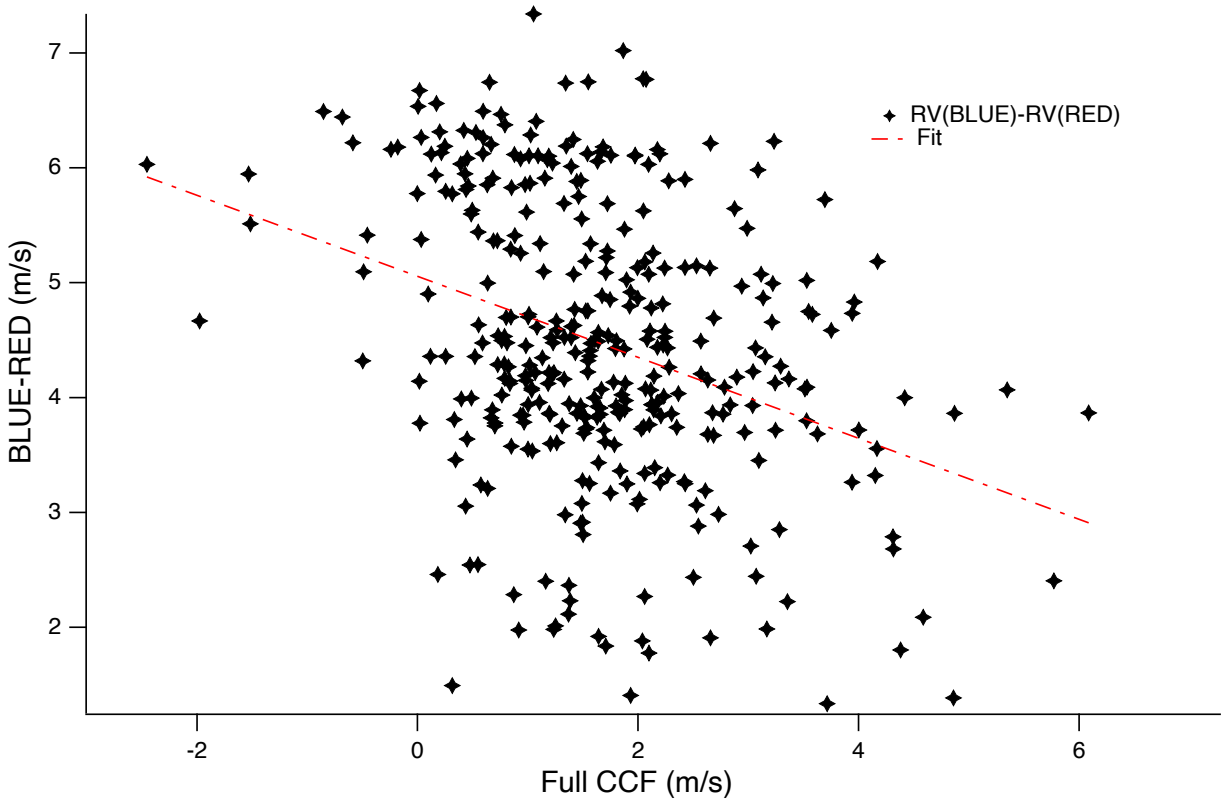


Figure 3.27: Relationship between the difference RV(BLUE)-RV(RED) and the absolute value of RV from the CF method applied to the full CCF. A linear fit RV = a+(BLUE-RED)b gives a slope b=-0.35 $\pm$ 0.05.

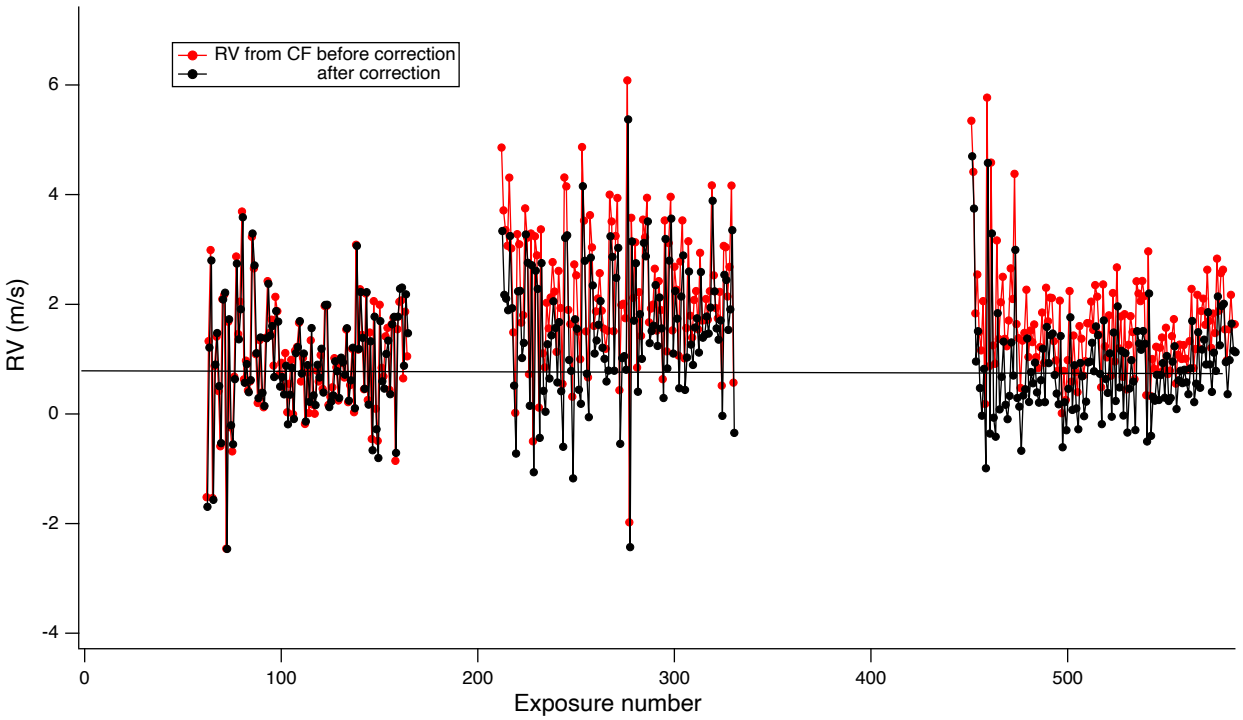


Figure 3.28: RV from CF before and after application of the gradient shown in Fig. 3.26. The RVs display smaller departures from constancy during the three nights

Chapter 4

Study of the HD 40307 planetary system

As it was already mentioned in the previous section the CCF matrix which could be downloaded from ESO Archive has an additional column equal to piled-up order CCFs, we will call it CCF_{tot}. Such CCF_{tot} is robust to several artefacts: any pattern, which is fixed, in the wavelength frame of the spectrometer, while the star spectrum is moving by up to $\pm$ 30 kms^{-1}, will be detrimental (producing a bias on RV) for all methods which compare directly the motion of the spectrum, as the Connes algorithm or others (like template matching of an observed spectrum). This is the case of PRNU (Pixel-to-pixel Non Unformity, DCNU (Dark Current Non Unformity), optical fringes, and micro-tellurics (large tellurics are just avoided by standard Binary Masks). The CCF_{tot} is robust (but not perfect) because all these effects are distributed among many orders and many lines, and there is some averaging out of their adverse effects when all lines of the spectrum are combined into a single CCF.

In Chapter 2 we proposed to use the optimal retrieval, photon noise limited, algorithm based on the first formula of Connes (CF1) with robust CCF_{tot}. The CCF of a star is computed for two spectra obtained at two epochs, yielding CFF1 and CCF2. One trick is that the CCF is not computed on the wavelength scale provided by the laboratory calibrated spectrometer, but on the wavelength scale produced after BERV correction (Barycentric Earth Radial Velocity). It means that it provides the spectrum of the star as if it were observed form the barycenter of the Solar System. As already mentioned in sub-section 3.3 in the Solar System barycentric frame of reference, the radial velocity of a star experiences variations caused by the potential presence of exoplanets as well as stellar activity. Therefore, there is only a small displacement of the two curves CCF1 and CCF2, suitable to use the Pierre Connes first formula CF1 valid for small displacements. To the best of our knowledge,

this method has never been used nor described before. Still, the use of the BERV-corrected spectra has been used within template matching methods, for instance by Dumusque (2018); Cretignier et al. (2020); Artigau et al. (2022). These works are rapidly overviewed in Chapter 5. We call this method of applying the 1^{st} formula of Pierre Connes to the CCF, the EPiCA method (Exoplanets Pierre Connes CCF Algorithm).

Ideally, in order to compare two methods yielding changes of RV, one could use a planet-less star with a constant RV, and compute the standard deviation of a series of measurements of RV changes around the mean which should be 0. Actually, we wished to use ESPRESSO data which seem to provide an absolute wavelength scale of excellent quality (Pepe et al., 2021), possibly the best in the world. Therefore, we have used a series of 1151 exposures spread over 7 days of star HD 40307, more about it can be found in section 1.4.2.

Since this star has some planets, there are some RV variations due to these planets. In such a case, the comparison criteria between the two methods may be the χ^2 characterizing the difference between data and model; or it may be the consistency of retrieved RV with a planetary system. Indeed, we will attempt to retrieve some planetary parameters (mainly the periods) from the two time series obtained with the two methods, find a least-square best fit planet model, and scrutinize the residuals between data and best-fit model. We have already compared the two time-series ("pipeline" and CF applied to CCF_{tot}) in Figure 3.9. Below we plot again the two time series (concatenated, plotted as a function of exposure number).

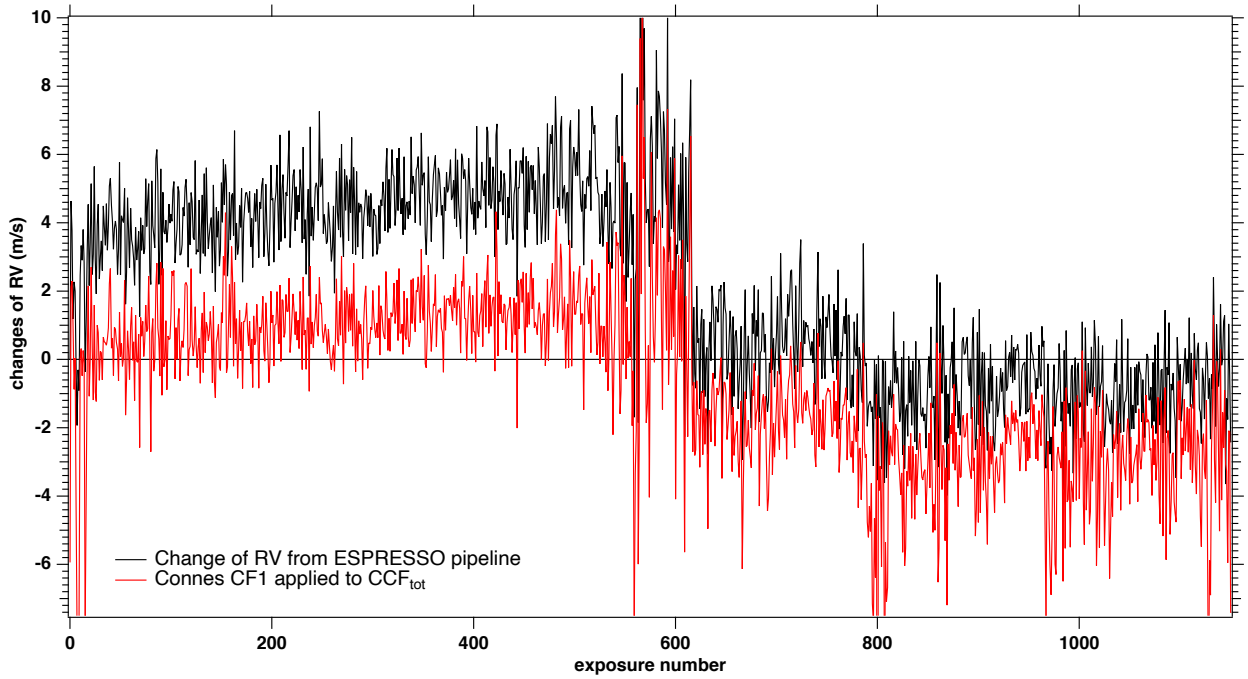


Figure 4.1: The data (RV changes in ms^{-1}) are plotted as a function of time with two colors. Black points are derived by subtracting from all pipeline RV data contained in each spectrum fits file the value of the first exposure: 31.3668 kms^{-1}. Red points are obtained by the Connes formula 1 (CF1) applied to the CCFs of the current exposure and the first exposure of night 24 (exposure n^{o}10 when all nights are concatenated), taken as a reference.

Both time series show a rather sharp decrease after exposure 615, which is the end of the night 24. This is due to the presence of planets, as we shall see later. The offset between the two series is partially due to the difference of reference exposure: n^{o}0 for ESPRESSO pipeline, n^o 10 (first of night 24) for Connes CF1 applied to CCF$_{tot}$. But the offset is not constant, as can be seen when plotting the difference between the two time series, figure 4.2

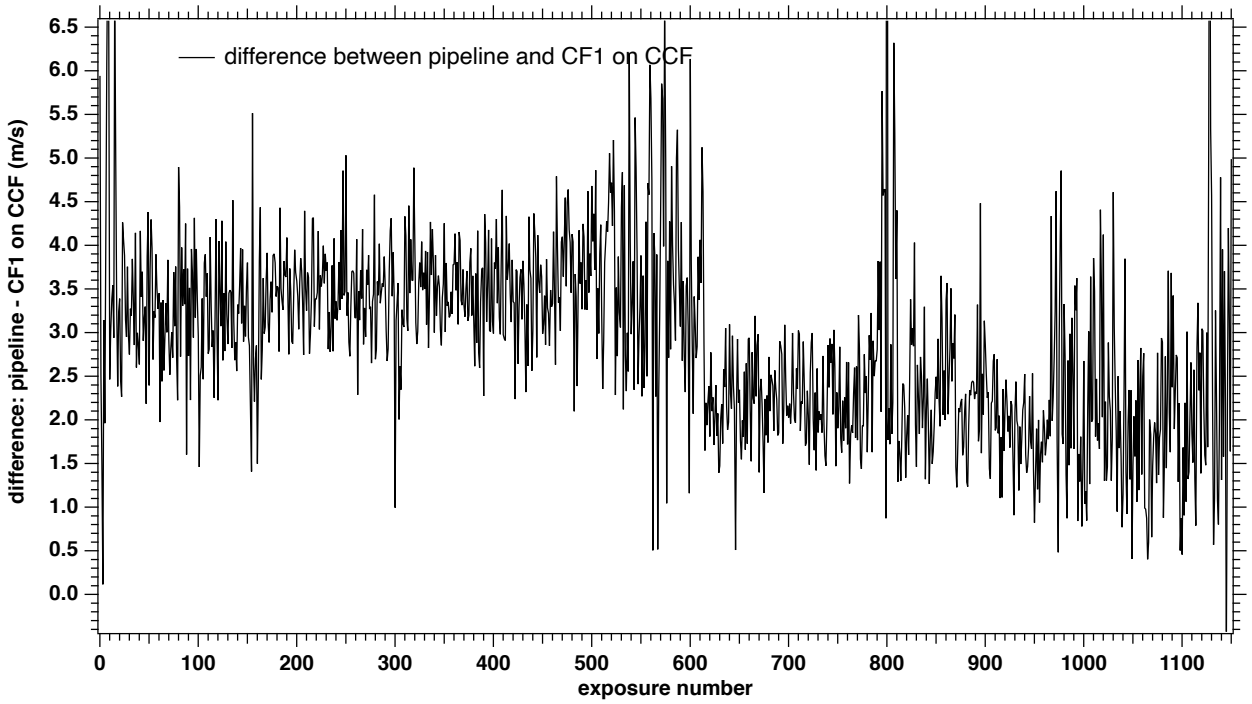


Figure 4.2: Difference of the two time series of RV changes: ESPRESSO pipeline - CF1 applied to CCF_{tot}. This difference shows an abrupt decrease ($\sim$1.3 ms^{-1}) after exposure n^{o}615 and corresponds to transition from night 25 to night 26, which remains after that.

Therefore, one of the two methods is giving some wrong results by about 1.3 ms^{-1} (or may be both of them), which do not correspond to a true change of the dynamical radial velocity dR/dt. We suspect that this is the effect of stellar activity, and more precisely of granulation shift changes, with a time scale of a few hours. For the time being, we are going to fit both time series with a model with 3 planets as was discovered previously by Mayor et al. (2009).

A 3-planet best fit to the official pipeline ESPRESSO RV time series.

In Table 4.1 we summarize parameters of 3 planets around HD 40307, more about their discoveries and some debates can be found in section 1.4.2.

	HD 40307b	HD 40307c	HD 40307d
Amplitude and error [ms^{-1}]	1.84$\pm$ 0.14	2.29$\pm$ 0.13	2.31$\pm$ 0.14
Nominal period and error [day]	4.3114$\pm$0.0002	9.6210$\pm$0.0008	20.412$\pm$0.004
Zero crossing time T_0 at epoch 2009 (bjd-2.4e6)	54562.77$\pm$0.08	54551.53$\pm$0.15	54532.42$\pm$0.2

Table 4.1: Characteristics of three planets in HD 40307 system from HARPS 2009 campaign. The values used were taken from http://exoplanet.eu and corresponds to the Coffinet et al. (2019)

Best-fit strategy and result of RV official pipeline ESPRESSO RV time series.

Our analysis is done only on ESPRESSO 2018 data, without archive HARPS data. ESPRESSO data set consists only of 6 days, so it is not possible to detect signal from the 4^{th} (HD 40307f), since in would induce only a small drift over 6 days. We have assumed that planets are on circular orbits as it is mentioned by Table 9 from Díaz et al. (2016).

Therefore, we kept fixed the amplitudes and periods of 3 planets. We know the time T_0, the origin of the sinusoidal wave at epoch 2009 from the exoplanet.eu website (see table 4.1). With these time values of zero crossing in 2009, we can extrapolate in time all 3 sine waves with the defined periods up to the epoch of 2018 observations, getting the curve $RV_{nominal}$ (pale grey) on Figure 4.3, compared to ESPRESSO data. We see a major discrepancy, easily explained: a small error on any of the periods, propagated over 10 years, will make $RV_{nominal}$not representing well the data of 2018. Therefore, we have to adjust the exact times of zero crossing for each of the sine functions to get a good fit of data. Actually, we take as unknown in the fitting process the exact phase of each of the sine function at a reference time (bjd_{ref}= 2458474.5), taken as the time 0 of the plot of Figure 4.4 It is adjusted by a classical Levenberg-Marquardt scheme within IGOR software. We also have to let free a constant offset parameter w_0, added to the three sine functions, since we deal only with variations of RV.

Best fit parameters:

Coefficient values $\pm$ one standard deviation

w_0=1.5583 $\pm$ 0.515 - ms^{-1}

w_1=-0.38916 $\pm$ 0.00861 - phase of planet b at time 0 of plot (modulo 1)

w_2=-0.098627 $\pm$ 0.0149 - phase of planet c at time 0 of plot (modulo 1)

w_3=0.3994 $\pm$ 0.0418 - phase of planet d at time 0 of plot (modulo 1)

On the Figure 4.4 is plotted the best fit curve, in red, while the 3 sine waves for the 3 planets are represented by dashed lines around the zero line. It is clear that the drop in RV between night 25 (day 3.2) and the next day is mainly due to planet b, with small contributions from planets c and d.

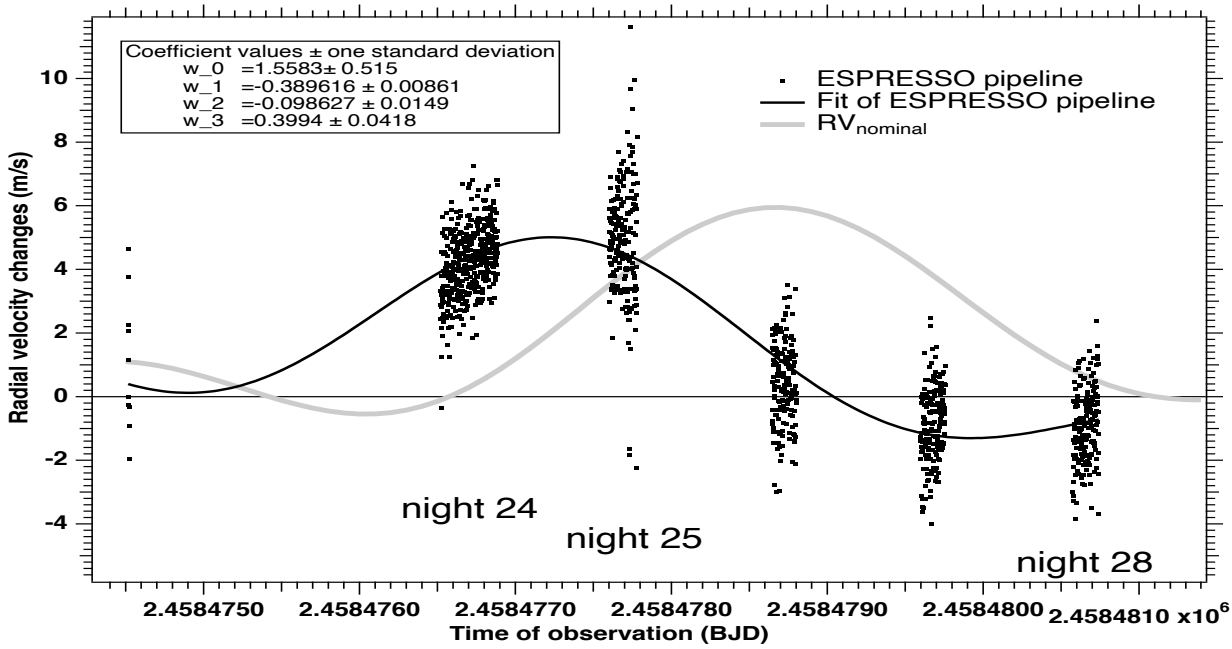


Figure 4.3: The black dots represent the time series of changes of RV values contained in the official ESPRESSO pipeline, obtained by subtracting the RV value of the first exposure of the first night. The thick pale grey solid line is the extrapolation of the sum of the 3 sine waves as determined form 2008 HARPS data. It does not fit well the data, because the extrapolation over 9 years is sensitive to small errors in the periods. The solid black line is the best fit to the pipeline time series.

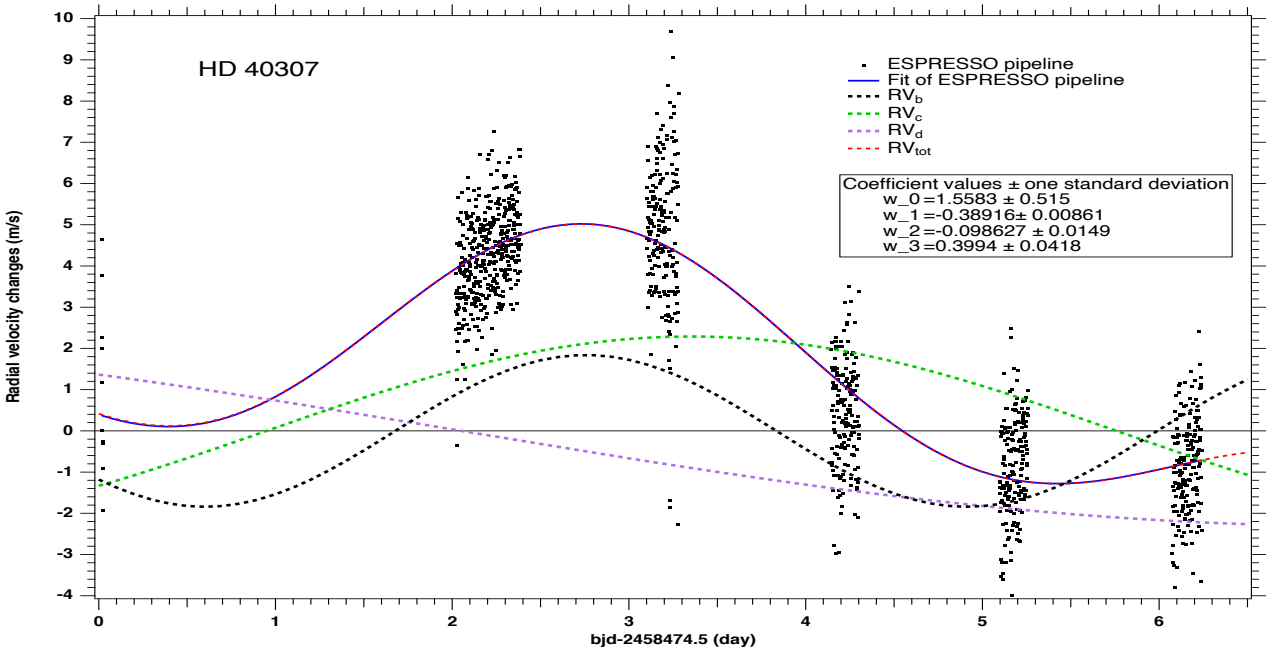


Figure 4.4: Same as figure 4.3. Now the best fit curve is in violet (solid line), while the dashed lines are each of the 3 sine waves representing the 3 planets : b (black), c(green), d (violet). The best fit curve is the sum of the three planets plus a constant $w_0 = 1.5583 \pm 0.515$ ms^{-1} represented in red dashed line RV$_{tot}$. The RV$_{tot}$ line (red dashed) is superimposed on the best fit curve and just prolongs it after the last data point.

Refining the periods by comparing 2008 and 2018 data. For each of the sine curve, knowing the phase at T_{ref} allows then to compute the time T_{2b} (respectively T_{2c}, T_{2d}) of start of the period (sin=0 and becoming positive) just before T_{ref}. The values are displayed in Table 4.2 with their uncertainties. We find in the literature that a similar sinus crossing happened for planet b at T_{0b} = 2454562.77$\pm$ 0.08 BJD. Therefore, we may determine a precise value of the average orbital period of planet b between epochs April 2008 and December 2018, knowing that it must be an integer number of periods between T_{0b} and T_{2b}. The elapsed time difference T_{2b}-T_{0b}= 3909.0964 $\pm$ 0.088 days. Here we take as the uncertainty the quadratic sum of the two uncertainties, respectively 0.08 and 0.037 days for T_{0b} and T_{2b}. Dividing the elapsed time difference by the nominal period 4.3114 gives a number of orbits of planet b K_{per}=906.688 (Table 4.2) , while it should be an integer number. We designate by K_1 and K_2=K_1+1 the two integer numbers encompassing K_{per}.

With 907 full orbits between T_{0b} and T_{2b}, the period is found to be P_{b907} = 4.30992 $\pm$ 1.0 10^{-4} days, while with 906 full orbits, the period is P_{b906} = 4.31467 $\pm$ 1.0 10^{-4} days (Table 4.2). These two values encompass the original value of the period found at discovery announcement 4.3115 $\pm$ 0.0006 and the more recent value Per_b=4.3114$\pm$ 0.0002. Therefore we favour the P_{b907} which is about 2.2 times nearer this value Per_b=4.3114 than P_{b906}. Still, this new accurate value P_{b907} is different from the nominal 2008 period Per_b=4.3114$\pm$ 0.0002 by 0.0015 d, outside the claimed error bar err_{Per_b}=0.0002 by a factor of 7. The planet b may have changed its average period from interactions with the other planets between 2008 and 2018. Much more likely, the uncertainty of 0.0002 claimed for 2008 data was underestimated.

The same procedure is done for planets c and d, results are in Tables 4.3 and 4.4 respectively. Clearly, K_{per} is nearer an integer number than for planet b, and the most likely values are K_{per}=407 for planet c, and K_{per}=193 for planet d.

	Planet b pipeline	Planet b CF1
Amplitude and error [ms^{-1}]	1.84± 0.14	1.84± 0.14
Nominal period and error [day]	4.3114±0.0002	4.3114±0.0002
Phase at BJD$_{ref}$ (epoch 2018)	0.6108± 0.0086	0.6266± 0.0087
Zero crossing time T$_2$ at epoch 2018 (bjd-2.4e6)	58471.866±0.037	58471.866±0.037
Elapsed time T$_2$-T$_1$ [day]	3909.1±0.09	3909.03±0.09
K$_{per}$ number of periods in T$_2$-T$_1$	906.688	906.673
K1	906	906
K2	907	907
Period and error for K1 [day]	4.3147±0.0001	4.3146±0.0001
Period and error for K2 [day]	4.3099±0.0001	4.3098±0.0001

Table 4.2: New estimation of period for planet HD 40307b.

	Planet c pipeline	Planet c CF1
Amplitude and error [ms^{-1}]	2.29± 0.13	2.29± 0.13
Nominal period and error [day]	9.6210±0.0008	9.6210±0.0008
Phase at BJD$_{ref}$ (epoch 2018)	0.90041±0.0145	0.82599±0.0055
Zero crossing time T$_2$ at epoch 2018 (bjd-2.4e6)	58465.837±0.1396	58465.837±0.1396
Elapsed time T$_2$-T$_1$ [day]	3915.747±0.88	3915.0232±0.159
K$_{per}$ number of periods in T$_2$-T$_1$	406.85	406.92
K1	406	406
K2	407	407
Period and error for K1 [day]	9.6411±0.0005	9.643±0.0004
Period and error for K2 [day]	9.6174±0.0005	9.619±0.0004

Table 4.3: New estimation of period for planet HD 40307c.

	Planet d pipeline	Planet d CF1
Amplitude and error [ms^{-1}]	2.31± 0.14	2.31± 0.14
Nominal period and error [day]	20.412±0.004	20.412±0.004
Phase at BJD$_{ref}$ (epoch 2018)	0.39676±0.0419	0.33147±0.0555
Zero crossing time T$_2$ at epoch 2018 (bjd-2.4e6)	58466.4	58466.4
Elapsed time T$_2$-T$_1$ [day]	3933.98±1.15	3933.98±1.15
K$_{per}$ number of periods in T$_2$-T$_1$	192.7287	193.07
K1	192	193
K2	193	194
Period and error for K1 [day]	20.489±0.00455	20.4194±0.0059
Period and error for K2 [day]	20.383±0.00457	20.3142±0.0059

Table 4.4: New estimation of period for planet HD 40307d.

A 3-planets best fit to the CF1 applied to CCF$_{tot}$ time series. In this paragraph we want to find out if application of CF1 on CCF$_{tot}$ will produce improvement in precision. Figure 4.5 illustrates the fit by model with 3 planet of data computed by CF1 on CCF$_{tot}$.

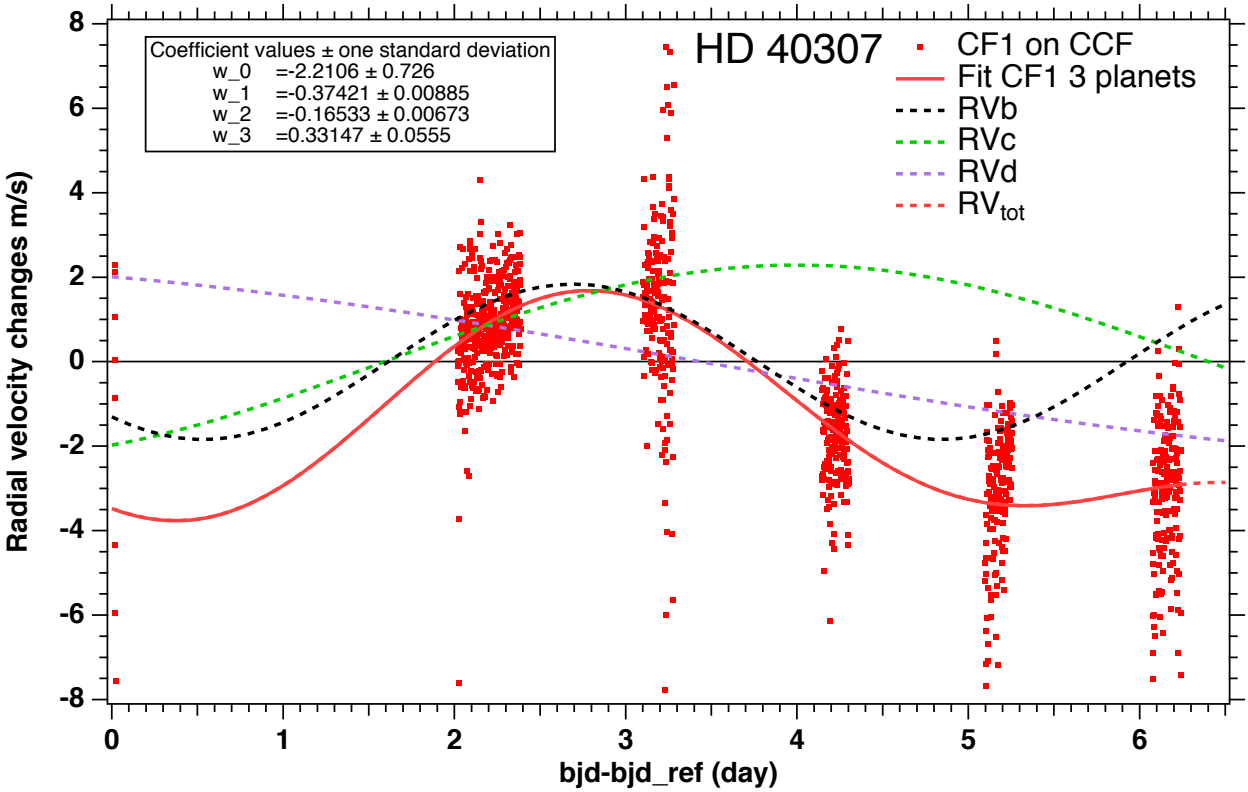


Figure 4.5: Same as figure 4.4, but for the time series obtained with CF1 on CCF_{tot} (taking the first exposure as the reference). The best fit curve is in red (solid line), while the dashed lines are each of the 3 sine waves representing the 3 planets : b (black), c (green), d (violet). The best fit curve is the sum of the three planets plus a constant $w_0 = 2.2106 \pm 0.726$ ms^{-1}.

On figure 4.6 are compared the best fit curves for the 3 planets obtained with the two methods, pipeline and CF1 CCF_{tot} . While there is not much difference for planet b, there is some substantial time shift for planet c.

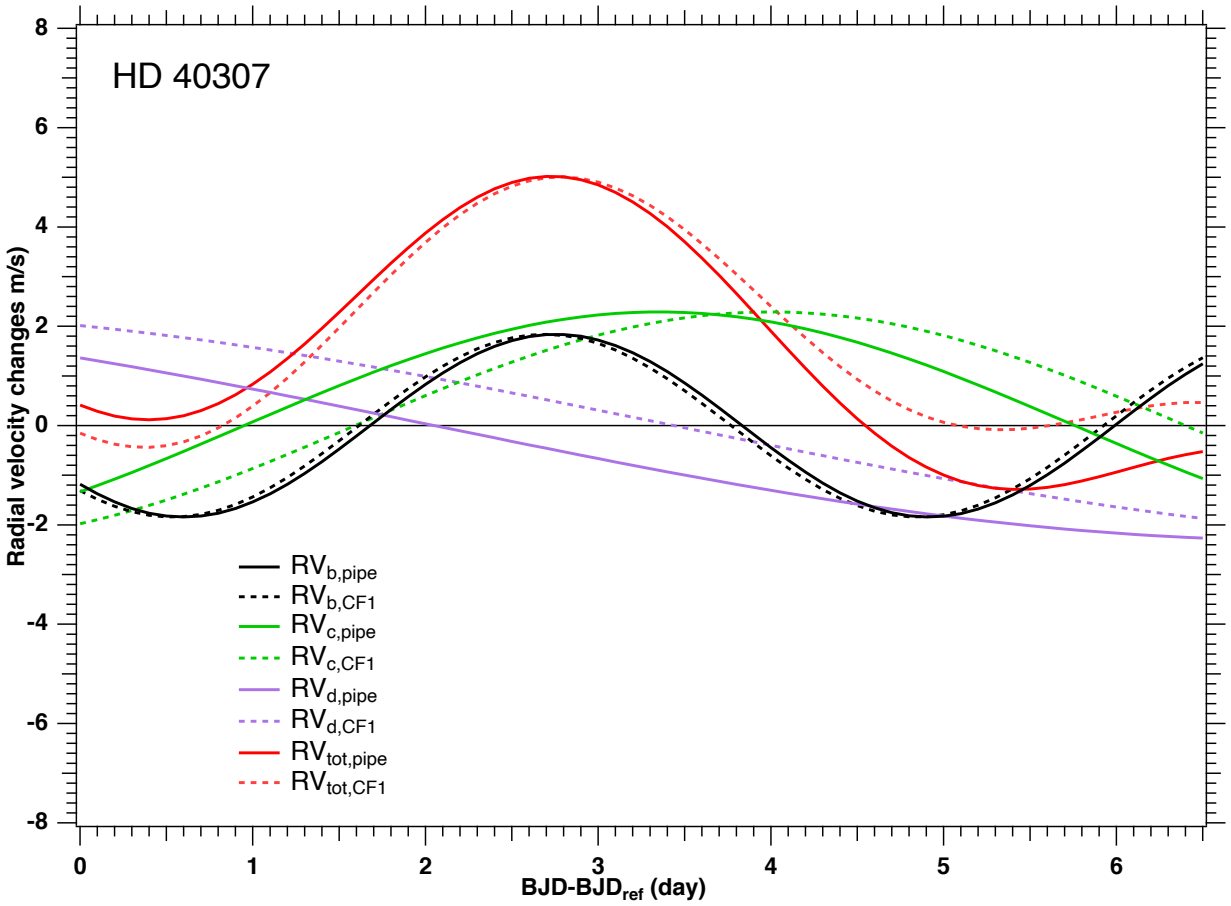


Figure 4.6: Comparison of best fit models of the two time series. Solid lines are for the ESPRESSO pipeline. Dashed lines are for the CF1 applied to CCF$_{tot}$. The red curve is for the total of 3 planets (+ the offset w$_0$), while planets b, c, d have different colors.

The same exercise to compute the period from the elapsed time between two zero crossings (from negative to positive) of the sine wave was done from the best fit data to CF method. The results (also shown in Table 4.2, 4.3, 4.4 for the two methods are not significantly different, and for the CF1 method we find also the same offset with the nominal period found at 2008 epoch. We have summarized in Table 4.5 the new results of the periods of the 3 planets both for the best fit to pipeline data and CF1 data. The results of both methods are similar. We have also indicated the difference of period between the nominal (old) value, and the new value found from the CF1 method. As mentioned above, this difference for planet b is 7.5 times larger than the previously claimed error bar. For planets c and d, the difference is 2.5 and 2 larger than the claimed error bar. In summary, joining the 2008 and 2018 data is providing more accurate periods than those which were determined with 2008 observations only.

	HD 40307b	HD 40307c	HD 40307d
Amplitude and error [ms^{-1}]	1.84± 0.14	2.29± 0.13	2.31± 0.14
Nominal period and error [day]	4.3114±0.0002	9.6210±0.0008	20.412±0.004
New period and error [day] from pipeline data	4.3099 ±0.0001	9.6174 ±0.0005	20.383±0.0045
New period and error [day] from CF1 method	4.30984±0.0001	9.619 ±0.0004	20.4194±0.0006
Difference nominal- CF1 [day]	0.0015	0.002	0.007

Table 4.5: Characteristics of three planets in HD 40307 system from HARPS 2009 campaign. New periods are computed based on pipeline data and on CF1 method results.

Comparison of the residuals to 3-planets best fit between the two methods. On figure 4.7 are represented (black points) are the residuals (δ_{RV} time series after subtraction of the best fit model) for the official RV pipeline of ESPRESSO, and similarly (red points) on figure 4.8 the residuals for the CF1 to CCF$_{tot}$ method. By eye, it seems that the second method is better aligned on the Zero line (solid black line) than for the δ_{RV} time series.

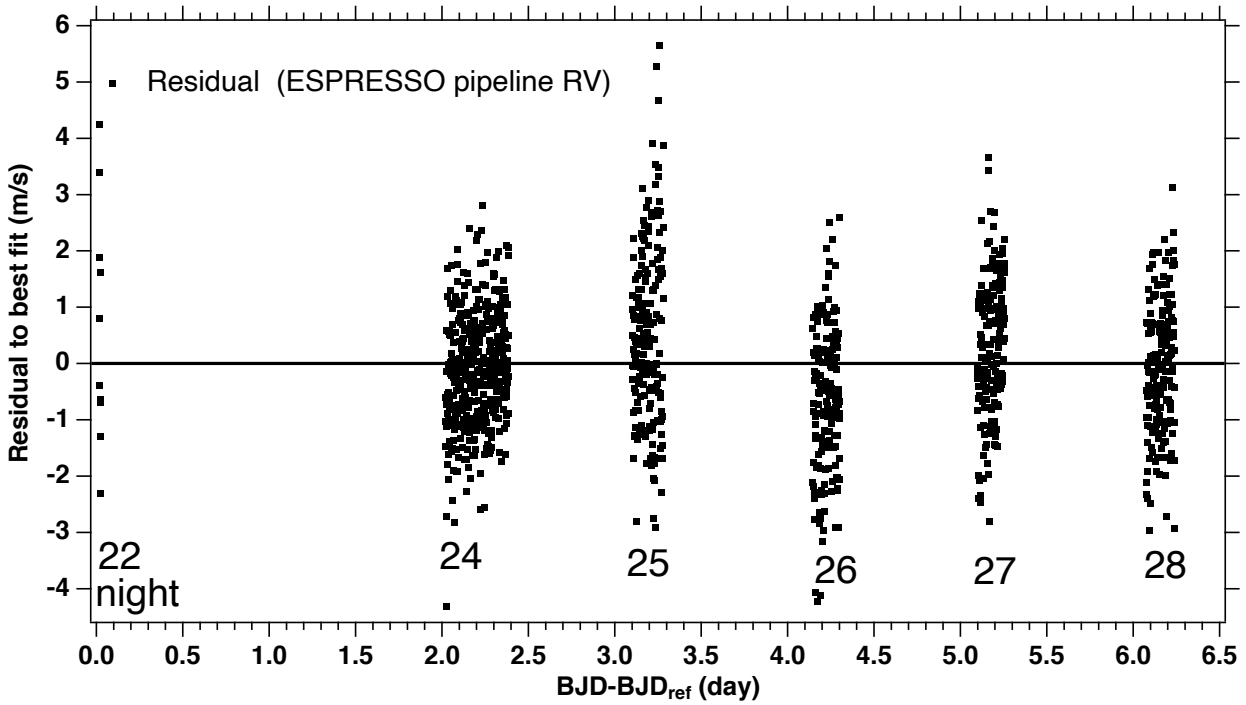


Figure 4.7: Black points are residuals after subtraction of the best fit model from the δ_{RV} official ESPRESSO pipeline. It seems by eye that the average of groups of points do not always lie on the zero line. There is a bias that depends on the day of observation (numbers represent the night of December when was observed HD 40307.)

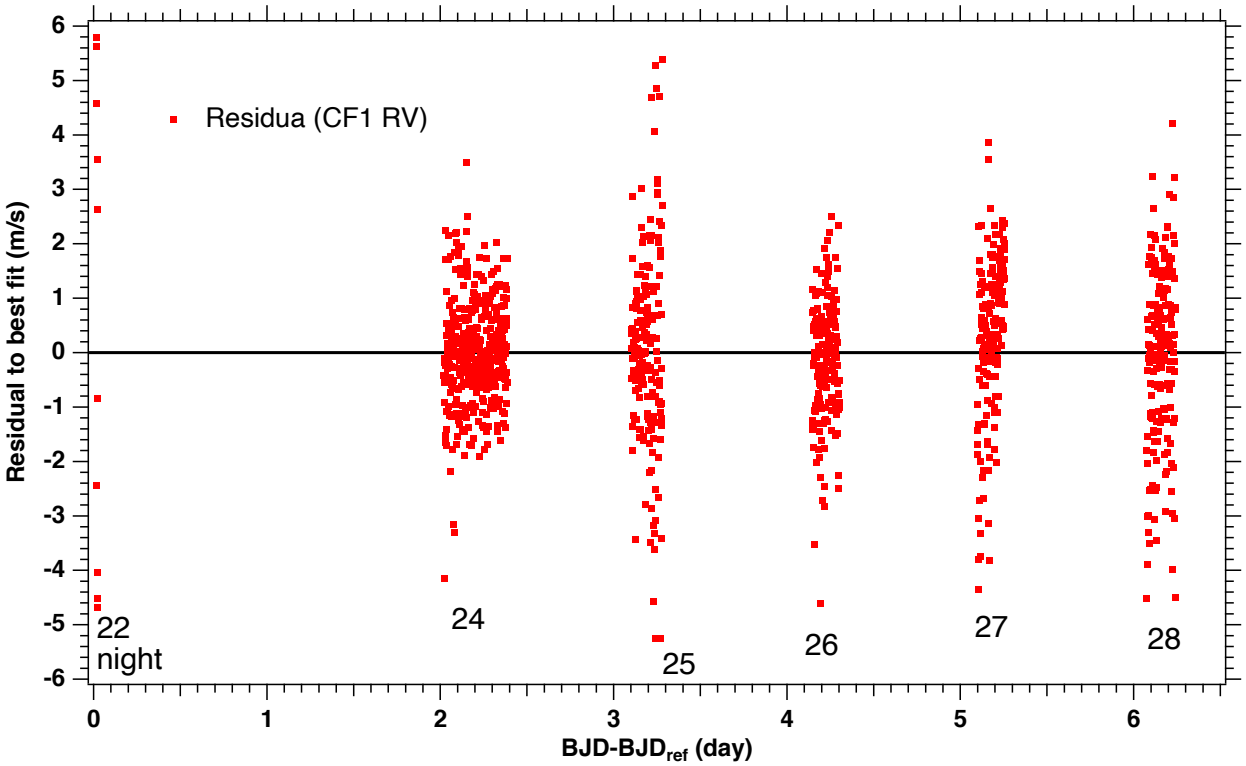


Figure 4.8: Red points are residuals after subtraction of the best fit model from the CF1 to CCF$_{tot}$ time series. It seems by eye better aligned on the Zero line (solid black line) than for the δ_{RV} time series of figure 4.7

This visual impression is confirmed by the following statistical analysis, summarized in Table 4.6 and on figure 4.9. The first night (December 22) contained only 10 exposures, and is not in the Table 4.6. The second night is the longest one and contained 406 exposures. It shows the smallest standard deviation, quite similar for the two methods. (*) The third night (200 exposures) had a specific problem. An instrumental noise due to a blue detector instability spoiled somewhat the second part of the night (visible on all figures with data, residuals, and standard deviation). This is why we computed the standard deviation and mean residual per night for both the first 100 exposures only and for the 200 exposures of the night, as two separate entries of Table 4.6

Date	N^o of exposures, exposure N^o	S$_{dev}$ pipeline	S$_{dev}$ CF1	Residual pipeline	Residual CF1
24	406 [10,415]	1.014	1.03	-0.094	-0.014
25	100 [416,515] (*)	1.13	1.04	0.45	0.098
25	200 [416,615] (*)	1.87	2.26	0.482	0.072
26	174 [616,789]	1.26	1.11	-0.616	-0.147
27	176 [790,965]	1.19	1.91	0.344	0.203
28	185 [966,1150]	1.13	1.68	-0.09	-0.12
mean absolute residual ms^{-1}				0.32	0.11
standard deviation of residual ms^{-1}				0.43	0.14

Table 4.6: Statistics per night on Standard deviation and Mean residual, both in ms^{-1}. Keeping the 200 exposures for night 25, the mean absolute residual (averaged over the 5 nights), and the standard deviation are computed in the two last lines

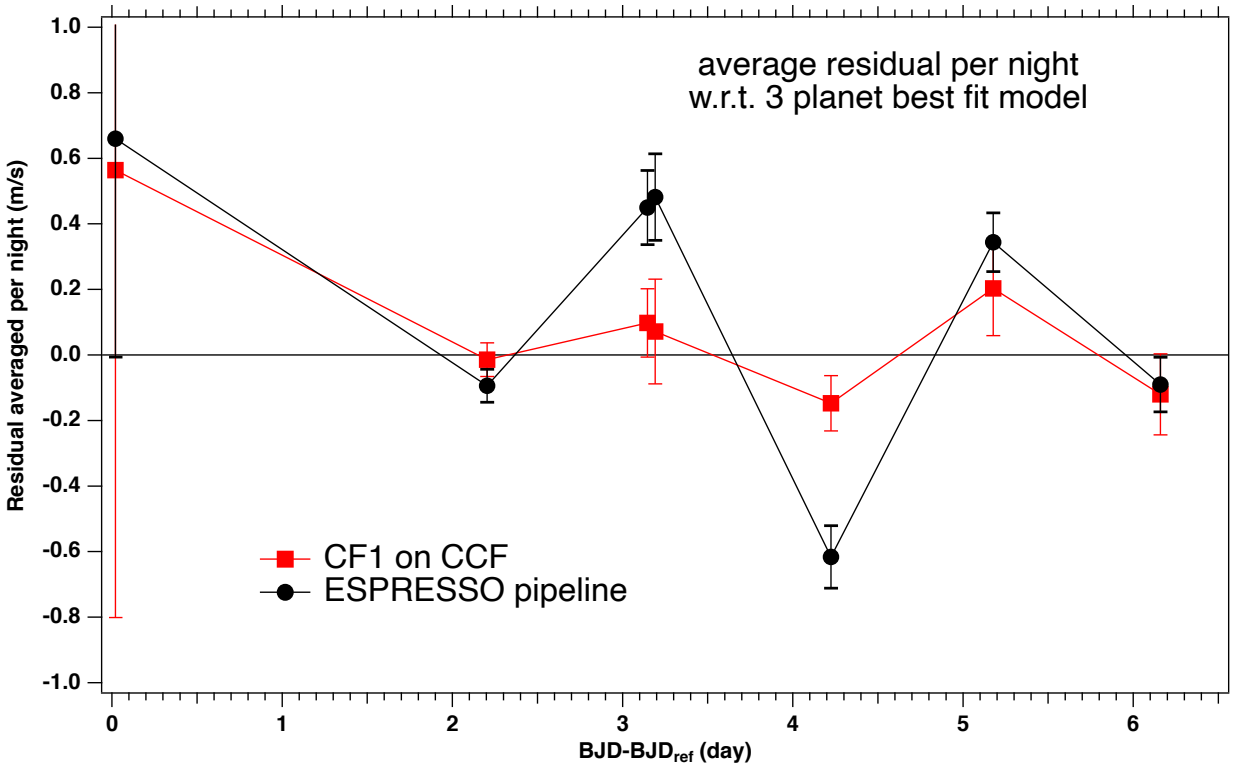


Figure 4.9: Mean residual for each night as a function of time (BJD-BJD$_{ref}$). The error bars are computed as the standard deviation divided by sqrt(number of exposures). The two points for night 25 (around 3.2 days) are not independent.

The mean residual (which can be qualified as variable bias) is much nearer zero for the CF1 on CCF$_{tot}$ method than for the official pipeline.

The mean absolute value over the 5 nights (excluding the first one) is 0.11 ms^{-1} for the CF1 to CCF$_{tot}$ method instead of 0.32 ms^{-1} for the official ESPRESSO pipeline. Similarly, the standard deviation of the mean value is 0.14 ms^{-1} and 0.43 ms^{-1} respectively for the two methods.

This is a measure of the quality of the fit to a 3 planet model : a factor of 3 in favour of the new method. In our mind it is a strong indication that this new method of applying CF1 to CCF$_{tot}$ is giving

a result which is nearer the truth than the official ESPRESSO pipeline , and such a method would deserved to be explored further and used more widely in the analysis of exoplanet RV signals.

It should also be kept in mind that such low values, 0.11 ms^{-1}, or even 0.32 ms^{-1}, are anyway excellent, and reflects the high quality of ESPRESSO data, which is certainly even better after the upgrade of the blue detector cryostat in June 2022. With an average discrepancy of 0.11 ms^{-1} between observations and a dynamical planetary model, we are approaching the precision required to detect an Earth-like planet around a solar-type star. However, these results are covering only 6 days of observations, and this EPiCA method should be applied to much longer periods, in order to explore its domain of validity and the gain that can be achieved.

In the next chapter (chapter 5), we examine a direct product of the EPiCA method: the variation of the RV returned for each point of the CCF_{tot}.

Chapter 5

The EPiCA method as a tool to detect RV changes associated with stellar activity

In this chapter, we wish to discuss the effect of changes of line shapes due to stellar activity (granulation blue shift changes and attenuation by stellar activity) on the retrieved "apparent" RV (not due to planets), and the use of the EPiCA method as a simple tool of diagnostic, to indicate at least that there has been a change of shape between two stellar spectra that are compared. Since CCF_{tot} is an "image", combination of all spectral lines in the stellar spectrum, it must reflect at least a physical effect that is dominating in the average "image", even if it is not present in all the spectral lines. It is a very well known fact that the shape of the lines are changing with time. Dravins et al. (1981), anlyzing high resolution solar lines, used the bisector as a tool to characterize such changes. In essence, we have the same aim as bisector analysis, which have used later by many authors (e.g. Queloz et al. (2001); Baştürk et al. (2011)). The bisector analysis consider only the middle of each segment, while with EPiCA, (most) all points of the CCF_{tot} may be used as a diagnostic.

5.1 Convective granulation blue shift (GBS) and its time variations.

As we already reported in section 2.1, the (unattenuated) convective blue shift depends very much on the particular solar line which is used. Analyzing with HARPS the solar light scattered by the Moon (to get a disc integrated signal), González Hernández et al. (2020) (and references therein) have provided evidence that the convective blue shift is larger for weak lines than for deep lines, represented by a linear correlation with the equivalent width EW of the line (figure 2.3 in 2.1).

For instance, lines with EW = 50 mÅ have a blue shift of about 633-230 $\simeq$ 400 ms^{-1}, while lines with EW $\geqslant$ 200 mÅ have practically no blueshift. This does represent a static situation. One may wonder then how this plot would change when there is some magnetic activity which attenuates the convective blueshift, as described by Meunier (2021) and references therein. If the attenuation, measured in ms^{-1}, were the same for all the lines, then the linear fit would just be translated upwards. But then it would be impossible to discriminate the attenuation of convective blueshift from a genuine Doppler effect linked to an actual change of dynamical RV, dR/dt. Fortunately, the attenuation/inhibition of convective blue shift is also a function of EW, or depth of line, as shown by Cretignier et al. (2020). These authors have analyzed a series of 47 HARPS RV measurements of star α Cen B over 81 days (see Figure 5.1. The rotation period of the star is 35.8 days. The calcium activity index $\log(R^{`}_{HK})$ follows the rotation of the star, with one side of the star being much more active than the other. The retrieved RV is at a minimum of $\simeq$ -4 ms^{-1} when the inactive face of the star is observed, and goes up to $\simeq$ +4-5 ms^{-1} when the active face is observed. This is the so-called attenuation of the convective blue shift by stellar activity (Meunier, 2021).

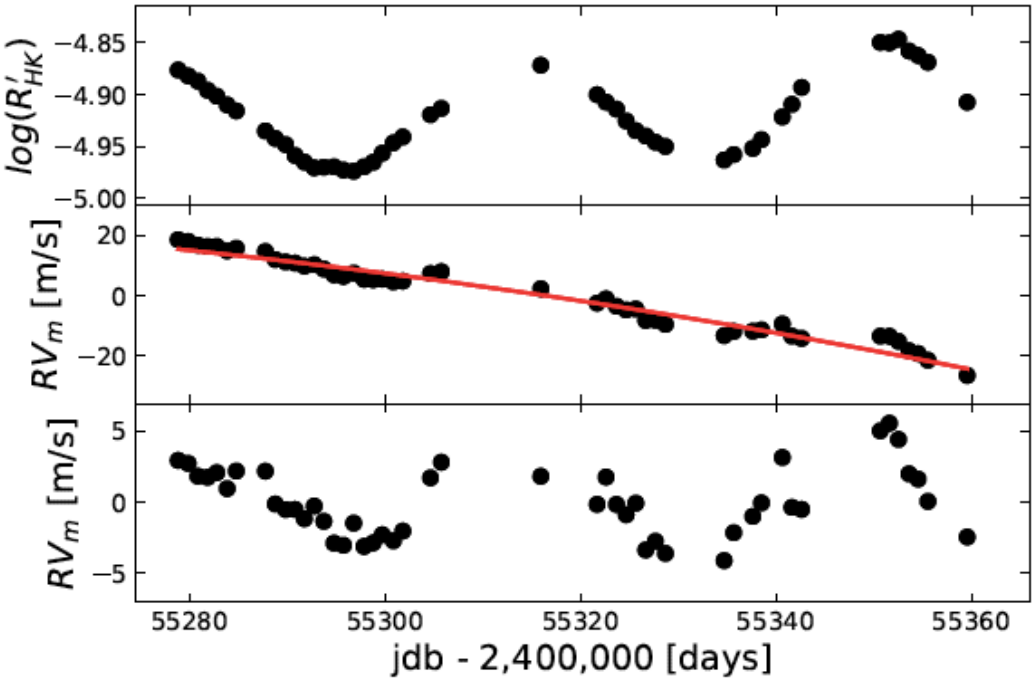


Fig. 1. *Top*: calcium activity index $\log(R'_{\rm HK})$ time series of α Cen B in 2010. The observations span two consecutive rotational phases, where $P_{\rm rot}$ = 35.8 days. *Middle*: RV time series over the same time interval. We show a quadratic trend fitted to the data (red) that corresponds to the binary contribution of α Cen A and a possible activity cycle (Dumusque 2012). *Bottom*: RV residuals after removing the quadratic trend, which present an RV rms of 2.1 m s^{-1}. These RV residuals are strongly correlated to the calcium activity index and therefore are significantly affected by stellar activity.

Figure 5.1: Figure and caption from Cretignier et al. (2020).

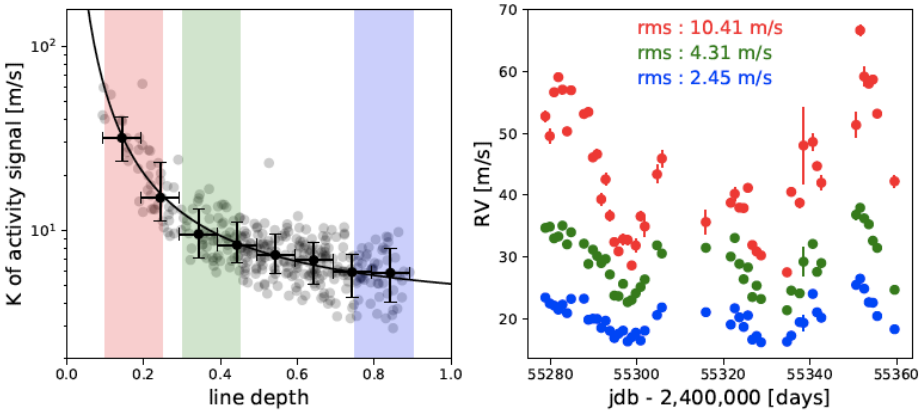


Fig. 8. *Left*: radial velocity semi-amplitude of the stellar activity signal as a function of the line depth d for the correlated symmetric-unblended lines (grey dots). The data are binned into nine bins of size 0.1 (black dots), and the best fit for a second-order polynomial in $1/d$ is shown by the black line. *Right*: radial velocity of the subselections of lines falling in the area of the left plot panel. The RV amplitude grows from strong (blue) to shallow (red) lines.

Figure 5.2: Figure 8 and caption from Cretignier et al. (2020). On the left plot of RV semi-amplitude generated by stellar activity as a function of the line depth, only the lines which are symmetric and not containing a parasitic blending line on one side or the other. On the right plot, we read that for line depths of 10-25% of the continuum (pink vertical band on the left), the time variation of the line shifts amounts to 30 to 60 ms^{-1}, with a rms (rather a standard deviation) of 10.41 ms^{-1}.

The authors then perform a line-by-line analysis of RV variations due to stellar activity, and show (their figure 8 reproduced here as Fig. 5.2 that the weaker is the line depth, the larger is the amplitude of the RV changes (inverse law). The depth of the line is linked to the EW. Therefore, linking the figure 5 of González Hernández et al. (2020) and this figure of Cretignier et al. (2020), we may state that the lines which have the largest average convective GBS are also the lines which RV are changing the most, a result which is not totally surprising after all.

This differential effect of activity offers the possibility to make some corrections to the RV (spurious) signal, a mitigation. We quote Cretignier et al. (2020): "Finally, we investigated a few simple possibilities to mitigate stellar activity based on the relation between strength of the activity effect and line depth. By splitting the correlated lines into two groups, those shallower and stronger than a threshold d, we were able to mitigate stellar activity from a level of 2.17 ms^{-1} down to 0.84 ms^{-1}."

5.2 EPiCA method (applied to CCF_{tot}) and examination of DRV as a function of RV grid.

In the standard use of ESPRESSO results (or from other spectrometers), to each exposure/spectrum is assigned an absolute radial velocity RV in the pipeline. Therefore, the simplest way to examine changes of RV potentially indicating the presence of an exo-planet is to analyse the RV time series. In order to improve the precision of RV changes, and achieve a pRV (precise RV), a variety of analysis schemes have been proposed in the past. One of them (Artigau et al., 2022) is quite recent, and deserves to be described here, to examine the similarities and differences with our EPiCA scheme. Artigau et al. (2022) propose a LBL (Line-by-Line) analysis of a time series of spectra obtained on two stars (Proxima Centauri with HARPS and Barnard with SPIRou). For SPIRou, they correct from telluric absorption with their software APERO, initialized with TAPAS. Then they build a template spectrum by using the median of many spectra, with wavelength scale corrected from the BERV (therefore, as if the star was observed from the barycenter of the solar system). Then, separately for each of the lines (or "domaine" between two maxima), they apply the Connes algorithm between a current spectrum and the template. One estimate DRV_i of RV and its uncertainty σ_i is determined for each line i. This allows to eliminate outliers, roughly defined as a threshold on $\frac{DRV_i}{\sigma_i}$, for instance eliminating all lines yielding with $\frac{DRV_i}{\sigma_i} \geqslant 5$. Then an average DRV is obtained by the optimal combination of all kept DRV_i with weights $\frac{1}{\sigma_i^2}$, as usual.

In the basic version of EPiCA algorithm applied to CCF_{tot}, such a filtering of outliers is not possible, because CCF_{tot} comes from the pipeline. As we have said, the CCF_{tot} is an average image of all the lines which are used to determine it.

On the other hand, the EPiCA algorithm gives a detailed access to the variation of DRV along the CCF_{tot} RV grid. This would not be possible with the LBL method applied to one single exposure, because this requires a very high S/N which in most cases is not achieved with individual lines of individual exposures, but is provided by CCF_{tot} by cumulating many lines together. And, as we shall see further, the examination of the DRV curve vs CCF_{tot} RV grid gives immediately a diagnostic about the change (or not) of the shape of the lines: if DRV is constant for the whole CCF_{tot}, it means that on average the line shape has not changed between the two epochs of the two compared spectra. We will see that DRV is showing strong variations along the CCF_{tot}, indicating that there is indeed a change

of shape of the lines between two epochs, likely as a consequence of stellar changes in the convective blue shift pattern, and/or its attenuation by stellar activity. In such a case, we are in a rather awkward situation: we use an algorithm which is valid only if there is no change of shape, and the result of the algorithm tells you that there is indeed a change of shape! We note that the alternate algorithm consisting of determining RV from the center of a symmetric Gaussian may have the same problem. This is generally discussed with a close examination of the bisector (Queloz et al., 2001; Baştürk et al., 2011). Some symmetric deformation may be considered as an indicator of stellar activity (Artigau et al., 2022). But, as we shall see later, the change of shape actually observed may be actually very asymmetric between the red side and the blue side of CCF_{tot}, and changing from one day to the next. As we have said before, if there is no change in shape, the DRV estimate should be uniform across the RV grid. If it is not uniform, then it means that the line shape has changed between the two epochs of the two spectra generating the two CCF_{tot}.

5.3 Simulations of EPiCA method in the case of a pure Doppler shift

We have simulated what we should see, for the curve DRV as a function of RV grid, with the EPiCA method. We have used order 104, and stacked 100 spectra together to increase the S/N ratio. We computed the CCF for this order (with classical CCF algorithm, without any weights) on the nominal RV grid, with a sampling of 0.5 kms^{-1}. Then we computed another spectrum, shifted artificially by 100 ms^{-1} (actually stretched), and then resampled on the same wavelength grid as the original spectrum. We recomputed the CCF on this shifted spectrum on the same RV grid. The two CCFs are displayed on Figure 5.3. The first formula of Connes is applied to the two CCF, providing an estimate of the Doppler shift for each point of the RV grid and its uncertainty bar.

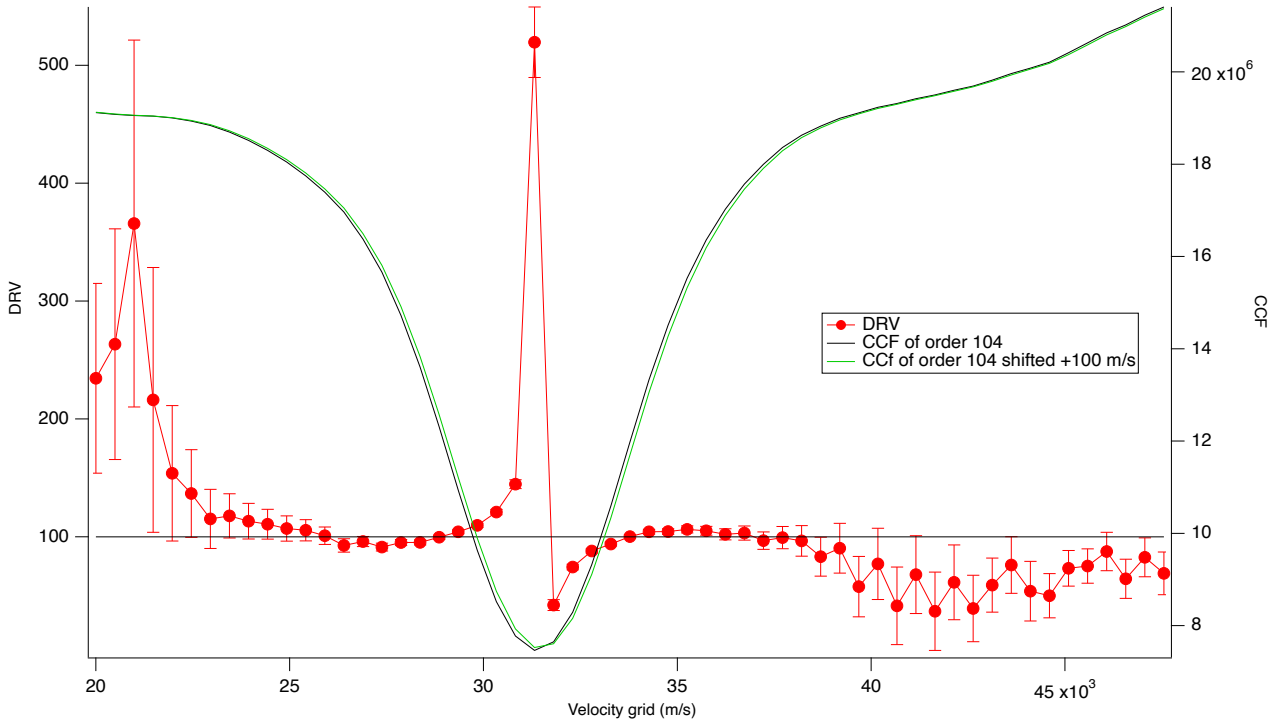


Figure 5.3: The EPiCA method applied point by point along the CCF grid to the simulated data: case of two identical CCFs shifted by 100 ms^{-1}

As expected, the values of DRV are about uniform along the RV grid, within the error bars. However, there are deviations from uniformity near the center of the CCF, with very large values over two points, with deviations of uniformity of opposite sign, and a smooth symmetric pattern when approaching the center of the CCF: increasing on the blue side, decreasing on the red side. We interpret these behaviours as follows:

- the change of sign near the center is due to the change of sign of the derivative of the CCF.

- the large deviations near the center are likely due to the fact that the derivative of the CCF which is at the denominator, becomes very small. It is is not computed very accurately, because it comes from a finite difference algorithm, and the RV grid, with 500 ms^{-1} between two points in rather crude.

- the smooth symmetric pattern deviating from uniformity comes from the fact that we related DRV=Δy_c and shift dS=Δx by the 1st formula of Connes (equation 2.15) , which is also: -Δy_c = $y'_0 \Delta x$. Actually, it considers only the first term of the Taylor expansion of function y(x) around y_0. Adding the second term of the Taylor expansion will translate into:

$$-\Delta y_c = \Delta y = y'_0 \Delta x + \frac{y''_0 \Delta x^2}{2} \tag{5.1}$$

Indeed, we note that the deviation seems to be 0 at the points where the CCF curve has an inflexion

point, where the second derivative is 0. It is possible, and may be worthwhile, to use the Taylor expansion second term of (equation 5.1), and extract the shift Δx from this second degree equation. However, its usefulness will be limited by noise on the second derivative.

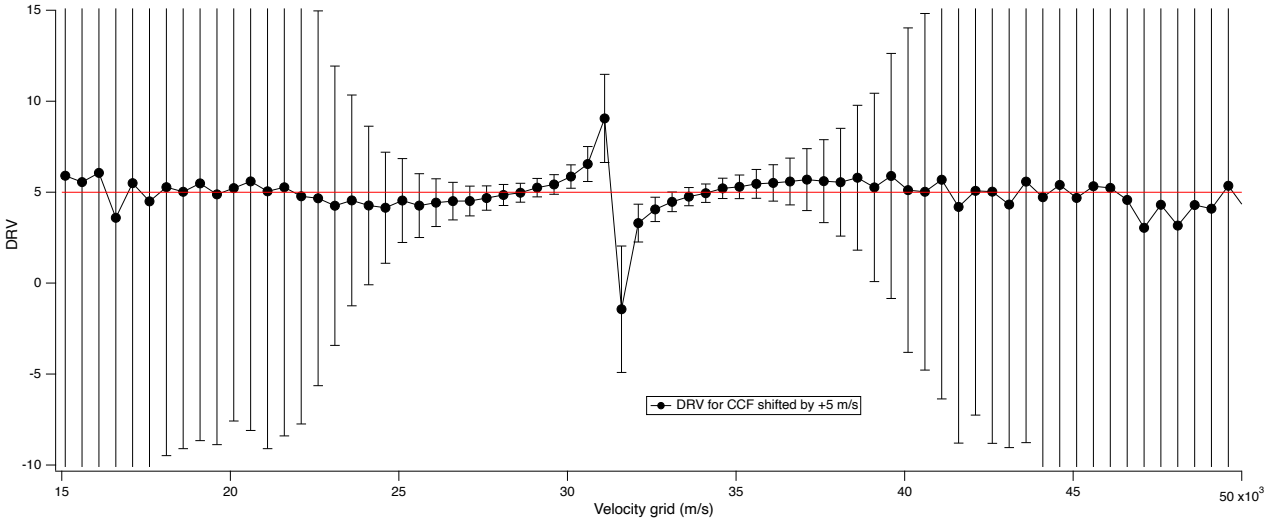


Figure 5.4: The EPICA method applied point by point along the CCF grid to the simulated data: case of two identical CCFs shifted by +5 ms^{-1}

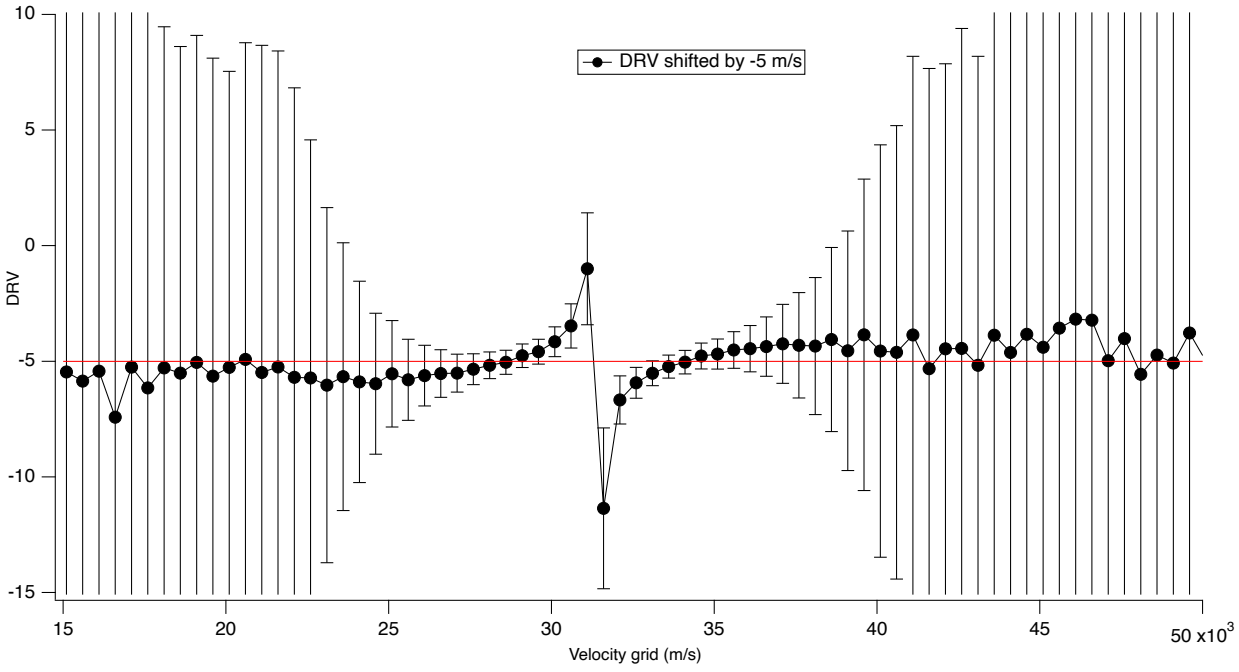


Figure 5.5: Same as figure 5.4 for two identical CCFs shifted by -5 ms^{-1}

We made also other simulations, by displacing directly the CCF by + or – 5 ms^{-1}, and computed the DRV with EPiCA method: we see (Figures 5.4 and 5.5) similar results to the simulation at + 100 ms^{-1} near the center, and a similar trend on both sides as with the 100 ms^{-1} simulation, but well attenuated (within the error bars), for two reasons: - the linear interpolation imposed by the re-

sampling is much less severe because Δx is 20 times smaller

- Δx^2 is 400 times smaller, and the second term of Taylor expansion becomes completely negligible.

We may conclude from this simulation exercise that the DRV values are indeed almost uniform in the case of a pure Doppler shift, except near the center, especially over 2 or 3 points. Also, out of the central region, there is a small deviation, symmetric about the center, but of opposite sign. Finally, the biases (deviations from uniformity) are still present for small shifts, but very much attenuated when shifts are small (5 ms^{-1} or smaller). These two caveats will be important when looking at the results of EPiCA on real data.

5.4 The EPiCA method applied to the HD 40307 24-28 December 2018 series

As said in Chapter 4, the star HD40307 has 3 known planets, therefore the RV of the star (in the Solar System Barycentric system) is changing accordingly. Figure 4.5 displays the measured changes of RV as a function of time during the one week observation campaign of 2018 and we reproduce it also here for convenience of the reader (Figure 5.6, because it is important to look at this figure when interpreting the DRV(i) results obtained by EPiCA computation when comparing two CCF, where i is the grid point of the RV grid on which are computed the CCF.

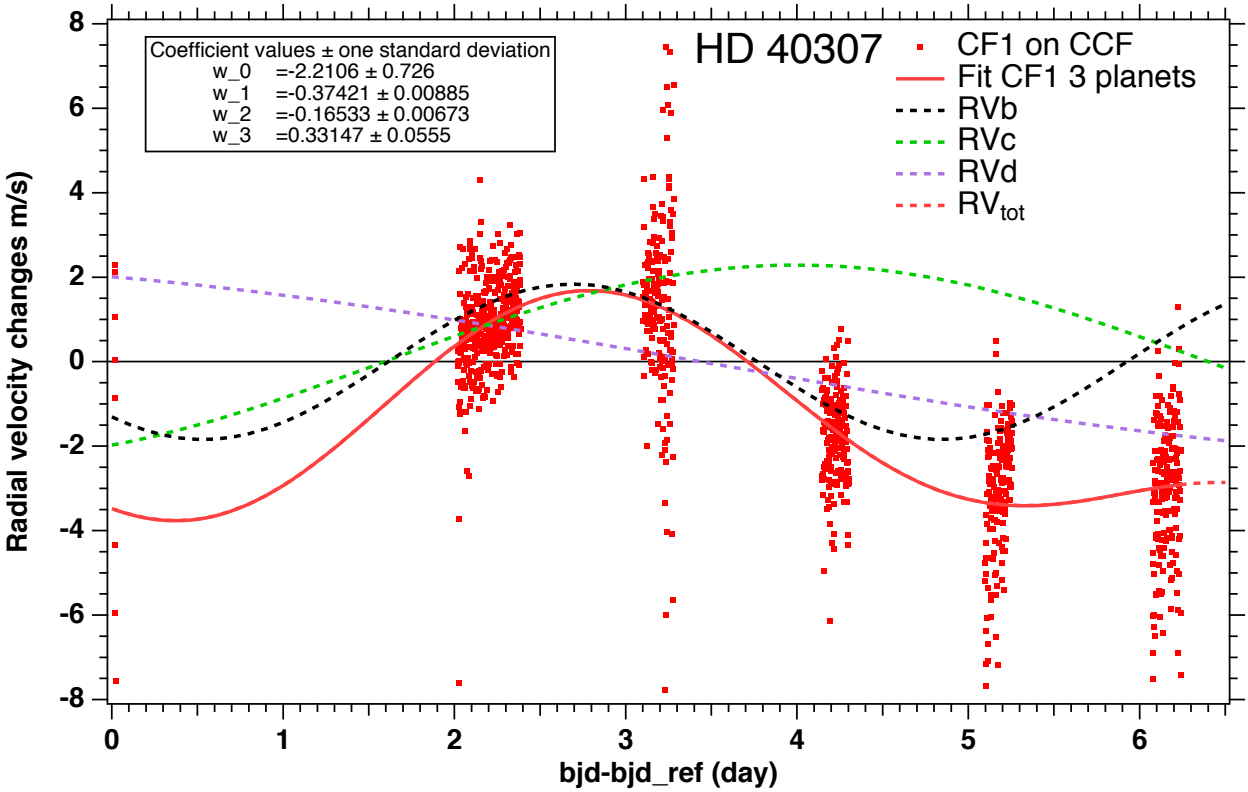


Figure 5.6: Duplicate of Figure 4.5: Time series obtained with CF1 on CCF_{tot} (taking the first exposure as the reference). The best fit curve is in red (solid line), while the dashed lines are each of the 3 sine waves representing the 3 planets : b (black), c (green), d (violet). The best fit curve is the sum of the three planets plus a constant w_0= 2.2106 $\pm$ 0.726 ms^{-1}.

