Astronomy & Astrophysics manuscript no. Cretignier_2020_RASSINE
February 13, 2020

RASSINE: An interactive tool to normalise spectra

I. Description and performance of the code

M. Cretignier[1], J. Francfort[2], R. Allart[1], X. Dumusque[1], and F. Pepe[1]

[1] Astronomy Department of the University of Geneva, 51 ch. des Maillettes, 1290 Versoix, Switzerland
e-mail: michael.cretignier@unige.ch
[2] Département de Physique Théorique, Université de Genève, 24 quai Ansermet, CH–1211 Genève 4, Switzerland
e-mail: jeremie.francfort@unige.ch

Received XXX ; accepted XXX

ABSTRACT

Aims. Provide an open-source code allowing an intuitive and robust spectra normalisation.
Methods. We developed RASSINE, a *Python* code to normalise *1d* spectra through the concepts of convex hulls. The code uses 6 parameters that can be easily fine-tuned. The code also provides a complete user-friendly interactive interface, including graphical feedback, helping the user to chose the parameters as easily as possible. To facilitate even further the normalisation, RASSINE can provide a first guess for the parameters which are derived directly from the spectrum thanks to calibrations already performed on the parameters.
Results. We found out that such normalisation were accurate and precise enough to provide reliable line depth for solar and α Cen B spectra. The accuracy of RASSINE was estimated to 2.0% on α Cen B HARPS spectra where this low accuracy is in part due to molecular band absorption and line density in the bluest part of the spectrum. Excluding wavelength shorter than 4500 Å increases the accuracy up to 1.2%. For HARPS spectra of the Sun, an accuracy of 0.20% can be reached which is 3 times better than with the commonly used method of polynomial fit. The precision on individual spectra normalisation is estimated around $\sim$ 0.15% which can be reduced to the photon-noise limit (0.10%) when a time-series of spectra is given as input for RASSINE.
Conclusions. With an accuracy higher than polynomial fitting method and a precision compatible with the photon noise, RASSINE is a tool that could find applications in numerous cases, for example as stellar parameters determination, transmission spectroscopy of exoplanets or activity-sensitive lines detection.

Key words. Spectrum normalisation – Alpha shape – Python code – RASSINE

1. Introduction

A spectrum is a fundamental observable used to study astronomical objects such as galaxies, stars and exoplanets. It describes the distribution of photons per wavelength bins and can either be used in term of absolute quantity to determine the luminosity of the objects or colors can be analyzed separately through photometric bands. A rich content of information is also brought by the absorption or emission lines for which the spectrum has to be normalised by a continuum. This happens for instance in the framework of stellar abundances studies (e.g. Blanco-Cuaresma et al. 2014; Sousa et al. 2015; Adibekyan et al. 2016) or exoplanets atmospheres (e.g. Wyttenbach et al. 2015; Allart et al. 2017). For radial velocity (RV), the spectrum does not need to be continuum-normalised but a color correction has to be applied (e.g Bourrier & Hébrard 2014; Malavolta et al. 2017), itself related to the continuum of the spectrum. Normalised spectra are also necessary to construct the binary masks used in Doppler spectroscopy to extract RV using the cross-correlation technique (Pepe et al. 2003).

When studying absorption lines, a non-trivial step consists in normalising the spectrum by its continuum, where the latter can differ from a blackbody curve due to Rayleigh scattering reddening. Such differential chromatic response can also be directly induced by the spectrograph itself through the optical elements and the CCD quantum efficiency that varies with wavelength, which cannot be modeled by any radiative transfer code. Due to the large number of high resolution spectra obtained with spectrographs such as from ESPRESSO (Pepe et al. 2014), EXPRESS (Fischer et al. 2017), NEID (Schwab et al. 2016), PEPSI (Strassmeier et al. 2015), CRIRES, NIRPS (Bouchy et al. 2017) and the huge amount of spectra already available produced by CORALIE (Queloz et al. 2000), HARPS (Pepe et al. 2002b; Mayor et al. 2003; Pepe et al. 2003), HIRES (Pasquini et al. 2010) or HARPN (Cosentino et al. 2012), an effective process to normalise them in a unified and coherent way appears as an important step.

Current methods often deal with *2d* spectra from echelle-grating spectrographs since they represent a narrower band of the spectrum where the continuum presents a smaller number of inflection points. A filter is often used in order to smooth the data and remove as many stellar lines as possible. This filter can be a rolling maximum or moving average, an asymmetric sigma clipping or even a Fourier filtering. A low order polynomial can eventually be used to fit the continuum and often the blaze function simultaneously to the smoothed version of the spectrum.

In the case of *1d* spectra with large wavelength coverage, high order polynomial functions are mandated to account for the numerous inflection points. However, approximating the spectral continuum, which is the product of the stellar continuum, atmo-

spheric absorption and CCD response, by a polynomial function is an unrealistic model overly simplistic.

If the continuum of a stellar template is known, the continuum of a given spectrum can be obtained by deriving the ratio between this spectrum and the template, and then fitting any trend observed in this ratio. The main advantage of this method consists in considerably reducing the order of the polynomial needed, but the intrinsic disadvantages of polynomial fit remain (Škoda 2008). Such method allow for instance to correct for color variation induced by different airmass observations (Malavolta et al. 2017).

Using the methods cited above, most scientists can at some level perform a sufficiently good normalisation, namely by iteratively performing smoothing, frequency filtering and polynomial fitting (Tody 1986, 1993). However, a lot of issues are usually encoutered when fitting a polynomial (Škoda 2008) as selecting the sub-selection of spectral fluxes used in the fit. Also, the chain of operations necessary to normalise spectra can suffer from a lack of consistency and is often not described in great details by authors.

RASSINE (Rolling Alpha Shape for a Spectral Improved Normalisation Estimation) is a Python code in free access[1] written to guide the user with the different steps necessary to perform an efficient spectrum normalisation, with animated and graphical feedback provided. The RASSINE code uses the convex hull and alpha shape theories to model the upper envelope of a spectrum, often equivalent to the continuum for solar-type stars. Those theories, as well as the algorithm behind the code is described in Sect.2. We note that the approach presented in this paper is similar to the one recently published in Xu et al. (2019), where the authors showed the higher performance of the alpha shape method compared to classical iterative methods. The accuracy and precision scores of the code, as well as its ability to extract the correct continuum around broad lines, are described in Sect.3. RASSINE uses 6 free parameters to model the spectrum continuum. An automatic mode of the algorithm, described in Appendix B, can provide first guesses for five of them.

2. Theory

2.1. Convex hull and alpha shape

The problem of finding the upper envelop of a spectrum is closely related to the concepts of convex hull and alpha shape, the latter being a generalization of the former. We briefly review those ideas here, restricting our description to the two-dimensional case

The convex hull of a set of points S in the plane can be understood instinctively as follows: take an elastic band and stretch it around the set S. The final shape that the band will take is the convex hull of the set. Mathematically speaking, as defined by Asaeedi et al. (2013), if the set is made of points x_i, the convex hull $C(S)$ can be written as :

$$C(S) = \left\{ y_i = \sum \alpha_i x_i \in \mathbb{R}^2 \;\middle|\; \alpha_i \geq 0, \sum \alpha_i = 1 \right\}. \quad (1)$$

The convexity of the convex hull forbids the existence of inflexion points. Since a black body radiation curve presents inflexion points, we need to consider the concept of alpha shapes, which is a generalisation of the convexe hull theory and allows for concave hulls.

When considering a convex hull, all the internal angles are at most 180°. In the alpha shape framework, this condition is relaxed such that each internal angle is at most 180° + α. This allows to model the inflexion points present in the upper envelope of a spectrum. This upper envelope is then modeled by a shape that is intermediate between a convex hull (α = 0) and a full alpha shape (α = 180).

2.2. Outline and structure of the code

2.2.1. Brief overview

The main assumption of the code is that the local maxima of the spectrum are corresponding to the continuum. This assumption is relatively well satisfied for solar-type spectra, except perhaps in the bluest part of the visible because of scattering and the high density of spectral lines. However, this assumption does not hold anymore for M dwarfs spectra, for which the extremely high-density of lines completely erase the stellar continuum at nearly all wavelengths.

RASSINE can be subdivided in 5 steps called the SNAKE sequence:

1. Smoothing of the spectrum,
2. Neighborhood's maxima detection,
3. Alpha shape algorithm,
4. Killing outliers,
5. Envelop interpolation.

The smoothing is necessary to increase the signal-to-noise ratio (S/N) per bin element which allows to deal with spectra at different noise levels. This step is also relevant, since it prevent adjusting the continuum to the upper envelope of the noise, which is not physically correct. After the smoothing, the maxima corresponding to the upper envelope are then chosen using a modified gift wrapping algorithm, which removes the local maxima formed by blended lines. An alpha-shape, that can be compared to a rolling pin rolling over the remaining local maxima, is then used to find the best continuum. We note that the radius of the alpha-shape changes according to a penalty law, which allows to correctly account for broad absorptions. Outliers points are then rejected according to given criteria that prevent biasing the different interpolation performed at the end.

2.2.2. Smoothing of the spectrum

This first step of smoothing consists in removing the high frequency noise present in the stellar continuum, to prevent being contaminated by it when looking for the best fit to the continuum. RASSINE proposes two different smoothing options, either degrading the spectrum resolution, or using a Savitzky-Golay filter (Savitzky & Golay 1964).

For the first smoothing option, the spectrum resolution is degraded by convolving the spectrum with either a rectangular or Gaussian kernel. We note that this convolution will change the depth of spectral line, however, since the goal is to fit at best the stellar continuum, this will not be an issue. The kernels can be choosen with the parameter ***par_smoothing_kernel***[2]. The strength of the smoothing is controlled by the width of the kernel. The associated parameter ***par_smoothing_box*** corresponds to the window width of the smoothing process in units of the spectral sampling.

[1] https://github.com/MichaelCretignier/Rassine_public

[2] In the rest of this article, we will use the convention of the bold and cursive syntax for the parameters, where the name of the parameter is written as in the config file of the code.

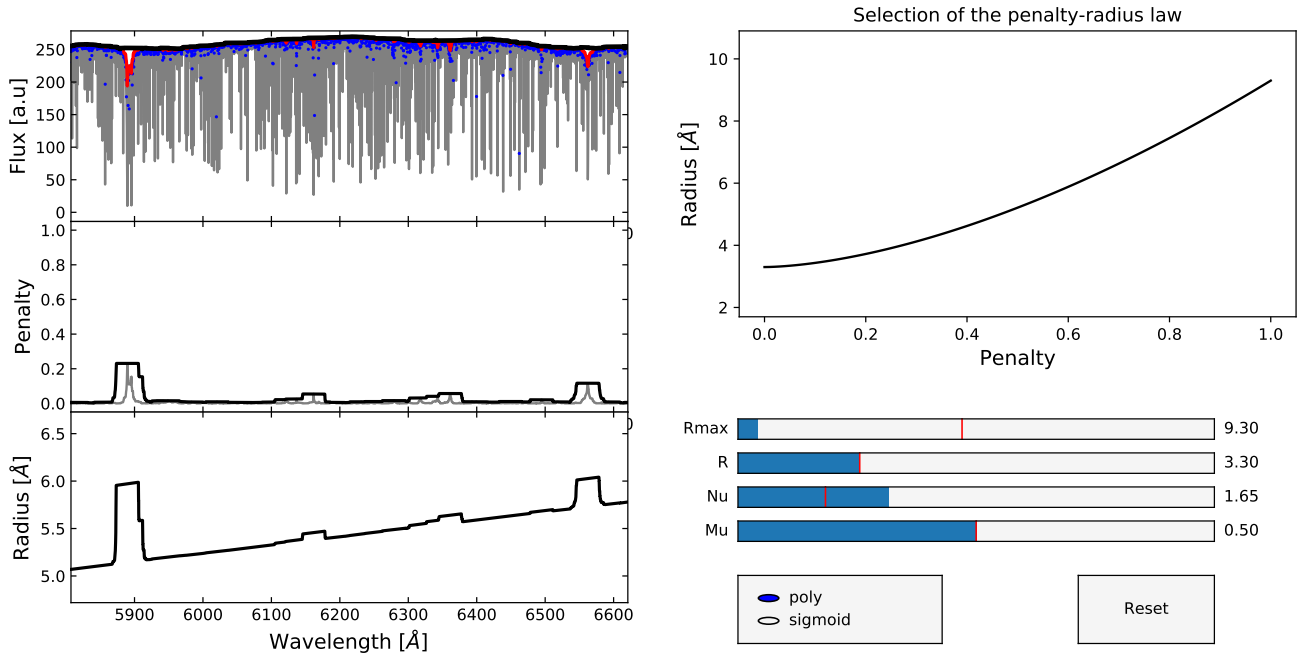

Fig. 1: Graphical interface of the second RASSINE interactive plot. In this case, the values for the parameters are bad (R_{max} too small, ν too high) but chosen to highlight the dependency of the radius with wavelength. **Top left:** The local maxima (blue dots) extracted from the smoothed spectrum (gray curve). A continuum S_1 using a small window (red curve) and S_2 using a large window (black curve) are computed with a rolling maxima on the spectrum. **Middle left:** The penalty is computed from the relative difference between S_1 and S_2 (gray curve). The penalty is rounded up and looks like a step function (black curve). **Bottom left:** Final radius values for the rolling pin along the wavelength axis. Broader absorption lines like the sodium doublet NaD around 5890 Å or the H_α at 6563 Å are penalized such that the rolling pin increases in size. The observed linear trend corresponds to the dispersion, as all spectral lines have at first order the same line width in velocity. **Top right:** Penalty map relation between the radius of the rolling pin and the penalty value. The curve is parametrised with a minimum radius R, a maximum radius R_{max} and an exponent ν, which represent the exponent of the power law. The second parameter μ is only useful when the sigmoid function is selected for the penalty map and represents the center of the sigmoid (ν being proportional to the sigmoid width).

The second smoothing option consists of a Savitzky-Golay filter, which can be described as a low order polynomial rolling fit on the data. The order of the polynomial is chosen to be of degree three. The advantage of this filter comprises its ability to keep almost invariant the global shape of the spectrum, whereas it filters very well the oscillations on smaller scale than the window. However, this filter remains more sensitive to strong flux variation which can be due e.g. to cosmic rays, lamp contaminatio or hot pixels. To counteract this issue, an asymmetric sigma-clipping is performed on the absolute difference between the smoothed and initial spectrum and outliers in the smoothed spectrum are brought back to the initial flux values to detect and reject these points more easily at a later stage.

2.2.3. Local maxima

The second step performed by the code is to search for the local maxima. A point is called local maximum if its value is the highest in its closest neighbourhood. A parameter ***par_vicinity*** is introduced, corresponding to the size of the half-window defining the neighbourhood. The window is expressed in unit of the wavelength sampling (e.g. 0.01 Å for HARPS spectra). If the previous smoothing step was performed correctly, this parameter remains quite irrelevant, but it gets important when low S/N spectra are studied and allows to speed up the reduction.

