UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

Bachelor Project of Physics

Laser Frequency Comb Calibrations and Radial Velocity Extractions

Computational methods for exoplanet detection.

Jakob S. Harteg

Advisor: Troels C. Petersen

Submitted: June 15, 2022

Abstract

This project report explores data from The EXtreme PREcision Spectrograph (EXPRES) and explains devised methods for performing laser frequency comb (LFC) wavelength calibrations and radial velocity (RV) extractions from star spectra. For the calibration a per-line rms of 10.0 m/s was reached using a cubic-spline interpolation. Relative radial velocity shifts for four different planets were extracted, although with faulty errors, and coincided within an order of magnitude with the results of Lily L. Zhao et al. [9].

Contents

Note about the project

In the beginning of February 2022, Troels C. Petersen (UCPH) and Lars A. Buchhave (DTU) started a collaboration to improve computational methods for LFC calibrations of spectrographs, and Troels invited me and a few other students to come along for the ride. During the project, I have been exploring real data and developing a crude method for performing LFC calibrations and radial velocity extractions. With this project report, I intend to share and explain the process of performing such analyses, thinking of the reader as a fellow physics student perhaps interested in picking up where I left off.

1 Introduction

The radial velocity method was used to discover the first exoplanets and continues to be one of the main methods for the discovery and characterization of exoplanets today[6]. With new extreme-precision radial velocity (EPRV) spectrographs such as the EXtreme PREcision Spectrograph (EXPRES), data on which this project is based, we are slowly approaching the precision necessary for the discovery of Earth-sized planets around Sun-like stars. For this to succeed, it is however necessary to understand and mitigate many effects, such as the movement of the Earth relative to the center-of-mass of the solar system (barycentric corrections), light scattering in the Earth's atmosphere (tellurics), and light scattering inside the spectrograph (blaze). A general wavelength calibration of the spectrograph is also needed, the quality of which of course directly influences the precision of the final radial velocities that can be obtained. For that purpose, EXPRES utilizes a Laser Frequency Comb (LFC), which is a rather new technique. While to extract radial velocities (RV) we measure the apparent wavelength shift between observations over a period of time, preferably more than a year, and utilize the well-known Doppler effect.

The full procedure from raw data to results, also referred to as the *pipeline* in the literature, is extensive and complex, which is why I have ignored many aspects of it during this project. I have worked directly on the LFC calibration and RV extractions and tried to orientate myself about the most important corrections, thus that is what I will describe in this report. The full pipeline is described in detail in [7].

2 Theory

2.1 Radial velocity method for exoplanet detection

The radial velocity method is one of the few current methods of detecting exoplanets. Two celestial bodies in orbit around each other, such as a star and a planet, orbit their common center of mass (barycenter). This means that the star, although typically much more massive than the planet, is also in movement relative to an outside observer. The larger the planet is relative to the star, the faster the star will appear to be moving. This motion we can measure through the Doppler effect, as the light of the star observed on Earth will be blue-shifted when the star is moving toward us and red-shifted when moving away. If there is a planet around a star, we should therefor observe a periodic Doppler shift. The method is illustrated in figure 1.

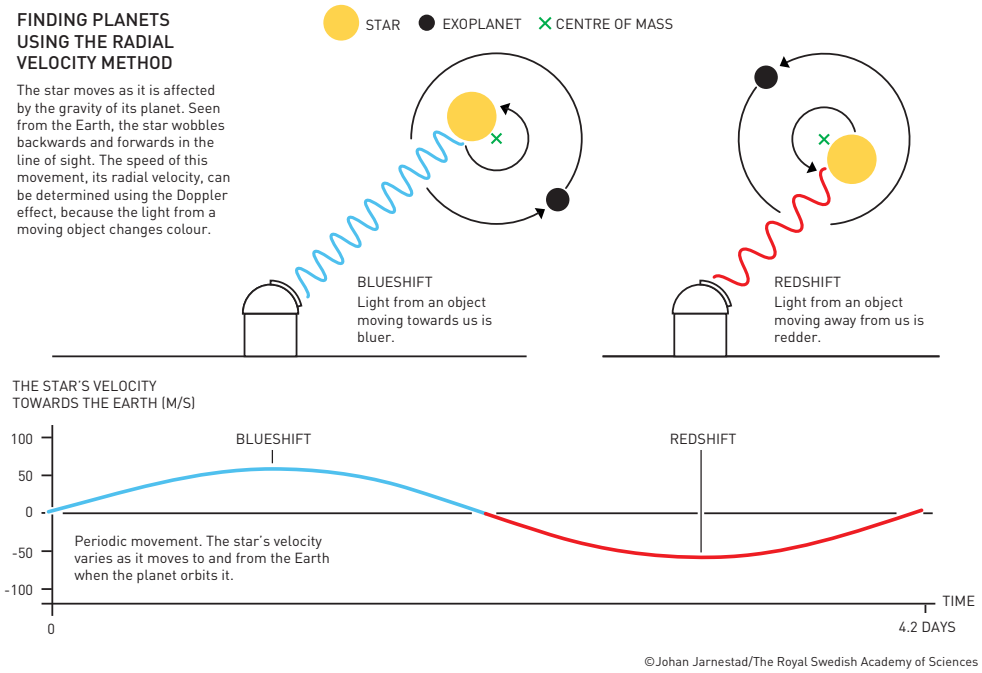

Figure 1: Illustration of the radial velocity method for exoplanet detection. © Johan Jarnestad/The Royal Swedish Academy of Sciences.

The induced motion is called *radial velocity semi-amplitude* and can be calculated by

$$K_1 = \sqrt{\frac{G}{(1-e^2)}} m_2 \sin i \, (m_1 + m_2)^{-1/2} \, a^{-1/2} \tag{1}$$

where G is Newton's gravitational constant, e is the eccentricity of the orbit, a the semi-major axis, i the inclination and m_1 and m_2 the mass of the star and the planet respectively [6]. With this formula, we can calculate that Jupiter causes the Sun to move back and forth with a velocity of $K_1 = 12.5$ m/s, for an outside observer in the plane of orbit, while the Earth induces a radial velocity of only $K_1 = 8.95$ cm/s. If we look at another star system exactly in the plane of orbit the inclination i is 90° degrees. If we however observe the system at an angle (which most of the time we are) the radial velocity will appear smaller by a factor of $\sin i$.

2.1.1 Doppler shift

The radial velocity method relies on the well-known Doppler effect. Ignoring terms of c^{-4} and higher, the general shift caused by a relative displacement between the source and an observer at zero gravitational potential is given by

$$\lambda = \lambda_0 \frac{1 + \frac{1}{c}\mathbf{k} \cdot \mathbf{v}}{1 - \frac{\Phi}{c^2} - \frac{v^2}{2c^2}}, \tag{2}$$

where λ is the observed wavelength, λ_0 the emitted wavelength, Φ the Newtonian gravitational potential at the source ($\Phi = GM/r$ at a distance r of a spherically symmetric mass M), $\mathbf{k}$ the unit vector pointing from the observer to the source, $\mathbf{v}$ the velocity of the source relative to the observer and c the speed of light. This formula accounts for both special relativistic effects and the gravitational Doppler shift described by general relativity[5].

As we are dealing with velocity shifts on the order of meters or centimeters per second, we can safely ignore special relativistic effects and thus cross out the third term in the denominator. The gravitational Doppler shift is independent of the motion of the star but instead depends on its mass and radius. The Sun has a mass loss rate of about 3×10^{-12} % per year[3]. If we assume a similar mass loss rate for the stars used in this study (which all have masses close to that of the Sun), we can safely assume that the gravitational effect is negligable and omit the second term from the denominator.

If the unit vector $\mathbf{k}$ were pointing directly toward us, it would mean that we were observing the system in the plane of orbit. This is however unlikely. Since we don't know the inclination angle, a possible simplification is to omit $\mathbf{k}$ and treat the resulting $\mathbf{v}$ as a minimum radial velocity. Thus, we are left with

$$\lambda = \lambda_0 \times \Big(1 + \frac{v}{c}\Big), \tag{3}$$

which is to say that the observed wavelength is simply the emitted wavelength scaled by a factor $(1 + v/c)$. This formula allows us to compute the minimum relative velocity shift, v, between two observations, λ and λ_0.