Night 24 contains 406 exposures, and there are 200 exposures for night 25; and Night 26 has 174 exposures; Night 27 has exposures 176; and Night 28 has 185 exposures. Night 25 is somewhat affected by some instability of the CCD blue detector of ESPRESSO in the second half of the series of 200 exposures.

Taking as a reference the stack of exposures 0-99 of night 24, we have computed with the EPiCA method the DRV(i) for the three remaining stacks of 100 exposures of night 24. The three DRV(i) curves are plotted on figure 5.7, with a zoom on the central part, the weight function is also shown. The 2 to 4 points near the center are likely suffering from the division by the very small derivative, as shown in the simulations of the previous section, and we discard them form the discussion.

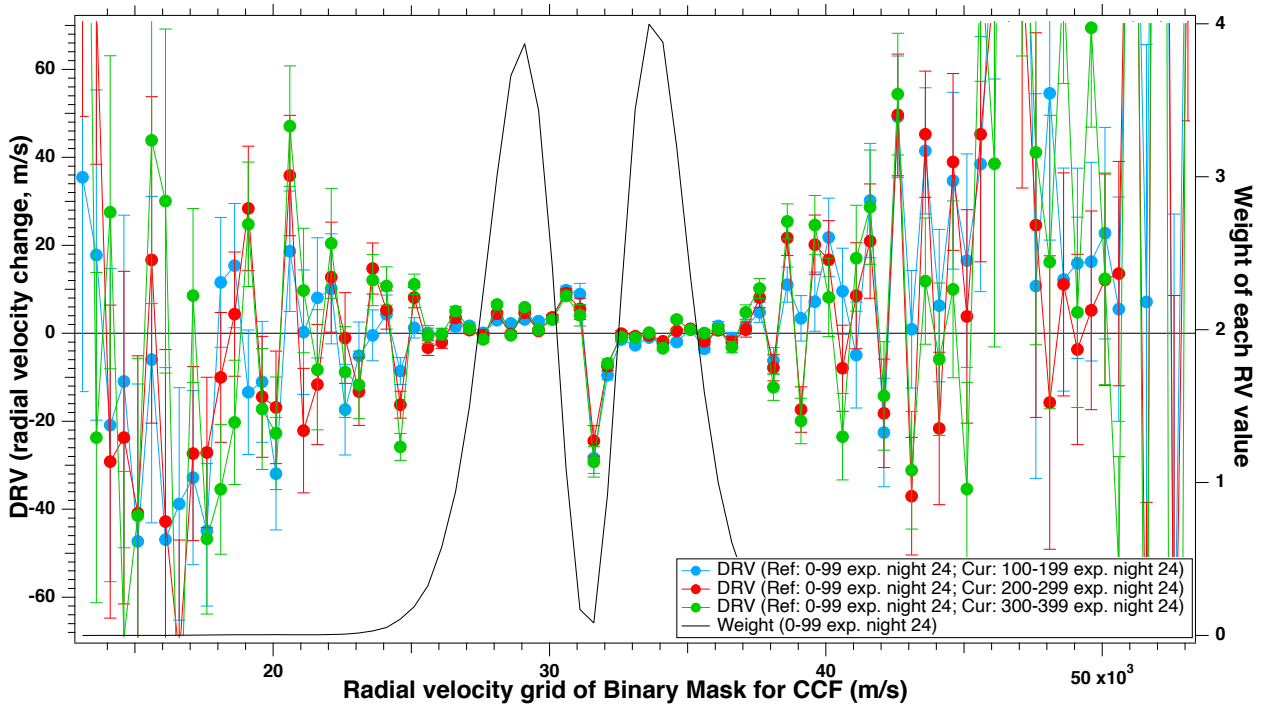


Figure 5.7: Results of the EPiCA method (DRV), shift in velocity space. Reference - first 100 exposures of night 24 December; Data - 3 sets of 100 consecutive exposures. The pattern of velocity change is stable, we can note that on the left part (between 27 kms^{-1} and 30 kms^{-1}) is slightly above zero, while the right part (between 32.5 kms^{-1} and 35.5 kms^{-1}) is almost at zero or slightly below.

The three curves of night 24 (ref. 1^{st} 100 exp from night 24) are rather similar. The main surprise is the difference of average level between the blue side of the CCF_{tot} (RV of BM < 31.5 kms^{-1}) and the red side of the CCF_{tot} (RV of BM > 31.5 kms^{-1}), while the simulation show the same level on both sides of the center.

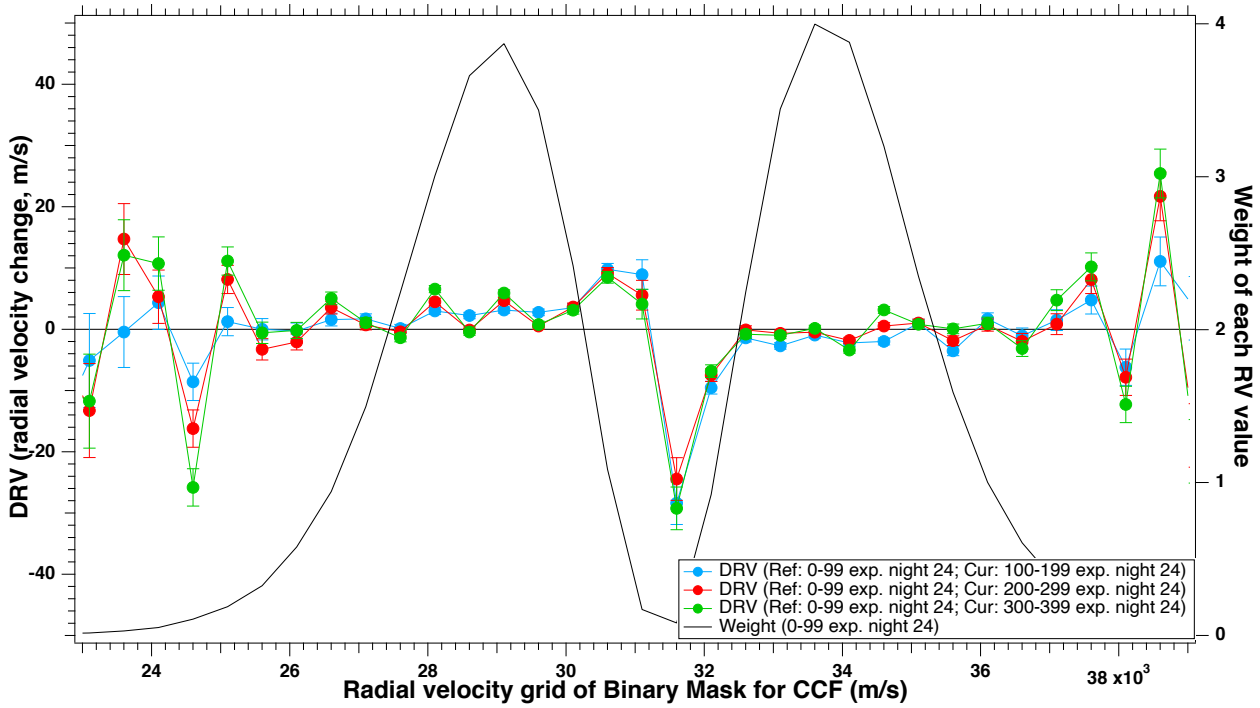


Figure 5.8: Same as Figure 5.7, zoom into region from 23 to 39 kms^{-1}.

Also, while DRV(i) seems rather uniform on the red side around 0 (where the error bar is small), there is a slight increasing trend of DRV(i) on the blue side with increasing i, and also the average is > 0. These features are suggesting that the CCF$_{tot}$ has not only made a pure Doppler shift, but that its shape has also changed. Overall, along the night of 24 December, the EPiCA analysis showed that the RV of the star is increasing along the night (figure 5.6), as predicted by the planetary motion and the whole results of RV gaussian fit or EPiCA analysis point by point. The surprising feature is that the blue side and the red side of the CCF do not indicate the same shifts. The shifts of stacked CCF is larger for the blue side than for the red side. This change of shape of the CCF$_{tot}$ on such a short time scale is likely linked to stochastic fluctuations in the granulation pattern.

Then we plot on figure 5.9 the two DRV(i) curves for night 25, taking as a reference the second packet of 100 exposures of night 24 (rather than the 1st one, because apparently there was already a change of shape between the 1st and 2nd packet of 100 exposures). Both curves are quite similar; this time (still discarding 3 points near the center), the DRV(i) blue side and red side are at about the same level, and are also uniform versus i. This suggests that there has been no significant change of shape between the end of night 24 and night 25. The overall level is near 0, consistent with the RV behaviour for all exposures displayed on figure 5.9.

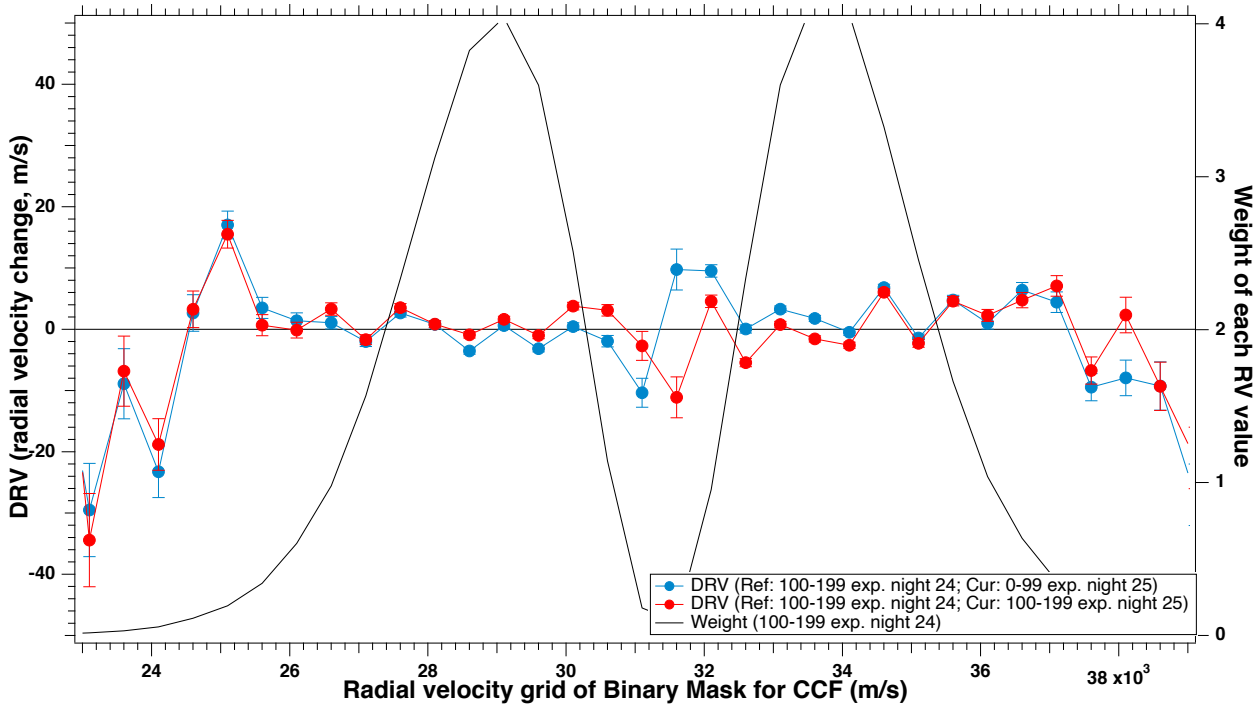


Figure 5.9: Results of the EPiCA method (DRV), shift in velocity space. Reference - second 100 exposures of night 24 December; Data - 2 sets of 100 consecutive exposures of night 25 December. We can see that pattern of the 2 curves is similar, almost identical, but curve which corresponds to the second part of the night 25 December is slightly closer to the 0 line.

As shown on figure 5.10, the two DRV(i) curves for night 26 show a quite different scheme than for the night 25. They are similar between them, but they are not uniform at all, showing a pronounced descending trend on both blue and red sides, with somewhat different average values. The resulting RV coming from this average is consistent with the drop of RV noticed between night 25 and 26, due to the effects of planets. Most interestingly, this peculiar behaviour between night 25 and 25 corresponds exactly to the abrupt change of the difference between the two time series (drop by $\simeq 1.3$ ms^{-1}), the classical Gaussian fit and the EPiCA method, as displayed on figure 4.2.

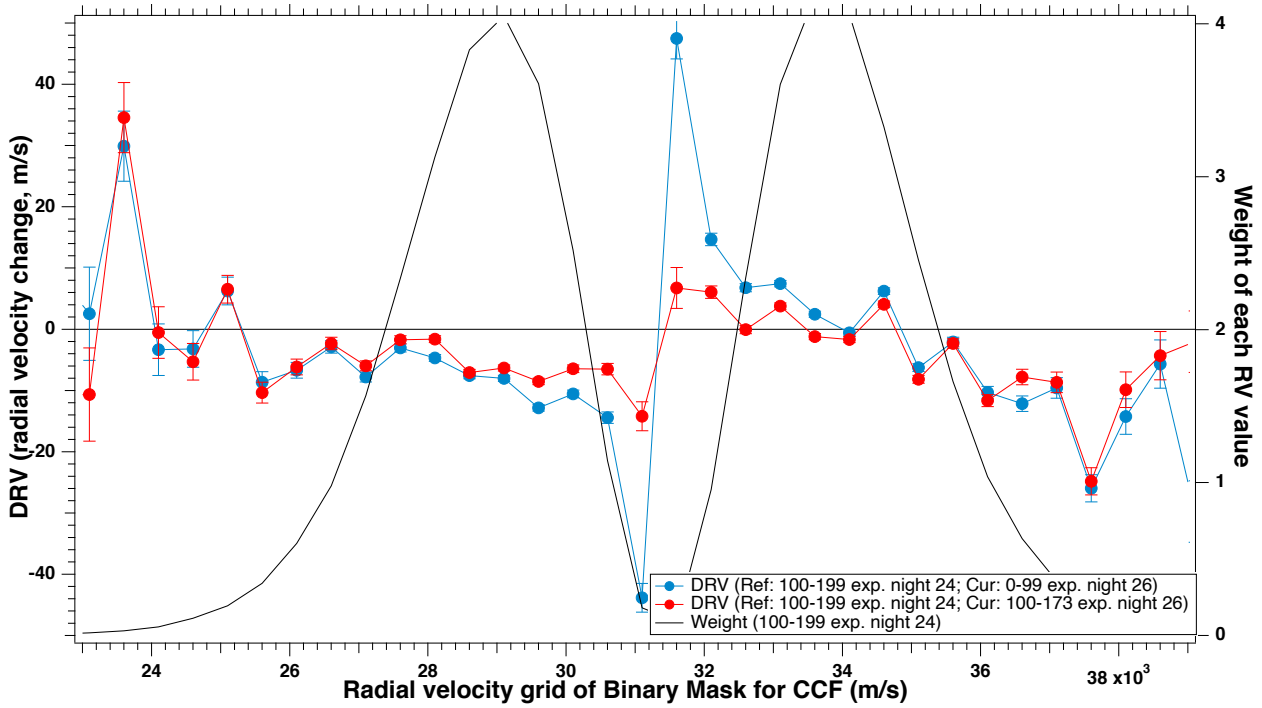


Figure 5.10: Results of the EPiCA method (DRV), shift in velocity space. Reference - second 100 exposures of night 24 December; Data - 2 sets of consecutive exposures of night 26 December (0-99 and 100-173). On this figure we can note the change of the pattern from the one we see in the figure 5.9. Now one can note the presence of a pronounced slope in the curve for the left part and an even bigger slope on the right side. The linear fit shows that the slope is twice times stronger in the red part, than in the left for both curves; And slope for the blue curve is stronger than for the red curve.

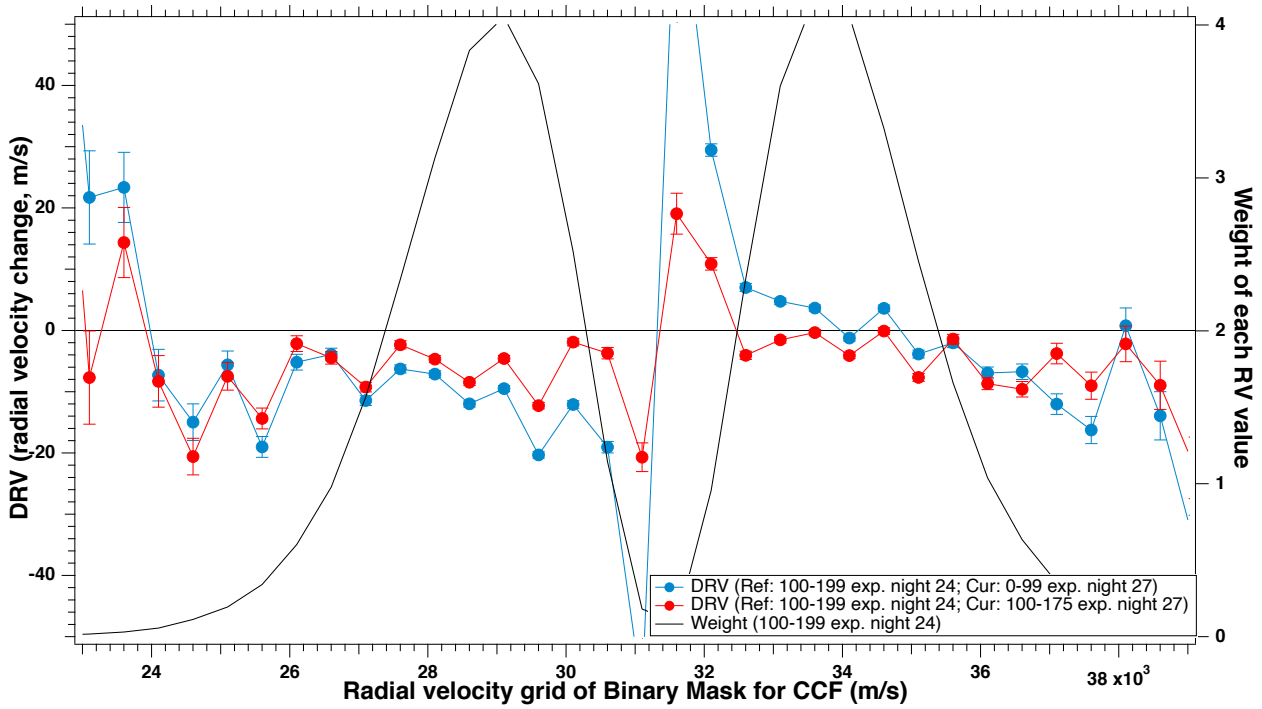


Figure 5.11: Results of the EPiCA method (DRV), shift in velocity space. Reference - second 100 exposures of night 24 December; Data - 2 sets of consecutive exposures of night 27 December (0-99 and 100-175). The slope which was first time noted in figure 5.10 remains. But the slope is less pronounced for the red curve (second half of the night), than for the blue (first half of the night).

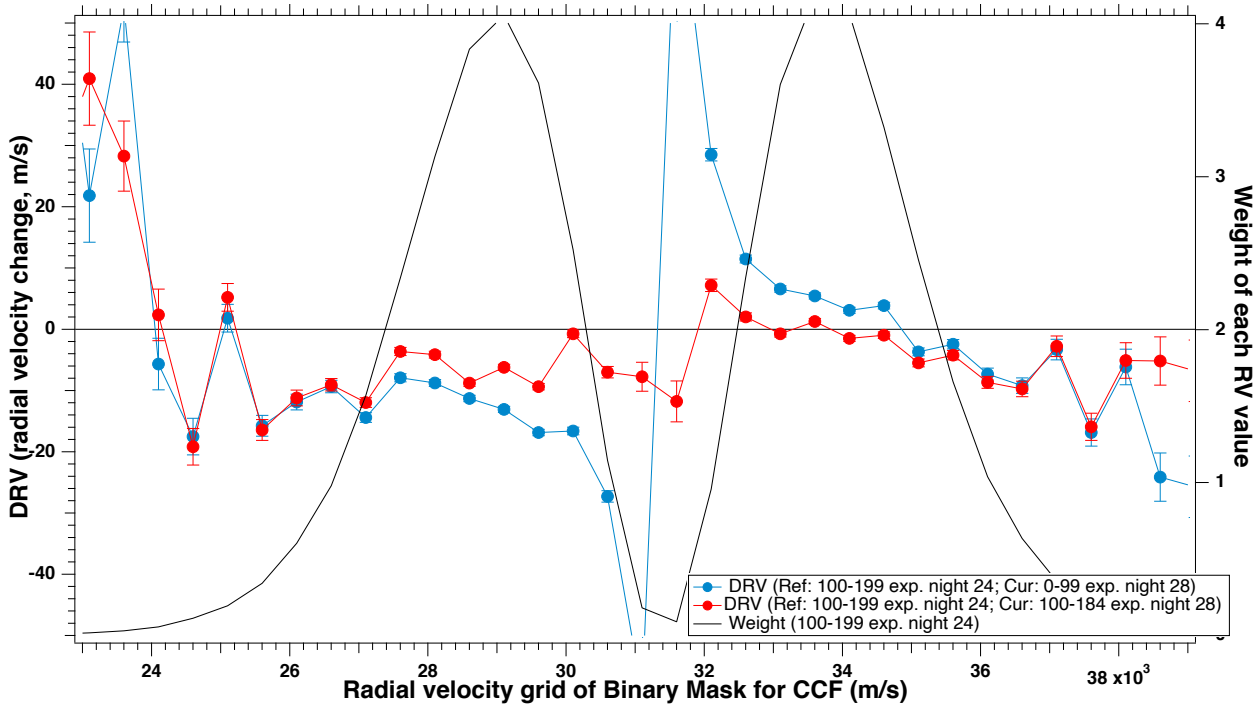


Figure 5.12: Results of the EPiCA method (DRV), shift in velocity space. Reference - second 100 exposures of night 24 December; Data - 2 sets of consecutive exposures of night 28 December (0-99 and 100-184). The difference between the first half of the night and the second half of the night is noticeable for the left part, while the first part of the night remains the negative slope, second half of the night has a small positive slope.

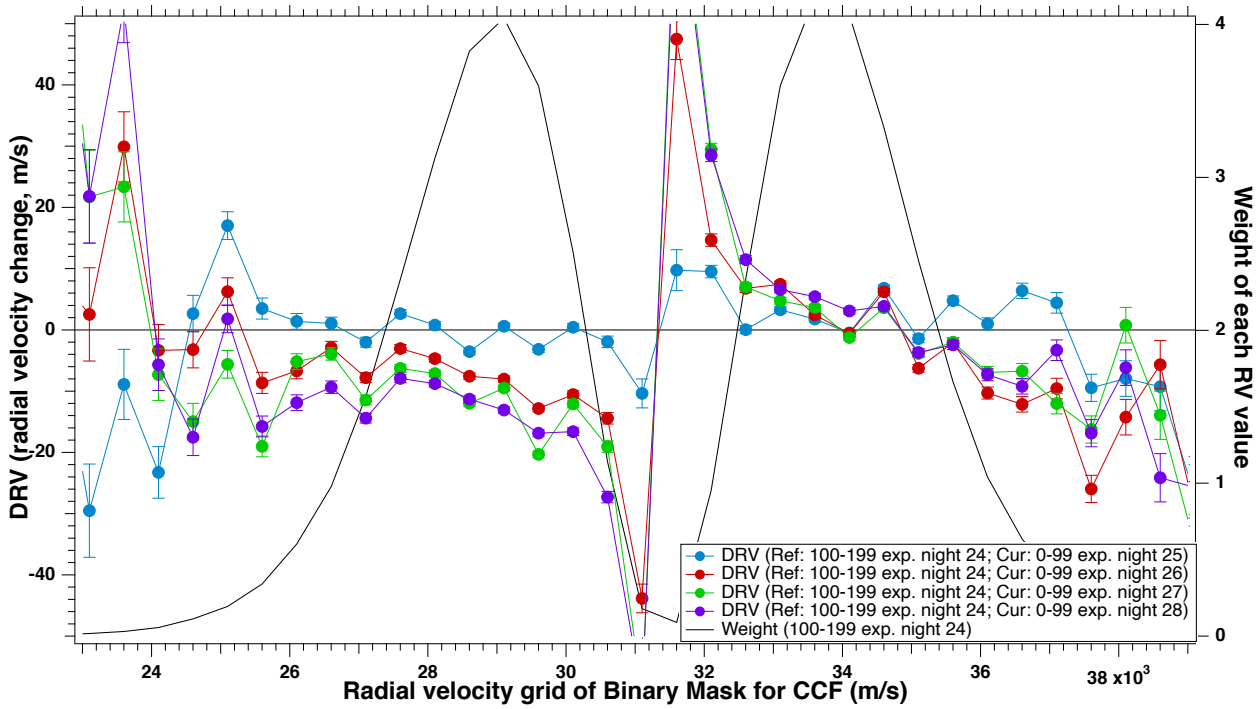


Figure 5.13: Results of the EPiCA method (DRV), shift in velocity space. Reference - second 100 exposures of night 24 December; Data - 4 sets of second half of the exposures of nights 25, 26, 27 and 28 December.

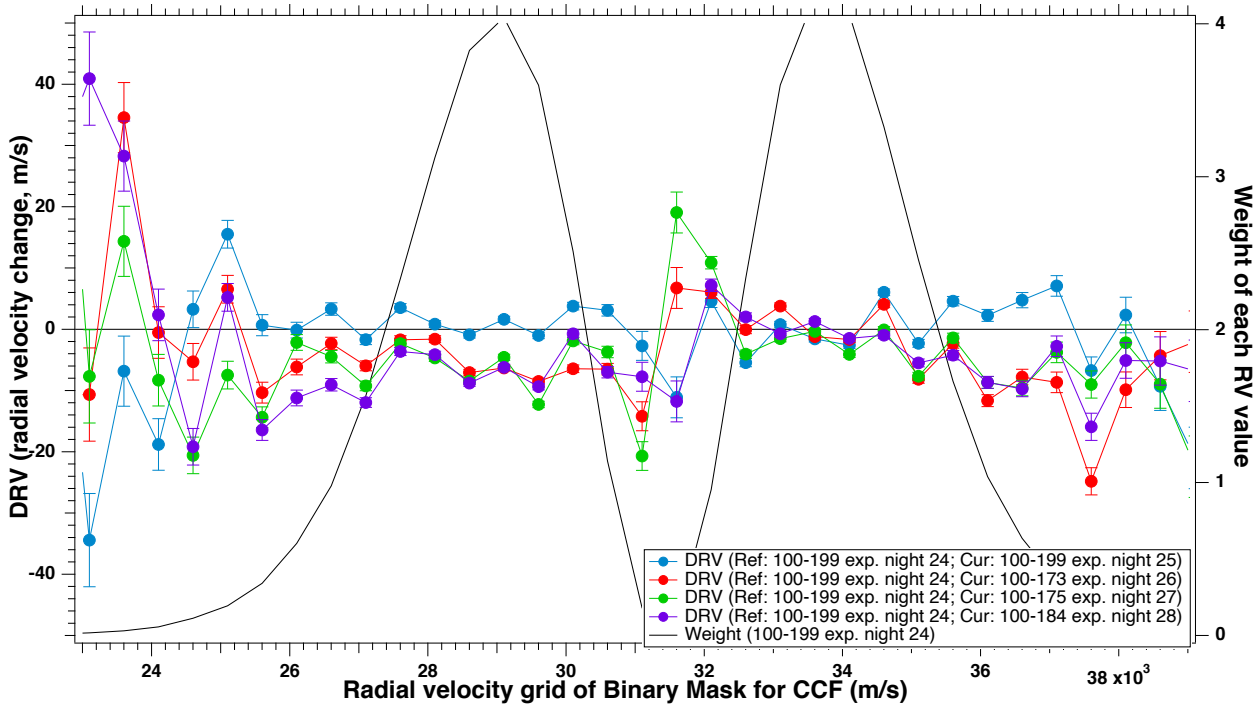


Figure 5.14: Results of the EPiCA method (DRV), shift in velocity space. Reference - second 100 exposures of night 24 December; Data - 4 sets of second half of the exposures of nights 25, 26, 27 and 28 December.

The two DRV(i) curves for each of nights 27 and 28 are shown respectively in figures 5.11 and 5.12, taking the same reference CCF_{tot} as for nights 25 and 26. The shapes are similar to the night 26 DRV(i) curve.

The four curves for nights 25, 26, 27, 28 are plotted together on figures 5.13 (first half of all nights) and 5.14 (second half of all nights) for comparison between nights. On both figures, it is seen that, while the DRV(i) curve is about flat for night 25 on both blue and red side (taking the reference during night 24), for the three other nights, the DRV(i) curve is about uniform on the blue side, but has a significant slope on the red side of CCF_{tot}.

As said before, since CCF_{tot} is an "image", combination of all spectral lines in the stellar spectrum, it must reflect at least a physical effect that is dominating in the average "image", even if it is not present in all the lines. Therefore, the DRV(i) behaviour tells us that the star spectrum is not only shifted by a Doppler shift, but that its shape has changed in one or two days time. We now explore one simple way to characterize this change of shape.

5.5 Characterization of shape changes by ratioing CCF_{tot}.

The simplest way to characterize a change of shape is just to make the ratio of two CCF_{tot} taken at various epochs. It is better to normalize the two curves to each other before ratioing them. In fact, if there is a simple and pure shift between the two curves, the ratio should present a characteristic shape that we examine first. Hereafter we label the DRV(i) curve for night xx and first (resp. second) series of hundred exposures by C-xx-1 (resp. C-xx-2).

We have taken as a reference the curve C-24-1. Then we shift it by various amounts, (by -20, -10, -5, +5 ms^{-1}) and we plot the ratio of shifted curves to the original C-24-1 on Figure 5.15. Superimposed is the measured ratio C-24-2 over C-24-1. We can see that it does not fit any of the simulated Doppler shift curves. This is the proof that the CCF_{tot} has changed its shape over the course of 2 hours during night 24. Looking at the behaviour dictated by the dynamical model with 3 planets, we should have over 2 hours a modest increase DRV of the radial velocity, of about 0.3 ms^{-1}. Instead, the (black) ratio curve left side indicates a positive DRV shift of 1-3 ms^{-1} (interpolating within the theoretical curves) while the right side indicates a negative shift of about 2-3 ms^{-1}. This is consistent with the DRV(i) analysis shown in figure 5.8.

Actually, the "image" of all lines indicates a narrowing of the line over 2 hours, since the blue side is shifting to the red, and the red side is shifting to the blue. The two other curves collected in the same night, (displayed in pale grey on figure 5.16) indicate a similar behaviour as the first one.

The pattern of the two ratio curves (to the reference C-24-1) of the following night 25 (figure 5.17) is similar to the previous ones. This is in agreement with the motion of the star induced by the 3 planets, according to Figure 5.6 (same RV as the end of night 24). Given the scale on the left side, we may estimate the SNR on the ratio at about $\simeq$3000.

Keeping as a reference the CCF_{tot} of the first 100 exposures of night 24, C-24-1, we plotted the ratio curves for night 26, 27, and 28 respectively on figures 5.18, 5.19 and 5.20. For each night, we put two curves, representing respectively the first 100 exposures of the night, and the remaining part of the night. We also kept for comparison the ratio of C-24-2 over C-24-1 after normalization (black line).

The two "deformation" curves for night 26 are similar to each other, but they are very different from the deformation along the night 24. It coincides in time with a dynamical change due to planets.

As seen on Figure 5.18, the left side of night 26 ratio curves show a clear sign of negative DRV, which is in agreement with the motion induced by 3 planets according to Figure 5.6 (decrease of RV of about DRV=-3 ms^{-1}). The shape of the right side does not coincide with any of the theoretical curves for a pure shift only. It is possible that supergranulation changes (on a one day time scale) might add to the dynamical RV change a change of RV of the order of $\sim$ 1 ms^{-1}, as modeled by Meunier and Lagrange (2020) (fig.A4, case of medium supergranulation). It is also consistent with the difference of RV, retrieved either form the Gaussian fit to CCF_{tot} (nominal pipeline) or from the EPiCA method, as shown previously in figure 4.2. The difference shows an abrupt decrease by $\simeq$ 1.3 ms^{-1} from night 25 to night 26, which persists more or less for the following nights 27 and 28. Consistently, the ratio curves for night 27 and 28 are of similar shapes than the ratio curve of night 26. We note though that for nights 27 and 28, there are marked differences between the two curves for the same night, indicating some stellar activity at the time scale of one-two hours during these nights.

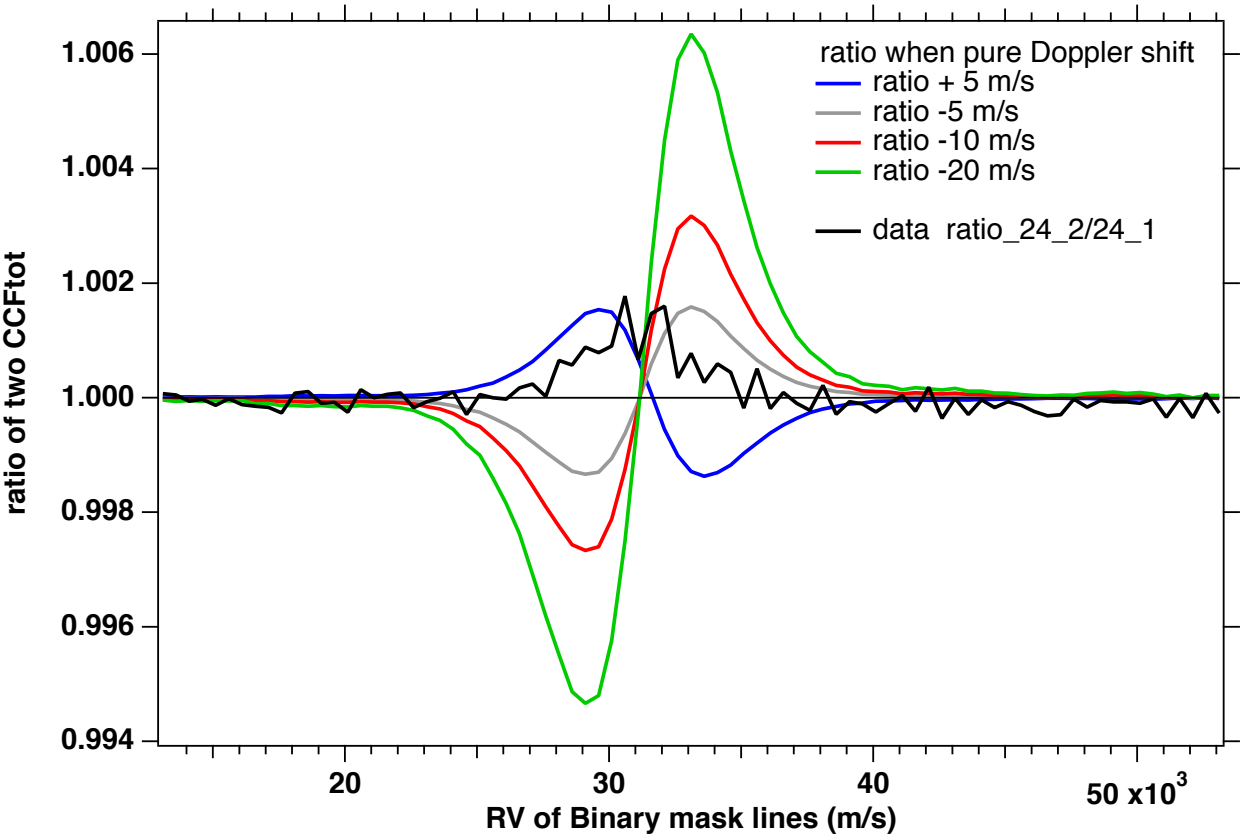


Figure 5.15: Simulation of the ratio curves C-24-1shifted/ C-24-1 for a pure Doppler shift (-20,-10,-5, +5 ms^{-1}). These shapes are characteristic of a pure (Doppler) shift. Superimposed is the measured ratio C-24-2 over C-24-1

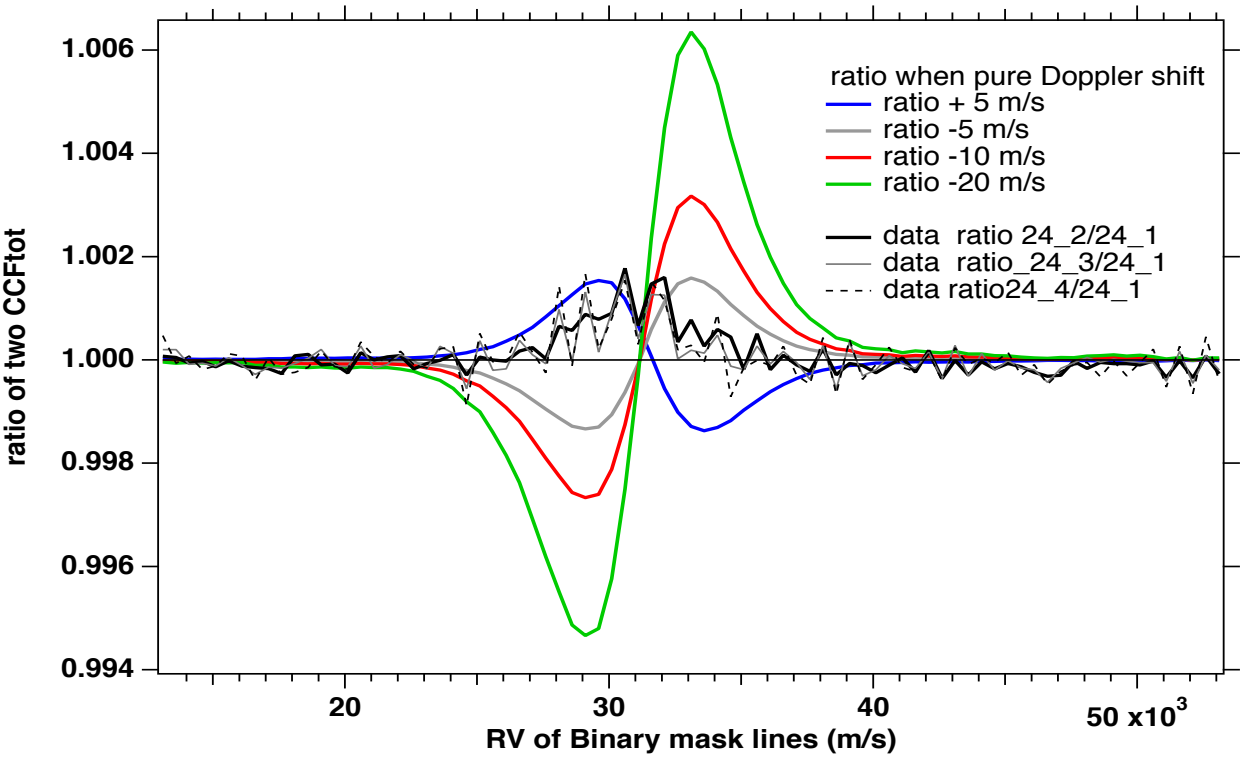


Figure 5.16: Same as Fig 5.15 with addition of measured ratios for the last periods of night 24 which are somewhat similar to the ratio curve C24-2/C-24-1

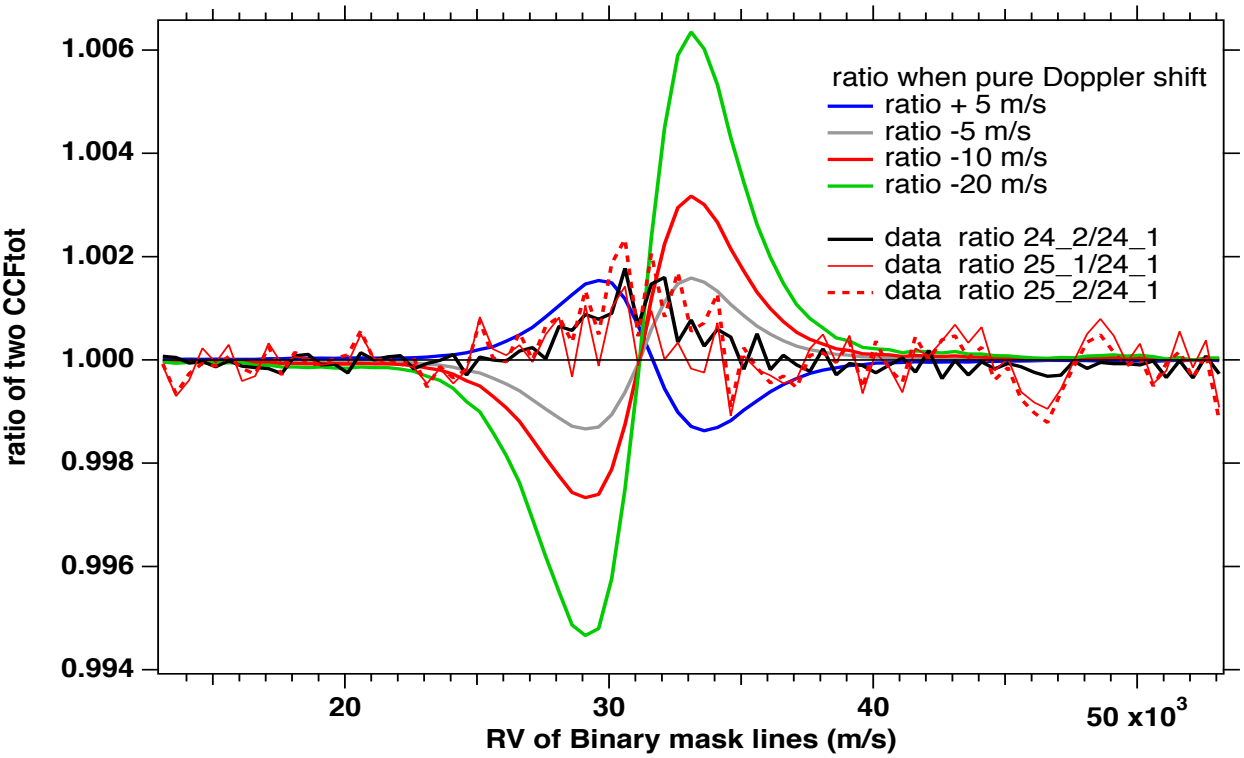


Figure 5.17: Same as Fig 5.15 with addition of measured ratios for the two periods of night 25 (thin red) which are somewhat similar to the previous ratio curves. Again, the left side indicates a positive DRV, while the red side shape indicates rather a somewhat ambiguous or negative DRV.

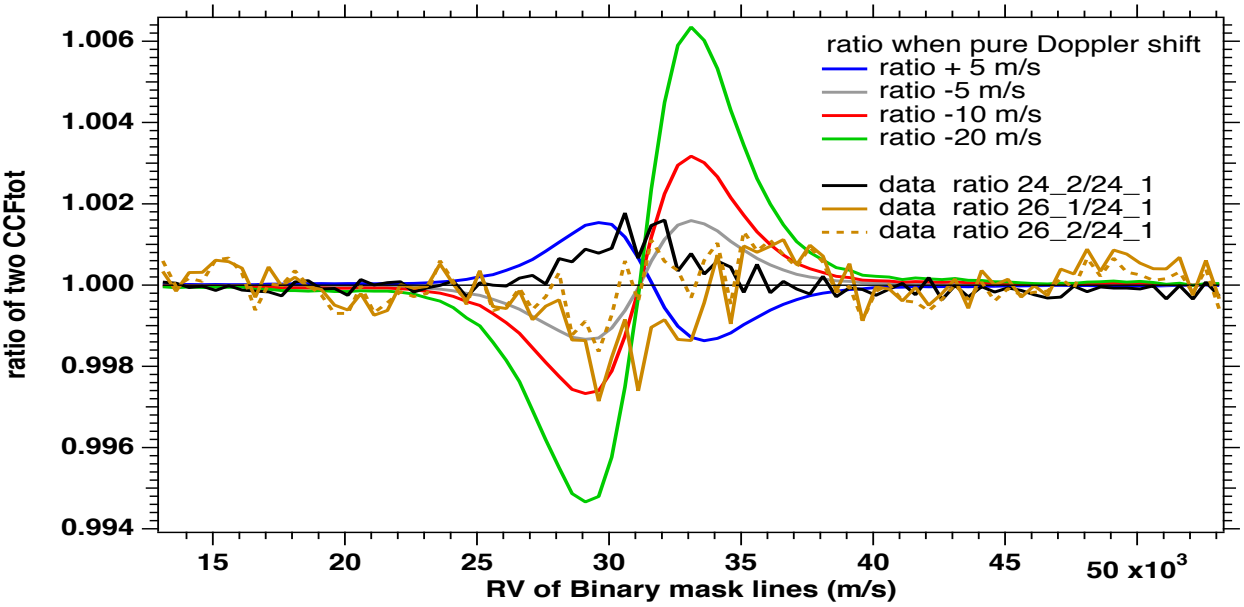


Figure 5.18: Same as Fig 5.15 (except for the third and last parts of night 24) with addition of measured ratios for the two periods of night 26 (brown curves). They are very similar. The left side of night 26 curves show a clear sign of negative DRV, which is in agreement with the motion induced by 3 planets according to Figure 5.6 (decrease of RV of about DRV=-3 ms^{-1}). The shape of the right side does not coincide with any of the theoretical curves for a pure shift only.

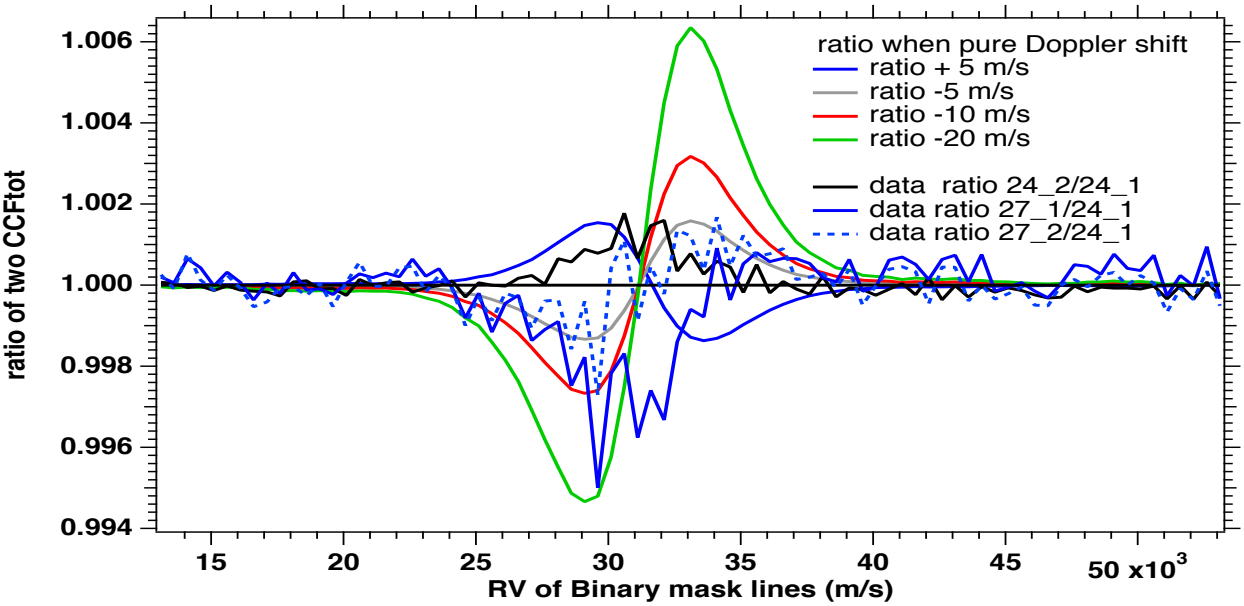


Figure 5.19: Same as Fig 5.18 but for the two periods of night 27 (blue curves). The two blue curves are similar on the left side and show a clear sign of negative DRV, which is in agreement with the motion induced by 3 planets according to Figure 5.6 (decrease of RV of about DRV= -4 ms^{-1} from night 24 to 27). The shapes of the right side are different, and does not coincide with any of the theoretical curves for a pure shift only.

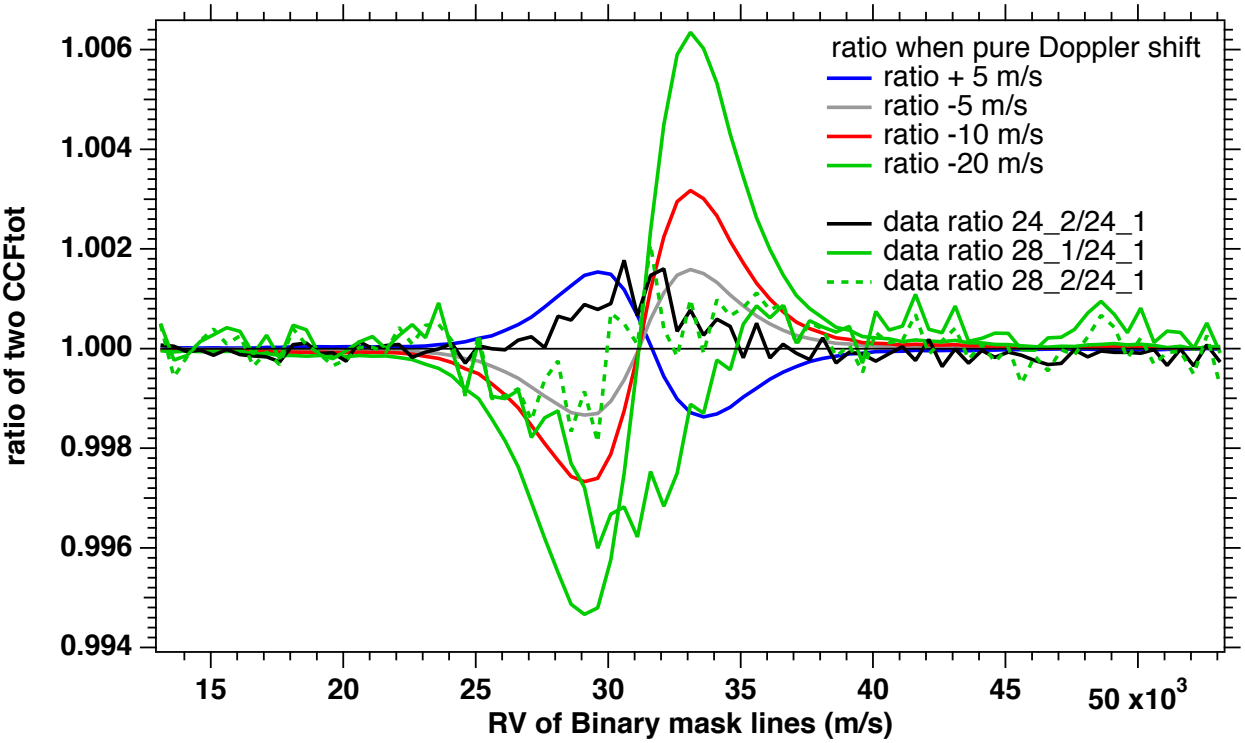


Figure 5.20: Same as Fig 5.18 but for the two periods of night 28 (green). The two green curves are on the left side more different than the blue curves from night 27 in the previous figure); they still show a clear sign of negative DRV, which is in agreement with the motion induced by 3 planets according to Figure 5.6 (decrease of RV of about DRV= -3 ms^{-1} from night 24 to 27). The shapes of the right side are different, and do not coincide with any of the theoretical curves for a pure shift only.

5.6 Summary, discussion, perspectives

We have applied the Pierre Connes 1st formula (a shift finding algorithm) to all individual points of the velocity grid over which is computed the CCF_{tot}, i.e. the total CCF contained in the ESPRESSO product, a global "image" of all spectral lines of the star used to build up the CCFs. We found that the DRV(i), the Doppler velocity shift, may be sometimes very far from being constant along the points of the CCF RV grid. This is the proof that the spectral lines are changing their shape, in addition to any Doppler shift, at a time scale of $\simeq$2 hours or a little shorter. We believe that this is the result of stochastic changes of the convective granulation blue shift (GBS) averaged over the stellar disc , and possibly of supergranulation between night 25 and 26. We have also studied the change of shape of CCF_{tot} in another way, by rationing all CCFs to a period of reference (first 2 hours of night 24). Indeed, the observed ratios do not fit the theoretical expectations for the case of pure shift only. Looking at figures 5.9, 5.10, 5.11, 5.12, it seems that the blue side (the left side, RV $\leqslant$ 31.5 kms^{-1}) of the DRV(i) curve (figure 5.10) is more uniform (smaller gradient over the low uncertainty points)

than the red side (RV $\geqslant$ 31.5 kms^{-1}). Correspondingly, the observed changes of shape of the blue side (revealed by the ratios) are much better following theoretical changes of the ratios predicted for pure shifts only, than the red side. Therefore, one could think of using only the blue side of the CCF$_{tot}$ in order to determine better a Doppler shift DRV for each exposure, a feature that can be obtained easily with the Connes shift finding algorithm. In other words, it seems that the blue side is more robust than the red side to perturbing stellar processes like GBS. We note that the blue side of the lines is weighted more by the ascending material in granules than the descending material in inter-granules lanes. It could be likely some evidence that the pattern of granules is more stable than the inter-granule pattern.

Potential improvements

In the previous chapter we found some evidence that using the Connes formula to CCF$_{tot}$ provided somewhat better results (fitting better a Kepler's law for the reflex star motion) than the standard algorithm (fitting the CCF$_{tot}$ by a symmetric Gaussian). Using the blue side only of CCF$_{tot}$ would be one step further in the same direction.

We have used data of a one–week campaign on star HD40307 in December 2018. As was recognized later, at that time the blue detector suffered from some instability problems affecting the data, that showed up for instance in the second half of the night 25. Therefore, it would be very useful to perform the same analysis with the refurbished blue detector, preferably on a bright star (to increase the SNR), with no planets (and a priori no Doppler shift due to reflex motion). Tau Ceti would be a good candidate.

The CCF$_{tot}$ is the sum of CCFs of all orders. But for this star, the signal is very weak in the blue, shorter of ESPRESSO order $\simeq$ 40, and may introduce some noise. Therefore, we could try to limit the piling up of CCF, either to the red side only (orders 90 to 169), or to orders 40 to 169.

Instead of applying the shift finding algorithm to CCF$_{tot}$, we could ultimately combine a line-by-line analysis and EPiCA. This requires a high SNR, that might be achieved by stacking the HD40307 by packets of 10 or 20 or using Tau Ceti). The objective of the first step is to try to identify a set of spectral lines that are showing the smallest changes of shape with stellar activity. Then, it is possible to pile up all CCF of all orders of this set of "quiet" lines, and use the CCFquiet to determine a DRV which would be less sensitive to stellar activity.

Another approach, somewhat similar to Cretignier et al. (2020) and Artigau et al. (2022), would be :

Step 1. to make two packets of spectral lines, divided by thresholds in EW: Packet 1. small EW; large GBS, large changes of GBS, unblended (list of Artigau et al. (2022)). Packet 2. large EW; small GBS, small changes of GBS.

Step 2. Then compute the CCF for the two packets respectively: CCFw and CCFs, from all orders together (for weak and strong lines respectively).

Step 3. Apply EPiCA when comparing two exposures (one may be a reference , like a template done from data), and compute the two retrieved RV, RVw and RVs, and their difference DRV(act) , as an index of change of convective GBS.

Step 4. Make an RV correction for each line, based on the particular value of this index, and look-up tables of GBS established as a function of DRV(act) over an ensemble of observations.

Difference with bisector analysis:

The analysis of the bisector is a well-known method that has been used to detect changes of the shape of spectral line, therefore with an aim similar to our aim. We note a potentially substantial difference between the methods. By essence, the bisector is making a link between both sides of the spectral lines (or the CCF image), while the PC algorithm is applied independently to each point of the line (or the CCF image), providing therefore more flexibility in the analysis of departures from a single shift between two epochs.

Chapter 6

Bias: different causes of and how to correct for them

In the field of RV retrieval with cross-dispersing, it is a common practice to correct the observed spectrum from the blaze function. In this section we investigate the bias on RV introduced by NOT correcting from the blaze function. We also found out that the official ESPRESSO blaze function is displaced, and dividing by the official blaze function would introduce a systematic bias, detrimental to the quality of RV retrieval. Therefore, we decided to use the ESPRESSO without any blaze correction.

6.1 The blaze function bias

Common characteristics of all cross-dispersed echelle spectrometer: each order presents an overall bell shape. It is a common practice to correct for the blaze before fitting a spectral line by a Gaussian, because fitting by a "flat" Gaussian a spectral line with tilted continuum introduces a bias in the center of the Gaussian, the wavelength and its Doppler shift w.r.t. the line of the BM.
We note that this bias also occurs when a stellar line falls on the side of another stellar line, therefore correcting for the blaze does not prevent from some systematic shifts (a bias) for blended line.

- The sign of the bias depends on the sign of the slope I(p), intensity I (electrons or adu) versus pixel p. Therefore it is of opposite sign on the right and left side of the order.
- The RV measurements is between 2 epochs. If the 2 epochs are near in time (this is the case of our ESPRESSO study), no variation of Doppler shift, a line does not move very much to the

point of change of blaze, so the bias is constant and not relevant for RV variations.

Finally, we decided not to correct for the blaze function, as it is discussed in the next part of this section.

Nevertheless we have attempted to use the official blaze function of Espresso. However, we found out that (at least for some orders) it did not represent well the data, the de-blazed spectrum presenting a strong slope (intensity vs wavelength).

We have investigated the bias introduced by a tilted continuum, developed the relevant equations, applied them to some ESPRESSO data, and estimated the corresponding bias, in ms^{-1}. We found out that it depends much on the extent over which the spectral line is fitted by a Gaussian.

With the choice of Lafarga et al. (2020) type scheme to use a width of the slot/hole = about 1 pixel for the CCF, and a sampling grid of the CCF of about one pixel, the CCF for each line is in fact very similar the original line. Therefore, the bias on a CCF (approximated by a Gaussian) should be about the same as the bias on a line (approximated by a Gaussian). Therefore, we may continue to study the problem directly on the lines, and not the CCFs. This is particularly true when the shift is determined separately for each spectral line. In the case of the full-Lafarga scheme, in which all CCFs for one order are piled up together with appropriate weights, spectral lines are on both sides of one bell-shaped order (raw spectra), therefore the various biases from various spectral lines will more or less compensate each other.

6.2 Mathematical formulation of the bias introduced by a continuum with a slope and application to ESPRESSO data.

Here we establish the equations allowing to estimate analytically the displacement of the minimum of a center of a Gaussian, when a “flat” Gaussian (untilted) is tilted by multiplication by a linear function. Of course, the minimum of the “flat” Gaussian is at its center. After tilting, the tilted curve has a minimum which is displaced from its original position, by an amount (the shift) which is a measure of the bias phenomenon, when the minimum of the tilted curve (raw data) is taken as the RV of the spectral line, instead of the “deblazed” data with a flat continuum.

6.2.1 Analytical formulation.

We start with a model of a spectral line with a flat (constant) continuum. Here we choose to have the continuum at unity, and the spectral line can be described as a Gaussian in absorption.

$$f(x) = 1 + Aexp(-(\frac{x - x_0}{w})^2) \tag{6.1}$$

Here we choose to have A negative, to be consistent with the Igor software formulation of the fit by a Gaussian, which uses this convention. A is the (relative) depth of the spectral line, also called the contrast in the context of the Binary Mask. x_0 is the center of the line, from which will be determined the shift/bias, and w is the width of the line , both in pixel coordinates.
In order to simulate the effect of the blaze (or other spectral distortion), we multiply this model by a linear function to get a "tilted" spectral line, with a continuum which has now a slope b:

$$g(x) = a + bx \tag{6.2}$$

At position x_0, the value of the tilted linear continuum has the value Y_2= a+bx_0 and

$$g(x) = Y_2 + b(x - x_0) \tag{6.3}$$

Illustration of flat Gaussian and tilted by linear function Gaussian is shown on Figure 6.1.

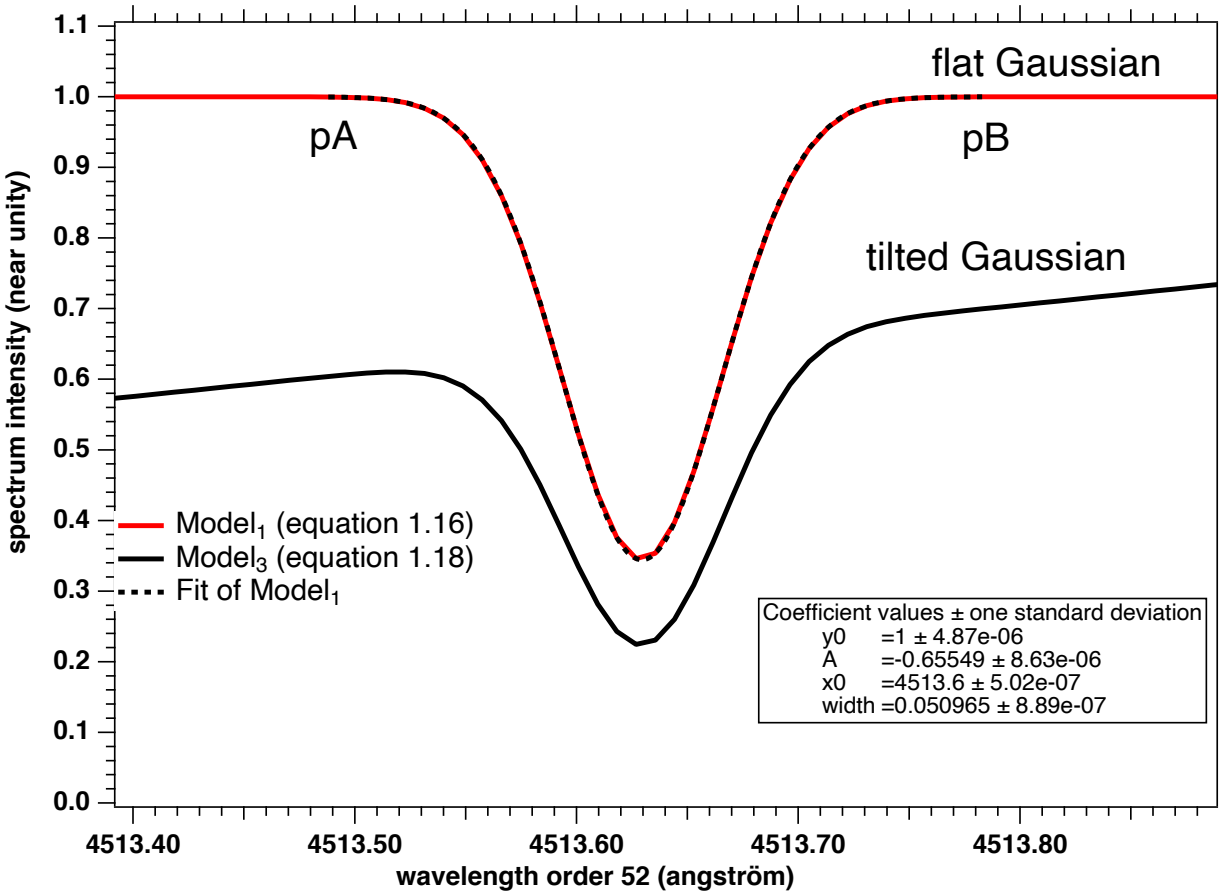


Figure 6.1: Red curve: a flat Gaussian from 6.1. When fitted between points pA and pB by a flat Gaussian, the parameters are found with a very large accuracy (as expected). The fitting curve is the black dashed line. Black curve: a tilted Gaussian obtained by multiplying the flat Gaussian by a linear function.

Now we find where is the minimum x_{min} of U(x)=f(x)g(x) near x_0, and we will call the shift, or bias (indifferently in the following), introduced by a blaze or distortion, as bias=x_{min}- x_0.

$$exp(-(\frac{x-x_0}{w})^2)(-\frac{2A(x-x_0)}{w^2}(Y_2+b(x-x_0)+Ab))+b=0 \tag{6.4}$$

We divide by AY_2 and pose X=x-x_0:

$$exp(-(\frac{X}{w})^2)(-2\frac{X}{w^2}(1+\frac{b}{Y_2}X)+\frac{b}{Y_2})+\frac{b}{Y_2}\frac{1}{A}=0 \tag{6.5}$$

The desired shift (bias) is the value of X which satisfies equation 6.5. The 3 parameters on which depends the bias (or shift between flat and tilted) are:

- The depth of absorption A (A is negative and A= - contrast): coef[0]
- the (Gaussian) width of the spectral line (or CCF) w (in pixels): coef[2]

- the ratio of b/Y_2 (coef[1]) where Y_2 is the value of the continuum at line center after the multiplication of the model by a linear function a+bx (x, in pixels), and b, the slope of the continuum, i.e. equivalent to (dY/Y)/dX= d(log(Y))/dx; therefore, it is the (local) logarithmic slope of the spectrum continuum at the center of the line.

In order to find the solution of 6.5 for a given set of parameters, we used the Igor function FindRoots.

Numerical example: For the Gaussians of figure 6.1, flat and tilted, we have: A= -0.65549; w=5.87 pixel; Y_2= 0.65; b/Y_2= 0.0043224. This last value of the log slope of the continuum is taken as 3 times the value of an observed line at this place, which is the beginning of the order with a rather large slope among the data.

For these parameters, the solution of 6.5 is a bias of -0.039 pixels. Given the wavelength size of one pixel, dλ per pixel (the dispersion) and the velocity of light, it corresponds to a bias of -22.6 ms^{-1}. The size of one pixel in ms^{-1} is c dλ/λ. It is plotted as a function of the pixel number for orders 52 and 146 on figure 6.2. It does not vary very much, between 370 and 593 ms^{-1}, with a mean value around $\sim$ 482 ms^{-1}. It is also quite similar for all orders; this is a classical optical feature of cross-dispersed spectrometers with an echelle grating.

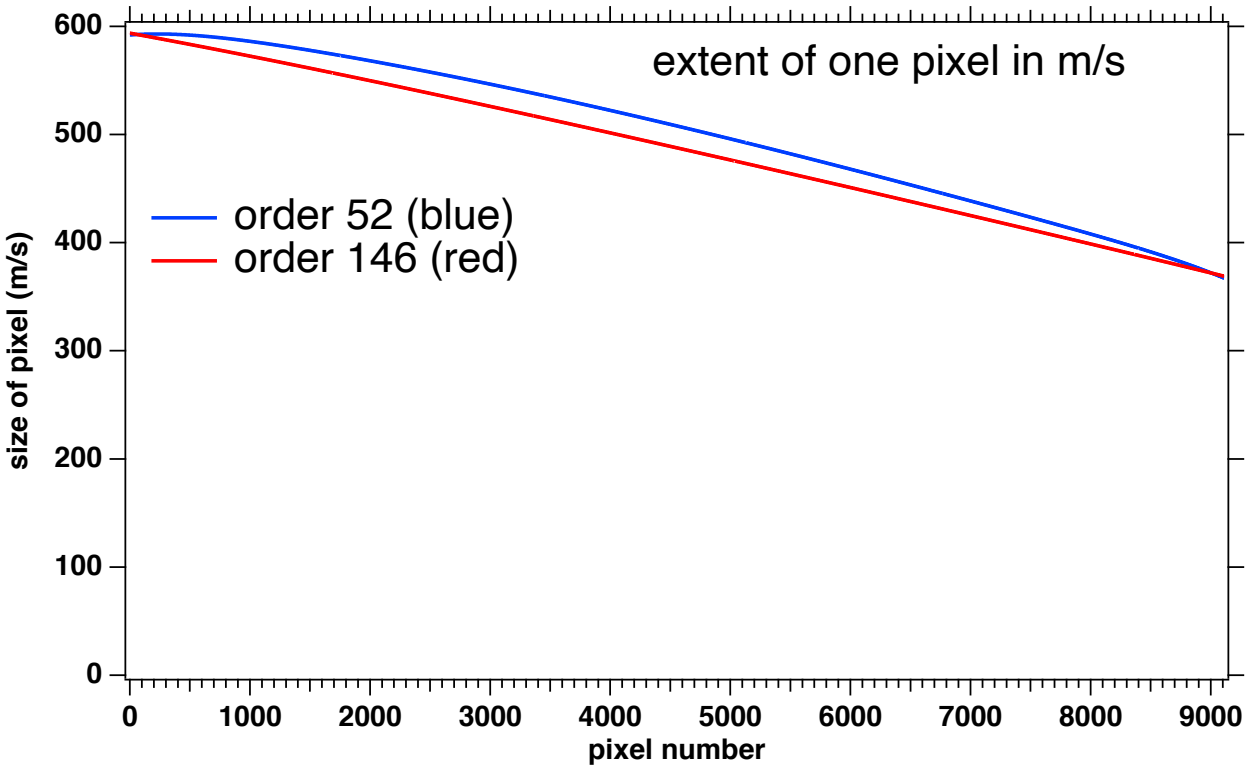


Figure 6.2: The extent (or size) of one pixel in RV units (ms^{-1}) =cdλ/λ is displayed for order 52 (blue) and 146 (red). The two curves are quite similar, a feature valid for all orders.

6.2.2 Dependence of shift on the three parameters.

We wish to explore the dependence of shift on the three parameters.

- A, depth (or contrast) of the spectral line, is varied from 0.1 to 0.8 covering most of observed values.
- w, the (Doppler) width of spectral line, in units of pixels. The value of w depends on the element atimoc mass and temperature of the photosphere. It is varied between 1 and 21 pixels, or $\sim$ 0.5 to 10.5 kms^{-1}. We note that the width in pixels, for the same element (say, iron, FeI or FeII), the width in pixel is proportional to wavelength: greater in the red than in the blue.
- b/Y_2, the logarithmic slope of the continuum (relative increase of the continuum per pixel). Values considered are 0.0005, 0.0015 and 0.005

The numerical example above is for b/Y_2= 0.0043224 is 3 times larger than rather extreme case of continuum slope due to the blaze. However, when one line falls on the flank of another line, the slope of its continuum may be substantially larger. As example of a such line, we can use line around 1381 pixels shown at Figure 6.3. We can compute the position of the line center by fitting Gaussian, but also we can compute it by fitting tilted Gaussian and result of such fit has to show a real center of line. Also, we can compute slope of continuum for this line and make a correction by the continuum, to obtain non-tilted line, after that if we fit corrected line by Gaussian we should obtain approximately same value as for a tilted Gaussian fit. Difference in line center positions can be also calculated according to formula 6.5.

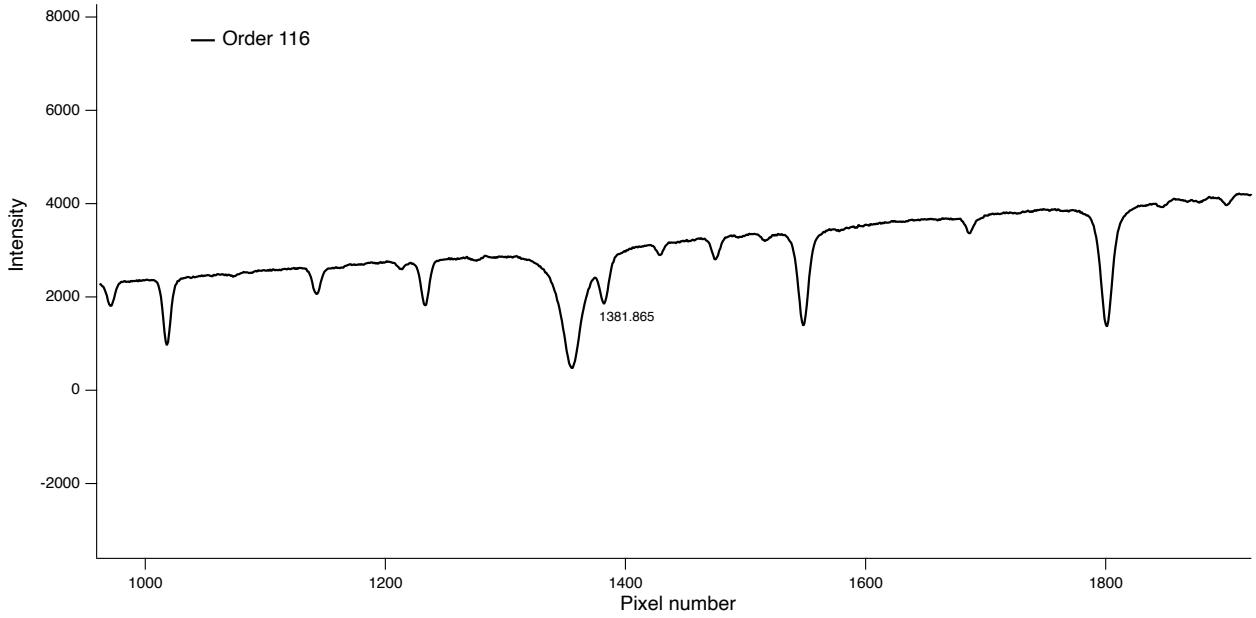


Figure 6.3: Example of the line which falls on the flank of nearby line.

After a first fit of non-slope corrected line by Gaussian we obtain value of center 1381.86 pixels; by fitting with tilted Gaussian we obtain 1382.27. After a linear fit for the slope we obtain b/Y_2 equal to 0.0088, we can make continuum correction for this particular line, to eliminate blaze effect. This correction is made by division of data by the linear function a+bx, where a and b are taken from linear fit result. After a such correction continuum of line is equal 1 and slope is 0, this is illustrated on figure 6.4.

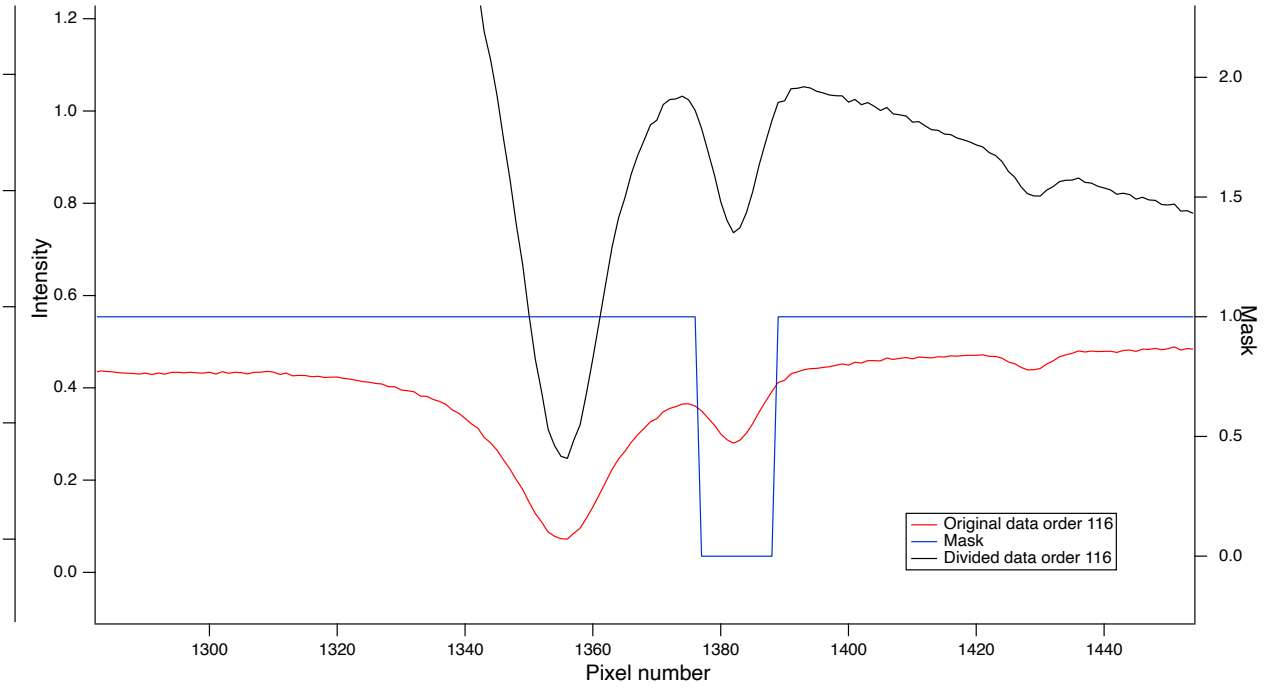


Figure 6.4: Example of the line which falls on the flank of nearby line. Red – original data, left scale. Black – data after continuum correction of line, left scale. Blue – mask used for linear fitting, right scale.

Now if we fit the corrected line with a Gaussian only, we should obtain a value of center similar to the one which we obtained after the fit by tilted Gaussian on original data. Indeed, Gaussian fit of corrected data return value of center equal to 1382.29. We also make a check of our formula 6.5 – if the difference between fit by Gaussian of original data and fit by Gaussian of corrected data will be the same as predicted by the formula. The shift predicted by the formula is equal to -0.21, the shift calculated on data basis is equal to -0.4286 pixels. We can conclude, that real bias is twice larger than the expected bias caused only by blaze.

It is noted that correcting for the blaze would not allow avoiding this problem. To make a check of formula 6.5 we choose a line which is separated from other lines, so this line should be affected only by the blaze. Such line is the one with center around 1800 pixels, order 116. The value of b/Y_2 is equal to 0.00033, while the predicted shift of line center is equal to -0.00459. If we fit original data by a Gaussian, the line center is located at 1800.69023, after continuum adjustment the line center is located at 1800.69573, difference is equal to -0.0055, which is in a good correspondence with predicted bias.