2.2.4. Penalty for broad absorption regions

At first order, all the spectral lines should have the same width in velocity space, determined by the projected stellar rotation, macroturbulence and instrumental resolution. To get an average line width, we can use the FWHM parameter of the cross correlation function (CCF, Baranne et al. (1996); Pepe et al. (2002a)). This average line width is one of the most important parameter for RASSINE, and it can be stored, in $\mathrm{km\,s^{-1}}$, in the ***par_fwhm*** variable. As we will see in the next sub-section, the idea behind RASSINE is to roll a rolling pin, of a size larger than typical spectral lines, on-top of the spectrum to adjust its continuum. If the size of the rolling pin is always the same, the continuum will be badly fitted in region of strong absorption. When lines begin to saturate the Voigt profile's wings begin to dominate and the line width diverges from the typical value of the weak regime. As examples of strong absorption lines, we can mention the H_α, CaII H&K, sodium doublet NaD or triplet magnesium MgIb.

To prevent the rolling pin to fall in regions of broad absorption, a penalty map is used to increase the radius of the rolling pin as soon as such a region is reached. The interactive graphical interface for this step is presented in Fig.1. First of all, we adjust two approximated continuum of the spectrum by using a rolling maximum algorithm. The first one S_1 (red curve first panel Fig.1), using a small window about 40 times the FWHM of the CCF (converted in Å by using λ_{min}) and the second one S_2, using a window 10 times larger than S_1 (black curve first

panel Fig.1). To prevent the rolling maximum algorithm to be sensitive to anomalous flux intensities (e.g. cosmics, contamination between fibers), we first reject outliers using a rolling sigma clipping on a 5 Å window, with the specificity that in our sigma-clipping, the mean is replaced by the median and the standard deviation by 1.5 time the interquartile range (1.5 time the 75% - 25% quartiles, (Upton & Cook 1996)). Then, for each wavelength, the penalty $p = (S_2 - S_1)/S_2$ is computed and normalised by the minimum and maximum value over all the spectrum to lie between 0 and 1. This penalty curve (gray curve in the second panel of Fig.1) is then transformed into a step function (black curve in the second panel of Fig.1) to make the code more efficient.

The next step consist in choosing the parameter ***par_reg_nu***, which appears in the function used to map the penalty p to the radius r of the rolling pin (black curve bottom panel Fig.1). Two functions to map the penalty to the radius of the rolling pin are then provided: a polynomial function and a sigmoid one. For polynomial mapping (black curve right panel Fig.1), the radius is given by

$$r(p, \lambda) = C(\lambda)[R + (R_{max} - R) \cdot p^{\nu}] \quad (2)$$

where $C(\lambda)$ is the chromatic law[3], ν is a real (positive) parameter specified directly with ***par_reg_nu***, and R, R_{max} are two parameters of the model: ***par_R*** and ***par_Rmax***. They define the minimum radius and the maximum one in wavelength units (Å). Both parameters must be specified at the minimum wavelength of the spectrum, since the values will be scaled by the C(λ) in the penalty law. The values for ***par_R*** and ***par_Rmax*** have to be provided by the user regarding the typical line width of its spectrum as well as the broadest absorption gap, but both parameters can be estimated in automatic mode (see appendix B). If $\nu > 1$, the penalty map is convex and only high penalties will modify the radius substantially. On the opposite, if $\nu < 1$, the map is concave and only low penalties will leave the radius unchanged. The second available law, the sigmoid one (see Fig.D.2 in appendix D for an example), allows to produce a two-radius regime due to its step-like shape. The transition from the smallest and largest radius is given by the sigmoid center μ and the smoothness of the transition is given by the sigmoid width ν. The best value for a given spectrum will depend on the shape on this latter. Smoothed spectra can be efficiently reduced with $\nu < 1$ whereas spectra with long oscillation due to instrumental effects or bad one dimensional reconstruction are reduced better with $\nu > 1$. A good value for several cases was found using the polynomial mapping and $0.7 < \nu < 1.3$. Note that this parameter is the only one for which no automatic algorithm is available.

2.2.5. Alpha shape (rolling pin)

This part is the main step of the code. Recall that we have now: a collection of selected local maxima, from which we would like to extract a continuum. Moreover, for each of this local maxima, a radius for the alpha shape was assigned in the previous part (Sect. 2.2.4). The idea is the following: the rolling pin starts at the first maxima in the smallest wavelength. It will then roll clockwise being anchored to this maxima until it touches another maxima. This maxima is then taken to be the next anchor point, the radius is updated and the code continues iteratively. Such an algorithm is called gift wrapping or Jarvis walk (Jarvis 1973) and will be presented below. Note that other algorithms of convex hull exist (Graham 1972; Eddy 1977; Phan 2007), but Jarvis walk remains the most intuitive one.

The alpha-shape, or more intuitively the rolling pin, is a circle that rolls on-top of the spectrum. However, if we take the raw spectrum, we will have a problem as flux units, normally expressed in analog-to-digital units (ADU) are much larger than the wavelengths ones, generally expressed in Å. We therefore have to renormalise the flux so that variations in the wavelength and flux directions are comparable. This is done by transforming the flux according to the following relation $f = f_{raw} \frac{\Delta\lambda}{\Delta f_{raw}}$, with f_{raw} the raw spectrum's flux, and $\Delta\lambda$ and Δf_{raw} the difference between the maximum and minimum wavelength and raw flux values, respectively. In other words, after the first scaling the following equality is satisfied : $\lambda_{max} - \lambda_{min} = f_{max} - f_{min}$.

The second stretching is controlled by the parameter ***par_stretching***, which is positive and larger than 1. Namely, the flux is rescaled one more time, leading to the relation :

$$\lambda_{max} - \lambda_{min} = \textbf{\textit{par\_stretching}} \cdot (f_{max} - f_{min}). \quad (3)$$

The scaling of the axes is critical for the success of the alpha shape step performed later. Indeed, an inappropriate value can lead to unsuitable normalisation where the continuum is not on-top of the spectrum, but is rather going through it (see Fig.2). Hopefully, a large range of values produces good enough results and wrong values are easily recognizable in the final product. The same conclusions was raised in Xu et al. (2019). Actually, the ***par_stretching*** parameter is quite similar by analogy to the tension of the veil covering the spectrum, smaller is the value of the parameter higher is the tension. We can wonder why the veil goes through the spectrum as in Fig.2 (blue curve). This can be explained considering that the veil is deployed from the left to the right and that the algorithm is only applied to local maxima. If the stretching parameter is too small, the good local maxima

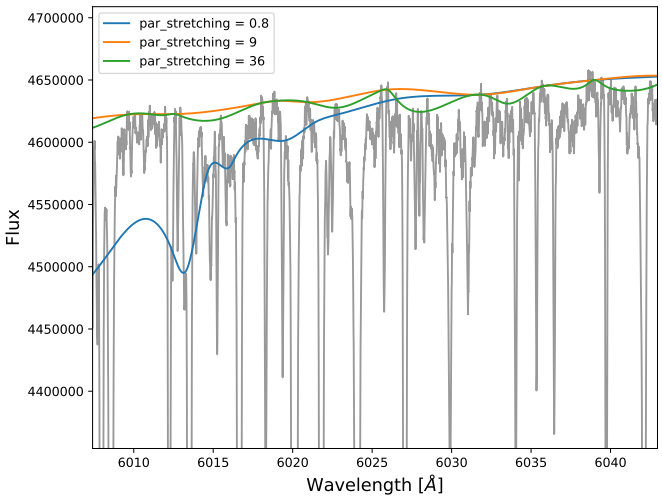

Fig. 2: Effect of different values for ***par_stretching*** on the normalisation of the spectrum (gray curve). When ***par_stretching*** is too low (blue curve), the rolling pin fall on the blended local maxima. As soon as ***par_stretching*** is sufficiently high (orange curve), the rolling provides a good normalisation. A wide range of values (orange and green curves) can provide a good enough normalisation under the 1% level.

[3] As line width are similar at first order in velocity space, the classical Doppler effect formula tells us that line width in wavelength-space will be a linear function of wavelength. Therefore, the radius of the rolling-pin in wavelength should satisfy the following chromatic law $C(\lambda) = \lambda/\lambda_{min}$

are unreachable and the rolling pin chooses a blended local maxima in order to keep rolling. Once the stretching is well selected, we roll the alpha-shape, or rolling pin, on top of the spectrum to select all the local maxima corresponding to the continuum. Details about this process, and how local maxima are selected is described in Appendix A.

2.2.6. Outliers detection

An upper envelope remains very sensitive to outliers with anomalous flux intensities. Even if until now, we took a lot of precaution to suppress as many of these points as possible, we perform a last check by looking at three types of outliers: 1) border-maxima, 2) high- or low-maxima and 3) close-maxima :

1. The rolling pin will fall on the borders of the spectrum on spurious maxima, simply because it starts and ends on them. If the spectrum was measured on an infinite wavelength grid, the continuum would touch other maxima which are actually not available. It is necessary to prevent the continuum from falling on those spurious maxima. However, we do not want these spurious maxima to be suppressed. Instead, the values of the j first and j last maxima are equalise to the value of the one with the index ***count_cut_lim*** + 1 and len(spectrum)−***count_cut_lim*** − 1. By default, ***count_cut_lim*** is set to 3, therefore RASSINE flattens the three first and three last points.
2. Sharp peaks are removed by computing the left and right derivative at each selected maxima, with respect to the two neighbouring ones. If a peak is too high, the left derivative will be positive and steep, while the right one will be negative and steep. By default, the five points with the highest absolute difference between the left and right derivatives are displayed by the code, and the user can choose to keep or delete them. If one point is suppressed, the next one with the highest absolute derivative difference will replace it. By default, RASSINE suppresses the points higher than the 99.5 percentiles of the distribution. The process can be run iteratively with the number of iteration controlled by the optional parameter ***count_out_lim***.
3. Close-maxima are problematic for the cubic interpolation used at the end (see Sect. 2.2.7) since interpolating on a non-equidistant grid can sometimes lead to weird behaviours. Moreover, since these points do not bring any relevant information they are removed automatically. To detect them the code computes the distance between each pair of neighbouring maxima still present in the final selection. Then, an asymmetric sigma clipping is performed on the distribution of the distance difference to identify outliers. For each group of detected close-maxima, the algorithm keeps the point which maximises the equidistance with the previous and next maxima.

2.2.7. Envelop interpolation

We are now left with several reliable local maxima, none or few of them being outliers. To generate the continuum, RASSINE interpolates linearly and cubically on them. Some points may still be too close, producing wiggles in somes cases due to the reason explained in Sect. 2.2.6 regarding item 3).

RASSINE also provides two other normalisation procedures (similar to the former ones). Such methods are necessary for noisy spectra for which the upper envelop is always more or less correlated with the noise level and thus the spectrum S/N. This tends to produce continuum slightly too high as already mentioned by Xu et al. (2019). In this case, instead of taking the flux value of the local maxima to build the continuum, an average of the flux in a window determined by the parameter ***denoising_dist*** (half-window) is taken. By default, the window is made of five wavelength elements. In summary, RASSINE produces four continuum (linear/cubic + denoised/undenoised) and the user is free to decide which of them is the best. In Sect. 3.4, we will show that in most cases the linear interpolation gives the best results, hence, we recommend to use this continuum. For some applications, it could be interesting to have a continuum differentiable everywhere and thus the cubic interpolated continuum might be better. An example of the RASSINE final result, performed in complete automatic mode (see appendix B), is shown for a HARPS solar spectrum in Fig.3 and for four other spectra in appendix E, Fig.E.1.

2.3. Comparison with the AFS code (Xu et al. 2019)

A similar alpha shape strategy has been developed by Xu et al. (2019) with a code called AFS and written in R, showing the higher performance of the alpha shape compared to iterative methods. Interestingly enough, our codes, which are rather similar in the approach, were not thought to be used on the same objects. Indeed, whereas the purpose of RASSINE was initially to normalise *1d* spectra, AFS was developed to remove the blaze function on the *2d* echelle-spectra. Both codes are then quite different in their conception.

Another important difference between the two codes is the fact that the size of our radius is varying along the wavelength axis (because of the penalty law), whereas their radius is fixed. As a consequence, even if their code contains only 3 parameters, they have necessarily to be fine-tuned depending on the spectral features present in the spectral order. As a comparison, RASSINE contains 6 parameters for the full spectrum, but 5 of them are calibrated and can be automatically adjusted to provide a first guess.

AFS is by nature more sensitive to broad absorption lines than RASSINE. Indeed, as an example, they showed in their Fig.10 that for the sodium doublet at 5890 Å, their alpha shape process selects the flux between the two lines as part of the continuum, whereas this location is in fact still a wing absorption, making both broad lines appear slightly shallower. The same issue occurred in our case for the solar spectrum, but we managed to deal with it by simply increasing the tension (see Sect. 3.1). This trouble is emerging in their case precisely because the authors work with a *2d* spectrum, which means that the flux is mixed with the blaze. As the interlines region is precisely situated at the maximum of the blaze, the alpha shape is forced to select the region as part of the continuum and only a post-correction can solve the issue.

2.4. Normalisation of a spectra time-series

When several spectra of the same star are considered, for instance during a series of consecutive nights, it is expected that the positions in wavelength of the local maxima forming the continuum are the same. Different local maxima from one night to the next would be due to instrument or atmospheric effects, but not from the star. Using different local maxima from one spectrum to the next will induce wiggles in spectra ratios or differences, which are clearly not desired (see blue curve bot-

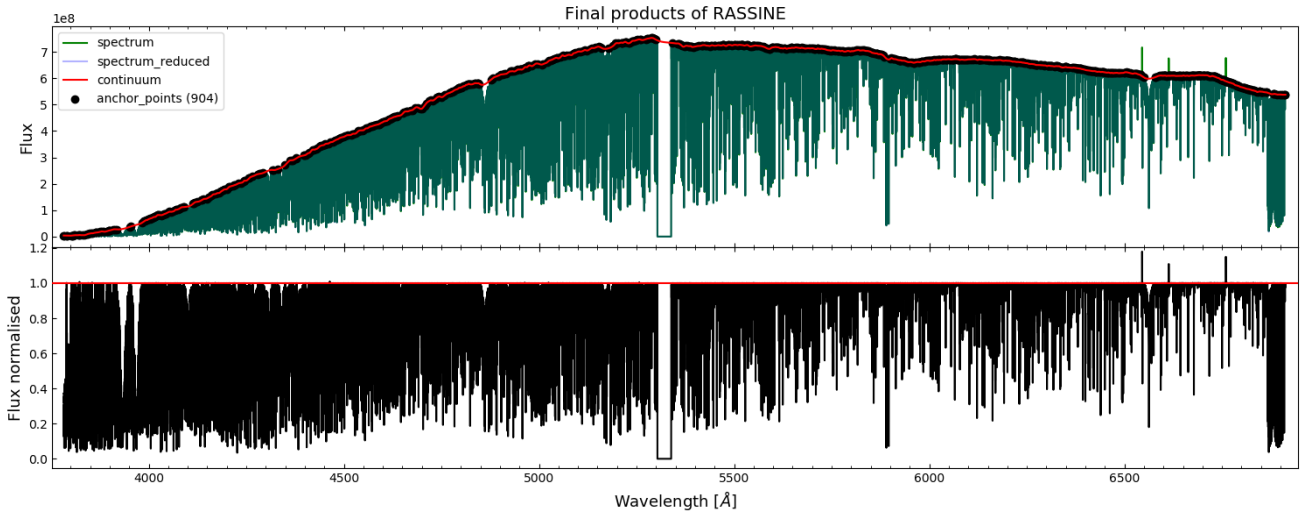

Fig. 3: Final plot interface. **Top:** The initial solar spectrum given as input to RASSINE (green curve). The continuum (red curve) as computed by RASSINE in complete automatic mode (see appendix B) based on a cubic interpolation of a selection of relevant local maxima also called anchor points (black dots). The continuum avoids three spurious cosmic peaks that we simulated around 6550, 6600 and 6750 Å and jumps above the large absorption regions. **Bottom:** Output continuum normalised by RASSINE.

tom panel Fig.4). These numerical artefacts are mostly produced by the cubic interpolation (blue curve bottom panel) and can already be dumped using the linear interpolation (blue curve top panel). Nevertheless, some of them are still present and need to be cleared.