2.2 Description of the instrument

The EXtreme PREcision Spectrograph or EXPRES is an extreme-precision spectrograph situated at the Lowell Observatory's 4.3m Lowell Discovery Telescope (LDT) near Flagstaff, Arizona, USA. The LDT allows for up to 280 partial nights of observation per year.

As is common for many spectrographs, at the heart of EXPRES is a Charge Coupled Device (CCD). A CCD is a silicon-based multi-channel photon detector consisting of a large number of small light-sensitive areas called pixels. The CCD in EXPRES is an STA1600LN CCD backside illuminated image sensor with a $10,560 \times 10,560$ array containing 9µm×9µm pixels, designed with a wavelength range of 3800−7800Å. When a photon hits a pixel it is converted into a charge, and each pixel can thus supply independent measurements. Since a one-dimensional sensor would be impractical, EXPRES is constructed in such a way that it wraps the spectrum inside the CDD, meaning that the spectrum starts in the top row of the sensor, and continues in the second row. Short wavelengths are thus to be found at the top of the CCD and long wavelengths at the bottom.

EXPRES is housed in a vacuum enclosure to minimize changes in temperature and pressure, which could otherwise cause the spectra to change position on the CCD and thus lead to errors in the RV measurements.

Calibration device: Wavelength calibrations are performed with the use of a Laser Frequency Comb (LFC), produced by Menlo Systems, which is a laser source with a spectrum consisting of a series of discrete, equally spaced frequency lines. The LFC however also needs calibration for which a Thorium Argon (ThAr) lamp with known frequencies is used.

Spectral resolution: EXPRES has a maximum spectral resolution of $R = 150,000$, where R is defined as $R = \lambda/\Delta\lambda$. Inverting this we get what is called the resolution element of the instrumental spread Function (or line spread function), $\Delta\lambda = \lambda/R$. For a wavelength of say $\lambda = 5000$Å, this comes out to $\Delta\lambda = 5000\text{Å}/150,000 \approx 0.033$Å, and it describes the "blurring" of monochromatic beams on the detector. Absorption features narrower than $\Delta\lambda$ can, in a well-behaved spectrograph, be approximated as a normalized, symmetric Gaussian function with FWHM = $\Delta\lambda$. Specifically for EXPRES, a super-Gaussian is a better description of the line spread function[9]. A Super-Gaussian is a regular Gaussian but with an extra parameter that allows the top to be flattened. LFC lines being monochromatic and thus very narrow will appear on the detector as a super-Gaussian with $\Delta\lambda$ ranging from 3.9-5 pixels across the detector. By fitting, the center of the peak can be identified to a fraction of a pixel. The LFC lines are separated by about 10 pixels to remain distinct after this blurring. For star spectra, however, some absorption lines will be too close together and appear "blended" on the detector as a result[9].

Barycentric correction: Barycentric corrections are derived from the EXPRES exposure-meter, which is essentially a smaller, less precise spectrograph. Described in detail in [1]. EXPRES as a whole is described in technical detail in [4].

2.3 Description of the data

EXPRES data are meant to serve as an example of the data being produced by next-generation spectrographs. The data used in this project was supplied by Lily L. Zhao and is by no means raw data, but data that has already gone through a lot of processing [9].

For the development of an RV extraction method, observations from four stars were used:

- HD 34411 (188 observations, 58 nights, Oct. 08, 2019 - Nov. 27, 2020)
- HD 10700 (174 observations, 34 nights, Aug. 15, 2019 - Nov. 27, 2020)
- HD 26965 (114 observations, 37 nights, Aug. 20, 2019 - Nov. 27, 2020)
- HD 101501 (45 observations, 22 nights, Feb. 10, 2019 - Nov. 26, 2020)

Most days have three to four observations and there are significant time gaps in the data as well. LFC exposure files were provided by Lars A. Buchhave.

2.3.1 Data structure

The data used in this project consists of already packaged FITS (Flexible Image Transport System) files, which is a portable file standard widely used in the astronomy community to store images and tables. There is a FITS file for each observation, containing a variety of measurements for each pixel on the CCD. The rows of the CCD data are referred to as orders. There are 86 orders each of which contains values from 7920 pixels. Drawing a coordinate system on the CCD, we are thus moving through pixels as we go along the x-axis and through orders as we go along the y-axis.

This would give the CCD the very elongated dimensions of 86×7920, but as mentioned earlier, the CCD is square. Due to the optics of EXPRES, orders hit the CCD at an angle and for this reason, *order tracing* is necessary. Order tracing reduces each order from 2d array to a 1d array, which means the final image comes out much shorter in the vertical/order dimension. Described in detail in section 3.2.1 of [7].

Furthermore, the CCD is not equally sensitive everywhere, and there are areas along the edges that are deemed useless. The data comes with a mask that shows which pixels should be used.

2.3.2 Noise and corrections

Photon noise and read noise are the two largest contributors to the noise on a given pixel on the EXPRES CCD. These two quantities are measured and summed in quadrature for each pixel. Photon noise is assumed to be Poisson distributed and the standard deviation is then the square root of photon counts. Read noise is calculated empirically and is assumed to be consistent throughout each night of observation[7].

Although manufacturers have tried their best to limit it, the CCD still gets hits by scattering light, being the strongest in the center of the detector. This has been modeled and subtracted from the spectrum by measuring the photon count in between orders. The blaze function is available in the data file and allows for recovering the original counts for each pixel by multiplying the blaze with the spectrum.

Tellurics in the context of spectrographs refers to the contamination that ground-based spectrographs must cope with, which occurs as the light passes through our atmosphere encountering molecules such as oxygen and water vapor on the way. The data I use is already corrected for this with a technique called SELENITE[9].

The barycentric correction is vital, as this is what removes the movement of the Earth around the center-of-mass of the solar system from the data. How this is done exactly I have however not delved into during this project.

In the end, the largest source of error comes from stellar activity; variations in the light from a given star due to various physical processes on the surface. Dark spots, granulation, oscillation and rotation are a few examples. Disentangling stellar activity signals from RV is practically a research field of its own.

3 Data Analysis

In this section, I will describe and explain methods for performing LFC calibrations and RV extractions. I did not end up having access to star spectra with associated LFC exposures, such that I could calibrate the star spectra myself before computing the radial velocities. As a result, part one of the data analysis does not directly carry into part two, although, with the right data, it could.

3.1 Calibration

The calibration is needed to map each pixel on the CCD to a specific wavelength. Such a map is referred to as a wavelength solution. For this, we first shine a light source with known frequencies, preferably many discrete lines, onto the CCD. The pixels on the CCD on which each line appears can then be mapped to the known wavelength of the given line. EXPRES uses a Thorium Argon (ThAr) lamp with 4,000 lines across 82 orders for an initial rough wavelength solution. These values are then linearly interpolated to provide a wavelength solution for all pixels across the CCD. The data I have worked with came with this wavelength solution.

For a better wavelength solution, EXPRES uses a laser frequency comb (LFC), which generates about 20,000 lines across 50 orders, with frequencies given by the relation

$$v_n = v_{\text{rep}} \times n + v_{\text{offset}} \tag{4}$$

for integers n. The repetition rate v_{rep} and offset frequency v_{offset} are referenced against a GPS-disciplined quartz oscillator, providing calibration stability corresponding to a fractional uncertainty of less than 8×10^{-12} for integration times greater than 1s[7]. The values I have used, $v_{\text{rep}} = 14e9$ and $v_{\text{offset}} = 6.19e9$, I found in the Excalibur github[8].

The following steps are followed to perform the LFC calibration: 1) determine the CCD pixel location of each LFC line, 2) use ThAr wavelength solution to find the corresponding mode n in Eq. (4) to know the true wavelength of the line, and finally 3) find a way to estimate a wavelength solution in between LFC lines.

Figure 2 is an illustration (not real data) of this process. The left panel shows LFC lines appearing in different locations on the CCD (with exaggerated variations and errors for clarity) and the corresponding wavelength solution on the y-axis given by Eq. (4). The middle and right panels illustrate two ways of modeling a wavelength solution in between LFC lines, polynomial fitting and cubic-spline interpolation respectively. With such a wavelength solution, you have a direct map between any location on the CCD (x-axis) and a wavelength solution (y-axis). This is to be performed in every order of the CCD.