The shift is the solution of equation 6.5 that may be converted to ms^{-1} (here we took a mean value of $481.5 ms^{-1}$ per pixel as representative). On Figure 6.5, 6.6, 6.7 we show the shift of the minimum of a

flat Gaussian line when it is multiplied by a linear function, as a function of the Doppler width of the line (in pixels) computed for different A and b/Y_2

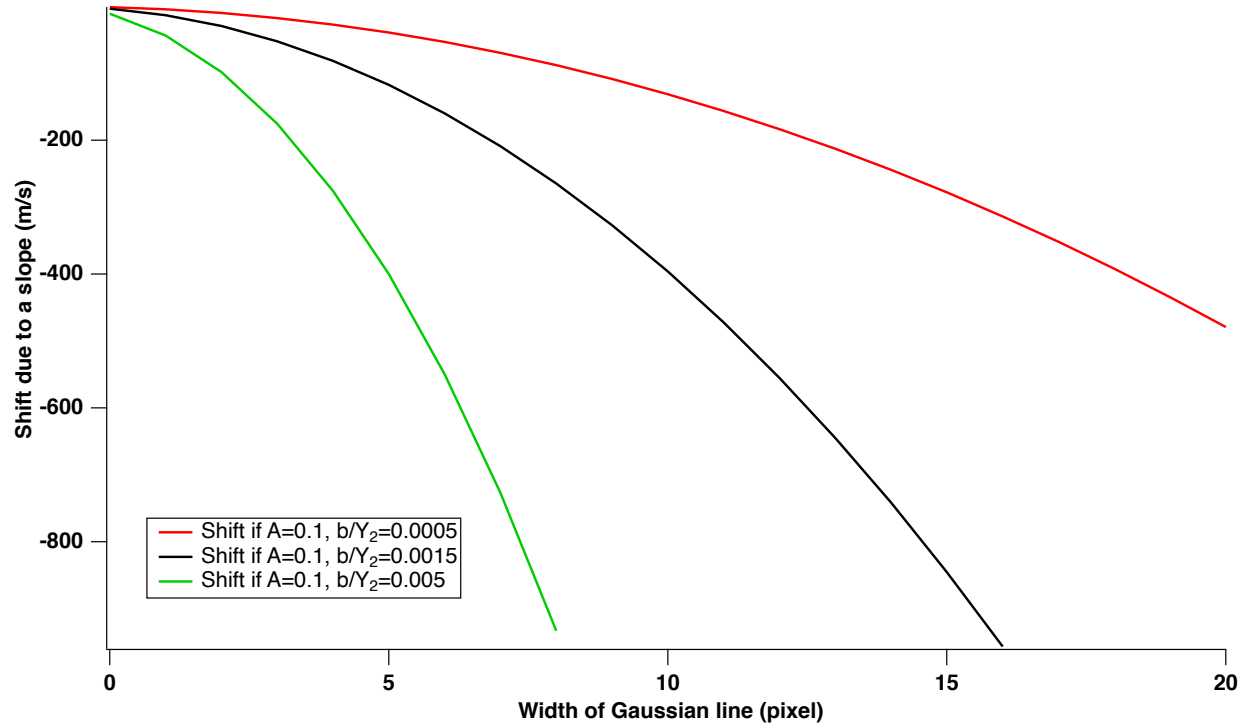


Figure 6.5: Shift of the minimum of a flat Gaussian line when it is multiplied by a linear function, as a function of the Doppler width of the line (in pixels). Red: A=0.1, b/Y_2=0.0005, black: A=0.1, b/Y_2=0.0015, green: A=0.1, b/Y_2=0.005.

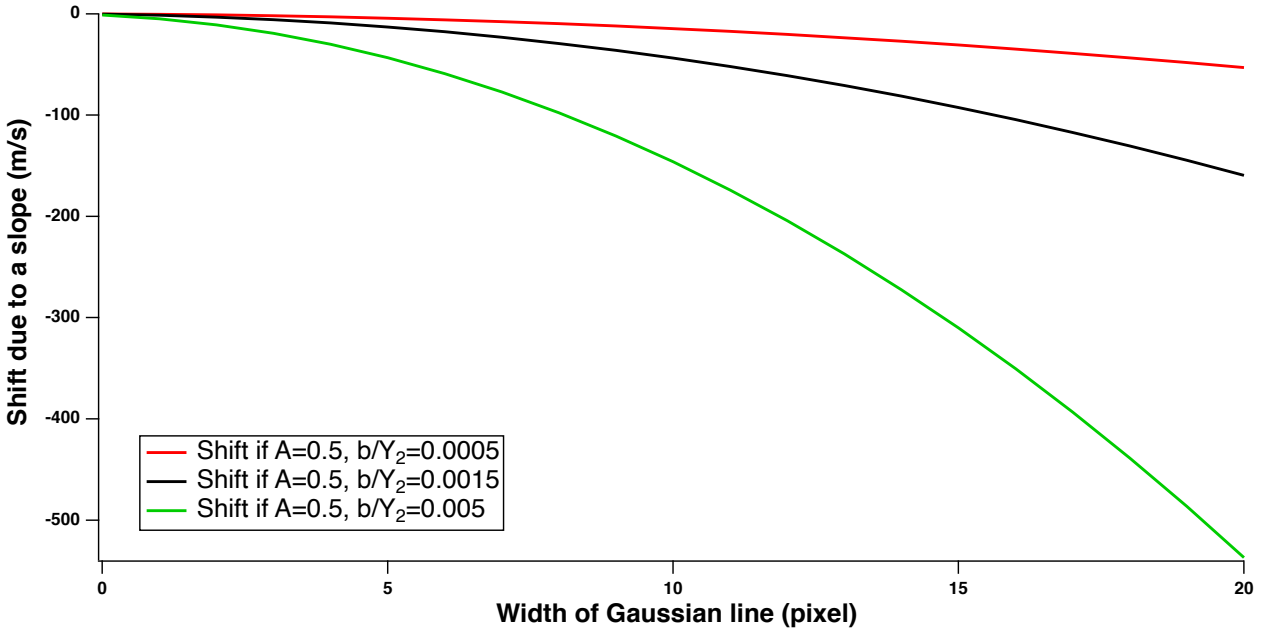


Figure 6.6: Shift of the minimum of a flat Gaussian line when it is multiplied by a linear function, as a function of the Doppler width of the line (in pixels). Red: A=0.5, b/Y_2=0.0005, black: A=0.5, b/Y_2=0.0015, green: A=0.5, b/Y_2=0.005.

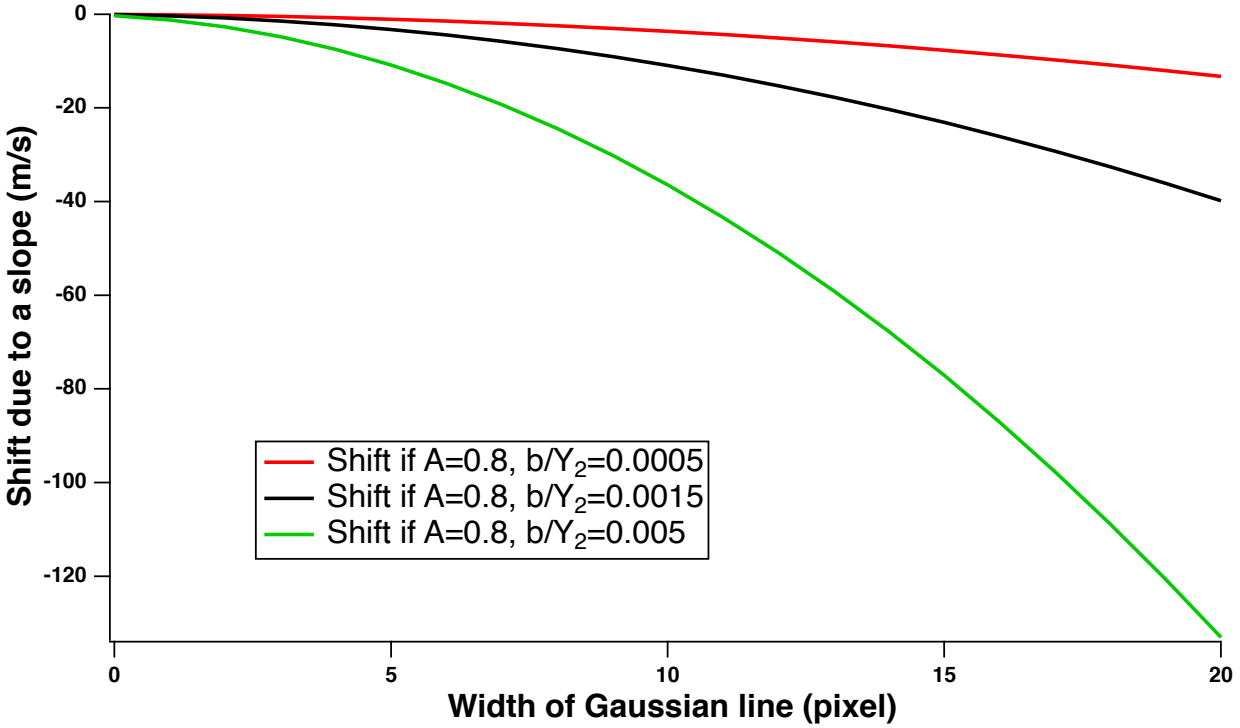


Figure 6.7: Shift of the minimum of a flat Gaussian line when it is multiplied by a linear function, as a function of the Doppler width of the line (in pixels). Red: A=0.8, b/Y_2=0.0005, black: A=0.8, b/Y_2=0.0015, green: A=0.8, b/Y_2=0.005.

After this study of parameters we can conclude:

- The shift is increasing with the width of the line (in pixel).
- The shift is increasing with the slope
- The shift is decreasing with the value A of depth of the line.

6.3 The blaze function of ESPRESSO.

6.3.1 The theoretical blaze function of an echelle grating.

Equation of echelle grating The grating equation taking into account possibly negative angles in the oriented plane of diffraction is:

$$d(\sin i + \sin i') = m\lambda \tag{6.6}$$

with m an integer, called the diffraction order. We can take the convention that the angle of incidence i is positive. Then angles i and i' have the same convention of sign: if they are on the same side of the normal, they are both positive. If i' is on the other side, it is negative. An echelle-grating is most

often used in the so-called Littrow mounting, in which i and i' are almost equal, the angle i such that the incident beam is just perpendicular to the facet of a groove.

As a matter of fact, in the grating formula, the index of refraction does not appear. This is because λ in equation 6.6 represents the wavelength of the light in the medium where it propagates. It must be realized that, while photons do not change their energy when going through the spectrometer, the fact that they propagate in air changes their wavelength (even at low pressures). It has an impact on the direction (determined by angle i' emission from the grating) they take when leaving the echelle spectrometer because the constructive interference on grating lines is built in air at ambient pressure. This is the main reason to put high resolution spectrometers under vacuum, like HARPS and ESPRESSO; in which case, the wavelength λ of equation 6.6 is in vacuum.

Echelle geometry is shown on Figure 6.8.

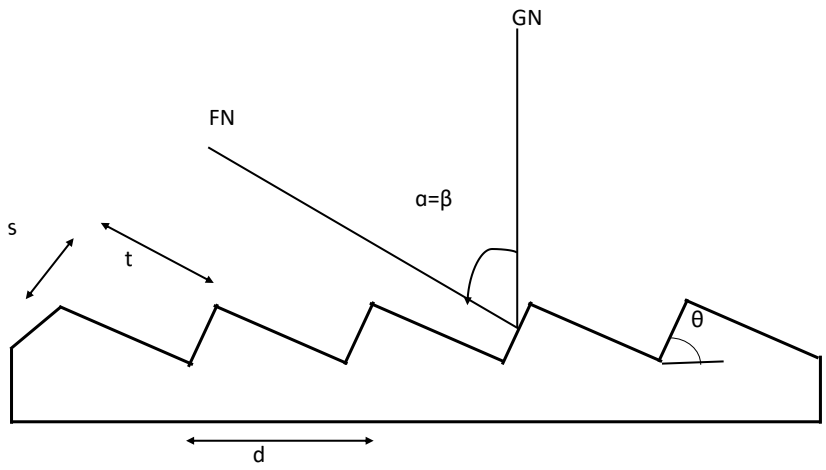


Figure 6.8: Echelle geometry. The groove spacing d, step width t and step height s are shown. GN is the grating normal, FN is the facet normal.

We are with an echelle-grating with a cross-disperser. Experimentally, we see that a flat spectrum is decomposed on many orders, and each order has a maximum intensity around the center, decreasing on both sides. This is the so-called blaze effect. The deeply rooted reason is that, when light is reflected on each small facet s, it suffers diffraction: a parallel beam of light going through an aperture will present an angular pattern of diffraction after the aperture. Here, each facet may be considered as an "infinite slit", an aperture with the width of the facet (perpendicular to the propagation of light).

The light has an angular distribution after the facet, with the form of a squared sinus cardinal, (sin x/x)^2; Calling d_{emer} the small angle with the normal to the facet, the distribution of light as a function of d_{emer} (the blaze function) is given by :

$$I(x) = I_0 sinc^2(\frac{\pi s}{\lambda} d_{emer}) \tag{6.7}$$

where sinc(x)=$\frac{\sin(x)}{x}$

The NIR/ACS (near-Infrared/Atmospheric Chemistry Suite) experiment on board Exomars is performing atmospheric sounding by the technique of solar occultations, in orbit around planet Mars. It is an echelle grating spectrometer, with orders sorted by an AOTF (Acousto-Optic Tunable Filter) system. During the analysis, J.L. Bertaux developed a code allowing to adjust the parameters of the sinc formulation of the blaze function to the observed bell-shaped spectra. Actually, the number of grooves per mm is known, and the parameters that can be adjusted are: the exact value of angle α for the maximum returned intensity, and the effective size of the facet, which may be slightly different from the geometrical size, and which impacts on the width of the blaze function for each order. We have applied this algorithm to some ESPRESSO orders.

6.3.2 The official ESPRESSO blaze function.

In view of the fantastic quality of the Espresso spectrometer at VLT (Pepe et al., 2021), we found that in order to evaluate the improvements on RV brought by atmospheric correction, it would be a good idea to use publicly available data from ESPRESSO.

We are using ES_S2BA data product, which are extracted spectrum, order by order. The reason is that Rosine Lallement found out, in the frame of another project (ESO Edibles project with UVES) that the atmospheric correction can not be conducted perfectly on combined-order spectra, because they are re-sampled on a fixed wavelength grid with an interpolation on the raw data, and this is detrimental to the quality of the atmospheric correction process, especially near the atmospheric line centers.

During the study we found a discrepancy in the ESO manual description of the ESPRESSO data product. Table 6 from document ESO Phase 3 Data Release Description is shown in Figure 6.9 and a Figure 6.10 of the content of order Espresso 152 (true optical grating order 86).

The tar archive[9] that is delivered as an ancillary file contains pipeline products which are needed by the ESPRESSO DAS recipes. The content of the tar file is listed in Table 6.

Table 6. Content of the tar archive

ORIGFILE begins with	**Type**	**Product category** HIERARCH ESO PRO CATG	Num-ber	Description
ES_S2DA	fits	S2D_A	1..N	Extracted spectra for each order, not flux-calibrated
ES_S2BA	fits	S2D_BLAZE_A	1..N	Extracted spectra for each order, not flux-calibrated, divided by blaze function
ES_S2SA	fits	S2D_SKYSUB_A	0..N	Extracted spectra for each order, sky-subtracted, not flux-calibrated, divided by blaze function; only if fibre B is on sky
ES_PFBA	fits	BLAZE_A	1	Blaze function

The ORIGFILE product names follow a naming convention which is

ES_<TYPE>_<OBS_ID>_<DP_ID>_<RES>_<BINX>x<BINY>_<TEL>.fits

See Table 7 for details. An example ORIGFILE name would be

```
ES_SFLX_2147913_2018-11-09T07:18:46.594_UHR_1x1_U3.fits
```

Figure 6.9: Table 6 from ESPRESSO manual with description of product categories.

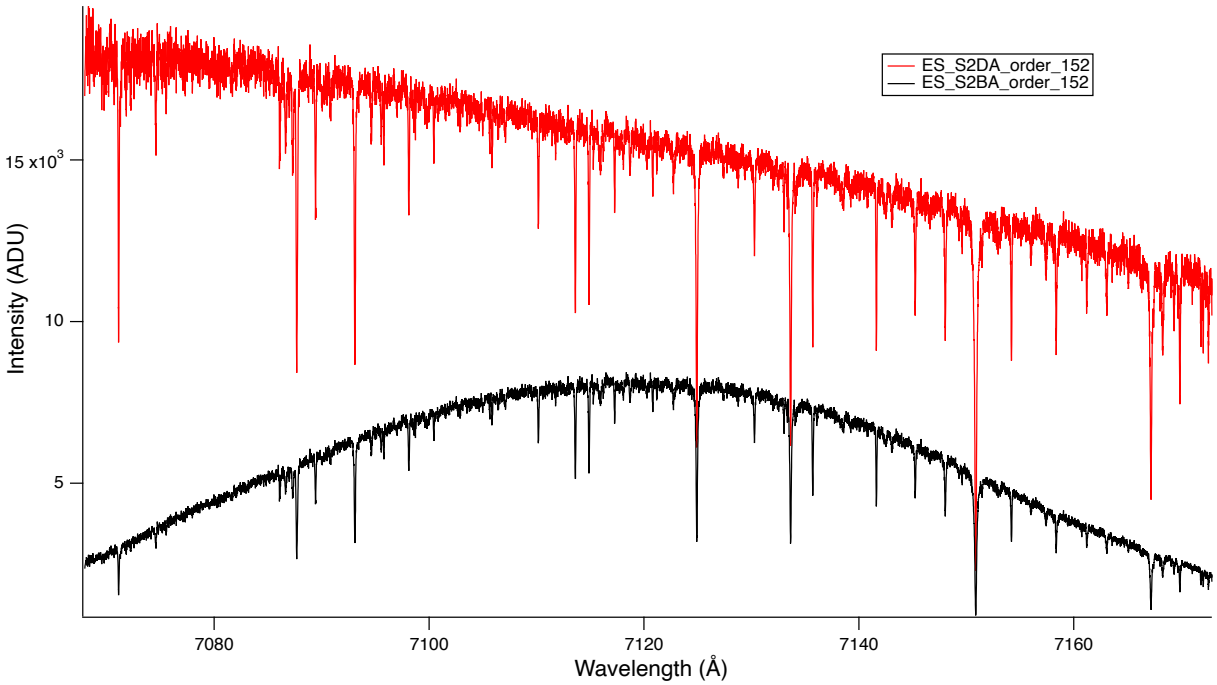


Figure 6.10: For the order ESPRESSO 152 (true optical order 86), the data product from ES_2BA (black) is the original spectrum (with its characteristic bell-shape due to the blaze) and ES_2DA (red) is the spectrum after division by the ESPRESSO blaze function.

As said above, the classical theory for an echelle grating is that it is the diffraction pattern of an

infinite slit; the infinite slit being one facet of a groove of the echelle grating, the facet is perpendicular to the incident beam in a Littrow mounting. It is a sinus cardinal function, sin x/x, peaking at center of emerging light beam, with value 1 at maximum where x=0. We have computed this theoretical function for the parameters of ESPRESSO spectrometer, and find a reasonable agreement with the overall shape of one order spectrum contained in ES_S2BA data product. We conclude that the data contained in ES_S2BA data product is NOT corrected from the blaze function, in contradiction with the Description columns in the Table, which says: "Extracted spectra for each order, not flux calibrated, divided by the blaze function." Most likely, it is a typo in the Table 6, with an inversion in the column Description between the two first entries, ES_S2BA and ES_S2DA.
After we discovered this, we contacted Burkhard Wolff from ESO mentioning this discrepancy between data and manual. Burkhard Wolff confirmed our finding and made correction at the ESPRESSO user manual.

The official ESO blaze function of ESPRESSO does not flattens the spectrum.

The curves of figure 6.10 are reproduced on figure 6.11, with some additional curves. The blue dashed line is the blaze function (left scale) extracted from ESPRESSO data file ES_PFBA_181224A_HR_1x1.fits, and its maximum value is about 0.5. The division of the black data spectrum by the ESPRESSO blaze function for order 152 gives the red curve found in ES_S2DA. The green curve is our estimate of the blaze function (Blaze 86) determined as a best fit to the continuum of the black curve, in windows delineating the continuum with the blue mask 152 curve (method developed by Bertaux for ACS/NIR). The blue data curve is then the deblazed spectrum obtained with our method, the result of the division of the black spectrum by our blaze 86 blaze function for true order 86 (ESPRESSO order 152).

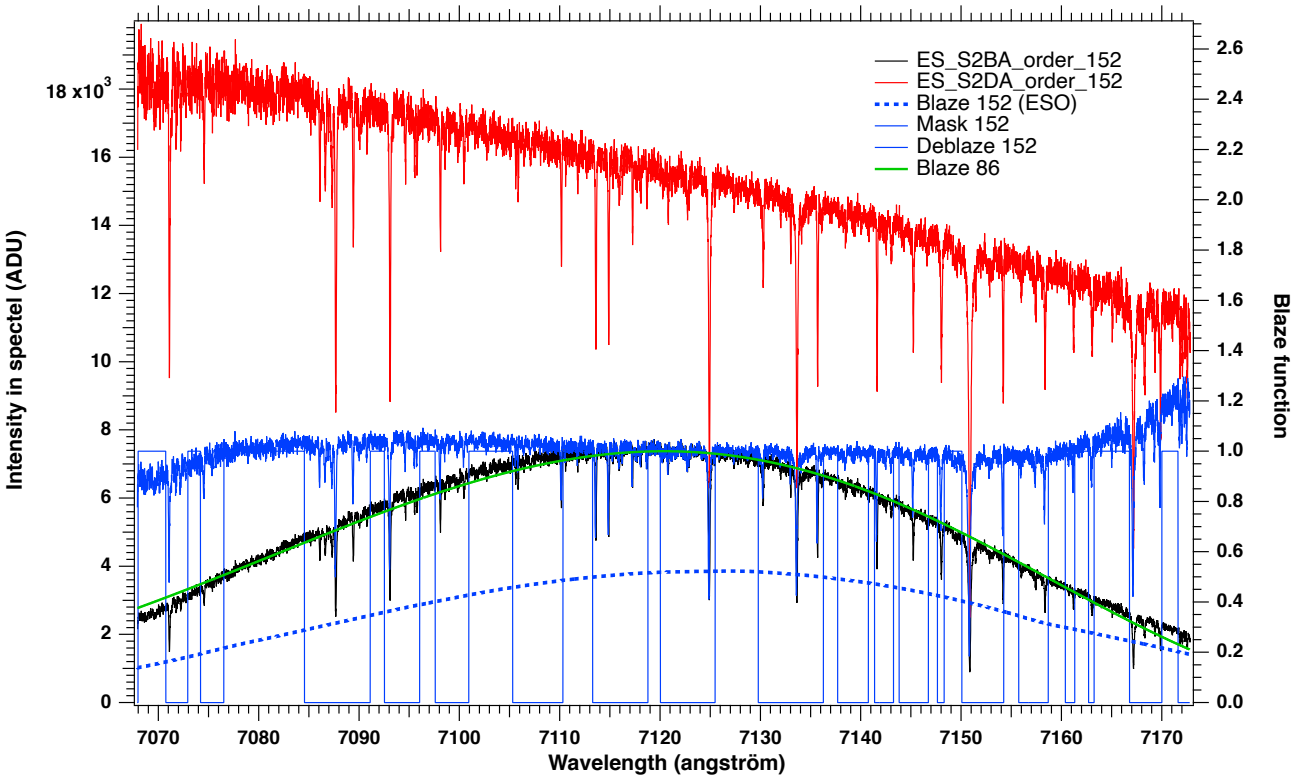


Figure 6.11: Blue dashed, right scale: ESO provided blaze function for order 152. Green curve: our best fit blaze function, peaking at 1 (right scale). Black curve, left scale: Order 152, ES_S2BA data for order 152. Red curve, left scale: ES_S2DA data, order 152. Blue curve, left scale: original spectrum (black) divided by our best fit blaze function.

The red curve is an unrealistic representation of the star spectrum because it varies by a factor of 2 on a short wavelength range. However, as pointed out by François Bouchy, this de-blazed spectrum should not be used directly for further analysis. It is a number of ADU per pixel (corrected from blaze), but the size of the pixel is varying substantially across one grating order. Therefore, one should divide the ES S2DA spectrum by the wavelength size of each pixel. This indeed gives a spectrum which is almost flat, in units of ADU per angström, as shown on figure 6.12.
This information was captured late in the course of those studies. Instead of using this flat deblazed spectrum, we used the original spectra, non-deblazed: raw extracted spectra ES S2BA.

One must realize that changing the values of the raw extracted spectra ES_S2BA is not neutral, since the error bars associated to ES_S2BA must have been correctly computed, but any change on the values of the raw spectra must be accompanied by a change on the error bars. Therefore, one should also be cautious when using the official blaze function of ESO (the maximum value of about 0.5 is somewhat puzzling, but maybe is due to the image slicer), and compute correctly the new error bar. The relative error bar of the values in ADU per pixel should be transported to fluxes per unit wavelength, and re-transformed into an absolute error bar for correct further error propagation

computations.

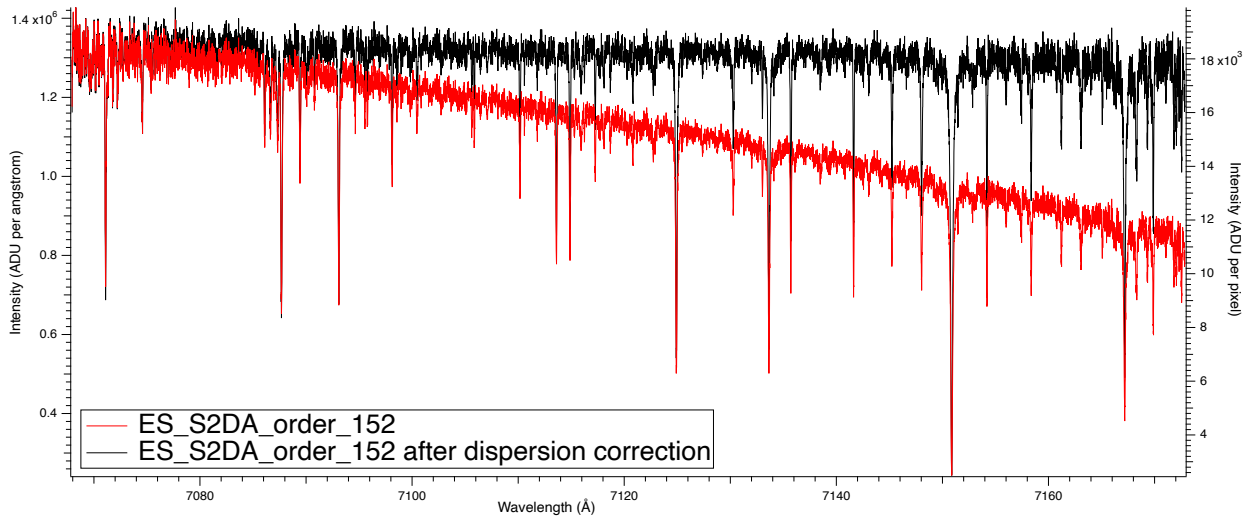


Figure 6.12: Black curve, left scale: ES_S2DA data, order 152 after dispersion correction, intensity in ADU per angstrom; Red curve, right scale: ES_S2DA data, order 152, intensity in ADU per pixel.

6.4 Bias caused by the Gaussian fitting.

When fitting a tilted Gaussian, discretized, by a flat Gaussian, it also introduces a bias, which depends also on the number of points over which is made the fit. This bias is different from the bias (or shift) of the minimum of a Gaussian when it is tilted which was addressed earlier on an analytical way. Also, the problem is different with a spectral line, which shape differs from a real Gaussian.

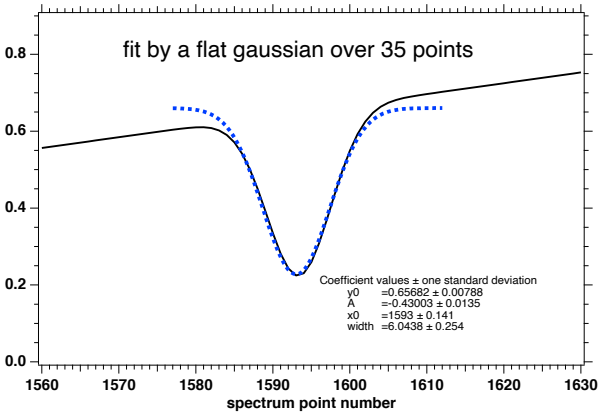
Fitting a tilted Gaussian by a flat Gaussian. We have selected as a tilted Gaussian the model mod_{3tilt} of figure 6.1, which is inspired by a real line of order 52, but with a slope of the continuum enhanced by a factor of 3:

$$mod_{3tilt} = mod_1 * (0.65 + 3 * 0.65 * 12.644 * (p - 1593.3)/8775.69) \tag{6.8}$$

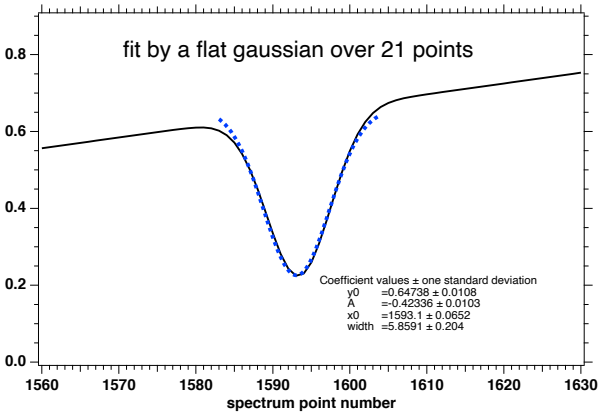
where p is the pixel number.

The fit is made when plotted as a function of the pixel number, with a size of pixel of 567.729 ms^{-1}. It allows converting the bias and the error, found in units of pixels, into their corresponding values in ms^{-1}. The spectral line is fitted by a flat Gaussian over different number of points: 9, 15, 21 and 35 points, as is illustrated on Figure 6.13

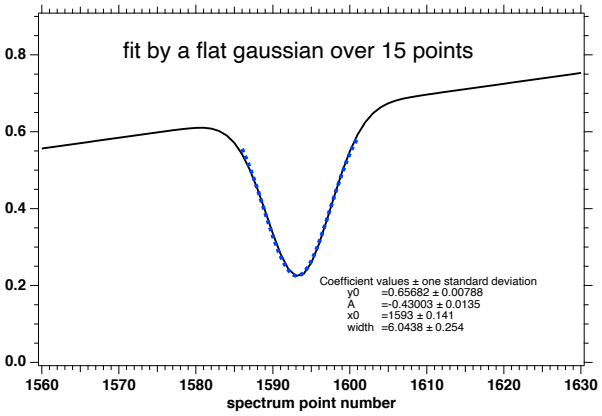
The bias and the error returned by the Gaussian fit both depend substantially on the number of



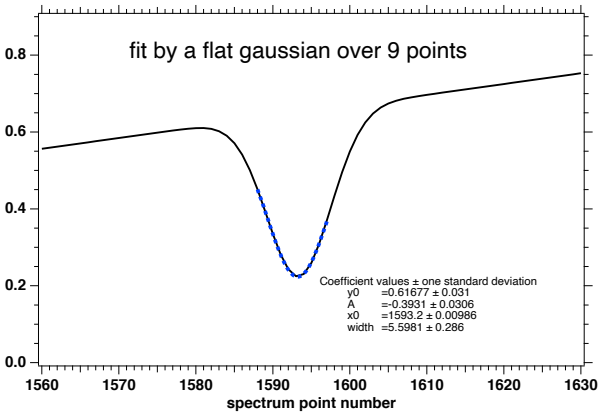
(a) Fit by 35 points



(b) Fit by 21 points



(c) Fit by 15 points



(d) Fit by 9 points

Figure 6.13: Fit of tilted Gaussian (black curve) by a flat Gaussian (blue dashed line).

points on which is fitted the tilted Gaussian. They are reported in Table 6.1, and plotted as a function of the number of fitted points on Figure 6.14 In spite of the fact that the model to be fitted is noise-free, the error is increasing with the number of points to be fitted. This may be understood, because the returned error accounts also for the departure of the model from the data (here a noise-free tilted Gaussian). In Table 6.1 is also indicated the χ^2 per point (a measure of the departure from the modeled data), which increases a lot with the number of points. Here we find another example of the well-known feature of non-linear least square fit: when a data set is fitted by a wrong model (fitting a flat Gaussian by a tilted Gaussian), it introduces a bias. From this exercise, it is rather clear that there is no interest to increase the number of fitted points.

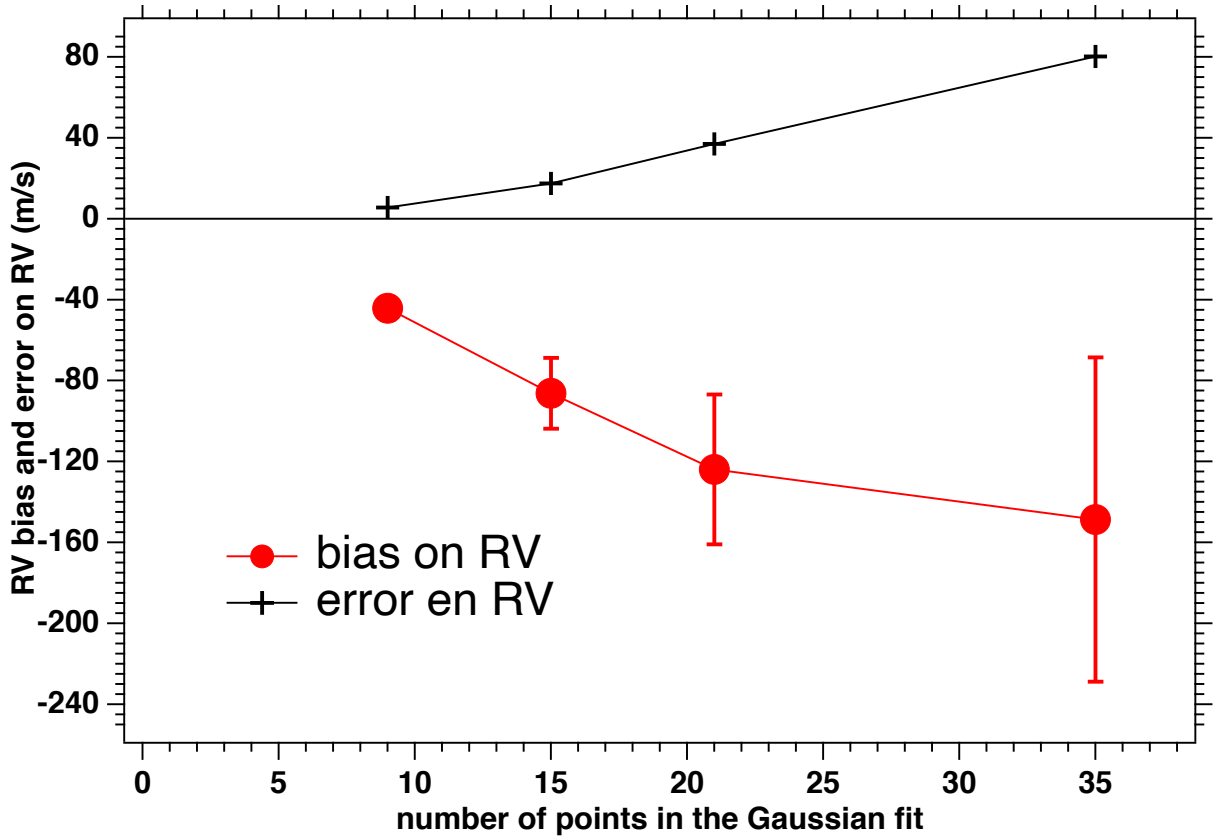


Figure 6.14: Bias and error on RV retrieval when a tilted Gaussian is fitted by a flat Gaussian, as a function of the fitted number of points.

Number of points	Bias (ms^{-1})	Error (ms^{-1})	χ^2 per point
9	-44.26	5.6	9.5e-07
15	-86.28	17.4	2.4e-05
21	-123.90	36.99	1.3e-04
35	-148.70	80.19	6.7e-04

Table 6.1: Table of bias and error in accordance to the number of point used for fitting.

Fitting a tilted Gaussian by a tilted Gaussian. We have also built a dedicated function

representing a tilted Gaussian, with free tilting parameters (coefficients of the linear function representing the continuum). The parameters are retrieved with an excellent accuracy (normal), as shown on Figure 6.15. The fit is also nearly perfect when using 9, 15, 21 and 35 points.

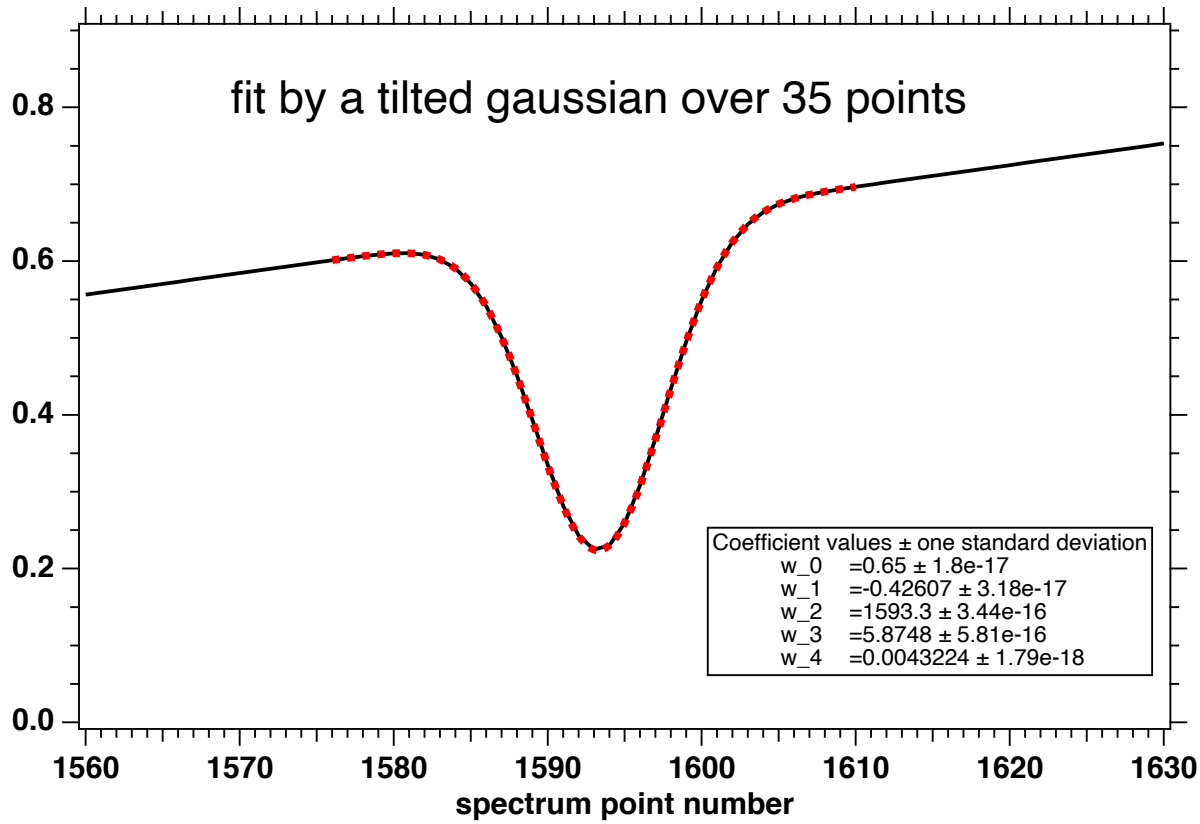


Figure 6.15: The same model of a tilted Gaussian is fitted over 35 points by a tilted Gaussian with a dedicated function. The parameters are retrieved with an excellent accuracy (see the $\pm$ in the frame), and there no bias at all. The fit is shown for a fit over 35 points.

The original flat Gaussian (before tilting) has its minimum at pixel 1593.3 and the minimum of the tilted Gaussian fitting is found at pixel 1593.3

Fitting an actual CCF by a flat Gaussian and a tilted Gaussian

In the previous section we studied an ideal Gaussian spectral profile (or CCF). Now we explore the problem of bias with real data, and the fit of the CCF built up from piling up all CCFs of one single order, with appropriate weights. We have selected order 104 as a test case. If fact, we cannot determine a bias from the absolute true radial velocity of the star, because we do not know it. However, we may determine the retrieved RV value, as a function on the number of points used in the fit (relative biases). Also, we study two ways to fit the CCF: either fitting with a flat Gaussian, or with a tilted

Gaussian, as shown on figures 6.16 and 6.17, with fits of one single exposure (n^{o}0) of order 104.

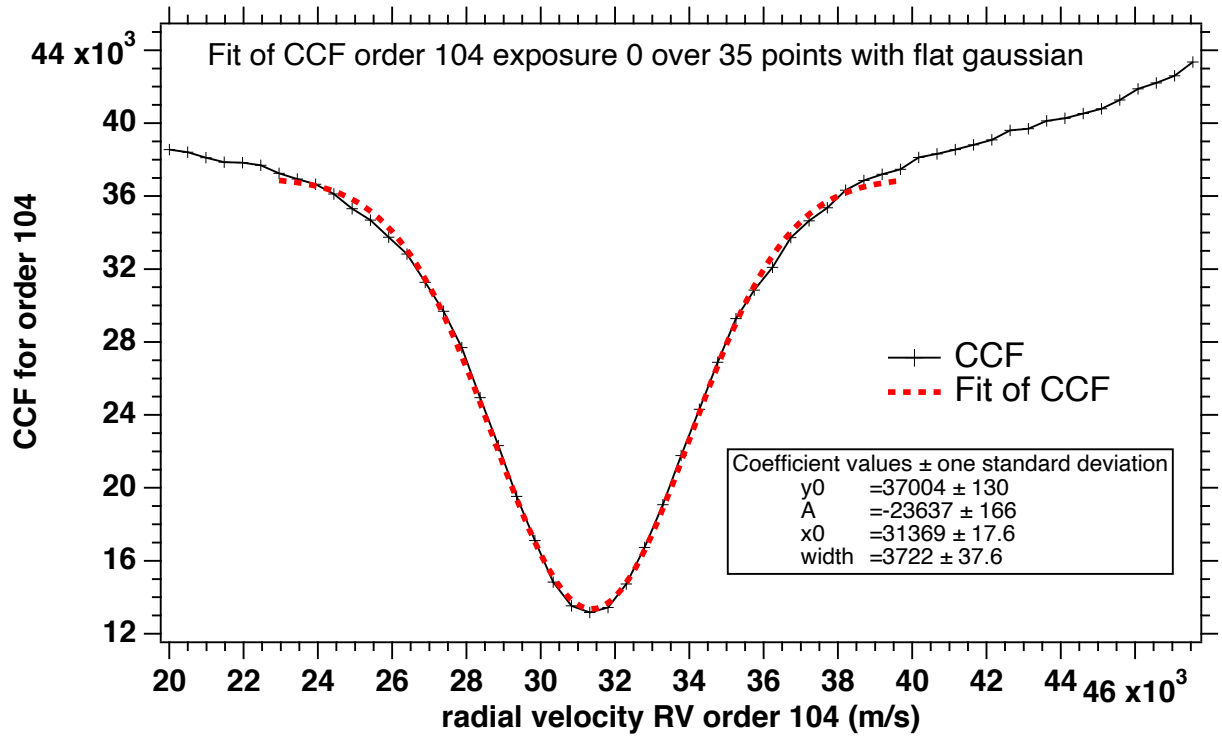


Figure 6.16: Black line and +: the CCF of order 104 and exposure 0 of the series of ESPRESSO measurements. Red dashed line: the best fit over 35 points by a flat Gaussian.

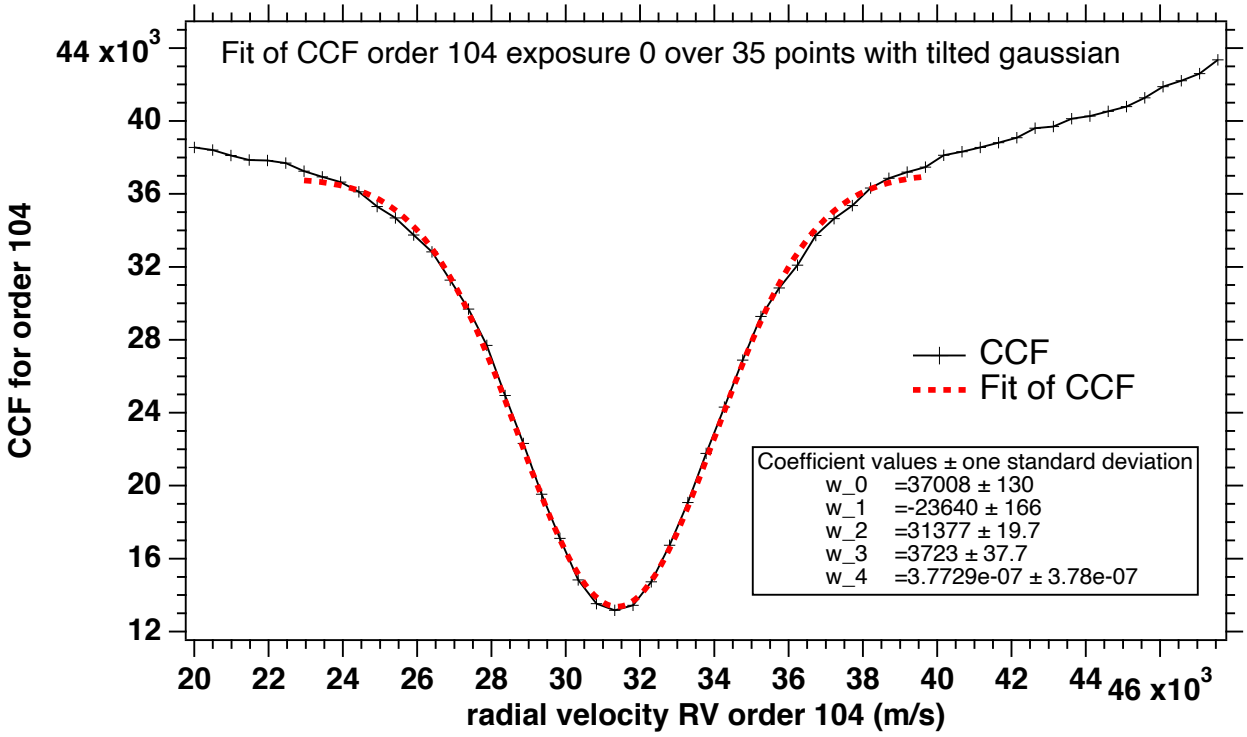


Figure 6.17: Same as 6.16, but with a fit by a tilted Gaussian

In both cases, some departures are seen on the shoulders of the CCF curve. In order to better characterize the two fitting methods (flat or tilted), they were applied to the whole series of 200

measurements of ESPRESSO on star HD40307 during the night of 25 December 2018 which were used for telluric corrections, and the average value of some retrieved parameters were compared for the two methods.

On figure 6.18 are plotted the average radial velocity (only for order 104) for the two methods, as a function of the number of fitted points. The error bars are the average of the returned error bars divided by $\sqrt{200}$=14.1

In the case of a fit with a flat Gaussian, the retrieved RV varies over a (rather small) range of $\sim$6 ms^{-1}, from 31369.4 to 31375.2 ms^{-1}. On the contrary, with a fit with a tilted Gaussian, there is a systematic trend where the retrieved RV increases with the number of fitted points, with a span of $\sim$ 20 ms^{-1} from 9 to 35 fitting points. We interpret this feature by the fact that a flat Gaussian is symmetrical, while the tilted Gaussian is more flexible, and captures (at least partially) any asymmetry of the CCF, which reflects the asymmetry of spectral lines: this is linked to the bisector effect, and physically is linked (at least partially) to the stellar granulation, as discussed by many authors (e.g, (González Hernández et al., 2020)).

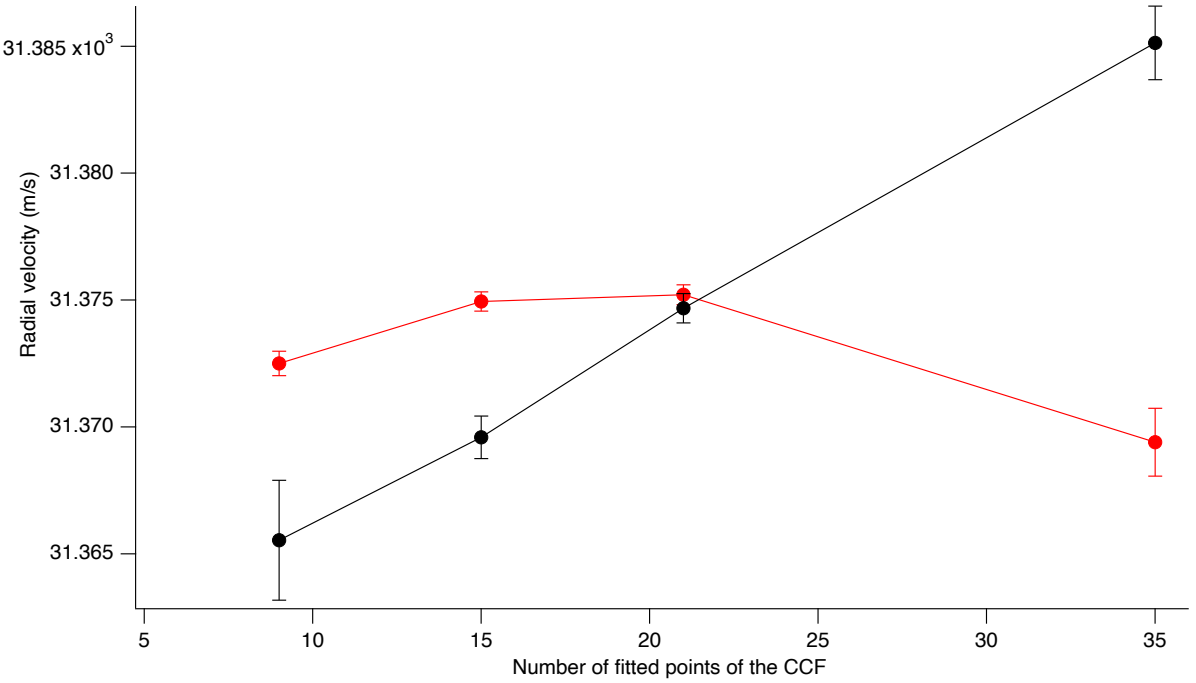


Figure 6.18: Red circles: RV retrieved with a flat Gaussian (with error bars), as a function of the fitted number of points. Black squares: RV retrieved with a tilted Gaussian.

From the whole series of the 200 measurements with order 104, yielding series of retrieved RV and error bar on RV, we may determine the average error and the standard deviation of the RV series. They are plotted as a function of the fitted number of points on figure 6.19, for the two cases: fit with

a flat Gaussian, and fit with a tilted Gaussian.

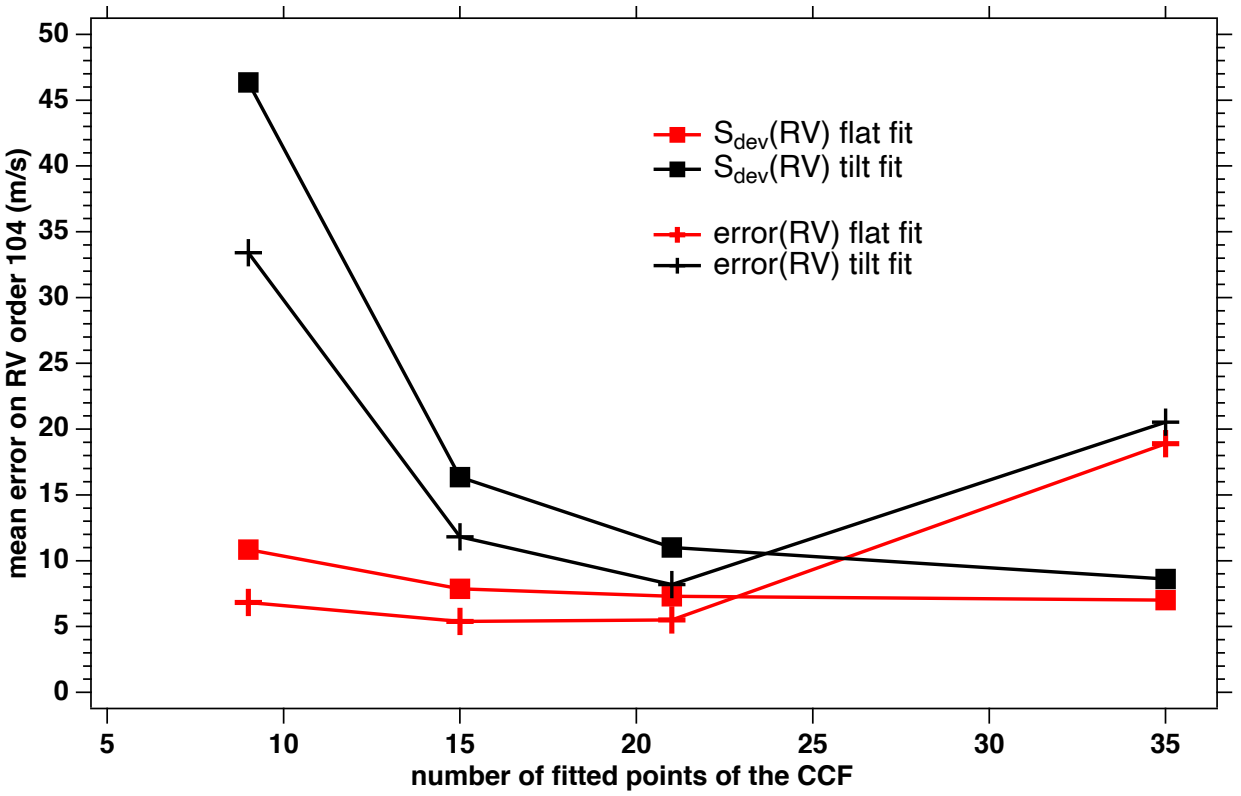


Figure 6.19: Squares: standard deviation S_{dev} (ms^{-1}) of the series of 200 retrieved values of RV, as a function of fitted number of points. Crosses: average returned formal error. Red colour is for the fit with a flat Gaussian, black is for the fit with a tilted Gaussian.

We can conclude:

- The standard deviation of the series of the 200 RV retrieved from order 104 is larger than the average error, except for both fits over 35 points, with a rather large error. The standard deviation for the fit with a tilt Gaussian decreases from 46 ms^{-1} for 9 fit points to 16 ms^{-1} for 15 fit points, to 11 ms^{-1} for 21 fit points and to 8 ms^{-1} for 35 fit points. Errors are also decreasing with increase of number of points, up until fit over 35 points. In case of fit with a flat Gaussian standard deviation is also decreasing with increase of fit points, but not as much. For the fit over 9 points it will be 10 ms^{-1}, for the fit over 15 points – 7.8 ms^{-1}, for the fit over 21 – 7.3 ms^{-1} and for the fit over 35 points – 7 ms^{-1}. Errors are almost equal for fits over 9, 15 and 21 points, but we also see the increase of error for the fit over 35 points.

- Both the error and the S_{dev} are larger for the tilt fit than for the flat fit. One likely explanation is that since the CCF is very flat, and adding a slope of the continuum is just degrading the accuracy of the retrieval, by adding one more degree of freedom in the fitting process.

- Considering the flat fit, selecting a fit over 15 or 21 points are providing equally small error bars, slightly smaller than the fit over 9 points. There are two contradictory factors. Increasing the number of points for the fit implies more data to be fitted, and the accuracy should increase. But since the CCF shape departs from a pure Gaussian (because of the asymmetry of spectral lines), the χ^2 per point increases with the number of points and as a result the formal error increases with the number of points. We note that, on the contrary, the standard deviation of the RV series continues to decrease when going form 21 to 35 points (but not as much).

6.5 The bias due to Rayleigh extinction.

In this section we estimate numerically the RV bias which is due to atmospheric Rayleigh extinction, when the stellar spectrum is not corrected for this attenuation effect.

Here we consider a spectral line with a Gaussian shape, imprinted on a stellar continuum constant at value 1. The Rayleigh extinction changes the stellar continuum to the value Tr(λ), which has a slope, and which moves the minimum of the spectral line to a new position, introducing a bias. We found out that the RV bias is a function of the width of the line, the depth of the line, and the logarithmic slope of the continuum. With the λ^{-4} dependence of Rayleigh optical thickness, we expect that the bias will be much greater in the blue than in the red. Furthermore, it will also vary with the air mass, and this is a problem, because when making several measurements of the same star to find an exoplanet, the measurements are made at various air masses, introducing different biases. With X = air mass, the atmospheric Rayleigh optical thickness from the telescope to the star may be written:

$$\tau(X,\lambda) = X\tau(1,\lambda_0)(\lambda_0/\lambda)^4 \tag{6.9}$$

where $\tau(1,\lambda_0)$ is the Rayleigh optical thickness at the reference wavelength λ_0, often taken at 5000 or 5500 Å.

The transmitted flux, after Rayleigh extinction is :

$$Tr(X,\lambda) = exp(-\tau(X,\lambda) \approx 1 - \tau(X,\lambda) = 1 - X\tau(1,\lambda_0)(\lambda_0/\lambda)^4 \tag{6.10}$$

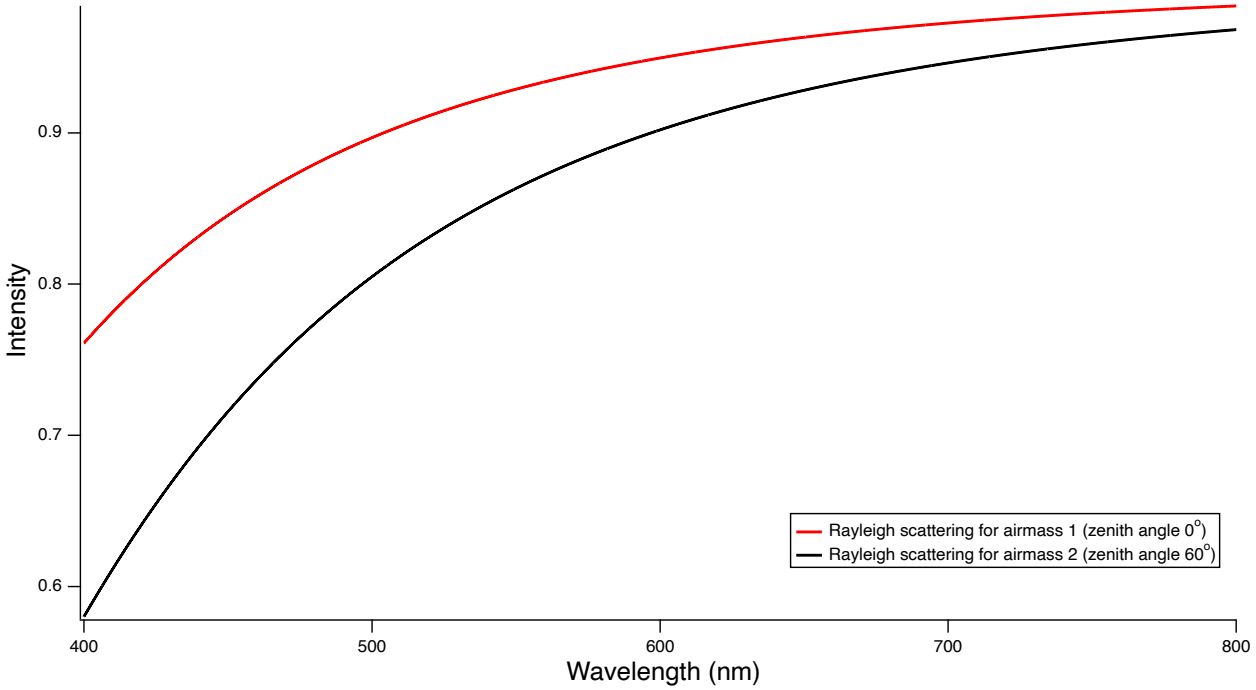
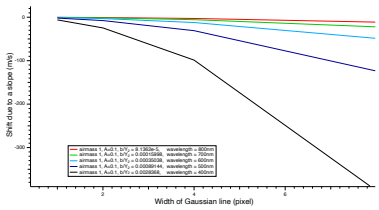
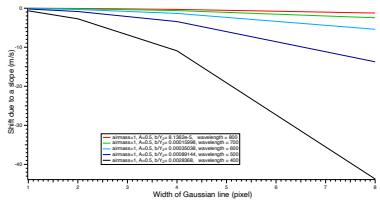


Figure 6.20: Rayleigh scattering computed by TAPAS for the night of observation t ESO/VLT. In red – Rayleigh scattering for airmass equal 1 or zenith angle 0^0, in black – Rayleigh scattering for airmass equal 2 or zenith angle 60^0.

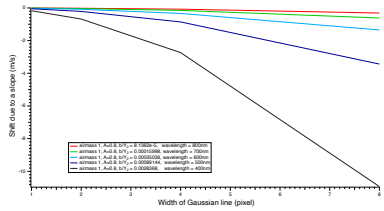
We can compute bias caused by Rayleigh scattering.



(a) airmass=1,A=0.1

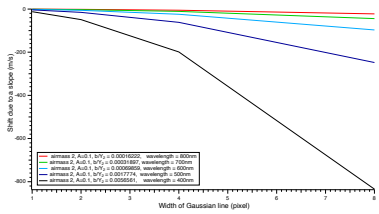


(b) airmass=1,A=0.5

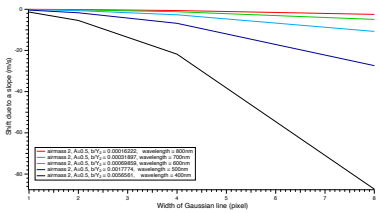


(c) airmass=1,A=0.8

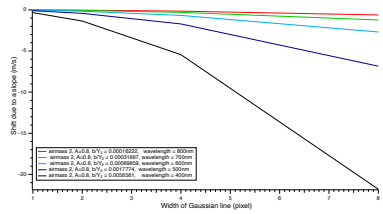
Figure 6.21: Evaluation of bias induced by Rayleigh scattering for airmass=1



(a) airmass=2,A=0.1



(b) airmass=2,A=0.5



(c) airmass=2,A=0.8

Figure 6.22: Evaluation of bias induced by Rayleigh scattering for airmass=2

First of all, for wavelengths for which Rayleigh scattering is not negligible, especially for 400 nm, the shift caused by it can achieve similar level as the shift due to the blaze. We have to mention 2 points: first, the shift is decreasing with increase of line depth, and second, the shift is increasing with

increase of airmass.

6.6 Conclusions

After the analysis presented in this section we can conclude that for our purpose it is possible to use non-deblazed data.

As reasons for this conclusion we can list:

- For our current work we are using CCF method, for a whole order, so the shift due to the blaze slope should be compensated because of symmetry
- Dataset or HD 40307 is one-night observations, for such short period of time spectrum of star is not displaced, so change of blaze function with wavelength is negligible
- It was shown that blaze function provided with official pipeline produce artificial slope into data and data should be additionally corrected after de-blaze
- Shifts produced by blaze function are smaller than the one produced by the biases of laboratory measurements of spectral line positions

The study of the effect of Rayleigh extinction came rather late in the course of this PhD. work, and we did not included it in our analysis of HD 40307. Still, since the effect is large for the blue part of the spectrum, and for weak and wide spectral lines, it is recommended to account for the Rayleigh extinction and correct the spectra, which is easily done by using the TAPAS on line web service.

Chapter 7

Study of Radial Velocities on BTA SAO RAS spectrograph

This chapter is dedicated to the RV study based on the data obtained by FFOREST/BTA SAO RAS spectrograph. As a member of the team, I was involved in processing the observational data and further analyzing it. .

7.1 Instrument

The 6-m Big Telescope Azimuthal (BTA) of SAO RAS is located at an altitude of 2070 m above sea-level near the Pastukhova mountain in Karachay-Cherkessia Republic (Republic of the Russian Federation, in north Caucasus). The high-resolution fiber-fed spectrograph FFOREST (Fiber-Feed Optical Russian Echelle SpecTrograph) with resolving power 50 000-100 000 has been constructed several years ago (Valyavin et al., 2014, 2015, 2020); the optical configuration of the spectroscopic unit is so-called "white pupil". Originally "white pupil" or "Pupille Blanche" was proposed by Baranne (1988).

In classical spectrographs the camera aperture is larger than the disperser size, proportionally to the field size. With increase of angular size one has to use a large aperture for a camera, which is difficult, and cameras start to have complicated design due to the fact that pupil is far from camera and it generates vignetting. In the case of "white pupil" such a problem disappears, the disperser is re-imaged onto the camera and monochromatic beams intersect at the new pupil which is superimposed on the camera entrance. This is done in BTA SAO RAS spectrograph by 2 off-axis collimators; one

of the collimators operates with an echelle grating in the autocollimation mode; the second collimator forms the pupil plane at its focus by constructing an undispersed image of the echelle grating there (Valyavin et al., 2014). The "white pupil" configuration requires a more complex optical system than in a classical approach, but it has advantages that make it widely used in high-precision and high-resolution spectrographs. For example, such an optical scheme allows to reduce the size of the spectrograph, and it allows to use a focusing camera with a smaller aperture than with a traditional optical mounting. This cuts the cost of spectrograph, because the focusing camera is one of the most expensive part of a spectrograph. The second advantage is the minimization of scattered light. At first "white pupil" was used at ELODIE (OHP), CORAVEL (ESO) and now it can be seen almost in every spectrograph, for example UVES, HARPS, ESPRESSO (ESO/VLT) or FFOREST (SAO).

It is possible to use two fibers ducts simultaneously, which allows simultaneous observations of the target spectrum and thorium-argon (Th-Ar) lamp for wavelength calibration; or observations of the target spectrum and sky-background spectrum, similarly to the HARPS realization. The spectrograph is located at the foot of telescope due to the almost absence of mechanical vibrations (the foundation is affected only by telescope motion), the temperature remains constant to within several tenths of a degree over several days. The spectrograph is not in vacuum, but in sarcophagus with a system to dampen sudden pressure and temperature fluctuations. The index of refraction of air is a function of pressure and temperature (actually of the air density p/kT) and therefore the positions of the lines on the detector are changing with pressure and temperature, mimicking spuriously a Doppler shift, if not accounted for. With the Edlen's formula giving the index of refraction of air, we computed that a variation of 1 mbar around 1013 mbar would correspond to an apparent Doppler shift of 82 ms^{-1} at constant temperature, while a variation of 1^{o}C at constant pressure would correspond to a shift of 304 ms^{-1}. This imposes very stringent temperature and pressure control systems to get a reasonable stability of the spectrograph. Those spurious shifts may be measured by a very accurate wavelength calibration system.

The operating wavelength range of FFOREST is from 4000 to 7500 Å; the scheme of the spectrograph allows further upgrades to extend its operating wavelength range up to 10000 Å. For wavelength calibration Th-Ar lamps are used, and halogen lamps are used for flat-fielding. There is a plan to use a laser frequency combs (LFC) for wavelength calibration, since it is a more advanced method and allows to provide precise calibration along the entire wavelength range.The spectrograph is operating,

but still there are plans for improvement and changes. For example, one of the plans is to implement a Fabry-Perot system into the design of the spectrograph to improve the accuracy of the calibration.

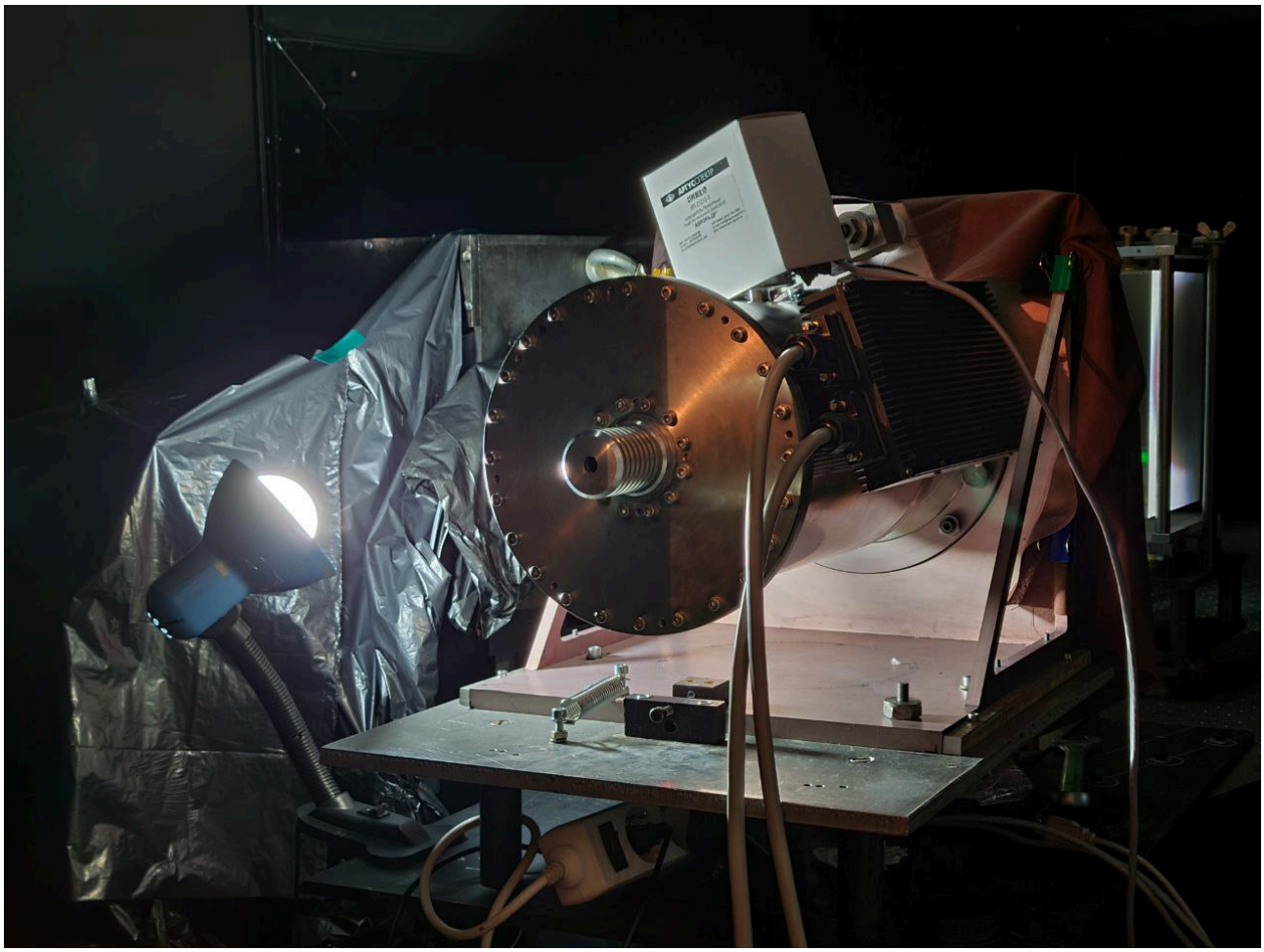


Figure 7.1: FFOREST spectrograph. Photo taken on 23 December 2021 during conference at SAO RAS.

7.2 Data

HD 42176 (Kelt-2A) is a F7V type star with visual magnitude 8.71 in a binary system with a companion Kelt-2B. It was observed during one night on 1st January 2021 by the FFOREST spectrograph. Data were obtained with a spectral resolution around 70 000. The main aim objective of this observations was to observe one transit of exoplanet Kelt-2A b and study the Rossiter-McLaughlin (RM) effect[1]; 25 measurements were obtained during one night, one of the measurements was affected by the weather conditions (clouds) and is not useful for radial velocities measurements. Kelt-2A b is a hot Jupiter with mass 1.524 M_J transiting the primary star of the binary system (Beatty et al., 2012). According

[1] PI: Valyavin

to Beatty et al. (2012) the total duration of a transit is 0.212 days, which is approximately 300 minutes and the RM effect has and RV semi-amplitude 160.5 ± 8.7 ms^{-1}.

7.3 Usage of so-called "first" formula of Connes

The Pierre Connes algorithm is a good tool to evaluate the stability of a spectrograph. This was done for SAO RAS spectrograph after the first observations. As it is described in chapter 7.1, this spectrograph is not in vacuum and thermal stability is provided by the general thermal stability of the room at foundation of telescope housing. From time to time, to cool down the CCD detector assembly, one member of the observation team has to put liquid nitrogen into the Dewar container. Such interventions may affect the thermal and mechanical stability of spectrograph, and this is why it is important to make precise calibration corrections and evaluate the differences between observation epochs. As a reference for wavelength calibration SAO RAS is using Thorium-Argon (Th-Ar) lamp, observations of Th-Ar are made simultaneously with the target star spectrum, but fed to a second fiber, allowing to track position of given wavelengths on CCD pixels at the exact moment of observations. We applied as we call it the first formula of Connes for 11 exposures of Th-Ar taken by spectrograph of SAO RAS; the first 4 are taken during night of 31 December 2020 from 14:39 to 14:46; 1 at night of 1 January 2021; 2 January 2021 and 3 January 2021; 3 at night 4 January from 14:25 to 15:45 and the last exposure was done during night 6 January 2021. As a reference exposure we used the first exposure of 31 December, we choose one strong line from each grating order to track its displacement. For each order we calculated the shift in ms^{-1} (between 2 exposures) and error. We kept most usually one line per order. After that we made a weighted average of all selected lines on all orders (a total of 35 lines) over 35 orders. Any displacement detected as a fraction of pixel may be converted into a shift in ms^{-1}, with the approximate scale of 1.9 kms^{-1} for a displacement of one pixel. At 4877.7613 Å, one pixel covers 1888.75 ms^{-1} (per pixel).

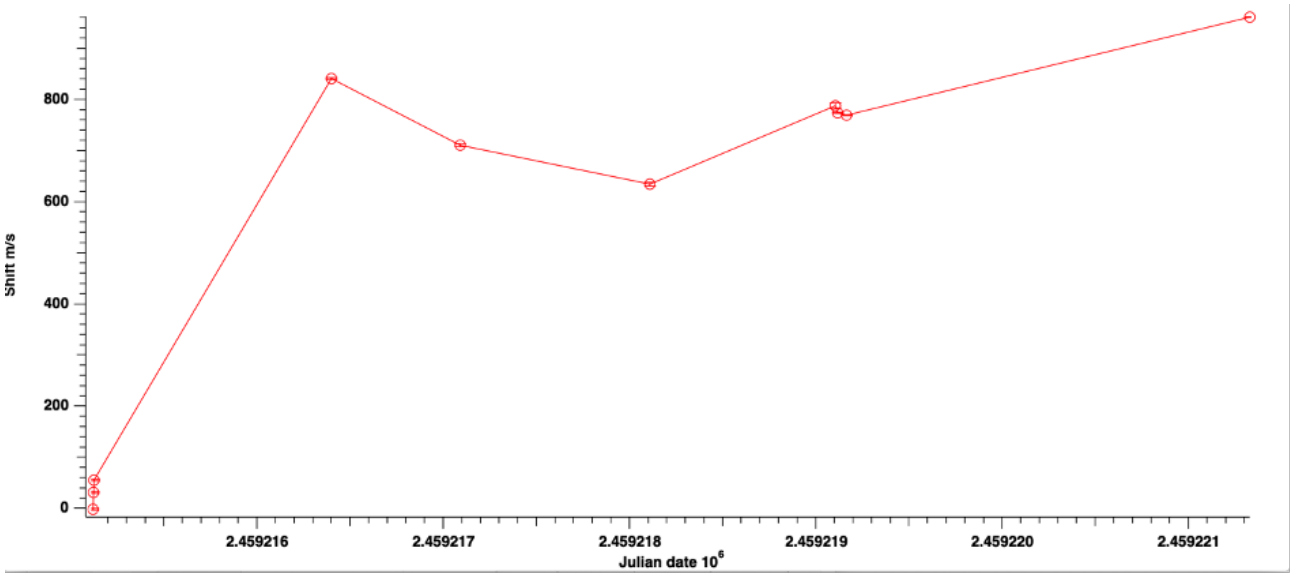


Figure 7.2: Mechanical displacement of Th-Ar calibration lines of SAO RAS spectrograph over 7 days. Bottom axis – Julian Date. The last digit change by one unit each day. Vertical axis – the mechanical shift (averaged over 35 grating orders) is given in ms^{-1}, with an approximate scale of 1.9 kms^{-1} per pixel displacement.

This is a limited data set, which still allows some remarks. The 4 exposures of the first day show a shift of $\sim$60 ms^{-1} in 5 minutes. But apparently this trend must have slowed down afterwards. Probably an effect of starting the operations of the spectrometer, with unstabilized thermal situation of the instrument at start, which may be linked to the filling of liquid Nitrogen into the cooling system of the CCD detector. The shift in wavelengths is 840 ms^{-1} over 1.5 days. Such shifts are crucial to be measured accurately for velocity measurements for exoplanetary detection or characterization. The whole wavelength calibration of the cross-dispersed spectrograph (or wavelength assignment to every pixel) depends on it.

7.4 Rossiter-McLaughlin effect

The Rossiter-McLaughlin effect is a subtle spectroscopic effect which was predicted by Holt (1893) and first observed by Rossiter (1924)) and McLaughlin (1924). It was predicted in the context of eclipsing binary systems of stars, which can be detected by the sudden change of total light when one star is passing in front of the other star. Assuming for simplicity that the rotation axis of the eclipsed star is in the plane of the sky, there is one half of the star which is approaching, while the other half is receding (Figure 7.3). Assuming in addition that the orbital motion of the eclipsing star around the other is prograde (in the same direction as the rotation axis (spin) of the eclipsed star), we may foresee

the following situation. When the eclipsed star is entering eclipse, the receding limb is visible and the approaching limb is covered by the companion; the spectral lines of the eclipsed star are shifted to the red part of spectrum, and the overall radial velocities of all lines are positive; at the center of eclipse the lines are symmetrical, and after that in the second part of the eclipse the receding limb is covered and the approaching limb is visible, lines are shifted to the blue part of spectrum and radial velocities are negative.