To tackle this issue, RASSINE contains a function in its library (called *intersect_all_continuum*) which detects and forms clusters of local maxima. A set of local maxima with an occurrence rate over all the spectra higher than a given threshold is selected (see the graphical interface Fig.D.7). On the opposite, the local maxima which are selected only in a few number of nights are rejected. Also, if for one spectrum, two maxima are falling in the same cluster, the farthest one from the cluster center is removed. Maxima which are "missing" from the cluster are added. By doing this, we ensure that all the continuum are formed with the same anchor points which considerably reduces the wiggles (see green curves Fig.4). We note that, for the case where the user has to work with time-series of spectra at moderate S/N (from 75 up to 150), if the smoothing step is not efficient enough, there is a risk for the clustering algorithm to not work properly since the local maxima will be spuriously positioned. In this specific case, the user can follow the procedure described in appendix C.

A second function can be used (called *matching_diff_continuum*) if one wishes to reduce even more the wiggles. This latter searches for the spectrum with the highest S/N at 5500 Å and uses it as a reference. The spectra difference is computed with all the other spectra turn-by-turn, the wiggles are fitted on each spectrum difference using a Savitzky-Golay filtering (black curves Fig.4) and removed from the spectra (red curve Fig.4). The window of the filter is chosen in a graphical interface (see Fig.D.8). Note that because this step can also suppress true variations, we advise to use it with caution. Nevertheless, this option seems to produce the most precise spectra time-series (see Sect. 3.4 and Fig.8).

3. Results

3.1. Broad lines absorption

As a first test, we investigated if broad absorption lines were correctly normalised or if RASSINE was modifying their profile. To do so, we chose as test cases 61 Cyg A, the Sun and HD142 which are three stars probing a wide range of spectral-types and CCF FWHM values (see Table B.1 in appendix). The HARPS spectra were normalised in automatic mode (see appendix B) with intermediate tension (***par_stretching*** = 'auto_0.5') and polynomial law (ν = 1). We focused on two broad lines of first interest which are the sodium doublet NaD and H_α. In Fig.5, we compared the spectrum normalised by the alpha shape with a more classical method described as follows. Two spectral windows, as free as possible of stellar lines, were selected on the right and left side of each broad lines, each window being different for each star. The average flux was measured in each window before the continuum was estimated as being the straight line connecting these two obtained points.

On 61 Cyg A, which is a K5V star, we observe that the classical method leads to a too low continuum level for both broad absorption lines. This produces a normalised spectrum with several values higher than 1. This is due to the fact that, for cool stars, the regions selected as continuum are necessarily contaminated by absorption lines due to the high lines density. The normalisation produced by RASSINE appears to be better for this star, the continuum level being higher. Note that this latter could still be too low regarding the true stellar continuum, which is free of blended lines.

For the Sun, we observed at first a strong discrepancy of 2% for the sodium doublet between the RASSINE continuum, and the one derived using the classical method. We note that this second continuum is compatible with the Kitt peak (Wallace et al. 2011) and IAG (Reiners et al. 2016) solar atlases. This discrepancy can be explained by the fact that the interline region between the sodium doublet was considered by RASSINE as being part of the continuum in the automatic mode. To solve this issue, we increased the tension (***par_stretching*** = 'auto_0.0') and decreased the coefficient ν down to 0.7 for the polynomial map-

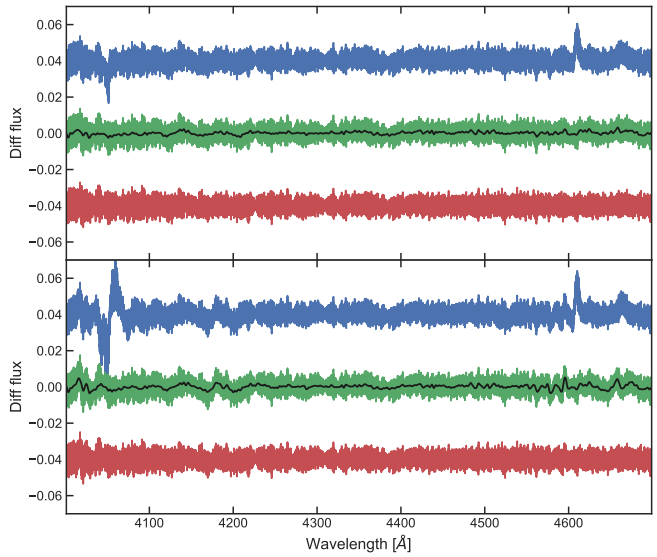

Fig. 4: Visualisation of the wiggles produced in a difference of normalised spectra (blue curve) with an arbitrary zero point. The two spectra come from the same star and are taken one night apart. **Top:** Difference between the spectra normalised by the linear interpolated continuum. **Bottom:** Same as top with the cubic interpolation. The wiggles are more important for the cubic spline as expected. These are induced when a local maxima (used to build the continuum curve) is selected in one spectrum but not in the other. We resolved this issue by developing a clustering algorithm on spectra time-series which can be called with the function *intersect_all_continuum*. The resulting effect improves considerably the stability (green curve). The remaining fluctuation (black curve) are under the 1% level. The residual curve after suppressing these fluctuations is indicated in red and can also be obtained from the RASSINE library function *matching_diff_continuum*.

ping. After doing so, the continuum found by RASSINE was similar to the one found using the classical method. However, both methods led to a smaller H_α width than the IAG atlas which could be explained by different solar activity level.

For HD142, a F7 dwarf, good agreement is found for H_α. However a discrepancy is found in the continuum obtained by the classical method in the sodium doublet. A clear quadratic drift is visible in the right part of the spectrum which is not induced by the method itself, since a linear fit was performed. Such low frequency variation are thus inherent to the *1d* spectrum and are either due to the instruments optics or produced during the *1d* construction. This latter option is more likely since this wavelength range is precisely situated in the overlapping region between the HARPS orders 56 and 57.

3.2. Evaluation of the accuracy of RASSINE

Even if not the primary interest of the RASSINE code, we tested the accuracy of the normalisation to evaluate if the derived lines depths are accurate. Since the alpha shape method implemented in RASSINE is an upper envelope approach, it is known that for low S/N spectra the continuum level may be too high (Xu et al. 2019), which will provide overestimated line depths. Obtaining accurate line depths is a difficult task. Probably the best method is to rely on stellar templates. However, those templates have been generated using a few atmospheric parameters, that are themselves derived from spectra that have been normalized in some way. Another strong difficulty is to obtain a stellar template with atmospheric parameters as close as possible to the observed star. Indeed, due to the high dimensionality of the stellar atmospheric parameter space, templates library often mainly focus on a good coverage of the effective temperature and gravity surface parameters, at the expense of abundances diversity. Also, the resolution of an observed spectra is generally not uniform, which could introduce wrong line depths since a stellar template is convoluted with a kernel at fixed resolution. As an example, for HARPS, the resolution is varying typically between 95′000 and 110′000 (private communication). For this reason, the computation of the accuracy cannot be performed by measuring a difference of line depth.

To measure the accuracy, we compared a spectrum normalised by RASSINE with a normalised synthetic spectrum of reference. We used the POLLUX database (Palacios et al. 2010) which is a library of spectra containing high resolution stellar templates normalised. We evaluated the RASSINE accuracy on two stars: α Cen B and the Sun. We used the spectra from the α Cen B 2008 data set, but same results are obtained with the 2010 dataset. The atmospheric parameters chosen for the templates are given in Table 1 and were derived from MARCS (de Laverny et al. 2012) and ATLAS (Kurucz 2005) models. For the Sun, it is also possible to use the IAG solar atlas (Reiners et al. 2016) in order to compare the accuracy of the alpha shape strategy versus iterative fitting methods. Solar atlas as well as synthetic spectra were degraded at the HARPS resolution, shifted to match the rest frame of the stellar spectra and interpolated on the same wavelength grid. Since the IAG spectra begins at 4047 Å, we removed shorter wavelength range from the HARPS spectrum. On the opposite, since HARPS spectrum end at 6910 Å, we suppressed longer wavelength in the IAG one.

To measure the accuracy, we extracted flux values of the normalised stellar template f_i at the same wavelength positions that the anchor points used to build our continuum. As by definition our continuum will take the value of 1 at these locations, the standard deviation of the continuum difference $1 - f_i$ provides a good metric for the RASSINE continuum accuracy. Note that because this metric is only based on the anchor points (selected local maxima), no conclusion can be raised about the accuracy value of the continuum between anchor points and thus the accuracy score for the linear and cubic interpolated continuum are the same. Considering that the continuum is a smooth function between two anchor points, which is a good approximation here as the average distance between consecutive anchors points is small (4.5Å), the derived accuracy can be considered as an average value representative of the alpha shape strategy for the whole spectrum between 4047 and 6910 Å. Such metric is however completely insensitive to a global offset between our continuum and the one from the template since such artefact will affect the mean and not the standard deviation of the distribution, but this is strongly unlikely for high S/N spectra. The accuracy was then extracted as the sigma-width of the distribution, taking from the percentile definition. The 1-sigma width being equivalent to the half distance between the 84^{th} and 16^{th} percentiles. The 2-sigma width is defined in the same way with the 97.5^{th} and 2.5^{th} percentiles and will be the default value when we will refer to *accuracy*

Additional spectral lines were sometimes present in the synthetic spectrum and thus produced easily identifiable outliers in the continuum difference (see gray dots in Figure.6). We re-

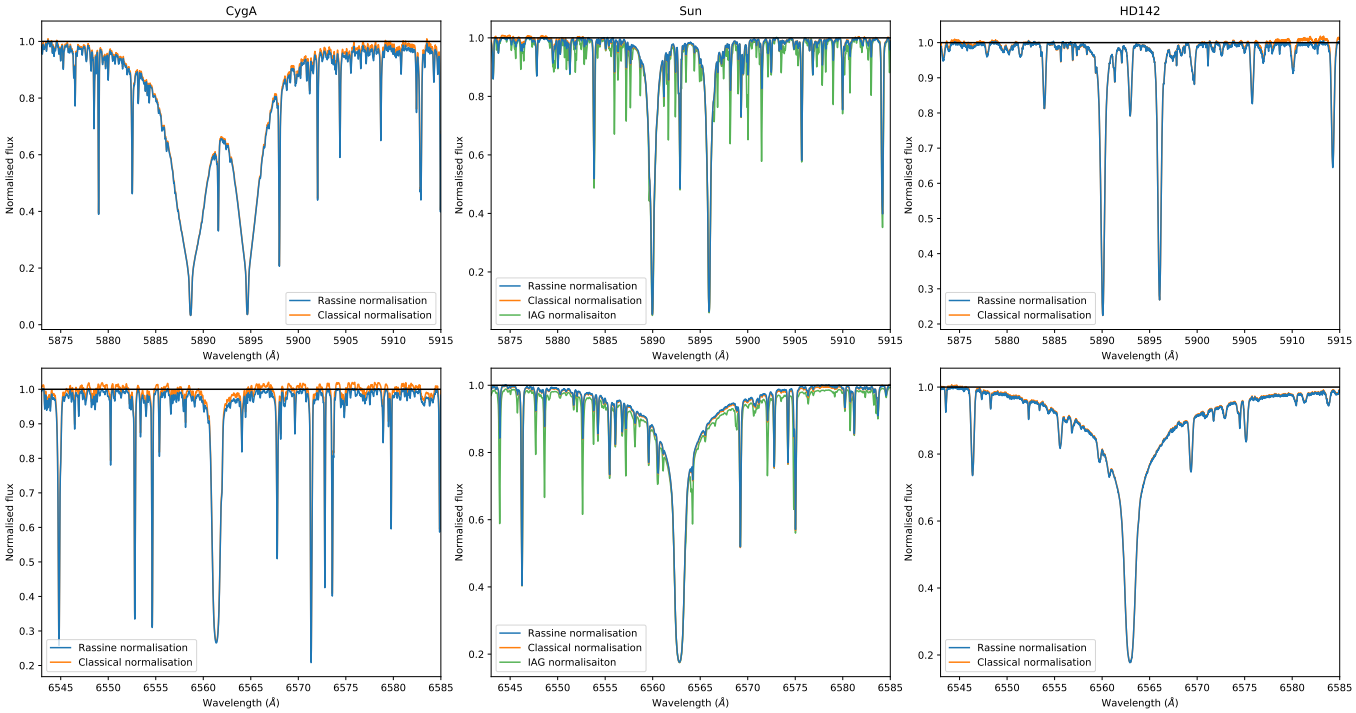

Fig. 5: Comparison between the RASSINE normalisation and classical method (linear fit) around spectral broad lines for three different spectral type stars : K5V, G2V and F7V, namely from left to right : α Cyg A, the Sun and HD142. **Top:** Sodium doublet NaD lines. **Bottom:** H_α line. Both α Cyg A and HD142 were normalised in automatic mode with intermediate tension (***par_stretching*** = 'auto_0.5') and polynomial mapping $\nu = 1$. For the Sun, higher tension was needed in order to not select the interline region as part of the continuum. High tension (***par_stretching*** = 'auto_0.0') and $\nu = 0.7$ appeared has satisfying. A comparison with the IAG solar atlas (Reiners et al. 2016) degraded at the HARPS resolution $R \sim 115'000$ is also displayed and show that ours normalisations are matching well the NaD lines, but lead to a thinner H_α width compared to the IAG. This could be explained by instrument systematics, different activity level or different continuum region selected to perform the normalisation. In any case, the continuum obtained with RASSINE is similar to the one using the classical approach on the same spectrum. The deeper lines observed in the IAG are tellurics lines.

jected those points by performing a rolling median on 20 adjacent points, and rejected all points in the residuals that were 10 times the median absolute deviation (MAD) further away from zero. This process was performed iteratively until no more outliers were detected. Also, because the IAG atlas was taken at a different BERV value, a few local maxima were situated in tellurics lines. We removed these points by performing the same outliers rejection that for the synthetic spectra.