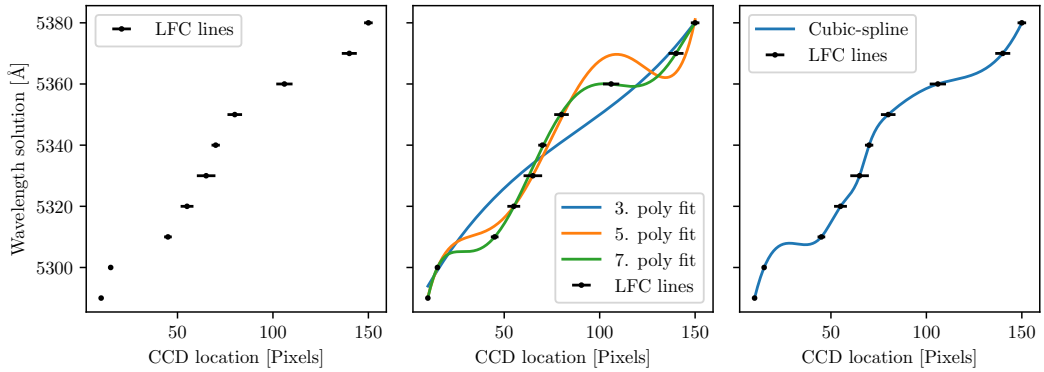

Figure 2: Illustration of the LFC calibration process (not real data). See text for explanation.

3.1.1 Determining LFC line locations

The right side of figure 3 shows an LFC exposure plotted in its entirety. It is clear that the LFC does not cover all orders of the CCD, but starts shortly before order 40 and stops around 76. To determine the location of the LFC lines I follow this procedure: 1) Find peaks using scipy peak finder, 2) make data slices around each peak with the size of the average distance between peaks, 3) using iminuit do a chi2 minimization fit to each peak with a super-Gaussian plus a linear background. See figure 3 left side.

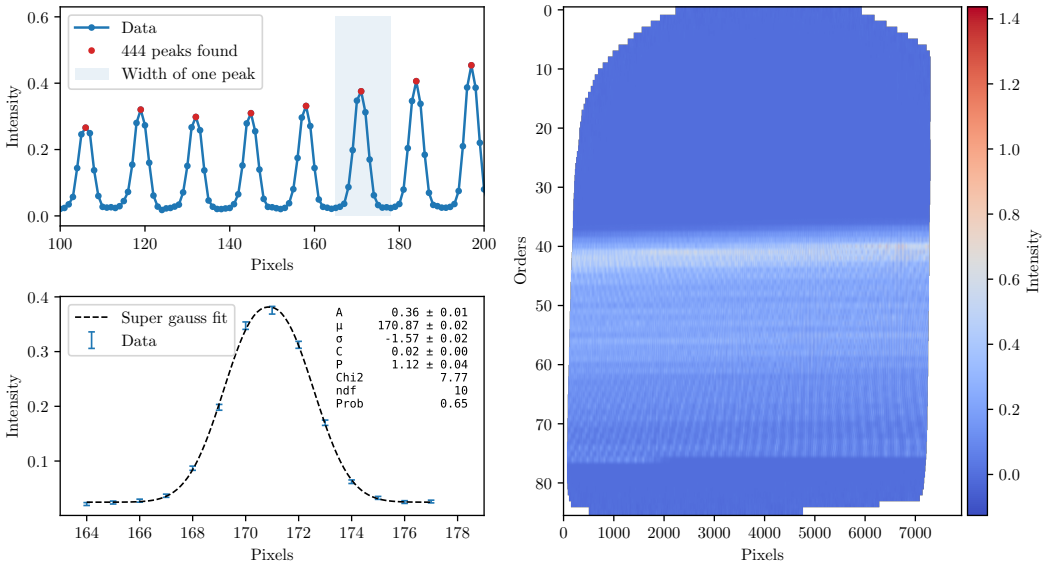

Figure 3: Upper left: illustration of a few LFC lines in order 65. Peaks are identified with scipy peak finder. Lower left: each peak is fitted with a super gauss to find the exact top of the peak with uncertainties. Right: Measured intensities for the LFC across the CCD (unitless).

A super-Gaussian, defined in Eq. (5), is a regular Gaussian but with an extra parameter, here denoted P, that allows the top of the Gaussian to be flattened. The last two terms here add a linear background and an offset.

$$f(x; A, B, C, P, \mu, \sigma) = A \exp\left(-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)^P\right) + B(x-\mu) + C \tag{5}$$

The fit is then a standard chi2 minimization:

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - f(x; A, B, C, P, \mu, \sigma)}{\sigma_i}\right]^2, \tag{6}$$

where N is the number of data points, x is pixel-space, y_i and σ_i are the measured intensity and uncertainty respectively. The fit returns the values and uncertainties for the parameters A, B, C, P, μ, σ when the chi2 is minimized.

We are mainly interested in μ, which gives the location of the LFC peak on the CCD (in pixel-space). With the initial rough wavelength solution derived from the ThAr lamp, I can determine what the approximate wavelength of the LFC peak should be. To find the better

wavelength solution I then go look up the closest frequency given by Eq. (4), and we now have a map of 20,000 points on the CCD with a good wavelength solution.

For a complete solution, we also need to estimate the wavelength solution between the LFC lines and to do that, I have explored two approaches: polynomial fitting and cubic-spline interpolation. Before moving on though, it is worthwhile to study the chi2 values from the many fits we just performed, as they can help us evaluate whether the errors of the original data are correct.

3.1.2 Errors in the calibration data

According to Zhao, the line-spread-function of EXPRES can best be represented by a super-Gaussian [9], and so as we fit a super-Gaussian to the peaks, the chi2 will grow if the measured values do not agree with the predicted values within the uncertainties. For a given data point, if the measured value differs by exactly the uncertainty, that data point would contribute 1 to the sum. However, fitting with the super-Gaussian (including a linear background) we also have 6 fitting parameters that allow some wiggle room for the model to fit the data, which we must compensate for. For a given data series, such as an LFC peak, we can expect the chi2 to roughly equal the number of data points in the sum *minus* the number of parameters in the fit, i.e. the number of the degrees of freedom:

$$N_{\text{dof}} = N_{\text{data-points}} - N_{\text{fit-parameters}} = 13 - 6 = 7, \tag{7}$$

as I use roughly 13 data points in each fit (the LFC line spacing does vary a bit across the CCD).

So if we plot all chi2 values we get from fitting the LFC lines, as I have done in figure 4, we should see a typical chi2 distribution with a peak around 7. Using the uncertainties as provided in the data file, the chi2 distribution is however very flat (blue curve). It peaks somewhere around 25, which suggests that the uncertainties are about a factor $\sqrt{3}$ too small (square-root because the chi2 of course is squared and $25/3 \sim 8$). Multiplying the uncertainties by $\sqrt{3}$ and fitting the peaks again gives a chi2-distribution with a peak roughly around 7-8 (green curve).