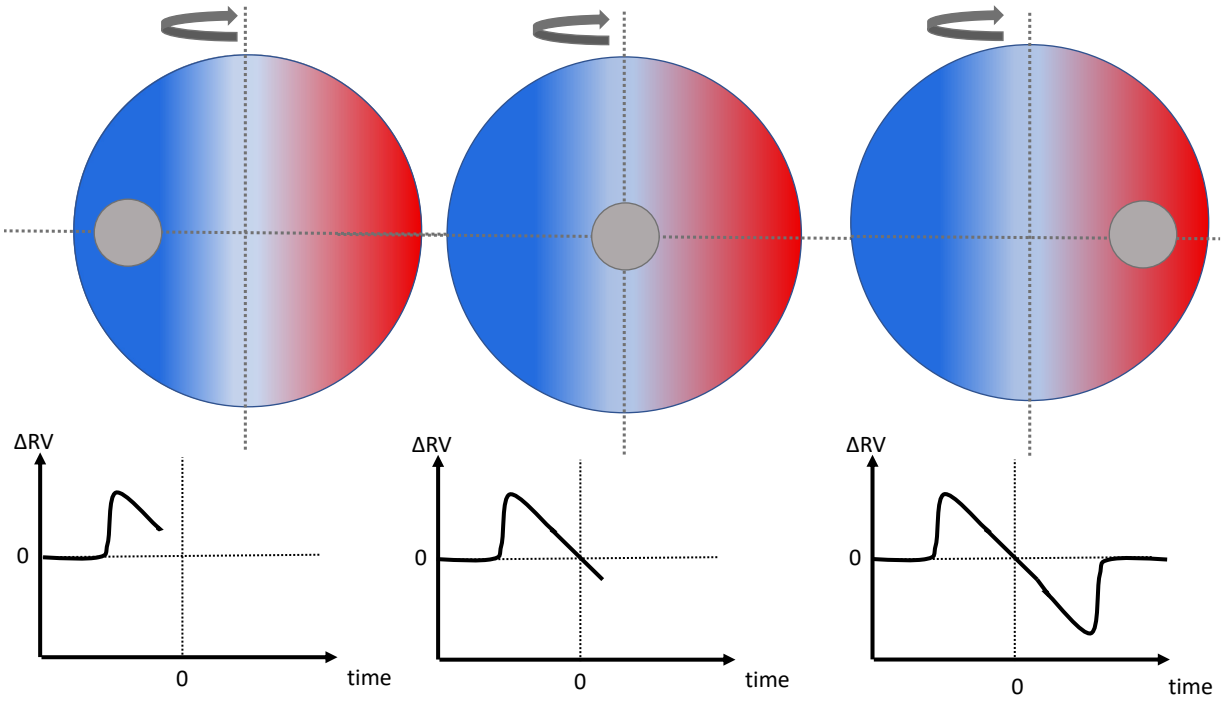


Figure 7.3: Illustration of the Rossiter-McLaughlin effect.

Rossiter (1924) made the analysis of Beta Lyrae eclipsing binary system observations; he found that with a first least squares adjustment of orbital elements, two sets of large residuals appeared during 1.6 days, one preceding and one following the principal minimum. The first set of residuals was positive and the second was negative. This was the indicator of a secondary oscillation, which was not due to orbital motion. A similar analysis was done by McLaughlin (1924)) for Algol star. He found out that this so-called rotational effect can be used as a method of determination of the actual scale of the eclipsing system. Also, it is possible to measure the sky-projected angle between the spin and orbital axes.

In the case of exoplanets, the effect was first detected by Queloz et al. (2000): it was done on the data of HD 209458 star and its hot Jupiter exoplanet. For star HD 80606 Hébrard et al. (2010) made the analysis of Rossiter-McLaughlin effect with SOPHIE (OHP) observations and concluded that

the scenario with aligned spin-orbit can be rejected. This fact, along with the analysis of twenty other systems leads to the conclusion that there are two different evolution scenarios, depending on the gravitational interaction with a third body. WASP-17b was the first exoplanet for which a retrograde orbit was detected (Anderson et al., 2010), using the shape of Rossiter-McLaughlin effect. The Rossiter-McLaughlin (RM) effect has an important role in the frame of planet migration theory: since the misaligned hot Jupiters migrated via few-body gravitational interactions, it is possible that all hot Jupiters migrated in the same way.

As a part of work with BTA SAO RAS spectrograph the full cycle of extraction and analysis of data was done. The RM effect study was done on the data for Kelt-2A 7.2. There is no automatic pipeline for this SAO spectrograph. My personal contribution was in the processing of all the spectrograph data, which consisted of: each exposure was manually extracted from CCD image, calibrated with Th-Ar series taken simultaneously with a target star, flat-fielding was done by the flat field map, data was cleaned from cosmic rays and bad pixels removed or "repaired". Then the first spectrum was taken as a template (instead of a mask) for a cross-correlation function method and the radial velocities were calculated. All this work was done "manually" using DECH software (Galazutdinov, 2022), in an interactive way. The result of the study is published (Valyavin et al., 2022).

To begin with, the observed RV-curve has a classical RM shape (Figure 7.4, top graph). But there was a problem with the duration of the observed effect. In section 7.2 it is mentioned that Beatty et al. (2012) found a duration of transit to be around 300 minutes, while the duration obtained from first fitting is around 200 minutes (see figure 7.4, top graph). To investigate the cause of such a difference, the additional observations by the EXPANSION network (Sokov et al., 2018) were requested, and it was shown by the transit photometry curves that the duration of transit is indeed around 300 minutes. Therefore in a second attempt the RV time series was fitted with the assumption that it may contain some residual systematic variations. A best fit is obtained by the combination of the RM effect and a single sinusoidal variation (Figure 7.4, 2^{nd} and 3^{rd} graphs). The RM effect has a semi-amplitude of 92 ± 10 ms^{-1}, which is consistent with preliminary estimates from spectral line widths (20 ± 8 kms^{-1}). The sinusoidal variation has a the period of $\sim$ 3 hours and a semi-amplitude of 63 ± 13 ms^{-1}. The RV residuals which remain after subtracting this two-components model reveal no other variations and r.m.s. is about 40 ms^{-1} (Figure 7.4, bottom graph). The fit was computed using PlanetPack software (Baluev, 2013, 2018). The nature of sinusoidal 3-hours variation is still unclear, and the observations

of Kelt-2A are continuing. The result of this analysis is published in (Valyavin et al., 2022).

Still, in view of the RV effect that we detected attached to the presence of (almost) undetectable optical fringes in the spectra collected with ESPRESSO, we might think that weak optical fringes may also play a role in these observations, and create sinusoidal trend.Another suggestion is that during its transit over the disc, the planet may have masked an active area that was decreasing the convective GBS. Observations must be repeated along with simulations.

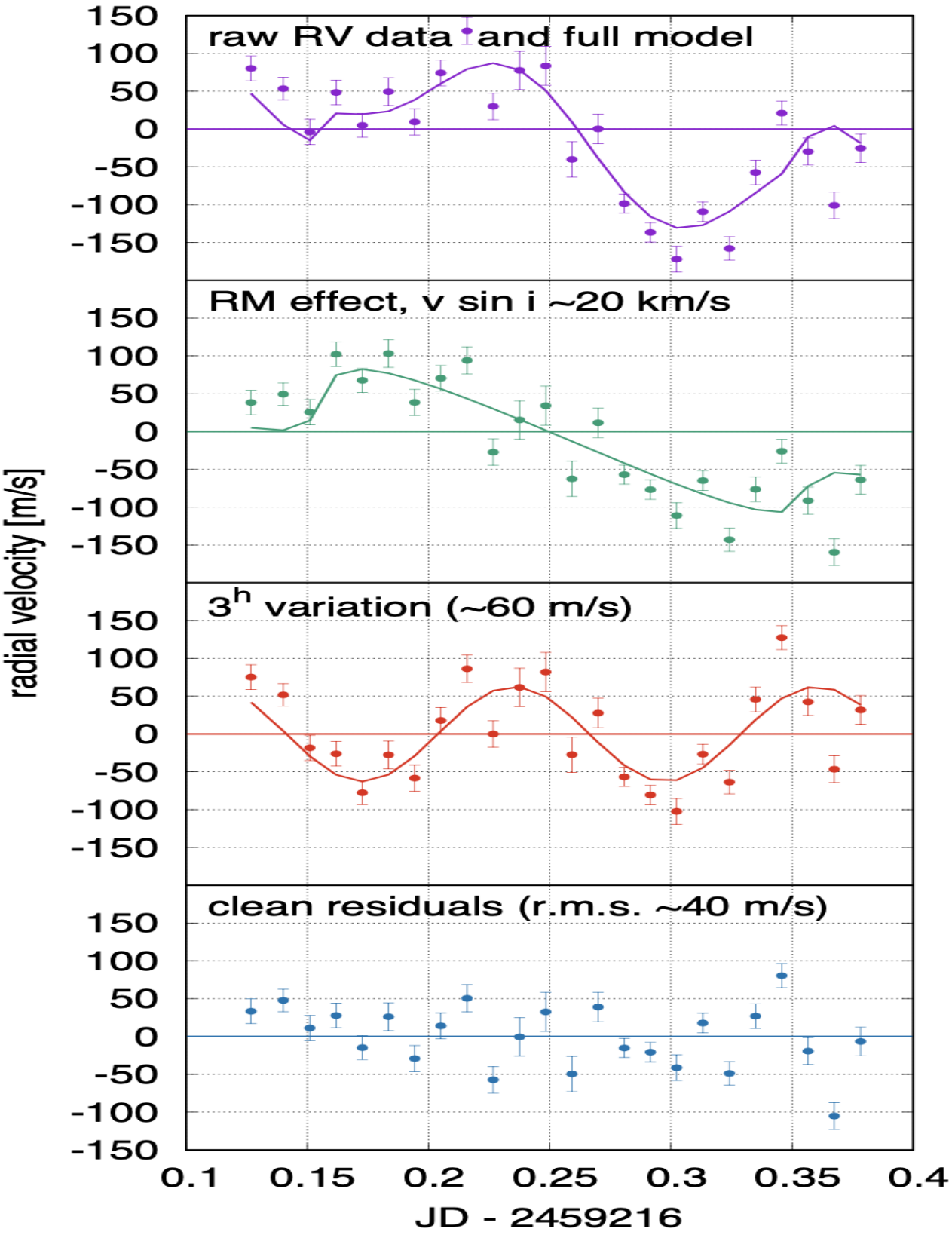


Figure 7.4: Top graph: raw RV data; 2^{nd} graph: model of RM effect; 3^{rd} graph: model of 3h sinusoidal variations; bottom graph: RV residuals remain after subtracting 2-component model. Figure from Valyavin et al. (2022)

Chapter 8

Exoplanetary atmospheres

The atmospheres of exoplanets can be studied with various methods, for example transmission spectroscopy, occultation spectroscopy, reflection spectroscopy or a full-orbit phase curve. During my work I focused on transmission spectroscopy.

Using the model from Marcq et al. (2017) I worked on a forward transmission model, in order to simulate transit observations by the James Webb Space Telescope or ARIEL.

In this chapter we will discuss transmission spectroscopy in 8.1, the diversity of exoplanets focusing on hot telluric exoplanet atmospheres as modeled by Marcq et al. (2017) in 8.2, and models for forward modeling, both already available and including my own in 8.3.

8.1 Transmission spectroscopy

8.1.1 Method

One of the methods to study exoplanetary atmosphere is transmission spectroscopy. During the transit an exoplanet passes in front of the host star disk and blocks part of the stellar light, what is called primary transit. The geometry of transit is shown on figure 8.2. The fraction of the light blocked by the exoplanet, resulting in a slight drop in total brightness, is equal to the effective area of the planet's sky-projection relative to the area of the star and is called the transit depth δ (assuming uniform brightness over the whole stellar disk. It works well away from the limb).

$$\delta = \frac{{R_p}^2}{{R_s}^2} \tag{8.1}$$

where R_p is a planet radius and R_s is a star radius.

Since wavelength-dependent opacities (e.g. absorbing gases, aerosols) in the planet atmosphere absorb stellar light at different effective planet radii a transit spectrum is a variation of transit depth as a function of wavelength.

$$\delta(\lambda) = \frac{{R_p}^2(\lambda)}{{R_s}^2} \tag{8.2}$$

Figure 8.1 shows an example of a transmission spectrum of the cloudless atmosphere of an exoplanet GJ 1214 b, figure is taken from Gandhi et al. (2020). The authors used AURA model (Pinhas et al., 2018) to produce transmission spectrum.

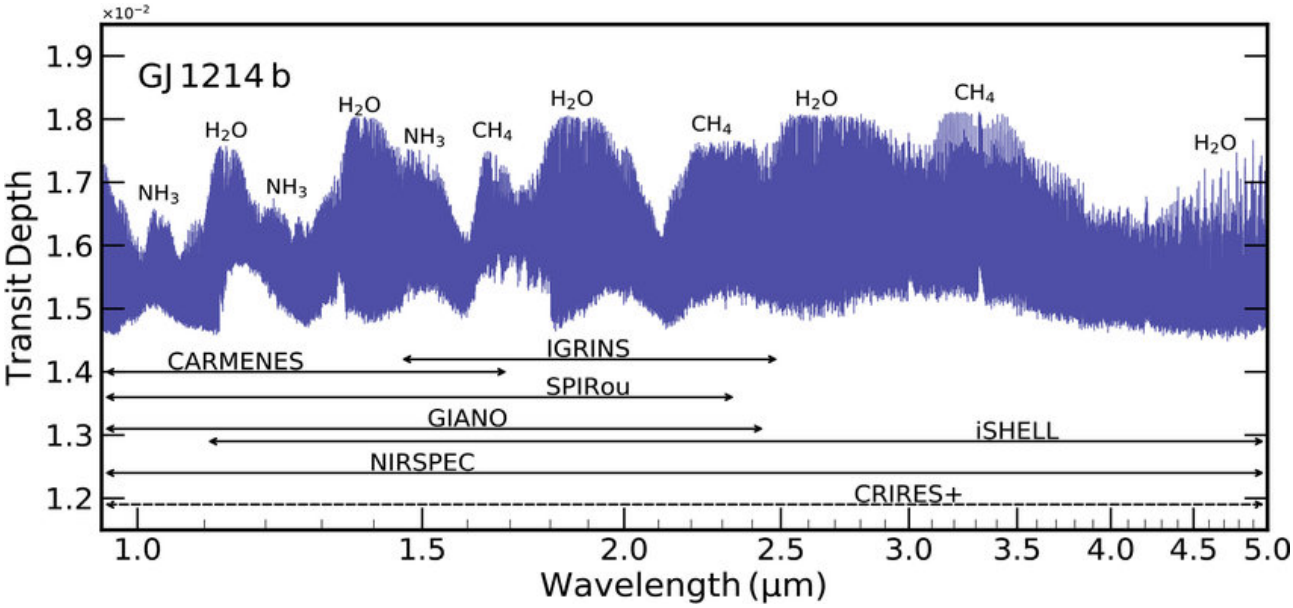


Figure 8.1: Example of transmittance spectrum. The simulated cloud free transmission spectrum of GJ 1214 b showing the transit depth with wavelength. Figure is taken from Gandhi et al. (2020)

On Figure 8.2 H stands for, the atmospheric scale height which is the change in altitude over which the pressure drops by a factor of e. It is calculated with assumption of hydrostatic equilibrium and ideal gas law:

$$H = \frac{k_B T}{\mu g} \tag{8.3}$$

where k_B is the Boltzmann constant, T is the planet's mean temperature, μ is the mean molecular mass and g is the local gravity acceleration.

Using the scale height, it is possible to predict the typical size of transmission features (in terms of amplitude).

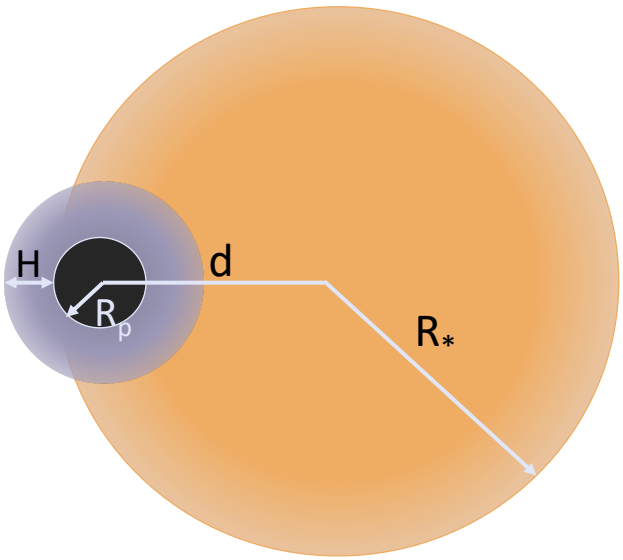


Figure 8.2: Illustration of transit geometry. Light from a star passes through the planet's atmosphere. R_p and R_s are planet and stellar radii respectively, H is the atmospheric scale height, d is the separation of centers in the plane of the sky. Figure is analogue to the one from Kreidberg (2018)

According to Kreidberg (2018), amplitude of transmission features can be expressed on the basis of transit depth as a function of wavelength

$$\delta_\lambda = \frac{(R_p + nH)^2}{R_s^2} - \frac{{R_p}^2}{{R_s}^2} \approx \frac{2nR_pH}{{R_s}^2} \tag{8.4}$$

where n is the the number of "opaque" scale heights at wavelengths with high opacity. According to Stevenson (2016) $n \approx 2$ for cloud-free atmospheres at low spectral resolution.

8.1.2 Instruments

Observations of exoplanet atmospheres are possible both from the ground and from space.

Ground-based observations In case of ground-based observation, as it was already described at chapter 1, Earth atmosphere is unavoidable and can cause systematics. The line of sight properties as airmass, clouds and precipitable water vapor can affect observations and produce some trends in the light curve, which can drown out the signal from the exoplanet atmosphere. One of the

possible solutions is to observe simultaneously target star and one or few additional stars with similar brightness and in the same field of view. For these additional stars, which are also called reference or comparison stars, the optical paths pass through similar parts of the Earth's atmosphere, so the systematics and trends caused by the Earth's atmosphere will be almost identical as one for the target star. If one divides the target star light curve by the sum of the reference star light curves it will remove most of systematics. Such approach is quite widely used, for example Nikolov et al. (2021) used it for FORS2 VLT data.

A second way to obtain transmission spectrum of exoplanet atmosphere from Earth's surface is to use high-resolution spectroscopy. The main idea is that the telluric spectrum and exoplanet atmosphere spectrum are Doppler shifted because of the exoplanet orbital motion, so exoplanet atmosphere spectrum is separated from telluric spectrum and stellar spectrum and it is possible to cross-correlate it with a spectrum template to reveal the atmospheric composition and orbital velocity. As a ground-based spectrographs we can highlight SPIRou (Donati et al., 2020), for example in Boucher et al. (2021) the HD 189733 b atmosphere was characterized by analysis of SPIRou data. CRIRES (Kaeufl et al., 2004) and CRIRES$^+$ (Holmberg and Madhusudhan, 2022) can be another example of spectrographs used for transmission spectroscopy from ground. Both SPIRou and CRIRES are working in infrared domain, in visible wavelength domain ESPRESSO was recently used (Tabernero et al., 2021), (Allart, 2021). Among future spectrographs NIRPS (Bouchy et al., 2022) can be distinguished.

Observations from space Observations from space are not affected by the Earth's atmosphere, so that the majority of systematics which cause problems in ground-based spectroscopy disappears. It is also possible to observe in wavelength domains which are highly contaminated by Earth's atmosphere. The Hubble Space Telescope (HST) and the Spitzer Space Telescope were the two telescopes in space which could observe atmospheres of exoplanets, up until last year, when the James Webb Space Telescope was launched. The Space Telescope Imaging Spectrograph (STIS) and the Wide Field Camera 3 (WFC3) on the board of Hubble are working from the ultra-violet to the near-infrared domain ($\sim$ 0.1μm to 1.7μm). The IRAC instrument on Spitzer covers from 3.5μm to 4.5μm. Combination of instruments working on board of JWST covers range from 0.6μm to 28.5μm (currently past 18 μm MIRI has problems with the low efficiency of the grating, low quantum

efficiency of the detectors, and rising thermal background). ARIEL (Tinetti et al., 2022) is the next space telescope dedicated to exoplanetary atmosphere observations, it should cover wavelength range from 0.5μm to 7.8μm (with resolving power R=20-100 depending on the wavelength sub-range).

8.1.3 Results of atmospheric studies

The first detection of absorbing species was using the sodium (Na) doublet at 589 nm in the hot-Jupiter HD 209458b in 2002 by Charbonneau et al. (2002). After this first discovery, Na (Nikolov et al., 2014, 2016) and potassium (K) (Sing et al., 2011) have been discovered in other hot-Jupiters. Atomic hydrogen has been discovered by Lyman-α observations with Hubble Space Telescope (Vidal-Madjar et al., 2003; Bourrier et al., 2013). Molecular hydrogen was proposed as a possible cause of Rayleigh scattering in the spectrum of HD 209458b (Lecavelier Des Etangs et al., 2008). Water produces a strong absorption feature around 1.4μm, wavelength covered by HST WFC3. It is the most detectable molecule for hot-Jupiters (Tinetti et al., 2007; Deming et al., 2013; Huitson et al., 2013; Kreidberg et al., 2014; Line et al., 2016; Tsiaras et al., 2018). It also can be detected in atmospheres of sub-Neptunes and super-Earths (Fraine et al., 2014; Tsiaras et al., 2019).

Using data from JWST researchers started to find new species, for example CO_2 (JWST Transiting Exoplanet Community Early Release Science Team et al., 2023), SO_2 in the atmosphere of WASP-39b (Tsai et al., 2023).

8.2 Telluric exoplanet atmosphere model description

According to the statistics, planets with radius from 0.5 to 3 $R_\oplus$ are the most common. And one of the most interesting type of planets with such radii are the terrestrial planets, since such planets can help us to understand how our planet evolved and why it is so different from e.g. Venus or Mars in our solar system. By studying such planets, it is possible to find which initial conditions are most important for atmospheric formation and evolution, e.g. if it depends more on planet size or on the received flux.

At some point during their formation, telluric exoplanets go through the so-called *magma ocean* stage: temperature is high enough for the mantle to be in a liquid (molten) state. During this stage, the planet can cool itself by radiating more heat than it absorbs from its host star. Atmospheres of

such exoplanets are expected to consist mostly of H_2O, CO_2 and small fraction of N_2 (Lupu et al., 2014),(Schaefer et al., 2016),(Turbet et al., 2020), (Bower et al., 2022).

In Marcq et al. (2017) authors found out that 2 cases for such exoplanet exist:

- If the surface temperature is higher than a threshold temperature T_ε, cooling is fast and only thin clouds can form very high in the atmosphere;
- If the surface temperature is relatively low, it leads to a slower cooling, a colder upper atmosphere and formation of thick water clouds at the tropopause level.

The model described in Marcq et al. (2017) whose output is used in the current study prescribes a vertical structure divided in three physical layers, from the surface up to the mesopause:

Unsaturated troposphere which follows a dry adiabat lapse rate;

Moist troposphere (optionally) if water saturation is reached at some point. This layer follows a moist adiabat;

Isothermal mesosphere at a temperature close to the skin temperature (near 200 K for Earth-like conditions).

Temperature profiles as a function of altitude are shown on Figure 8.3. Temperature-Pressure profiles are shown on Figure 8.4. In the moist troposphere (if present), liquid water droplets are present in non negligible amounts. Their radiative effect is twofold: they lower the asymptotic outgoing longwave radiation (OLR) at low surface temperatures; and they extend the blanketing effect to higher surface temperatures (increase of T_ε compared to the cloudless case). The blanketing effect can be described as OLR $= \sigma {T_{\text{eff}}}^4 = \varepsilon\sigma {T_{\text{surf}}}^4$ where ε is an effective emissivity. The smaller the value of ε, the stronger the blanketing effect. T_ε can be used as a threshold to distinguish the low-temperature regime ($T_{\text{surf}} < T_\varepsilon$) from the high-temperature regime ($T_{\text{surf}} > T_\varepsilon$).

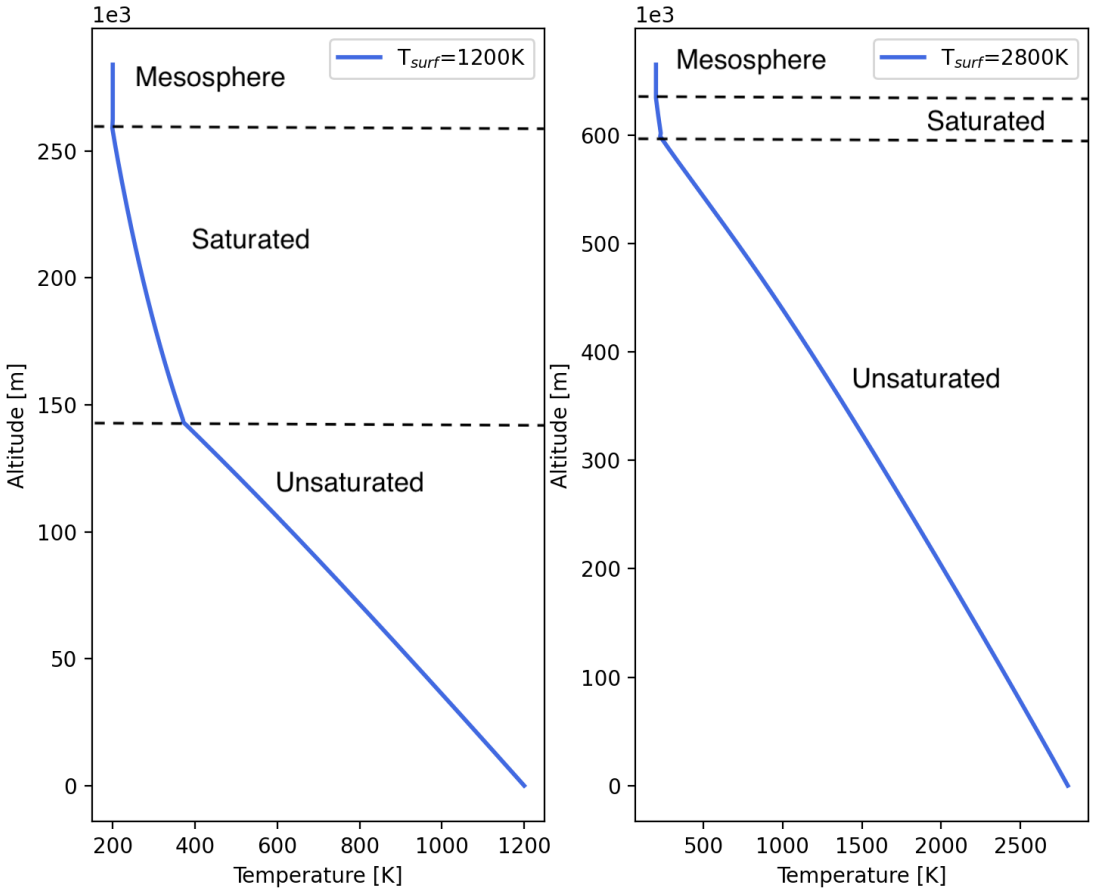


Figure 8.3: $T(z)$ profile. Left: $T_{\rm surf} = 1200\,{\rm K} < T_\varepsilon \approx 1690\,{\rm K}$ for $P_{\rm surf}({\rm H_2O}) = 300$ bar and $P_{\rm surf}({\rm CO_2}) = 100$ bar. Right: $T_{\rm surf} = 2800\,{\rm K} > T_\varepsilon \approx 1690\,{\rm K}$ for $P_{\rm surf}({\rm H_2O}) = 300$ bar and $P_{surf}({\rm CO_2}) = 100$ bar.

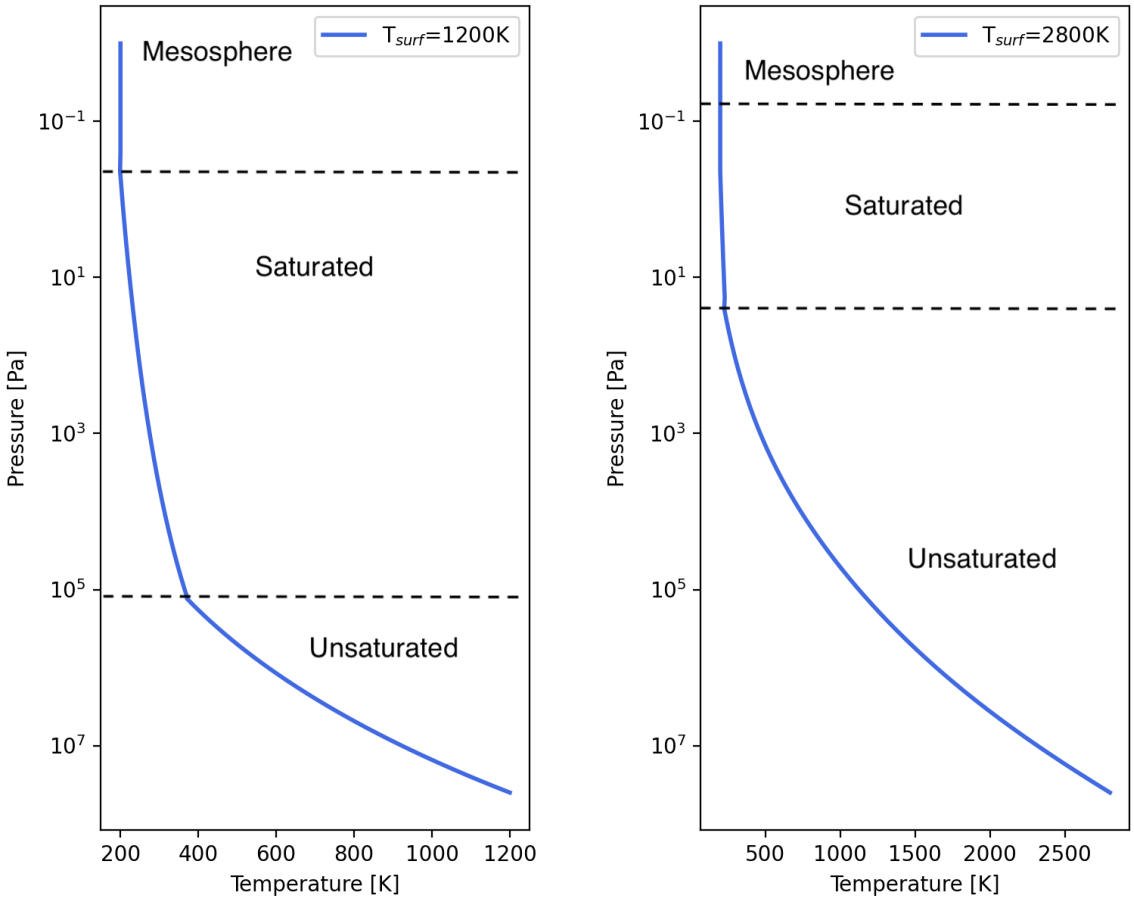


Figure 8.4: Temperature-Pressure profiles. Left: $T_{\rm surf} = 1200\,{\rm K} < T_\varepsilon \approx 1690\,{\rm K}$ for $P_{\rm surf}({\rm H_2O}) = 300\,{\rm bar}$ and $P_{\rm surf}({\rm CO_2}) = 100\,{\rm bar}$. Right: $T_{\rm surf} = 2800\,{\rm K} < T_\varepsilon \approx 1690\,{\rm K}$ for $P_{\rm surf}({\rm H_2O}) = 300\,{\rm bar}$ and $P_{\rm surf}({\rm CO_2}) = 100\,{\rm bar}$

Inputs of the model consist of: R_T – the planetary surface radius, $T_{\rm surf}$ – the surface temperature, $T_{\rm top}$ – the mesospheric temperature up to the mesopause at 0.1 Pa , $P_{\rm surf}({\rm H_2O})$ – H_2O surface pressure, $P_{\rm surf}({\rm CO_2})$ – CO_2 surface pressure, $P_{\rm surf}({\rm N_2})$ – N_2 surface pressure.

The output of the model includes, among other values dealing with radiative budget, arrays of altitude, partial pressures and local densities of H_2O, CO_2 and N_2 that characterize the vertical structure of these atmosphere. Since the radiative outputs of the model from (Marcq et al., 2017) are computed only for emission and reflection spectra at a low spectral resolution in order to assess the radiative budget of the planet, they could not be used here, and we built an improved model for high spectral resolution transmission described hereafter.

8.3 Transmission model description

Several models exists for the 1-D forward transmission model, for example radiative-convective model Exo-REM (Baudino et al., 2015; Charnay et al., 2018; Blain et al., 2020) for brown dwarfs and giant exoplanets, Exo_k (Leconte, 2021), petitRADTRANS (Mollière et al., 2019). In our study we decided to create a model to compute high resolution spectra as precise as it could be, later we will point to the difference between our model and others. The comparison is done between our resulting transmission spectra and Exo_k. During the transit stellar light crosses the exoplanet's atmosphere, rays enter the atmosphere with impact parameter b, distance of the closest approach is r_{min}, θ is the polar angle, vertical height is equal to $r - R_p$, as shown in Figure 8.5. Transmission depth can be computed as:

$$\delta_\nu = \frac{(\pi R_p^2 + \alpha_\nu)}{\pi R_{star}^2} \tag{8.5}$$

where α_ν is equal to

$$\alpha_\nu = 2\pi \int_0^{z_{max}} (R_p + z) \left(1 - e^{-\tau_\nu(z)}\right) \, dz = 2\pi \int_{R_p}^{R_p + z_{max}} b \left(1 - e^{-\tau_\nu(b)}\right) \, db \tag{8.6}$$

where z_{max} is equal to the maximum height where $\tau_\nu \ll 1$, τ_ν is the slant optical thickness at wavenumber ν and impact parameter b=R$_p$+z. For each value of b, several layers are crossed by stellar light rays, the path of a ray can be computed by variation of θ from 0 to arccos $\left(\frac{R_p}{R_P + z(P_{top})}\right)$. It should be mentioned that we compute the integral using the SciPy sub-package `integrate`, while for example in Exo_k, computations are performed upon discretized layers.

The optical thickness τ_ν depends on wavenumber, because different gases absorb at different bands – ranges of wavenumbers, and on height – because atmosphere consists of several layers and for each layer the local density of absorbing gas may vary. Two main approach exist to compute it, first is called "Line-By-Line method" (Goody and Yung, 1989), second is a "Correlated-k method" (Goody et al., 1989; Lacis and Oinas, 1991). In the Line-By-Line method, the transmission function is calculated by integration of the frequency domain. For each frequency, the absorption line (absorption coefficient) of the selected gas is calculated, usually there are many spectral points and large spectral intervals for modeling, so this practical implementation is computationally long and expensive. It was noticed that at many points in the spectral intervals, similar values of the absorption coefficient are observed.

It is therefore possible to compute the distribution density of the absorption coefficient values in the spectral interval, after which it is possible to switch from integration over frequency to integration over the absorption coefficient values, that is, instead of counting by frequency, we count by the value of the absorption coefficient. This is the Correlated-k method. Correlated-k method is faster than the Line-By-Line, but less accurate depending on the discretization values chosen for the absorption coefficients. Correlated-k method can be used as a first step with future modeling by Line-By-Line method if it is needed. For creation of high-resolution spectra Line-By-Line method is preferred.
In current study it was decided to stick with Line-By-Line method, since we want to create a precise model. Calculation of optical thickness in our model is performed following:

$$\begin{aligned}\tau_\nu(b) = 2\int_0^{\theta_{max}} (&\sigma_{H_2O}(\theta,b,\nu)n_{H_2O}(b,\theta) + \sigma_{CO_2}(\theta,b,\nu)n_{CO_2}(b,\theta)+\\ &+\sigma_{R,H_2O}(\theta,\nu)n_{H_2O}(b,\theta) + \sigma_{R,CO_2}(\theta,\nu)n_{CO_2}(b,\theta)+\\ &+Q_{ext}(\nu)\kappa\rho_{H_2O}(b,\theta))ds(\theta,b)\end{aligned} \tag{8.7}$$

where $\sigma_{H_2O}(\theta,b,\nu)$ and $\sigma_{CO_2}(\theta,b,\nu)$ are H_2O and CO_2 absorption cross-sections, $n_{H_2O}(\theta)$ and $n_{CO_2}(\theta)$ are H_2O and CO_2 local number densities, $\sigma_{R,H_2O}(\theta,\nu)$ and $\sigma_{R,CO_2}(\theta,\nu)$ are H_2O and CO_2 Rayleigh scattering cross-sections, Q_{ext} the extinction efficiency of clouds droplets, $\kappa = 0.038 \times 130$ in m^2/kg, with the factor 0.038 representing the ratio between liquid water content and the local water vapor mass density, and 130 m^2/kg the extinction of 10 μm water droplets in (Kasting, 1988). Also, $ds(\theta,b) = b \times d(tan(\theta))$

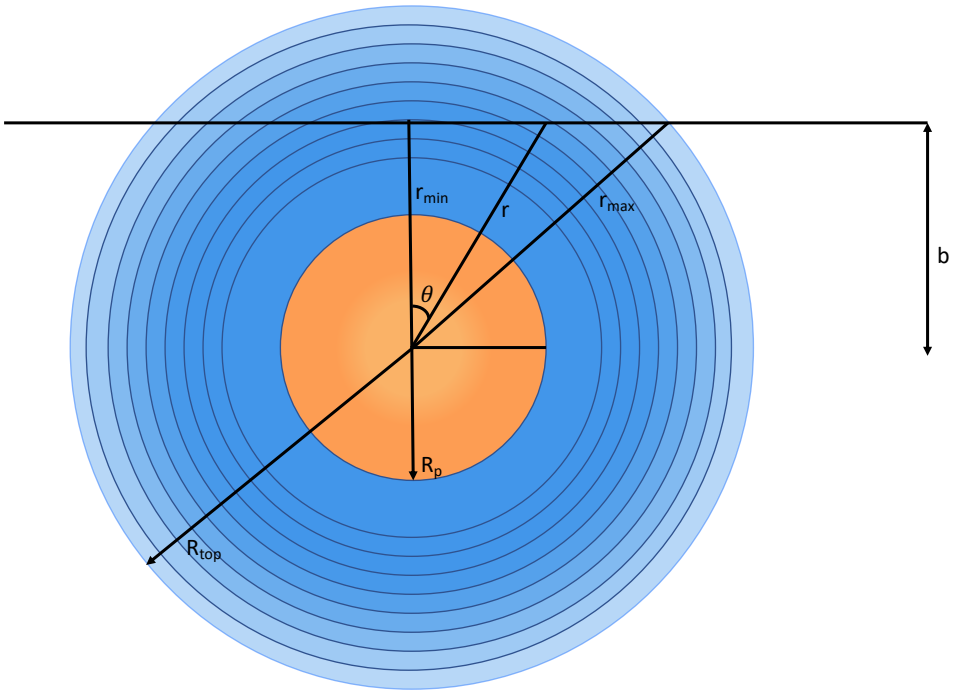


Figure 8.5: Sketch illustrating scheme. R_p – radius of the planet, $b = r_{min}$ – impact parameter, θ - angle between normal and ray path

Gaseous line absorption We used HITRAN2020 database (Gordon et al., 2022) for the H_2O and CO_2 cross-sections. Calculation of absorption cross-sections is done using the HAPI - HITRAN Application Programming Interface (Kochanov et al., 2016). First HAPI is used for downloading a HITRAN database for usage on computer (local database) from the HITRAN website (https://hitran.org). In addition, it is possible to use HAPI functions on any HITRAN format database; after it is possible to compute cross-sections based on line-by-line data using HAPI, it supports Doppler, Lorentzian, Voigt profiles and several others, for our model we chose a Voigt profile. Profile is calculated for the selected molecule for the selected wavenumber range, for selected temperature and pressure, and it is also possible to choose the diluent - parameter to set the gas mixture composition (by default it is `air`, but it is possible to set it to `self` or for example `CO2`). Example of calculated cross-sections by HAPI is shown on Figure 8.6. In Figure 8.7 we illustrate the broadening of H_2O caused by CO_2. As it was mentioned above, it is possible to set `Diluent` parameter at HAPI function to create absorption cross-section for mixture of gases, to check if it produces correct data, we decided to take HITRAN-like database from Brown et al. (2007) with data for CO_2 broadened water at region 200-900 cm^{-1} and see if obtained values will be in agreement. As we can see absorption cross-section values produced using HAPI database and by Brown et al. (2007) database match each other, so it possible to use

HAPI interface without additional/separate database-search.

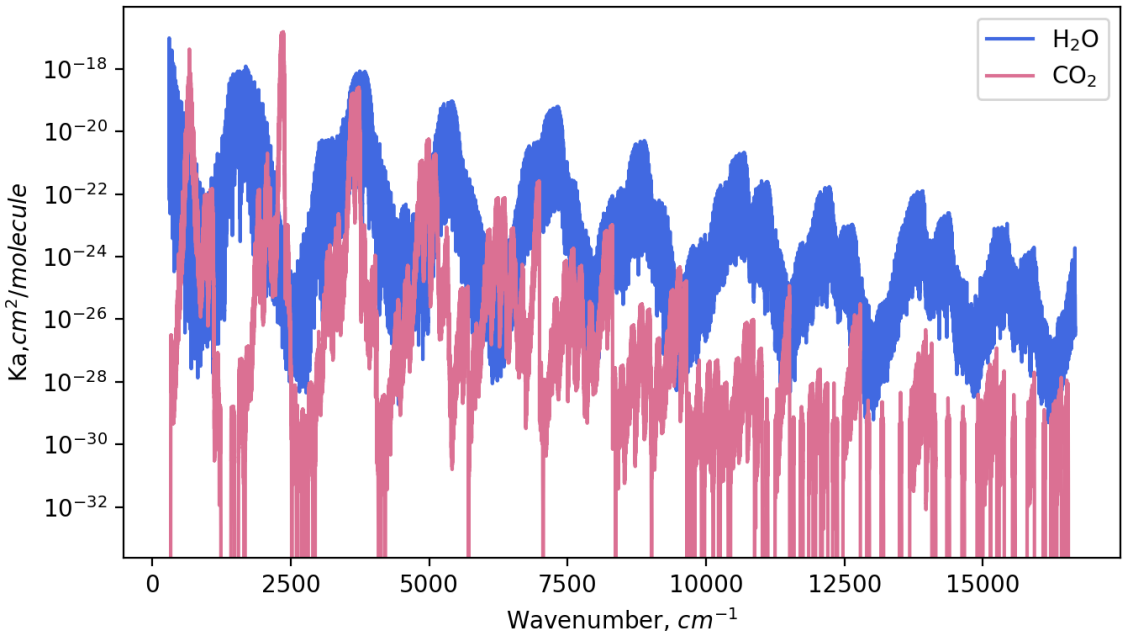


Figure 8.6: Absorption cross-sections calculated by HAPI for H_2O (blue) and CO_2 (pink) at wavenumber range from 300 to 16666 cm^{-1}. Broadening is done by air, $T = 296$ K, $P = 1$ atm.

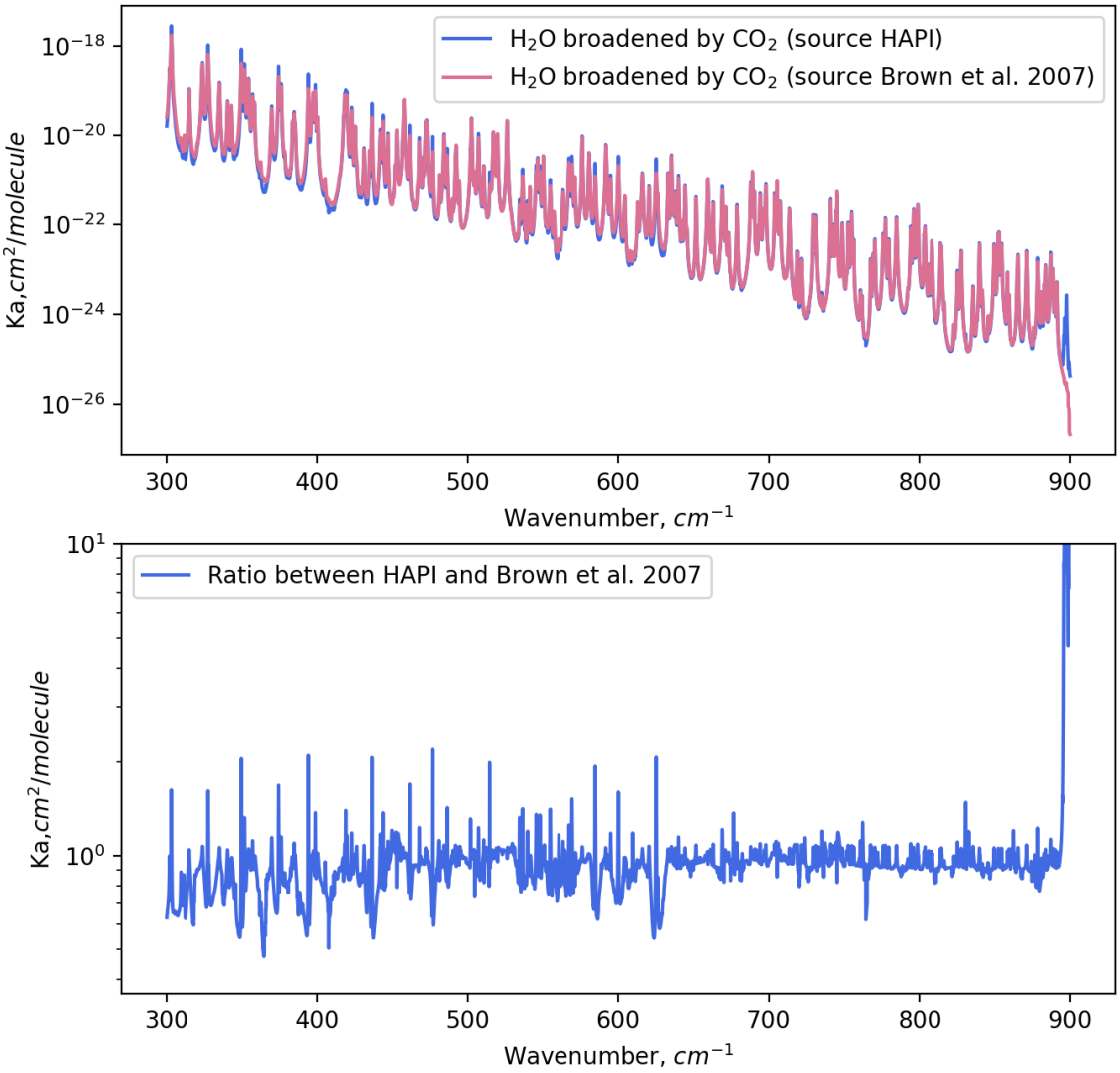


Figure 8.7: Absorption cross-sections calculated by HAPI for H_2O broadened by CO_2, $T = 296$ K, $P = 1$ atm, at wavenumber range from 300 to 900 cm^{-1}. Top graph: blue – source HITRAN database with `diluent` parameter, pink – database from (Brown et al., 2007). Bottom graph: ratio between cross-sections computed from HITRAN database and (Brown et al., 2007).

To accelerate computations we performed pre-computations of cross-sections stored in look-up tables. We made a grid of (Pressure, Temperature, Wavenumber), where pressure is a grid of 10 pressure values from 10^{-2} to 10^{7} Pa, temperature is a grid of 26 temperature values from 200 to 2800 K, and wavenumber is a grid of wavenumbers from 300 to 16666 cm^{-1} with step 0.01 cm^{-1}. So it is possible to make precise interpolation in case of (P,T,W) point in between grid layers.

Rayleigh scattering Rayleigh cross-sections of H_2O and CO_2 were assumed to follow λ^{-4} dependency: $\sigma_R(\lambda) = \sigma_0 \, (\frac{\lambda_0}{\lambda})^4$ with σ_0 and λ_0 stated at 8.1.

Molecule	σ_0 [cm^2/molecule]	$\lambda_0[\mu m]$	Bibliographical reference
H_2O	$2.5\ 10^{-27}$	0.6	Kopparapu et al. (2013)
CO_2	$1.24\ 10^{-26}$	0.532	Sneep and Ubachs (2005)

Table 8.1: The Rayleigh scattering cross-sections and corresponding wavelengths for H_2O and CO_2.

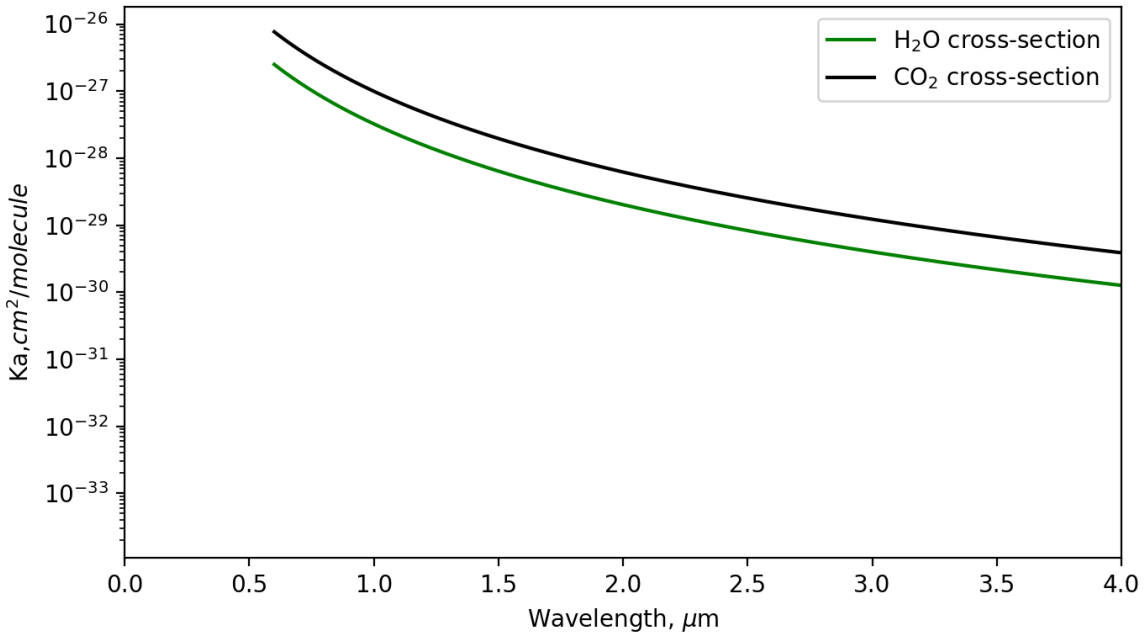


Figure 8.8: Rayleigh cross-section as a function of wavelength for H_2O and CO_2.

It is also possible to follow (Caldas et al., 2019) for Rayleigh scattering equation and parameters described at the Appendix D of this paper.

Cloud optical properties The cloud opacity is taken following (Marcq et al., 2017). According to approximated Mie theory calculations from Kasting (1988) and assuming spherical H_2O droplets with a mean 5 μm radius we take Q_{ext} – the extinction efficiency – equal to 1 (in the wavenumber range with which we are working). As previously mentioned, the mass loading of clouds is equal to 0.038×130 m^2/kg based on Kasting (1988).

8.4 Result of modeling

For the first test case we decided to make a comparison with Exo_k model for the pure Rayleigh atmosphere. The pure Rayleigh atmosphere is done by using only N_2 as a background gas in Exo_k and Rayleigh cross-sections of N_2 from Caldas et al. (2019) for the current developed model. The result of first test case is shown in figure 8.9, we can see that the models are in agreement and shows same behavior.

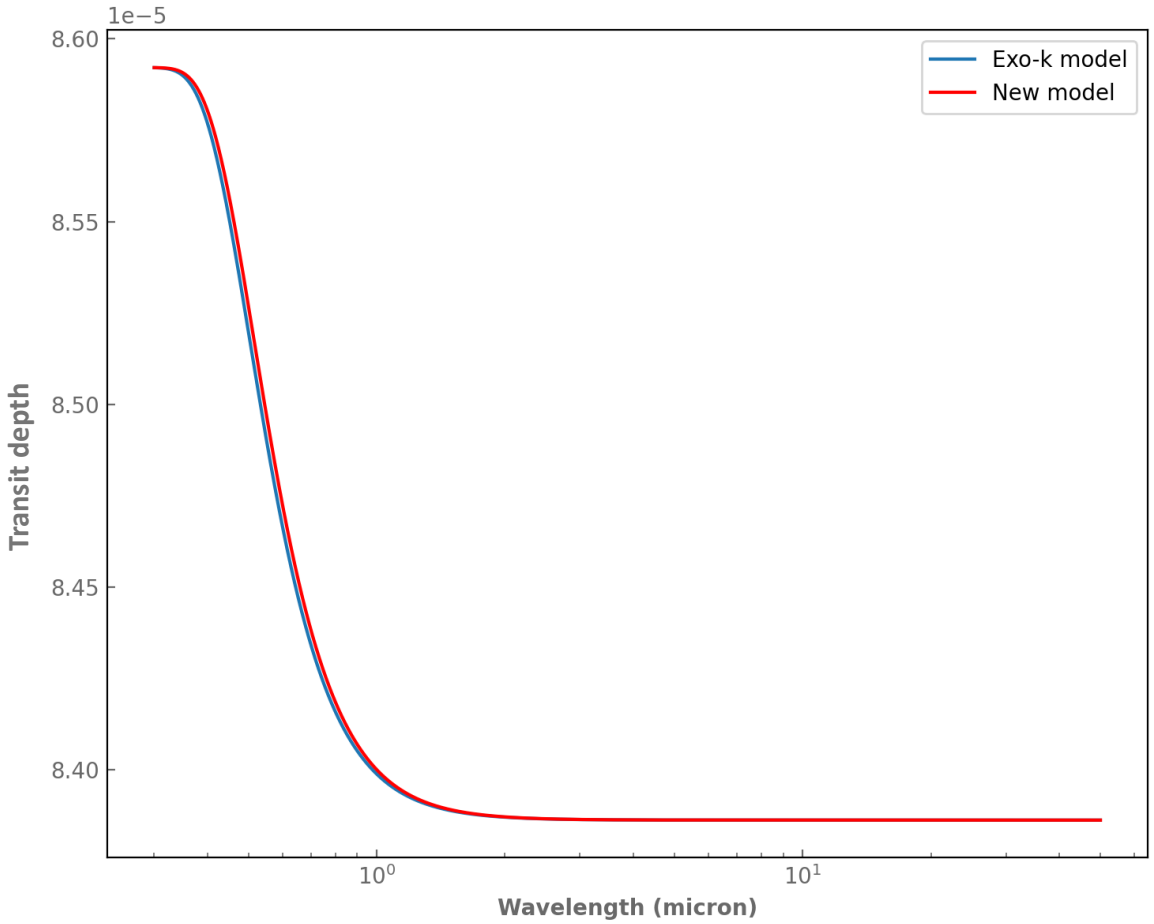


Figure 8.9: The transit depth as a function of wavelength for the pure Rayleigh atmosphere. Blue line - Exo_k model, red line - developed model.

As a second test case it was decided to model the simple 1-layer pure H_2O atmosphere. For economy of time computation was done first only for band at 2000cm^{-1} - 3333 cm^{-1}. As temperature limits were chosen 200 K at the top of atmosphere and 600 K at the bottom; as pressure limits - 10^{-1} Pa and 10^4 Pa correspondingly. As a source of k-table for Exo_k the POKAZATEL (Polyansky et al., 2018) computed for R=1000 was used. As a source of cross-sections for developed model the HITRAN database (with usage of HAPI) was used, R=1000. The model was computed for the

Earth-size exoplanet, around Sun like star. The result of computation is shown on Figure 8.10. On Figure 8.11 the zoom on a small part of spectrum, we can notice small discrepancies which may be due to the fact of the differences between original k-table and cross-sections, and differences between the spectral binning.

For case shown on Figure 8.12 for the pure 1-layer H_2O atmosphere, band 4168 cm^{-1}-5261 cm^{-1}, discrepancies may be more pronounced, for example line around 2 μm, which is not presented at the current model, but is at the Exo_k model result. This kind of discrepancies are caused by the difference between original k-table and cross-sections table.

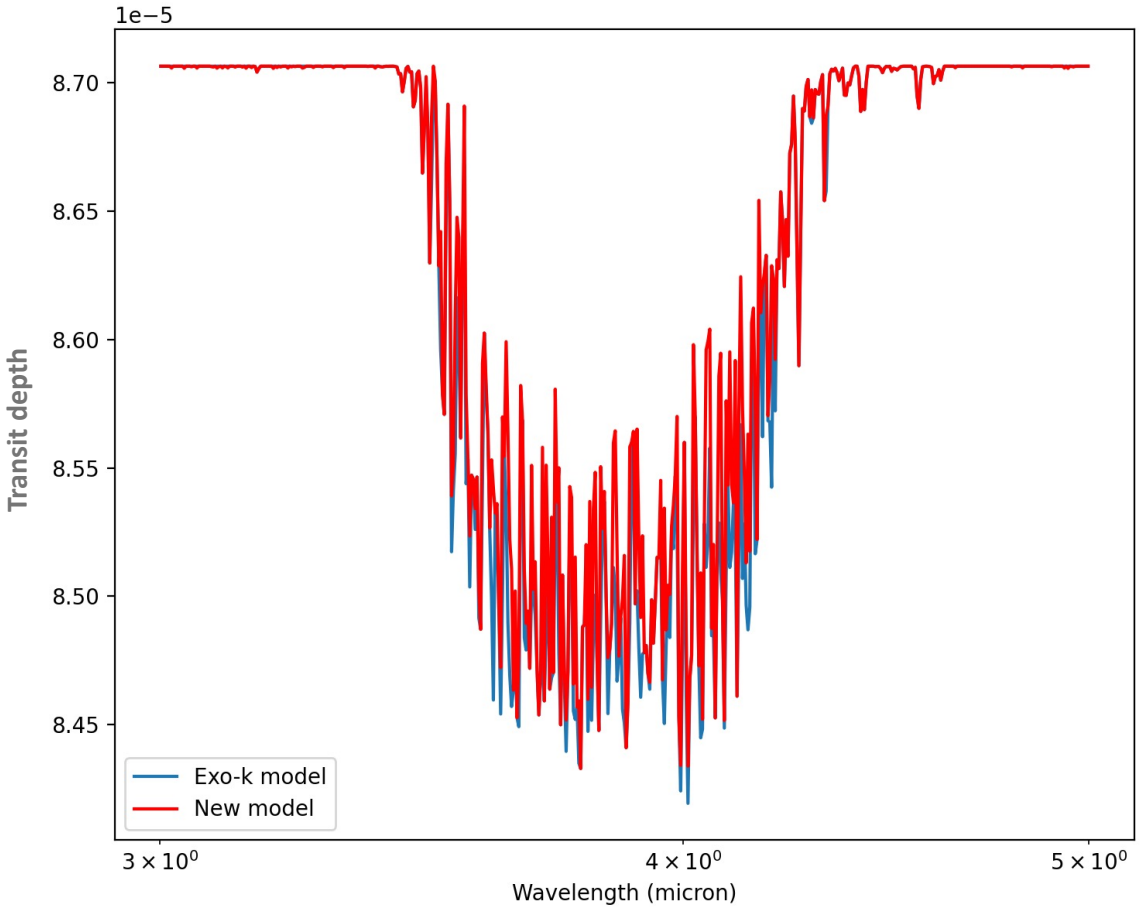


Figure 8.10: The transit depth as a function of wavelength for the 1-layer pure H_2O atmosphere, band 2000 cm^{-1}-3333 cm^{-1}. Blue line - Exo_k model, red line - developed model.

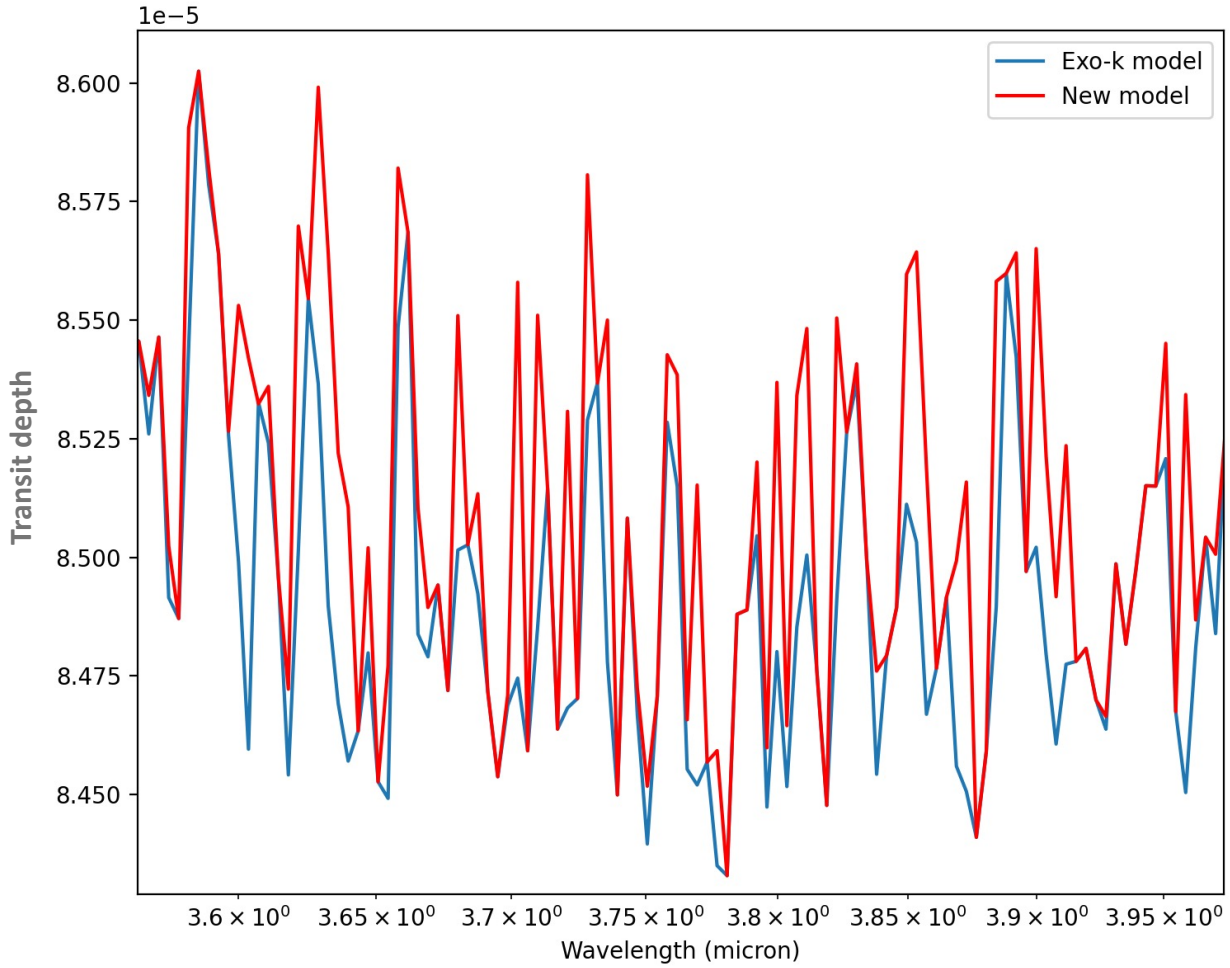


Figure 8.11: Zoom on Figure 8.10. Blue line - Exo_k model, red line - developed model.

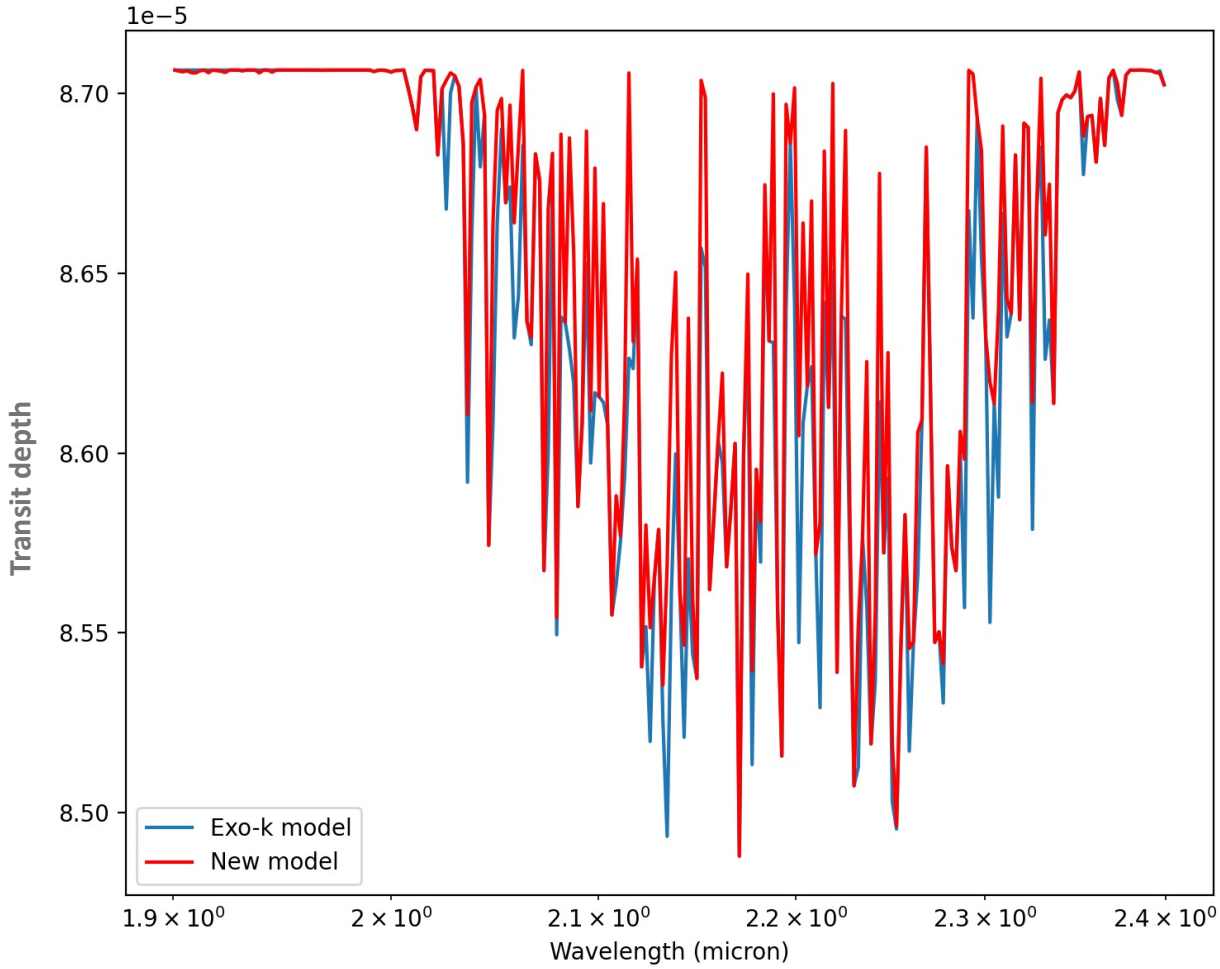


Figure 8.12: The transit depth as a function of wavelength for the 1-layer pure H_2O atmosphere, band 4168cm^{-1}-5261cm^{-1}. Blue line - Exo_k model, red line - developed model.

After we have proved that the developed model returns an adequate result we can proceed to the mixing of gases, since the most interesting for us is a case of the atmosphere consisting of H_2O and CO_2. By both models it is possible to produce mixed atmosphere, for the first test it was decided to produce band from previous test 4168 cm^{-1}-5261 cm^{-1}. On Figure 8.13 we show 2 atmospheres, pure H_2O and pure CO_2 produced by Exo_k. On Figure 8.14 we demonstrate models result for the 1-layer 70% H_2O and 30% CO_2 atmosphere.

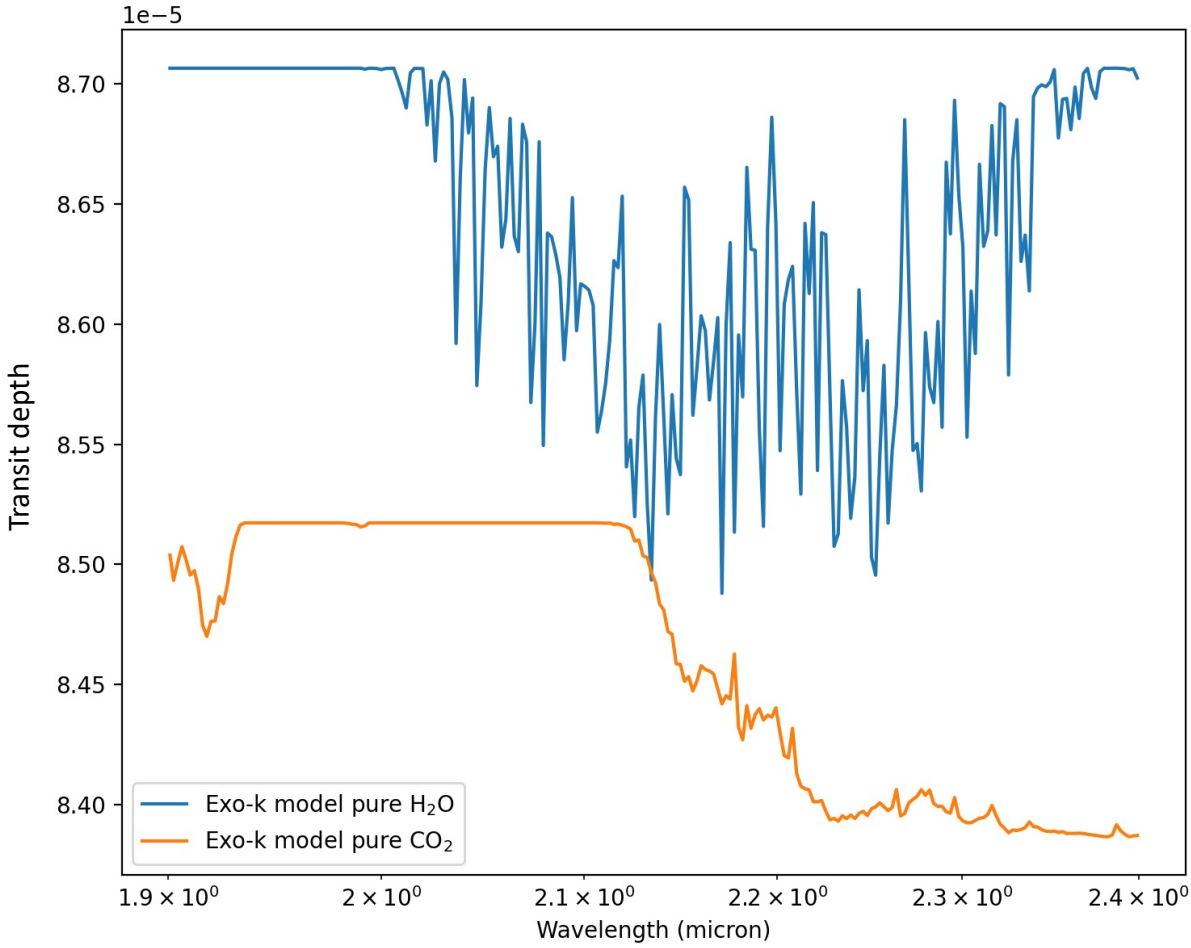


Figure 8.13: The transit depth as a function of wavelength for the 1-layer pure H_2O and pure CO_2 atmospheres, band 4168 cm^{-1}-5261 cm^{-1}. Blue line – Exo_k model for pure H_2O atmosphere, orange line – Exo_k model CO_2 atmosphere

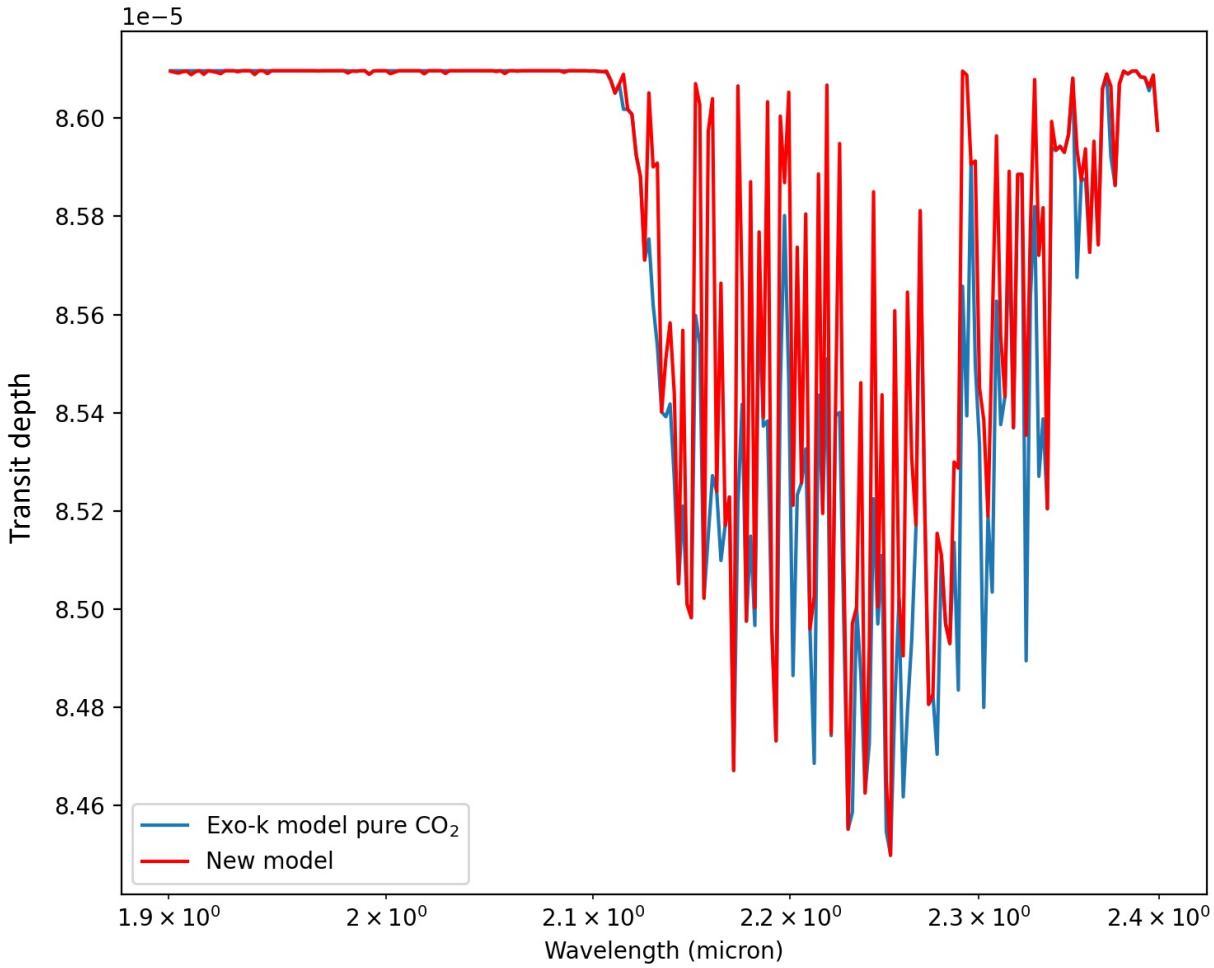


Figure 8.14: The transit depth as a function of wavelength for the 1-layer 70% H_2O and 30% CO_2 atmosphere, band 4168 cm^{-1}-5261 cm^{-1}. Blue line - Exo_k model, red line - developed model.

During our study, it became clear that exoplanets with such thin atmospheres would not be accessible by transmission spectroscopy, the transit depth being extremely small for ground-based spectroscopy capabilities. Furthermore, even in the case of spectroscopy from space, it will be possible to observe only the upper atmospheric layers (at pressures lower than $\sim$ 0.1 bar), yet the main strong point of the model from Marcq et al. (2017) is in the deep layers (not probed by transmission spectroscopy).

It was then decided in a second time to switch to thermal emission and reflected starlight spectroscopy; first – to benefit from the full advantages of the Marcq et al. (2017) model.; second – it will be easier to detect features in near infrared emission for planets with such hot atmospheres. Unfortunately, there was too little time to develop corresponding orbital phase curve contrasts (emission and reflection) models.

Chapter 9

Exoplanet statistics

During my master study and at the beginning of PhD I have worked on the exoplanetary statistics with IKI RAS team. The main idea we focused on is the observational biases such as observational selection. At first it was mainly dedicated to the statistics of transit exoplanets, afterwards we included statistics of RV-exoplanets in our consideration. The differences in spectrograph sensitivities, differences in activity levels of host stars, different durations of observations, different numbers of radial velocity measurements, and different data processing methods lead to a strong inhomogeneity of the existing catalogs of RV exoplanets. This inhomogeneity causes a significant distortion of the distributions of RV planets with respect to minimum masses ($m = M \sin i$) and orbital periods. To construct an unbiased (real) distribution of planets by minimum masses and orbital periods, it is necessary to take into account the different detection efficiency of RV planets with different minimum masses and orbital periods. One approach to regularization of inhomogeneous RV exoplanet data is the "detectability window" method. "Detectability window" is an $(n \times n)$ matrix in the "minimum mass – orbital period" diagram, the elements of which represent the probability of finding a planet with a given orbital period and minimum mass $V(m, P)$. To build a distribution that is not distorted by observational selection, each discovered RV planet on the "Minimum Mass – Orbital Period" diagram should be taken with a statistical weight inverse to the corresponding value of the probability of its detection – the matrix element $V(m, P)$. Two main approaches have been tried to construct the detectability window.

The first is an empirical approach based on the analysis of already discovered RV planets. Within the framework of this approach, it is assumed that each set of measurements of the radial velocity of a star forms the so-called "single detectability window" around this star - the area on the "minimum

mass – orbital period" diagram in which a planet with a minimum mass m and period P is guaranteed to be detected (detection probability is 1), and outside this area the planet will be missed (detection probability is 0). A planet with a minimum mass m and period P was considered detectable if several inequalities were simultaneously satisfied for it, for the period and semi-amplitude of the radial velocity. The second approach is based on the Lomb-Scargle periodogram method. For each star, based on the available measurements of its radial velocity, a Lomb-Scargle periodogram is constructed. The simulated signal from an exoplanet with period P and minimum mass m is added to the available values of the measured radial velocity. If the addition of the signal leads to the appearance on the periodogram of a reliable clear peak (FAP $< 0.1\%$) corresponding to the period of the added planet, it is considered that such a planet would have been detected (detection probability $p = 1$). If the addition of the signal does not lead to the appearance of a significant peak on the periodogram, it is considered that a planet with such an orbital period P and a minimum mass m would be skipped ($p = 0$).