Table 1: Atmospheric parameters chosen to generated our reference synthetic spectra. Spectra were generated either by a MARCS (M) or ATLAS (A) model specifying an effective temperature T_{eff} in K, a surface gravity $\log g$, a metalicity [Fe/H] and [α/Fe] value and a micro-turbulence velocity ξ_t in $\mathrm{km\,s^{-1}}$.

Star	Model	T_{eff}	$\log g$	[Fe/H]	[α/Fe]	ξ_t
α Cen B	M	5250	4.5	0.25	0	1
Sun	A	5800	4.5	0	0	2

We observe in Fig.6 that our distribution of continuum difference is always positive, meaning that the continuum fitted by RASSINE remains always below the synthetic continuum (in absolute flux) ; this is coherent with the alpha shape strategy. Since, the only possibility for our continuum level to be higher than the synthetic one would be that at least one of the local maxima selected as anchor point is similar to an "emission-peak" either induced by a cosmic or a contamination. However, as seen in Sect. 2.2.6, such kind of outliers should be removed during the cleaning process. Overall, RASSINE will tend to select too many anchor points as being part of the continuum, an effect that is enhanced when reducing the tension parameter. As an example, in the case of the continuum fitted to the solar spectrum, RASSINE considers the maximum of the inter-region between the CaII H & K lines as being part of the continuum, which is clearly not the case in the synthetic template due to the strong absorption in this region. This discrepancy produces the strong excursion of 0.35 around 3900 Å, seen for α Cen B in Fig. 6. Note that for the Sun, the discrepancy is as high as 0.2 at the same wavelength. This excursion is however not shown in this figure as the HARPS spectrum have been cut below 4047 Å to match the IAG solar atlas.

A summary of the accuracy scores can be found in Table. 2. The accuracy is lower for α Cen B in the blue part of the spectrum with respect to the Sun due to the CH molecular band at 4300 Å, also called G-band, and the CN violet molecular band at 3883 Å which are deeper in cool stars and sensitive to stellar activity (Berdyugina & Usoskin 2003). The accuracy for α Cen B is estimated around 2.0% which is 6 times worse than the accuracy for the Sun around 0.29%. Such a conclusion is expected since cooler stars present more blended lines, in addition to molecular bands, which implies less local maxima probing the

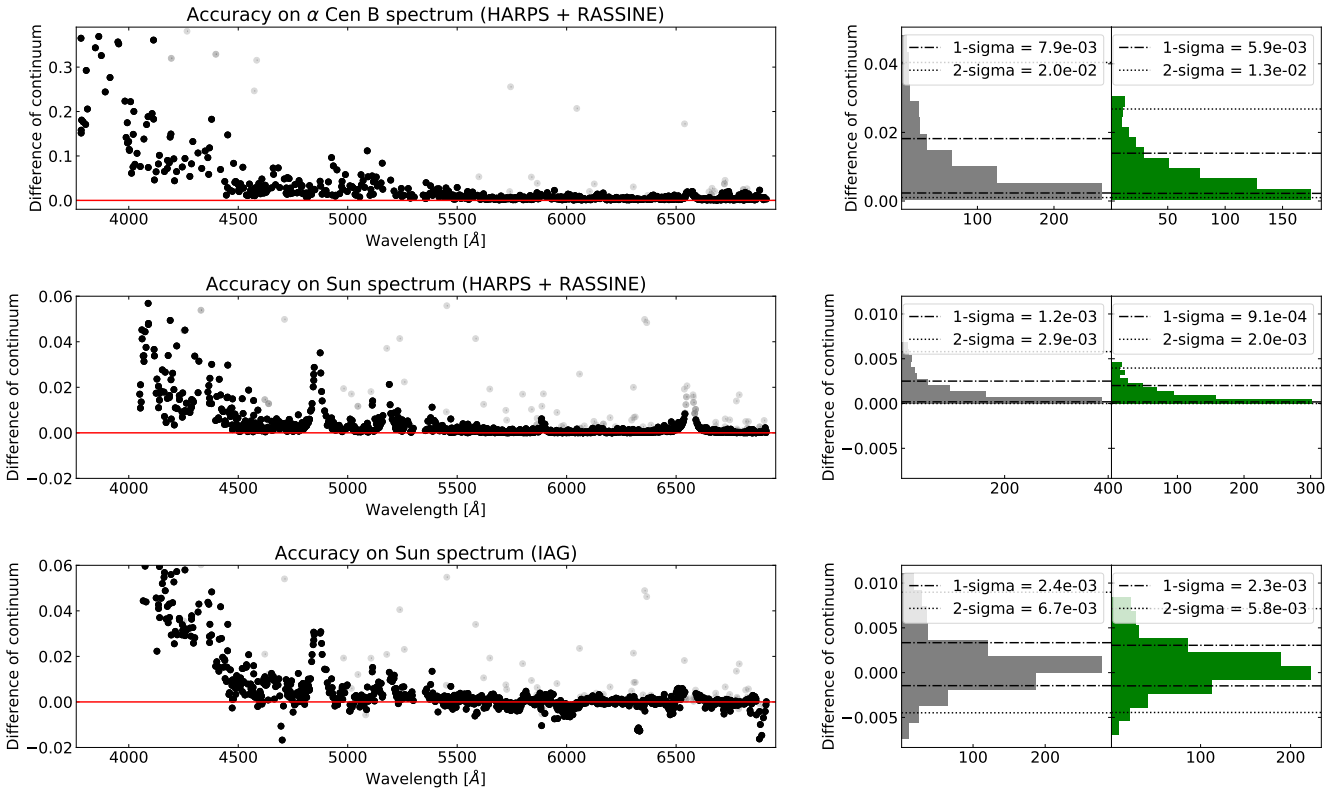

Fig. 6: Computation of the RASSINE accuracy by measuring the continuum difference level between our spectra and synthetic stellar templates. **First row:** Difference of continuum for α Cen B between the MARCS template and the normalised spectrum of RASSINE as a function of wavelength. A few continuum difference cannot be extracted correctly in the synthetic template due to the presence of stellar lines absent in our spectra (gray dots). The distribution of the continuum difference is plotted on the right as well as the distribution for $\lambda > 4500$ Å (green curve). **Second row:** Difference of continuum for the Sun between the ATLAS template and the normalised spectrum of RASSINE. **Third row:** Difference of continuum for the Sun between the ATLAS template and the IAG normalised spectrum.

continuum. For this reason, we also computed the accuracy score excluding wavelength shorter than 4500 Å. In this more restrictive range, the accuracy are respectively of 1.3% for α Cen B and 0.26% for the Sun.

When comparing the continuum of the IAG solar atlas, fitted using a low order polynomial fit, with the synthetic continuum, we see that the distribution of the difference gives values above and below zero, meaning that sometimes the flux continuum level is higher, and sometime lower than the reference. The accuracy found is also nearly two times worse that the one obtained with RASSINE, being 0.67% compared to 0.29%, respectively. Excluding the wavelength shorter than 4500 Å produces an accuracy of 0.58%, three times lower than the 0.20% of obtained with RASSINE. Finally, a clear excursion of 3% around H_β at 4861 Å is visible in both IAG and HARPS spectra which indicates that the difference is induced by the synthetic template which does not reproduce very well the wings of this spectral feature.

3.3. Evaluation of the precision of RASSINE

Another important score to determine is the code's precision. For this purpose, we normalised several very high S/N spectra of α Cen B, taken during successive nights. We selected 9 consecutive nightly-binned spectra at minimum activity in 2008, from BJD = 2′454′550 to BJD = 2′454′558. We also selected 13 consecutive nights in 2010, from BJD = 2′455′288 to BJD = 2′455′301, keeping in mind that this data set is known to be contaminated by a large active region (Dumusque et al. 2015) which could change the depth of spectral lines (Thompson et al. 2017; Wise et al. 2018). Since we are looking from night-to-night variation and that the rotational period is around 36 days (DeWarf et al. 2010), we expect the change induced by stellar activity to be rather small. This will be confirmed later since both data sets produce the same scores. We also selected 11 consecutive days of HARPS solar spectra from BJD = 2′458′507 to BJD = 2′458′518.

All the α Cen B nightly stacked spectra have similar S/N at 5500 Å ranging in 2008 (and 2010) from 1050 up to 1369 (and from 1293 up to 2390) respectively. For the Sun, all the daily stacked spectra have S/N ranging from 1951 to 5013. For each of them, the lines depths were measured and compared between adjacent nights or days. Because spectra are close in time, the line depths are not expected to change from one night to the next, thus the measured variation is a direct measurement of RASSINE's precision. The spectra were reduced in complete automatic mode with a Savitzky-Golay filtering, a ***par_smoothing_box*** of 6 wavelength elements, a polynomial penalty mapping ($\nu = 1.0$) and with an intermediate tension (***par_stretching*** = 'auto_0.5').

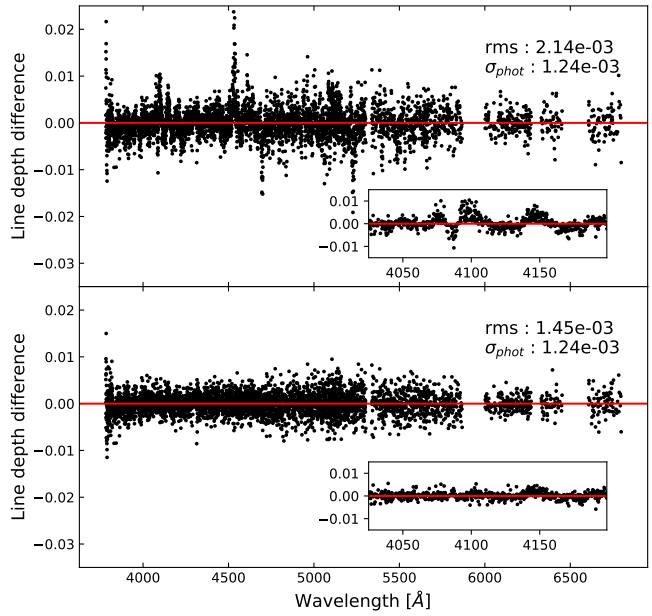

Fig. 7: Line depth difference between two α Cen B spectra take one night apart (BJD = 2′454′557 and BJD = 2′454′558) . The uncertainties, coming from the photon noise, are not displayed for sake of clarity but their median value σ_{phot} is indicated as well as the weighted RMS used to measure the dispersion. Both spectra are normalised by the RASSINE cubic continuum which is the least precise one (see Fig.8). **Top:** Spectra normalised by the cubic continuum. It matches the 8^{th} blue square data point in the left panel of Fig.8 and was the least precise one with a precision of 0.21%. Part of the jitter is not Gaussian, as displayed in the inner plot which is a zoom at 4100 Å where the discrepancy is observed to be about 1.0%. The structures observed are induced by the cubic spline interpolation. **Bottom:** Same as top after performing clustering and low frequency filtering (see Sect. 2.4), it matches the 8^{th} red square data point in the right panel Fig.8. The remaining jitter of 0.15% is dominated by the interference pattern (see Fig.9).

Line depths were estimated by fitting a parabola on the points present inside a ±2.5 km s^{-1} window around the core of each spectral line. The lines to be fitted were defined as in Cretignier et al. (2019) and all lines contaminated by a telluric by more than 2% were rejected. The telluric spectrum used was generated by Molecfit (Smette et al. 2015) as described in Cretignier et al. (2019). For the Sun, the stellar lines to be fitted were defined by the HARPS G2 mask. For each line, the uncertainty on its depth were derived considering photon noise.

Two examples of line depth differences between two adjacent nights are displayed in Fig.7. The precision is computed by measuring the weighted RMS of the line depth differences, and are displayed in Fig. 8. Due to the poor constrain of the interpolation on the edges of the spectrum, the weighted RMS was computed excluding the first and last 20 Å of the HARPS spectra. The photon noise is higher in 2008 than in 2010 and induces an average systematics on the line depth measurement of 0.12 % and 0.09 % respectively, whereas for the Sun this latter is of 0.05 %. A summary of the precision scores can be found in Table. 2. We observe that the precision is worse for the cubic interpolation than for the linear interpolation. For the former, the RMS significantly changes depending on the pair of nights studied which is related to the amount of wiggles produced in the spectra difference. In average, the precision in line depth of RASSINE for the linear interpolated continuum is found to be 0.17 % and 0.15 % for 2008 and 2010, respectively, approximatively 50% larger than photon noise. A value slightly better of 0.12 % is obtained for the Sun.

As shown in Fig.4 and Fig.7, the continuum fitted by RASSINE presents systematics which are not white noise but structured in wavelength. These structures are produced when a local maxima is selected in a zone where no local maxima were detected for the other spectrum (see Sect. 2.4) and constitute a major limitation to achieve high precision. We discuss in the next subsection ways to improve this precision further.

3.4. Evaluation of the precision of RASSINE for a spectra time-series

Until now, we have investigated the precision and accuracy of RASSINE for the normalisation of individual spectra. Each spectrum was thus reduced independently with respect to the others, without sharing any information between the different nights. However, if all the spectra are coming from the same star and thus formed a spectra time-series, it is possible to gather the information of all the continuum in order to improve the code's stability. We already presented this aspect in Sect. 2.4.

We performed the previous analysis of the line depth measurements on the 9 and 13 nightly adjacent spectra of the 2008 and 2010 α Cen B dataset as well on the 11 daily adjacent solar spectra, applying this time the *intersect_all_continuum* and *matching_diff_continuum* functions on the three data sets separately. A summary of the precision scores can be found in Table. 2. As described in Sect. 2.4, the former produces clusters of local maxima by identifying those with the highest occurrence rate whereas the second is a low frequency filter performed on the previous output (see Fig.4). The precision in line depth obtained after implementing these additional corrections are displayed in green and red in Fig. 8, respectively. Again the linear interpolation appears being more precise than the cubic one due to the wiggles produced by the latter. Performing the clustering on local maxima (i.e. applying the *intersect_all_continuum function*) improves significantly the precision on line depth down to 0.12%, and performing in addition the low frequency filetering (i.e. applying the *matching_diff_continuum function*) improves the precision even better down to 0.10%, a level compatible with the photon-noise for α Cen B. For the Sun, the precision is always higher than the stellar photon noise except the 8^{th} night pair for which the precision is compatible with it. This different behaviour is due to the fact than the same flat field was used for both consecutive days, since HARPS flat fields are themselves limited to S/N of ∼1000, they introduced the observed limit around 0.10% observed for both stars. Considering this, RASSINE is compatible with the photon-noise limit.