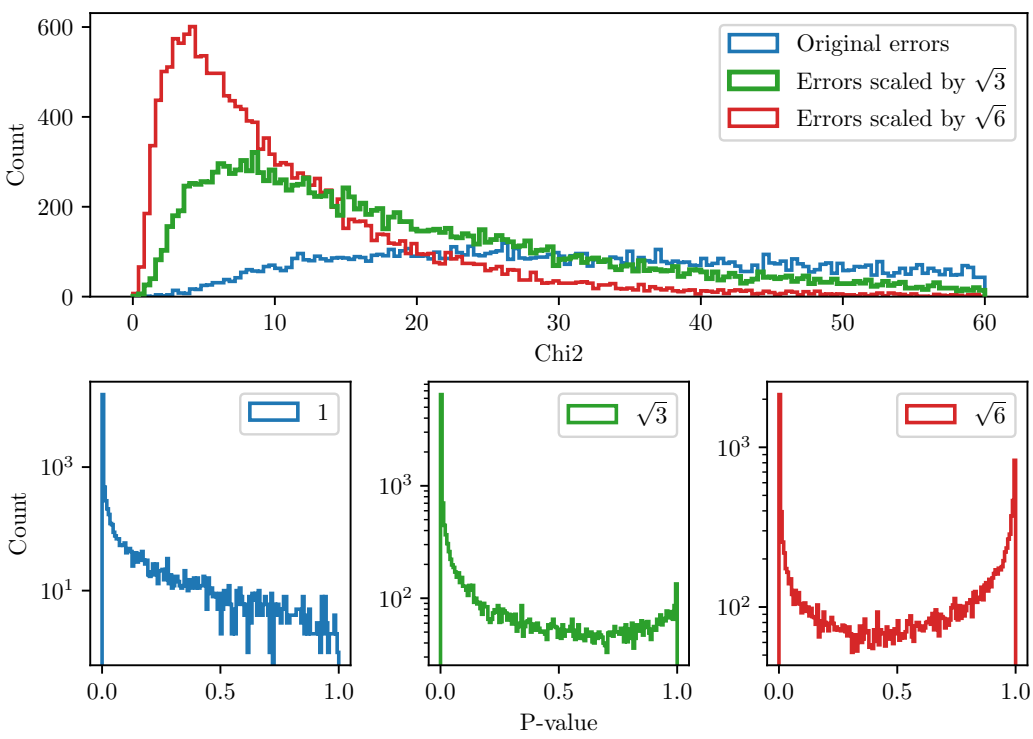

Figure 4: Chi2-values and p-values from individual LFC peak super-Gaussian fits with errors multiplied by different scale-factors (1, $\sqrt{3}$ and $\sqrt{6}$).

Additionally, we can also look at the distribution of the p-values. With proper uncertainties, we should expect neither too many awful fits (p=0) nor too many perfect fits (p=1), but rather a roughly flat distribution. Although the chi2-distribution check is more clear-cut, the p-value distribution for uncertainties scaled by $\sqrt{3}$ (green) is better. Scaling by $\sqrt{6}$ is too much: the chi2 peaks around 4 and the p-distribution shows many more perfect fits.

3.1.3 Poly-fit calibration

Since the LFC peak locations, as seen in the right plot in figure 3, appear to exhibit a certain periodic behavior, my initial approach to compute a wavelength solution between LFC lines was to fit the LFC peak locations with a polynomial. Looking at the residuals of fitting the LFC line locations with polynomials of increasing degree revealed smaller and smaller periodic variations, until reaching 5th degree, see figure 5. The p-value drops to 0 for 6th degree, and the per-line rms (given by Eq. (8)) explodes. Only LFC lines with a chi2 (from the super-Gaussian fit) smaller than 100 were used for the analysis and errors were scaled by $\sqrt{3}$.

The residuals plotted in figure 5 show the difference between the theoretical wavelength for each LFC line (Eq. (4)) and the wavelength that the calibration predicts for that location. You can get a feel for the quality of the calibration by looking at the residuals, but a more formal quantification is to compute the per-line rms for a given LFC exposure as:

$$\text{RMS/line } [\text{ms}^{-1}] = \sqrt{\sum_{n=1}^{N} \frac{\left[\frac{(\lambda_{n,\text{pred.}} - \lambda_{n,\text{theory}})}{\lambda_{n,\text{theory}}} \times c\right]^2}{N}} \tag{8}$$

for N lines, where c is the speed of light in m/s.

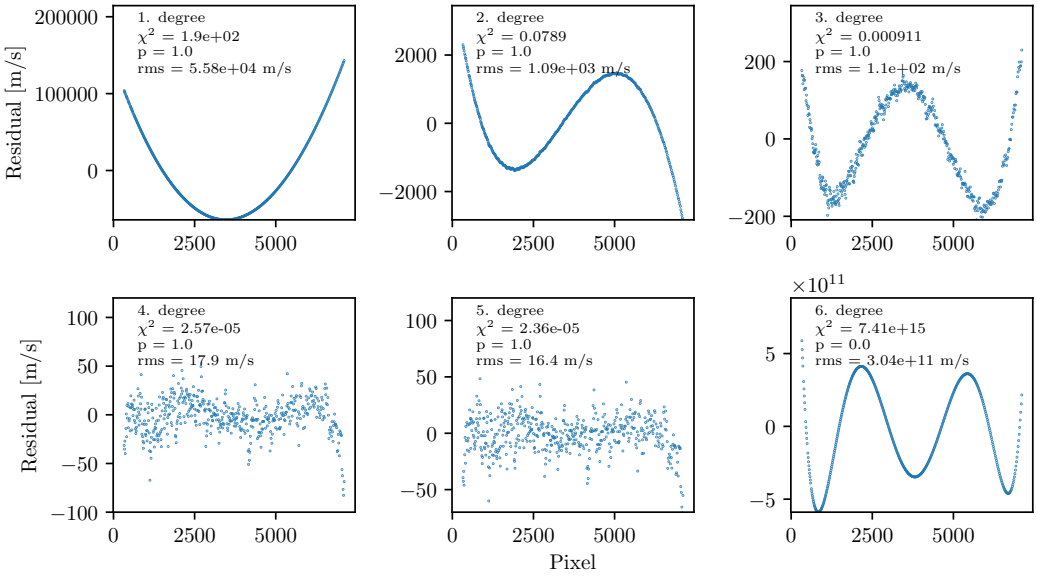

Figure 5: Example of residuals from fitting LFC line locations with polynomials of increasing degree in one order (44). The x-axis shows pixels and the y-axis residuals in m/s in all plots. Errors used for determining line locations have been scaled by $\sqrt{3}$. The noted rms in each plot is the per-line rms as defined in Eq. (8).

3.1.4 Interpolation calibration

Although there are large periodic trends in the residuals in figure 5, on the very small scale, the changes seem rather random. A better approach might therefore be to use interpolation, as this is more flexible and copes better with very complicated and non-repetitive behavior.

A cubic-spline interpolation would force all the residuals to be zero, so in order to evaluate the quality of the method, we can omit for instance every second peak from the interpolation and then compute the residual between the omitted peaks and the resulting interpolation function. Then we can flip it around and interpolate the peaks we left out before and compute the remaining residuals.

Figure 6 shows a comparison of the residuals from a 5th degree poly-fit and the cubic-spline. For the cubic-spline I get a per-line rms of 10.0 m/s, while the poly-fit gives 5.73×10^8 m/s, which suggests that the cubic-spline approach is the better one. It is also worth noting that because the interpolation was done on only half the data points at a time, it will be even better when performed on all data points, as it would be when used for calibrating data before an RV analysis.

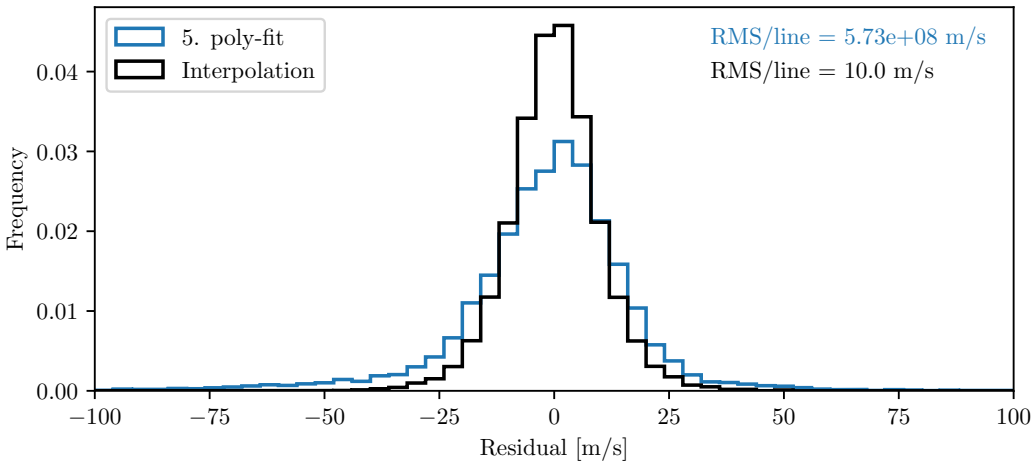

Figure 6: Residuals from calibrations performed through a 5th degree poly-fit and a cubic-spline interpolation. The per-line rms as defined in Eq. (8) is given in the top-right corner in the color of the corresponding method.

3.2 Radial velocity extraction

To extract radial velocities we need to measure the Doppler shift between spectra from different days of observation. One way to do that is to compute the cross-correlation, which is a measure of the similarity of two data series as a function of the displacement of one relative to the other.

We can do this either for individual absorption features, chunks of the spectrum a few angstroms wide, or entire orders at a time. I have chosen primarily to work with the individual features.