With both approaches, a single detectability window is constructed from single detectability windows by averaging. The resulting detectability windows were compared and used to regularize the RV exoplanets distributions. This results and future development of it will be published. First results can be found at (Ivanova et al., 2021b).

Additional work on statistics Despite the fact that my main work in statistics was with the IKI team, I participated in a paper dedicated to the problem of minimum masses or apparent mass, or projected masses). The problem with minimum masses is that when the parameters of a planet are measured using the radial velocity method, the mass equal to $m \sin i$ is obtained, and not the real mass m of the planet (except when the planet is observed in a transit configuration). Thus, the transformation of the distribution of planets by minimum masses to the distribution by real masses m is necessary, such a method is proposed in the article. In addition to the fact that the article proposed a method similar to the classical method of atmospheric physics (onion peeling), the article also showed that the so-called desert of sub-Saturns really exists, even after the transition to the real-mass distribution, shown on Figure 9.1. More about the method can be found at Bertaux and Ivanova (2022).

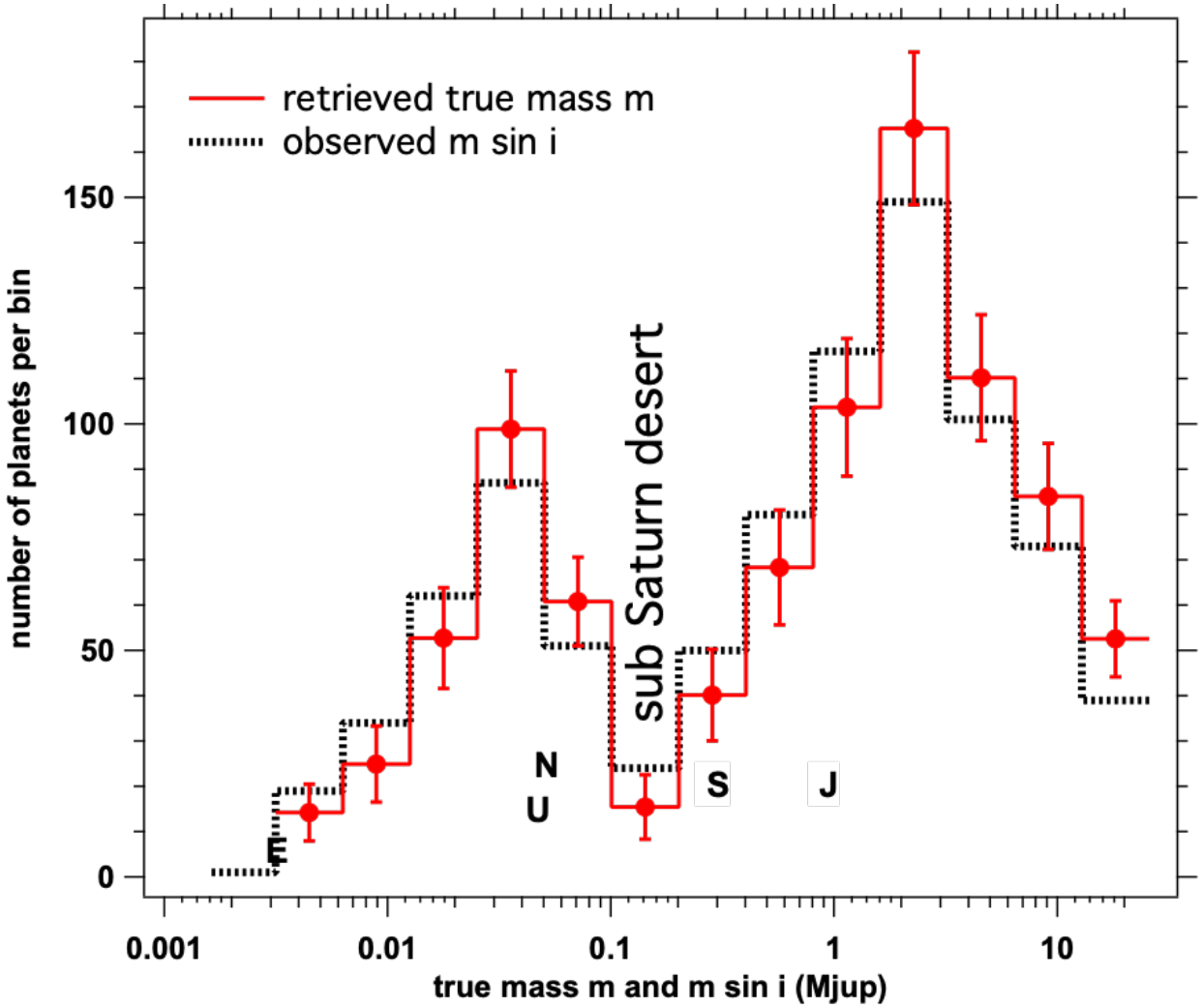


Figure 9.1: Dashed black line: distribution of msin i of 885 planets detected by the radial velocity method, from 1 $M_{\oplus}$ = 0.00314 M_{jup} up to 25.7 M_{jup}. Solid red line and red filled circles: distribution of true masses obtained by inversion with our new method with error bars. Letters E, U, N, S, J are representing the masses of Earth, Uranus (14.54 $M_{\oplus}$), Neptune (17.2 $M_{\oplus}$), Saturn (95 $M_{\oplus}$), and Jupiter (317.8 $M_{\oplus}$), respectively. The lowest bin in the central valley has limits 0.1–0.2 M_{jup} (32–64 $M_{\oplus}$). This sub-Saturn desert is even more depleted after inversion of the msin i distribution.

Conclusion and Perspectives

Ground-based spectroscopy helps us detecting exoplanets, refining the characteristics of already discovered planets, and even detecting exoplanet atmospheres. The radial velocity method was the first to detect an exoplanet in 1995 and is still widely used today. At present, spectrographs such as ESPRESSO show the utmost in instrument accuracy, excel in instrument calibrations, and maximize the reduction of errors due to instrument physics. If it is no longer possible to improve the instrument, then it is necessary to come up with new approaches to working with data.

One of these approaches is the correction of telluric absorption. In the course of my study, a method was developed for correcting telluric absorption in the visible wavelength range. The method was tested on ESPRESSO spectrograph data and we find that the precision of radial velocity measurements is improved in the second case, with a reduction of the average formal error from 1.04 ms^{-1} down to 0.78 ms^{-1} in the case of these ESPRESSO data and this stellar type for the red arm. Using an estimator of the minimal error based on photon noise limit applied to the full CCF, the error is reduced from 0.89 ms^{-1} down to 0.78 ms^{-1}. This corresponds to a significant decrease of about 35% of observing time to reach the same precision in the red part.Both in the visible and near IR range, we wish to improve the telluric correction based on TAPAS output in at least three directions:

- To make the correction process as automated as possible. Ultimately, it might be done at the time of data collection. There is one caveat however. This is because on the very day of observation, the meteorological field of ECMWF which is available (through TAPAS) is a (short term) prediction. A few days later, the re-analysis of all meteorological measurements provides a latest version of ECMWF. At present we have not checked if the difference of the two fields would reflect in a significant change of the TAPAS correction, but it may be a limiting factor to the possibility to make “real time” TAPAS telluric corrections in a dedicated pipeline.

- To detect serious discrepancies between some observed telluric lines and TAPAS/HITRAN

predictions. A typical example is the H_2O lines around 7185.51 Å (figures 1.20 and 1.21), and also the fact that there is a difference of 3% between the O_2 columns fitting the γ band and the B band of O_2. The source of discrepancy may be either a wrong parameter in the HITRAN data base, or a temperature/pressure vertical profile which is not well representing the reality of the day of observations. In the above mentioned case, since the discrepancy was noted by two observers at two different observations, the HITRAN discrepancy is more likely. Finding the actual cause of discrepancy may be done by statistical analysis of several observations of the same line, preferably at different observatory locations. However, TAPAS was operated with the 2016 version of the HITRAN data base, and the use of HITRAN 2020 could mitigates the atmospheric discrepancies noted above.

- To check the existence of residuals that could induce spurious radial velocity variations, and a one-year signal. This can be done based on one additional hot star observation during the night, its correction and subsequent series of simulations using the hot star residuals and a synthetic stellar spectrum to detect and either suppress some lines from the binary mask or build an appropriate correction, to be applied to all cool star exposures. Work is needed to check whether or not the gain in observing time for all exoplanet target exposures of the night is above the additional time for the hot star. Ultimately, an elaborated technique based on a library of residuals measured towards series of hot stars, similarly to the APERO procedure for SPIROU in the near-infrared may be necessary.

In the course of our studies of RV measurements of star HD 40307 hosting 3 planets, we also got interested in the algorithms used in order to extract one RV from a single spectrum, and algorithms extracting the variation of RV between two epochs. A new approach to the use of the Pierre Connes formulas was also investigated, overcoming the usual restriction to small displacement only by the use of the wavelength scale of the star in the Barycentric system (which is in the ESPRESSO product). It was proposed to apply it to the total CCF (CCF_{tot}) when trying to retrieve the change of RV between two exposures, and we call it the EPiCA algorithm.

The simplest way to examine the changes of RV of a star observed at different epochs is just to examine the series of RV which are in the official ESPRESSO data product (one RV per exposure), obtained by a Gaussian fit to CCF_{tot}.

There are ways however to get a better precision. The most sophisticated way is probably the Line-by-

Line analysis as performed by Artigau et al. (2022). It requires though the use of several sophisticated analysis and software tools that may be difficult to operate by the standard ESPRESSO user.

Our EPiCA algorithm should be perceived as a tool of intermediate complexity between these two extremes. It does not require to look at the details of the spectra, since it works on the CCF_{tot} which is contained in the data product, and is a "summary image" of all the spectral lines of the spectrum. The corresponding software has been coded in Python and is available by pip or github (https://github.com/aeictf/EPiCA).

By providing change of RV, DRV(i), as a function of the CCF grid point, the EPiCA algorithm allows to detect simply and immediately if there has been only a pure Doppler shift between two epochs, or if there has been also a *change* of the shape of spectral lines. In such a case, this is a warning that stellar activity (for example the convective granulation blue shift) has modified the spectral lines shapes.

We do not claim that the EPiCA method gives more precise results than the Line-by-line analysis. But we think it gives better results than the simplest method of Gaussian fit to CCF_{tot}. Using the data of the HD 40307 as an example, it was shown that the application of the new method has advantages over the classical implementation of Gaussian fit to CCF_{tot}. It reduces the residuals (averaged over one night) from the dynamical planetary best fit by a factor of 3, giving more confidence in the retrieved best fit parameters with the EPiCA method. These residuals come down to 0.11 ms^{-1}, approaching the realm of an Earth-like planet around a solar-type star (admittedly, over a limited time frame covering one week only).

We have shown that the two time series of RV, Gaussian fit, and EPiCA, have a variable difference, amounting to $\sim$1.3 ms^{-1}. We strongly suspect that it the sign that both methods are reacting differently to the distortions caused by stellar activity, EPiCA method being less perturbed than the Gaussian fit. It has also been shown that this method can be used to study stellar activity. By applying the CF method separately on the blue side and the red side of the CCFs, we found out that there is a significant difference (up to $\sim$5 ms^{-1}) between the two times series of retrieved RV, and also that this difference is variable in time (figure 3.19) . It is the sign of distortion of the spectral lines connected to stellar activity. More specifically, we think that variations of granulation shifts are the reason of this blue-red difference in the RV time series. Indeed, the blue side of the line is formed in regions of upwelling flow, while the red side is formed in downwelling regions. Therefore, any change in the

relative areas of upwelling and downwelling granules will be reflected in a distortion of the spectral line.

One positive feature of the Connes approach is that each point of the CCF (or a spectral line) gives a signal. The challenge will be to distinguish, in this signal, what is due to a Doppler shift displacement (for instance due to the presence of a planet), or some sign of stellar activity. Meunier (2021) made review on the RV effects of several kinds of stellar activity. Not only the blue-red difference could be studied, but also the actual distortion/displacement of the line (or CCF) all along the CCF (or line). Probably a study of line-by-line with the Connes approach would be an ultimate investigation. However, the SNR on a single line might be an obstacle. The goal would be to try to establish a relation between a signature of granulation activity and the corresponding overall Doppler shift of lines or CCF, in order to provide a correcting factor of the stellar activity. We have briefly sketched such a scheme at the end of Chapter 5.

In the field of transmission spectroscopy, model of transit spectroscopy have been made for an early stage telluric exoplanet – magma ocean. During the study we came to the conclusion that it is preferable to develop thermal emission spectroscopy models for this type of planets. In fact it was already started in Acuña et al. (2023), the authors developed their model on the basis of Marcq et al. (2017). As possible extension to their work, we can propose:

- exploring a wider parameter space than done in Acuña et al. (2023);
- aiming for an improved spectral resolution and/or spectral range for the thermal component ;
- including the reflected starlight component, paving the way for full spectroscopic phase curves.

Published papers

1. **Improved precision of radial velocity measurements after correction for telluric absorption**

A. Ivanova, R. Lallement, J.L. Bertaux

Article published in Astronomy & Astrophysics Volume 673, A56, May 2023

DOI: https://doi.org/10.1051/0004-6361/202245089

2. **Toward a 3D kinetic tomography of Taurus clouds I. Linking neutral potassium and dust**

A. Ivanova, R. Lallement, J. L. Vergely, C. Hottier

Article published in Astronomy & Astrophysics Volume 652, A22, August 2021

DOI: https://doi.org/10.1051/0004-6361/202140514

3. **The "Detectability Window" Method to Take into Account Observational Selection in the Statistics of Exoplanets Discovered through Radial Velocity Measurements**

A. Ivanova, O. Yakovlev, V. Ananyeva, I. Shashkova, A. Tavrov, J. -L. Bertaux Article published in Astronomy Letters Volume 47, January 2021

DOI: https://doi.org/10.1134/S1063773721010059

Published or Submitted paper as co-author

1. **EXPLANATION: Exoplanet and Transient Events Investigation Project**

Valyavin, G. et al.

Article published in Astrophysical Bulletin Volume 77, 4, December 2022

DOI: https://doi.org/10.1134/S1990341322040186

2. **EXPLANATION: Exoplanet and Transient Event Investigation Project Optical Facilities and Solutions**

Valyavin, G. et al.

Article published in Photonics for Solar Energy Systems IX Volume 9, 12, December 2022

DOI: https://doi.org/10.3390/photonics9120950

3. **A numerical inversion of m sin i exoplanet distribution: the sub-Saturn desert is more**

depleted than observed and hint of a Uranus mass gap

J.-L. Bertaux, A. Ivanova

Article published in Monthly Notices of the RAS Volume 512, 4, June 2022

DOI: https://doi.org/10.1093/mnras/stac777

Appendix A

The non detection of stellar oscillations on HD 40307

One potential cause of RV time variation is the presence of stellar oscillations, similar to the well-known 5 minutes solar oscillations (pressure modes). Their presence is detrimental to the detection of planets, because they may bias the RV at a given time. It is common practice to use an integration time larger than the assumed period, if it is not known. HARPS team strategy has been to get at least 15 minutes exposures when searching for exo-planets. Here we show our Lomb-Scargle analysis of the time series on star HD 40307, with negative results. Before, and to put the HD 40307 in context, an example of another star found in the literature with positive results is presented.

The case of HD 88595. This star HD 88595 has been observed along a coordinated campaign with both photometric measurements from space with CHEOPS and RV measurements with ESPRESSO (Sulis et al., 2023). This star has no planet, and the barycentric RV variations may be assigned only to oscillations, and changes of granulation/supergranulation. We take from Sulis et al. (2023) the two RV time series on Figure A.1:

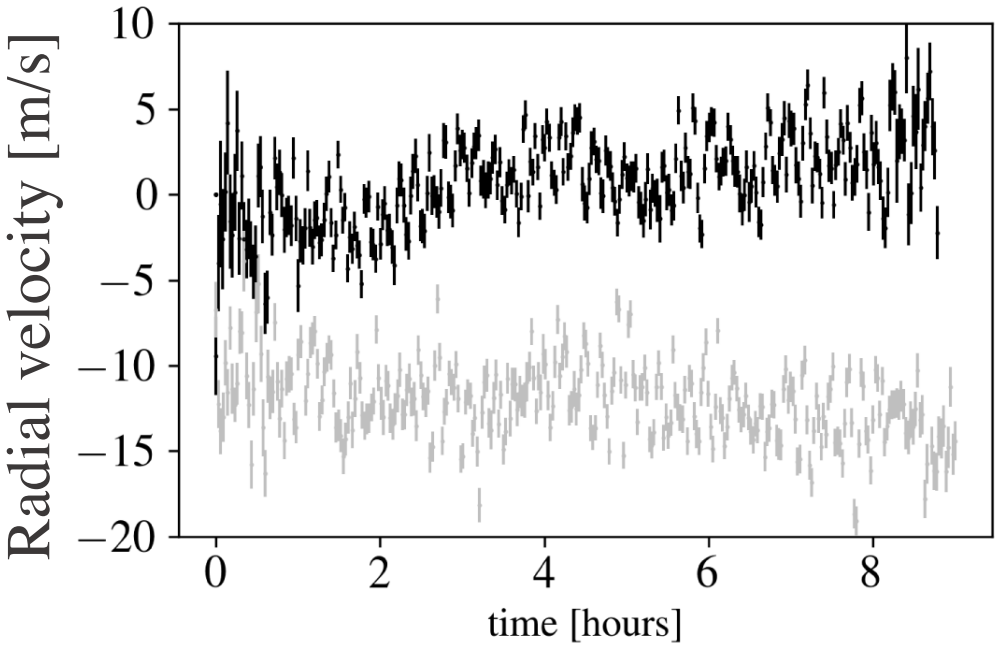


Figure A.1: ESPRESSO radial velocities time series for star HD 88595 during two different nights (about 8 days apart), black and grey (extracted from Sulis et al. (2023)). Periodic stellar oscillations are clearly visible to the naked eye is some portions of the time series.

The stellar oscillations are clearly visible with a period of about 10 minutes, and should be averaged out if one want to study only the granulation changes. Indeed, the periodogram of this time series (figure A.2, right), there is one single conspicuous peak around 1534 μHz, corresponding to about 652 s or 11 minutes, about the expected period for this F7 star. By contrast, the other star studied in this campaign, HD 67458 (F5 spectral type), does not display any significant peak (periodogram on the left side of Figure A.2).

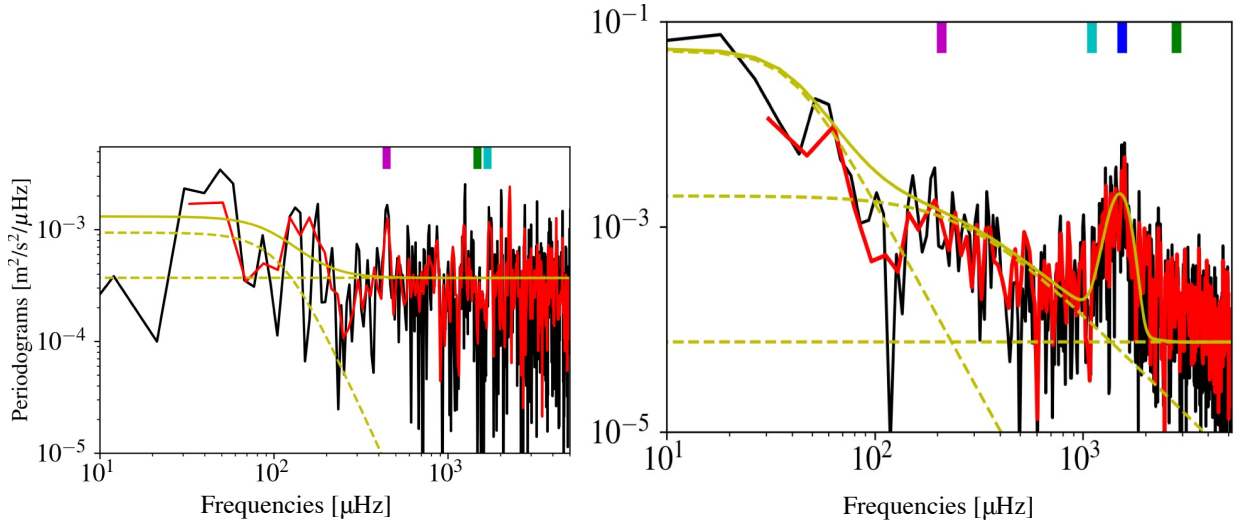


Figure A.2: Periodograms of ESPRESSO RV times series of stars HD 67458 (left) and HD 88595 (right). The abscissae are frequency in μHz. There is a distinct peak only the right, centered around 1534 μHz, corresponding to about 652 s or 11 minutes. The peak has a significant width (like the Sun). Extracted from Sulis et al. (2023).

The case of HD 40307. As already mentioned previously, the 8 days ESPRESSO campaign for this star was organized to detect stellar oscillations (PI Bouchy, reported in Pepe et al. (2021)). A high cadence was selected (one exposure every ∼71 second). We have made a Lomb-Scargle periodogram analysis of the longest series obtained during night 24.

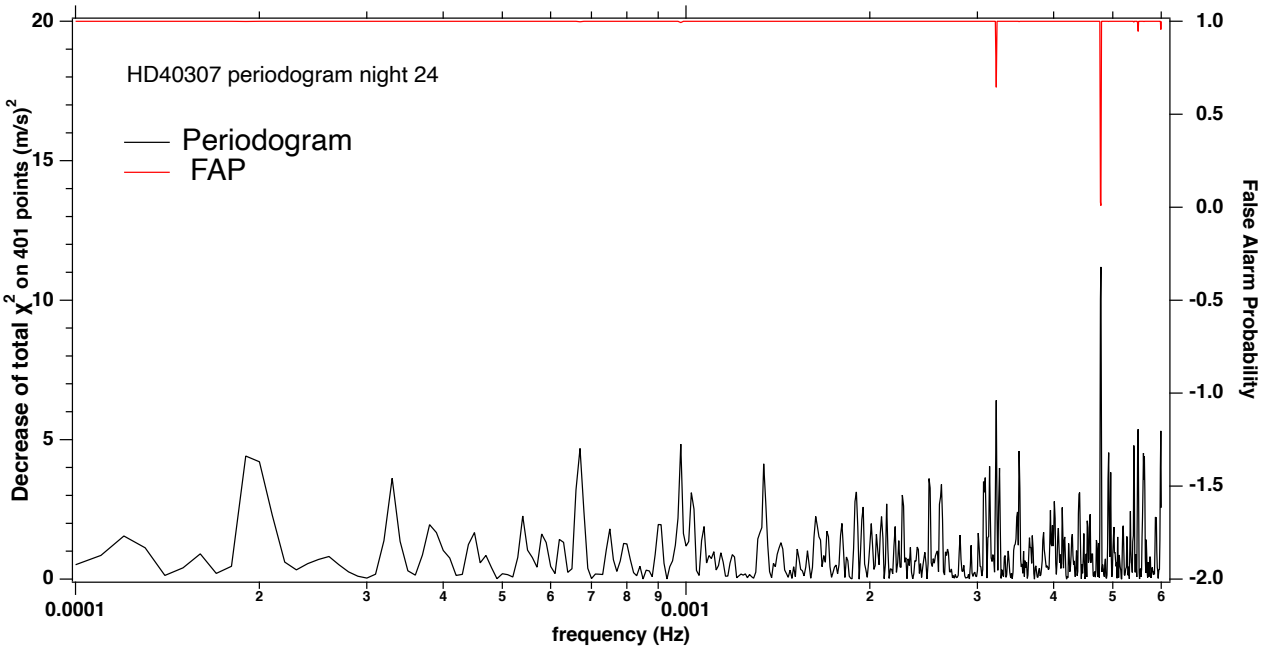


Figure A.3: Lomb-Scargle periodogram on the longest series of 401 measurements of night 24 December 2018. RV is determined form a Gaussian fit to CCF. The frequency step is 10^{-5} Hz. There is one single peak at frequency 0.00478 Hz, corresponding to a period of 209.5 s (3.5 mn), with a FAP (False Alarm Probability) of less than 1%.

However, at variance with the observed peak on HD88595, here the peak is very narrow (only on two points) and is certainly of spurious origin, probably associated to measurements windows timing. Several periodograms have been performed similarly to the one above, on various RV time series obtained in different conditions (parameters defining the Gaussian fit, Connes formula on CCFs of individual orders contained in the ESPRESSO product, weighting of orders with standard deviations), but always with the data of night 24. We noted that a peak is not found for all conditions. For instance, it seems that, using Connes formula on CCFs of individual orders to establish the time series, there is no longer a peak, as shown on Figure A.5. Peaks (spurious, as on figure A.4) are found only with the Gaussian fit to CCF, and never with Connes formula on CCFs.

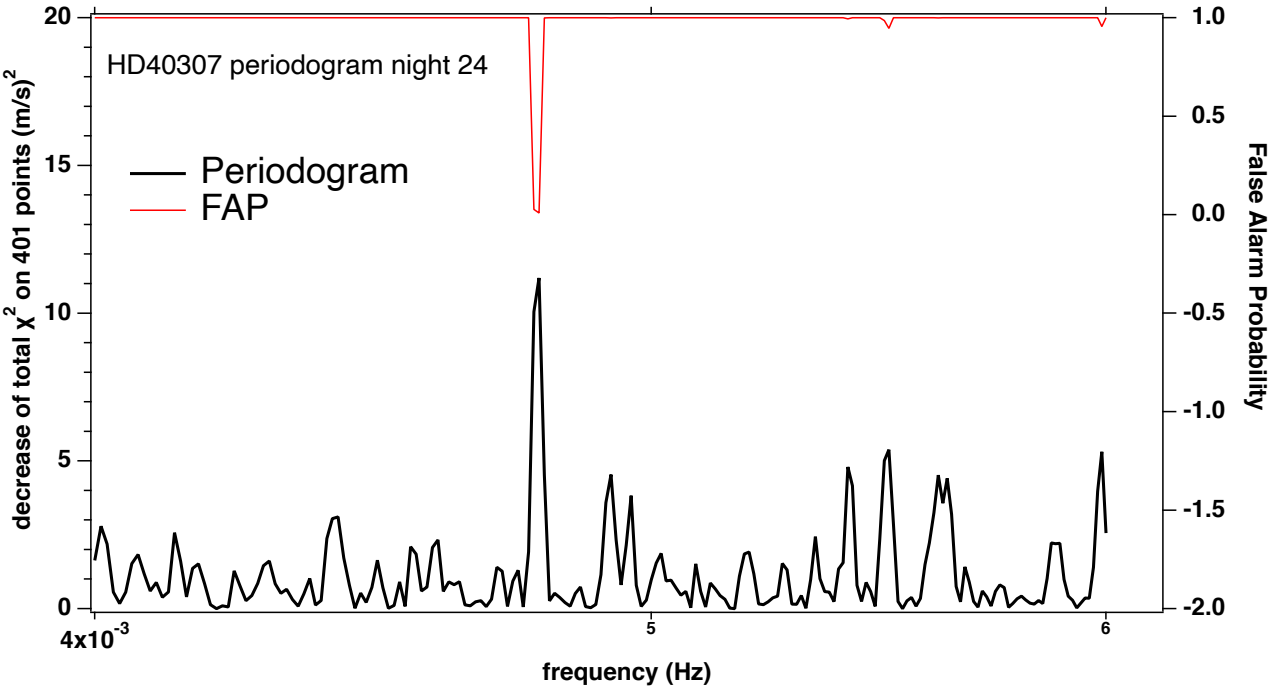


Figure A.4: Zoom on peak in A.3. Periodogram on the longest series of 401 measurements of night 24 December 2018. RV is determined form a Gaussian fit to CCF. The frequency step is 10^{-5} Hz. There is one single peak at frequency 0.00478 Hz, corresponding to a period of 209.5, with a FAP (False Alarm Probability) of less than 1%.

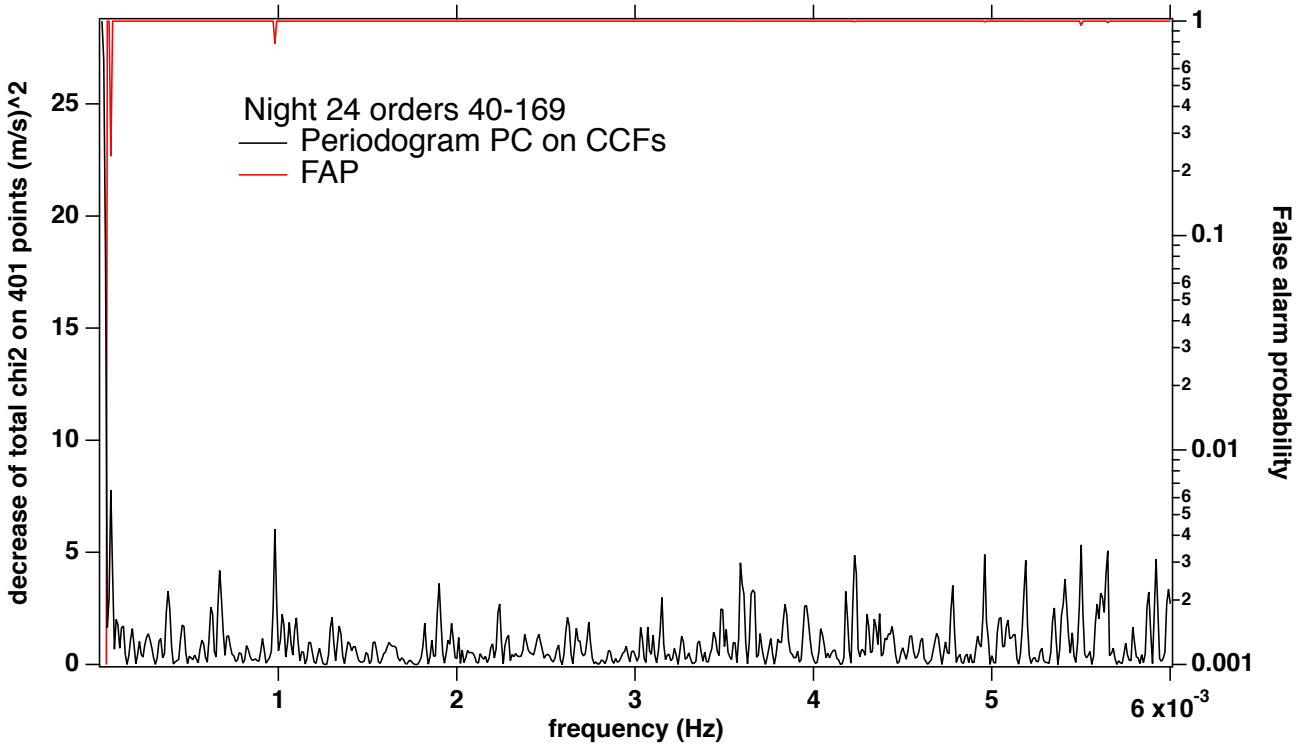


Figure A.5: Lomb-Scargle periodogram on the longest series of 401 measurements of night 24 December 2018, orders (40-169, blue and red). The time series was produced using Connes formula on CCFs of individual orders, and displayed (black) in figure 3.19. No peak is detected.

Bibliography

L. Acuña, M. Deleuil, and O. Mousis. Interior-atmosphere modelling to assess the observability of rocky planets with JWST. , 677:A14, Sept. 2023. doi: 10.1051/0004-6361/202245736.

R. Allart. WASP-127b: a misaligned planet with a partly cloudy atmosphereand tenuous sodium signature seen by ESPRESSO. In *European Planetary Science Congress*, pages EPSC2021–438, Sept. 2021. doi: 10.5194/epsc2021-438.

R. Allart, L. Pino, C. Lovis, S. G. Sousa, N. Casasayas-Barris, M. R. Zapatero Osorio, M. Cretignier, E. Palle, F. Pepe, S. Cristiani, R. Rebolo, N. C. Santos, F. Borsa, V. Bourrier, O. D. S. Demangeon, D. Ehrenreich, B. Lavie, M. Lendl, J. Lillo-Box, G. Micela, M. Oshagh, A. Sozzetti, H. Tabernero, V. Adibekyan, C. Allende Prieto, Y. Alibert, M. Amate, W. Benz, F. Bouchy, A. Cabral, H. Dekker, V. D'Odorico, P. Di Marcantonio, X. Dumusque, P. Figueira, R. Genova Santos, J. I. González Hernández, G. Lo Curto, A. Manescau, C. J. A. P. Martins, D. Mégevand, A. Mehner, P. Molaro, N. J. Nunes, E. Poretti, M. Riva, A. Suárez Mascareño, S. Udry, and F. Zerbi. WASP-127b: a misaligned planet with a partly cloudy atmosphere and tenuous sodium signature seen by ESPRESSO. *Astronomy & Astrophysics*, 644:A155, Dec. 2020. doi: 10.1051/0004-6361/202039234.

R. Allart, C. Lovis, J. Faria, X. Dumusque, D. Sosnowska, P. Figueira, A. M. Silva, A. Mehner, F. Pepe, S. Cristiani, R. Rebolo, N. C. Santos, V. Adibekyan, G. Cupani, P. Di Marcantonio, V. D'Odorico, J. I. González Hernández, C. J. A. P. Martins, D. Milaković, N. J. Nunes, A. Sozzetti, A. Suárez Mascareño, H. Tabernero, and M. R. Zapatero Osorio. Automatic model-based telluric correction for the ESPRESSO data reduction software. Model description and application to radial velocity computation. *Astronomy & Astrophysics*, 666:A196, Oct. 2022. doi: 10.1051/0004-6361/202243629.

D. R. Anderson, C. Hellier, M. Gillon, A. H. M. J. Triaud, B. Smalley, L. Hebb, A. Collier Cameron, P. F. L. Maxted, D. Queloz, R. G. West, S. J. Bentley, B. Enoch, K. Horne, T. A. Lister, M. Mayor, N. R. Parley, F. Pepe, D. Pollacco, D. Ségransan, S. Udry, and D. M. Wilson. WASP-17b: An Ultra-Low Density Planet in a Probable Retrograde Orbit. *Astrophysical Journal*, 709(1):159–167, Jan. 2010. doi: 10.1088/0004-637X/709/1/159.

G. Anglada-Escudé and R. P. Butler. The HARPS-TERRA Project. I. Description of the Algorithms, Performance, and New Measurements on a Few Remarkable Stars Observed by HARPS. *Astrophysical Journals*, 200(2):15, June 2012. doi: 10.1088/0067-0049/200/2/15.

É. Artigau, N. Astudillo-Defru, X. Delfosse, F. Bouchy, X. Bonfils, C. Lovis, F. Pepe, C. Moutou, J.-F. Donati, R. Doyon, and L. Malo. Telluric-line subtraction in high-accuracy velocimetry: a PCA-based approach. In A. B. Peck, C. R. Benn, and R. L. Seaman, editors, *Observatory Operations: Strategies, Processes, and Systems V*, volume 9149 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 914905, July 2014. doi: 10.1117/12.2056385.

É. Artigau, C. Cadieux, N. J. Cook, R. Doyon, T. Vandal, J.-F. Donati, C. Moutou, X. Delfosse, P. Fouqué, E. Martioli, F. Bouchy, J. Parsons, A. Carmona, X. Dumusque, N. Astudillo-Defru, X. Bonfils, and L. Mignon. Line-by-line Velocity Measurements: an Outlier-resistant Method for Precision Velocimetry. *Astronomical Journal*, 164(3):84, Sept. 2022. doi: 10.3847/1538-3881/ac7ce6.

Ö. Başẗurk, T. H. Dall, R. Collet, G. Lo Curto, and S. O. Selam. Bisectors of the HARPS cross-correlation function. The dependence on stellar atmospheric parameters. , 535:A17, Nov. 2011. doi: 10.1051/0004-6361/201117740.

R. V. Baluev. PlanetPack: Radial-velocity time-series analysis tool. Astrophysics Source Code Library, record ascl:1311.004, Nov. 2013.

R. V. Baluev. PlanetPack3: A radial-velocity and transit analysis tool for exoplanets. *Astronomy and Computing*, 25:221–229, Oct. 2018. doi: 10.1016/j.ascom.2018.10.005.

A. Baranne. White Pupil Story or Evolution of a Spectrographic Mounting. In *Very Large Telescopes and their Instrumentation, Vol. 2*, volume 30 of *European Southern Observatory Conference and Workshop Proceedings*, page 1195, Oct. 1988.

A. Baranne, M. Mayor, and J. L. Poncet. Coravel— A new tool for radial velocity measurements. *Vistas in Astronomy*, 23(4):279–316, Jan. 1979. doi: 10.1016/0083-6656(79)90016-3.

A. Baranne, D. Queloz, M. Mayor, G. Adrianzyk, G. Knispel, D. Kohler, D. Lacroix, J. P. Meunier, G. Rimbaud, and A. Vin. ELODIE: A spectrograph for accurate radial velocity measurements. *Astronomy & Astrophysicss*, 119:373–390, Oct. 1996.

J. L. Baudino, B. Bézard, A. Boccaletti, M. Bonnefoy, A. M. Lagrange, and R. Galicher. Interpreting the photometry and spectroscopy of directly imaged planets: a new atmospheric model applied to β Pictoris b and SPHERE observations. *Astronomy & Astrophysics*, 582:A83, Oct. 2015. doi: 10.1051/0004-6361/201526332.

T. G. Beatty, J. Pepper, R. J. Siverd, J. D. Eastman, A. Bieryla, D. W. Latham, L. A. Buchhave, E. L. N. Jensen, M. Manner, K. G. Stassun, B. S. Gaudi, P. Berlind, M. L. Calkins, K. Collins, D. L. DePoy, G. A. Esquerdo, B. J. Fulton, G. Fűrész, J. C. Geary, A. Gould, L. Hebb, J. F. Kielkopf, J. L. Marshall, R. Pogge, K. Z. Stanek, R. P. Stefanik, R. Street, A. H. Szentgyorgyi, M. Trueblood, P. Trueblood, and A. M. Stutz. KELT-2Ab: A Hot Jupiter Transiting the Bright (V = 8.77) Primary Star of a Binary System. *Astrophysical Journal, Letters*, 756(2):L39, Sept. 2012. doi: 10.1088/2041-8205/756/2/L39.

M. Bedell, D. W. Hogg, D. Foreman-Mackey, B. T. Montet, and R. Luger. WOBBLE: A Data-driven Analysis Technique for Time-series Stellar Spectra. *Astronomical Journal*, 158(4):164, Oct. 2019. doi: 10.3847/1538-3881/ab40a7.

J.-L. Bertaux and A. Ivanova. A numerical inversion of m sin i exoplanet distribution: the sub-Saturn desert is more depleted than observed and hint of a Uranus mass gap. *Monthly Notices of the RAS*, 512(4):5552–5571, June 2022. doi: 10.1093/mnras/stac777.

J. L. Bertaux, R. Lallement, S. Ferron, C. Boonne, and R. Bodichon. TAPAS, a web-based service of atmospheric transmission computation for astronomy. *Astronomy & Astrophysics*, 564:A46, Apr. 2014. doi: 10.1051/0004-6361/201322383.

D. Blain, B. Charnay, and B. Bézard. 1D atmospheric modelling of K2-18b. In *European Planetary Science Congress*, pages EPSC2020–245, Sept. 2020. doi: 10.5194/epsc2020-245.

I. Boisse, A. Eggenberger, N. C. Santos, C. Lovis, F. Bouchy, G. Hébrard, L. Arnold, X. Bonfils, X. Delfosse, M. Desort, R. F. Díaz, D. Ehrenreich, T. Forveille, A. Gallenne, A. M. Lagrange, C. Moutou, S. Udry, F. Pepe, C. Perrier, S. Perruchot, F. Pont, D. Queloz, A. Santerne, D. Ségransan, and A. Vidal-Madjar. The SOPHIE search for northern extrasolar planets. III. A Jupiter-mass companion around HD 109246. *Astronomy & Astrophysics*, 523:A88, Nov. 2010. doi: 10.1051/0004-6361/201014909.

A. Boucher, A. Darveau-Bernier, S. Pelletier, D. Lafrenière, É. Artigau, N. J. Cook, R. Allart, M. Radica, R. Doyon, B. Benneke, L. Arnold, X. Bonfils, V. Bourrier, R. Cloutier, J. Gomes da Silva, E. Deibert, X. Delfosse, J.-F. Donati, D. Ehrenreich, P. Figueira, T. Forveille, P. Fouqué, J. Gagné, E. Gaidos, G. Hébrard, R. Jayawardhana, B. Klein, C. Lovis, J. H. C. Martins, E. Martioli, C. Moutou, and N. C. Santos. Characterizing Exoplanetary Atmospheres at High Resolution with SPIRou: Detection of Water on HD 189733 b. *Astronomical Journal*, 162(6):233, Dec. 2021. doi: 10.3847/1538-3881/ac1f8e.

F. Bouchy, F. Pepe, and D. Queloz. Fundamental photon noise limit to radial velocity measurements. *Astronomy & Astrophysics*, 374:733–739, Aug. 2001. doi: 10.1051/0004-6361:20010730.

F. Bouchy, S. Udry, M. Mayor, C. Moutou, F. Pont, N. Iribarne, R. da Silva, S. Ilovaisky, D. Queloz, N. C. Santos, D. Ségransan, and S. Zucker. ELODIE metallicity-biased search for transiting Hot Jupiters. II. A very hot Jupiter transiting the bright K star HD 189733. , 444(1):L15–L19, Dec. 2005. doi: 10.1051/0004-6361:200500201.

F. Bouchy, F. Wildi, and J. I. González Hernández. The new Near-Infrared Adaptive-Optics assisted high-resolution NIRPS spectrograph on the ESO 3.6m. In *European Planetary Science Congress*, pages EPSC2022–937, Sept. 2022. doi: 10.5194/epsc2022-937.

V. Bourrier, A. Lecavelier des Etangs, H. Dupuy, D. Ehrenreich, A. Vidal-Madjar, G. Hébrard, G. E. Ballester, J. M. Désert, R. Ferlet, D. K. Sing, and P. J. Wheatley. Atmospheric escape from HD 189733b observed in H I Lyman-α: detailed analysis of HST/STIS September 2011 observations. *Astronomy & Astrophysics*, 551:A63, Mar. 2013. doi: 10.1051/0004-6361/201220533.

D. J. Bower, K. Hakim, P. A. Sossi, and P. Sanan. Retention of Water in Terrestrial Magma Oceans

and Carbon-rich Early Atmospheres. *Planetary Science Journal*, 3(4):93, Apr. 2022. doi: 10.3847/PSJ/ac5fb1.

L. R. Brown, C. M. Humphrey, and R. R. Gamache. CO_2-broadened water in the pure rotation and ν_2 fundamental regions. *Journal of Molecular Spectroscopy*, 246(1):1–21, Nov. 2007. doi: 10.1016/j.jms.2007.07.010.

A. Caldas, J. Leconte, F. Selsis, I. P. Waldmann, P. Bordé, M. Rocchetto, and B. Charnay. Effects of a fully 3D atmospheric structure on exoplanet transmission spectra: retrieval biases due to day-night temperature gradients. *Astronomy and Astrophysics*, 623:A161, Mar. 2019. doi: 10.1051/0004-6361/201834384.

F. Castelli and R. L. Kurucz. New Fe ii energy levels from stellar spectra. *Astronomy & Astrophysics*, 520:A57, Sept. 2010. doi: 10.1051/0004-6361/201015126.

J. Chaffee, F. H. and R. E. White. A survey of interstellar neutral potassium .I. Abundances and physical conditions in clouds toward 188 early-type stars. , 50:169–198, Nov. 1982. doi: 10.1086/190824.

D. Charbonneau, T. M. Brown, R. W. Noyes, and R. L. Gilliland. Detection of an Extrasolar Planet Atmosphere. *Astrophysical Journal*, 568(1):377–384, Mar. 2002. doi: 10.1086/338770.

B. Charnay, B. Bézard, J. L. Baudino, M. Bonnefoy, A. Boccaletti, and R. Galicher. A Self-consistent Cloud Model for Brown Dwarfs and Young Giant Exoplanets: Comparison with Photometric and Spectroscopic Observations. *Astrophysical Journal*, 854(2):172, Feb. 2018. doi: 10.3847/1538-4357/aaac7d.

A. Chelli. Optimizing Doppler estimates for extrasolar planet detection. I. A specific algorithm for shifted spectra. , 358:L59–L62, June 2000.

S. A. Clough, M. J. Iacono, and J.-L. Moncet. Line-by-Line Calculations of Atmospheric Fluxes and Cooling Rates: Application to Water Vapor. *Journal of Geophysics Research*, 97(D14):15,761–15,785, Oct. 1992. doi: 10.1029/92JD01419.

S. A. Clough, M. W. Shephard, E. J. Mlawer, J. S. Delamere, M. J. Iacono, K. Cady-Pereira, S. Boukabara, and P. D. Brown. Atmospheric radiative transfer modeling: a summary of the AER

codes. *Journal of Quantitiative Spectroscopy and Radiative Transfer*, 91(2):233–244, Mar. 2005. doi: 10.1016/j.jqsrt.2004.05.058.

A. Coffinet, C. Lovis, X. Dumusque, and F. Pepe. New wavelength calibration of the HARPS spectrograph. *Astronomy & Astrophysics*, 629:A27, Sept. 2019. doi: 10.1051/0004-6361/201833272.

P. Connes. Absolute Astronomical Accelerometry. *Astrophysics and Space Science*, 110(2):211–255, Mar. 1985. doi: 10.1007/BF00653671.

N. J. Cook, É. Artigau, R. Doyon, M. Hobson, E. Martioli, F. Bouchy, C. Moutou, A. Carmona, C. Usher, P. Fouqué, L. Arnold, X. Delfosse, I. Boisse, C. Cadieux, T. Vandal, J.-F. Donati, and A. Deslières. APERO: A PipelinE to Reduce Observations-Demonstration with SPIRou. *Publications of the ASP*, 134(1041):114509, Nov. 2022. doi: 10.1088/1538-3873/ac9e74.

M. Cretignier, X. Dumusque, R. Allart, F. Pepe, and C. Lovis. Measuring precise radial velocities on individual spectral lines. II. Dependance of stellar activity signal on line depth. , 633:A76, Jan. 2020. doi: 10.1051/0004-6361/201936548.

M. Cretignier, X. Dumusque, N. C. Hara, and F. Pepe. YARARA: Significant improvement in RV precision through post-processing of spectral time series. *Astronomy & Astrophysics*, 653:A43, Sept. 2021. doi: 10.1051/0004-6361/202140986.

S. Dalal, R. D. Haywood, A. Mortier, W. J. Chaplin, and N. Meunier. Predicting convective blueshift and radial-velocity dispersion due to granulation for FGK stars. , 525(3):3344–3353, Nov. 2023. doi: 10.1093/mnras/stad2393.

D. Deming, A. Wilkins, P. McCullough, A. Burrows, J. J. Fortney, E. Agol, I. Dobbs-Dixon, N. Madhusudhan, N. Crouzet, J.-M. Desert, R. L. Gilliland, K. Haynes, H. A. Knutson, M. Line, Z. Magic, A. M. Mandell, S. Ranjan, D. Charbonneau, M. Clampin, S. Seager, and A. P. Showman. Infrared Transmission Spectroscopy of the Exoplanets HD 209458b and XO-1b Using the Wide Field Camera-3 on the Hubble Space Telescope. *Astrophysical Journal*, 774(2):95, Sept. 2013. doi: 10.1088/0004-637X/774/2/95.

R. F. Díaz, D. Ségransan, S. Udry, C. Lovis, F. Pepe, X. Dumusque, M. Marmier, R. Alonso, W. Benz, F. Bouchy, A. Coffinet, A. Collier Cameron, M. Deleuil, P. Figueira, M. Gillon, G. Lo Curto,

M. Mayor, C. Mordasini, F. Motalebi, C. Moutou, D. Pollacco, E. Pompei, D. Queloz, N. Santos, and A. Wyttenbach. The HARPS search for southern extra-solar planets. XXXVIII. Bayesian re-analysis of three systems. New super-Earths, unconfirmed signals, and magnetic cycles. *Astronomy & Astrophysics*, 585:A134, Jan. 2016. doi: 10.1051/0004-6361/201526729.

J. F. Donati, M. Semel, B. D. Carter, D. E. Rees, and A. Collier Cameron. Spectropolarimetric observations of active stars. , 291(4):658–682, Nov. 1997. doi: 10.1093/mnras/291.4.658.

J. F. Donati, D. Kouach, C. Moutou, R. Doyon, X. Delfosse, E. Artigau, S. Baratchart, M. Lacombe, G. Barrick, G. Hébrard, F. Bouchy, L. Saddlemyer, L. Parès, P. Rabou, Y. Micheau, F. Dolon, V. Reshetov, Z. Challita, A. Carmona, N. Striebig, S. Thibault, E. Martioli, N. Cook, P. Fouqué, T. Vermeulen, S. Y. Wang, L. Arnold, F. Pepe, I. Boisse, P. Figueira, J. Bouvier, T. P. Ray, C. Feugeade, J. Morin, S. Alencar, M. Hobson, B. Castilho, S. Udry, N. C. Santos, O. Hernandez, T. Benedict, P. Vallée, G. Gallou, M. Dupieux, M. Larrieu, S. Perruchot, R. Sottile, F. Moreau, C. Usher, M. Baril, F. Wildi, B. Chazelas, L. Malo, X. Bonfils, D. Loop, D. Kerley, I. Wevers, J. Dunn, J. Pazder, S. Macdonald, B. Dubois, E. Carrié, H. Valentin, F. Henault, C. H. Yan, and T. Steinmetz. SPIRou: NIR velocimetry and spectropolarimetry at the CFHT. *Monthly Notices of the RAS*, 498(4):5684–5703, Nov. 2020. doi: 10.1093/mnras/staa2569.

D. Dravins, L. Lindegren, and A. Nordlund. Solar granulation - Influence of convection on spectral line asymmetries and wavelength shifts. , 96(1-2):345–364, Mar. 1981.

X. Dumusque. Measuring precise radial velocities on individual spectral lines. I. Validation of the method and application to mitigate stellar activity. , 620:A47, Nov. 2018. doi: 10.1051/0004-6361/201833795.

H. Fan, L. M. Hobbs, J. A. Dahlstrom, D. E. Welty, D. G. York, B. Rachford, T. P. Snow, P. Sonnentrucker, N. Baskes, and G. Zhao. The Apache Point Observatory Catalog of Optical Diffuse Interstellar Bands. , 878(2):151, June 2019. doi: 10.3847/1538-4357/ab1b74.

P. Figueira, F. Kerber, A. Chacon, C. Lovis, N. C. Santos, G. Lo Curto, M. Sarazin, and F. Pepe. Comparing radial velocities of atmospheric lines with radiosonde measurements. *Monthly Notices of the RAS*, 420(4):2874–2883, Mar. 2012. doi: 10.1111/j.1365-2966.2011.20015.x.

P. Figueira, G. Lo Curto, and A. Mehner. *Very Large TelescopeParanal Science Operations ESPRESSO User Manual*. 2021.

J. Fraine, D. Deming, B. Benneke, H. Knutson, A. Jordán, N. Espinoza, N. Madhusudhan, A. Wilkins, and K. Todorov. Water vapour absorption in the clear atmosphere of a Neptune-sized exoplanet. *Nature*, 513(7519):526–529, Sept. 2014. doi: 10.1038/nature13785.

G. A. Galazutdinov. DECH: A Software Package for Astronomical Spectral Data Processing and Analysis. *Astrophysical Bulletin*, 77(4):519–529, Dec. 2022. doi: 10.1134/S1990341322040034.

F. Galland, A. M. Lagrange, S. Udry, A. Chelli, F. Pepe, D. Queloz, J. L. Beuzit, and M. Mayor. Extrasolar planets and brown dwarfs around A-F type stars. I. Performances of radial velocity measurements, first analyses of variations. , 443(1):337–345, Nov. 2005. doi: 10.1051/0004-6361:20052938.

S. Gandhi, M. Brogi, S. N. Yurchenko, J. Tennyson, P. A. Coles, R. K. Webb, J. L. Birkby, G. Guilluy, G. A. Hawker, N. Madhusudhan, A. S. Bonomo, and A. Sozzetti. Molecular cross-sections for high-resolution spectroscopy of super-Earths, warm Neptunes, and hot Jupiters. *Monthly Notices of the RAS*, 495(1):224–237, June 2020. doi: 10.1093/mnras/staa981.

J. I. González Hernández, R. Rebolo, L. Pasquini, G. Lo Curto, P. Molaro, E. Caffau, H. G. Ludwig, M. Steffen, M. Esposito, A. Suárez Mascareño, B. Toledo-Padrón, R. A. Probst, T. W. Hänsch, R. Holzwarth, A. Manescau, T. Steinmetz, T. Udem, and T. Wilken. The solar gravitational redshift from HARPS-LFC Moon spectra⋆. A test of the general theory of relativity. *Astronomy & Astrophysics*, 643:A146, Nov. 2020. doi: 10.1051/0004-6361/202038937.

R. Goody, R. West, L. Chen, and D. Crisp. The correlated-k method for radiation calculations in nonhomogeneous atmospheres. , 42:539–550, Dec. 1989. doi: 10.1016/0022-4073(89)90044-7.

R. M. Goody and Y. L. Yung. *Atmospheric radiation : theoretical basis*. 1989.

I. E. Gordon, L. S. Rothman, Y. Tan, R. V. Kochanov, and C. Hill. HITRAN2016: Part I. Line lists for H_2O, CO_2, O_3, N_2O, CO, CH_4, and O_2. In *72nd International Symposium on Molecular Spectroscopy*, page TJ08, June 2017. doi: 10.15278/isms.2017.TJ08.

I. E. Gordon, L. S. Rothman, R. J. Hargreaves, R. Hashemi, E. V. Karlovets, F. M. Skinner, E. K. Conway, C. Hill, R. V. Kochanov, Y. Tan, P. Wcisło, A. A. Finenko, K. Nelson, P. F. Bernath, M. Birk, V. Boudon, A. Campargue, K. V. Chance, A. Coustenis, B. J. Drouin, J. M. Flaud, R. R. Gamache, J. T. Hodges, D. Jacquemart, E. J. Mlawer, A. V. Nikitin, V. I. Perevalov, M. Rotger, J. Tennyson, G. C. Toon, H. Tran, V. G. Tyuterev, E. M. Adkins, A. Baker, A. Barbe, E. Canè, A. G. Császár, A. Dudaryonok, O. Egorov, A. J. Fleisher, H. Fleurbaey, A. Foltynowicz, T. Furtenbacher, J. J. Harrison, J. M. Hartmann, V. M. Horneman, X. Huang, T. Karman, J. Karns, S. Kassi, I. Kleiner, V. Kofman, F. Kwabia-Tchana, N. N. Lavrentieva, T. J. Lee, D. A. Long, A. A. Lukashevskaya, O. M. Lyulin, V. Y. Makhnev, W. Matt, S. T. Massie, M. Melosso, S. N. Mikhailenko, D. Mondelain, H. S. P. Müller, O. V. Naumenko, A. Perrin, O. L. Polyansky, E. Raddaoui, P. L. Raston, Z. D. Reed, M. Rey, C. Richard, R. Tóbiás, I. Sadiek, D. W. Schwenke, E. Starikova, K. Sung, F. Tamassia, S. A. Tashkun, J. Vander Auwera, I. A. Vasilenko, A. A. Vigasin, G. L. Villanueva, B. Vispoel, G. Wagner, A. Yachmenev, and S. N. Yurchenko. The HITRAN2020 molecular spectroscopic database. *Journal of Quantitiative Spectroscopy and Radiative Transfer*, 277:107949, Jan. 2022. doi: 10.1016/j.jqsrt.2021.107949.

D. F. Gray. The Third Signature of Stellar Granulation. , 697(2):1032–1043, June 2009. doi: 10.1088/0004-637X/697/2/1032.

K. Gullikson, S. Dodson-Robinson, and A. Kraus. Correcting for Telluric Absorption: Methods, Case Studies, and Release of the TelFit Code. *Astronomical Journal*, 148(3):53, Sept. 2014. doi: 10.1088/0004-6256/148/3/53.

G. Hébrard, J. M. Désert, R. F. Díaz, I. Boisse, F. Bouchy, A. Lecavelier Des Etangs, C. Moutou, D. Ehrenreich, L. Arnold, X. Bonfils, X. Delfosse, M. Desort, A. Eggenberger, T. Forveille, J. Gregorio, A. M. Lagrange, C. Lovis, F. Pepe, C. Perrier, F. Pont, D. Queloz, A. Santerne, N. C. Santos, D. Ségransan, D. K. Sing, S. Udry, and A. Vidal-Madjar. Observation of the full 12-hour-long transit of the exoplanet HD 80606b. Warm-Spitzer photometry and SOPHIE spectroscopy. *Astronomy & Astrophysics*, 516:A95, June 2010. doi: 10.1051/0004-6361/201014327.

A. E. Hedin. Extension of the MSIS thermosphere model into the middle and lower atmosphere. *Journal of Geophysics Research*, 96(A2):1159–1172, Feb. 1991. doi: 10.1029/90JA02125.

U. Heiter, K. Lind, M. Bergemann, M. Asplund, Š. Mikolaitis, P. S. Barklem, T. Masseron, P. de

Laverny, L. Magrini, B. Edvardsson, H. Jönsson, J. C. Pickering, N. Ryde, A. Bayo Arán, T. Bensby, A. R. Casey, S. Feltzing, P. Jofré, A. J. Korn, E. Pancino, F. Damiani, A. Lanzafame, C. Lardo, L. Monaco, L. Morbidelli, R. Smiljanic, C. Worley, S. Zaggia, S. Randich, and G. F. Gilmore. Atomic data for the Gaia-ESO Survey. *Astronomy & Astrophysics*, 645:A106, Jan. 2021. doi: 10.1051/0004-6361/201936291.

M. Holmberg and N. Madhusudhan. A First Look at CRIRES+: Performance Assessment and Exoplanet Spectroscopy. *Astronomical Journal*, 164(3):79, Sept. 2022. doi: 10.3847/1538-3881/ac77eb.

J. R. Holt. Spectroscopic Determination of Stellar Rotation. *Astronomy and Astro-Physics (formerly The Sidereal Messenger)*, 12:646, Jan. 1893.

C. M. Huitson, D. K. Sing, F. Pont, J. J. Fortney, A. S. Burrows, P. A. Wilson, G. E. Ballester, N. Nikolov, N. P. Gibson, D. Deming, S. Aigrain, T. M. Evans, G. W. Henry, A. Lecavelier des Etangs, A. P. Showman, A. Vidal-Madjar, and K. Zahnle. An HST optical-to-near-IR transmission spectrum of the hot Jupiter WASP-19b: detection of atmospheric water and likely absence of TiO. *Monthly Notices of the RAS*, 434(4):3252–3274, Oct. 2013. doi: 10.1093/mnras/stt1243.

A. Ivanova, R. Lallement, J. L. Vergely, and C. Hottier. Toward a 3D kinetic tomography of Taurus clouds. I. Linking neutral potassium and dust. , 652:A22, Aug. 2021a. doi: 10.1051/0004-6361/202140514.

A. Ivanova, R. Lallement, and J. L. Bertaux. Improved precision of radial velocity measurements after correction for telluric absorption. *Astronomy & Astrophysics*, 673:A56, May 2023. doi: 10.1051/0004-6361/202245089.

A. E. Ivanova, O. Y. Yakovlev, V. I. Ananyeva, I. A. Shashkova, A. V. Tavrov, and J. L. Bertaux. The "Detectability Window" Method to Take into Account Observational Selection in the Statistics of Exoplanets Discovered through Radial Velocity Measurements. *Astronomy Letters*, 47(1):43–49, Jan. 2021b. doi: 10.1134/S1063773721010059.

JWST Transiting Exoplanet Community Early Release Science Team, E.-M. Ahrer, L. Alderson, N. M. Batalha, N. E. Batalha, J. L. Bean, T. G. Beatty, T. J. Bell, B. Benneke, Z. K. Berta-Thompson, A. L. Carter, I. J. M. Crossfield, N. Espinoza, A. D. Feinstein, J. J. Fortney, N. P.

Gibson, J. M. Goyal, E. M. R. Kempton, J. Kirk, L. Kreidberg, M. López-Morales, M. R. Line, J. D. Lothringer, S. E. Moran, S. Mukherjee, K. Ohno, V. Parmentier, C. Piaulet, Z. Rustamkulov, E. Schlawin, D. K. Sing, K. B. Stevenson, H. R. Wakeford, N. H. Allen, S. M. Birkmann, J. Brande, N. Crouzet, P. E. Cubillos, M. Damiano, J.-M. Désert, P. Gao, J. Harrington, R. Hu, S. Kendrew, H. A. Knutson, P.-O. Lagage, J. Leconte, M. Lendl, R. J. MacDonald, E. M. May, Y. Miguel, K. Molaverdikhani, J. I. Moses, C. A. Murray, M. Nehring, N. K. Nikolov, D. J. M. Petit dit de la Roche, M. Radica, P.-A. Roy, K. G. Stassun, J. Taylor, W. C. Waalkes, P. Wachiraphan, L. Welbanks, P. J. Wheatley, K. Aggarwal, M. K. Alam, A. Banerjee, J. K. Barstow, J. Blecic, S. L. Casewell, Q. Changeat, K. L. Chubb, K. D. Colón, L.-P. Coulombe, T. Daylan, M. de Val-Borro, L. Decin, L. A. Dos Santos, L. Flagg, K. France, G. Fu, A. García Muñoz, J. E. Gizis, A. Glidden, D. Grant, K. Heng, T. Henning, Y.-C. Hong, J. Inglis, N. Iro, T. Kataria, T. D. Komacek, J. E. Krick, E. K. H. Lee, N. K. Lewis, J. Lillo-Box, J. Lustig-Yaeger, L. Mancini, A. M. Mandell, M. Mansfield, M. S. Marley, T. Mikal-Evans, G. Morello, M. C. Nixon, K. Ortiz Ceballos, A. A. A. Piette, D. Powell, B. V. Rackham, L. Ramos-Rosado, E. Rauscher, S. Redfield, L. K. Rogers, M. T. Roman, G. M. Roudier, N. Scarsdale, E. L. Shkolnik, J. Southworth, J. J. Spake, M. E. Steinrueck, X. Tan, J. K. Teske, P. Tremblin, S.-M. Tsai, G. S. Tucker, J. D. Turner, J. A. Valenti, O. Venot, I. P. Waldmann, N. L. Wallack, X. Zhang, and S. Zieba. Identification of carbon dioxide in an exoplanet atmosphere. *Nature*, 614(7949):649–652, Feb. 2023. doi: 10.1038/s41586-022-05269-w.

H.-U. Kaeufl, P. Ballester, P. Biereichel, B. Delabre, R. Donaldson, R. Dorn, E. Fedrigo, G. Finger, G. Fischer, F. Franza, D. Gojak, G. Huster, Y. Jung, J.-L. Lizon, L. Mehrgan, M. Meyer, A. Moorwood, J.-F. Pirard, J. Paufique, E. Pozna, R. Siebenmorgen, A. Silber, J. Stegmeier, and S. Wegerer. CRIRES: a high-resolution infrared spectrograph for ESO's VLT. In A. F. M. Moorwood and M. Iye, editors, *Ground-based Instrumentation for Astronomy*, volume 5492 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 1218–1227, Sept. 2004. doi: 10.1117/12.551480.

J. F. Kasting. Runaway and moist greenhouse atmospheres and the evolution of Earth and Venus. *Icarus*, 74(3):472–494, June 1988. doi: 10.1016/0019-1035(88)90116-9.

R. V. Kochanov, I. E. Gordon, L. S. Rothman, P. Wcisło, C. Hill, and J. S. Wilzewski. HITRAN Application Programming Interface (HAPI): A comprehensive approach to working with

spectroscopic data. *Journal of Quantitiative Spectroscopy and Radiative Transfer*, 177:15–30, July 2016. doi: 10.1016/j.jqsrt.2016.03.005.

R. K. Kopparapu, R. Ramirez, J. F. Kasting, V. Eymet, T. D. Robinson, S. Mahadevan, R. C. Terrien, S. Domagal-Goldman, V. Meadows, and R. Deshpande. Habitable Zones around Main-sequence Stars: New Estimates. *Astrophysical Journal*, 765(2):131, Mar. 2013. doi: 10.1088/0004-637X/765/2/131.

L. Kreidberg. Exoplanet Atmosphere Measurements from Transmission Spectroscopy and Other Planet Star Combined Light Observations. In H. J. Deeg and J. A. Belmonte, editors, *Handbook of Exoplanets*, page 100. 2018. doi: 10.1007/978-3-319-55333-7_100.

L. Kreidberg, J. L. Bean, J.-M. Désert, M. R. Line, J. J. Fortney, N. Madhusudhan, K. B. Stevenson, A. P. Showman, D. Charbonneau, P. R. McCullough, S. Seager, A. Burrows, G. W. Henry, M. Williamson, T. Kataria, and D. Homeier. A Precise Water Abundance Measurement for the Hot Jupiter WASP-43b. *Astrophysical Journal, Letters*, 793(2):L27, Oct. 2014. doi: 10.1088/2041-8205/793/2/L27.

R. L. Kurucz. Including all the lines. *Canadian Journal of Physics*, 89:417–428, Apr. 2011. doi: 10.1139/p10-104.

R. L. Kurucz. ATLAS9: Model atmosphere program with opacity distribution functions. Astrophysics Source Code Library, record ascl:1710.017, Oct. 2017a.

R. L. Kurucz. Including all the lines: data releases for spectra and opacities. *Canadian Journal of Physics*, 95(9):825–827, Sept. 2017b. doi: 10.1139/cjp-2016-0794.

R. L. Kurucz. Including All the Lines: Data Releases for Spectra and Opacities through 2017. In *Workshop on Astrophysical Opacities*, volume 515 of *Astronomical Society of the Pacific Conference Series*, page 47, Aug. 2018.

A. A. Lacis and V. Oinas. A description of the correlated-k distribution method for modelling nongray gaseous absorption, thermal emission, and multiple scattering in vertically inhomogeneous atmospheres. , 96:9027–9064, May 1991. doi: 10.1029/90JD01945.

M. Lafarga, I. Ribas, C. Lovis, M. Perger, M. Zechmeister, F. F. Bauer, M. Kürster, M. Cortés-Contreras, J. C. Morales, E. Herrero, A. Rosich, D. Baroch, A. Reiners, J. A. Caballero, A. Quirrenbach, P. J. Amado, J. M. Alacid, V. J. S. Béjar, S. Dreizler, A. P. Hatzes, T. Henning, S. V. Jeffers, A. Kaminski, D. Montes, S. Pedraz, C. Rodríguez-López, and J. H. M. M. Schmitt. The CARMENES search for exoplanets around M dwarfs. Radial velocities and activity indicators from cross-correlation functions with weighted binary masks. *Astronomy & Astrophysics*, 636:A36, Apr. 2020. doi: 10.1051/0004-6361/201937222.

R. Lallement, C. Babusiaux, J. L. Vergely, D. Katz, F. Arenou, B. Valette, C. Hottier, and L. Capitanio. Gaia-2MASS 3D maps of Galactic interstellar dust within 3 kpc. , 625:A135, May 2019. doi: 10.1051/0004-6361/201834695.

K. W. F. Lam, F. Faedi, D. J. A. Brown, D. R. Anderson, L. Delrez, M. Gillon, G. Hébrard, M. Lendl, L. Mancini, J. Southworth, B. Smalley, A. H. M. Triaud, O. D. Turner, K. L. Hay, D. J. Armstrong, S. C. C. Barros, A. S. Bonomo, F. Bouchy, P. Boumis, A. Collier Cameron, A. P. Doyle, C. Hellier, T. Henning, E. Jehin, G. King, J. Kirk, T. Louden, P. F. L. Maxted, J. J. McCormac, H. P. Osborn, E. Palle, F. Pepe, D. Pollacco, J. Prieto-Arranz, D. Queloz, J. Rey, D. Ségransan, S. Udry, S. Walker, R. G. West, and P. J. Wheatley. From dense hot Jupiter to low-density Neptune: The discovery of WASP-127b, WASP-136b, and WASP-138b. , 599:A3, Mar. 2017. doi: 10.1051/0004-6361/201629403.

A. Lecavelier Des Etangs, A. Vidal-Madjar, J. M. Désert, and D. Sing. Rayleigh scattering by H_2 in the extrasolar planet HD 209458b. *Astronomy & Astrophysics*, 485(3):865–869, July 2008. doi: 10.1051/0004-6361:200809704.

J. Leconte. Spectral binning of precomputed correlated-k coefficients. *Astronomy & Astrophysics*, 645:A20, Jan. 2021. doi: 10.1051/0004-6361/202039040.

C. Leet, D. A. Fischer, and J. A. Valenti. Toward a Self-calibrating, Empirical, Light-weight Model for Tellurics in High-resolution Spectra. *Astronomical Journal*, 157(5):187, May 2019. doi: 10.3847/1538-3881/ab0d86.

F. Liebing, S. V. Jeffers, A. Reiners, and M. Zechmeister. Convective blueshift strengths of 810 F to M solar-type stars. , 654:A168, Oct. 2021. doi: 10.1051/0004-6361/202039607.

M. R. Line, K. B. Stevenson, J. Bean, J.-M. Desert, J. J. Fortney, L. Kreidberg, N. Madhusudhan, A. P. Showman, and H. Diamond-Lowe. No Thermal Inversion and a Solar Water Abundance for the Hot Jupiter HD 209458b from HST/WFC3 Spectroscopy. *Astronomical Journal*, 152(6):203, Dec. 2016. doi: 10.3847/0004-6256/152/6/203.

R. E. Lupu, K. Zahnle, M. S. Marley, L. Schaefer, B. Fegley, C. Morley, K. Cahoy, R. Freedman, and J. J. Fortney. The Atmospheres of Earthlike Planets after Giant Impact Events. *Astrophysical Journal*, 784(1):27, Mar. 2014. doi: 10.1088/0004-637X/784/1/27.

E. Marcq, A. Salvador, H. Massol, and A. Davaille. Thermal radiation of magma ocean planets using a 1-D radiative-convective model of H_2O-CO_2 atmospheres. *Journal of Geophysical Research (Planets)*, 122(7):1539–1553, July 2017. doi: 10.1002/2016JE005224.

M. Mayor and D. Queloz. A Jupiter-mass companion to a solar-type star. *Nature*, 378(6555):355–359, Nov. 1995. doi: 10.1038/378355a0.

M. Mayor, S. Udry, C. Lovis, F. Pepe, D. Queloz, W. Benz, J. L. Bertaux, F. Bouchy, C. Mordasini, and D. Segransan. The HARPS search for southern extra-solar planets. XIII. A planetary system with 3 super-Earths (4.2, 6.9, and 9.2 $M_\oplus$). *Astronomy & Astrophysics*, 493(2):639–644, Jan. 2009. doi: 10.1051/0004-6361:200810451.

D. B. McLaughlin. Some results of a spectrographic study of the Algol system. *Astrophysical Journal*, 60:22–31, July 1924. doi: 10.1086/142826.

N. Meunier. Stellar variability in radial velocity. *arXiv e-prints*, art. arXiv:2104.06072, Apr. 2021. doi: 10.48550/arXiv.2104.06072.

N. Meunier and A. M. Lagrange. The effects of granulation and supergranulation on Earth-mass planet detectability in the habitable zone around F6-K4 stars. , 642:A157, Oct. 2020. doi: 10.1051/0004-6361/202038376.

N. Meunier, A. M. Lagrange, and M. Desort. Reconstructing the solar integrated radial velocity using MDI/SOHO. , 519:A66, Sept. 2010. doi: 10.1051/0004-6361/201014199.

N. Meunier, A. M. Lagrange, S. Borgniet, and M. Rieutord. Using the Sun to estimate Earth-like

planet detection capabilities. VI. Simulation of granulation and supergranulation radial velocity and photometric time series. , 583:A118, Nov. 2015. doi: 10.1051/0004-6361/201525721.

P. Mollière, J. P. Wardenier, R. van Boekel, T. Henning, K. Molaverdikhani, and I. A. G. Snellen. petitRADTRANS. A Python radiative transfer package for exoplanet characterization and retrieval. *Astronomy & Astrophysics*, 627:A67, July 2019. doi: 10.1051/0004-6361/201935470.

C. Moutou, S. Dalal, J. F. Donati, E. Martioli, C. P. Folsom, É. Artigau, I. Boisse, F. Bouchy, A. Carmona, N. J. Cook, X. Delfosse, R. Doyon, P. Fouqué, G. Gaisné, G. Hébrard, M. Hobson, B. Klein, A. Lecavelier des Etangs, and J. Morin. Early science with SPIRou: near-infrared radial velocity and spectropolarimetry of the planet-hosting star HD 189733. , 642:A72, Oct. 2020. doi: 10.1051/0004-6361/202038108.

G. Nave, S. Johansson, R. C. M. Learner, A. P. Thorne, and J. W. Brault. A New Multiplet Table for Fe i. *Astrophysical Journals*, 94:221, Sept. 1994. doi: 10.1086/192079.

N. Nikolov, D. K. Sing, F. Pont, A. S. Burrows, J. J. Fortney, G. E. Ballester, T. M. Evans, C. M. Huitson, H. R. Wakeford, P. A. Wilson, S. Aigrain, D. Deming, N. P. Gibson, G. W. Henry, H. Knutson, A. Lecavelier des Etangs, A. P. Showman, A. Vidal-Madjar, and K. Zahnle. Hubble Space Telescope hot Jupiter transmission spectral survey: a detection of Na and strong optical absorption in HAT-P-1b. *Monthly Notices of the RAS*, 437(1):46–66, Jan. 2014. doi: 10.1093/mnras/stt1859.

N. Nikolov, D. K. Sing, N. P. Gibson, J. J. Fortney, T. M. Evans, J. K. Barstow, T. Kataria, and P. A. Wilson. VLT FORS2 Comparative Transmission Spectroscopy: Detection of Na in the Atmosphere of WASP-39b from the Ground. *Astrophysical Journal*, 832(2):191, Dec. 2016. doi: 10.3847/0004-637X/832/2/191.

N. Nikolov, G. Maciejewski, S. Constantinou, N. Madhusudhan, J. J. Fortney, B. Smalley, A. L. Carter, E. J. W. de Mooij, B. Drummond, N. P. Gibson, C. Helling, N. Mayne, T. Mikal-Evans, D. K. Sing, and J. Wilson. Ground-based Transmission Spectroscopy with VLT FORS2: Evidence for Faculae and Clouds in the Optical Spectrum of the Warm Saturn WASP-110b. *Astronomical Journal*, 162(3):88, Sept. 2021. doi: 10.3847/1538-3881/ac01da.

F. Pepe, M. Mayor, F. Galland, D. Naef, D. Queloz, N. C. Santos, S. Udry, and M. Burnet. The CORALIE survey for southern extra-solar planets VII. Two short-period Saturnian companions to HD 108147and HD 168746. *Astronomy & Astrophysics*, 388:632–638, June 2002. doi: 10.1051/0004-6361:20020433.

F. Pepe, S. Cristiani, R. Rebolo, N. C. Santos, H. Dekker, A. Cabral, P. Di Marcantonio, P. Figueira, G. Lo Curto, C. Lovis, M. Mayor, D. Mégevand, P. Molaro, M. Riva, M. R. Zapatero Osorio, M. Amate, A. Manescau, L. Pasquini, F. M. Zerbi, V. Adibekyan, M. Abreu, M. Affolter, Y. Alibert, M. Aliverti, R. Allart, C. Allende Prieto, D. Álvarez, D. Alves, G. Avila, V. Baldini, T. Bandy, S. C. C. Barros, W. Benz, A. Bianco, F. Borsa, V. Bourrier, F. Bouchy, C. Broeg, G. Calderone, R. Cirami, J. Coelho, P. Conconi, I. Coretti, C. Cumani, G. Cupani, V. D'Odorico, M. Damasso, S. Deiries, B. Delabre, O. D. S. Demangeon, X. Dumusque, D. Ehrenreich, J. P. Faria, A. Fragoso, L. Genolet, M. Genoni, R. Génova Santos, J. I. González Hernández, I. Hughes, O. Iwert, F. Kerber, J. Knudstrup, M. Landoni, B. Lavie, J. Lillo-Box, J. L. Lizon, C. Maire, C. J. A. P. Martins, A. Mehner, G. Micela, A. Modigliani, M. A. Monteiro, M. J. P. F. G. Monteiro, M. Moschetti, M. T. Murphy, N. Nunes, L. Oggioni, A. Oliveira, M. Oshagh, E. Pallé, G. Pariani, E. Poretti, J. L. Rasilla, J. Rebordão, E. M. Redaelli, S. Santana Tschudi, P. Santin, P. Santos, D. Ségransan, T. M. Schmidt, A. Segovia, D. Sosnowska, A. Sozzetti, S. G. Sousa, P. Spanò, A. Suárez Mascareño, H. Tabernero, F. Tenegi, S. Udry, and A. Zanutta. ESPRESSO at VLT. On-sky performance and first results. *Astronomy & Astrophysics*, 645:A96, Jan. 2021. doi: 10.1051/0004-6361/202038306.

R. C. Peterson and R. L. Kurucz. New Fe I Level Energies and Line Identifications from Stellar Spectra. III. Initial Results from UV, Optical, and Infrared Spectra. *Astrophysical Journals*, 260(2): 28, June 2022. doi: 10.3847/1538-4365/ac596b.

P. Petit, T. Louge, S. Théado, F. Paletou, N. Manset, J. Morin, S. C. Marsden, and S. V. Jeffers. PolarBase: A Database of High-Resolution Spectropolarimetric Stellar Observations. , 126(939): 469, May 2014. doi: 10.1086/676976.