In addition, we note that regardless of the algorithm used to build the continuum of the two last pairs of night for α Cen B in 2008, the precision in line depth is sensitively worse than for the other pairs of night. After investigations, it turned out that 2008 HARPS spectra were contaminated by an interference pattern produced by a 1.5mm width filter placed in the parallel beam of the Tungsten lamp used to perform flat fielding. The periodicity $\Delta\lambda$ of the pattern at the wavelength λ is directly related

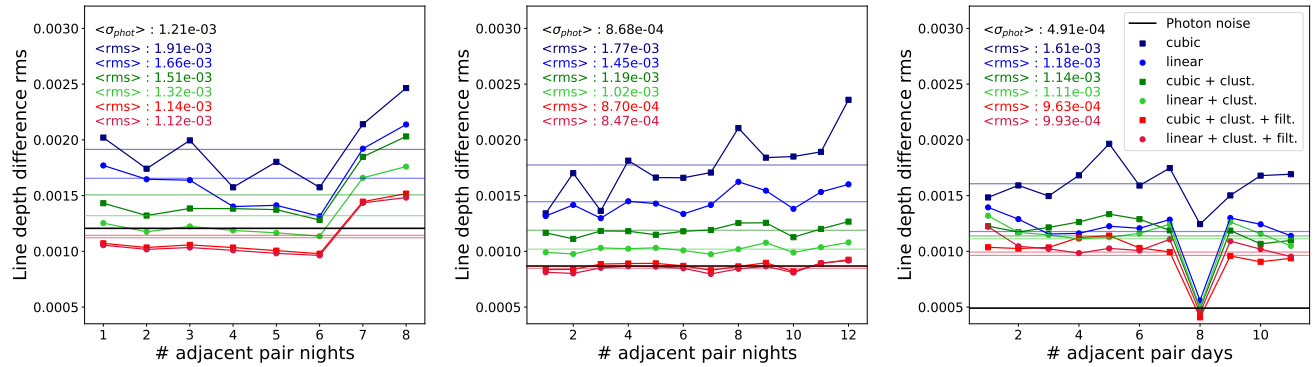

Fig. 8: Measurement of the RASSINE precision. A series of adjacent nights were selected. For each night, the line depth was computed on the full spectrum, rejecting telluric regions. The weighted RMS (color dots) was extracted from the distribution of the line depth difference where the difference is taken between two adjacent nights (see for example Fig.7). The average RMS $\langle rms \rangle$ on the night-to-night differences is measured (solid lines). The values of $\langle rms \rangle$ for each reduction is reported in the caption. As we can see, the best estimation of RASSINE, using a linear spline with clustering and filtering (see Sect. 2.4) gives a precision at the level of the photon noise. **Left:** 2008 α Cen B dataset (see the text for the precise dates). **Middle:** 2010 α Cen B dataset. **Right:** Sun dataset. The 8^{th} pair of days provides an exquisite precision compatible with the photon noise of the stellar spectra. After investigation, it turned out that the same flat field was used for both days. It revealed that the 0.10% precision limit observed for the other pairs of days is in fact the photon noise of the flat fields.

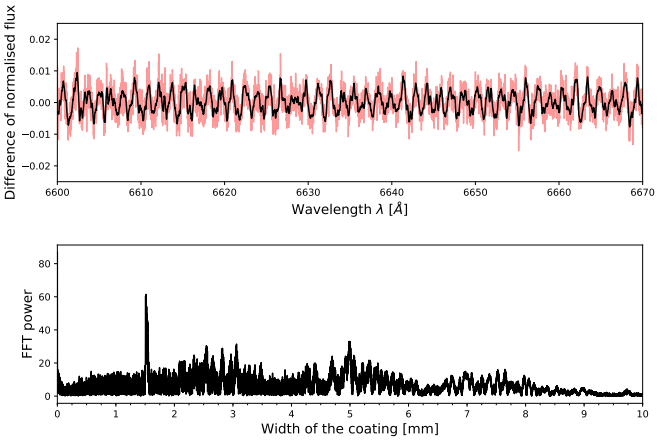

Fig. 9: **Top:** Zoom in the red 6640 Å region of the normalised spectra difference corresponding to the last red dot of the 2008 dataset shown in the left panel of Fig.8. A clear pattern of interference is visible in the spectra difference (red curve) even more noticeable once the difference is smoothed (black curve). **Bottom:** Fourier transform of the spectra difference. A power excess is visible and corresponds to an interference pattern created by an optical element of width 1.51 mm.

to the width w of the optical element producing the interference according to the law :

$$w = \frac{\lambda^2}{2\Delta\lambda}. \quad (4)$$

An upgrade of the instrument in August 2009 modified the position of this filter from parallel to diverging beam, which solved this issue. This pattern is affecting the flat field and is introduced in the stellar spectrum when correcting from it. We note that this pattern stays stable for couples of nights and then changes significantly, producing the interference pattern in the spectra difference seen in Fig.9 with a peak-to-peak of 1% in normalised flux units. In the data analysed from 2008, we note that the interference pattern stays stable for the 7 first nights, and then changes for night 8 and changes again for night 9. The variation of this interference pattern during the two last night is responsible for the degradation of the precision observed in Fig.8.

4. Conclusion

In this paper, we presented the concepts behind the RASSINE Python code, a tool developed to facilitate the normalisation of stellar spectra. The realisation of the code was motivated by the challenge to develop a coherent and robust normalisation algorithm of *1d* spectra, allowing to deal with different spectral types and S/N, presenting a limited number of free parameters to adjust. RASSINE proposes a graphical user-friendly interface (see appendix D) that allows the user to choose the best parameter values. In addition, calibrations based on the S/N level of the spectrum and the FWHM of the stellar CCF were performed to obtain first guesses for 5 of the 6 relevant RASSINE parameters (described in appendix B). The code has been tested with CORALIE, HARPS, HARPN and ESPRESSO, 4 instruments spanning different instrumental resolutions and has produced each time visually satisfying results. In this paper, we focused on high S/N spectra (S/N > 500) to test the accuracy and precision of RASSINE, which are by hypothesis the objects for which alpha shape strategy is expected to be efficient. This does not mean that spectra with a lower SNR cannot be reduced with RASSINE. If the smoothing step is performed correctly, the code can theoretically perform well on low S/N spectra (S/N < 100). However, due to the nature of the algorithm, finding the continuum for spectra with S/N lower than 50 remains challenging.

We tested the accuracy and the precision of the automatic mode of the code on the 2008 and 2010 α Cen B data set, keeping in mind that no guaranty is given for this mode to provide the best values of accuracy. We showed that the linear interpolation provided better continuum that the cubic interpolation in all cases since this latter is producing unwanted artefacts. The accuracy on the continuum level was measured on the 2008 α Cen B

Table 2: Summary of the different scores reached with RASSINE's automatic mode. Only the accuracy of continuum level and precision on line depth obtained using the linear interpolated continuum are displayed. We also show for comparison with precision the average error on line depth induced by photon noise. The three values of precision are respectively for individual spectra normalisation, spectra normalised with clustering and clustering + filtering. All the values are in %.

Scores	α Cen B 2008	α Cen B 2010	Sun	Sun (IAG)
Accuracy	2.0	2.0	0.29	0.67
Accuracy ($\lambda > 4500$Å)	1.4	1.2	0.20	0.58
Precision	0.17	0.14	0.12	-
Precision (clust.)	0.13	0.10	0.11	-
Precision (clust. + filt.)	0.11	0.08	0.10	-
σ_{phot}	0.12	0.09	0.05	-

data set and on solar spectra by comparing our continuum with stellar template from MARCS and ATLAS model. An accuracy of 2.0% was derived for the former star, whereas a value a 0.29% was found for the Sun, 2.3 times better than the 0.67% obtained with low polynomial fit of the IAG solar atlas. By considering wavelengths longer thanb 4500 Å, the accuracy of the spectra increases up to 1.3% and 0.20%, three times better than for the IAG around 0.59%. The algorithm's precision on line depth was found to be better than the accuracy. For α Cen B, the precision on the 2010 data set which do not contain the interference pattern is about 0.14%. Additional algorithms were developed in order to stabilised even more the continuum, in particular when the goal is to study a spectra time-series. A clustering algorithm, allowing to always select the same local maxima on all spectra, improved the precision down to $\sim$ 0.12%. The precision can be boosted after applying a low frequency filtering on spectra difference which seems to produce a precision limit for RASSINE around $\sim$ 0.10%, a value reached for α Cen B and solar spectra. This limit is the photon noise limit of Alpha Cen B spectra, and the photon limiit of the flat fields for the Sun. Therefore, we can conclude that RASSINE normalisation was always found compatible with the dominant photon-noise level. All the accuracy and precision scores are summarized in Table 2.

Such tool can find applications in numerous situations such as computations of stellar atmospheric parameters (Blanco-Cuaresma 2019), the development of tailored cross-correlation mask for each star in the context of radial velocities (Bourrier et al. 2020), line-by-line variability related to stellar activity (Cretignier et al. 2019) or to correct the interference observed in transmission spectra. Such normalisation algorithm could also improve the RV derived from high-resolution spectra, reducing the jitter in RV time-series by providing a better color correction and facilitating further correction algorithms to mitigate stellar activity or telluric contamination.

5. Acknowledgments

We are grateful to Nathan Hara and his constructive comments. We thanks Lila Chergui and Yannick Demets for their help regarding English. This work has made use of the VALD database, operated at Uppsala University, the Institute of Astronomy RAS in Moscow, and the University of Vienna, and data coming from the ESO archive (Alpha Cen B) and from the HELIOS solar telescope at La Silla Observatory. This work has been carried out within the frame of the National Centre for Competence in Research "PlanetS" supported by the Swiss National Science Foundation (SNSF). M.C, J.F and R.A. acknowledges the financial support of the SNSF. F.P. greatly acknowledges the support provided by the Swiss National Science Foundation through grant Nr. 184618. X.D is grateful to the Branco-Weiss Fellowship for continuuous support. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 851555).

References

Adibekyan, V., Delgado-Mena, E., Figueira, P., et al. 2016, A&A, 592, A87
Allart, R., Lovis, C., Pino, L., et al. 2017, A&A, 606, A144
Asaeedi, S., Didehvar, F., & Mohades, A. 2013, arXiv e-prints [arXiv:1309.7829]
Baranne, A., Queloz, D., Mayor, M., et al. 1996, A&AS, 119, 373
Berdyugina, S. V. & Usoskin, I. G. 2003, A&A, 405, 1121
Blanco-Cuaresma, S. 2019, MNRAS, 486, 2075
Blanco-Cuaresma, S., Soubiran, C., Heiter, U., & Jofré, P. 2014, A&A, 569, A111
Bouchy, F., Doyon, R., Artigau, É., et al. 2017, The Messenger, 169, 21
Bourrier, V., Ehrenreich, D., Lendl, M., et al. 2020, arXiv e-prints, arXiv:2001.06836
Bourrier, V. & Hébrard, G. 2014, A&A, 569, A65
Cosentino, R., Lovis, C., Pepe, F., et al. 2012, in Proc. SPIE, Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV, 84461V
Cretignier, M., Dumusque, X., Allart, R., Pepe, F., & Lovis, C. 2019, arXiv e-prints, arXiv:1912.05192
de Laverny, P., Recio-Blanco, A., Worley, C. C., & Plez, B. 2012, A&A, 544, A126
DeWarf, L. E., Datin, K. M., & Guinan, E. F. 2010, ApJ, 722, 343
Dumusque, X., Glenday, A., Phillips, D. F., et al. 2015, ApJ, 814, L21
Eddy, W. F. 1977, ACM Trans. Math. Softw., 3, 398
Fischer, D., Jurgenson, C., McCracken, T., et al. 2017, in American Astronomical Society Meeting Abstracts, Vol. 229, American Astronomical Society Meeting Abstracts #229, 126.04
Graham, L. 1972, Information processing letters
Gray, D. F. 2005, The Observation and Analysis of Stellar Photospheres
Jarvis, R. 1973, Information processing letters
Kurucz, R. L. 2005, Memorie della Societa Astronomica Italiana Supplementi, 8, 14
Malavolta, L., Lovis, C., Pepe, F., Sneden, C., & Udry, S. 2017, MNRAS, 469, 3965
Mayor, M., Pepe, F., Queloz, D., et al. 2003, The Messenger, 114, 20
Palacios, A., Gebran, M., Josselin, E., et al. 2010, A&A, 516, A13
Pasquini, L., Cristiani, S., García López, R., et al. 2010, in Proc. SPIE, Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III, 77352F
Pepe, F., Bouchy, F., Queloz, D., & Mayor, M. 2003, in Astronomical Society of the Pacific Conference Series, Vol. 294, Scientific Frontiers in Research on Extrasolar Planets, ed. D. Deming & S. Seager, 39–42
Pepe, F., Mayor, M., Galland, F., et al. 2002a, A&A, 388, 632
Pepe, F., Mayor, M., Rupprecht, G., et al. 2002b, The Messenger, 110, 9
Pepe, F., Molaro, P., Cristiani, S., et al. 2014, Astronomische Nachrichten, 335, 8
Phan, T. 2007, Annales Mathematicae et Informaticae
Queloz, D., Mayor, M., Weber, L., et al. 2000, A&A, 354, 99
Reiners, A., Mrotzek, N., Lemke, U., Hinrichs, J., & Reinsch, K. 2016, A&A, 587, A65
Savitzky, A. & Golay, M. J. E. 1964, Analytical Chemistry, 36, 1627
Schwab, C., Rakich, A., Gong, Q., et al. 2016, in Proc. SPIE, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, 99087H
Smette, A., Sana, H., Noll, S., et al. 2015, A&A, 576, A77
Sousa, S. G., Santos, N. C., Adibekyan, V., Delgado-Mena, E., & Israelian, G. 2015, A&A, 577, A67

Strassmeier, K. G., Ilyin, I., Järvinen, A., et al. 2015, Astronomische Nachrichten, 336, 324
Thompson, A. P. G., Watson, C. A., de Mooij, E. J. W., & Jess, D. B. 2017, MNRAS, 468, L16
Tody, D. 1986, in Proc. SPIE, Vol. 627, Instrumentation in astronomy VI, ed. D. L. Crawford, 733
Tody, D. 1993, in Astronomical Society of the Pacific Conference Series, Vol. 52, Astronomical Data Analysis Software and Systems II, ed. R. J. Hanisch, R. J. V. Brissenden, & J. Barnes, 173
Upton, G. & Cook, I. 1996, Oxford University Press, 443
Škoda, P. 2008, in Astronomical Spectroscopy and Virtual Observatory, ed. M. Guainazzi & P. Osuna, 97
Wallace, L., Hinkle, K. H., Livingston, W. C., & Davis, S. P. 2011, ApJS, 195, 6
Wise, A. W., Dodson-Robinson, S. E., Bevenour, K., & Provini, A. 2018, AJ, 156, 180
Wyttenbach, A., Ehrenreich, D., Lovis, C., Udry, S., & Pepe, F. 2015, A&A, 577, A62
Xu, X., Cisewski-Kehe, J., Davis, A. B., Fischer, D. A., & Brewer, J. M. 2019, arXiv e-prints, arXiv:1904.10065

Appendix A: Alpha shape algorithm

Rolling the alpha-shape, or rolling pin, on-top of the spectrum given our conditions is only an issue of trigonometry. Assuming our rolling pin is situated on a local maximum (see Fig.B.1), we first obtain all the local maxima (to the right) such that the distance d to the current local maximum satisfies $d < 2r$. If no such points exist, the radius is increased by a factor 1.5 until a close enough point is found. Let's call the current point P and the potential next point N. We compute the vector going from P to N, called $\boldsymbol{\delta}$ whose norm is δ. We want to find the coordinates of the center C of the circle of radius r (the current value), touching by P and N. The distance h between this center and the segment δ is given by Pythagora's theorem

$$r^2 = h^2 + \frac{\delta^2}{4}. \tag{A.1}$$

The vectorial components of the center C are then given by :

$$\boldsymbol{C} = \boldsymbol{P} + \frac{1}{2}\boldsymbol{\delta} + \frac{h}{\delta}\bar{\boldsymbol{\delta}}. \tag{A.2}$$

The second term goes from P to the middle of the segment $\boldsymbol{\delta}$, say the point M. Note that PCN is isosceles and CM is perpendicular to PN. The last term of the equation is a vector perpendicular to δ that is normalised to have norm h. In components terms, $\bar{\boldsymbol{\delta}} = (-\delta_y, \delta_x)$.