Due to a lack of access to data consisting of star spectra with associated LFC captures, I have worked on RV extractions using already calibrated data provided by Zhao. This data has been calibrated using a technique called Excalibur [10].

The data is visualized in 7, where the top left shows an extract of wavelength vs. intensity data from an observation of HD 34411. The file also includes a model of the continuum function, with which we can normalize the spectrum through division, shown in the bottom left. Plotted on the right side are all continuum normalized data within the EXCALIBUR mask, i.e. data marked as having a proper calibration.

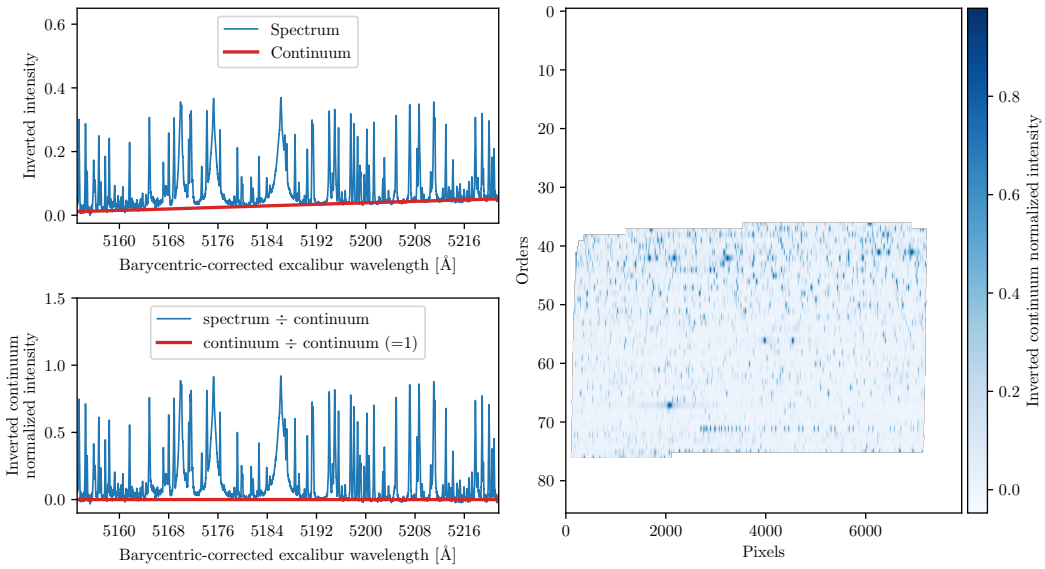

Figure 7: Overview of Excalibur calibrated data from an observation of HD 34411. Upper left: extract of wavelength solution vs. intensity. Lower left: continuum normalized spectrum. Right: all continuum normalized data within the Excalibur mask.

3.2.1 Finding and matching features across observations

To measure how much individual absorption features move in between observations, the first challenge is to find the "same" features in both observations. To do that, I have devised the following procedure:

- Load intensities from the data column `"spectrum"` and errors from `"uncertainty"` as

well as Excalibur calibrated barycentric-corrected wavelengths from `"bary_excalibur"`, all masked by `"excalibur_mask"`.

- Normalize intensities and errors with the continuum function from `"continuum"`.
- Invert intensities to turn absorption features into positive peaks by $y = 1 - y$.
- Locate peaks using `find_peaks` from `scipy.signal` with minimum peak distance of 5 pixels and a minimum peak prominence of 0.25 (unitless).
- Finally slice data around each peak with a width of 30 data points.

And then to match features/peaks between two observations:

- Iterating through the peaks of observation1, for each peak, find the closest peak in observation2. With Excalibur calibrated data, peaks should not shift so much that they overlap. However, the algorithm laid out so far does sometimes match peaks that are far apart or do not resemble each other in shape at all. To bypass such matches we can add two filters:
  - Maximum peak distance: We could filter out all matches where the distance between the peaks is equivalent to a radial velocity greater than 12.5 m/s (the RV Jupiter induces in the Sun). However, since we are dealing with discrete data, the difference sometimes comes out much larger than it really is and a narrow cut of 12.5 m/s would remove many good matches. Instead, setting a very generous cut of 0.5 Å, equivalent to about 20-30 km/s depending on the wavelength, filters out the few very bad matches, but leaves the rest. When analyzing non-barycentric-corrected data, I set this up to 1Å.
  - Maximum difference between the areas under the graph of two features (the sum of the intensity values in the feature): Peaks with similar shapes will give a low difference. This filter is most useful when analyzing non-barycentric-corrected, where features often move so much that they overlap. Figure 8 shows an example of a bad and a good match, where the bad match can be avoided by setting the max area difference down to 0.1 (unitless).
- A way to formally select the best values for these filters is yet to be worked out.

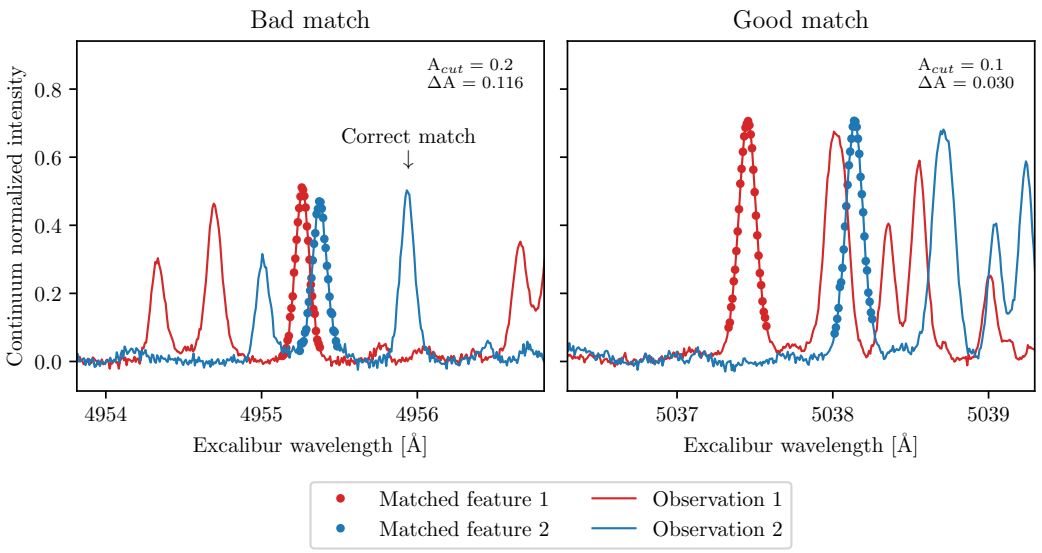

Figure 8: Example of match filtering with non-barycentric-corrected data, where features move a lot. The maximum area difference cut was set to 0.2 and 0.1 on the left and right plots respectively. The cut on the left plot was not strict enough to avoid the bad match that was made, with an area difference of 0.116. While on the right plot, the correct feature was selected spite another feature being closer in wavelength.

3.2.2 Computing velocity shift as the cross-correlation

At this point, I have a list of matching features in different observations. We can now compute the cross-correlation for each match as a chi2 minimization fit with the radial velocity as our fit parameter. The fit will converge when it has found the radial velocity that shifts one peak onto the other one.

Before moving on though, we have to cubic-spline interpolate the spectra data. This is needed for two reasons: 1) The shifts we are looking for are much smaller than the individual pixels on the CCD, so we need to be able to shift by sub-pixel amounts, and 2) to compute the difference in intensity values between peaks, the intensity values must have the same wavelength solution, but, since EXPRES is calibrated independently for each observation, the wavelength solutions are different.

So I cubic-spline interpolate the spectra data from the first observation in the match, but before interpolating the second observation, I shift the wavelength solution by multiplying the shift factor from Eq. (3). Now I can evaluate the two interpolation functions on a common wavelength range, using $N = 1000$ steps[1], and I am ready to compute the chi2:

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - f(x; v)}{\sigma_i} \right]^2 \tag{9}$$

where y_i are the unshifted interpolated intensities from the first observation, σ_i the errors on the intensity of the first observation also sampled through a cubic-spline interpolation, and the function $f(x; v)$ is the cubic-spline interpolation function given by interpolating the intensity values of the second observation with wavelength values shifted by Eq. (3), evaluated on the wavelength range common to both features:

$$f(x; v) = \textbf{interp}[x \times (1 + v/c), y]\,(x_{\text{common}}) \tag{10}$$

I then compute the cross-correlation and obtain the radial velocity, v, as a minimization of Eq. (9) using iminuit.