A. Pinhas, B. V. Rackham, N. Madhusudhan, and D. Apai. Retrieval of planetary and stellar properties in transmission spectroscopy with AURA. , 480(4):5314–5331, Nov. 2018. doi: 10.1093/mnras/sty2209.

O. L. Polyansky, A. A. Kyuberis, N. F. Zobov, J. Tennyson, S. N. Yurchenko, and L. Lodi. ExoMol

molecular line lists XXX: a complete high-accuracy line list for water. *Monthly Notices of the RAS*, 480(2):2597–2608, Oct. 2018. doi: 10.1093/mnras/sty1877.

L. Puspitarini and R. Lallement. Distance to northern high-latitude HI shells. , 545:A21, Sept. 2012. doi: 10.1051/0004-6361/201219284.

D. Queloz, A. Eggenberger, M. Mayor, C. Perrier, J. L. Beuzit, D. Naef, J. P. Sivan, and S. Udry. Detection of a spectroscopic transit by the planet orbiting the star HD209458. *Astronomy & Astrophysics*, 359:L13–L17, July 2000. doi: 10.48550/arXiv.astro-ph/0006213.

D. Queloz, G. W. Henry, J. P. Sivan, S. L. Baliunas, J. L. Beuzit, R. A. Donahue, M. Mayor, D. Naef, C. Perrier, and S. Udry. No planet for HD 166435. , 379:279–287, Nov. 2001. doi: 10.1051/0004-6361:20011308.

A. Reiners, N. Mrotzek, U. Lemke, J. Hinrichs, and K. Reinsch. The IAG solar flux atlas: Accurate wavelengths and absolute convective blueshift in standard solar spectra. , 587:A65, Mar. 2016. doi: 10.1051/0004-6361/201527530.

R. A. Rossiter. On the detection of an effect of rotation during eclipse in the velocity of the brigher component of beta Lyrae, and on the constancy of velocity of this system. *Astrophysical Journal*, 60:15–21, July 1924. doi: 10.1086/142825.

L. S. Rothman. History of the HITRAN Database. *Nature Reviews Physics*, 3(5):302–304, Jan. 2021. doi: 10.1038/s42254-021-00309-2.

L. Sbordone, P. Bonifacio, F. Castelli, and R. L. Kurucz. ATLAS and SYNTHE under Linux. *Memorie della Societa Astronomica Italiana Supplementi*, 5:93, Jan. 2004. doi: 10.48550/arXiv.astro-ph/0406268.

L. Schaefer, R. D. Wordsworth, Z. Berta-Thompson, and D. Sasselov. Predictions of the Atmospheric Composition of GJ 1132b. *Astrophysical Journal*, 829(2):63, Oct. 2016. doi: 10.3847/0004-637X/829/2/63.

D. K. Sing, J. M. Désert, J. J. Fortney, A. Lecavelier Des Etangs, G. E. Ballester, J. Cepa, D. Ehrenreich, M. López-Morales, F. Pont, M. Shabram, and A. Vidal-Madjar. Gran Telescopio Canarias OSIRIS transiting exoplanet atmospheric survey: detection of potassium in XO-2b

from narrowband spectrophotometry. *Astronomy & Astrophysics*, 527:A73, Mar. 2011. doi: 10.1051/0004-6361/201015579.

A. Smette, H. Sana, S. Noll, H. Horst, W. Kausch, S. Kimeswenger, M. Barden, C. Szyszka, A. M. Jones, A. Gallenne, J. Vinther, P. Ballester, and J. Taylor. Molecfit: A general tool for telluric absorption correction. I. Method and application to ESO instruments. *Astronomy & Astrophysics*, 576:A77, Apr. 2015. doi: 10.1051/0004-6361/201423932.

M. Sneep and W. Ubachs. Direct measurement of the Rayleigh scattering cross section in various gases. *Journal of Quantitiative Spectroscopy and Radiative Transfer*, 92(3):293–310, May 2005. doi: 10.1016/j.jqsrt.2004.07.025.

E. N. Sokov, I. A. Sokova, V. V. Dyachenko, D. A. Rastegaev, A. Burdanov, S. A. Rusov, P. Benni, S. Shadick, V.-P. Hentunen, M. Salisbury, N. Esseiva, J. Garlitz, M. Bretton, Y. Ogmen, Y. Karavaev, A. Ayiomamitis, O. Mazurenko, D. Alonso, and S. F. Velichko. Transit timing analysis of the exoplanet TrES-5 b. Possible existence of the exoplanet TrES-5 c. *Monthly Notices of the RAS*, 480(1):291–301, Oct. 2018. doi: 10.1093/mnras/sty1615.

K. B. Stevenson. Quantifying and Predicting the Presence of Clouds in Exoplanet Atmospheres. *Astrophysical Journal, Letters*, 817(2):L16, Feb. 2016. doi: 10.3847/2041-8205/817/2/L16.

S. Sulis, M. Lendl, H. M. Cegla, L. F. Rodríguez Díaz, L. Bigot, V. Van Grootel, A. Bekkelien, A. C. Cameron, P. F. L. Maxted, A. E. Simon, C. Lovis, G. Scandariato, G. Bruno, D. Nardiello, A. Bonfanti, M. Fridlund, C. M. Persson, S. Salmon, S. G. Sousa, T. G. Wilson, A. Krenn, S. Hoyer, A. Santerne, D. Ehrenreich, Y. Alibert, R. Alonso, G. Anglada, T. Bárczy, D. Barrado y Navascues, S. C. C. Barros, W. Baumjohann, M. Beck, T. Beck, W. Benz, N. Billot, X. Bonfils, L. Borsato, A. Brandeker, C. Broeg, J. Cabrera, S. Charnoz, C. Corral van Damme, S. Csizmadia, M. B. Davies, M. Deleuil, A. Deline, L. Delrez, O. D. S. Demangeon, B. O. Demory, A. Erikson, A. Fortier, L. Fossati, D. Gandolfi, M. Gillon, M. Güdel, K. Heng, K. G. Isaak, L. L. Kiss, J. Laskar, A. Lecavelier des Etangs, D. Magrin, M. Munari, V. Nascimbeni, G. Olofsson, R. Ottensamer, I. Pagano, E. Pallé, G. Peter, G. Piotto, D. Pollacco, D. Queloz, R. Ragazzoni, N. Rando, H. Rauer, I. Ribas, M. Rieder, N. C. Santos, D. Ségransan, A. M. S. Smith, M. Steinberger, M. Steller, G. M. Szabó, N. Thomas, S. Udry, N. A. Walton, and D. Wolter. Connecting photometric and

spectroscopic granulation signals with CHEOPS and ESPRESSO. *Astronomy & Astrophysics*, 670: A24, Feb. 2023. doi: 10.1051/0004-6361/202244223.

H. M. Tabernero, M. R. Zapatero Osorio, R. Allart, F. Borsa, N. Casasayas-Barris, O. Demangeon, D. Ehrenreich, J. Lillo-Box, C. Lovis, E. Pallé, S. G. Sousa, R. Rebolo, N. C. Santos, F. Pepe, S. Cristiani, V. Adibekyan, C. Allende Prieto, Y. Alibert, S. C. C. Barros, F. Bouchy, V. Bourrier, V. D'Odorico, X. Dumusque, J. P. Faria, P. Figueira, R. Génova Santos, J. I. González Hernández, S. Hojjatpanah, G. Lo Curto, B. Lavie, C. J. A. P. Martins, J. H. C. Martins, A. Mehner, G. Micela, P. Molaro, N. J. Nunes, E. Poretti, J. V. Seidel, A. Sozzetti, A. Suárez Mascareño, S. Udry, M. Aliverti, M. Affolter, D. Alves, M. Amate, G. Avila, T. Bandy, W. Benz, A. Bianco, C. Broeg, A. Cabral, P. Conconi, J. Coelho, C. Cumani, S. Deiries, H. Dekker, B. Delabre, A. Fragoso, M. Genoni, L. Genolet, I. Hughes, J. Knudstrup, F. Kerber, M. Landoni, J. L. Lizon, C. Maire, A. Manescau, P. Di Marcantonio, D. Mégevand, M. Monteiro, M. Monteiro, M. Moschetti, E. Mueller, A. Modigliani, L. Oggioni, A. Oliveira, G. Pariani, L. Pasquini, J. L. Rasilla, E. Redaelli, M. Riva, S. Santana-Tschudi, P. Santin, P. Santos, A. Segovia, D. Sosnowska, P. Spanò, F. Tenegi, O. Iwert, A. Zanutta, and F. Zerbi. ESPRESSO high-resolution transmission spectroscopy of WASP-76 b. *Astronomy & Astrophysics*, 646:A158, Feb. 2021. doi: 10.1051/0004-6361/202039511.

G. Tinetti, A. Vidal-Madjar, M.-C. Liang, J.-P. Beaulieu, Y. Yung, S. Carey, R. J. Barber, J. Tennyson, I. Ribas, N. Allard, G. E. Ballester, D. K. Sing, and F. Selsis. Water vapour in the atmosphere of a transiting extrasolar planet. *Nature*, 448(7150):169–171, July 2007. doi: 10.1038/nature06002.

G. Tinetti, P. Eccleston, T. Lueftinger, J.-C. Salvignol, S. Fahmy, and C. Alves de Oliveira. Ariel: Enabling planetary science across light-years. In *European Planetary Science Congress*, pages EPSC2022–1114, Sept. 2022. doi: 10.5194/epsc2022-1114.

S.-M. Tsai, E. K. H. Lee, D. Powell, P. Gao, X. Zhang, J. Moses, E. Hébrard, O. Venot, V. Parmentier, S. Jordan, R. Hu, M. K. Alam, L. Alderson, N. M. Batalha, J. L. Bean, B. Benneke, C. J. Bierson, R. P. Brady, L. Carone, A. L. Carter, K. L. Chubb, J. Inglis, J. Leconte, M. Line, M. López-Morales, Y. Miguel, K. Molaverdikhani, Z. Rustamkulov, D. K. Sing, K. B. Stevenson, H. R. Wakeford, J. Yang, K. Aggarwal, R. Baeyens, S. Barat, M. de Val-Borro, T. Daylan, J. J. Fortney, K. France, J. M. Goyal, D. Grant, J. Kirk, L. Kreidberg, A. Louca, S. E. Moran, S. Mukherjee,

E. Nasedkin, K. Ohno, B. V. Rackham, S. Redfield, J. Taylor, P. Tremblin, C. Visscher, N. L. Wallack, L. Welbanks, A. Youngblood, E.-M. Ahrer, N. E. Batalha, P. Behr, Z. K. Berta-Thompson, J. Blecic, S. L. Casewell, I. J. M. Crossfield, N. Crouzet, P. E. Cubillos, L. Decin, J.-M. Désert, A. D. Feinstein, N. P. Gibson, J. Harrington, K. Heng, T. Henning, E. M. R. Kempton, J. Krick, P.-O. Lagage, M. Lendl, J. D. Lothringer, M. Mansfield, N. J. Mayne, T. Mikal-Evans, E. Palle, E. Schlawin, O. Shorttle, P. J. Wheatley, and S. N. Yurchenko. Photochemically produced SO_2 in the atmosphere of WASP-39b. *Nature*, 617(7961):483–487, May 2023. doi: 10.1038/s41586-023-05902-2.

A. Tsiaras, I. P. Waldmann, T. Zingales, M. Rocchetto, G. Morello, M. Damiano, K. Karpouzas, G. Tinetti, L. K. McKemmish, J. Tennyson, and S. N. Yurchenko. A Population Study of Gaseous Exoplanets. *Astronomical Journal*, 155(4):156, Apr. 2018. doi: 10.3847/1538-3881/aaaf75.

A. Tsiaras, I. P. Waldmann, G. Tinetti, J. Tennyson, and S. N. Yurchenko. Water vapour in the atmosphere of the habitable-zone eight-Earth-mass planet K2-18 b. *Nature Astronomy*, 3:1086–1091, Sept. 2019. doi: 10.1038/s41550-019-0878-9.

M. Tuomi, G. Anglada-Escudé, E. Gerlach, H. R. A. Jones, A. Reiners, E. J. Rivera, S. S. Vogt, and R. P. Butler. Habitable-zone super-Earth candidate in a six-planet system around the K2.5V star HD 40307. *Astronomy & Astrophysics*, 549:A48, Jan. 2013. doi: 10.1051/0004-6361/201220268.

M. Turbet, E. Bolmont, V. Bourrier, B.-O. Demory, J. Leconte, J. Owen, and E. T. Wolf. A Review of Possible Planetary Atmospheres in the TRAPPIST-1 System. *Space Science Reviews*, 216(5):100, July 2020. doi: 10.1007/s11214-020-00719-1.

W. D. Vacca, M. C. Cushing, and J. T. Rayner. A Method of Correcting Near-Infrared Spectra for Telluric Absorption. *Publications of the ASP*, 115(805):389–409, Mar. 2003. doi: 10.1086/346193.

G. Valyavin, G. Beskin, A. Valeev, G. Galazutdinov, S. Fabrika, V. Aitov, O. Yakovlev, A. Ivanova, R. Baluev, V. Vlasyuk, I. Han, S. Karpov, V. Sasyuk, A. Perkov, S. Bondar, F. Musaev, E. Emelianov, T. Fatkhullin, S. Drabek, V. Shergin, B.-C. Lee, G. Mitiani, T. Burlakova, M. Yushkin, E. Sendzikas, D. Gadelshin, L. Chmyreva, A. Beskakotov, V. Dyachenko, D. Rastegaev, A. Mitrofanova, I. Yakunin, K. Antonyuk, V. Plokhotnichenko, A. Gutaev, N. Lyapsina, V. Chernenkov, A. Biryukov, E. Ivanov, A. Belinsky, E. Sokov, A. Tavrov,

O. Korablev, M.-G. Park, V. Stolyarov, V. Bychkov, S. Gorda, A. Popov, and A. Sobolev. EXPLANATION: Exoplanet and Transient Events Investigation Project. *Astrophysical Bulletin*, 77(4):495–508, Dec. 2022. doi: 10.1134/S1990341322040186.

G. G. Valyavin, V. D. Bychkov, M. V. Yushkin, G. A. Galazutdinov, S. V. Drabek, V. S. Shergin, A. N. Sarkisyan, E. A. Semenko, T. E. Burlakova, V. M. Kravchenko, D. O. Kudryavtsev, A. M. Pritychenko, P. G. Kryukov, S. L. Semjonov, F. A. Musaev, and S. N. Fabrika. High-resolution fiber-fed echelle spectrograph for the 6-m telescope. I. Optical scheme, arrangement, and control system. *Astrophysical Bulletin*, 69(2):224–239, Apr. 2014. doi: 10.1134/S1990341314020102.

G. G. Valyavin, V. D. Bychkov, M. V. Yushkin, G. A. Galazutdinov, S. V. Drabek, V. S. Shergin, A. N. Sarkisyan, E. A. Semenko, A. V. Perkov, D. A. Sazonenko, D. E. Kukushkin, A. V. Bakholdin, T. E. Burlakova, V. M. Kravchenko, D. O. Kudryavtsev, A. M. Pritychenko, P. G. Kryukov, S. L. Semjonov, F. A. Musaev, and S. N. Fabrika. High-Resolution Fiber-Fed Echelle Spectrograph for the SAO 6-m Telescope. In Y. Y. Balega, I. I. Romanyuk, and D. O. Kudryavtsev, editors, *Physics and Evolution of Magnetic and Related Stars*, volume 494 of *Astronomical Society of the Pacific Conference Series*, page 305, Apr. 2015.

G. G. Valyavin, F. A. Musaev, A. V. Perkov, V. N. Aitov, V. D. Bychkov, S. V. Drabek, V. S. Shergin, D. A. Sazonenko, D. E. Kukushkin, G. A. Galazutdinov, E. V. Emelyanov, G. V. Yakopov, T. E. Burlakova, J. L. Bertaux, A. V. Tavrov, O. I. Korablev, M. V. Yushkin, A. F. Valeev, D. R. Gadelshin, K. M. Kim, I. Han, and B. C. Lee. High-Resolution Fiber-Fed Spectrograph for the 6-m Telescope of the Special Astrophysical Observatory of the Russian Academy of Sciences: Assessment of Efficiency. *Astrophysical Bulletin*, 75(2):191–197, July 2020. doi: 10.1134/S1990341320020157.

A. Vidal-Madjar, A. Lecavelier des Etangs, J. M. Désert, G. E. Ballester, R. Ferlet, G. Hébrard, and M. Mayor. An extended upper atmosphere around the extrasolar planet HD209458b. *Nature*, 422 (6928):143–146, Mar. 2003. doi: 10.1038/nature01448.

S.-i. Wang, R. H. Hildebrand, L. M. Hobbs, S. J. Heimsath, G. Kelderhouse, R. F. Loewenstein, S. Lucero, C. M. Rockosi, D. Sandford, J. L. Sundwall, J. A. Thorburn, and D. G. York. ARCES: an echelle spectrograph for the Astrophysical Research Consortium (ARC) 3.5m telescope. In M. Iye and A. F. M. Moorwood, editors, *Instrument Design and Performance for Optical/Infrared*

Ground-based Telescopes, volume 4841 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 1145–1156, Mar. 2003. doi: 10.1117/12.461447.

D. E. Welty, L. M. Hobbs, and V. P. Kulkarni. A High-Resolution Survey of Interstellar NA i D 1 Lines. , 436:152, Nov. 1994. doi: 10.1086/174889.

R. G. West, C. Hellier, J. M. Almenara, D. R. Anderson, S. C. C. Barros, F. Bouchy, D. J. A. Brown, A. Collier Cameron, M. Deleuil, L. Delrez, A. P. Doyle, F. Faedi, A. Fumel, M. Gillon, Y. Gómez Maqueo Chew, G. Hébrard, E. Jehin, M. Lendl, P. F. L. Maxted, F. Pepe, D. Pollacco, D. Queloz, D. Ségransan, B. Smalley, A. M. S. Smith, J. Southworth, A. H. M. J. Triaud, and S. Udry. Three irradiated and bloated hot Jupiters:. WASP-76b, WASP-82b, and WASP-90b. , 585:A126, Jan. 2016. doi: 10.1051/0004-6361/201527276.

M. Zechmeister, A. Reiners, P. J. Amado, M. Azzaro, F. F. Bauer, V. J. S. Béjar, J. A. Caballero, E. W. Guenther, H. J. Hagen, S. V. Jeffers, A. Kaminski, M. Kürster, R. Launhardt, D. Montes, J. C. Morales, A. Quirrenbach, S. Reffert, I. Ribas, W. Seifert, L. Tal-Or, and V. Wolthoff. Spectrum radial velocity analyser (SERVAL). High-precision radial velocities and two alternative spectral indicators. *Astronomy & Astrophysics*, 609:A12, Jan. 2018. doi: 10.1051/0004-6361/201731483.

A&A 652, A22 (2021)
https://doi.org/10.1051/0004-6361/202140514

Astronomy & Astrophysics

Toward a 3D kinetic tomography of Taurus clouds

I. Linking neutral potassium and dust⋆

A. Ivanova[1,2], R. Lallement[3], J. L. Vergely[4], and C. Hottier[3]

[1] LATMOS, Université Versailles-Saint-Quentin, 11 Bd D'Alembert, Guyancourt, France
e-mail: rosine.lallement@obspm.fr
[2] Space Research Institute (IKI), Russian Academy of Science, Moscow 117997, Russia
[3] GEPI, Observatoire de Paris, PSL University, CNRS, 5 Place Jules Janssen, 92190 Meudon, France
[4] ACRI-ST, 260 route du Pin Montard, 06904, Sophia Antipolis, France

Received 8 February 2021 / Accepted 22 April 2021

ABSTRACT

Context. *Gaia* parallaxes and photometric measurements open a three-dimensional (3D) era for the Milky Way, including its interstellar (IS) matter. Three-dimensional Galactic dust distributions are constructed in various ways, based on *Gaia* data and photometric or spectroscopic surveys.
Aims. The assignment of radial motions to IS dust structures seen in 3D, or 3D kinetic tomography, would be a valuable tool allowing one to connect the structures to emission lines of the associated gas, which are now measured at increasingly higher spectral and angular resolutions, and rich in information on physical and chemical processes. To this end, one of the potential techniques is to establish a link between dust clouds and Doppler velocities of absorption lines imprinted in stellar spectra by the gas associated with the dust. This requires a relatively close correlation between the absorber column and the dust opacity. We have investigated the link between the strength of interstellar K I absorption and the opacity of the dust in front of stars in the Taurus area, and we have tested the feasibility of assigning velocities to 3D dust clouds on the basis of K I absorption data.
Methods. We have obtained high spectral resolution and high signal-to-noise spectra of 58 early-type stars in the direction of the Taurus, Perseus, and California molecular clouds. We have developed a new, dual interstellar and telluric profile-fitting technique to extract the interstellar K I $\lambda\lambda$ 7665, 7699 Å absorption lines from stellar spectra and applied it to the new data and to archived spectra of 58 additional targets. In parallel, we have updated 3D dust maps reconstructed through the inversion of individual stellar light extinctions. To do so, we supplemented the catalog of extinction estimates based on *Gaia* and 2MASS photometry with recently published extinction catalogs based on stellar spectroscopic surveys. We used the 3D map and the set of velocity components seen in absorption to assign radial velocities to the dust clouds distributed along their paths in the most consistent way.
Results. We illustrate our profile-fitting technique and present the K I velocity structure of the dense ISM along the paths to all targets. As a validation test of the dust map, we show comparisons between distances to several reconstructed clouds with recent distance assignments based on different techniques. Target star extinctions estimated by integration in the 3D map are compared with their K I 7699 Å absorptions and the degree of correlation is found comparable to the one between the same K I line and the total hydrogen column for stars distributed over the sky that are part of a published high resolution survey. We show images of the updated dust distribution in a series of vertical planes in the Galactic longitude interval 150–182.5° and our estimated assignments of radial velocities to the opaque regions. Most clearly defined K I absorptions may be assigned to a dense dust cloud between the Sun and the target star. It appeared relatively straightforward to find a velocity pattern consistent will all absorptions and ensuring coherence between adjacent lines of sight, at the exception of a few weak lines. We compare our results with recent determinations of the velocities of several clouds and find good agreement. These results demonstrate that the extinction-K I relationship is tight enough to allow one to link the radial velocity of the K I lines to the dust clouds seen in 3D and that their combination may be a valuable tool in building a 3D kinetic structure of the dense ISM. We discuss limitations and perspectives for this technique.

Key words. ISM: clouds – dust, extinction – ISM: kinematics and dynamics – ISM: structure

1. Introduction

The history and present evolutionary state of the Milky Way are currently being deciphered at all spatial scales, boosted by the remarkable *Gaia* measurements (Gaia Collaboration 2021), and complemented by massive photometric and spectroscopic stellar surveys (for a recent review of results on the Milky Way history see Helmi 2020). On the large scale, and in addition to astrometric measurements, Milky Way evolutionary models also benefit from new, accurate photometric data and subsequent realistic estimates of stellar parameters and extinction of starlight by dust. New, massive measurements of extinctions and faint sources' photometric distances have been produced (e.g., Sanders & Das 2018; Anders et al. 2019; Queiroz et al. 2020). This has favored the development of catalogs of dust cloud distances and three-dimensional (3D) dust distributions (Green et al. 2019; Chen et al. 2019; Lallement et al. 2019; Rezaei Kh. et al. 2020; Zucker et al. 2020; Leike et al. 2020; Guo et al. 2021; Hottier et al. 2020). On the smaller scale of star-forming regions, the combination of accurate distances and proper motions with realistic age determinations marks a new era for their detailed study (see,

⋆ Based on observations obtained at the Télescope *Bernard Lyot* (TBL) at Observatoire du Pic du Midi, CNRS/INSU and Université de Toulouse, France.

e.g., recent studies of Großschedl et al. 2021; Roccatagliata et al. 2020; Kounkel et al. 2018; Galli et al. 2019).

The assignment of radial velocity to each structure represented in 3D would allow one to go one step further in a number of on-going analyses, and, again, in a wide range of spatial scales. Sophisticated combined N-body and hydrodynamical simulations of the stellar-gaseous Galactic disk are being developed (see, e.g., Khoperskov et al. 2020, and references therein) and used to reproduce the 6D phase-space (stellar positions and velocities) distributions provided by *Gaia*. Exciting debates are currently ongoing about the respective roles of external perturbations due to dwarf galaxy passages or internally driven resonances associated with bar and spiral arms (Antoja et al. 2018; Khoperskov et al. 2021). In parallel, chemo-dynamical models are being developed and aim to reproduce the observed enrichment sequences and dichotomies in the distributions of elemental compositions (e.g., Haywood et al. 2019; Khoperskov et al. 2021; Sharma et al. 2020; Katz et al. 2021). Up to now, efforts have been concentrated on comparisons with the newly observed spatial distributions and chemical and dynamical characteristics of the stellar populations. However, stellar history and interstellar (IS) matter evolution are tightly coupled and additional constraints on the models may be brought by IS matter distribution and dynamics. On the smaller scale of star-forming regions, dense IS matter and stars are dynamically coupled, as was recently quantified by Galli et al. (2019) who found a difference of less than 0.5 km s^{-1} between the radial velocities of young stellar objects (YSOs) and the associated CO in Main Taurus regions. This shows that the velocity distribution of the interstellar gas and dust may help to reconstruct the star formation history. Extremely detailed spectro-imaging surveys have been performed or are in progress whose aims are state-of-the-art studies of physical and chemical processes at work in the various phases of the interstellar medium (ISM) in such regions (see, e.g., Pety et al. 2017). The assignment of all emission lines, which contain very rich information on the processes, to their source regions, can help to constrain and refine the models. Finally, and for all interstellar medium studies, the assignment of velocities to interstellar clouds located in 3D would allow one to connect the structures to their multiwavelength emission through velocity cross-matching. Extended 3D kinetic maps would display multiple structures sharing the same radial velocity and located at different distances, if any, and be useful to clarify the models.

Here we use the term 3D kinetic tomography for the assignment of velocities to structures occupying a volume in 3D space, that is, it does not include the association of both absorption and emission lines along the same individual line of sight. A structure represented in 3D may be a dust cloud reconstructed by inversion, a voxel in the case of discretized 3D maps of dust or gas, or a cloud having some extent in 2D images and localized in distance. Several recent works have been devoted to 3D kinetic tomography. Tchernyshyov & Peek (2017) developed a method using HI and CO spectral cubes (i.e., position-position-velocity matrices) on the one hand and a 3D reddening map (i.e., a position-position-position matrix) on the other hand. The authors adjusted the radial velocity in each voxel of a discretized volume around the Galactic plane until consistency between the three data sets was achieved, assuming conversion factors between CO, HI temperatures, and dust opacity. Tchernyshyov et al. (2018) used the series of measurements of a near-infrared (NIR) diffuse interstellar band (DIB) in SDSS/APOGEE stellar spectra and photometric distances of the target stars to derive a planar map of the radial velocity. The authors used variations in measured DIB absorption profiles for stars at slightly different distances and directions to extract the contribution of the local interstellar matter. Zucker et al. (2018) used the distance-reddening posterior distributions from the Bayesian technique of Green et al. (2018) and ^{12}CO spectra to assign radial velocities to the main structures in Perseus. To do so, the authors associated each opacity bin in Perseus with a linear combination of velocity slices.

Three-dimensional kinetic tomography is not straightforward, and one can reasonably foresee that using different tracers and different techniques will help to achieve accurate results in large volumes. The difficulties are of various types. The omnidirectional spatial resolution of 3D dust maps computed for large volumes has not yet fully reached a level that allows one to identify the parsec or subparsec counterparts of extremely small details in the ultra high angular and spectral resolution radio data. The Pan-STARRS/2MASS map of Green et al. (2019) has a very high angular resolution, similar to the one of emission data; however, the resolution along the radial direction is much poorer than in the plane of the sky. The hierarchical technique presented in Lallement et al. (2019) was still limited to a 25 pc wide 3D kernel, and the updated map presented in the present article and based on the same technique is limited to 10 pc resolution. The method used by Leike et al. (2020) allowed the authors to reach 1 pc spatial resolution, however for a limited volume due to the computational cost. It is hoped that this difficulty should gradually decrease thanks to future releases of *Gaia* catalogs and the extension of ground surveys. If emission spectra are the sources of the Doppler velocities, a second difficulty is the existence of clouds at different distances and sharing the same radial velocity, a degeneracy that sharply increases at low latitudes. The techniques of Tchernyshyov & Peek (2017) and Zucker et al. (2018) may, in principle, disentangle the contributions; however, this requires prior knowledge of the ratio between the dust opacity and the species used in emission. For example, Zucker et al. (2018) assumed that the dust located closer than 200 pc from the Sun in front of IC348 in Perseus does not contribute to the CO emission, despite the non-negligible (≥35%) contribution to the extinction of a foreground cloud at 170 pc (see Fig. 12 for the visualization of this foreground cloud). One way to help break the degeneracy is the additional use of distance-limited IS absorption data. In this case, target stars must be distributed within and beyond the volume containing the main IS cloud complexes. Finally, if extinction maps are part of the tomography, an additional difficulty is due to the fact that dust traces both dense molecular and atomic gas phases, and, as a consequence, emission or absorption data used for the Doppler shift measurements must trace these two phases. This is illustrated in Fig. 1 that displays ^{12}CO columns in the Taurus area and superimposed HI 21 cm intensity contours, both for the LSR velocity range of the local IS matter (here $-10 \leq V \leq +10$ km s^{-1}). The marked and well-known differences between the respective amounts of molecular and atomic gas clearly show that using solely one of these two species for a cross-identification with dust structures everywhere in the image is inadequate, unless one restricts the study to predominantly molecular regions, as in Zucker et al. (2018), in which case CO is a convenient unique tracer.

The objective of the present work is to test the use of neutral potassium (K I) absorption in 3D kinetic ISM tomography. Absorption by interstellar neutral potassium is expected to allow both dense atomic and molecular phases to be traced, according to the results of the extensive study devoted to this species by (Welty & Hobbs 2001; herafter WH01). Using a large number of Galactic K I measurements, extracted from very high

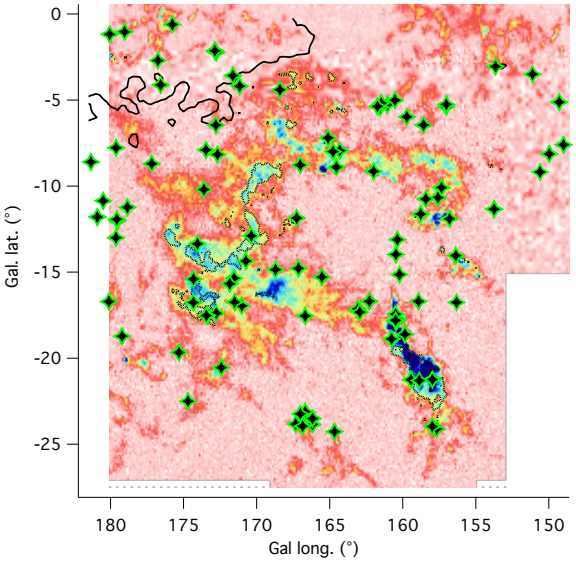


Fig. 1. Target stars from our program and from archives superimposed on a ^{12}CO map of the Taurus-Perseus-California area (Dame et al. 2001). Superimposed are HI iso-contours for T(HI) = 1, 2, and 3.5 K (thin, dot-dashed, and thick line lines, respectively), based on HI4PI Collaboration (2016) data. Both CO and HI columns were restricted to LSR velocities between −10 and +10 km s^{-1}.

resolution spectra, the authors studied the link between K I and various species, and found a quadratic dependence of K I with hydrogen (H_{tot}= HI +2 H_2). Because H_{tot} is known to correlate with the extinction, we may expect a similar relationship between neutral potassium and extinction. We note that according to the WH01 study, the K I - H_{tot} correlation is better than for Na I with H_{tot}, which favors the use of the former species. Additional advantages of using potassium by comparison with neutral sodium is its higher mass, hence its narrower and deeper absorption lines, which are more appropriate for studies of the velocity structure. In comparison with DIBs, the very small width of the K I lines allows for a much better disentangling of multiple velocity components and a significantly higher accuracy of the absolute values for each one. The caveat is that, with K I being detected in the optical, highly reddened stars are not part of potential targets and one must restrict observations to the periphery of the dense clouds. Using NIR DIBs, on the contrary, allows one to use targets located behind high opacity clouds.

Our test consisted in the following: gathering high spectral resolution, high quality stellar spectra of target stars located in front of, within, and behind the Taurus, Perseus, and California clouds, extracting from the spectra interstellar K I absorption profiles, and measuring the radial velocities of the absorption lines; updating 3D extinction maps to achieve better spatial resolution; comparing the K I absorption strengths with the integrated extinction along the path to the targets; attempting a synthetic assignment of the measured K I radial velocities to the clouds contributing to this extinction; and evaluating the extent of information on the velocity pattern that can be deduced from K I data alone.

Neutral potassium is detectable by its $\lambda\lambda$7665, 7699 Å resonance doublet. Unfortunately, the first, stronger transition is located in the central part of the *A*-band of telluric molecular oxygen. This is why the second, weaker transition should be used alone or the data must be corrected for the telluric absorption. Here we present a novel method to extract a maximum of information from the two transitions that avoids using telluric-corrected spectra obtained by division by a spectrum of a standard star or by a model. Such corrected spectra are often characterized by strong residuals in the regions of the deepest telluric lines that make the determination of the continuum and the subsequent profile-fitting difficult. To avoid this, we used two distinct, consecutive adjustments. First, we used the atmospheric transmittance models from the TAPAS online facility[1] (Bertaux et al. 2014), and selected those that are suitable for the respective observing sites. We used these models and their adjustments to the data to (i) refine the wavelength calibration, (ii) determine the instrumental function (or line spread function, hereafter LSF) as a function of wavelength along the echelle order, and (iii) derive the atmospheric transmittance that prevailed at each observation, prior to entry into the instrument. In a second step, this adjusted transmittance and the spectral resolution measurements were used in a global forward model in combination with a multi-cloud model of the two IS K I 7665 and 7699 Å transitions.

The spectral data were recorded with the Pic du Midi TBL-Narval spectrograph during two dedicated programs and are complemented by archival data from the Polarbase facility, recorded with Narval or the ESPaDOnS spectro-polarimeter at CFHT (Donati et al. 1997; Petit et al. 2014) and by several other pieces of archival data. The extracted K I lines and their radial velocities were compared with the locations of the dust clouds in an updated 3D dust map. This new 3D distribution was derived from the inversion of a catalog of individual extinction measurements that comes out as an auxiliary result of the Sanders & Das (2018) extensive analysis devoted to stellar populations' ages, a study based on six massive spectroscopic surveys. This new Bayesian inversion used the 3D distribution recently derived from *Gaia* and 2MASS as a prior (Lallement et al. 2019).

The article is organized in the following way. Spectral data are presented in Sect. 2. The telluric modeling is then described in Sect. 3, and the K I extraction method is detailed in Sect. 4. In Sect. 5 we describe the construction of the updated 3D dust map and test the link between the integrated extinction along the path to the star and the K I absorption intensity. Section 6 shows the velocity assignments to the dense structures that were found compatible with the whole data set and appeared as the most coherent ones. In the last section we discuss the results and the perspectives of such an approach, as well as its future association with the information from emission data.

2. Spectral data

2.1. TBL Narval data

High resolution, high signal spectra were recorded in service mode with the Narval Echelle spectro-polarimeter at the *Bernard Lyot* Telescope (TBL) facility during two dedicated programs (L152N07 and L162N04, PI: Lallement). The Narval spectra cover a wide wavelength range from 370 to 1058 nm. Target stars were selected in direction and distance to sample the volumes of Taurus, Perseus, and California clouds, and they were chosen preferentially among the earliest types. The observations were done in the "star only" (pure spectroscopic) mode that provides a resolving power on the order of 75 000. Most of the observations were performed at low airmass.

Spectra of good and excellent quality were obtained for a total of 58 new targets listed in Table A.1. We used the fully reduced and wavelength-calibrated spectra from the Narval pipeline. The pipeline provides the spectrum in separate echelle

[1] http://cds-espri.ipsl.fr/tapas/

orders, and here we use the order that covers the ≃764.5–795.0 nm interval and contains the two K I transitions at 7664.911 and 7698.974 Å. The first transition is also included in an adjacent order; however, we used a unique order and non resampled data for the K I extraction. We discuss our reasoning in Sect. 3. This is also valid for the archive Narval and ESPaDOnS data described below. The whole spectrum was used to check for the spectral type, the stellar line widths, and the interstellar Na I absorption lines (see Sect. 4). Further studies of the full set of atomic and molecular absorption lines and of the diffuse interstellar bands are in progress.

2.2. Archival data

We performed an online search in the POLARBASE database and found ten Narval and 15 ESPaDOnS spectra of target stars that have been observed as part of other programs and are suitable for our study. Their spectral resolution R is comprised between ≃55 000 and ≃85 000, with the lower value for the polarimetric mode. We also searched for useful spectra in the ESO archive facility and found spectra for seven UVES, two X-shooter, and one FEROS targets of interest ($R \simeq 70\,000$, 18 000, and 80 000 respectively). We also used six spectra from the Chicago database recorded with the ARCES spectrograph at $R \simeq 31\,000$ (Wang et al. 2003; Fan et al. 2019). Finally, one spectrum recorded with the OHP 1.93 m Elodie spectrograph $R \simeq 48\,000$ has been added. Elodie does not cover the K I doublet; however, we derived the Doppler velocity of the main absorbing cloud from neutral sodium lines (5890 Å doublet). All targets and corresponding sources are listed in Table A.1. Heliocentric wavelength scales were either directly provided by the archive facility or derived from the observing date in case of spectra being released in the topocentric reference frame. In the special case of ARCES data, we used the observing date and the telluric features. Finally, in addition to these archival data that entered our telluric correction and profile-fitting, we complemented the observation list with Chaffee & White (1982) Doppler velocities of Taurus clouds for eight targets, also based on neutral potassium lines and high resolution observations. One additional result by Welty et al. (1994) based on sodium lines has also been included. We did not include the WH01 result on HD 23630 (ηTau) and its measured K I radial velocity +7.9 km s^{-1} due to the target imprecise distance and to redundancy with HD 23512 for which we infer a similar velocity of +7.3 km s^{-1}. The locations of the set of targets are shown in Fig. 1, superimposed on a ^{12}CO map (Dame et al. 2001), here restricted to LSR velocities between −10 and +10 km s^{-1}. A few targets are external to the map and are not shown. Also added are HI 21 cm iso-contours for the same velocity range (HI4PI Collaboration 2016).

3. Spectral analysis

3.1. The dual interstellar and telluric profile-fitting

It is well known that simultaneously fitting two different transitions of the same absorbing species with different oscillator strengths gives additional constraints on the characteristics of the intervening clouds, and it helps to disentangle their respective contributions in the frequent case of overlapping lines. In the case of K I, the stronger transition is located in a spectral region strongly contaminated by telluric O_2, and taking advantage of this transition requires taking the O_2 absorption profile into account. While, obviously, nothing can be said about the interstellar contribution to the absorption at the bottom of a fully saturated O_2 line, in most cases the interstellar K I absorption or a fraction of it falls in unsaturated parts and contains information. Here we have devised a novel method that allows one to extract all available information contained in the two transitions without using the division by a standard star spectrum or a telluric model prior to the extraction of the interstellar lines. This division creates strong residuals which makes the profile-fitting particularly difficult in the spectral region of the stronger transition and we want to avoid this. The method is made up of two steps. The first step consists in fitting a telluric model to the data and determining the wavelength-dependent LSF. During this step, the spectral regions around the K I lines are excluded. The second step is the dual profile-fitting itself. The previously derived telluric model is multiplied by as many Voigt profiles as necessary to represent the interstellar K I absorption lines and the velocity distribution of intervening clouds, and the total product is convolved by the wavelength-dependent LSF. The number of clouds and the interstellar parameters are optimized to fit this convolved product to the data.

3.2. The preliminary telluric correction and determination of the wavelength-dependent instrumental width

For the telluric correction, we used the TAPAS online facility: TAPAS uses vertical atmospheric profiles of pressure, temperature, humidity, and most atmospheric species, interpolated in the meteorological field of the European Centre for Medium-range Weather Forecasts (ECMWF), the HIgh-Resolution TRANsmission molecular absorption database (HITRAN), and the Line-by-Line Radiative Transfer Model (LBLRTM) software to compute the atmospheric transmission at ultra-high resolution (wavelength step $\simeq 4 \times 10^{-6}$ Å). See Bertaux et al. (2014) for details about TAPAS products and all references. Here we use TAPAS calculations for the observatory from which each spectrum under analysis was obtained. From the star spectrum and the TAPAS model, we extracted a wavelength interval of 80 Å containing the K I doublet, for which we performed the following steps. At first we created a Gaussian LSF for an estimated preliminary resolution, adapted to the central wavelength of the correction interval. Via this LSF, we then convolved the initial TAPAS transmission model for one airmass unit, and through cross-correlation we obtained a preliminary Doppler shift between the observed spectrum (data) and this model (convolved TAPAS). To do so, we excluded the potassium lines regions from the Doppler shift estimate, using a mask with null value from 7663 to 7667 and from 7697 to 7701 for both the data and model. We then used a comprehensive list of un-blended or weakly blended telluric O_2 lines located in the ≃80 Å interval containing the K I doublet to obtain a first estimate of the airmass at observation as well as a more refined estimate of the spectral resolution. To do so, Gaussian profiles were fitted in both the stellar and Earth transmission spectra at the locations of all these potentially useful lines and the results were compared one-by-one. Undetected lines that are too weak and/or lines with profiles that are too noisy in the data were removed from the list. The strongest lines were selected and all corresponding ratios between the data and model equivalent widths were computed. The average ratio provides a preliminary airmass (or relative optical thickness with respect to the initial model), and the line-width increase between the model and data provides an estimate of the mean resolution.

After all these preparatory steps, the main correction procedure starts. It can be broken down into the following three main steps.

1. The first step involves fitting to the data of the telluric model by means of the rope-length method (see below), masking the central parts of the strong telluric lines (where the flux is close or equal to zero). This step assumes a unique Doppler and a unique Gaussian LSF. The free parameters are the airmass factor, the Doppler shift, and the LSF width.
2. The second step consists of the computation of a decontaminated spectrum by division of the data by the adjusted transmission model and replacement of the ratio by an interpolated polynomial at the locations of the strong telluric lines. This provides a quasi-continuum without strong residuals. By quasi-continuum, we mean the telluric-free spectrum, that is the stellar spectrum and its interstellar features.
3. The third step involves a series of fits of the convolved product of the quasi-continuum and a TAPAS transmission to the data, gradually removing constraints on the width of the instrumental profile and on the Doppler shift, and allowing their variability within the fitted spectral interval, along with the airmass factor variability.

To expand on this, for the first and second steps, we started with the preliminary resolution estimated from the linewidths, and the preliminary O_2 column. We then ran an optimization process for a freely varying O_2 column (or equivalently a free airmass) and a freely varying LSF width, the convergence criteria being the minimal length of the spectrum obtained after division of the data by the convolved transmission spectrum. Here the length is simply the sum of distances between consecutive data points. Such a technique relies on the fact that the minimum length corresponds to the smoothest curve and therefore to minimum residuals after the data to model ratio, that is to say a good agreement on all the line shapes. It is important to note that this technique is sufficient for weak to moderate lines, but it is applied here only as an intermediate solution (see, e.g., the different situations in Cox et al. 2017; Cami et al. 2018). The third step is adapted to strong lines and corresponds to a forward modeling. We assume that the data, after division by the already well determined atmospheric transmission model, provide the telluric-free continuum outside the strong lines' centers, and that the interpolated polynomials may represent at first order the telluric-free data elsewhere. We then performed a series of fits to the data of convolved products of this adopted continuum by a transmission model, masking the interstellar K I regions. We let the Doppler shift free to vary along the spectral interval, then, in a final stage, we allowed for a variation of the LSF width. We found that this last stage is very important since the two lines are at very different locations along the echelle order, the first one being very close to the short wavelength limit of the order, and this results in a significant difference reaching a 30% relative variation of the LSF width. We tested the whole procedure for a unique echelle order and for adjacent orders merging. We found that the rebinning and the interpolations required during the order merging as well as the resulting nonmonotonous LSF wavelength variation introduce strong residuals at the locations of the deep lines and make the correction more difficult. For this reason, we have chosen to restrict the analysis to one order. After final convergence, we stored the wavelength-dependent LSF width and, importantly, the adjusted transmission model before its convolution by the LSF.

3.3. Application, examples of corrections

This method was applied to all stars, which are listed in Table A.1 on top of the two horizontal lines. A special TAPAS model was used for each separated observatory. An example of telluric model adjustment is represented in Figs. 2a,b. The fitted TAPAS transmission before entrance in the instrument is displayed on top of the data in Fig. 2a. At the end of the process, if we convolve this transmission by the LSF and divide data by the resulting convolved model, we obtain a partially corrected spectrum, represented in Fig. 2b. There are still narrow spikes at the locations of the strongest lines (see e.g., the first pair of O_2 lines in Fig. 2b). This is unavoidable in the case of deep lines since the flux in the data and model is close to zero. However, because we did not use the result of the division, but only kept the best telluric model as an ingredient to enter the K I profile-fitting, we are not impacted by such spikes, which is the main advantage of the method. There is an additional advantage: Dividing data by a convolved telluric transmission profile is not a fully mathematically correct solution because the convolution by a LSF of the product of two functions is not equivalent to the product of the two functions separately convolved. The latter computation gives an approximate solution, however, only if all features in

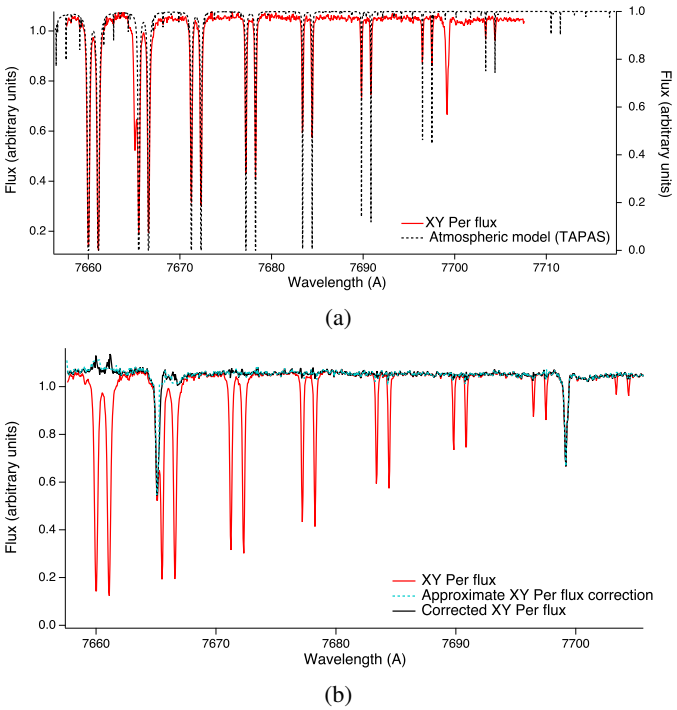


Fig. 2. Illustration of the telluric correction, prior to the dual interstellar-telluric profile-fitting. The TAPAS model was selected for the Pic du Midi observatory. (*a*) Atmospheric profile (black color, right axis) after adjustment to the observation, superimposed on the initial stellar spectrum (here the star XY Per, red color, left axis). We note that this is the atmospheric profile before convolution by the instrumental profile, which is used at the next step of profile-fitting. (*b*) Corrected stellar spectrum (black line) obtained by division of the raw data by the above atmospheric profile, after its convolution by the instrumental profile. It is superimposed on the initial stellar spectrum (red line). There are some spiky residuals in the strong absorption areas which are due to the division of weak quantities for both the data and models. Also shown is the quasi-continuum obtained from the corrected spectrum after some interpolation at these regions (turquoise dashed line). Here, the residuals are weak and there is only a very small difference between the two curves. We note that in this figure we want to illustrate the accuracy of the atmospheric model and the achieved level of adjustment: none of these corrected spectra will enter the dual telluric-interstellar profile-fitting, instead only the adjusted preinstrumental profile shown at the top will be used.

the initial spectrum are wider than the LSF. This is not the case in the presence of sharp stellar or interstellar features, and this may have an impact on the accuracy of the results. We note that this type of method can be used in exo-planetary research for the removal of telluric contamination of the spectra obtained from ground-based facilities.

4. Spectral analysis: interstellar K I doublet extraction

4.1. K I doublet extraction method

As a result of the telluric correction procedure, two files were produced for each target, the first one being the fitted TAPAS model, before convolution by the LSF. The second contains the LSF width as a function of wavelength. We used the second to extract the two separate LSFs appropriate for each K I line (7665 and 7700 Å). After this preparation, we then proceeded to the K I doublet profile-fitting in a rather classical manner (see, e.g., Puspitarini & Lallement 2012), except in two ways: (i) the inclusion of the telluric profile, now combined with KI Voigt profiles before convolution by the LSF; and (ii) the use of two distinct LSFs.

The profile-fitting procedure is made up of the following steps:

1. The first step includes the conversion of wavelengths into Doppler velocities for both the stellar spectrum and the fitted transmission model. The correction is done separately for the two K I transitions. We restricted the analysis to spectral regions wide enough to allow a good definition of the two continua.
2. The second step involves continuum fitting with a first, second, or third order polynomial, excluding the interstellar features. The normalization of the spectra at the two transitions was done through division by the fitted continua.
3. The third step consists in the visual evaluation of the minimum number of clouds with distinct Doppler velocities based on the shapes of the K I lines. Markers were placed at the initial guesses for the cloud velocities, using the 7699 Å line.
4. The fourth steps includes fitting to normalized data of convolved products of Voigt profiles and the telluric model by means of a classical Levenberg-Marquardt convergence algorithm in both K I line areas and cloud number increase, if it appears necessary. Here, we have imposed a maximum Doppler broadening (i.e., combination of thermal broadening and internal motions) of 2.5 km s^{-1} for each cloud. This arbitrary limit was guided by the spectral resolution of the data, the spatial resolution of the present 3D maps of dust, and our limited objectives in the context of this preliminary study (see below).

As is well known, the advantage of fitting a doublet instead of a single line is that more precise cloud disentangling and velocity measurements can be made since there must be agreement between shifts, line depths, and widths for the two transitions and therefore tighter constraints are obtained. An advantage of our technique is obtaining maximal information from the doublet regardless of the configuration of telluric and K I lines. On the other hand, disentangling clumps with velocities closer than ≃2 km s^{-1} is beyond the scope of this study, which is devoted to a first comparison with 3D dust reconstruction. More detailed profile-fitting combining K I and other species is planned. Accurate measurements of the K I columns for the various velocity components is similarly beyond the scope of this profile-fitting.

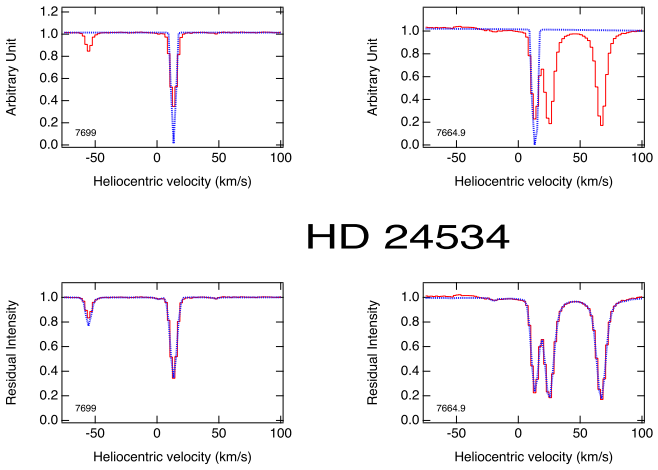


Fig. 3. Example of K I doublet dual interstellar-telluric profile-fitting: star HD 24534 (X Per). The spectra are shown around the 7699 (resp. 7665) Å line on the left (resp. on the right). *Top graphs*: the fitted interstellar model before convolution by the instrumental profile (blue line), superimposed on the raw data. *Bottom graphs*: the adjusted interstellar-telluric model (blue line) superimposed on the normalized data. For this target, the interstellar K I lines are Doppler shifted out of the strong telluric oxygen lines.

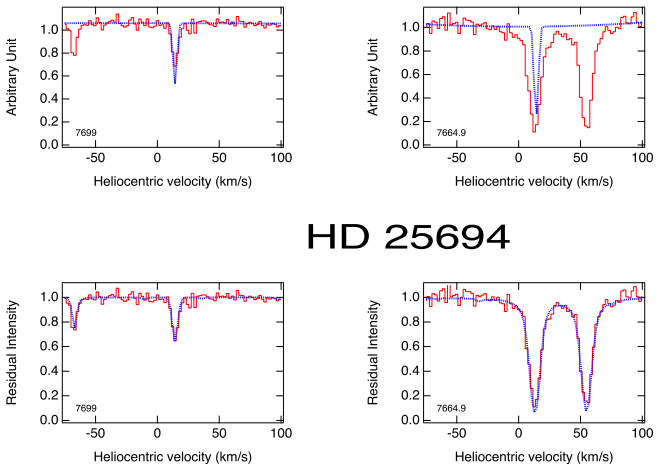


Fig. 4. Same as Fig. 3, but for the target star HD 25694. The 7665 Å K I line is blended with one of the telluric oxygen lines.

Despite the use of the K I doublet, such measurements are difficult because the lines are often saturated and the widths of the individual components are generally smaller than the LSF for most of the data. WH01 found a median component width (FWHM) lower than 1.2 km s^{-1}, compared to average values on the order of 3.5 km s^{-1} for the present data. Here we attempt to use solely the order of magnitude of the measured columns in our search for a counterpart of the absorption in the 3D dust map.

4.2. Profile-fitting results

Examples of profile-fitting results are presented in Figs. 3–6. We have selected different situations in terms of locations of the interstellar K I lines with respect to the strong telluric oxygen lines and in terms of cloud numbers. Figure 3 shows a case in which both K I lines are shifted out of the O_2 telluric lines and

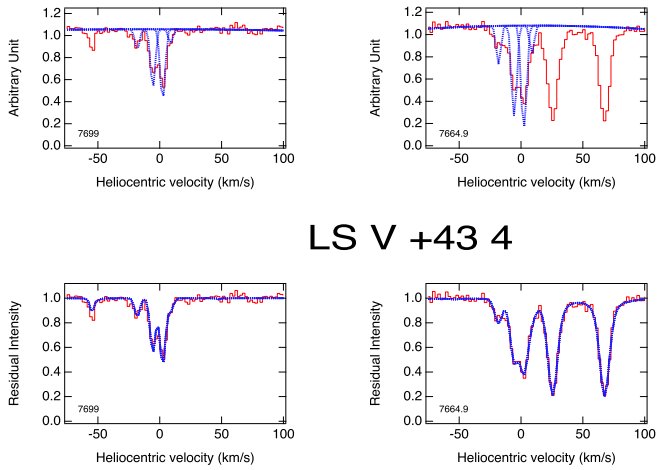


Fig. 5. Same as Fig. 3, but for the target star LS V +43 4. 4 individual clouds are necessary to obtain a good fit, i.e., with residuals within the noise amplitude.

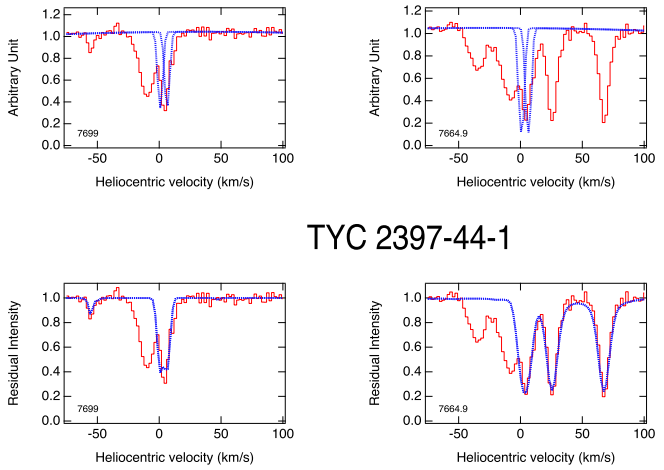


Fig. 6. Same as Fig. 3, but for the star TYC-2397-44-1, as an example of profile-fitting in the presence of strong, but Doppler shifted, stellar lines.

only one strong velocity component is detected. Figure 4 displays a case in which the K I 7664.9 absorption area coincides with the center of a strong O_2 telluric line. Still, the inclusion of an interstellar absorption contribution in this area is required to obtain a good adjustment and there is some information gathered from using the doublet instead of the K I 7699 line solely. Figure 5 corresponds to four distinct absorption velocities. Finally, Fig. 6 corresponds to a special case of a star cooler than the majority of other targets and possessing strong K I lines, here fortunately sufficiently Doppler shifted to allow measurements of the moderately shifted Taurus absorption lines.

The profile-fitting results are listed in Table A.2. We note that errors on velocities and K I columns have a limited significance and must be considered cautiously. Due to the partially arbitrary choice of cloud decomposition, our quoted errors on velocities correspond to this imposed choice. It has been demonstrated that there is a hierarchical structure of the dense, clumpy ISM in regions such as Taurus, and that a decomposition into more numerous substructures with relative velocity shifts on the order of 1 km s^{-1} is necessary if the spectral resolution is high enough (Welty et al. 1994, WH01). However, as we see in the next section, the 3D maps that are presently available do not allow one to disentangle the corresponding clumps that have a few parsec or subparsec sizes and this justifies the choice of our limitations during the profile-fitting. As a consequence of our allowed line broadening, uncertainties on the central velocities of the clouds may reach up to $\simeq$2 km s^{-1}. We have checked our fitted radial velocities with those measured by WH01 for the targets in common, ζPer, ξPer, and ϵPer, and we found agreement within the uncertainty quoted above by taking our coarser velocity decomposition into account. On the other hand, our quoted errors on the K I columns also correspond to our choices of decomposition and are lower limits. The order of magnitude of the columns is a precious indication for the following step of velocity assignment to structures revealed by the maps.

5. Updated 3D dust maps

5.1. Map construction

Massive amounts of extinction measurements toward stars distributed in distance and direction can be inverted to provide the location, in 3D space, of the masses of interstellar dust responsible for the observed extinction. Several methods have been used and the construction of 3D dust maps is in constant progress due the new massive stellar surveys and, especially, due to parallaxes and photometric data from the ESA *Gaia* mission. Here we use an updated version of the 3D extinction map presented by Lallement et al. (2019), which was based on a hierarchical inversion of extinctions from *Gaia* DR2 and 2MASS photometric measurements on the one hand, and *Gaia* DR2 parallaxes on the other hand. The inversion technique here is the same as for this previous map; however, the inverted extinction database and the prior distribution are different. In the case of the construction of the previous map, a homogeneous, plane-parallel distribution was used as the prior. Here, instead, the prior distribution is the previous 3D map itself. The inverted data set is no longer the set of *Gaia*-2MASS extinctions, instead it is now the extinction database made publicly available by Sanders & Das (2018), and it is slightly augmented by the compilation of nearby star extinctions used in Lallement et al. (2014). With the goal of deriving their ages for a large population of stars, Sanders & Das (2018) have analyzed, in a homogeneous way, data from six stellar spectroscopic surveys, APOGEE, GALAH, GES, LAMOST, RAVE, and SEGUE, and they estimated the extinction for a series of 3.3 million targets (if the catalog is restricted to the "best" flag, see Sanders & Das 2018). To do so, and in combination with the parameters inferred from the spectroscopic data, they used the photometric measurements from various surveys and in a large number of bands (J, H, and K from 2MASS; G, GBP, and GRP from *Gaia*; gP, rP, and iP from Pan-STARRS; g, r, and i from SDSS; and Jv ,Hv, and Kv from VISTA). They used a Bayesian model and priors on the stellar distributions as well as priors on the extinction derived from the Pan-STARRS 3D mapping (Green et al. 2018) or the large-scale model from Drimmel et al. (2003). Because the extinction determined by Sanders & Das (2018) is in magnitude per parsec in the *V* band, and since our prior map was computed as A_0 (i.e., the mono-chromatic extinction at 5500 Å), we applied a fixed correction factor to the prior distribution of differential extinction, a factor deduced from cross-matching A_0 extinctions used for this prior and the extinctions of Sanders & Das (2018) for targets in common. Due to the strong constraints brought by the spectroscopic information, the individual extinctions from the Sanders & Das (2018)

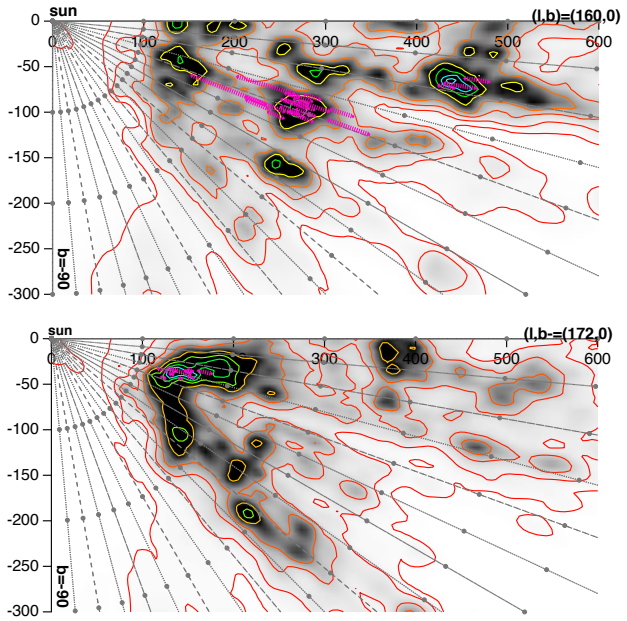


Fig. 7. Images of the inverted dust distribution in two vertical planes containing the Sun (located at 0,0) and oriented along Galactic longitudes of l = 160 and 172 degrees. Galactic latitudes are indicated by dashed black lines, every 5 degrees from b = 0 in the plane to b = −90 (South Galactic pole). Coordinates are in parsecs. The quantity represented in black and white is the differential extinction dA_V/dl at each point. Isocontours are shown for 0.0002,0.00045, 0.001, 0.002, 0.005, 0.01 , 0.015, and 0.02 mag per parsec. Superimposed are the closest and farthest locations of the molecular clouds from the Zucker et al. (2020) catalog that are at longitudes close to the one of the image (see text). A dashed pink line connects these two locations for each cloud.

catalog are more accurate than purely photometric estimates, and they are expected to allow one to refine the 3D mapping. This is why the minimal spatial correlation kernel for this new inversion, a quantity that corresponds to the last iteration of the hierarchical scheme, is 10 pc, compared to the 25 pc kernel of the *Gaia*-2MASS map (see Vergely et al. 2010 for details on the basic principles as well as limitations of the inversion, and see Lallement et al. 2019 for a description of the hierarchical inversion scheme). We note, however, that because the whole set of spectroscopic surveys is far from homogeneously covering all directions and distances, such a spatial resolution is not achievable everywhere and, in the case of regions of space that are not covered, the computed solution remains the one of the prior 3D map.

5.2. Taurus dust clouds in 3D

The Taurus area has been quite well covered by the spectroscopic surveys used by Sanders & Das (2018), in particular APOGEE and LAMOST (Majewski et al. 2017; Deng et al. 2012). This has allowed a more detailed mapping of the dust clouds in this region and justifies the use of the updated map for our kinetic study. In Fig. 7 we show images of the newly computed differential extinction in two planes perpendicular to the mid-plane and containing the Sun, oriented along Galactic longitudes of 160 and 172°, respectively. The color-coded quantity is the local differential extinction, or extinction per distance, a quantity highly correlated with the dust volume density. As mentioned above, an important characteristic of the maps and images is the minimum size of the cloud structures, directly connected to the minimum spatial correlation kernel imposed during the inversion. In the Taurus area, thanks to the target density the kernel is on the order of 10 pc almost everywhere in the regions of interest up to ≃500 pc. We have compared the locations of the dense structures reconstructed in Taurus, Perseus, and California with results on individual molecular cloud distances in the recent catalog of Zucker et al. (2020), based on PanSTARRS and 2MASS photometric data on the one hand, and *Gaia* DR2 parallaxes on the other hand. Superimposed on the first image along l = 160° (resp. 172°) are the closest and farthest locations of the clouds as measured by Zucker et al. (2020), restricting their list to cloud center longitudes comprised between 156 and 164° (resp. 166 and 176°). This implies that, here, we neglect the difference of up to 4° between the molecular cloud actual longitudes and the longitude of the vertical plane. We also solely used the statistical uncertainties on the distances quoted by Zucker et al. (2020), and we did not add the systematic uncertainty of 5% quoted by the authors. Despite these differences and reduced uncertainties, and despite our limited spatial resolution required by the full-3D inversion technique, it can be seen in Fig. 7 that there is good agreement between the locations of the dense structures obtained by inversion and the locations independently obtained by Zucker et al. (2020).

As mentioned in the introduction, large amounts of YSOs can now be identified and located in 3D space, and several studies have been devoted to their clustering and associations with known molecular clouds (Großschedl et al. 2021; Roccatagliata et al. 2020; Kounkel et al. 2018; Galli et al. 2019). If the stars are young enough, they remain located within their parent cloud and their proper motions are identical to the motion of the cloud. As an additional validation of the new 3D dust distribution, we have superimposed the locations of the Main Taurus YSO clusters found by Galli et al. (2019) in the dust density images used for the preliminary 3D kinetic tomography presented in the next section. The locations of these YSO clusters are also, in principle, those of the associated molecular clouds. L 1517/1519, L 1544, L 1495 NW, L 1495, B 213/216, B 215, Heiles Cloud 2, Heiles Cloud 2NW, L 1535/1529/1531/1524, L 1536, T Tau cloud, L 1551, and L 1558 are the main groups identified by the authors and are shown in Figs. 10–12. Although the dust map spatial resolution is not sufficient to disentangle close-by clouds, it can be seen from the series of figures that the YSO clusters coincide with very dense areas.

5.3. Link between extinction and K I absorption strength

Most of our targets do not have individually estimated extinctions. However, it is possible to use the 3D distribution of differential extinction and to integrate this differential extinction along the path between the Sun and each target star to obtain an estimate of the star extinction. We performed this exercise for our series of targets and compared the resulting extinctions A_V with the full equivalent widths of the 7699 Å K I absorption profiles. The comparison is shown in Fig. 8. Despite an important scatter, there is a clear correlation between the two quantities, a confirmation of the link between neutral potassium and dust as well as a counterpart of the link between K I and the H column shown by WH01. We note that we do not expect a tight correlation between the two quantities for several reasons. One of the reasons is the limited resolution of the map, and the fact that the extinction is distributed in volumes on the order of the kernel. As a consequence, high density regions in cloud cores are smeared out and high columns that occur for lines of sight crossing such cores have no comparably high counterparts in the

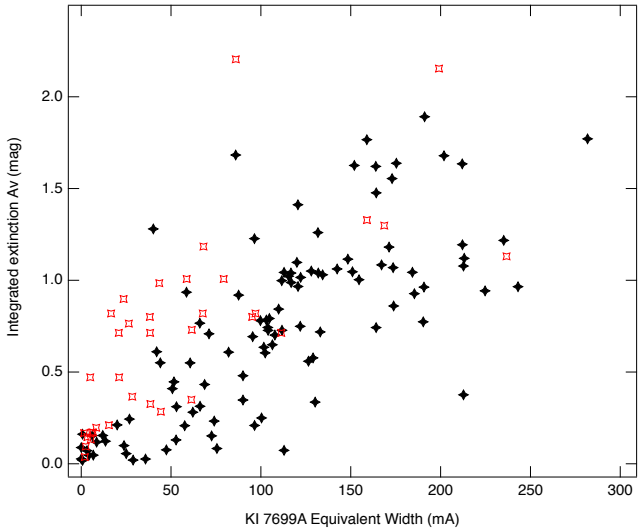


Fig. 8. Integrated extinction between the Sun and each target, using the 3D map, as a function of the full equivalent width of the 7699 Å K I absorption profile (black signs). Also shown are 7699Å K I equivalent widths from WH01 and extinctions estimated from the conversion of H columns (red signs, see text).

integrations. A second reason is the distance uncertainty. Despite their unprecedented accuracy, *Gaia* EDR3 uncertainties are on the order of a few parsec for the Taurus targets and this may influence the integral. An additional source of variability of a different type is the nonlinear relationship between the equivalent width we are using here and the column of atoms, and its dependence on the temperature, turbulence, and velocity structure (see, e.g., WH01). In order to compare the links between K I and H columns on the one hand, and between K I and dust columns on the other hand, we have used the 7699 Å K I equivalent widths measured by WH01 and we have converted H_{tot} columns taken from the same study into extinctions, assuming a constant ratio and $A_V = 1$ for $N(H_{tot}) = 6.\ 10^{21}/3.1$ cm^{-2}. Figure 8 displays the relationship between the two quantities, superimposed on our values of equivalent widths and extinctions. The scatter appears to be similar, as confirmed by the similarity of the Pearson correlation coefficients, 0.74 for our results and 0.72 for the quantities derived from the WH01 study. This suggests that, as expected, the link of K I with dust is comparable to its link with H. We note that our 105 data points correspond to Taurus stars, while the 35 ratios using K I and $N(H_{tot})$ from WH01 correspond to stars distributed over the sky. The average ratio between extinction and K I is smaller in our case, a large part of the difference is certainly linked to our limited spatial resolution and the corresponding smearing of the extinction. In the context of kinetic tomography, the obtained relationship is encouraging and it suggests that K I absorptions may be a convenient tool.

6. Preliminary 3D kinetic study

Our objective here is to study which constraints are brought about by our relatively small number of K I measurements in the context of kinetic tomography, more precisely what velocity assignments can be made to dense structures reconstructed in 3D by extinction inversion, using the measured absorption velocities and our most recent 3D dust maps exclusively. We started, by extracting from the 3D distribution of differential extinction, a series of images of dust clouds in vertical planes containing the Sun and oriented toward Galactic longitudes distributed between l = 150 and l = 182°, by steps of 2.5°, with the exception of l = 152.5°, due to the absence of targets, that is images similar to those in Fig. 7, now covering the whole Taurus-Perseus-California regions. About these images, it is important to keep in mind that the mapped structures have a minimum size of 10 pc, which prevents detecting smaller clumps. We used this set of images, the paths to the target stars, and the absorption results to link, in a nonautomated way, dust concentrations and velocities, plane after plane. To do this, for each image of the vertical plane, we selected the target stars whose longitudes are less than 1.3 degrees from the longitude of the plane and we plotted the projection of their line of sight on the plane. We did this for their most probable distance, in all cases available from the *Gaia* Early Data Release 3 (EDR3) catalog (Gaia Collaboration 2021) except for two targets. We then compared their observed K I absorption velocities as well as their associated approximate columns (in units of 10^{10} cm^{-2}) and the dense structures encountered along their trajectories. As we have already pointed out, this exercise would take too much time in the case of more numerous data and the results are partially arbitrary, but we want to illustrate here what can be done based only on a preliminary visual method. Work is underway to develop automated techniques.

Figure 9 displays the vertical plane along l = 150°. Planes along other longitudes are displayed in Figs. 10–12. In each figure, we restrict ourselves to radial velocities measured for stars with longitudes within 1.3° from the plane, listed at the bottom. To help visualize which star corresponds to which velocity assignment, we have numbered them and the numbers correspond to the objects listed in the text included in each figure. The velocity assignments to the dense clouds are indicated by arrows pointing outward (inward) for positive (negative) radial velocities. The length of the arrow is proportional to the velocity modulus. Two or more velocities at very close locations in the map indicate that information on the same cloud is provided by different stars. In some cases the arrows are slightly displaced to avoid the superimposition of different measurements based on stars at very close latitudes. In all cases, differences between measured velocities on the order of 1 km s^{-1} and sometimes up to 2 km s^{-1} may be due to profile-fitting uncertainties and do not necessarily imply a different clump.

The first result is linked to the exercise itself: it was surprisingly easy to associate velocities and clouds. For all K I columns on the order or above $\simeq 5 \times 10^{10}$ cm^{-2}, dense structures are found along the corresponding path to the target star and the cases with multiple velocities frequently offered obvious solutions. In general, K I starts to get detected for lines of sight that are crossing the 0.005 mag pc^{-1} differential extinction iso-contour, although this approximate threshold varies as a function of the distance to the plane and the signal-to-noise ratio (S/N) of the recorded spectrum. There is consistency between the near side of the reconstructed Main Taurus dust clouds and the distances to the targets with detected absorption. The closest star with detected K I is HD 26873 at 123.5 ± 1.3 pc, followed by HD 25204 at 124 ± 6.8 pc and DG Tau at 125.3 ± 2 pc. Only HD 23850 at 123.2 ± 7.3 pc has a smaller parallax distance, but the uncertainty on its distance is significant. Similarly, HD 280026 has no detected neutral potassium, but deep 5890 Å sodium absorption despite the most probable distance as short as 103 pc. However, there is no DR2 nor EDR3 *Gaia* parallax value, and this unique distance estimate is based on HIPPARCOS, with a large associated uncertainty of 30 pc. As a conclusion, it is likely that these last two targets are located within the near side boundary of the

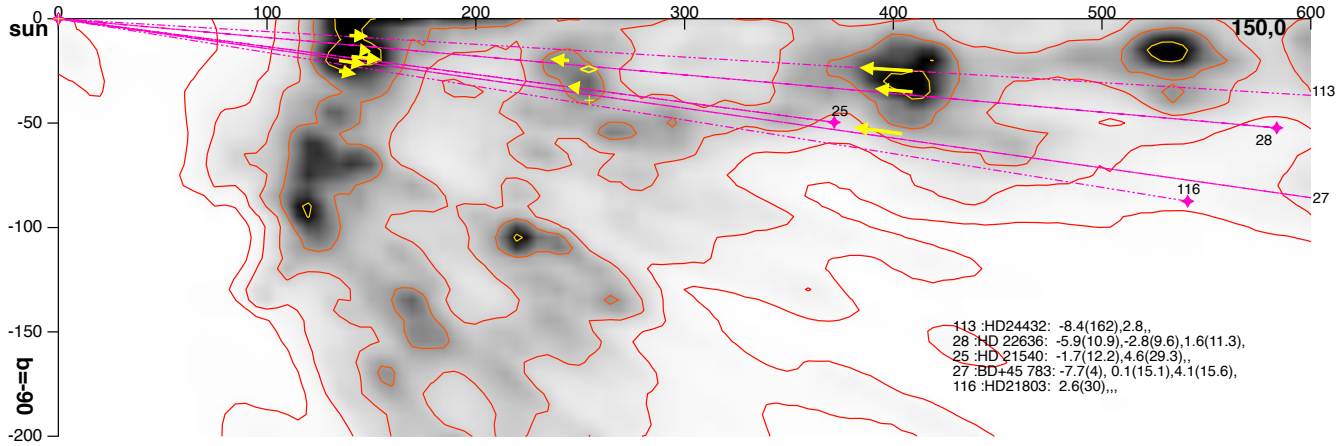


Fig. 9. Image of the inverted dust distribution in a vertical plane containing the Sun (located at 0,0) and oriented along Galactic longitude $l = 150°$. Coordinates are in parsecs. The quantity represented in black and white is the differential extinction. Iso-contours are shown for 0.0002, 0.00045, 0.001, 0.002, 0.005, 0.01 , 0.015, and 0.02 mag per parsec. The paths to the target stars whose longitudes are within 1.3 degrees from the longitude of the represented plane are superimposed. The stars are numbered as in the text and drawn at the bottom of the figure; additionally, the Doppler velocities and approximate columns of absorbing K I in 10^{10} cm^{-2} units are listed for each target (in parentheses after velocities). Stars' numbers are indicated in yellow or black. The list of targets is given by decreasing latitude. For distant targets falling outside the image, their numbers are indicated along their paths at the boundary of the figure. The preliminary velocity assignments to the dense clouds are indicated by arrows pointing outward (inward) for positive (negative) radial velocities. The length of the arrow is proportional to the velocity modulus.

Taurus complex, probably in the atomic phase envelop and very close to the molecular phase at ≃123 pc. These are encouraging results, showing that the 3D maps, the distances, and the velocity measurements are now reaching a quality that is actually allowing kinetic tomography. In many cases the assignment is not questionable (see, e.g., the star 55 in the $l = 165°$ image). In other numerous cases, the assignment based on one star strongly influences the solution for other stars. Finally, since there is continuity in the velocity pattern in consecutive planes, several consecutive planes can be used to follow both the shape and velocity changes of the same large structure.

We now discuss the results in more detail, plane by plane. The $l = 150°$ image shows two dense structures at 145 and 400 pc and a weaker cloud at 250 pc. Interestingly, the three distances are similar to the distances to Main Taurus, California, and Perseus, but these opaque regions are located much closer to the Galactic plane than the well known dense clouds. As can be seen from all figures (see, e.g., Fig. 10), this triple structure is still present at higher longitudes up to $l = 170°$. In the $l = 150°$ plane, a radial velocity decrease with distance, from ≃+5 down to ≃−8 km s^{-1}, is clearly associated with these three groups, that is to say there is a compression of the whole region along the radial direction. The intermediate structure at 250 pc has a very small or null LSR radial velocity. The $l = 155°$ image confirms the velocity pattern (despite the small number of stars) and starts to show the densest cloud complexes, which are better represented in the next vertical planes. Perseus reaches a maximal density at $l \simeq 157.5$ and, at this longitude, is connected to a series of structures located between 250 and 320 pc and reaching the Galactic Plane. California is the most conspicuous structure at $l \simeq 160$–$162°$, and it starts to be divided into two main parts at $l \simeq 162°$, the most distant one reaching 540 pc. Along $l = 157.5°$, the global velocity pattern slightly changes, with the appearance of a positive velocity gradient between Main Taurus and Perseus. The structure at 320 pc (above Perseus in the image) is peculiar, with a larger positive radial velocity (marked by a question mark), but further measurements are required to conclude. In the $l = 157.5°$ and $l = 160°$ images, we have located the dense molecular clouds L1448, L1451, and IC348 in Perseus at their distances found by Zucker et al. (2018) and we have compared the radial velocities we can assign to L1448 and IC348 with the radial velocities deduced by the authors based on CO data and 3D extinction. From the strong absorption toward BD+30 540 (star 48) and unambiguously assigned to the region with maximal opacity that corresponds to L1448, we derived $V_r = +5$ km s^{-1} for this cloud, close to the 4.8 km s^{-1} peak-reddening radial velocity of Zucker et al. (2018). From several targets crossing the IC348 area (see the next $l = 160°$ plane), we derived $V_r \simeq +8$ km s^{-1} for IC348, which is close to their average velocity of +8.5 km s^{-1}. Still in the $l = 160°$ plane, the structure in the foreground of IC348 at ≃140 pc (a distance in agreement with Zucker et al. (2018), see their Fig. 6) appears to have a complex velocity structure with smaller velocities for the dense central region. However, again, more data is needed to better map and quantify this property. California at 450 pc is characterized by a decreasing modulus of the inward motion with increasing distance from the Galactic plane.