Once the coordinates of the center are found, we need to compute the rotation angle, i.e. the angle the rolling pin has to roll to touch this particular point. Let (p_x, p_y) and (c_x, c_y) be the coordinates of the current point and of the circle center, respectively. Trigonometric considerations show that this angle is given by

$$\theta = \begin{cases} -\arccos\left(\frac{c_x - p_x}{r}\right) + \pi, & \text{if} \quad c_y - p_y \geq 0 \\ -\arcsin\left(\frac{c_y - p_y}{r}\right) + \pi, & \text{if} \quad c_y - p_y < 0. \end{cases} \tag{A.3}$$

This angle is computed for every candidate points, namely the ones that are closer than $2r$ to the current point (N and N' on the figure). The next selected maximum is the first candidate touched by the rolling pin, mathematically speaking the one with the smallest θ, N in this case. This process goes on iteratively until the code reaches the end of the spectrum.

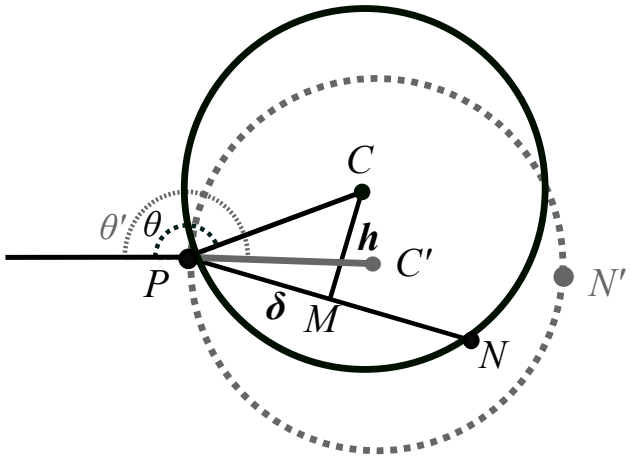

Fig. B.1: The rolling pin rotates around the point P. N and N' are two local maxima. The selected one is N, because the rolling pin falls on it before falling on N'. Mathematically speaking, this is represented by the condition $\theta < \theta'$.

Appendix B: Description of the automatic procedure

Even if RASSINE is a code with a complete interactive Python interface taking advantage of widgets and sliders, making the choice of the parameters as easy as possible, it can still be technical for an unfamiliar user to find the best parameters when dealing with several spectra of different stars having different spectral type or several spectra of the same star with different S/N. Hopefully, these parameters can often be approximated directly from the spectra. All of them, except the penalty law, can be replaced by the keyword "auto". In this case, several algorithms described below, can automatically guess initial values. The reader must still keep in mind that there is no guarantee regarding the quality of the final product in automatic mode. Such analysis should only be used to help the user getting a first guess for the continuum, or when a lot of spectra of the same star have to be reduced. Also, even if theoretically nothing prevent RASSINE to be used on *2d* echelle spectra, the user has to be aware that the automatic mode presented below was calibrated on *1d* spectra.

The first information one can get from a spectrum is the typical line width which is determined by the instrumental resolution, projected stellar velocity and macroturbulence (Gray 2005). As said previously, the easiest method to determine its value consists in computing the FWHM of the CCF which represents roughly an average line profile. If the parameter ***CCF_mask*** of RASSINE is put in "master" mode, the code computes the CCF and its corresponding FWHM. The first step is to build the cross-correlation mask that will then be used to produce the CCF by correlating this mask with the stellar spectrum. This is done by first getting a rude estimation of the spectrum continuum using a rolling maximum in a window of 30 Å. Because of the step-like shape of the obtained continuum, this latter is smoothed with a rectangular kernel of 15 Å. We then find all the local minima with a derivative criterion on a Savitsky-Golay smoothed version of the spectrum. This local minima corresponds to spectral lines and the difference between these local minima and the neighbour fitted continuum, provide an approximate depth for each spectral lines. The final mask is a collection of the wavelength of each minimun with its respective depth, as defined in Pepe et al. (2002a). The CCF obtained from cross-correlating the mask with the stellar spectrum is then fitted with a Gaussian and the FWHM extracted. A list of wavelength bands can be given as parameters ***mask_telluric*** in order to exclude some regions contaminated by telluric lines from the obtained CCF mask. A mask can also be given directly as input in the ***CCF_mask*** parameter, in which case the user also has to specify the systemic velocity of the star in order have the stellar spectrum and binary mask in the same rest frame. Due to more complexity, we do not recommend to use this option for users unfamiliar with RASSINE.

Once the FWHM is known, the first step consists in smoothing the spectra. If the ***par_smoothing_box*** parameter is set to "auto", a high frequency filter is applied in Fourier space to suppress the frequencies above the sigma width. Two filters are available: an error function and a top-hat with an exponentially

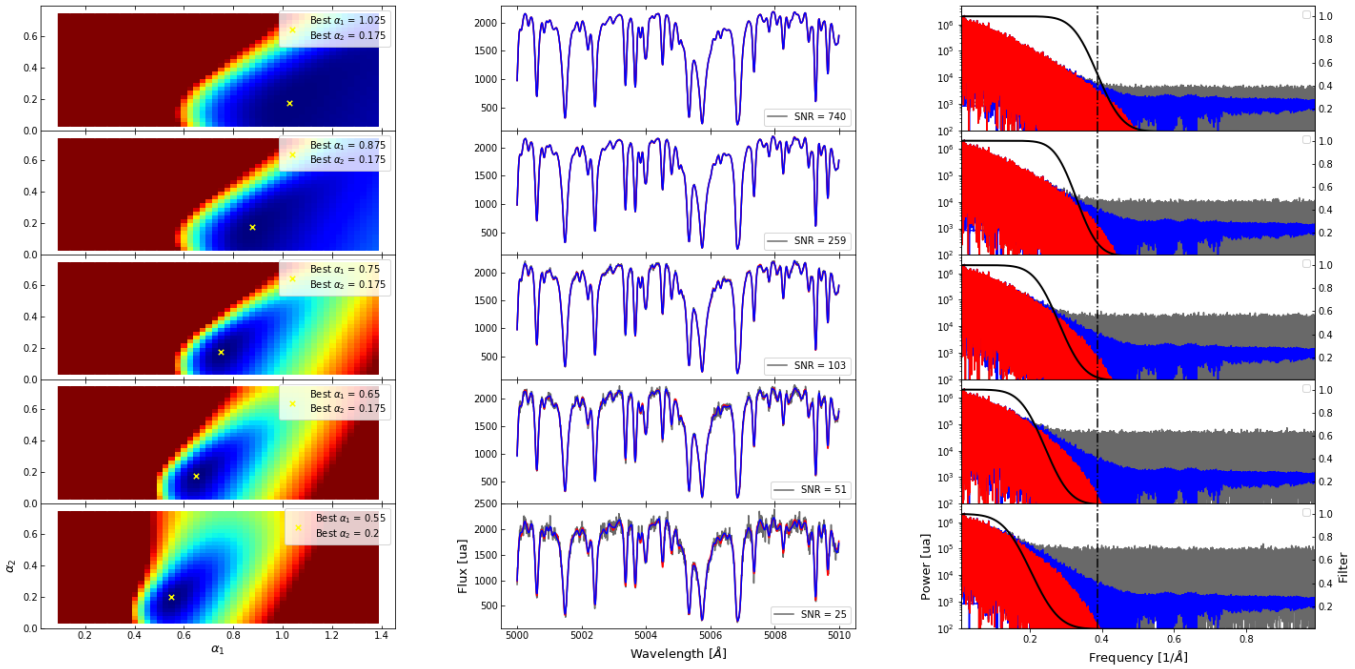

Fig. B.2: Calibration of α_1 and α_2 as a function of the spectrum S/N for the *erf* Fourier filter. The high S/N (~ 12000) spectrum of α Cen B was considered as noise-free, before injecting photon noise. **Left:** (α_1, α_2) parameter space for different S/N realisations. The simulations were run on a grid of evenly spaced value of α_1 and α_2. The colors encode the metric used here, namely the standard deviation between the smoothed and noise-free spectrum. The local minimum of the parameter space is indicated by a yellow cross. Noisier is the spectrum, lower is the best α_1. The best value for α_2 appears rather constant around 0.175. **Middle:** A high S/N spectrum (blue curve) is considered to be noise-free. Different realisations of Poisson noise are generated (gray curves) in order to determine the best couple (α_1,α_2) whose corresponding smoothed curve (red curve) will be the most similar to the original one. Only a subpart of the spectra is displayed for graphical considerations. **Right:** Fourier transform of the spectra of the middle column (blue curve). The high frequency noise (gray curve) increases significantly from $S/N = 740$ to $S/N = 25$ such that the center of the *erf* function is moving toward shorter frequencies. The *erf* filter corresponding to the best (α_1,α_2) couple is displayed (black curve). The red curve is the Fourier transform after smoothing by the function. The typical frequency scale σ^{-1}, where sigma is the width of the Gaussian fitted to the stellar CCF, is shown as a dashed vertical line.

decreasing tail. Their expressions are given by

$$f_{erf}(\omega) = C\ \mathrm{erf}\left(\frac{\omega_0-|\omega|}{\delta\omega}\right)$$

$$f_{hat}(\omega) = \begin{cases} 1 & |\omega| \leqslant \omega_0 \\ e^{-\frac{|\omega|-\omega_0}{\delta}} & |\omega| > \omega_0 \end{cases}, \quad \text{(B.1)}$$

where C is a normalisation constant (the filters should all satisfy $f(0) = 1$), ω_0 is the center of the filter which can be understood as the cut-off from which high-frequency modes are suppressed and $\delta\omega$ is the width of the filter (a small value of $\delta\omega$ will produce a sharp filter, while a high value will produce a smooth one).

By default, the error function filter is used if the ***par_smoothing_kernel*** parameter is set to "auto". Some care is then needed to get the correct values for ω_0 and $\delta\omega$. As mentioned before, a typical wavelength scale is given by the width of the Gaussian fitted on the CCF. This implies the existence of a given frequency scale σ^{-1}. Hence, we can parametrise $\omega_0 = \alpha_1\sigma^{-1}$ and $\delta\omega = \alpha_2\sigma^{-1}$, where the α_i's are two dimensionless parameters. To adjust them to the best values, a calibration curve was constructed using a high S/N (~ 12000) spectra of α Cen B and of the Sun, that were constructed by stacking a thousand of observations together. We took care of correcting the spectra from long RV trends in both cases in order to allow an optimal stacking.

We then simulated different S/N spectra by adding several levels of Poison noise in the spectrum (see Fig.B.2). The Fourier filter was performed in an optimisation grid of α_1 and α_2. The best values of the α's parameters were determined by measuring the standard deviation in the spectra difference between the noise-free and smoothed spectrum. By minimizing this quantity, we find the best values for α_1 and α_2 for each S/N value, which is extrapolated for every value of the S/N (see Fig.B.3) by fitting a polynomial function. The error bars $\Delta\alpha$ were defined as $\Delta\alpha = max_j(\alpha_j) - min_j(\alpha_j)$ with j the collections of the 5 percentiles best simulations. We see that the center of the filter is more important than its width since a clear dependence is observed for α_1 but not for α_2. If the automatic mode for the spectrum smoothing is chosen, the flux units of the input file has to be in analog to digital (ADU) units such that the S/N value can be extracted by taking simply the square root of the flux. We advise to use this automatic smoothing mode only if several spectra which span different order of magnitude in S/N have to be reduced.

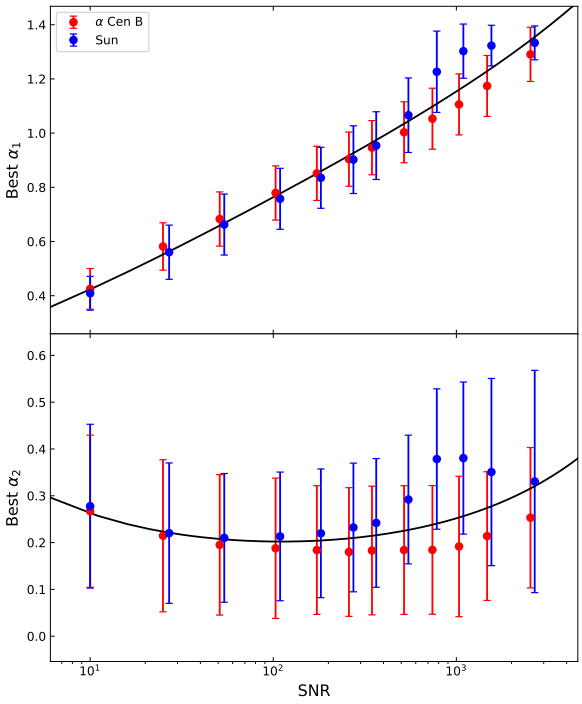

Fig. B.3: Calibration curves of the α_1 and α_2 parameters depending on the S/N value of a spectrum for the erf Fourier filter. **Top:** Calibration of α_1. The dots represent the best value according to the simulations (see Fig.B.2) which produce the least standard deviation (weighted, as explained in the text) between the noisy-free spectrum and the smoothed one. The error bars correspond to the 5% best values of the simulations. Two stars were used for the calibration (highlighted by the color of the points). The calibration curves were fitted by a third degree polynomial function in log-log space. **Bottom:** Calibration of α_2 with the same procedure. The calibration curves were fitted by a second degree polynomial function in log-log space. Clearly the width of the erf filter appears as less relevant than its center.