3.2.3 Computing velocity shifts for all features

We can now find and compute the cross-correlation for all feature matches between two observations. In general, I find between 500-1000 matches between two observations, which results in a fair amount of statistics. The computed radial velocities do build up a normal distribution around a central value, but there are many outliers. This is due to primarily two things: 1) Bad matches making it through the selection filters, and 2) stellar activity.

A better matching algorithm could supposedly be developed, but for now, I have just experimented with setting different cuts both in radial velocity and chi2 to weed out the bad ones. What I have found to work the best however is simply taking the median instead of the weighted average or the mean, as the median is much less affected by outliers. And as there is no reason to believe that stellar activity would cause more RV outliers in one particular direction, so the spread can be assumed to be symmetric, and the median is thus a good approximation. Figure 9 shows an example of computed radial velocitiy shifts for all features between two observations. It is clear that neither the weighted average (yellow) nor the regular mean (black) are good approximations of the total shift. The median (red), however, does a good job.

We also see a constellation of features likely to be affected by stellar activity on the far right, and a match on the left, which appears bad, but actually turns out to be "correct", but of a an absorption line so broad that the data-slice does not have enough information for the fit to converge properly (plotted in figure 14 in appendix A).

Also plotted in figure 9 are the respective errors on the weighted average, the mean, and the median. As it is the median that I have decided to work with, it would make sense to use the median error,

$$\sigma_{\text{median}} = \frac{\sigma}{\sqrt{N-1}} \times \sqrt{\pi/2},$$

for large N, where σ is the standard deviation, however, it not only comes out very large but also breaks the propagation of the original errors from the spectrograph. The weighted error carries on the information of the original errors, but it comes out very small. For the sake of continuing with the analysis, I will go on with the weighted error however.

[1] 1000 steps appeared as the best balance between run time and resulting uncertainty. See figure 13 in appendix A

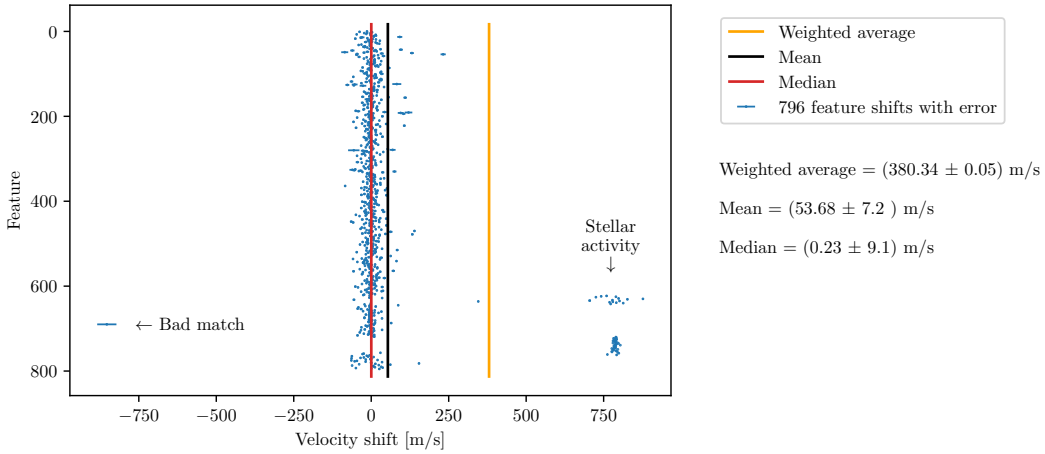

Figure 9: Comparison of weighted average, mean and median for computed rv shifts for two observations of HD 34411. The feature on the left side turns out to be a very broad feature, for which the standard data slice is too narrow and therefore yields bad results, (see figure 14 in appendix A) for details. While the features on the right side are likely to be caused by some kind of stellar activity.

3.2.4 Extracting relative shifts from an overconstrained system

With the devised method we can now compute the relative radial velocity shift between two observations and get out one number with an uncertainty. The next most obvious step would be to compute the shift between observations 1 and 2, 2 and 3, etc. Doing this however leads to correlated results, as the difference between say observations 1 and 10 would depend on all the observations in between, and if there is one bad one, it will affect all the rest. To circumvent this, we can first compute the relative shift between all observations. This will give us an *overconstrained system*, in the sense that there is more information than necessary. All these differences can then be reduced down to a single array, where each shift is relative to all the rest, not only the neighbor.

Computing the shifts between all observations yields an $N \times N$ upper triangular matrix, where each cell is the shift between observations i and j, and thus with a diagonal of zero. I will call this matrix ΔV_r^{ij}, see figure 10 for an example.

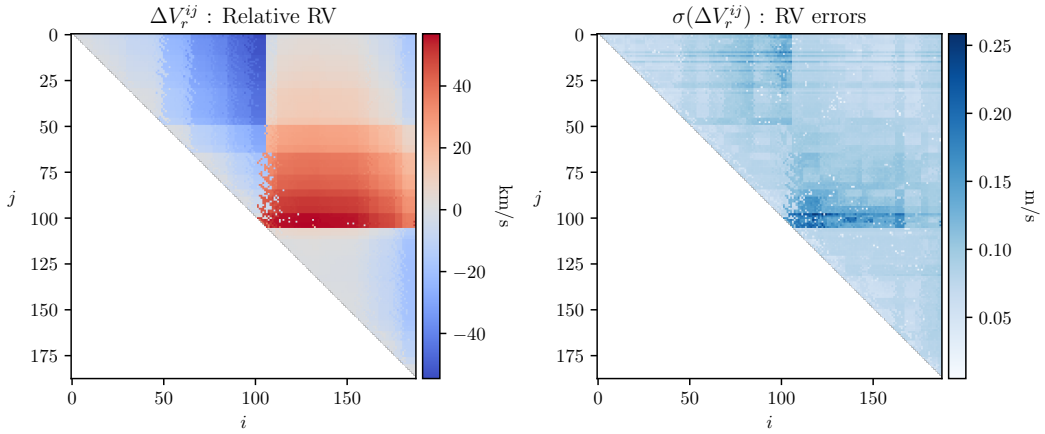

Figure 10: Radial velocity shifts matrix computed for 188 observations for HD34411 using Excalibur calibrated but non-barycentric-corrected data (column `excalibur`). Each cell shows the median radial velocity shift for all features found between observations i and j.

To reduce the matrix to one array, we can perform another chi2 minimization fit, defined bellow in Eq. (11), in which we fit an array of parameters we can call V_r^i of length N (the number of observations), initialized to zero. The chi2 will be at its minimum when it has found an array of velocities V_r^i that best describe all the differences in the matrix ΔV_r^{ij} and each of the resulting velocities are thus relative not only to its neighbors but to all the other observations as well, thereby avoiding the correlation.

$$\chi^2 = \sum_{i,j=0}^{N} \left[\frac{\Delta V_r^{ij} - (V_r^i - V_r^j)}{\sigma(\Delta V_r^{ij})} \right]^2 \quad : \quad i < j. \tag{11}$$

To test the method, I have analyzed non-barycentric-corrected data. The resulting ΔV_r^{ij} matrix is the one plotted in figure 10 while the extracted relative radial velocities (the fit parameters V_r^i) are plotted in figure 11 (black), where we see a clear signal of Earth's movement. I have fitted the data with a periodic function and found a period of (366.482 ± 3e−6) days and an amplitude of (28.132 ± 9e−7) km/s. This is very close to the actual orbital speed of the Earth, although we should not expect that, as the star is unlikely to

lie in the plane of the Earth's orbit. There are however three other things to notice: 1) the error of the period does not cover the discrepancy with the actual value of 365.24 days, 2) the very high chi2 value, and 3) the p-value of zero. These suggest very clearly that my errors are wrong. Nevertheless, the signal is clear and the period found is definitely on the right order of magnitude, from which I confirm that my method, in general, is working. The direct velocity shifts between observations 1 and 2, 2 and 3, etc., are plotted in blue in figure 11, and it is clear that the last step of computing the relative shift between all observations is vital.