In the $l = 165°$ vertical plane, one starts to see the disappearance of California and Perseus, and the change in California radial velocities, from negative to null values at about 480 pc. In this image and all the following ones (up to $l = 182.5°$), the main, densest areas of Main Taurus are the most interesting structures. The YSO clusters derived by Galli et al. (2019) in Taurus are found to be located in the densest parts, and the maps confirm the distribution in radial distance of the densest areas found by the authors and previous works, from 130 to 160 pc in the case of all clouds at latitudes below the $b \simeq -7°$ area. An interesting result is the existence of a series of measurements of small radial velocities we could associate with the most distant parts of the clouds (see the four images in Fig. 11). We cannot distinguish these velocities from the main flow in all stars, only excellent signal-to-noise and spectral resolution allow one to do it; however, their numbers strongly suggest the existence of a negative gradient directed outward, and this compression is very likely connected to the large number of star-forming regions in this area. The case of L1558 is peculiar. It is the only well known

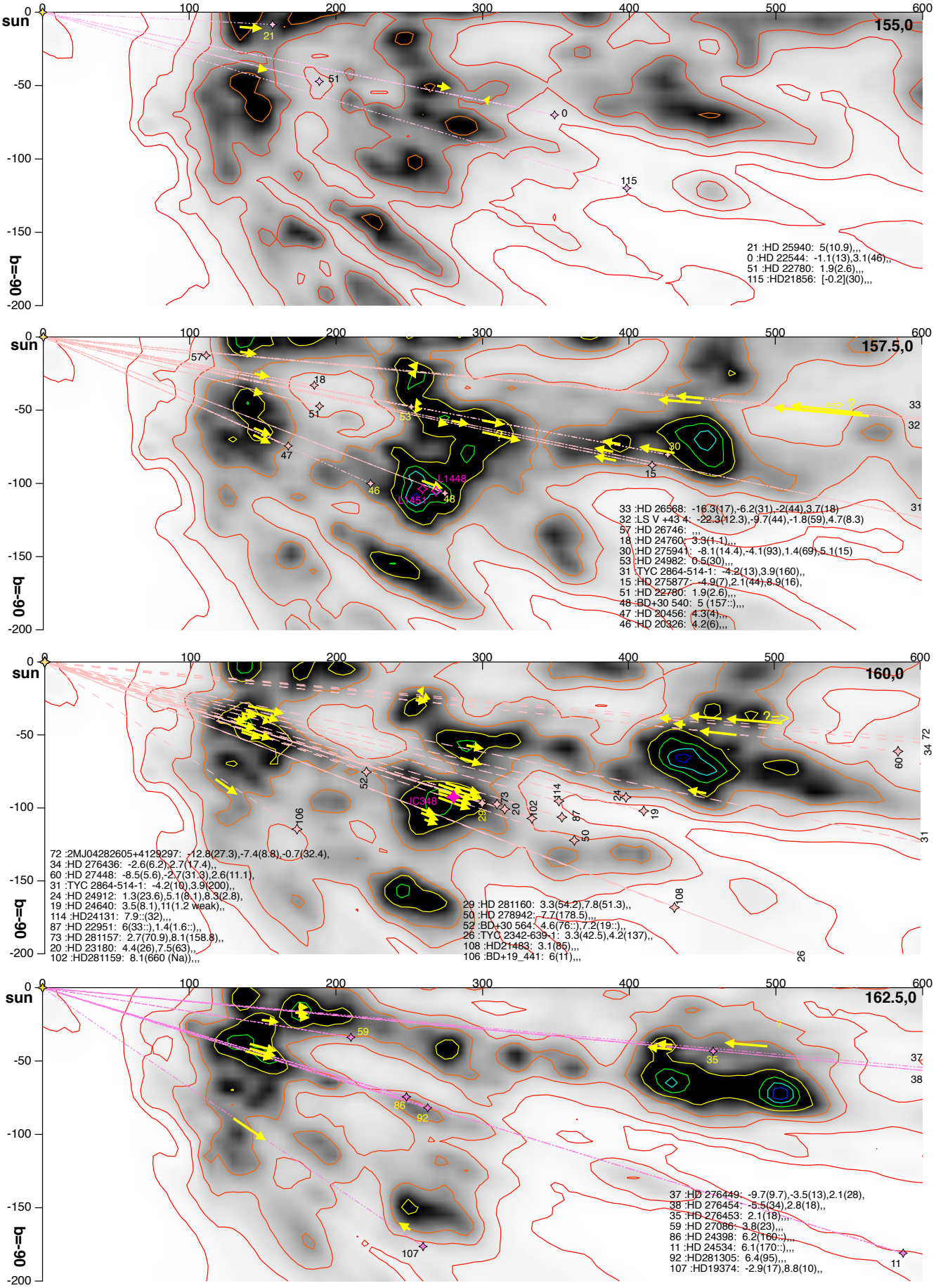


Fig. 10. Same as Fig. 9, but for longitudes l = 155, 157.5, 160, and 162.5°. For pink markers see text.

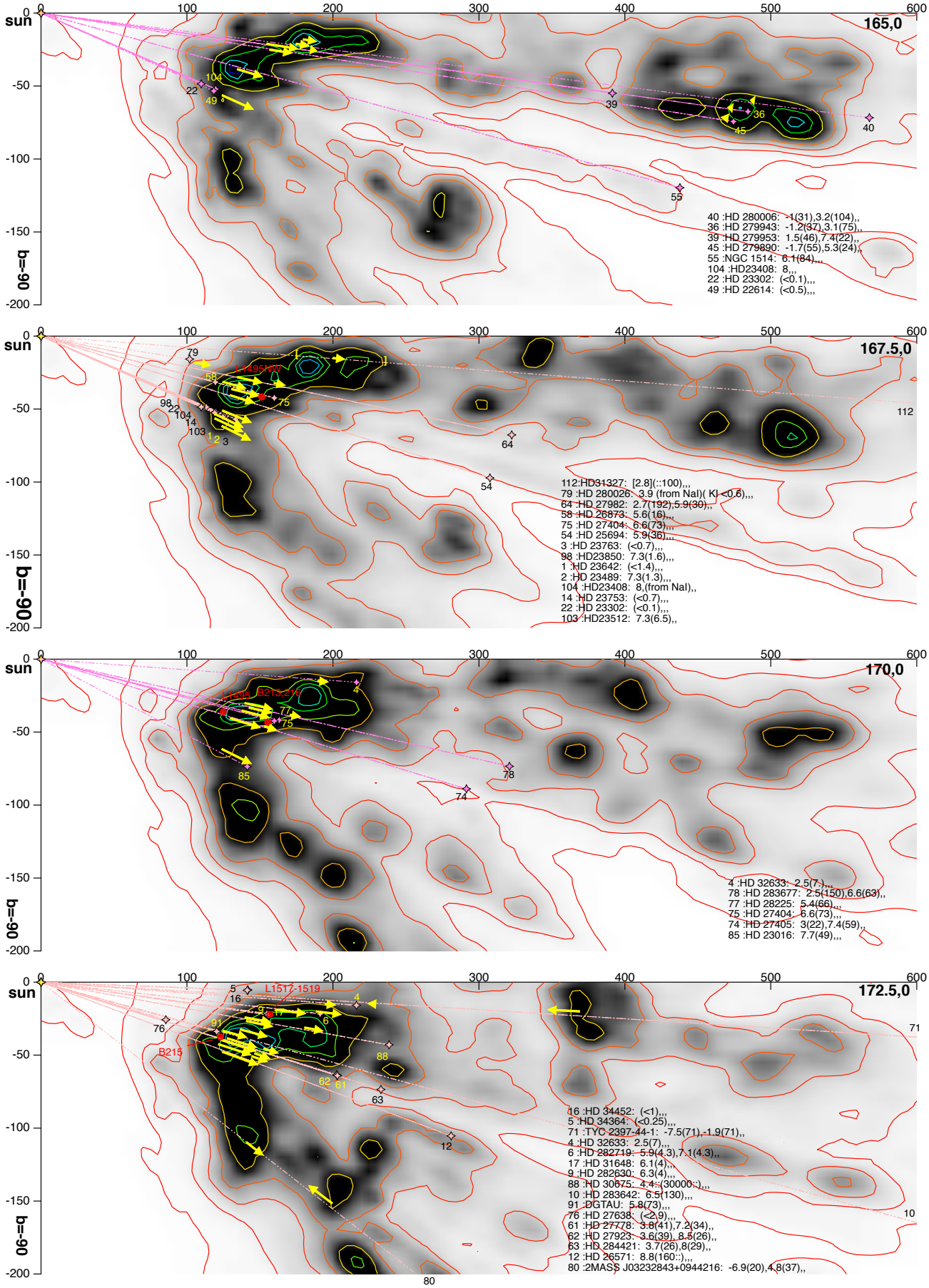


Fig. 11. Same as Fig. 9, but for longitudes l = 165, 167.5, 170, and 172.5°. For red dots see text.

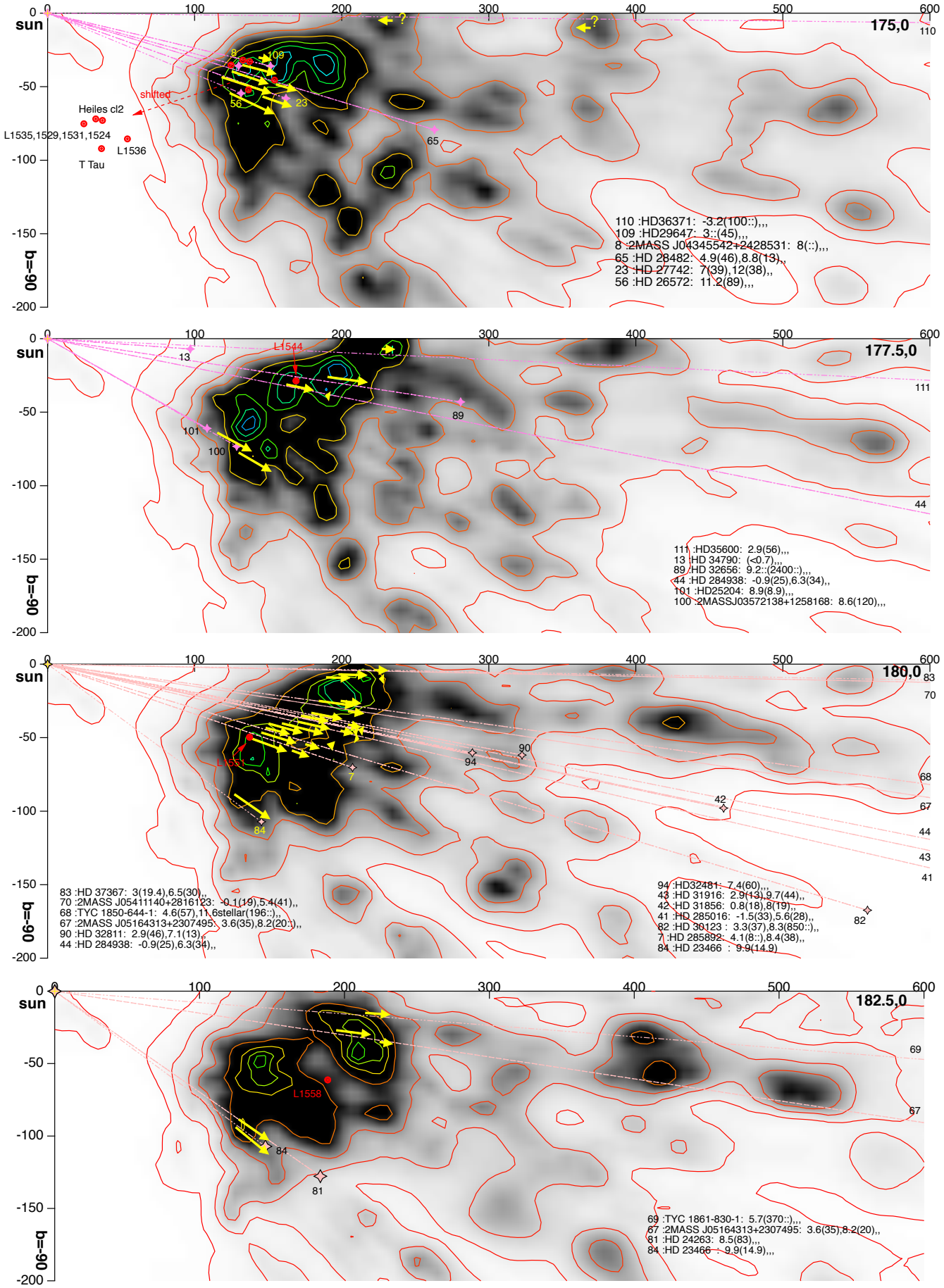


Fig. 12. Same as Fig. 9, but for longitudes l = 175, 177.5, 180, and 182.5°. For red dots see text.

structure found to be located in a moderately dense area; however, this may be an imperfection of the map due the lack of stars because it is located behind a particularly opaque area. From the point of view of the kinematics, we have used the positions and velocities in Cartesian coordinates of Galli et al. (2019) to compute the radial velocities in heliocentric and LSR frame, and we compared the results with our assignments when possible. The results are found to be compatible within 1.5 km s^{-1}. For L1495 (l = 170° figure), the most appropriate measurement is for star 74, with a velocity of +7.4 km s^{-1} for the main component, which is the closest to the Sun. According to Galli et al. (2019), the radial velocity of the cluster is +6.9 km s^{-1}. For B215, we can use several stars that provide velocities of +8, +8.5, +7.2, +5.8, and +6.5 km s^{-1}, that is an average of ≃+7km s^{-1}, while the estimated value of Galli et al. (2019) is +7.5 km s^{-1}. For L1517-1519, we can use star 9 with +6.3 km s^{-1}, while the authors find +4.8 km s^{-1}. For L1536, our low velocity value from star 65 is +4.9 km s^{-1}, which is compatible with +5.9 km s^{-1}. For T Tau, we can use the target star 23, which has two components at +7 and +12 km s^{-1}. The cluster found by Galli et al. (2019) is apparently located on the more distant, low velocity side with a value of +7.8 km s^{-1}. About L1544, star 44 has its high velocity component at +6.3 km s^{-1}, which is compatible with the +7.6 km s^{-1} found for the cluster. For all these cases, the comparisons seem to confirm the negative velocity gradient we have discussed above, that is to say a general compression and deceleration along the radial direction. L1551 does not follow the rule and there is no agreement between our estimated velocity pattern and the result from Galli et al. (2019): if the cluster and associated cloud is located as is shown in the l = 180° plane, it should be within the high velocity part of the two regions probed by star 7, that is at +8.4 km s^{-1}; however, its velocity of +5.6 km s^{-1} is closer to the low velocity component detected for this star, +4.1km s^{-1}. On the other hand, in this very thick region, the velocity pattern may be more complex and it may require more constraints.

7. Conclusions and perspectives

We have investigated the link between interstellar K I absorption and dust opacity, and we tested the use of K I absorption data in the context of 3D kinetic tomography of the ISM. To do so, we have obtained and analyzed high-resolution stellar spectra of nearby and mainly early-type stars in the Taurus-Perseus-California area, recorded with the Narval spectrograph at TBL/Pic du Midi. We have developed a new technique based on synthetic atmospheric transmission profiles that allows us to extract a maximum of information from the interstellar K I absorption doublet and we have applied the new technique to the Narval data as well as to archival data, mainly high-resolution spectra from the Polarbase archive. Atmospheric profiles were downloaded from the TAPAS facility for each observing site and adjusted to the data. During the adjustment, the instrumental width and its variation with wavelength was derived. A new interstellar-telluric profile-fitting using Voigt profiles in a classical way and the previously derived telluric profile was performed and the radial velocities of the main absorbing interstellar clouds were derived. The adjustment used the measured LSFs, one for each transition of the doublet. We present the results of this profile-fitting for 108 targets, complemented by results from the literature for eight additional stars.

In parallel, we computed an updated 3D distribution of interstellar dust, based on the inversion of a large catalog of extinction measurements for stars distributed at all distances. The maps were obtained in a similar way as the one based on *Gaia*/2MASS presented in Lallement et al. (2019), but the inverted data set was the extinction catalog computed by Sanders & Das (2018) as an auxiliary product of their study on stellar ages, augmented by the small compilation of nearby star extinctions used in Lallement et al. (2014). Importantly, instead of an initial, homogeneous, plane-parallel prior, the new prior was the *Gaia*-2MASS map itself. As a result, in regions that are not covered by the spectroscopic surveys analyzed by Sanders & Das (2018), the recovered 3D distribution is the unchanged prior, that is the *Gaia*/2MASS map, and in areas covered by the surveys there is additional information from the combination of spectral and photometric information. The spatial resolution of the dust maps is limited by the target star density and the correlation length imposed in the inversion. Here, the last iteration was done for a 10 pc wide kernel, and this implies that the recovered dust masses are spread out over volumes on the order of 10 pc or more.

As a first test of the link between K I and dust opacity, we compared the equivalent width of the 7699 Å absorption with the integral of the differential extinction between the Sun and the target star throughout the new 3D map. The relationship between the two quantities is compared with the results of WH01 for K I and the H column (HI +2 H_2) and we obtain similar correlation coefficients. This confirms that neutral potassium is tracing both atomic and molecular phases in dense structures. Based on the comparison between the dust cloud distribution and the absorption velocities, we have made a preliminary assignment of radial velocities to the dense structures showing up in the maps. According to the above description, our study had several limitations: the limited number of lines of sight (≃120 targets); the limited spatial resolution of the maps (≃10 pc); and the partially arbitrary decomposition of K I absorption profiles into discrete clouds, leading to a minimum velocity difference on the order of 1–2 km s^{-1}. However, despite of these limitations, we found that, except for very few measurements, it was possible and relatively easy to find a coherent solution for the connections between clouds and velocities, with some continuity in the direction and distance for the velocity field. Large-scale spatial gradients were retrieved, and some internal gradients could be determined. We present images of the dust distribution in a series of vertical planes containing the Sun and oriented along Galactic longitudes from 150 to 182.5°. Proposed velocity assignments and paths to the used target stars are indicated in each image, allowing one to visualize the velocity pattern and check the sources of the constraints. We show several comparisons between our maps and velocity assignments with the locations of the molecular cloud and young star clusters and their radial velocities from recent works by Zucker et al. (2018, 2020) and Galli et al. (2019). There is a good agreement between both positions and radial velocities.

This study shows that it is possible to obtain a first, spatially coherent 3D kinetic tomography of the ISM in out-of-plane nearby regions, using solely dust maps and K I absorption data. Such a preliminary tomography can be used as a prior solution for more ambitious, automated kinetic tomography techniques, using dust maps in combination with emission data, and/or with massive amounts of absorption data from future surveys. It shows that K I absorptions can also be a very efficient tool to check or validate tomographic results based on other absorption or emission spectral data. A main advantage of using K I as a tracer of Doppler velocities is the sharpness of the absorption lines and the resulting optimal disentangling of multiple clouds, at variance with DIBs that are broad. DIBs have already proven to be useful tools for kinetic measurements in the visible (e.g., Puspitarini et al. 2015), but they are most useful in the infrared wavelength range due to better access to a target beyond

opaque regions (Zasowski & Ménard 2014; Zasowski et al. 2015; Tchernyshyov et al. 2018). In this respect, they are superior to K I. On the other hand, another advantage of K I, as quoted above, is its association with both atomic and molecular phases, as shown by WH01, and extrapolated here to the dust opacity. For this reason, K I will be particularly well adapted to kinetic tomography using dust maps and devoted to extended regions consisting of both phases, including dense gas reservoirs devoid of CO. Ultimately, 3D kinetic tomography may be best achieved through the combination of various tracers, ideally in absorption and emission, and, obviously, using all results directly or indirectly related to *Gaia* catalogs. In all cases, assigning velocities to structures seen in 3D would allow one to link the structures to emission lines that possess particularly rich information on physical and chemical mechanisms at work and to shed light on evolutionary processes.

Acknowledgements. R.L. deeply thanks the Pic du Midi TBL staff for his efficiency and assistance. We greatly appreciated the constructive comments and also the criticisms of our referee, which led to a significant improvement of the article. A.I. wants to acknowledge the support of Ministry of Science and Higher Education of the Russian Federation under the grant 075-15-2020-780 (N13.1902.21.0039). J.L. Vergely acknowledges support from the EXPLORE project. EXPLORE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004214. This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France.

References

Anders, F., Khalatyan, A., Chiappini, C., et al. 2019, A&A, 628, A94
Antoja, T., Helmi, A., Romero-Gómez, M., et al. 2018, Nature, 561, 360
Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, AJ, 161, 147
Bertaux, J. L., Lallement, R., Ferron, S., Boonne, C., & Bodichon, R. 2014, A&A, 564, A46
Cami, J., Cox, N. L., Farhang, A., et al. 2018, The Messenger, 171, 31
Chaffee, F. H., J., & White, R. E. 1982, ApJS, 50, 169
Chen, B. Q., Huang, Y., Yuan, H. B., et al. 2019, MNRAS, 483, 4277
Cox, N. L. J., Cami, J., Farhang, A., et al. 2017, A&A, 606, A76
Dame, T. M., Hartmann, D., & Thaddeus, P. 2001, ApJ, 547, 792
Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, Res. Astron. Astrophys., 12, 735
Donati, J. F., Semel, M., Carter, B. D., Rees, D. E., & Collier Cameron, A. 1997, MNRAS, 291, 658
Drimmel, R., Cabrera-Lavers, A., & López-Corredoira, M. 2003, A&A, 409, 205
Fan, H., Hobbs, L. M., Dahlstrom, J. A., et al. 2019, ApJ, 878, 151
Gaia Collaboration (Brown, A. G. A., et al.) 2021, A&A 649, A1
Galli, P. A. B., Loinard, L., Bouy, H., et al. 2019, A&A, 630, A137
Green, G. M., Schlafly, E. F., Finkbeiner, D., et al. 2018, MNRAS, 478, 651
Green, G. M., Schlafly, E., Zucker, C., Speagle, J. S., & Finkbeiner, D. 2019, ApJ, 887, 93
Großschedl, J. E., Alves, J., Meingast, S., & Herbst-Kiss, G. 2021, A&A, 647, A91
Guo, H. L., Chen, B. Q., Yuan, H. B., et al. 2021, ApJ, 906, 47
Haywood, M., Snaith, O., Lehnert, M. D., Di Matteo, P., & Khoperskov, S. 2019, A&A, 625, A105
Helmi, A. 2020, ARA&A, 58, 205
HI4PI Collaboration (Ben Bekhti, N., et al.) 2016, A&A, 594, A116
Hottier, C., Babusiaux, C., & Arenou, F. 2020, A&A, 641, A79
Katz, D., Gomez, A., Haywood, M., Snaith, O., & Di Matteo, P. 2021, A&A, submitted [arXiv:2102.02082]
Khoperskov, S., Gerhard, O., Di Matteo, P., et al. 2020, A&A, 634, L8
Khoperskov, S., Haywood, M., Snaith, O., et al. 2021, MNRAS, 501, 5176
Kounkel, M., Covey, K., Suárez, G., et al. 2018, AJ, 156, 84
Lallement, R., Vergely, J.-L., Valette, B., et al. 2014, A&A, 561, A91
Lallement, R., Babusiaux, C., Vergely, J. L., et al. 2019, A&A, 625, A135
Leike, R. H., Glatzle, M., & Enßlin, T. A. 2020, A&A, 639, A138
Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94
Petit, P., Louge, T., Théado, S., et al. 2014, PASP, 126, 469
Pety, J., Guzmán, V. V., Orkisz, J. H., et al. 2017, A&A, 599, A98
Puspitarini, L., & Lallement, R. 2012, A&A, 545, A21
Puspitarini, L., Lallement, R., Babusiaux, C., et al. 2015, A&A, 573, A35
Queiroz, A. B. A., Anders, F., Chiappini, C., et al. 2020, A&A, 638, A76
Rezaei Kh., S., Bailer-Jones, C. A. L., Soler, J. D., & Zari, E. 2020, A&A, 643, A151
Roccatagliata, V., Franciosini, E., Sacco, G. G., Randich, S., & Sicilia-Aguilar, A. 2020, A&A, 638, A85
Sanders, J. L., & Das, P. 2018, MNRAS, 481, 4093
Sharma, S., Hayden, M. R., & Bland-Hawthorn, J. 2020, ArXiv e-prints [arXiv:2005.03646]
Tchernyshyov, K., & Peek, J. E. G. 2017, AJ, 153, 8
Tchernyshyov, K., Peek, J. E. G., & Zasowski, G. 2018, AJ, 156, 248
Vergely, J.-L., Valette, B., Lallement, R., & Raimond, S. 2010, A&A, 518, A31
Wang, S.-i., Hildebrand, R. H., Hobbs, L. M., et al. 2003, SPIE Conf. Ser., 4841, 1145
Welty, D. E., & Hobbs, L. M. 2001, ApJS, 133, 345
Welty, D. E., Hobbs, L. M., & Kulkarni, V. P. 1994, ApJ, 436, 152
Zasowski, G., & Ménard, B. 2014, in The Diffuse Interstellar Bands, eds. J. Cami, & N. L. J. Cox (Cambridge: Cambridge University Press), 297, 68
Zasowski, G., Ménard, B., Bizyaev, D., et al. 2015, ApJ, 798, 35
Zucker, C., Schlafly, E. F., Speagle, J. S., et al. 2018, ApJ, 869, 83
Zucker, C., Speagle, J. S., Schlafly, E. F., et al. 2020, A&A, 633, A51

Appendix A: Tables

Table A.1. Targets.

Star ID	Galactic longitude	Galactic latitude	Distance	Distance error	Origin
HD 22544	153.75	−11.34	356	16	ESPaDOnS*
HD 23642	166.53	−23.32	138.3	0.9	ESPaDOnS*
HD 23489	166.37	−23.50	136.1	0.5	ESPaDOnS*
HD 23763	166.68	−23.12	144.9	4.7	ESPaDOnS*
HD 32633	171.16	−4.19	216.7	1.8	ESPaDOnS*
HD 34364	172.77	−2.23	141.4	0.9	ESPaDOnS*
HD 282719	172.79	−6.45	192.4	0.7	ESPaDOnS*
HD 285892	179.21	−18.72	218.9	0.9	ESPaDOnS*
2M J04345542+2428531	174.32	−15.40	134.7	1.6	ESPaDOnS*
HD 282630	172.68	−8.16	157.2	0.5	Narval*
HD 283642	171.55	−15.35	653.3	7.1	ESPaDOnS*
HD 24534	163.08	−17.14	614	14	ESPaDOnS*
HD 26571	172.42	−20.55	300	10	ESPaDOnS*
HD 34790	176.57	−4.12	97.3	0.7	ESPaDOnS*
HD 23753	167.33	−23.83	130.0	1.7	ESPaDOnS*
DG Tau	171.84	−15.76	125.3	2.0	ESPaDOnS*
HD 275877	156.80	−11.90	424.9	5.5	Narval*
HD 34452	172.86	−2.13	141.8	2.1	Narval*
HD 31648	173.47	−7.90	156.2	1.3	Narval*
HD 24760	157.35	−10.09	188	15	Narval*
HD 24640	160.47	−13.97	423	50	Narval*
HD 23180	160.36	−17.74	331	43	Narval*
HD 25940	153.65	−3.05	157	7	Narval*
HD 23302	166.18	−23.85	120	6	Narval*
HD 27742	175.32	−19.67	172.1	1.4	Narval*
HD 24912	160.37	−13.11	409	44	Narval*
HD 21540	149.00	−7.60	375	13	Narval
TYC 2342-639-1	159.44	−21.25	562	7	Narval
BD+45 783	149.97	−8.13	632	22	Narval
HD 22636	149.29	−5.12	586	6	Narval
HD 281160	160.45	−17.89	315.0	2.6	Narval
HD 275941	157.59	−10.67	434	6	Narval
TYC 2864-514-1	158.77	−11.62	664	7	Narval
LS V +43 4	157.03	−5.38	2420	117	Narval
HD 26568	157.02	−5.25	840	17	Narval
HD 276436	161.05	−5.26	866	15	Narval
HD 276453	161.55	−5.41	459.2	4.3	Narval
HD 279943	164.31	−7.93	489	6	Narval
HD 276449	161.47	−5.17	1620	63	Narval
HD 276454	161.73	−5.37	964	22	Narval
HD 279953	164.77	−7.99	395.2	3.7	Narval
HD 280006	165.11	−7.20	572	71	Narval
HD 285016	179.63	−13.00	755	15	Narval
HD 31856	179.53	−12.02	470.0	5.5	Narval
HD 31916	179.57	−11.93	704	15	Narval
HD 284938	178.84	−11.23	730	41	Narval
HD 279890	164.54	−8.90	480.2	4.5	Narval
HD 20326	157.68	−24.11	244.9	2.1	Narval
HD 20456	157.98	−23.97	183.1	2.5	Narval
BD+30 540	157.97	−21.24	294.3	2.0	Narval
HD 22614	164.67	−24.27	129.8	0.6	Narval
HD 278942	159.85	−18.61	383	13	Narval
HD 22780	156.37	−14.04	194.5	5.1	Narval
BD+30 564	160.77	−18.88	233.0	1.3	Narval

Notes. Distances were calculated based on the parallaxes from *Gaia* EDR3. The asterisk represents archive data for which a profile-fitting was performed. The second part of the table below the two horizontal lines contains data which are published or could be found in other archives for which no profile-fitting was done. (Na) means that only sodium could be extracted. (7699) means that only the 7699 Å K I line could be extracted, otherwise data corresponds to the K I doublet. [(1)]Distance from Bailer-Jones et al. (2021).

Table A.1. continued.

Star ID	Galactic longitude	Galactic latitude	Distance	Distance error	Origin
HD 24982	158.42	−10.75	255.8	2.4	Narval
HD 25694	166.68	−17.54	322.5	2.2	Narval
NGC 1514	165.53	−15.29	453.8	3.5	Narval
HD 26572	174.69	−22.50	142.3	1.0	Narval
HD 26746	158.58	−6.45	111.8	0.4	Narval
HD 26873	167.14	−14.78	123.5	1.3	Narval
HD 27086	162.00	−9.14	212.9	1.2	Narval
HD 27448	159.72	−5.98	588	10	Narval
HD 27778	172.76	−17.39	212.2	1.3	Narval+UVES
HD 27923	173.38	−17.51	212.9	2.7	Narval
HD 284421	173.44	−17.52	244.2	1.0	Narval
HD 27982	167.26	−11.86	329.5	7.3	Narval
HD 28482	174.33	−16.76	274.7	1.9	Narval
HD 37752	184.35	−3.65	214.8	1.9	Narval
2M J05164313+2307495	181.34	−8.60	4658.3[(1)]	1336	Narval
TYC 1850-644-1	179.62	−7.77	673	8	Narval
TYC 1861-830-1	183.36	−4.49	1158	24	Narval
2M J05411140+2816123	180.06	−1.18	1041	38	Narval
TYC 2397-44-1	171.63	−3.58	2636	125	Narval
2M J04282605+4129297	160.54	−5.01	1453	46	Narval
HD 281157	160.49	−17.44	324.9	2.0	Narval
HD 27405	171.02	−16.95	304.7	2.5	Narval
HD 27404	168.71	−14.85	165.3	0.7	Narval
HD 27638	171.51	−16.74	89.3	0.7	Narval
HD 28225	170.74	−14.35	168.4	0.7	Narval
HD 283677	170.37	−12.89	329.2	2.2	Narval
HD 280026	167.02	−8.78	103	23	Narval
2M J03232843+0944216	173.19	−37.88	1081	21	Narval
HD 32481	180.90	−11.79	465.4	7.5	Narval (7699)
HD 281305	162.94	−17.32	273	3	Narval (7699)
HD 24263	182.07	−34.87	223.4	3.7	UVES*
HD 30123	180.11	−16.71	582.1	7.2	UVES*
HD37367	179.04	−1.03	1312	256	UVES*
HD 23466	181.28	−36.39	180.6	3.5	UVES*
HD 23016	168.99	−27.51	159.3	2.1	UVES*
HD 24398	162.29	−16.69	259	28	UVES*
HD 22951	158.92	−16.70	370	18	UVES*
HD 30675	173.61	−10.20	242.3	4.4	X-shooter*
HD 32656	177.18	−8.70	284.1	2.5	X-shooter*
HD 32811	180.50	−10.85	328.5	3.6	FEROS*
2M J03572138+1258168	177.34	−29.66	147.9	3.5	FEROS (Na, No measurable K I)
HD 25204	178.37	−29.38	124.4	6.8	FEROS (Na, No measurable K I)
HD 281159	160.49	−17.80	350.6	9.2	ELODIE (Na, No measurable K I)
HD 23512	166.85	−23.95	136.3	2.6	ARCES (7699)
HD 23408	166.17	−23.51	130.4	5.3	ARCES (Na, No measurable K I)
BD+19 441	159.60	−33.54	207.4	1.1	ARCES (7699)
HD 19374	162.98	−34.21	313.7	6.2	ARCES (7699)
HD 21483	158.87	−21.30	463.2	5.7	ARCES (7699)
HD 29647	174.05	−13.35	155.8	1.0	ARCES (7699)
HD 23850	167.01	−23.23	123.2	7.3	Welty & Hobbs 2001
HD 36371	175.77	−0.61	1100.4	241.4	Chaffee & White 1982
HD 35600	176.76	−2.70	906.9	34.8	Chaffee & White 1982
HD 31327	168.14	−4.40	942.6	28.9	Chaffee & White 1982
HD 24432	151.12	−3.50	1139.3	27.1	Chaffee & White 1982
HD 24131	160.23	−15.14	365.2	10.3	Chaffee & White 1982
HD 21856	156.32	−16.75	415.5	15.0	Chaffee & White 1982
HD 21803	150.61	−9.18	543.8	11.0	Chaffee & White 1982

[1] Distance from Bailer-Jones et al. (2021)

Table A.2. Velocity table.

Star ID	V_1	N_1	V_2	N_2	V_3	N_3	V_4	N_4
HD 22544	3.09(±0.2)	17.3(±25.12)	−1.10(±0.4)	17.2(±12.96)				
HD 23642		(<0.1)						
HD 23489		(<1.0)						
HD 23763		(<0.7)						
HD 32633	2.5(±0.09)	7(±1)						
HD 34364		(<0.25)						
HD 282719	5.9(±2)	4.3(±2)	7.1(±2)	4.3(±2)				
HD 285892	4.1(±0.1)	8::	8.4(±0.6)	38(±10)				
2MASS	8.0(±1)	::cool star, no derived column						
J04345542+2428531								
HD 282630	6.26(±0.09)	4(±1)						
HD 283642	6.45(±0.07)	130(±10)						
HD 24534	6.12(±0.01)	170(±20)						
HD 26571	8.83(±0.05)	160(±60)						
HD 34790		(<0.7)						
HD 23753		(<0.8)						
DG Tau	5.8(±1)	73(±10)						
HD 275877	−4.93(±0.87)	7(±5)	8.92(±0.43)	16(±5)	2.06(±0.15)	44(±10)		
HD 34452		(<1)						
HD 31648	6.1(±0.25)	4(±4)						
HD 24760	3.3(±1)	1.1(±0.3)						
HD 24640	3.51(±0.16)	8(±1)	11.0(±1)	1::				
HD 23180	4.40(±1.0)	26(±5)	7.50(±0.5)	63(±2.5)				
(26) et 7.5 (63)								
HD 25940	5.02(±0.05)	11(±3)						
HD 23302		(<0.1)						
HD 27742	7.0(±0.5)	39(±2)	12(±1)	38(±5)				
HD 24912	5.06(±1.67)	8(±10)	1.34(±0.36)	20(±7)	8.34 (±0.36)	3(±4)		
HD 21540	4.55(±0.21)	29(±8)	−1.69(±0.44)	12(±7)				
TYC 2342-639-1	4.23(±0.36)	137(±80)	3.30(±0.46)	42(±6)				
BD+45 783	0.1(±0.18)	15.1(±1)	4.1(±1)	15.6(±2)	−7.7 (±1.5)	4(±1)		
HD 22636	−2.8(±1)	9.6::	−5.9(±1)	10.9::	1.6(±1)	11.3::		
HD 281160	7.78(±0.07)	51(±9)	3.34(±0.17)	54(±50)				
HD 275941	−8.12(±1)	14.43(±39.7)	−4.12(±0.1)	93.41(±257.02)	1.39(±0.1)	69.22(±275.87)	5.09(±0.6)	15.15(±43.83)
TYC 2864-514-1	−4.21(±0.41)	13(±7)	3.88(±0.08)	160(±20)				
LS V +43 4	−9.68(±0.11)	44(±6)	−22.30(±0.22)	12.3(±4)	−1.77(±0.12)	59(±9)	4.73(±0.67)	8.3(±5)
HD 26568	−16.3(±0.1)	17(±6)	−6.2(±0.3)	31(±11)	−2.0 (±0.1)	44(±31)	3.7 (±0.1)	18 (±6)
HD 276436	−2.59(±1.74)	6(±7)	2.7(±0.47)	20(±8)				
HD 276453	2.11(±0.11)	18(±4)						
HD 279943	3.05(±0.75)	75(±30)	−1.21(±1.48)	37(±30)				
HD 276449	−9.70(±1.42)	9.7(±10)	−3.47(±1.41)	13(±20)	2.09(±0.84)	28(±10)		
HD 276454	−5.47(±0.13)	34(±8)	2.83(±0.22)	18(±6)				
HD 279953	1.49(±0.18)	46(±9)	7.36(±0.36)	22(±6)				
HD 280006	−0.96(±1.49)	31(±20)	3.18(±0.50)	100(±30)				
HD 285016	−1.47(±0.12)	33(±7)	5.63(±0.16)	28(±6)				
HD 31856	0.77(±0.25)	18(±5)	7.99(±0.24)	19(±5)				
HD 31916	2.89(±0.50)	13(±6)	9.71(±0.18)	44(±9)				
HD 284938	−0.90(±0.14)	25(±6)	6.27(±0.13)	34(±6)				
HD 279890	−1.74(±0.11)	55(±7)	5.28(±0.20)	24(±6)				
HD 20326	4.2(±3)	6::						
HD 20456	4.3(±3)	4::						
BD+30 540	5.03(±0.03)	157::(±10)						
HD 22614		(<0.5)						
HD 278942	7.73(±0.02)	180(±10)						
HD 22780	1.9(±2)	2.6(±1)						
BD+30 564	4.6(±0.3)	76::(±9)	7.2(±3)	19::				
HD 24982	0.50(±0.06)	30(±20)						
HD 25694	5.88(±0.16)	36(±10)						
NGC 1514	6.14(±0.03)	84(±10)						
HD 26572	11.21(±0.04)	89(±10)						
HD 26873	5.60(±0.06)	16(±20)						
HD 27086	3.82(±0.18)	23(±20)						
HD 27448	−8.51(±0.97)	5.5(±7)	2.59(±2.40)	11(±10)	−2.68(±0.83)	31(±20)		
HD 27778	3.8(±0.5)	41(±1)	7.2(±0.5)	34(±1)				
HD 27923	3.6(±0.1)	39(±11)	8.5(±0.1)	26(±9)				
HD 284421	3.74(±0.14)	26(±30)	7.95(±0.19)	29(±20)				
HD 27982	2.74(±0.6)	192(±1180)	5.94(±5.2)	29.7(±182)				

Notes. Velocities of clouds V_1, V_2, V_3, V_4 are represented in the LSR (in units of km s^{-1}). Column densities N_1, N_2, N_3, N_4 have dimensions of E10 cm^{-2}. The measurement error is shown in brackets. Similarly to Table A.1, the second part of the table below the two horizontal lines corresponds to data which are published or could be found in other archives, for which no profile-fitting was done.

Table A.2. continued.

Star ID	V_1	N_1	V_2	N_2	V_3	N_3	V_4	N_4
HD 28482	4.9(±0.5)	46(±5)	8.8(±2)	13(±5)				
HD 37752	9.32(±0.12)	14(±8)						
2MASS J05164313+2307495	3.6(±0.50)	35::	8.2(±0.90)	21::				
TYC 1850-644-1	11.62(±0.11	196(±30)	4.61(±0.22)	57(±10)				
TYC 1861-830-1	5.7(±0.05)	370(±60)						
2MASS J05411140+2816123	−0.14(±0.31)	19(±5)	5.36(±0.13)	41(±7)				
TYC 2397-44-1	−7.5(±0.30)	70(±20)	−1.9(±0.32)	71(±20)				
2MASS J04282605+4129297	−12.81(±0.38)	27(±8)	−0.67(±0.23)	32(±50)	−7.39(±0.88)	8.8(±20)		
HD 281157	2.69(±0.14)	71(±30)	8.14(±0.03)	160(±10)				
HD 27405	7.43(±0.11)	59(±9)	2.96(±0.89)	22(±100)				
HD27404	6.58(±0.08)	73(±10)						
HD 27638		(<2.9)						
HD 28225	5.41(±0.08)	66(±10)						
HD 283677	6.58(±0.13)	63(±30)	2.52(±0.29	150(±40)				
HD 280026	3.9 (±1) from Na	K I<0.6						
2MASS J03232843+0944216	−6.9(±0.25)	20(±6)	4.8(±0.2)	37(±8)				
HD 32481	7.4	60						
HD 281305	6.4	95::						
HD 24263	8.5(±0.05)	83(±2)						
HD 30123	3.3(±1)	37(±5)	8.3(±1)	850::				
HD 37367	3.03(±0.04)	19.4(±3)	6.54(±0.03)	30(±3)				
HD 23466	9.9(±0.03)	14.9(±1)						
HD 23016	7.72(±0.01)	49(±3)						
HD 24398	6.23(±0.01)	160(±20)						
HD 22951	6(±1)	33::	1.4(±1)	1.6::				
HD 30675	4.4(±0.20)	30000::						
HD 32656	9.2(±0.1)	2400::						
HD 32811	7.07(±0.6)	13(±20)	2.87(±0.4)	46(±30)				
2MASS J03572138+1258168	8.6	120 (Na)						
HD 25204	8.9	8.9 (Na)						
HD 281159	8.1	660:: (Na)						
HD 23512	7.3	6.5						
HD 23408	8.0							
BD+19 441	6.0	11						
HD 19374	−2.9	17	8.8	10				
HD 21483	3.05	85						
HD 29647	3.0	45						
HD 23850	7.3	1.6						
HD 36371	−3.2	100::						
HD 35600	2.9	56						
HD 31327	2.83	100						
HD 24432	−8.40	162	2.80					
HD 24131	7.88	32						
HD 21856	−0.22	30						
HD 21803	2.64	30						

A&A 673, A56 (2023)
https://doi.org/10.1051/0004-6361/202245089

Astronomy & Astrophysics

Improved precision of radial velocity measurements after correction for telluric absorption⋆

A. Ivanova[1,2], R. Lallement[3], and J.-L. Bertaux[1,4]

[1] LATMOS, Université Versailles-Saint-Quentin, 11 Bd D'Alembert, 78280 Guyancourt, France
e-mail: anastasia.ivanova@latmos.ipsl.fr
[2] Space Research Institute (IKI), Russian Academy of Science, Profsoyusnaya street, 84/32, Moscow 117997, Russia
[3] GEPI, Observatoire de Paris, PSL University, CNRS, 5 place Jules Janssen, 92190 Meudon, France
e-mail: rosine.lallement@obspm.fr
[4] LATMOS, Sorbonne Université, 4 place Jussieu, 75005 Paris, France

Received 29 September 2022 / Accepted 5 February 2023

ABSTRACT

Context. The detection of planets around other stars by measurement of stellar radial velocity (RV) variations benefits from improvements to dedicated spectrographs that can lead to the achievement of a precision of 1 m s^{-1} or better. Spectral intervals within which stellar lines are contaminated by telluric lines are classically excluded from the RV processing.
Aims. We aim to estimate the potential improvement of the precision of radial velocity measurements from telluric-absorption removal and subsequent extension of the useful spectral domain.
Methods. We developed a correction method based on the online web service TAPAS that allows a synthetic atmospheric transmission spectrum to be determined for the time and location of observations. This method was applied to telluric H_2O and O_2 absorption removal from a series of 200 consecutive ESPRESSO exposures of the K2.5V star HD40307 available in ESO archives. We calculated the radial velocity using the standard cross-correlation function (CCF) method and Gaussian fit of the CCF, with uncorrected spectra and the ESPRESSO standard stellar binary mask on one hand, and telluric-corrected spectra and an augmented binary mask with 696 additional lines on the other.
Results. We find that the precision of radial velocity measurements is improved in the second case, with a reduction of the average formal error from 1.04 m s^{-1} down to 0.78 m s^{-1} in the case of these ESPRESSO data and this stellar type for the red arm. Using an estimator of the minimal error based on the photon noise limit applied to the full CCF, the error is reduced from 0.89 m s^{-1} down to 0.78 m s^{-1}. This corresponds to a significant decrease (by about 35%) in the observing time required to reach the same precision in the red part.

Key words. methods: data analysis – techniques: spectroscopic – planets and satellites: detection

1. Introduction

After the discovery of the first exoplanet (Mayor & Queloz 1995), the radial velocity (RV) method brought us another ≃2000 exoplanets[1]. Although the method of transit photometry is currently the most productive, the RV method is still important, because in addition to enabling the discovery of new exoplanets and the determination of their apparent mass $m \sin i$, it can be used to confirm exoplanets discovered with other methods and to measure the masses of transiting exoplanets ($\sin i \simeq 1$). However, the RV method is limited by technical capabilities: the size and availability of the telescope and the stability and accuracy of the wavelength calibration of the spectrograph. Therefore, any improvement of the processing associated to the radial velocity retrieval will result in improved precision or a decrease in the telescope time required to achieve a defined-goal precision (when this latter is limited by photon shot noise).

The RV method (Connes 1985; Baranne et al. 1996) is based on measuring the Doppler shifts of the stellar lines caused by the reflex motion of the star around the center of mass of the star-planet system. As spectra for RV measurements are taken from Earth, the absorption lines produced by Earth's atmosphere are also implanted on the star spectrum along with the stellar lines. Absorptions by Earth's atmosphere, or telluric absorptions, are unavoidable and produce a system of absorption features whose wavelengths are essentially fixed in the wavelength reference system of the spectrograph. On the contrary, even if the radial velocity of a star with respect to the barycenter of the Solar System is constant (case of no planet), the motion of the Earth (rotation and orbital) induces a change in the Doppler shift of the star as seen from the telescope, which is referred to as barycentric Earth radial velocity (BERV).

The classical method of measuring RV in spectra is to cross-correlate them with a template, or binary mask (BM; Baranne et al. 1979). Binary masks were initially mechanical (Baranne et al. 1979), with holes at the positions of stellar absorption lines, and required the refocusing of the whole spectrum light that passed through the mask onto a single photometric detector. Later on, masks were numerical, and in the form of a set of boxcar functions (e.g., Baranne et al. 1996). Subsequent improvements to the cross-correlation function (CCF) method consisted in assigning different weights to different lines of the BM (Pepe et al. 2002; Lafarga et al. 2020).

Most RV measurements excluded from consideration those spectral regions with, or adjacent to, telluric absorption. Figure 1

⋆ Based on data obtained from the ESO Science Archive Facility, Prog. ID:0102.D-0346(A).
[1] https://exoplanetarchive.ipac.caltech.edu

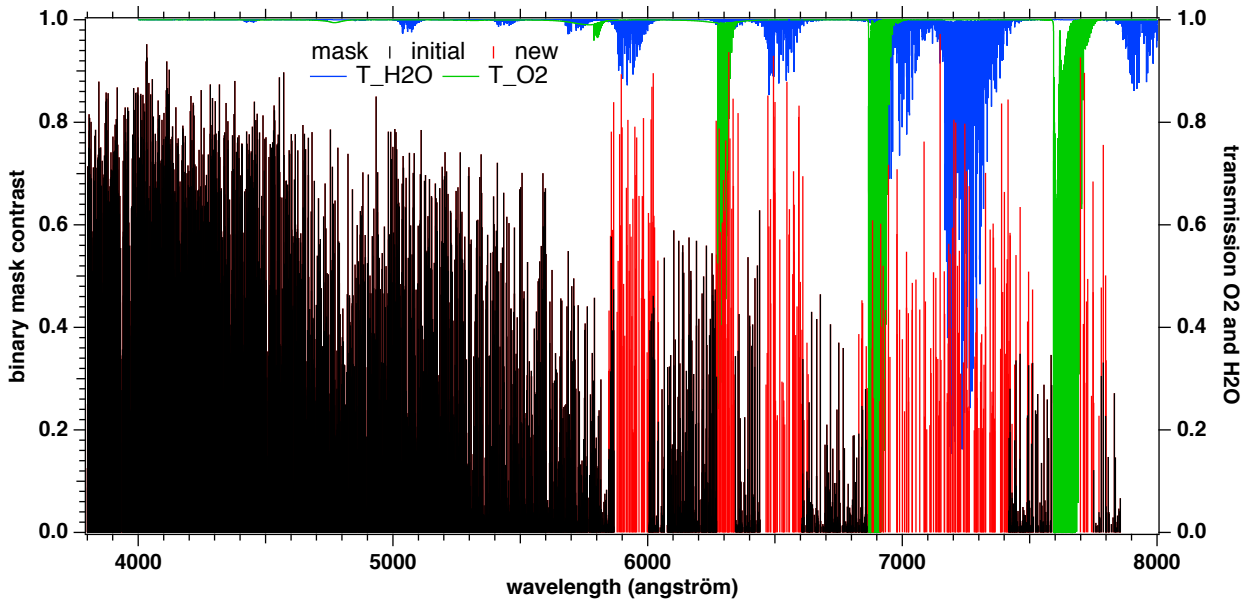


Fig. 1. Binary masks (left scale) and telluric lines (right scale). Black bars represent the contrast (relative depths) of the stellar lines of the standard K2 binary mask BM1 used in the ESPRESSO pipeline. Red bars show the contrast of the newly added lines forming the mask BMc. TAPAS telluric H_2O and O_2 transmission spectra are shown in blue and green, respectively.

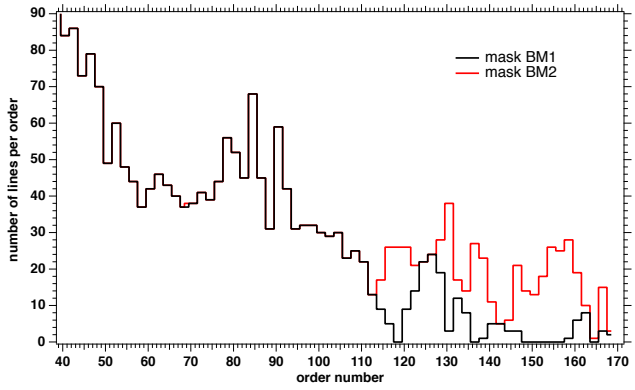


Fig. 2. Number of stellar lines per order from the standard BM1 (black) and the augmented binary mask BM2 = BM1 + BMc (red) restricted to the lines used in our study that have a contrast of ≥0.2. For the red arm, the new mask contains 724 more lines than the 1060 lines of the original mask (Table 2).

illustrates those wavelength intervals that are excluded from computation because of telluric contamination by restricting the official ESPRESSO BM lines to spectral regions clear of strong to moderate telluric lines. If we convert these regions into orders of the new, high-accuracy ESPRESSO spectrograph at ESO-VLT, 16 orders out of 170 are not used in RV calculation for this reason, while 30 others have a relatively small number of lines of the binary mask (Fig. 2). Our goal is to show that it is possible to correct the spectra for telluric absorption based on a highly detailed state-of-the-art synthetic atmospheric transmission spectrum available online and a rather simple fitting technique, and to use the previously discarded spectral regions for RV measurement. To check the correction and test its benefits, we took advantage of a quite unusual series of very short ESPRESSO exposures made in a single night.

Several methods have been developed to correct spectra for telluric absorption. We describe the different types of the existing methods and the differences between them in Sect. 2, where we also describe our method and how it fits into this typology. In Sect. 3, we describe the archive ESPRESSO data that we used for this study. Section 4 describes the steps toward the removal of telluric contamination from all exposures. In Sect. 5, we describe the classical CCF method, and explain how we constructed a new binary mask relevant to corrected parts of the spectrum. In Sect. 6, we describe our results when the CCF technique is applied to various parts of the spectrum, and show how we quantified the improvement in RV retrieval precision. In Appendix A, we describe the method used to estimate the uncertainties on RVs as a function of photon noise and spectrum quality, and how uncertainties derived in this way compare with our results. Finally, a comparison of radial velocity fluctuations seen with the blue arm and the red arm of ESPRESSO is presented in Sect. 7, followed by conclusions.

2. Overview of telluric correction existing techniques

One prominent technique is the division of each spectrum by the spectrum of a fast-rotating star (the so-called telluric standard) observed as close as possible in time and direction to the science target. This method has the disadvantage of requiring substantial additional observing time and the spectrum is affected by features present in the standard star. It is well adapted to multi-object spectrographs; in this case a number of fibers are devoted to the telluric target stars, but it is not well adapted to high-precision radial velocity measurements. A second, data-driven category of techniques builds on a series of auxiliary observations. Among these, Artigau et al. (2014) used a series of hot, fast-rotating stars to construct libraries of empirical telluric absorption spectra and performed a principal component analysis (PCA) of the dataset. These authors showed that the simultaneous adjustment of the first five principal components of an observed star spectrum allows them to obtain a good correction, to extend the spectral domain for RV extraction, and to increase the RV accuracy. More recently, Leet et al. (2019)

used a series of fast-rotator spectra observed in various different humidity conditions to empirically extract the whole series of water vapor lines, including the very weak, microtelluric absorptions. Bedell et al. (2019) developed a modern, sophisticated method of learning template spectra for stars and telluric absorptions directly from the data, a method adapted to a series of observations of cool stars covering sufficiently wide Doppler shifts between the stars and Earth. Recently, Cretignier et al. (2021) developed a PCA on a spectral time series, using prior information to disentangle contaminations from real Doppler shifts, that is, to eliminate both instrumental systematic errors and atmospheric contamination.

A third category of telluric decontamination relies on realistic atmospheric transmission spectra. Extremely detailed models of the global atmosphere and of time- and location-dependent atmospheric profiles are now available, the most detailed being the Global Data Assimilation System (GDAS) and the European Centre for Medium-Range Weather Forecasts (ECMWF). These give the altitude distribution of pressure, temperature, and H_2O concentration. Several tools have been constructed over recent decades to compute the resulting transmission spectra based on these profiles and the HITRAN molecular database (Rothman 2021; Gordon et al. 2017). The MolecFit facility developed by Smette et al. (2015) and the Telfit tool by Gullikson et al. (2014) make use of GDAS atmospheric profiles, while the TAPAS online service (see Bertaux et al. 2014) is extracting vertical profiles from the ECMWF. Both MolecFit and TAPAS use the state-of-the-art Line-By-Line Radiative Transfer Model (LBLRTM) code and the HITRAN database to compute the atmospheric transmission. Telfit and MolecFit also include software allowing users to correct spectral observations based on fits of the atmospheric transmission to the data, by adjusting the wavelength solution, a polynomial continuum and molecule abundances. To do so, MolecFit automatically retrieves the atmospheric profile at the time and place of the observations from an ESO repository. The advantage of this third category is the lack of a requirement for additional measurements and the adaptability of the methods to any instrument or observatory.

Very recently, Allart et al. (2022; hereafter A22) presented a fourth type of method for correcting stellar spectra from telluric absorptions. Similarly to the previous techniques, A22 use the HITRAN spectral data base to compute the transmittance of H_2O and O_2; however, instead of a worldwide time-variable grid of detailed atmospheric altitude profiles, the authors model the atmosphere as a single layer, characterized by a unique pressure p, unique temperature T, and a global H_2O column. The line profiles are less accurate, because the resulting atmospheric transmission does not take into account the different widths that correspond to different altitudes, contrary to the realistic atmosphere, but the advantage of this technique is the absence of a dependence on a meteorological field. A22 adjust p, T, and H_2O abundance by piling up 20 H_2O lines in a CCF.

Our work enters into the third category. Here, for the model of the Earth's atmosphere and the simulated transmission spectra, we used the TAPAS facility (Bertaux et al. 2014). TAPAS[2] is a free online web service that simulates with exquisite detail the atmospheric transmission spectra due to the main absorbing species as a function of date, observing site, and either zenith angle or target coordinates. TAPAS uses the ESPRI/AERIS ETHER (a French Atmospheric Chemistry Data Center) facility to obtain pressure, temperature, and constituent vertical profiles of the ECMWF, which are refreshed every 6 h. These are used to compute the atmospheric transmittance from the top of the atmosphere down to the observatory in each of 137 altitude layers based on the HITRAN database and LBLRTM (Clough & Iacono 1995; Clough et al. 2005). For each species and each altitude layer, the local temperature and pressure are used to compute air- and self-broadening as well as pressure line shift for each HITRAN transition. The basic mode gives access to spectra with wavelength grid steps on the order of 1 mÅ ($R \simeq 5 \times 10^6$), and the transmission spectra may be obtained separately for individual gases. TAPAS does not include correction tools and serves as a basis for line identifications or for various post-processing correction techniques. For example, Artigau et al. (2021) used TAPAS for a first data–model adjustment followed by a refined PCA-based correction for residuals. Bertaux et al. (2014) describe the TAPAS product, a synthetic atmospheric transmission spectrum for the time, place, and direction of any astronomical observation. These authors mention a potential application to the RV method of exoplanet searches, and a way to adapt the H_2O quantity of the model to a real observation is briefly outlined; this latter makes use of $T^X(H_2O)$, which is the model transmission $T(H_2O)$ elevated to the power X, which can be determined by comparison with the observed spectrum. In the present paper, we put the method sketched out in this latter paper into practice.

[2] https://tapas.aeris-data.fr/en/home/

3. Data

3.1. Brief description of the ESPRESSO spectrograph

The new-generation spectrograph ESPRESSO was chosen because of its high resolution, its excellent wavelength calibration, and, in particular, its impressive stability and the quality the data it provides. Indeed, ESPRESSO can achieve a precision on measurements of 30 cm s^{-1}, as reported in Pepe et al. (2021). These latter authors present a detailed description of the spectrograph, as well as an assessment of its performance. Installed by ESO at the Paranal observatory, the ESPRESSO spectrograph can be coupled to either one of the four VLT telescopes using optical fibers. It is an echelle-grating spectrograph with cross-dispersion installed in a vacuum chamber. It covers the wavelength range from 378.2 to 788.7 nm divided in two arms with two different detectors. We used the High Resolution mode, which provides a resolving power that varies between ≃100 000 and ≃160 000 depending on the position on the detectors. In this mode, an image slicer is used, which provides two images of the same echelle-grating physical diffraction order. All the useful parts of the spectra are numbered from 0 (start of the blue arm) to 169 (end of the red arm), and are called «orders» in the official nomenclature of ESPRESSO (used throughout this paper); not to be confused with true echelle-grating diffraction orders. In this nomenclature, one even-number order is followed by an odd-number order with the same wavelength coverage (due to the image slicer). The ESPRESSO processing pipeline has several versions of outputs. We selected the ES_S2BA-type data, that is, one spectrum per order, without flux calibration, not corrected for blaze function, with 9111 spectels in the red and 6541 in the blue. We used the wavelength assignment of each spectel in the laboratory system in which the telluric line absorption system should be fixed (except for the local wind). The blue arm contains orders 0 to 89 and covers the range 3781–5251 Å, while the red arm contains orders 90–169 and covers 5195–7908 Å. The intensity for each spectel is given in adu (analog to digital units). The electronic gain that was used for the archive data discussed here is ≃0.9 electron/adu.

3.2. Selected data

We used ESO VLT/ESPRESSO spectra of the K2.5V star HD 40307 ($V = 7.1$). The star has a mean radial velocity of $\simeq$31.3 km s^{-1} and at least four discovered exoplanets with periods from 4.3 to 197 days (Díaz et al. 2016). The target star HD 40307 was chosen for two main reasons: first, detailed measurements of radial velocities of this target with ESPRESSO were obtained as part of an asteroseismological campaign on the star[3]. The results of the full campaign are presented by Pepe et al. (2021).

Second, the ESO archive makes available all 1150 spectra of this star recorded over 5 nights. Here, we made use of the set of 200 consecutive exposures taken during one single night from 24 to 25 December 2018. The exposure time was set to 30 s for all data. The reason for the choice of such short and frequent observations by the proposing team was the attempted detection of stellar activity, namely *p*-mode oscillations. For the present study, this exceptional dataset of a very large number of exposures in a single night is ideal. First, the number of exposures is high enough to perform statistics on the RV measurements. Second, the total duration corresponds to $\simeq$240 m s^{-1} Doppler-shift variation of telluric lines with respect to stellar lines, which is a sufficiently large value to detect potential effects of the telluric correction on the absolute value of RV. Finally, the total duration of 4.3 h for the 200 selected exposures is too short to correspond to a detectable RV variation for the exoplanet given its period of 4.31 days, especially for the sequence of that particular night, which corresponds to a maximum-velocity region (data points around phase 130 in Fig. 25 of Pepe et al. 2021) and to an RV variation of less than a few tenths of a m s^{-1}. Because one criterion with which to verify the validity of our telluric correction is to check that the retrieved RV is constant over the period of observations, this is an important aspect.

We used the fully reduced spectra from the ESPRESSO pipeline. For each spectral order, the data are given separately for each of the two parts of the image slicer and we used these distinct spectra as if they were individual orders (e.g., orders 154–155 in the following sections correspond to a unique spectral order and the two images of the slicer). For RV measurements, all available spectra are shifted to the barycentric frame. For the purpose of telluric correction, we also used recorded spectra with a wavelength scale in the laboratory frame (see Sect. 4).

4. Correction of the data from telluric absorption

4.1. Principle of the correction

The main steps of the correction are as follows:

(1) First we construct a binary mask that eliminates stellar lines in order to let telluric lines only (hereafter referred to as the BMT). This mask is equal to 1 in spectral intervals free of stellar lines and zero elsewhere. It can be considered as a kind of complementary mask to the BM. This mask is constructed for the orders that are used in the next step.

(2) We then determine the atmospheric H_2O and O_2 columns present during each of the 200 exposures. These are obtained by fitting TAPAS synthetic atmospheric spectra in unmasked regions of specific appropriate orders. We checked that H_2O and O_2 are the two dominant gases in the wavelength range of ESPRESSO and that other atmospheric constituents produce negligible absorption.

(3) We then compute the product of the two H_2O and O_2 transmission spectra for each exposure and for all orders contaminated by tellurics using the previous results for the absorbing columns. The resulting transmission spectra are convolved by the appropriate order and wavelength-dependent ESPRESSO point spread function (PSF).

(4) Finally, we divide the contaminated data by the computed transmission spectra.

4.2. Preparation of the binary mask eliminating stellar lines

Telluric corrections are relatively easy in the case of hot stars. Their spectra are featureless and one can fit data directly to the product of a continuum and one or several atmospheric transmission spectra to adjust the columns of absorbing gases, the telluric line Doppler shift, and the PSF. The data are divided by this adjusted synthetic transmittance. For colder stars, this simple procedure is often not applicable because of the huge number of stellar lines. HD40307 is relatively cool with a temperature of $\simeq$4980 K. However, we took advantage of its low metallicity of $\simeq$–0.3 and the high resolution of ESPRESSO, resulting in lines narrow and weak enough to allow us to distinguish between tellurics and stellar features in some spectral regions where the former are separated from the latter, and to allow stellar continuum fitting, as in those regions the stellar continuum is well recovered between the lines. This means that, in such regions, it is possible to mask the stellar lines and fit the spectral regions free of stellar lines to atmospheric transmission model spectra.

The masks were constructed based on a semi-automated comparison between three spectra: a high-S/N spectrum of the star obtained by stacking ten consecutive exposures; a TAPAS transmission of atmospheric H_2O and O_2 adapted to the Paranal site and to the date of observations; and a stellar synthetic spectrum computed for a set of values of temperature, metallicity, and gravity adapted to those of HD 40307. This allows us to select areas that are devoid of non-negligible stellar lines and where telluric lines are dominant. Three masks were built for orders 130–131, 146–147, (for O_2 lines) and 154–155 (for H_2O lines), respectively. These can be seen in Figs. 3 and 4. The masked intervals are listed in Table 1.

4.3. Determination of H_2O or O_2 columns

Telluric lines vary during the night in response to airmass evolution along the line of sight to the star and also variations of atmospheric constituents. In particular, water vapor is highly variable and may change on timescales as short as a few minutes. Therefore, it is mandatory to measure the evolution of H_2O and O_2 during the series of measurements in order to obtain a good correction.

The individual exposures were divided into sections of about 2 nm and each section was fitted in the unmasked regions to a product of a second-order polynomial to represent the stellar continuum and a telluric transmittance convolved by a Gaussian PSF. The free parameters are the O_2 (for orders 130–131 and 146–147) or H_2O (for orders 154–155) columns and the coefficients of the second-order polynomial. The telluric transmittance varies as a power law of $T^X(H_2O)$, (respectively $T^X(O_2)$) where the factor X is proportional to the H_2O (respectively O_2) column (Bertaux et al. 2014). During the fitting procedure (which makes use of the Levenberg–Marquardt convergence scheme), the resolving power (and its associated Gaussian PSF) was fixed to a single value for the whole order estimated from Fig. 11 of Pepe et al. (2021). Examples of fits are shown in Figs. 3 (and 5)

[3] Prog. ID:0102.D-0346(A); PI: Bouchy.

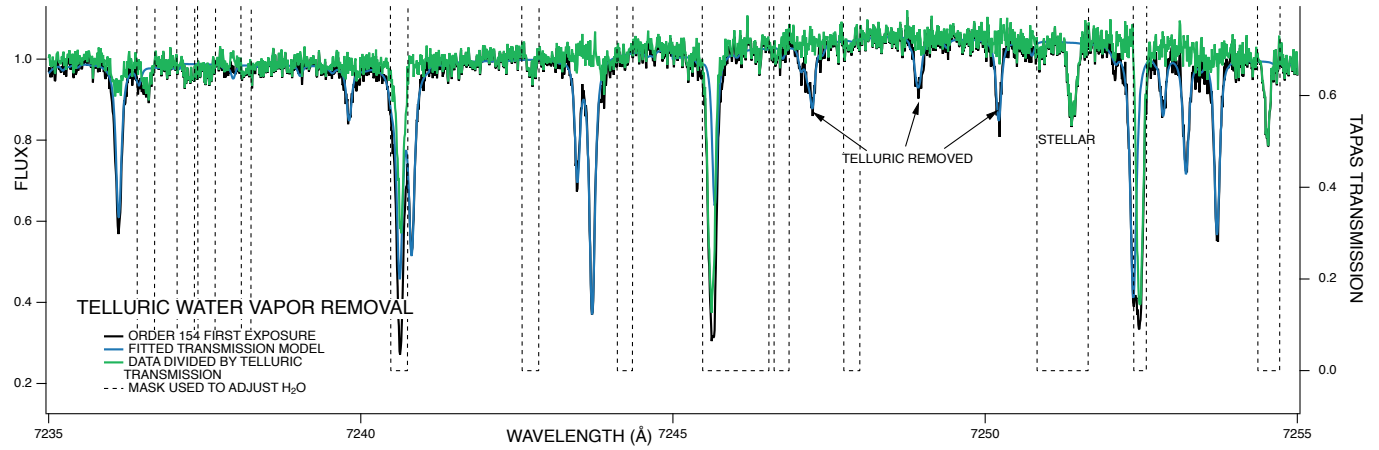


Fig. 3. Illustration of the procedure used to determine the H_2O column for each exposure (here exposure 0 and order 154) and the telluric H_2O removal. The mask of stellar lines, i.e., the BMT, is shown as a black dashed lines. The data and adjusted TAPAS transmittance are shown in black and light blue, respectively. The division of the data by the fit is shown in green. The correction works for isolated telluric lines and also for blended telluric and stellar lines (e.g., at 7240 A; see Fig. 5).

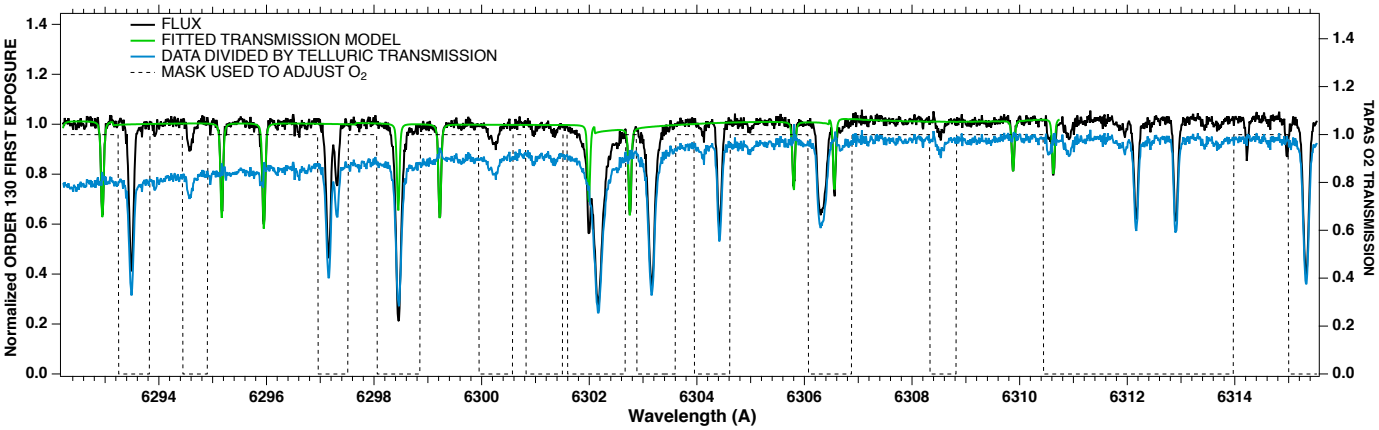


Fig. 4. Illustration of the procedure used to determine the O_2 column for each exposure (here exposure 0 and order 130). The BMT is shown as a black dashed line. The normalized data and the fitted TAPAS transmission are shown in black and green, respectively. The division of the data by the fit is shown in blue.

and 4 for the first exposure. The results for the various chunks are averaged along a full order to obtain H_2O or O_2 column temporal evolutions.

4.3.1. H_2O temporal evolution

The value of reference for the column chosen in the fitting procedure and the corresponding transmission spectrum were chosen as follows. We started with the spectrum predicted by the TAPAS website for the first exposure (25 December 2018, 00:24:00 GMT). The corresponding water vapor vertical column indicated by TAPAS COL(T) does not correspond to the column COL(P) estimated locally thanks to the infrared radiometer installed at Paranal, which continuously monitors the night sky and estimates this vertical column of H_2O in mm of precipitable water vapor. This information is included in the header of each exposure. This difference is not surprising given the strong temporal and local variability of the humidity and the fact that TAPAS interpolates temporally within ECMWF meteorological models refreshed every 6 h and spatially within nodes that may be distant with respect to the spatial scale of the variability. We used the two values to adapt the spectrum to the Paranal conditions by elevating it to the power of the ratio COL(P)/COL(T) and attributed a coefficient $X = 1$ to this reference spectrum.

The variation of the average fitted H_2O column for the 200 exposures is shown in Fig. 6. It is the product of the reference column COL(P) by the fitted coefficient X. It is compared with the vertical column estimated at Paranal. The comparison shows that the adjusted column of H_2O correlates well with the estimate from the infrared radiometer; in particular both have a similar amplitude of variation of about 15% and a similar time for minimum column. They also have very similar absolute levels, showing that the chosen reference was appropriate. More precisely, the fitted value for the first exposure is $X = 0.93$, which is not exactly 1 but is of the same order. During some time intervals, the values deduced from the spectral adjustment differ slightly from the on-site measurements. This is probably due to the sensitivity of the radiometric data to parameters other than the molecular column. In the subsequent correction for H_2O lines, we made use of the ESPRESSO spectral measurements that are the most direct.

4.3.2. O_2 temporal evolution

The retrieved O_2 column in the line of sight (LOS) is shown in Fig. 7. The chosen transmission spectrum of reference for the first exposure is exactly that predicted by the TAPAS website and was attributed a reference value $X = 1$ in the fitting procedure.

Table 1. Stellar mask for orders 130/131, 146/147, and 154/155.

Order 130		Order 146		Order 154	
Beginning	End	Beginning	End	Beginning	End
6266.54	6267.09	6829.64	6830.88	7150.50	7150.87
6267.54	6268.23	6831.31	6832.02	7151.57	7152.37
6270.10	6270.83	6833.66	6834.25	7152.76	7153.13
6271.45	6272.28	6834.71	6835.21	7153.25	7153.88
6272.64	6275.40	6838.34	6839.06	7153.90	7154.26
6275.65	6277.28	6840.10	6848.95	7155.06	7155.77
6278.95	6279.27	6850.08	6867.81	7156.12	7157.10
6281.28	6281.70	6872.62	6872.92	7157.16	7158.03
6282.00	6282.76	6873.10	6874.46	7158.45	7158.84
6283.31	6283.60	6875.97	6876.31	7159.57	7159.95
6284.03	6284.67	6877.07	6877.58	7160.24	7160.63
6286.67	6287.08	6878.03	6878.43	7161.93	7162.48
6287.55	6288.04	6881.27	6881.57	7162.58	7162.96
6289.86	6290.23	6882.01	6885.28	7163.37	7163.76
6292.05	6293.01	6887.38	6888.42	7164.10	7164.48
6294.31	6294.89	6896.07	6896.51	7164.95	7165.59
6295.50	6295.97	6899.70	6900.46	7165.97	7166.83
6298.02	6298.58	6902.67	6903.36	7166.87	7167.25
6299.12	6299.92	6904.46	6904.99	7167.32	7167.69
6301.01	6301.64	6907.02	6909.09	7168.73	7169.11
6301.88	6302.57	6909.99	6910.26	7170.87	7171.46
6302.66	6303.73	6910.87	6914.11	7171.88	7172.26
6303.94	6304.67	6914.92	6915.62	7172.27	7173.02
6305.01	6305.68	6915.88	6916.78	7175.50	7176.50
6307.13	6307.94	6917.59	6917.88	7176.59	7176.96
6309.39	6309.89	6918.16	6919.04	7177.70	7178.27
6311.50	6315.04	6919.69	6920.00	7178.70	7179.11
6316.06	6324.64	6920.33	6920.59	7180.54	7180.91
6326.72	6329.78	6921.04	6921.77	7181.54	7182.41
6330.06	6330.45	6921.99	6922.19	7182.98	7183.34
6331.54	6335.06	6923.05	6923.15	7183.74	7184.10
6336.50	6341.04	6924.15	6924.29	7185.59	7185.93
6345.60	6346.22	6925.49	6930.24	7186.69	7187.04
6351.05	6351.42			7188.80	7189.81

Notes. Wavelengths are in the stellar rest frame and in vacuum. For each order, the two limits of the masked region are listed.

The results of two series of fits are displayed, one for the order 130, which contains part of the O_2 γ band near 630 nm, and one for order 146, which contains part of the O_2 B band near 688 nm. Also shown is the evolution of the airmass for all exposures, as indicated in the headers. It can be seen that both fitted values follow the temporal evolution of the airmass very closely, which is expected. A small difference can also be seen between the two determinations. The γ band provides a more noisy determination but at the exact expected level (X=1 for the first exposure, which happens if the fitted column is exactly the reference value for the O_2 column of this exposure). The B band provides a less noisy determination, but with a small relative difference of about 3%. The reason for this small discrepancy is unclear. It is not likely to be a cross-section error in the HITRAN database. It might be due to small differences between the modeled and actual atmospheric profiles, or, more likely, to a lack of precision in the modeled PSF, in particular departures from the Gaussian approximation. In view of these results, we decided to use the B band result for the absolute value of X during the subsequent telluric-line removal, because the B band lines are stronger, and for the temporal evolution we chose to use quite simply the airmass as the multiplicative factor for the O_2 vertical column.

4.4. Correction for telluric lines based on the H_2O and O_2 adjusted columns

The next and final step of the correction procedure is, for the full wavelength range, the division of the data by a synthetic transmission computed for each exposure for the adjusted H_2O and O_2 values and adapted to the instrument resolution in an optimal way. The spectral resolution is not constant along an order and from one order to another, as is well known. Fortunately, the information on the resolution is available from Fig. 11 of Pepe et al. (2021). We digitized this figure and the associated color scale and derived the spectral resolution as a function of order and wavelength. The resolution for the three orders used for the O_2 and H_2O estimate are displayed in Fig. 8, showing variations with amplitudes of approximately 15% along an order. We used the results of this digitization to convolve the atmospheric transmission by the correct wavelength-dependent resolution,

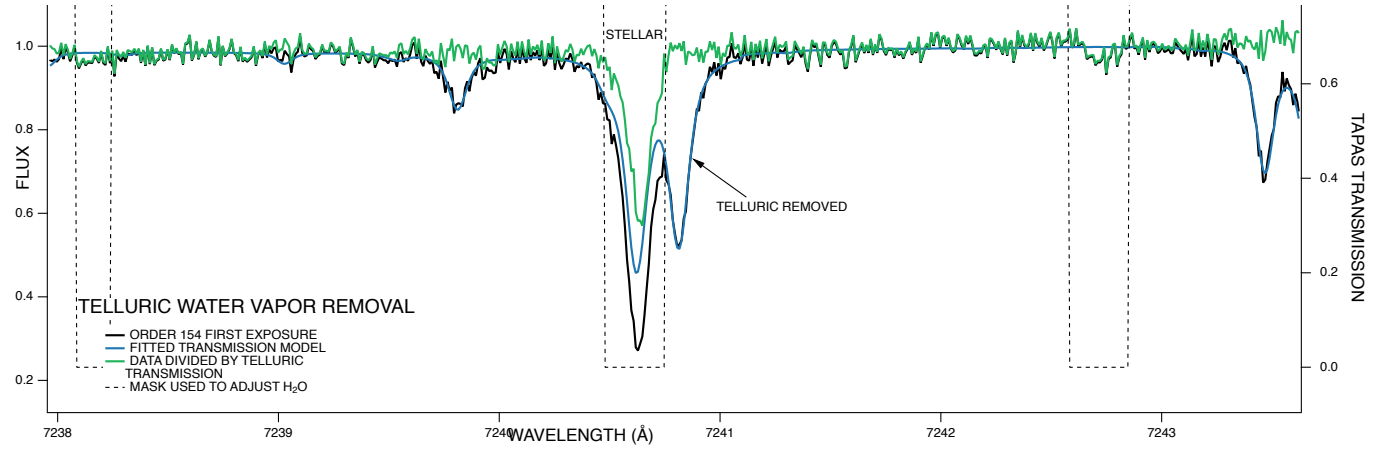


Fig. 5. Zoom onto Fig. 3. Even stellar lines heavily contaminated by H_2O lines may be kept for RV retrieval after correction from an adjusted telluric transmission. The BMT is shown as a black dashed lines.