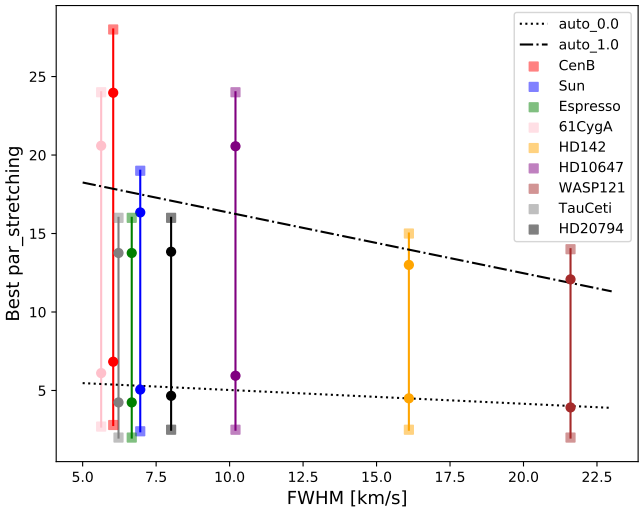

Fig. B.4: Calibration curves of the ***par_stretching*** parameter depending on the CCF FWHM value derived for a given spectrum. The calibration has been established using 9 stars (see Table B.1) probing the spectral type range from F6 to K5 (each star is indicated with a different color). The square dots indicate the minimum and the maximum values for the ***par_stretching*** parameter (chosen by eye). The circles dots are more "conservative" (16% and 84% of the parameter range). We fit the latter on a line to get the minimum value in automatic mode (***par_stretching*** = 'auto_0.0', dotted) and the maximum of the parameter (***par_stretching*** = 'auto_1.0', dash-dotted) as a function of the FWHM. The final choice is then let at the user discretion. Note the near constant line for the minimal value, and the decreasing line for the maximal value.

Once the spectrum has been smoothed, we need to stretch the flux units by rescaling the y axis with respect to the x-axis, which can be done by adjusting the ***par_stretching***. This parameter can be compared to a "tension" applied on the veil representing the stellar continuum. If the stretching parameter is too small, there is too much tension on the veil and the rolling pin can go through the spectrum as the blue curve in Fig.2 (we recall that the rolling pin considers only the local maxima, and there is no spectrum from its "point of view"). Similar issues can occur if the maximum radius for the rolling pin is too small. On the opposite, if the parameter is too large, the tension is not strong enough and the veil starts to fall inside the lines of the spectrum as the green curve in Fig.2.

There is a priori no precise way to determine the best value for this parameter. The calibration is presented in Fig.B.4 and was performed by eye with high S/N spectra of 9 stars listed in Table B.1. We need a typical length scale in order to calibrate the parameter of the stretching. The only such quantity is again the FWHM of the CCF. Furthermore, as raised previously, there is a large range of values working quite well to normalise the spectra, depending on the amount of tension we wish. Hence, the calibration provides, for a given value of the FWHM, a range over which the parameter can be taken. We fit two lines, one to get the minimum value in automatic mode (***par_stretching*** = 'auto_0.0') and one for the maximum value of the parameter (***par_stretching*** = 'auto_1.0'). The user has thus to specify after the auto keyword the level of tension in its continuum by also entering a number between 0 and 1. Note that 0 represents a strong tension whereas 1 represents a weak one. By default, an intermediate tension is used (***par_stretching*** = 'auto_0.5').

By observing Fig.B.4, one can notice interesting features. The minimal value of the parameter seems to be rather constant for all the stars, regardless of the value of the FWHM. This is the case because the flux of all the star are normalised in a same way, by scaling x and y axis on a same length. Recall that if the value of the parameter is too small (high tension on the veil), the rolling pin will go through the spectrum by reaching unsuitable local maxima (corresponding to blended lines). Those blended lines are present for most of the stars, either because of high stellar lines density or high rotational broadening. Hence a minimal value of 2 was found to be the same lower limit for all the stars. Regarding the upper value for the parameter, it decreases when the FWHM increases. If the value of the parameter is too high (small tension on the veil), the code will fall into absorption lines. Hence, if the value of the FWHM is already high, there is not much room to stretch the horizontal axis before having lines that are broad enough to make the rolling pin fall. Hence, the value of the parameter has to be smaller as the FWHM gets bigger.

The last two automatic parameters are the minimum and maximum radius of the rolling pin, R and R_{max}. The former is fixed as 100 times the σ value of the CCF transformed in Å with

Table B.1: Table of the stars and of the master spectra used to calibrate the ***par_stretching*** parameters. The stars were chosen to probe different instrumental resolutions and spectral types ranging respectively from $R \sim 55'000$ to $140'000$ and from F6V to K5V. The table contains from left to right: the name of the star, its spectral type, the number of exposures stacked to form the master spectrum, the S/N of the master spectrum at 5500 Å, the instrument which observed the star and the average date of the observations. At the end are displayed three parameters determined in automatic mode by RASSINE: the FHWM of the CCF in $\mathrm{km\,s^{-1}}$, the ***par_stretching*** with intermediate tension (***par_stretching*** = 'auto_0.5') and the ***par_Rmax*** value in Å.

Name	Spec. type	#	S/N	Instr.	Date	FWHM	***par_stretching***	***par_Rmax***
WASP-121	F6V	140	406	HARPS	2018/01	21.6	7.6	138
HD142	F7V	40	595	HARPS	2004/10	16.1	9.2	50
HD10647	F9V	56	784	HARPS	2004/10	10.2	10.8	86
Sun	G2V	1294	12445	HARPS	2019/01	6.96	11.7	78
HD20794	G6V	28	968	CORALIE	2016/12	8.01	11.7	76
τ Ceti	G8V	101	3036	HARPS	2017/07	6.22	11.9	76
α Cen B	K1V	1767	12244	HARPS	2010/03	6.07	11.9	72
α Hor	K2III	2	1081	ESPRESSO	2019/01	6.67	11.8	72
61 Cyg A	K5V	129	3494	HARPS	2013/03	5.63	12.1	65

the bluest wavelength of the spectrum, which prevents the rolling pin from falling inside the stellar lines. The maximum radius is computed with the same pre-continuum S_1 and S_2 used for the penalty (see Sect. 2.2.4). The idea being to detect the largest absorption region, we compute the difference between S_1 and S_2. R_{max} is then defined as the longest cluster for which $S_1 - S_2$ keeps the same sign.

Appendix C: Low S/N spectra time-series

Low S/N spectra are in general difficult data to deal with. In the case of alpha shape algorithm, we already presented the risk for the upper envelop to fit the noise envelop and not the continuum, justifying the low-pass filter performed as a first step on the spectra (see Sect. 2.2.2). Another concern might exist for the clustering algorithm presented in Sect. 2.4 in the case of spectra time-series. For low S/N spectra, the wavelength positions of the local maxima could be spuriously distributed which is contradictory to the main assumption used in the clustering algorithm, making it inefficient.

If the user wants to normalize N low-S/N spectra with the same anchor points, he should do the following. First, build a master spectrum, by stacking together all the individual spectra. Some care should be taken to shift them in the same rest frame. Then the user should run RASSINE on this master spectrum to normalise it and find the optimal anchor points. Finally, the user should run the function *intersect_all_continuum* giving as input the N spectra, and as optional argument *master_spectrum* the name of the master spectrum. This will enforce the code to use for each spectrum's normalisation the same anchor points as the master. After this, the user can run the *matching_diff_continuum* function with the N spectra and the master spectrum.

Appendix D: Graphical interfaces of RASSINE

In this section, We present the different graphical interfaces that are displayed to the user if the graphical feedbacks are activated (***feedback*** set to True). The first graphical interface (Fig.D.1) allow to smooth the spectrum, where a slider is used to select the kernel width and buttons can select the kernel itself. The second graphical interface (Fig.D.2) is used to select the penalty law, which means the minimum and maximum radius of the alpha shape, as well as the penalty law itself. After this step, the code displays the edges of the spectrum (Fig.D.3) and the user select the number of times to flatten the edges. Outliers based on the derivative criterion are then flagged and proposed to be visually inspected by the user (Fig.D.4). The user can also select manually (see Fig.D.5) the local maxima that he wants to keep or reject by clicking on them. The final product is presented in Fig.D.6. For spectra time-series, two more graphical interfaces are available. The first optional function (Fig.D.7) is launched with the intersect_all_continuum function. By adjusting the two slider, the user will see in direct the formation of the cluster. The second graphical interface (Fig.D.8) is used to select the length of the window used in the Savitzky–Golay filter performed on the spectra difference.

Appendix E: Collections of RASSINE reduction in automatic mode

We already tested RASSINE in automatic mode on spectra obtained with four different instruments which present various shape of instrumental response. In all the cases as presented in Fig.E.1, the continuum produced by RASSINE were visually satisfying. All of them were obtained without fine-tuning any parameters except one : the penalty law through the ***par_reg_nu*** parameter. RASSINE hence managed to deal with smoothed spectral continuum as with HARPN (first row), imperfect 1d spectrum reconstruction and CCD gap with HARPS (second row), telluric forest as with ESPRESSO (third row) and exotic continuum shape as with CORALIE (last row).

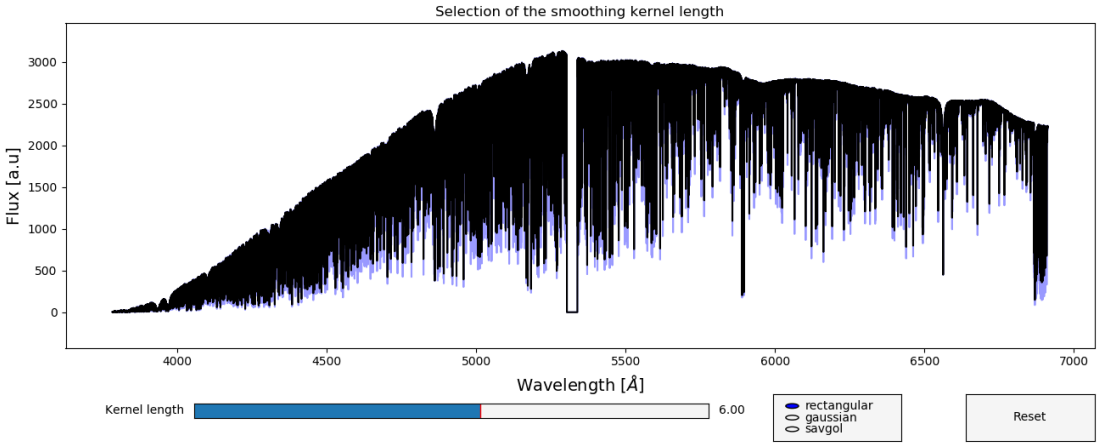

Fig. D.1: First graphical interface of RASSINE to select the smoothing kernel. The user has only to select the smoothing length and kernel shape.

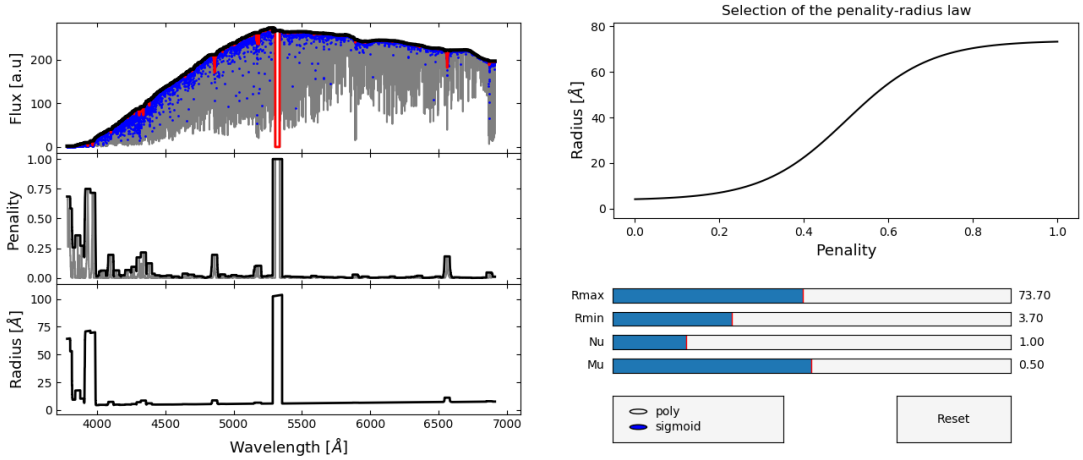

Fig. D.2: Second graphical interface of RASSINE to select the penalty law. The user has to select the functional form of the penalty law (polynomial or sigmoid) as well as the minimum and maximum values.

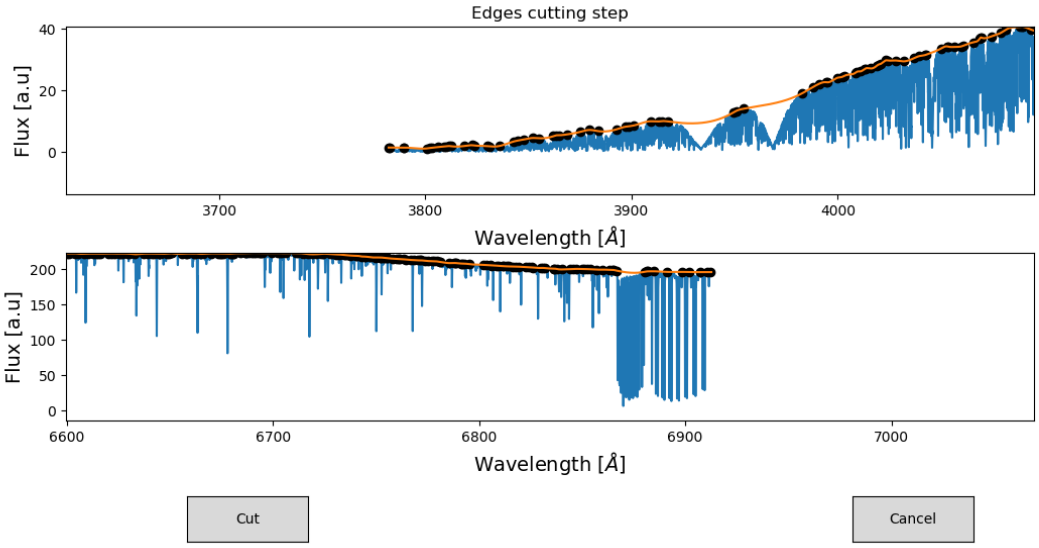

Fig. D.3: Third graphical interface of RASSINE to cut the borders of the continuum. The user has to click on the "cut" button until being satisfied.