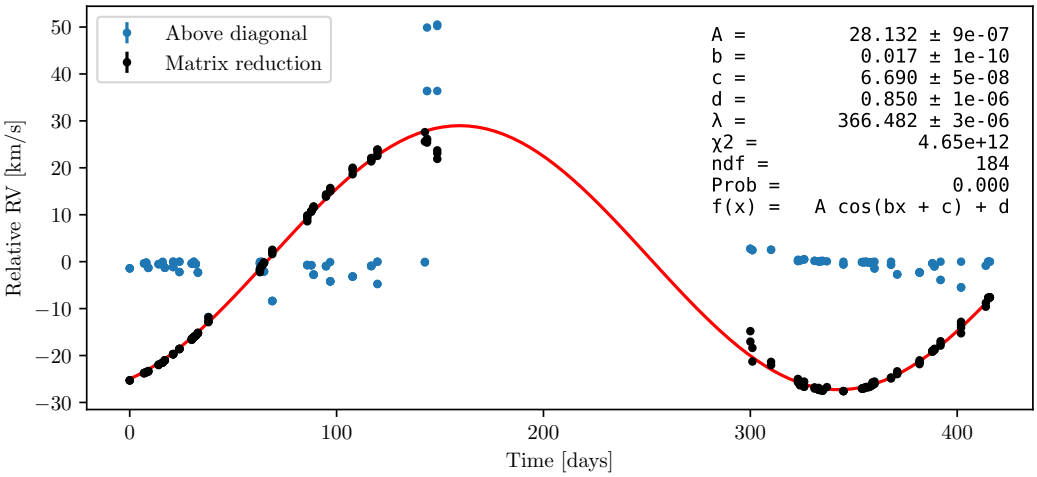

Figure 11: Final relative radial velocity results for 188 observations of HD 34411 using Excalibur calibrated but non-barycentric-corrected data (data column `excalibur`). Black: values computed through the overconstrained system approach. Blue: above diagonal of V_r^{ij}.

Although the number of computations necessary for the described analysis is high, it is possible to run it for 188 observations on a powerful laptop in about 2 hours. More run-times are listed in appendix A.

4 Results

Figure 12 shows my extracted relative radial velocity results for HD 34411 compared to those of Lily Zhao et al. (also available upon request [9]). As the Earth's movement has been removed, we are now on the scale of m/s instead of km/s. The method used by Lily Zhao et al. is a "chunk-by-chunk" method, where each spectrum is split into ∼2Å chunks for each of which an RV is found by shifting a template spectrum to match the observed spectrum. That is to say, a different method from mine. Our results however coincide within the same order of magnitude and with a similar rms of 1.67 m/s and 1.78 m/s, mine and theirs respectively. The residuals however do jump around quite a bit, and, of course, my mean error of 0.8 cm/s is not correct.

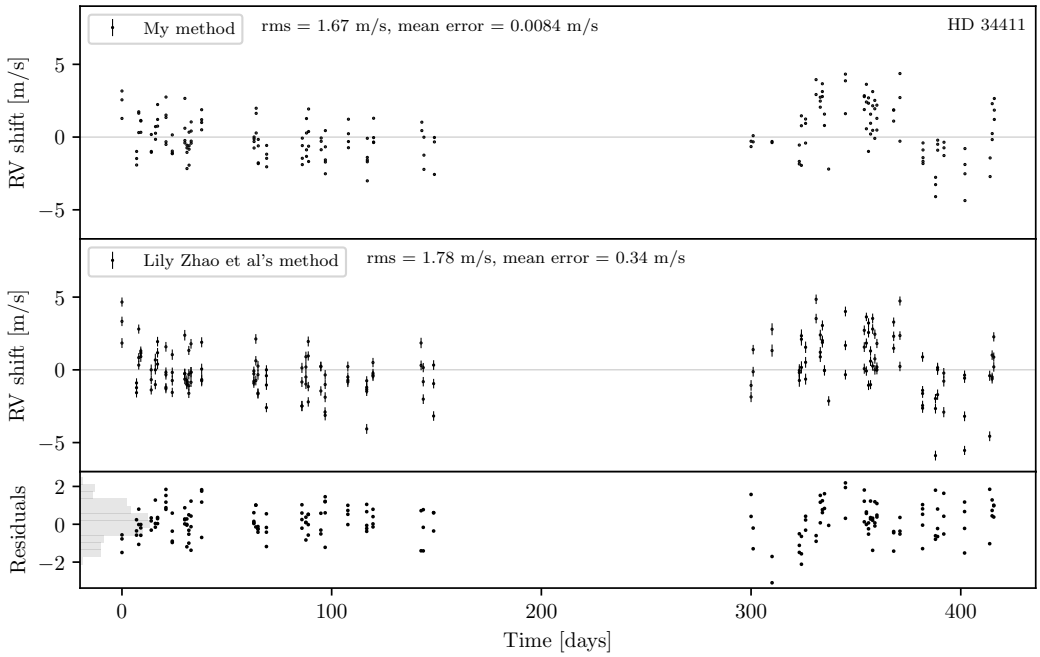

Figure 12: Top: my final extracted RV shifts for HD 34411 using Excalibur calibrated, barycentric-corrected data (data column `bary_excalibur`). Center: Lily Zhao et al.'s results using a chunk-by-chunk (CBC) analysis technique [9]. Bottom: residuals from subtracting Zhao's results from mine. 188 observations made between 2019-10-8 and 2020-11-27.

It is likely that the very small errors are due to the interpolation done when calculating the individual feature cross-correlations. The errors get smaller with increasing number of sampled data points, N, and looking again at figure 13 in appendix A, it appears that the error falls off as something like $1/\sqrt{N}$. A quick correction that one could make would thus be to "scale back" the errors by multiplying with $\sqrt{1000}$. Doing this, my mean error for HD 34411 comes out to 0.0084 m/s $\times \sqrt{1000} = 0.27$ m/s, which is about 20 % off from the results of Zhao et al.

I have analyzed data from three more stars and again compared the results with those of Zhao et al. As with HD 34411, my rms values come within a few m/s, but the residuals are big and my errors are too small. Rescaling is an improvement. Results are summarized in table 1, while plots can be found in appendix B.

Star	RMS	RMS (Zhao)	RMS deviation	Mean error (scaled)	Mean error (Zhao)	Error deviation
HD 34411	1.67	1.78	-20.6 %	0.27	0.34	-6.18 %
HD 101501	3.88	4.89	-20.7 %	1.1	0.35	214 %
HD 10700	1.48	1.86	-20.4 %	0.14	0.37	-62.2 %
HD 26965	3.07	3.19	-3.76 %	0.15	0.33	-54.5 %

Table 1: Summery of my results compared to those of Lily Zhao et al. All values listed are in m/s, except for the deviations, which are in percentage. Mean error (scaled) is the mean of the errors obtained from my method scaled by $\sqrt{1000}$.

5 Discussion

For the calibration, my best solution was the cubic-spline interpolation, with which I got a per-line rms of 10.0 m/s. This is not of much help in the pursuit of rv precision to centimeters per second and nowhere near that of Excalibur, which, obviously using more advanced techniques, achieves an rms of 3.3 cm/s [10].

The results from analyzing the non-barycentric-corrected data were not on point. I measured one Earth-year to be ∼366.5 days with a very small uncertainty. One day off. If EXPRES should be capable of measuring stellar RV to a precision of a dozen centimeters per second, I think I should be able to get this value right. However, considering the crudeness of my method and especially the resulting very large spread of the RVs computed for individual features (as seen in figure 9), it is perhaps not surprising, but rather justifies the decade-long development of much more advanced RV extraction methods.

Nevertheless, the signal is there. And, comparing my results for the barycentric-corrected data to that of Zhao et al. is also quite positive. RVs on the same order and many patterns in their results coincide with mine. The residuals from subtracting their results from mine are however quite large. For all four stars analyzed my rms also comes out smaller than theirs. This might be due to my match filtering, which is biased toward selecting features that are closer together if those features do not differ too much in shape.