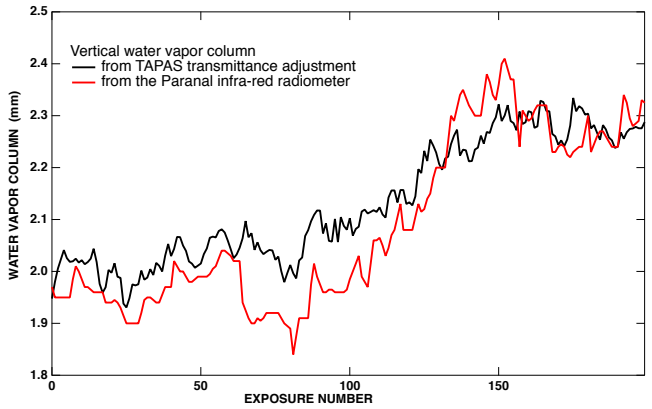


Fig. 6. Comparison between the vertical column of water vapor adjusted through spectral fitting (black curve) and the column estimated from the infrared monitor at Paranal (in red) for the 200 consecutive exposures.

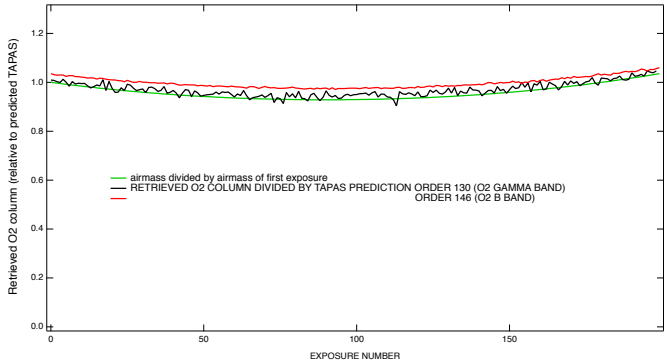


Fig. 7. Comparison between the column of O_2 adjusted through spectral fitting of the B band and the γ band (black and red curves) and the air mass (green curve). The retrieved columns of O_2 follow the air mass relatively well, which means that, unlike the H_2O variations, there is no significant variation of the O_2 vertical column.

that is, taking into account the true PSF for each order and each wavelength. The last operation, namely the division of the data by the convolved synthetic transmission, is not fully mathematically correct. In principle, the recorded spectrum should be de-convolved first, and then divided by the transmission, and the result of the division should be re-convolved (Bertaux et al. 2014). However, for a high resolving power, the first simplified approach is a good approximation. Examples of the divisions are shown in Figs. 3, 5, and 4 for H_2O and O_2.

For the whole correction, we assumed an exactly Gaussian PSF and only varied its width according to the wavelength and the order. We did this for the following reasons. First, in the case of hot stars with smooth continua, it is possible to deconvolve isolated telluric lines to infer the PSF shape; however, here in the case of this K star, the result of the deconvolution would be somewhat hazardous. Second, there are no indications of departures from a Gaussian shape in the ESPRESSO documents and, to our knowledge, no mention of departures in articles based on ESPRESSO data. Third, and most important, significant departures would introduce features at the locations of the wings of the corrected lines that would repeat along all lines, and be especially strong for a deep absorption; we did not notice any specific, repetitive shape in the residuals in the wings that would indicate such departures. On the contrary, the residuals, when present, do vary from one line to another.

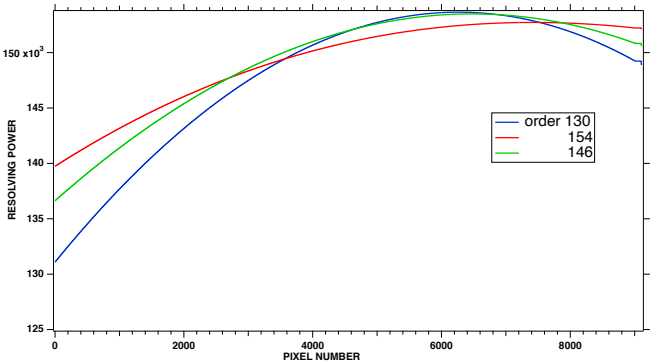


Fig. 8. Spectral resolution as a function of pixel number based on the digitization of Fig. 11 of Pepe et al. (2021). Shown are three examples for three orders of the red arm.

5. Standard and extended binary masks to be used in the classical cross-correlation function method

In order to quantify the quality gain obtained by correcting the spectra from telluric absorptions, we need to compare the retrieved RVs with exactly the same algorithm, applied either

Table 2. Number of lines in the binary mask.

Binary Mask	BM1	BMc	BM2 = BM1 + BMc
All orders 40–169	3654	725	4379
Blue arm 40–89	2594	1	2595
Red arm 90–169	1060	724	1784
Red arm 90–163	1050	696	1746

to the noncorrected data (i.e., the wavelength regions not contaminated by tellurics), and to the corrected data (including wavelength regions corrected for tellurics). There are several types of RV-retrieval algorithm, but for our purpose we do not need to use the absolute best one, as long as we use exactly the same one for the two cases.

The historical and classical approach of RV retrieval is the CCF with a binary mask (Baranne et al. 1996; Pepe et al. 2002), which is used in the official ESPRESSO pipeline for the official RV retrieval. This approach has the advantage of being robust. One of the recent papers describing this method in detail is Lafarga et al. (2020), from which we borrowed some details when coding our own version of this algorithm.

For processing the non-corrected data, we used the standard binary mask of ESPRESSO corresponding to the spectral type K2.5 of HD40307. This mask is made of a series of wavelengths in air and the relative depth of the star line (called the contrast), which is represented in Fig. 1 as a function of wavelength. When compared to the model transmission computed by TAPAS (Fig. 1), it is clear that this standard binary mask is avoiding regions seriously contaminated by atmospheric H_2O and O_2 lines. We selected all the lines of the standard mask with a contrast of >0.2, which constituted the BM1 mask that was used to process the noncorrected data. A second mask, BM2, was built by adding a series of new lines to the BM1 mask, specifically in regions contaminated by telluric absorption, constituting the mask BMc. There are many orders where there were no lines in BM1: 118, 136, 148, 150, 152, 154, 156, and 164 (and their odd number twins). In order to find the new lines of BM2 = BM1 + BMc, we stacked together the 200 spectra after telluric correction, and to this stacked spectrum we assigned the wavelengths of exposure number 100 in the reference frame of ESPRESSO. Then, for a set of observed stellar spectral features, we determined the precise wavelength (using a Gaussian fit) of the minimum and kept those lines that had a contrast of >0.2. In Fig. 2, we plot the number of lines per order for the two masks BM1 and BM2 as a function of order number. For some orders (116–123, 128–148, 158–159), the number of lines from BM1 was not zero but small; the lines were kept for BMc but their wavelengths were also taken from the stacked spectrum (after telluric correction). Table 2 lists the numbers of the initial lines (BM1), the additional selected lines (BMc), and the total lines of the binary mask. There is only one more line for the blue arm, and 724 more lines for the red arm, added to the 1060 lines of the original mask BM1. For the analysis of the time series of measurements, we dropped the last six orders (A band of O_2 and weak signal); the selected orders for the red arm contain 696 lines more than the 1050 lines of BM1. The gain in RV measurement precision is entirely due to the use of those 696 extra lines of the BMc mask. Table 3 lists the BMc lines for order 130. The full list of additional lines will be made available on request.

Figure 9 shows the order 130 of the stacked spectrum, as well as the TAPAS model transmission and the position of the lines represented by vertical bars of different colors for BM1 and BMc. This order is contaminated by the γ band of di-oxygen O_2, and as a result there are only three lines in this order in the BM1 mask (position indicated by vertical red lines at the bottom). We found 33 more lines for the BMc lines (vertical blue lines), giving a total of 36 lines for this order.

Figure 10 is a zoom onto Fig. 9 around a particular line at 6344.8 Å, which was already in BM1 and whose position is indicated by the black vertical line at the bottom of the figure. The blue line is the contrast (right scale) of the spectral line and is centered on the observed spectral line, which is displaced by a Doppler shift from the BM1 line. In order to allow it to serve as a line in a binary mask, we need to shift this new line back to a reference system with RV=0. Therefore, to carry out this back-shift we have to precisely select the RV of the star HD 40307. We chose $Vrad_0 = 31381.572\,m\,s^{-1}$, as determined with the mask BM1 with orders 92–115. When this line is back-shifted, it comes to the position of the red vertical bar, very near

Table 3. Binary mask lines for order 130/131.

Wavelength in vacuum	Contrast	FWHM m s^{-1}
6266.8465	0.8	4486.20
6268.0325	0.3	2602.77
6270.5443	0.8	12388.95
6271.9383	0.5	2677.01
6272.9907	0.3	2519.84
6276.3673	0.5	3034.65
6281.4607	0.3	3953.09
6282.3338	0.8	3967.68
6284.3536	0.6	12500.92
6286.8781	0.6	2914.92
6292.6845	0.4	2634.84
6294.5440	0.6	3009.47
6298.2113	0.5	2847.12
6298.3656	0.2	1939.91
6299.5157	0.7	3613.13
6303.2231	0.6	3603.95
6304.2168	0.6	3534.69
6305.4806	0.4	2433.84
6307.3757	0.8	17150.33
6313.2248	0.4	2445.00
6313.9634	0.4	2383.63
6316.3841	0.7	4334.12
6317.0346	0.5	2684.57
6317.5363	0.4	2565.93
6319.7486	0.9	10887.55
6320.4338	0.2	2874.01
6320.9591	0.2	3435.50
6324.4158	0.8	4638.91
6329.3274	0.4	2685.68
6331.8214	0.6	2797.14
6332.5758	0.2	2317.80
6337.0623	0.8	5807.31
6337.8318	0.3	2552.29
6338.5565	0.7	4091.04
6340.6073	0.3	2541.33
6340.8357	0.3	2592.87
6345.8836	0.6	3119.11
6356.7665	0.8	5692.01

Notes. Wavelengths are in vacuum.

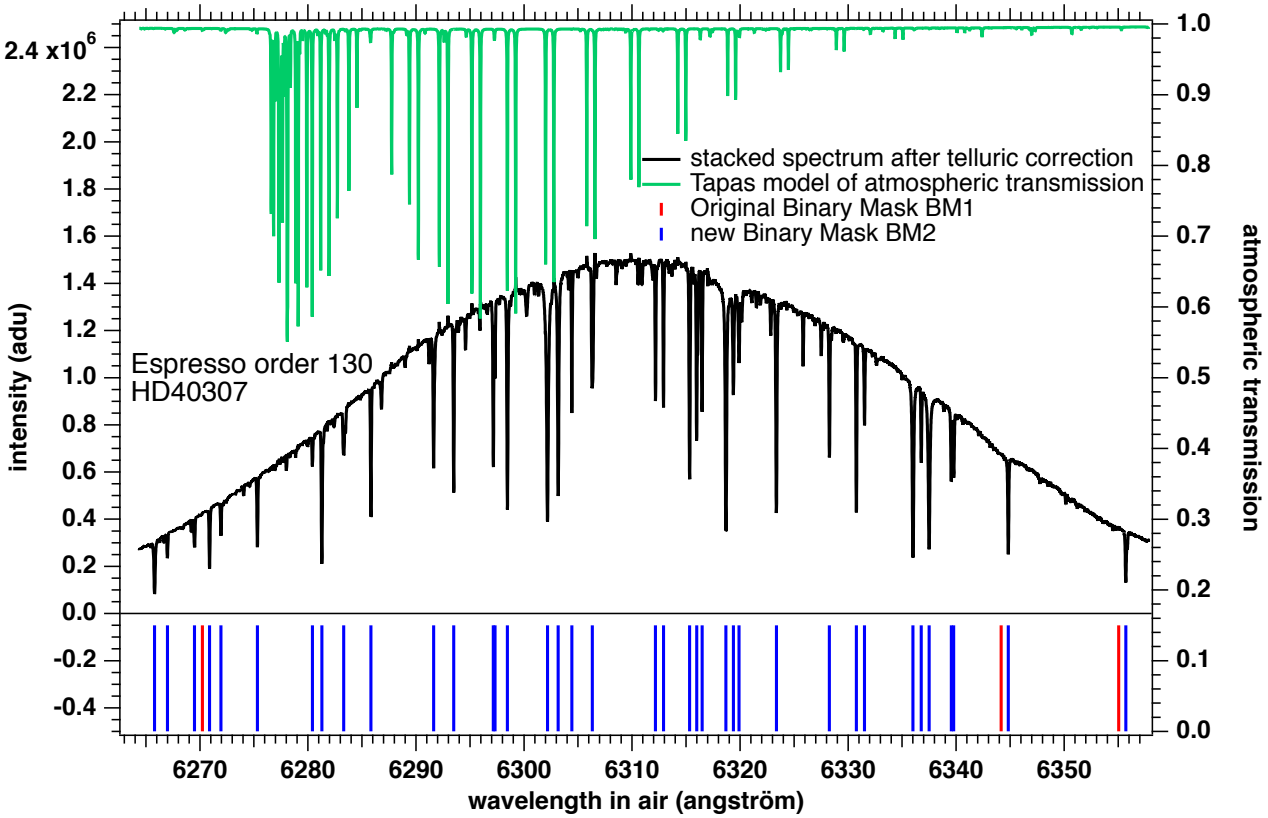


Fig. 9. Augmented mask of order 130. The intensity of the stacked spectrum (high S/N) after correction of telluric absorption is shown in black (left scale). The atmospheric transmission computed from TAPAS (green, right scale) displays the γ band of O_2. The original binary mask BM1 contains only three lines, whose wavelengths are indicated at the bottom by red vertical bars. These lines avoid the O_2-contaminated region. A new Binary Mask BM2 containing 36 lines is made from the observed stacked spectrum with stellar lines whose depths are ≥0.2. The wavelength assigned to the stacked spectrum is that of exposure number 100, in the middle of the series of exposures. The three red lines of BM1 are seen shifted to the left from their blue-line counterpart detected in the stacked spectrum.

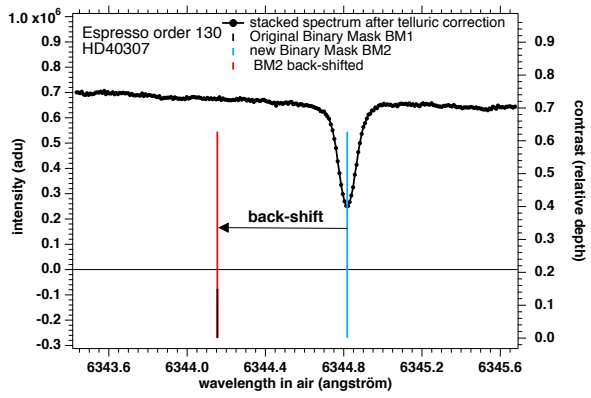


Fig. 10. Adjustment of the two masks. The vertical blue line gives the wavelength position of the center of one stellar line of the observed spectrum, after telluric correction and stacking of all exposures. Its height is the contrast of the line (relative depth, 0.64). The red line represent the same line, back-shifted to account for the radial velocity $Vrad_0$ of the star. The velocity $Vrad_0$ was determined with the original mask BM1 and data from orders 92–115. As a result, the position of the red line coincides well (but not perfectly) with the corresponding line of the original BM1 mask (vertical bar, bottom).

the original BM1 line. All new lines from the telluric-corrected stacked spectrum were back-shifted by the same stretch factor (the Doppler shift is actually a stretch) to constitute the mask BMc, and this was added to BM1 to form BM2.

In Fig. 11, we show the extreme case of the order 146, which is heavily contaminated by the O_2 B band. The correction is poor at the location of the strongest absorption lines, and strong residual spikes are visible. A better correction would require a detailed adjustment of the exact shape of the PSF, which is beyond the scope of this work. This order is also characterized by a low number of stellar lines. We could also simply add 18 lines to the binary mask. To do so, we avoided the regions with poor correction.

In the CCF scheme adopted here, and described by Lafarga et al. (2020), each order is treated separately. The CCFs of all the lines of one order are added together with a weight attached to each line. The total CCF is then fitted by a Gaussian whose central position (the minimum) is the radial velocity for this order. The uncertainty σ_{ord} attached to this determination is that returned by the Gaussian fit routine. Once the RVs for each order are determined, they are averaged together, accounting in a standard way for their uncertainties: the weighted average with weights $= 1/\sigma^2_{ord}$, and the error for the exposure σ_{exp}, such that $1/\sigma^2_{exp} = \Sigma\ 1/\sigma^2_{ord}$.

While Lafarga et al. (2020) took the ratio contrast/width as a weight attached to each line of a BM, we took only the contrast of each line as its weight for its CCF. We did this because the mask BM1 contains only the contrast and not the width. As we wanted to use exactly the same algorithm for comparing the RV results with BM1 and BM2, we were forced to use only the contrast for the two cases. We carried out a limited comparison with the two versions of the weight: either the contrast only,

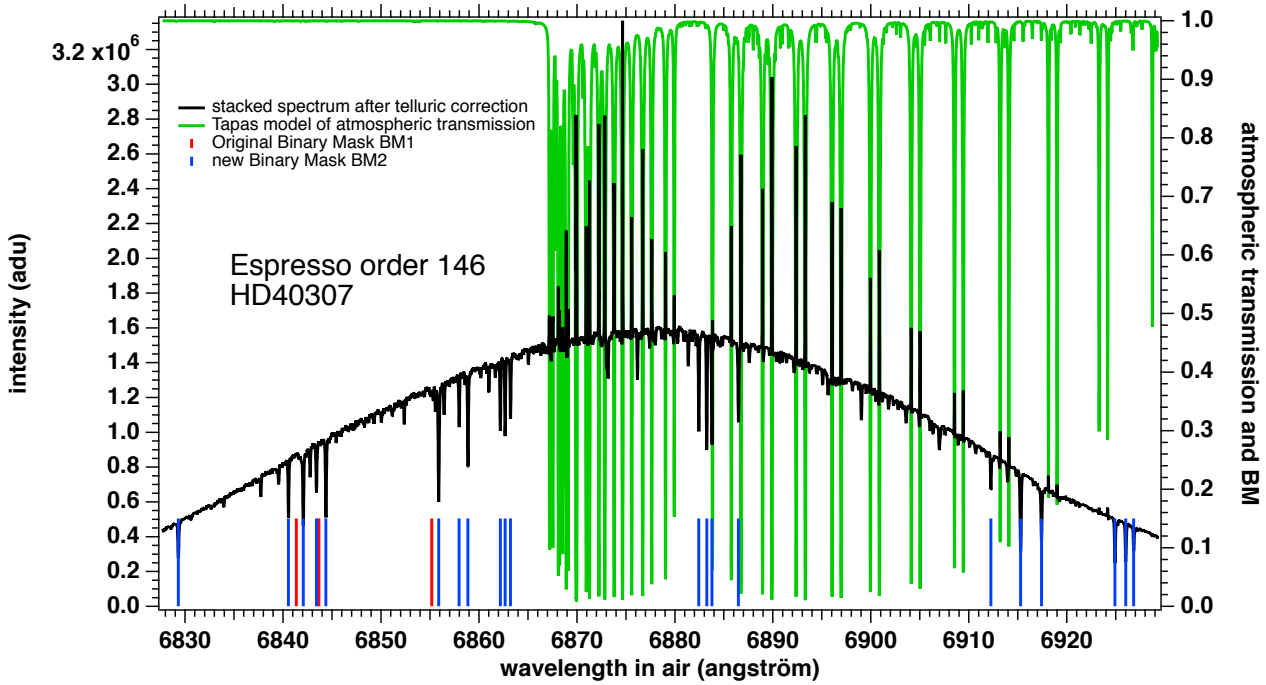


Fig. 11. Same as Fig. 9 for order 146, which is heavily contaminated by deep, narrow lines of the O_2 B band. The corrected spectrum exhibits conspicuous residuals (see text) at the location of the strongest absorption; however we were able to add 18 lines in the new mask in addition to the initial three lines.

or the contrast/width. We find a very modest decrease (2.9%) in the uncertainty in the latter case and a ≃3% decrease for the spread of all RV measurements (standard deviation) among the series of 200 exposures. Therefore, while clearly it would be slightly better to use the contrast/width weight, using only the contrast will not be detrimental for the comparison between the use of BM1 applied to original spectra and BM2 applied to telluric-corrected spectra.

The CCF of one line of the BM with one observed line is done by computing the signal contained in a boxcar centered on the BM line, and displaced by a variable RV (by steps of 492 m s^{-1}, the mean size of one spectel) around the observed minimum. Following the recommendation of Lafarga et al. (2020), we chose to use a boxcar width of approximately one spectel of the observed spectrum. As a result, the CCF is very similar to the observed line, because both the boxcar width and the step of the RV grid are of the size of one spectel. This method also has the advantage of providing independent points of the CCF. Indeed, the Gaussian fit algorithm used in Igor language (but this is the same for most coding languages used to code the Numerical Recipe library, Press et al. 1992) assumes that all points to be fitted are independent, and the error bars returned by the fit routine rely on this hypothesis. This would not be the case if the boxcar width was larger than one pixel, or if the RV grid step was smaller than one spectel. We find that one important parameter of the Gaussian fit to the total CCF for one order is the number of CCF steps over which the CCF is fitted, because we find that the exact position of the CCF minimum depends substantially on this number; it was fixed to ten points in all the presented results.

The blaze effect is relatively conspicuous in Fig. 9. We did not attempt to correct the blaze before performing the CCF. When fitting a tilted (from constant) signal, any blaze introduces a bias on the position of the minimum. However, we determined that with the observed slopes due to the ESPRESSO blaze, the bias does not exceed about 12 m s^{-1}. This systematic bias is constant along all exposures for one particular line and does not increase the spread of all measurements, nor the (shot noise associated) uncertainty on RV. Furthermore, being of opposite sign on the two sides of the order, the two biases approximately compensate for one another when performing the CCF for a full order. Also, we find out that for some lines of the BM1 mask (probably those determined from laboratory measurements of transitions), the bias is much larger – of the order of ≃100 m s^{-1} – due to convection and granulation, as pointed out by González Hernández et al. (2020), which is therefore much larger than the biasing blaze effect.

There are several adjustable parameters in this CCF algorithm (as described above), and therefore, for each exposure, we do not expect to find exactly the same RV value as the one contained in the archive product of HD 40307 (for which the exact values of the parameters are different from ours and are unknown to us). However, this is not our objective. Our purpose is to compare the RV series results obtained with BM1 and BM2 with exactly the same algorithm and its parameters. Also, we have not used all the lines of the original mask, keeping only those with a contrast larger than 0.2.

As suggested by our reviewer, we also computed another estimator of the error on RV based on a quality factor and photon noise as first devised by Connes (1985) and revived by Bouchy et al. (2001). To do so, we used the approach from Boisse et al. (2010), in which the authors consider a CCF for one order as an averaged noise-free spectral line. This provides an estimate of the minimal error. We detail this approach and present our calculations in Appendix A.

6. Results

In this section, we show a series of comparisons between two separate applications of the CCF code. On one hand, the code

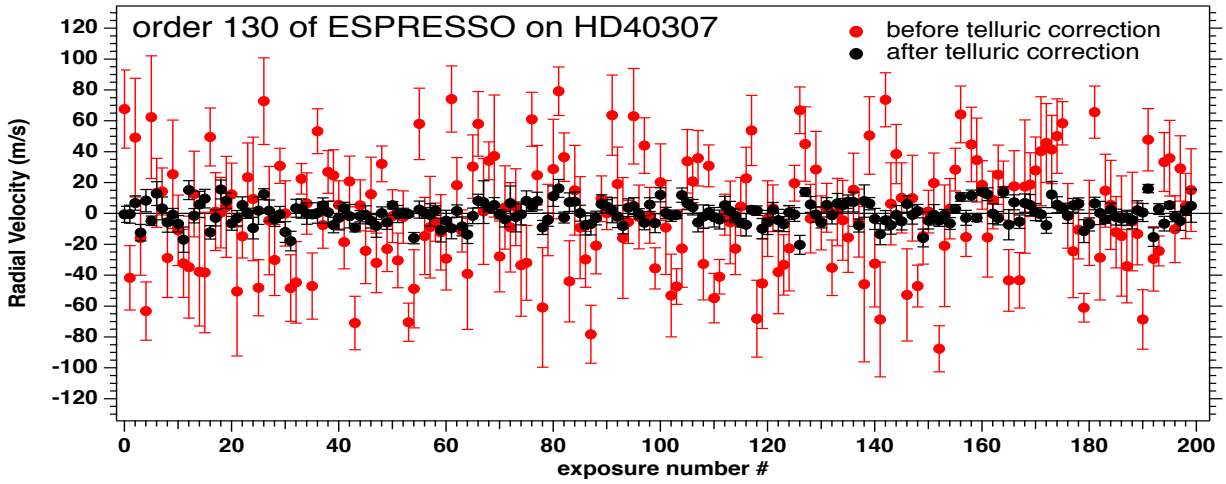


Fig. 12. Radial velocity of star HD40307 for order 130, which is contaminated by O_2. The result with the standard binary mask BM1 before telluric correction is shown in red; the result with our new binary mask BM2 after telluric correction is shown in black. The average radial velocities of 31 372.4 and 31 384.2 m s^{-1} were subtracted from the series before and after correction, respectively.

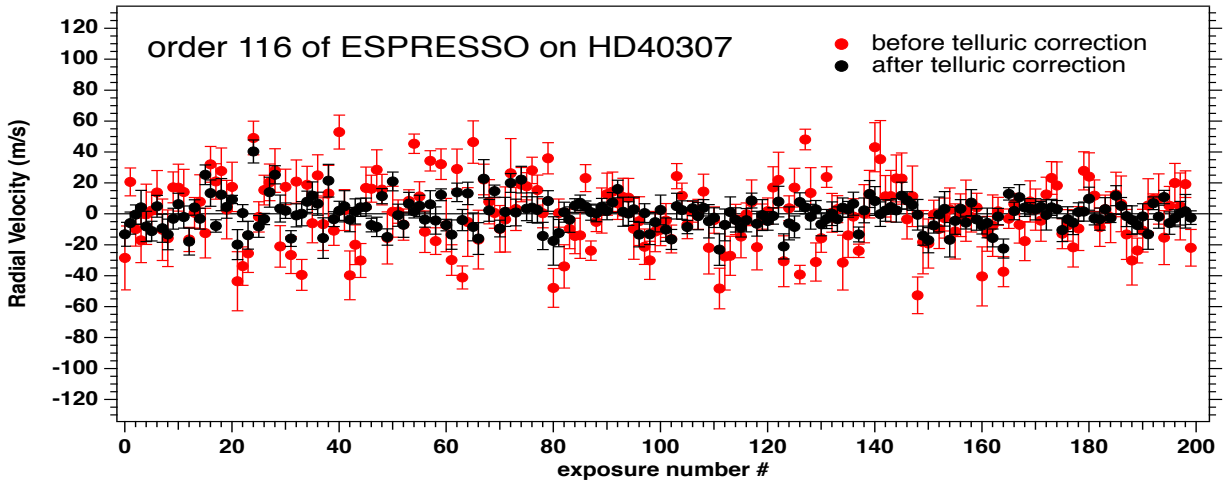


Fig. 13. Radial velocity of star HD40307 for order 116, which is contaminated by H_2O. The result with the standard binary mask BM1 before telluric correction is shown in red; the result with our new binary mask BM2 after telluric correction is shown in black. The average radial velocities of 31 386.4 and 31 388 m s^{-1} were subtracted from the series before and after correction, respectively.

is applied to the uncorrected data on the basis of the standard binary mask BM1. On the other hand, it is applied to corrected data on the basis of the combination BM2 of the standard mask and the newly created mask. We show three different cases: relatively straightforward H_2O correction (116–117), less straightforward O_2 correction (γ band, orders 130–131), and the more difficult case of the O_2 B band (orders 146–147).

6.1. Order 130 contaminated by O_2

The order 130 ($\lambda \simeq 6316$ Å) and its twin order 131 are contaminated by the γ band of O_2 (Fig. 9). Only three lines are contained in the official BM1 of ESPRESSO, while our new mask BM2, as described above, contains 38 lines for this order.

Figure 12 compares the RV measurements (for the whole series of 200 exposures) coming from our CCF algorithm before correction and after correction of tellurics. The average (over time) of RV has been subtracted from the whole series in order to ease the comparison of the dispersion. In principle, we should find a constant signal over the ≃4 h of measurements. We see by eye that the signal is approximately constant on average. Both the error bars and the spread of measurements (fluctuations, quantified by the standard deviation $S_{\rm dev}$) are very much decreased with the telluric-corrected data. Quantitatively, the average error bar of this order over the 200 measurements is reduced from 22.5 to 5.3 m s^{-1} for the new mask, while the spread of RV measurements $S_{\rm dev}$ is reduced from 35.4 to 7.0 m s^{-1}. This is of course a somewhat extreme example of what can be gained by the addition of new portions of star spectra once they are corrected for tellurics. Independently of this improvement, the constancy of the RV value over the exposures constitutes an additional test of the telluric correction.

6.2. Order 116 contaminated by H_2O

Similarly, the order 116 ($\lambda \simeq 5891$ Å) and its twin order 117 are contaminated by the blue end of a region contaminated by H_2O lines (Fig. 1). BM1 contains only 5 lines, while BM2 contains 20 lines for this order.

Figure 13 compares the RV measurements before correction and after correction of tellurics. Again, the time-averaged RV has been subtracted. The average error bar of this order is reduced

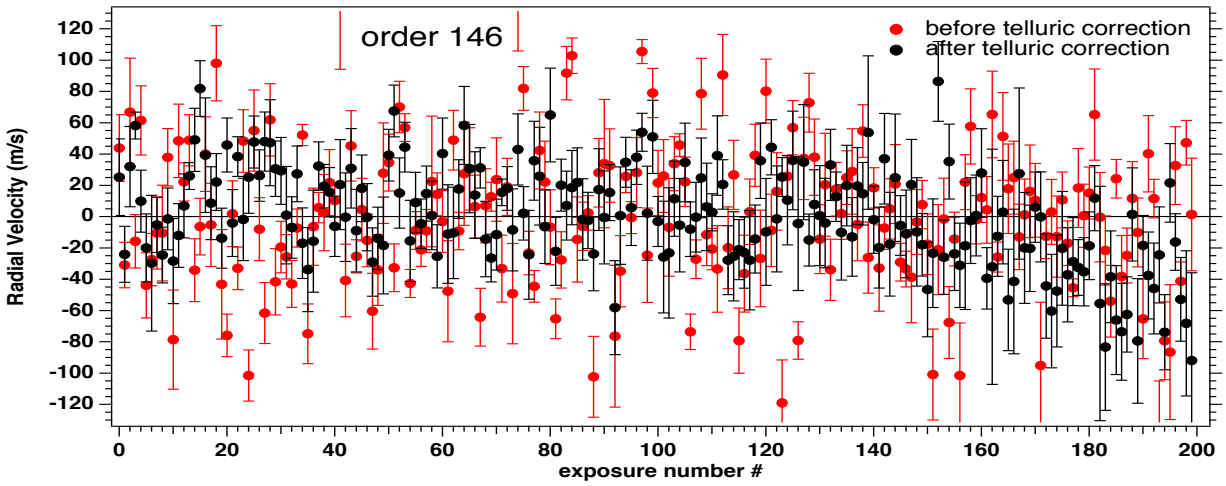


Fig. 14. Radial velocity of star HD40307 for the extreme case of order 146, which is heavily contaminated by O_2. The result with the standard binary mask BM1 before telluric correction is shown in red; the result with our new binary mask BM2 after telluric correction is shown in black. The average radial velocities of 31 386.4 and 31 388 m s^{-1} were subtracted from the series before and after correction, respectively.

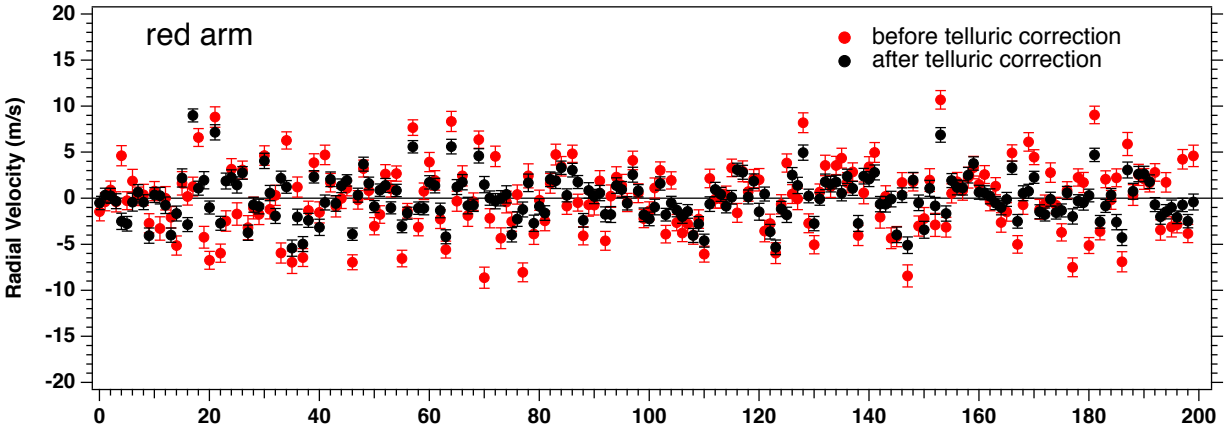


Fig. 15. Radial velocity of star HD 40307 for orders from 90 to 163. The result with standard binary mask BM1 before telluric correction is shown in red; the result with our new binary mask BM2 after telluric correction is shown in black. The average radial velocities of 31 382.5 and 31 384.4 m s^{-1} were subtracted from the series before and after correction, respectively.

from 12.8 to 7.1 m s^{-1} for the new mask, while the spread of RV measurements $S_{\rm dev}$ is reduced from 20.5 to 9.6 m s^{-1}. The vertical scale of RV covers the same extent of 260 m s^{-1}.

Figure 14 is identical to Fig. 12, but for order 146, which is heavily contaminated by the strong O_2 B band (see Fig. 11). We include this order as an extreme case because a number of telluric lines are saturated before convolution by the PSF and there are strong residuals. As mentioned above, a better correction would require a specific determination of the exact shape of the instrumental PSF, which is beyond the scope of this work. The standard binary mask contains only three stellar lines, while our binary mask created after telluric correction contains 18 stellar lines. Here, the spread decreased from 47 m s^{-1} to 32 m s^{-1}. However, the formal error increased from the rather high value of 21 to 25 m s^{-1}, a somewhat puzzling result. We note a similar behavior for order 144, which is not contaminated by telluric absorption.

6.3. Red arm

Figure 15 compares the series of RV measurements when the orders of the red arm are combined, with a weight of 1/err^2 where err is the formal error returned for each order. We note that the red arm contains the orders 90–169, while we kept only orders 90 to 163, dropping the four orders 164–167 which are heavily contaminated by the strongest A band from O_2, and contain saturated lines. Moreover, the last orders 168–169 are characterized by a low signal. The vertical scale of RV covers a 40 m s^{-1} full extent.

The average error bar of this red arm is reduced from 1.04 to 0.78 m s^{-1} for the new mask BM2 (Table 4). If we use the photon-noise estimation on the CCF (see Appendix A), the error bar decreases from 0.89 to 0.72 m s^{-1}, a reduction factor of 1.24. As discussed in the Appendix, the difference is due to our choice of using ten points of the CCF instead of the CCF in its entirety. Therefore, it is likely that the photon-noise estimator is more robust than the ten-point Gaussian fit. For both cases, this gain is entirely due to additional photons from the 696 new stellar lines from BMc included in the spectral regions made available by the telluric correction. The spread of RV measurements $S_{\rm dev}$ is reduced from 2.83 to 2.37 m s^{-1} (see Table 4). The reduction of the spread, by a factor of 1.19, is slightly smaller than the reduction of the formal error, which is reduced by a factor of 1.33.

Table 4. CCF statistical results.

Used exposures	0–199 No correction	0–199 With correction	0–99 With correction	100–199 With correction
Formal error err Red arm	1.04	0.78	0.79	0.78
Formal error err Blue arm	1.15	1.15	1.21	1.10
Formal error err Red + Blue arms	0.77	0.64	0.65	0.63
Standard deviation $S_{\rm dev}$ Red arm	2.83	2.37	2.50	2.23
Standard deviation $S_{\rm dev}$ Blue arm	5.78	5.56	4.13	6.71
Standard deviation $S_{\rm dev}$ Red + blue arms	3.53	2.59	3.03	3.97

Notes. Units are m s^{-1}.

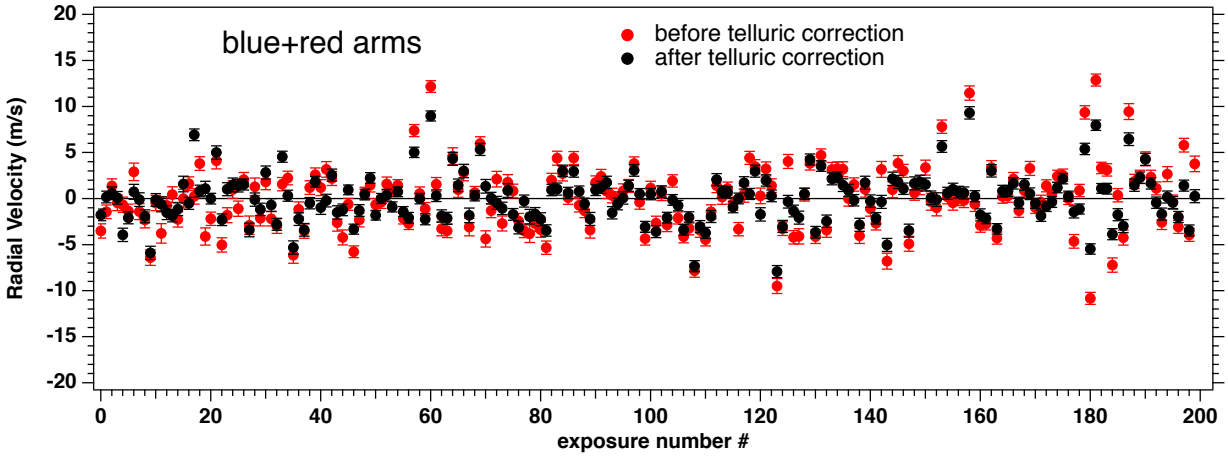


Fig. 16. Radial velocity of star HD 40307 for orders from 40 to 163. The result with standard binary mask BM1 before telluric correction is shown in red; the result with our new binary mask BM2 after telluric correction is shown in black. The average radial velocities of 31 386.9 and 31 385.9 m s^{-1} were subtracted from the series before and after correction, respectively.

6.4. Red+Blue arms

Finally, we performed the same comparison as above for a combination of red and blue arm data. More precisely, we combined data for orders 40 to 163 coming from both arms. Orders below 40 were dropped because of a very low signal for this K2.5 star. The comparison is shown in Fig. 16. The average error bar of blue+red arms is reduced from 0.77 to 0.64 m s^{-1} for the new mask BM2, while the spread of RV measurements $S_{\rm dev}$ is reduced from 3.53 to 2.59 m s^{-1} (see Table 4). The vertical scale of RV again covers a 40 m s^{-1} full extent. The reduction of the spread, by a factor of 1.36, is slightly larger than the reduction of the formal error, which is reduced by a factor of 1.20.

6.5. Absence of an RV trend in the RV time series

Figure 17 compares four time series of RV measurements: (1) The original RV measurements found in the archive (similar to Fig. 21 of Pepe et al. 2021). (2) The RV measurements from our processing after telluric correction restricted to blue orders (from 40 to 89, λ from 4319 to 5206 Å); this determination is almost identical to our processing of data not corrected for tellurics (not shown here). (3) Our RV measurements after telluric correction restricted to red orders (from 90 to 163, λ from 5195 to 7504 Å), and (4) the RV measurements after telluric correction for a combination of blue and red orders. All curves were obtained after subtraction of their mean value, and are plotted on the same scale. We recall that we did not do any filtering and used a threshold for the contrast, and for this reason we do not obtain the same final result for each RV point as Pepe et al. (2021). All the time series show a constant value, with some fluctuations, as discussed further in Sect. 7. A linear fit to the red arm RV values indicates a change of RV (a drift) of less than 0.064 m s^{-1} (the error bar encompasses 0), while the BERV had a change of 240 m s^{-1} over the whole series: the telluric system of lines has moved with respect to the star system of lines by this amount along the time series. The fact that there is no measurable trend over the time series is a serious indication that the telluric system has indeed been well removed from the telluric-corrected spectra. Unfortunately, the time series that we analyzed covers only 240 m s^{-1}, while the BERV could change by $\simeq$30 km s^{-1}, corresponding to a maximum drift of 8 m s^{-1}. The campaign of HD40307 observations covered only a limited excursion of BERV and we cannot significantly improve on this with this campaign. However, Allart et al. (2022) indeed showed on Tau Ceti that possible drifts due to uncorrected remnants of

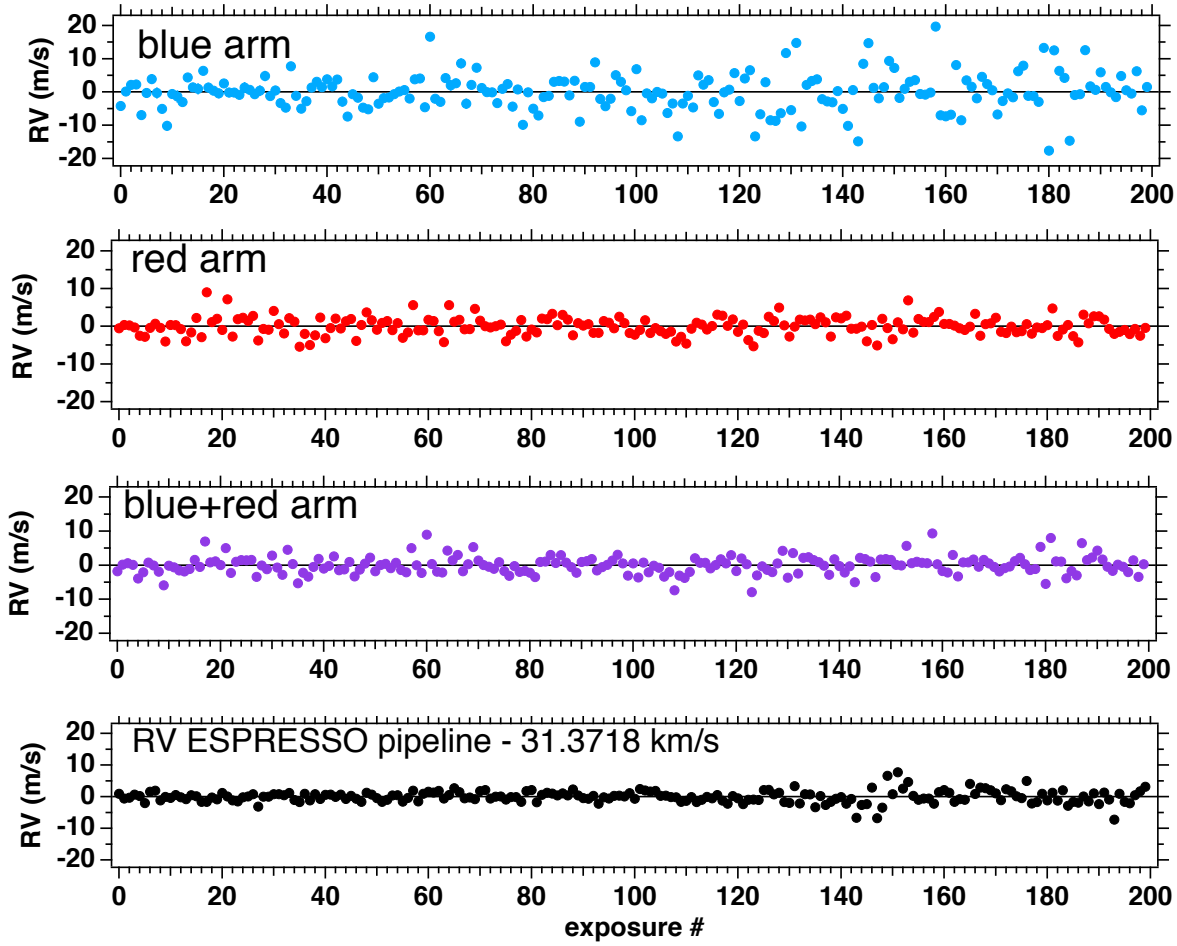


Fig. 17. Temporal and BLUE-RED dependence of the measured RV. Top blue graph: RV measurements from our processing after telluric correction restricted to blue orders. Top red graph: Same for the red orders. Bottom purple graph: Combination of blue and red orders above. Bottom black graph: RV measurements reported by Pepe et al. (2021).

tellurics are smaller than $\simeq 0.2\,\mathrm{m\,s^{-1}}$ over a large BERV excursion ($\pm 28\,\mathrm{km\,s^{-1}}$, their Fig. 12) in the case of their correction technique. As we perform an accurate computation of the atmospheric transmission (by taking the whole p-T vertical profile into account instead of a monolayer) and are using the same HITRAN molecular database, we may not expect our atmospheric correction scheme to produce more drift. The study of the exact amount of drift is an interesting subject, but is beyond the scope of this paper; it would require another target, because the planetary orbits are not yet known with enough precision to detect drifts on the order of $0.2\,\mathrm{m\,s^{-1}}$.

7. Temporal and wavelength dependence of radial velocity fluctuations

During our analysis, we discovered the presence of RV signal fluctuations, as previously mentioned in Pepe et al. (2021). What we refer to as jitter here is these rapid fluctuations of the RV signal, with amplitudes far above the formal error, without hypothesising as to the physical nature of the cause of this jitter. The existence of the time series of 200 exposures of the star HD 40307 offers the opportunity to identify some time periods where the dispersion of RV values is larger than usual (Fig. 17). In order to investigate these fluctuations further, we considered the blue and red arms of ESPRESSO data separately.

7.1. Temporal dependence

The difference in observed jitter between the blue arm and the red arm is striking. While the blue arm reproduces the time variation of the jitter presented in Pepe et al. (2021), with a sudden increase approximately after exposure 130, the red arm does not see such an increase. We divided the time series into two groups: a first group for exposures 0–99, and a second group for exposures 100–199. The change of the standard deviation S_{dev} from group 1 to group 2 is quite different for the blue arm and the red arm, as seen in Table 4. Quantitatively, going from group 1 to group 2, for the blue arm, S_{dev} increased substantially from 4.13 to $6.71\,\mathrm{m\,s^{-1}}$. At the same time, for the red arm, S_{dev} decreased (slightly) from 2.50 to $2.33\,\mathrm{m\,s^{-1}}$.

Our results in Fig. 17 imply that the fluctuations occur only on the blue detector. It is a little worrisome to see this dichotomy arriving between the blue arm and the red arm, because they have two different detectors, and because the telluric correction was mostly relevant to the red arm. Nevertheless, such a change in radial velocity detected by only one arm of the spectrograph is difficult to relate to stellar p-mode oscillations, because implied Doppler shifts of the stellar atmosphere would affect the RVs equally in the red and blue parts of the spectrum. Fortunately, our referee pointed out to us that until a major instrument intervention in May 2022, ESPRESSO had a blue cryostat thermal stability problem (Figueira et al. 2021). Therefore, this type of

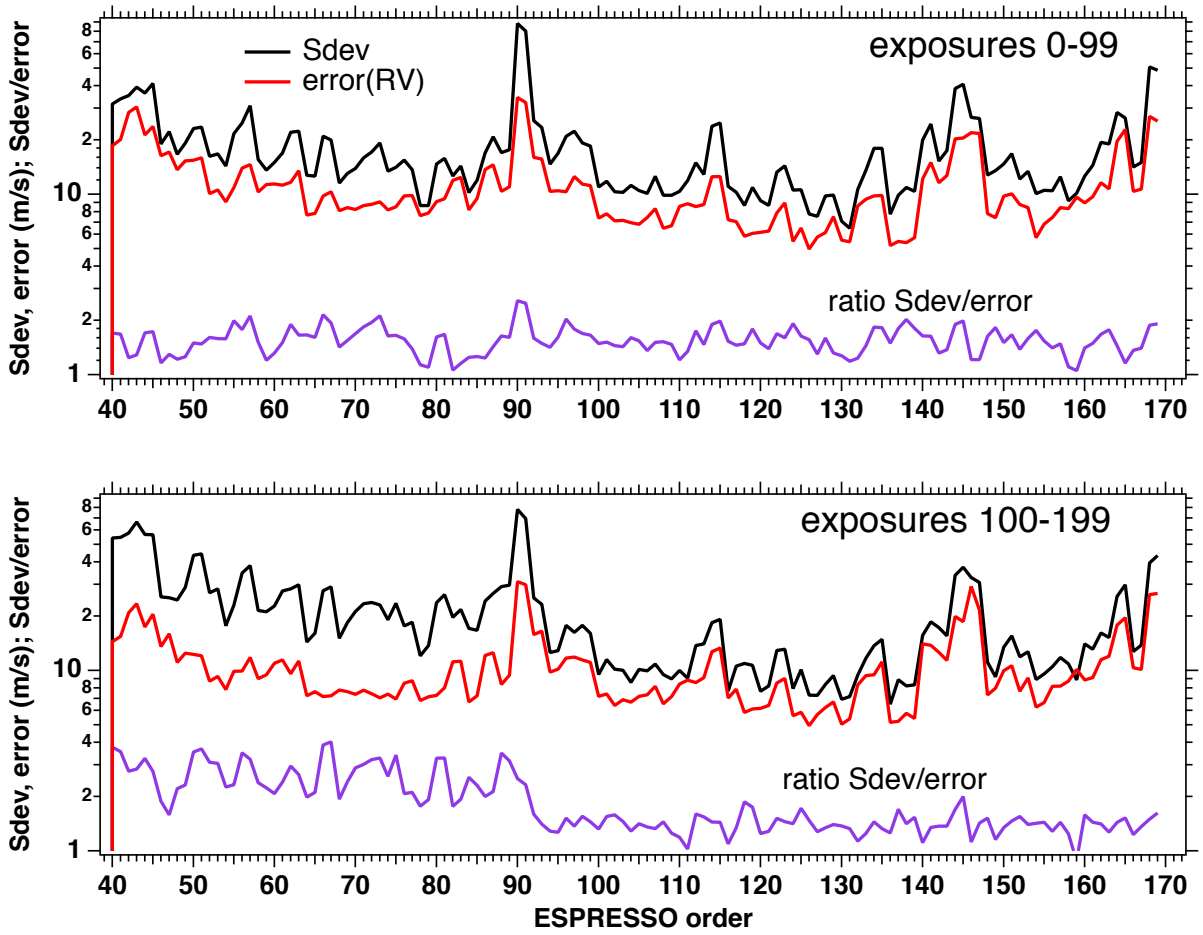


Fig. 18. Wavelength and temporal dependence of the observed RV. Top: standard deviation $S_{\rm dev}$ (black line) and formal error on RV measurements (red line) for each order from our processing, averaged over the first 100 exposures only. The ratio between the two quantities is shown in violet. Bottom: same as top but for the last 100 exposures. We note the high ratio found in the blue orders during the second half of the observational time, a sign of spurious jitter due to a blue cryostat thermal instability.

fluctuation (jitter), which affects only the blue arm that we have detected, is certainly mainly caused by the thermal instability of the blue detector, and is not related to the onset of *p*-mode oscillations. The fact that our separate blue arm and red arm measurements clearly reveal this behavior shows that our telluric correction, mostly applied in the red arm, does not introduce artefacts that would mask this difference in red and blue arm fluctuation levels.

7.2. Wavelength dependence

In order to refine the wavelength dependence of the jitter (to decipher whether or not it coincides with the blue arm only), the time series of retrieved RVs were also studied for each order separately. In particular, we examined the variation versus the ESPRESSO order of three variables: the mean (formal) error 'err' of the RV averaged over exposures, the standard deviation $S_{\rm dev}$ of the RV values among exposures, and the ratio $S_{\rm dev}$/err displayed in Fig. 18 as a function of order number (equivalent to wavelength) and separately for the two groups of exposures. Our reasoning here is that, in the frame of the theory of Gaussian errors, the standard deviation $S_{\rm dev}$ should be equal to the mean error of a series of measurements. This would hold if the error were well estimated, and if the spectra were not changing (otherwise from shot noise), implying a constant retrieved RV (within shot noise limits).

We note in Fig. 18 that both the error and $S_{\rm dev}$ are changing for three reasons: the intensity of the spectrum in the order (shot noise), the number of masked lines for this order, and their depths (contrast).

However, the value of $S_{\rm dev}$ is systematically larger than the value of the error, with a strong correlation (from order to order) between $S_{\rm dev}$ and the error. There are usually two reasons for which the standard deviation is found to be larger than the mean error: either the error bar has been underestimated or there are time fluctuations (jitter) in the data superimposed to a constant mean RV value, or a combination of both effects.

It can be shown mathematically (with the theorem of the variance of a sum of independent variables X and Y) that $S_{\rm dev}$ may be related to the formal error and the jitter amplitude as

$$S_{\rm dev}^2 = jitt^2 + k^2 {\rm err}^2, \quad (1)$$

where *jitt* is the amplitude of the jitter (affecting the RV), and k is a factor by which the formal error is underestimated with respect to the true error. This factor k is likely valid at all times and all orders, because it is intrinsic to the retrieval algorithm.

For the second group of exposures (100–199), a dichotomy is observed between the left part (blue arm) and the right part (red part). The behavior of the right part is quite similar to the whole top curve (0–99), both for the correlation between error (RV) and $S_{\rm dev}$, and the approximately constant value of the ratio

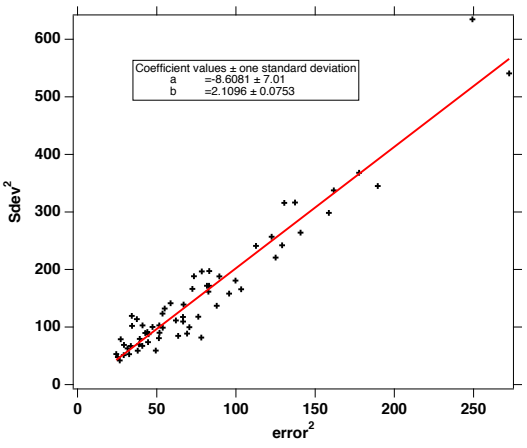


Fig. 19. Linear fit of the scatter plot of $S^2_{\rm dev}$ as a function of the error^2. Each point is for one red order of ESPRESSO. The slope allows us to determine the factor k of underestimation of formal error: k = true error/formal error.

around 1.41 (which is lower than for the first group, which is 1.55 for the red arm). On the contrary, for the blue arm, the $S_{\rm dev}$ curve is detached above the RV curve, there is no correlation between error(RV) and $S_{\rm dev}$, and the ratio is much larger, with a mean of 2.67. Therefore, the jitter, which appears mostly during the second half of the series of exposures, happens only in all the orders of the blue arm.

As mentioned above, the explanation of this blue jitter is related to the instrument, and linked to thermal instabilities of the cryostat containing the blue detector. The smallest values of the ratio $S_{\rm dev}$/err are found in the red arm for exposures 100–199 (Fig. 18). Let us assume that the jitter amplitude is the same for all red orders in this time frame. Figure 19 shows $S_{\rm dev}{}^2$ as a function of the squared formal error err^2, each point representing one order from 92 to 169. After the exclusion of a few outliers, there is a clear linear relationship of the form $a + bx$, where the slope $b = k^2$ is found to be 2.1 ± 0.07. This means that our formal error is underestimated (with respect to the true error) by a factor of $\simeq$1.4.

Extrapolating the linear fit to err = 0, the value of $S^2_{\rm dev}$ becomes equal to jitt^2. This value is found equal to $-8.6 \pm 7\,{\rm m^2\,s^{-2}}$. As a squared quantity cannot be negative, the jitter is likely weak, and undetectable in this portion of the spectrum and exposures 100–199.

8. Conclusions

We used an exceptional series of consecutive short exposures of the ESPRESSO spectrograph for a K2.5V type star to test and estimate the potential increase in precision of the RV measurement brought by telluric-line removal when this correction is based on a publicly available, location-, and time-matched atmospheric transmission spectrum and a simple and conventional fitting code to be performed locally.

We used theoretical transmissions of H_2O and O_2 downloaded from the TAPAS online facility and adjusted them to each exposure before dividing each recorded spectrum by this adjusted transmission. TAPAS conducts a high-quality computation of the telluric absorption based on the vertical T-P profile provided by the ECMWF meteorological field, the HITRAN spectroscopic data base and the LBLRTM code. To perform the adjustment of H_2O and O_2 columns, we took advantage of the characteristics of the target star, that is, a temperature of $\simeq$4980K and a low metallicity of $\simeq$–0.3, resulting in lines narrow and weak enough to allow easy stellar continuum fitting and distinction between tellurics and stellar features in some spectral regions. This process makes it possible to mask the stellar lines, adjust the telluric transmission in a very precise way, and divide the data by the adjusted telluric model, without direct use of a stellar synthetic spectrum. The quality of the adjustment is illustrated by the close correspondence between the fitted H_2O column and the column deduced for each exposure from the infrared radiometer located at Paranal, as well as the close correspondence between the adjusted O_2 column and the airmass.

We augmented the ESPRESSO binary mask with 696 additional stellar lines located within regions previously excluded due to the presence of tellurics (additional mask: BMc). We applied the same standard CCF method, with the same parameters, to the initial data and the standard binary mask as well as to the corrected data and the augmented mask.

We compared the average error on RV as well as the average spread of the results before and after the telluric correction and mask extension. For data from orders 90 to 163 (the ESPRESSO red arm), we find that the average error on RV measurements decreases from E(BM1) = 1.04 m s^{-1} for the standard binary mask BM1 to E(BM2) = 0.78 m s^{-1} when using spectra corrected for telluric absorption and the new binary mask adapted to corrected regions. The blue arm is almost unaffected by the telluric absorptions. The red arm error values of E(BM1) = 1.04 and E(BM2) = 0.78 m s^{-1} are for a 30 s exposure of the star HD40307, which is of spectral type K2.5, and has visual and red magnitudes of V = 7.1, R = 6.6, giving $\simeq$7094 electrons per pixel at the center of order 146 ($S/N \simeq 84$). For the same stellar spectral type, we expect that both errors will scale inversely to the S/N per pixel (from shot noise), and therefore their ratio will remain constant. For cooler stars, which have more flux in the red part for the same visual magnitude V, the reduction factor of the error will be larger. On the other hand, in the course of our study of the series of measurements of HD 40307, we noted that some jitter affects the RV signal, revealing some sudden increase in stellar activity, as reported by Pepe et al. (2021). We were surprised to find that this jitter has a greater affect on the bluest part of the spectrum than on the red part. We recently learned that this fluctuation was produced by thermal fluctuations of the blue arm cryostat.

When using the full spectrum (blue+red arms), the formal error after telluric correction is reduced from 0.77 to 0.64 m s^{-1} (or, if using the photon noise minimal uncertainty estimation, from 0.55 to 0.51 m s^{-1}). This improvement corresponds to a saving of about 45% in the telescope time required to reach the same precision (or 16% in the telescope time if using the photon-noise minimal uncertainty estimation). This is important when considering the huge need for telescope time when monitoring target stars in order to determine their populations of planets. Furthermore, it has been shown that long-term stellar activity (e.g., associated to the rotation period of the star) induces some spurious RV variations that are smaller in the red than in the blue, at least in the case of Proxima Centauri (Suárez Mascareño et al. 2020). This increases the benefit of telluric corrections, as they broaden the spectral intervals available for RV retrievals. Similarly, all stellar spectra taken before the correction of the blue detector instability (introducing a short-term artificial "jitter" in the RV data derived from the blue detector) could be reprocessed with our TAPAS-derived procedure for a better RV estimate.

The binary mask BMc built for the telluric contaminated regions is probably also valid for other spectral types close to

K2.5, but this needs to be explored. Cooler-type stars with more numerous lines (e.g., M stars) and/or stars with higher metallicity might be more difficult to treat; however, state-of-the-art atmospheric transmission spectra and synthetic stellar spectra allow us to predict atmospheric lines or fractions of lines free of stellar absorption. Moreover, it has been shown that in this case, fitting the observed spectrum with a forward model of the product of a stellar synthetic spectrum and a telluric transmission spectrum is also feasible (see, e.g., Puspitarini et al. 2015), allowing the quantity of H_2O to be determined and the quantity of O_2 in the atmosphere to be verified at the time of observations. Such procedures allow us to retrieve the stellar spectrum corrected for atmospheric absorption.

The present technique has the advantage of being applicable to any individual recorded spectrum; that is, it does not require time series or libraries of spectra. Also, it benefits from the very highly detailed transmission spectra publicly available from the TAPAS facility. Telluric lines associated with observatories located at different altitudes and in different regions have different shapes, and these differences are not entirely smoothed out after convolution by the spectrograph PSF. There are also changes according to the season. This is why computations made for a stratified atmosphere adapted to both location and date provides an optimal input for the corrections. As we show here, the O_2 transmission spectrum can be directly used after scaling to the airmass, while the H_2O transmission spectrum requires scaling. Moreover, the atmospheric water vapor is changing rapidly, often on timescales of shorter than 1 h, and the 6 h interval between ECMWF computations is too long to follow its variations. We show that the H_2O columns we have adjusted for all individual exposures very closely follow the values derived from the Paranal infrared radiometer and indicated in the data file for each exposure. This implies that, if an on-site measurement of precipitable water vapor column is provided, an initial scaling can be performed using the ratio between this measured column and the one predicted by TAPAS/ECMWF indicated in the accompanying information. Equivalently, one may enter the on-site value in the TAPAS request form and the computed profile will be adjusted accordingly. This provides a preliminary H_2O transmission very close to the true solution.

The correction method we present here requires that TAPAS transmission spectra be downloaded. TAPAS computations are now very fast and interruptions are very rare since the website has been upgraded. Because species other than H_2O follow the airmass, as we show here for O_2, it is possible to retrieve TAPAS transmissions for the different species at the beginning of the observing night or during an exposure and to adapt them to the airmass of each observation. Adjustments of H_2O are required for each exposure, as mentioned above, but they are also very fast provided one has prepared the mask appropriate to the stellar type in advance. For this work, the H_2O adjustment, the computation of the profiles order by order, and the correction take less than 1 min per exposure on a laptop with a 2.4 GHz Intel Core I9. TAPAS now covers the 300–3500 nm interval and is well adapted to both the near-UV, for example for the CUBES spectrograph under development (Covino et al. 2022), and to near-IR spectrographs such as SPIROU (Donati et al. 2020) or NIRPS (Bouchy et al. 2019). The available species in TAPAS are O_2, H_2O , O_3, CO_2, CH_4, and N_2O. Both NO_2 and NO_3 are also expected to become available soon.

Acknowledgements. We thank our referee for his/her very constructive comments and for having pointed out the instrumental reason of increased jitter for the blue detector data. We deeply thank Piercarlo Bonifacio for providing his personal ATLAS-SYNTHE environment and helping us to compute an appropriate stellar synthetic spectrum, used for comparison tests (template matching) and stellar line identification. We acknowledge a very useful support from Burkhard Wolff from ESO who helped us with ESPRESSO instrumental details and the binary mask. A.I. acknowledges the support of a Vernadski Scholarship for PhD students, sponsored by the French Government and the Ministry of Science and Higher Education of the Russian Federation under the grant 075-15-2020-780 (N13.1902.21.0039). This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France.

References

Allart, R., Lovis, C., Faria, J., et al. 2022, A&A, 666, A196
Artigau, É., Astudillo-Defru, N., Delfosse, X., et al. 2014, SPIE Conf. Ser., 9149, 914905
Artigau, É., Hébrard, G., Cadieux, C., et al. 2021, AJ, 162, 144
Baranne, A., Mayor, M., & Poncet, J. L. 1979, Vistas Astron., 23, 279
Baranne, A., Queloz, D., Mayor, M., et al. 1996, A&AS, 119, 373
Bedell, M., Hogg, D. W., Foreman-Mackey, D., Montet, B. T., & Luger, R. 2019, AJ, 158, 164
Bertaux, J. L., Lallement, R., Ferron, S., Boonne, C., & Bodichon, R. 2014, A&A, 564, A46
Boisse, I., Eggenberger, A., Santos, N. C., et al. 2010, A&A, 523, A88
Bouchy, F., Pepe, F., & Queloz, D. 2001, A&A, 374, 733
Bouchy, F., Doyon, R., Pepe, F., et al. 2019, in EPSC-DPS Joint Meeting 2019, 1860
Clough, S. A., & Iacono, M. J. 1995, J. Geophys. Res., 100, 16,519
Clough, S. A., Shephard, M. W., Mlawer, E. J., et al. 2005, J. Quant. Spec. Radiat. Transf., 91, 233
Connes, P. 1985, Ap&SS, 110, 211
Covino, S., Cristiani, S., Alcala', J. M., et al. 2022, arXiv e-prints, [arXiv:2212.12791]
Cretignier, M., Dumusque, X., Hara, N. C., & Pepe, F. 2021, A&A, 653, A43
Díaz, R. F., Ségransan, D., Udry, S., et al. 2016, A&A, 585, A134
Donati, J. F., Kouach, D., Moutou, C., et al. 2020, MNRAS, 498, 5684
Figueira, P., Lo Curto, G., & Mehner, A. 2021, Very Large Telescope Paranal Science Operations ESPRESSO User Manual
González Hernández, J. I., Rebolo, R., Pasquini, L., et al. 2020, A&A, 643, A146
Gordon, I. E., Rothman, L. S., Tan, Y., Kochanov, R. V., & Hill, C. 2017, in 72nd International Symposium on Molecular Spectroscopy (Urbana-Champaign: The University of Illinois), TJ08
Gullikson, K., Dodson-Robinson, S., & Kraus, A. 2014, AJ, 148, 53
Lafarga, M., Ribas, I., Lovis, C., et al. 2020, A&A, 636, A36
Leet, C., Fischer, D. A., & Valenti, J. A. 2019, AJ, 157, 187
Mayor, M., & Queloz, D. 1995, Nature, 378, 355
Pepe, F., Mayor, M., Galland, F., et al. 2002, A&A, 388, 632
Pepe, F., Cristiani, S., Rebolo, R., et al. 2021, A&A, 645, A96
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes in C. The Art of Scientific Computing (Cambridge: Cambridge University Press)
Puspitarini, L., Lallement, R., Babusiaux, C., et al. 2015, A&A, 573, A35
Rothman, L. S. 2021, Nat. Rev. Phys., 3, 302
Smette, A., Sana, H., Noll, S., et al. 2015, A&A, 576, A77
Suárez Mascareño, A., Faria, J. P., Figueira, P., et al. 2020, A&A, 639, A77

Appendix A: The RV uncertainty estimation based on photon-noise and CCF.

The photon noise is one limiting factor for the precision of RV time-change measurements, the other being the richness of the spectral features to be used to detect a Doppler shift of the star. A mathematical expression of the limitation was developed by Connes (1985) and tested for different types of stars and spectrographs by Bouchy et al. (2001). It uses the quality factor Q devised by Connes (1985) and revived in Bouchy et al. (2001), which represents the spectral-line richness of the spectrum; it is independent of the actual size of the spectral range, but is a function of the complexity of the spectral profile within that range. The estimation of the RV uncertainty δV_Q based on the quality factor Q is:

$$\delta V_Q = \frac{c}{Q\sqrt{N_{e^-}}}, \tag{A.1}$$

where c is the speed of light, Q the quality factor, and N_{e^-} is the total number of photo-electrons counted over the whole spectral range. The quality factor Q makes use of the derivative of the intensity with respect to the wavelength and should be estimated from a noise-free spectrum. The derivative is computed as the finite difference in intensity from one spectel to the next.
In reality, an observed spectrum is not noise free. Therefore, with noisy data the quality factor will be spuriously higher and the RV uncertainty will be lower than the true optimal uncertainty.
Boisse et al. (2010) described an approach to apply this formalism; not to a spectrum as in Connes (1985), but to a CCF. As the CCF for an order is constructed by piling up all the lines of this order, it can be considered as an average spectral line of the order and also as noise free. In Boisse et al. (2010), the quality factor of the spectrum is calculated as

$$Q = \frac{\sqrt{\sum W(i)}}{\sqrt{\sum A_0(i)}}, \tag{A.2}$$

where i is the pixel number of the spectrum and $W(i) = \frac{\lambda^2(i)(\partial A_0(i)/\partial\lambda(i))^2}{A_0(i)+\sigma_D^2}$, where A_0 is a reference noise-free spectrum intensity and σ_D^2 is the detector noise.
In the case of the CCF, the wavelength scale is replaced by a velocity scale. The quality factor Q_{CCF} shown in Boisse et al. (2010) becomes

$$Q_{CCF} = \frac{\sqrt{\sum_i \left(\frac{\partial CCF(i)}{\partial V(i)}\right)^2 / CCF^2_{noise}(i)}}{\sqrt{\sum_i CCF(i)}} \sqrt{N_{scale}}, \tag{A.3}$$

where $CCF_{noise}(i)$ is the quadratic sum of the photon noise and the read-out detector noise integrated inside the CCF mask holes for the velocity i, which is analogous to the $A_0(i)+\sigma_D^2$ in the W(i) formulation from Bouchy et al. (2001); and N_{scale} is a correction factor corresponding to the scale of the velocity step in detector pixel units. The velocity uncertainty is equal to

$$\delta V_{QCCF} = \frac{1}{Q_{CCF}\sqrt{\sum CCF(i)}}. \tag{A.4}$$

Uncertainties computed both from our Gaussian fit of the CCF δV_G and from the application of the above quality factor to the CCF δV_{QCCF} are shown in figures A.1 and A.2. In both cases, estimated uncertainties decrease after telluric correction. We verified that the fluctuations of δV_{QCCF} (Fig. A.2) are almost entirely due to intensity fluctuations reflected on $\sqrt{N_{e-}}$. This explains the parallelism of both curves, before and after telluric correction. In the last figure A.3, we show a comparison between our estimates based on a Gaussian fit δV_G and those based on the quality factor δV_{QCCF} for all orders (blue and red) simultaneously. We can see that, generally, the uncertainty estimation obtained from the quality factor δV_{QCCF} is smaller than the one obtained by a Gaussian fit and also displays smaller fluctuations. This is because Boisse et al. (2010) used the full CCF to perform their estimate while we used only ten points of the CCF for our Gaussian fit (we reiterate that in this way we attempt to avoid RV biases appearing when fitting a whole spectral line as described by González Hernández et al. (2020).

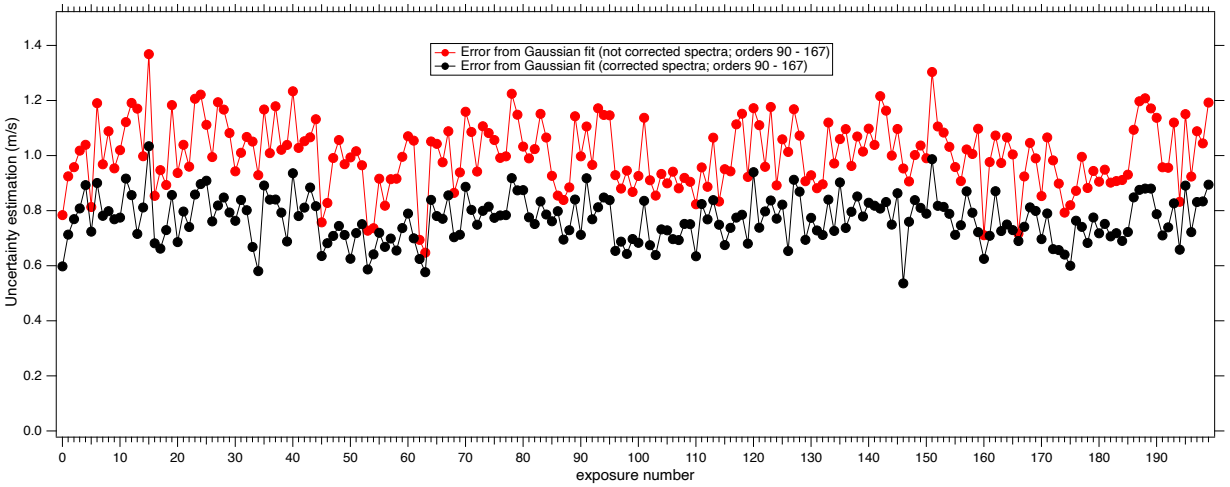


Fig. A.1. Estimated uncertainty $\delta V_{\rm G}$ based on the Gaussian fit for the red arm. The red and black curves show the uncertainty before and after telluric correction, respectively.

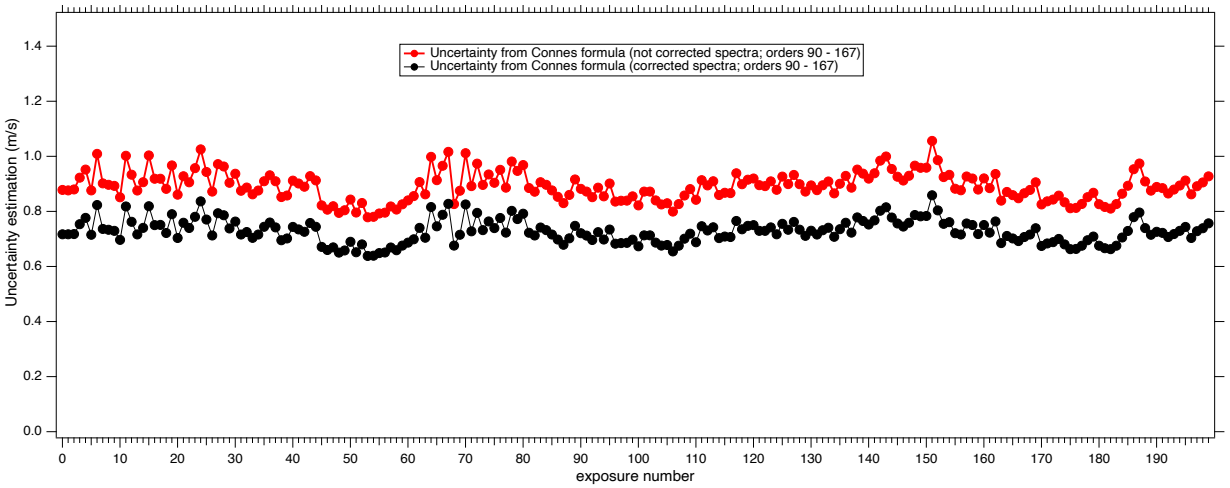


Fig. A.2. Estimated uncertainty δV_{QCCF} based on the quality factor for CCF from Boisse et al. (2010) for the red arm. The red and black curves show the uncertainty before and after telluric correction, respectively. Fluctuations are connected to the fluctuations of spectral intensity (see text).

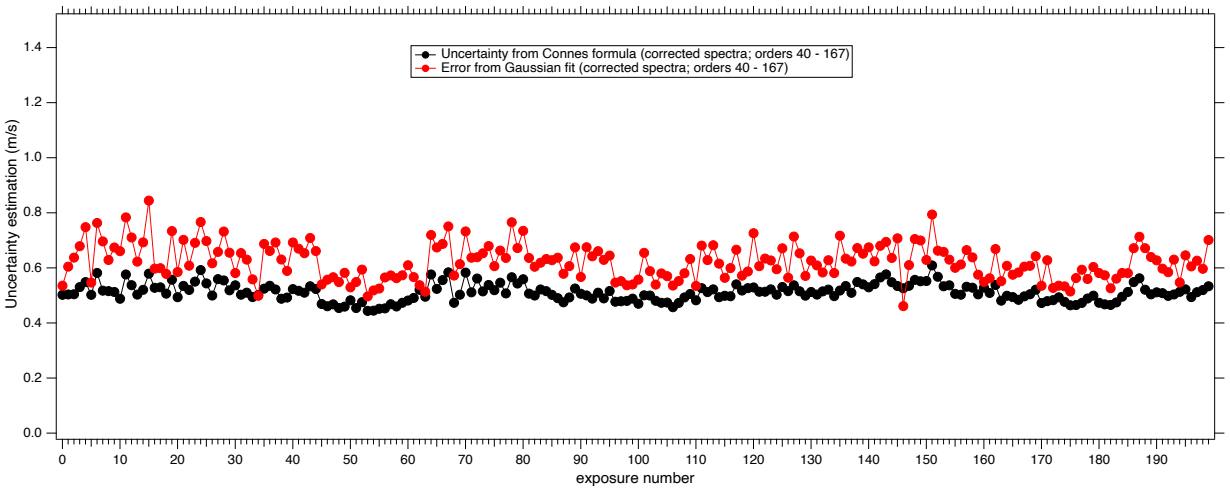


Fig. A.3. Comparison of estimated uncertainties: based on Gaussian fit of the CCF ($\delta V_{\rm G}$, in red) and on the quality factor for the CCF from Boisse et al. (2010), (δV_{QCCF}, in black), for red and blue arms combined.