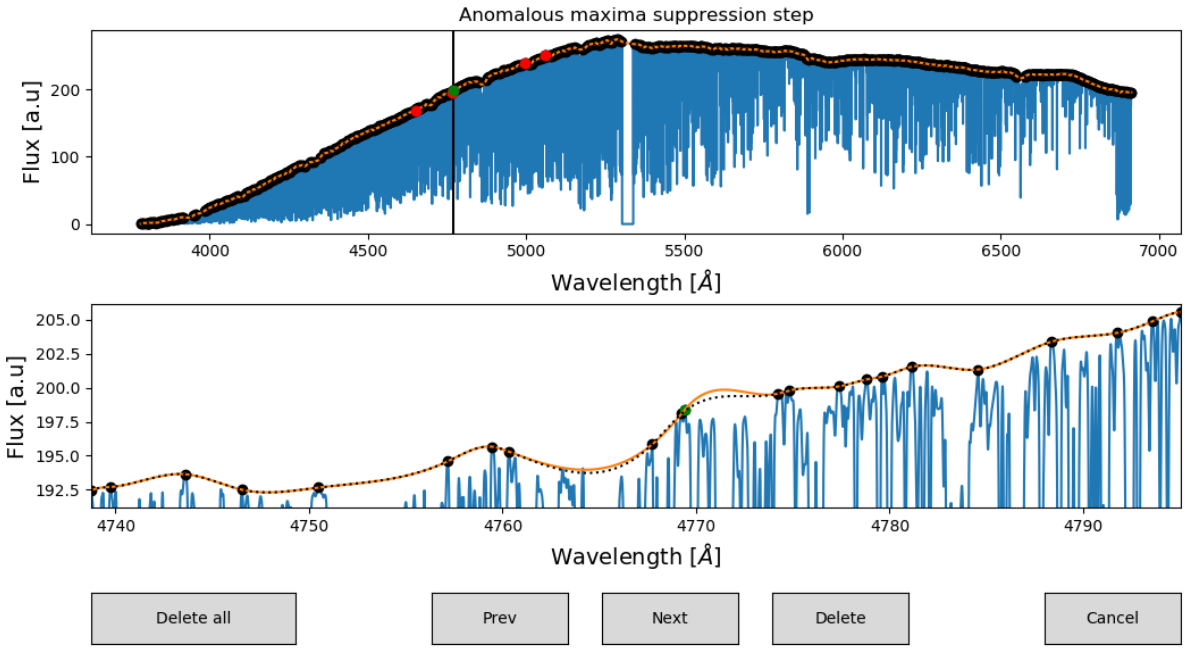

Fig. D.4: Fourth graphical interface of RASSINE to suppress anomalous local maxima. The user can navigate through 5 anomalous at the same time.

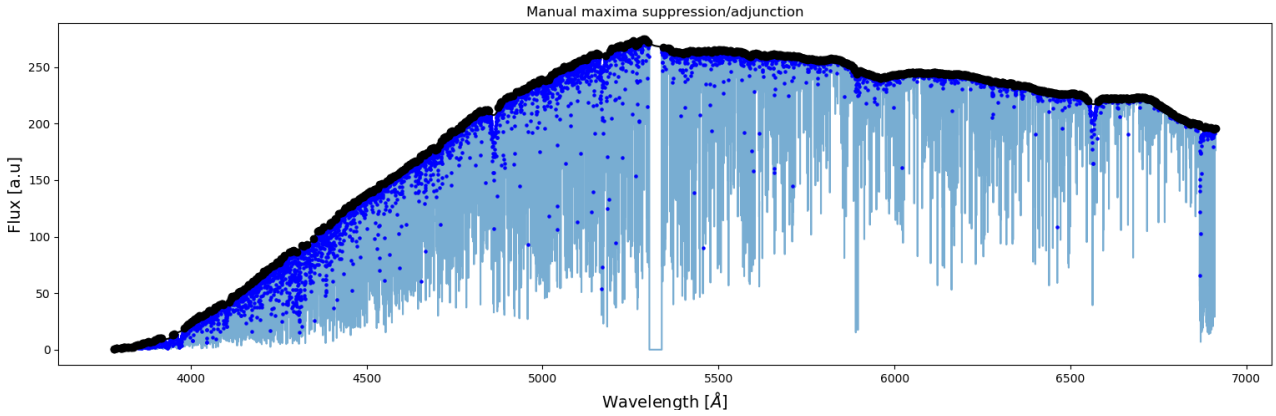

Fig. D.5: Fifth graphical interface of RASSINE to supress or add manually some local maxima by clicking on the points.

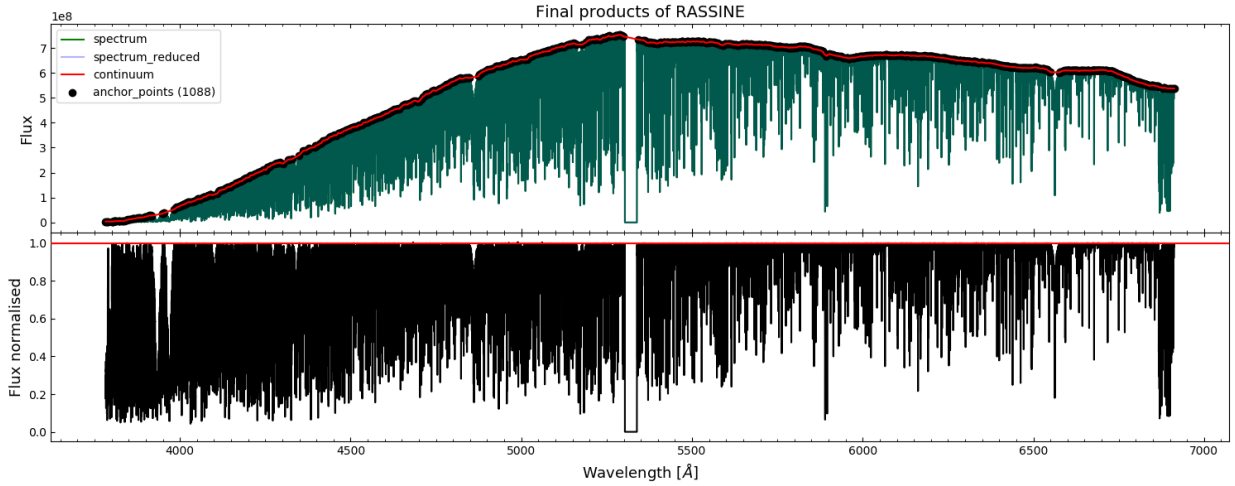

Fig. D.6: Last graphical interface of RASSINE presenting the final output.

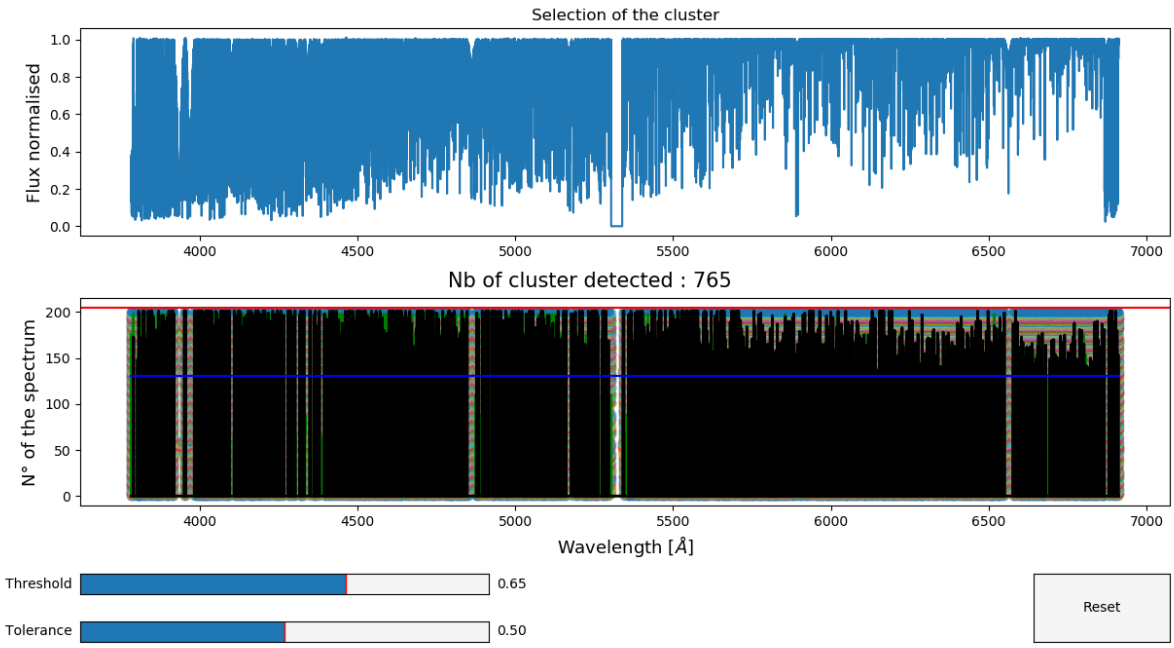

Fig. D.7: First optional function of RASSINE (*intersect_all_continuum*) to reduce spectra timeseries and stabilize the derived continuum. The anchors points of all the points are plotted.

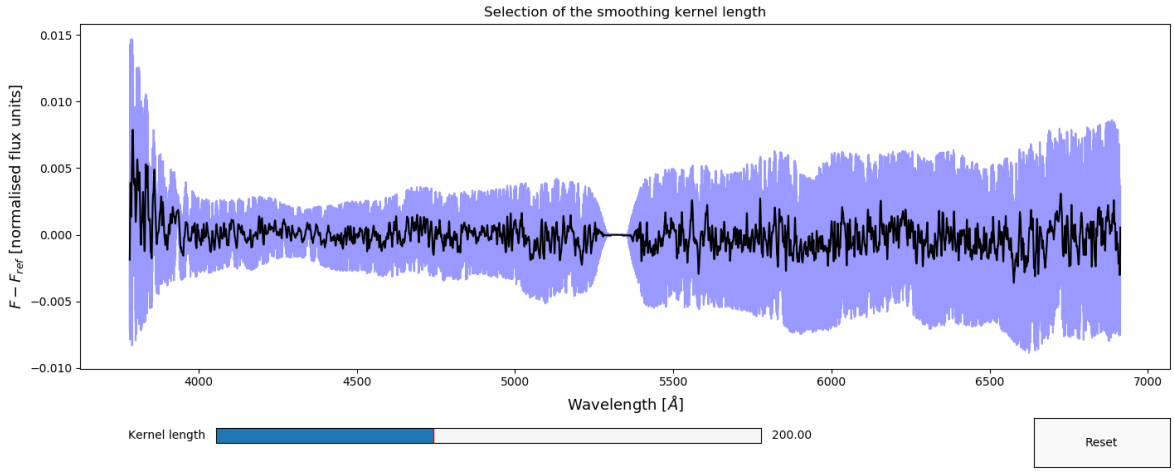

Fig. D.8: Second optional function of RASSINE (*matching_diff_continuum*) to reduce spectra timeseries and stabilize the derived continuum by applying a Savitzky–Golay filter on the spectra difference with a reference spectrum.

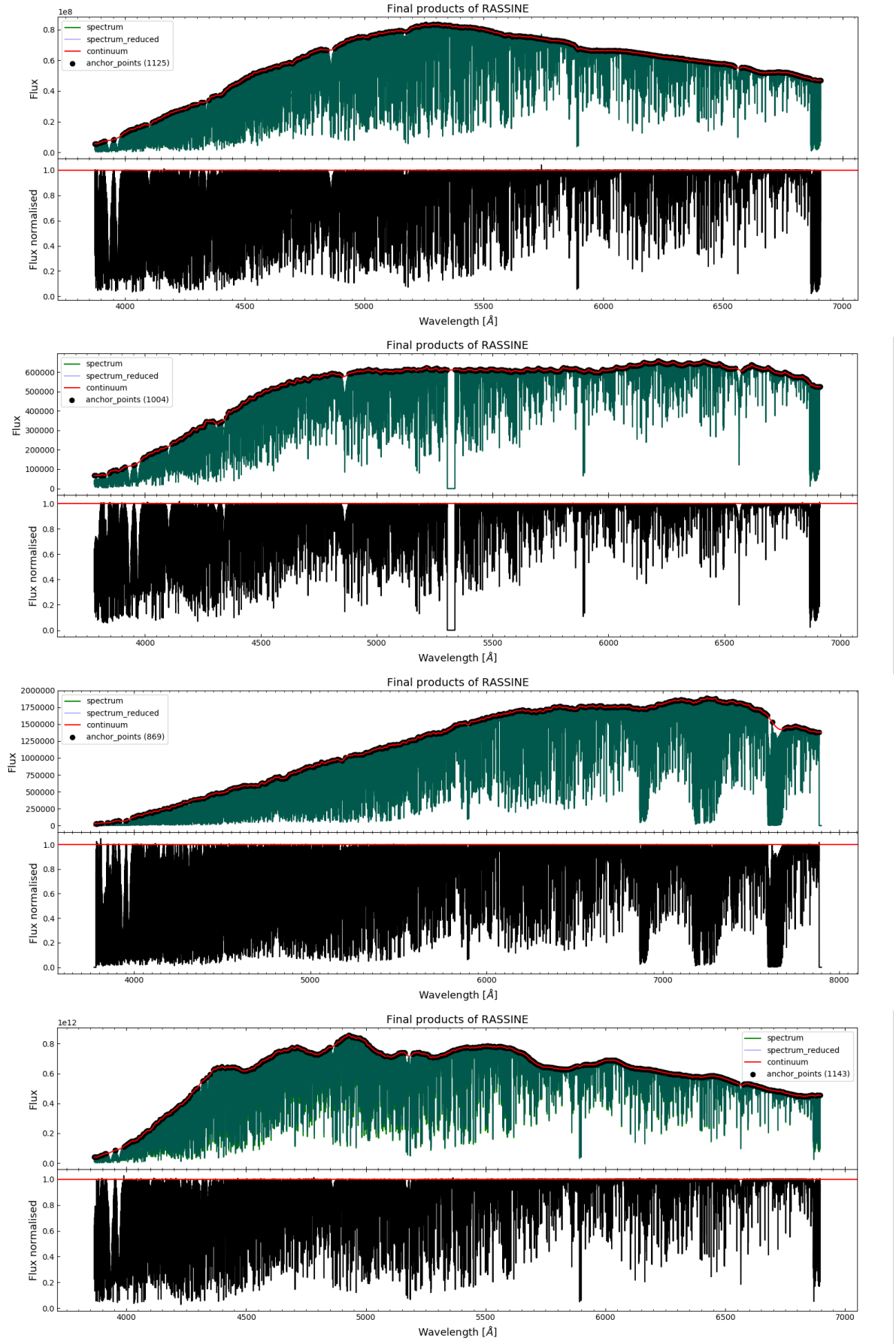

Fig. E.1: A collection of spectra normalised by RASSINE in automatic mode, $\nu = 1$, intermediate tension (***par_stretching*** = 'auto_0.5') and Savitzky-Golay filtering with ***par_smoothing_box*** = 6. Each spectra is coming from a different instrument, from top to bottom : HARPN, HARPS, ESPRESSO and CORALIE.