As mentioned, the small errors could be due to the interpolation and sampling of extra data points. A way around this could be to interpolate not both peaks but only one. Instead of sampling new data points, I would just evaluate the interpolating function in the data-point locations of the un-interpolated peak and thereby have intensity values on a common wavelength range. As there might be some bias from interpolating only one of the peaks, one could then flip it around and interpolate the other one instead, finally taking the mean of the two results. This would lead to twice as many computations but also a lot fewer data-points. Depending on the overhead of iminuit I imagine it could even out. At a first attempt, however, I often got statistically incompatible results when flipping whose turn it was to be interpolated. Further investigation is necessary.

Another possible approach is to compute the radial velocity shift for each feature across all observations individually. This way one could also analyze the stability of each feature over time, as Zhao et al. do, and weight it accordingly, and one could perhaps even discover other patterns of stellar activity. Identifying and matching features across all observations might however be challenging, as, for instance due to stellar activity, some features might move significantly and overlap with others. If one is simply interested in the stable features however, this would be an easy way to filter out the bad ones.

Having spent much more time during this project on exploring the data and devising solutions with my supervisor than reading papers, there is obviously much inspiration to gather from digging down more carefully in the literature. It is also conceivable that through the application of knowledge of stellar activity along with machine learning and a lot of data, features could be grouped and categorized according to the cause of their shift, which would be another way to sort out features that are unstable over time.

Finally, it is interesting to note that my solution and that of Zhao et al. do not yield completely overlapping results, although both solutions appear good. If two good, independent solutions do not give the same results, they must be extracting different information. It is therefore conceivable that a better solution exists, which extracts the combined information of both methods. That being said, there is no doubt, however, that their solution is better than mine and it is equally possible that the bias in my method is causing the deviation.

6 Acknowledgements

I would like to express my gratitude first and foremost to my supervisor, Prof. Troels C. Petersen, for his guidance and patience throughout this project and for sharing his plentiful experience and passion for crafting creative and clever solutions to complicated problems. Many of the solutions explored in this report arose based on ideas of his. Working with him has been a great source of joy and inspiration. Secondly, I would like to thank Prof. Lars A. Buchhave for sharing his extensive background knowledge of all things spectrographs and calibration. Finally, I would like to thank other students of Petersen's that have also taken part in the research, Liam J. Ward (MSc), Christian Michelsen (Ph.D), and Simon C. Debes (MSc), for the many great discussions and support.

References

[1] Ryan T Blackman, JM Joel Ong, and Debra A Fischer. The measured impact of chromatic atmospheric effects on barycentric corrections: Results from the extreme precision spectrograph. *The Astronomical Journal*, 158(1):40, 2019.

[2] John M Brewer, Debra A Fischer, Jeff A Valenti, and Nikolai Piskunov. Spectral properties of cool stars: extended abundance analysis of 1,617 planet-search stars. *The Astrophysical Journal Supplement Series*, 225(2):32, 2016.

[3] Bradley W Carroll and Dale A Ostlie. *An introduction to modern astrophysics*. Cambridge University Press, 3. edition, 2017.

[4] C Jurgenson, D Fischer, T McCracken, D Sawyer, A Szymkowiak, Allen Davis, G Muller, and F Santoro. Expres: a next generation rv spectrograph in the search for earth-like worlds. In *Ground-based and Airborne Instrumentation for Astronomy VI*, volume 9908, page 99086T. International Society for Optics and Photonics, 2016.

[5] Lennart Lindegren and Dainis Dravins. The fundamental definition of "radial velocity". *Astronomy & Astrophysics*, 401(3):1185–1201, 2003.

[6] Christophe Lovis, Debra Fischer, et al. Radial velocity techniques for exoplanets. *Exoplanets*, pages 27–53, 2010.

[7] Ryan R Petersburg, JM Joel Ong, Lily L Zhao, Ryan T Blackman, John M Brewer, Lars A Buchhave, Samuel HC Cabot, Allen B Davis, Colby A Jurgenson, Christopher Leet, et al. An extreme-precision radial-velocity pipeline: First radial velocities from expres. *The Astronomical Journal*, 159(5):187, 2020.

[8] Lily L. Zhao. Excalibur. `https://github.com/lilylingzhao/excalibur`, 2020. Accessed: 2022-05-15.

[9] Lily L. Zhao et al. Expres stellar-signals project. `https://exoplanets.astro.yale.edu/science/activity.php`. Accessed: 2022-05-15.

[10] Lily L. Zhao, David W Hogg, Megan Bedell, and Debra A Fischer. Excalibur: A nonparametric, hierarchical wavelength calibration method for a precision spectrograph. *The Astronomical Journal*, 161(2):80, 2021.

Appendix A Additional details

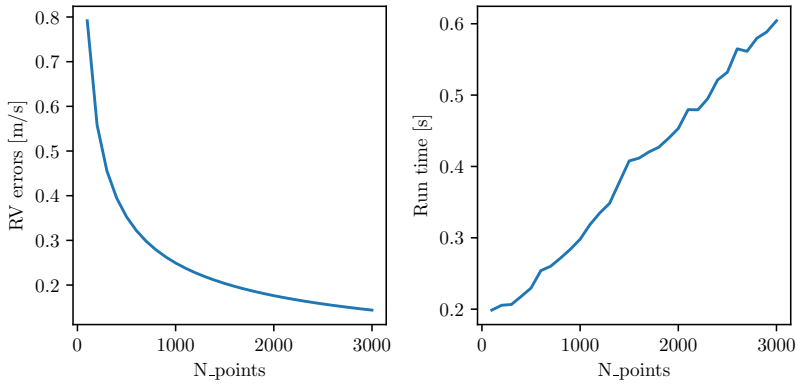

Figure 13: Resulting error in extracted radial velocity shift between two observations (left) and run time of that computation (right) as a function of number of sampled points in the interpolation.

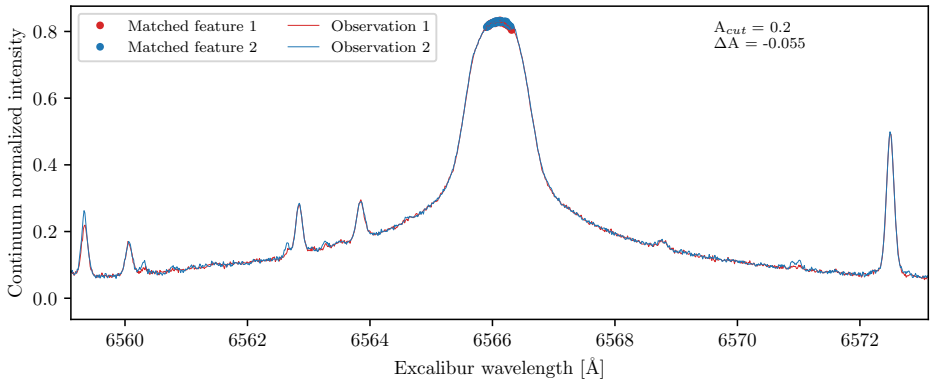

Figure 14: Example of a bad match that made it through the filters. This match is actually correct, but yields a bad result because it is too broad for the standard data slice size.

Computational run times

For running the analysis I have used a Macbook M1 pro (2021).

Calibration

- Fitting super-Gaussian to 17,068 LFC lines takes about 45 seconds (one core).
- Computing the interpolation function and fitting the polynomials takes less than a second.

RV extractions

- Finding all features in 188 files takes just about 1 minute (one core).
- The run time for computing the ΔV_r^{ij} matrix of course scales as N^2. The following times are for running the computation in parallel on 6 CPU cores:

HD 34411	$N_{\text{files}} = 188$	$N_{\text{features}} = 12{,}608{,}511$	:	121 minutes
HD 10700	$N_{\text{files}} = 174$	$N_{\text{features}} = 9{,}008{,}471$	:	91 minutes
HD 26965	$N_{\text{files}} = 114$	$N_{\text{features}} = 5{,}777{,}355$	:	62 minutes
HD 101501	$N_{\text{files}} = 45$	$N_{\text{features}} = 815{,}719$	:	9 minutes

- The subsequent reduction of the ΔV_r^{ij} matrix takes up to five minutes on one core.

Appendix B Additional results

Relative radial velocities computed for three more stars: HD 101501, HD 10700, and HD 26965. In each figure the top shows my final extracted RV shifts using excalibur calibrated, barycentric-corrected data (data column `bary_excalibur`), the center shows Zhao et al.'s results using a chunk-by-chunk (CBC) analysis technique [9], and the bottom shows residuals from subtracting Zhao's results from mine.

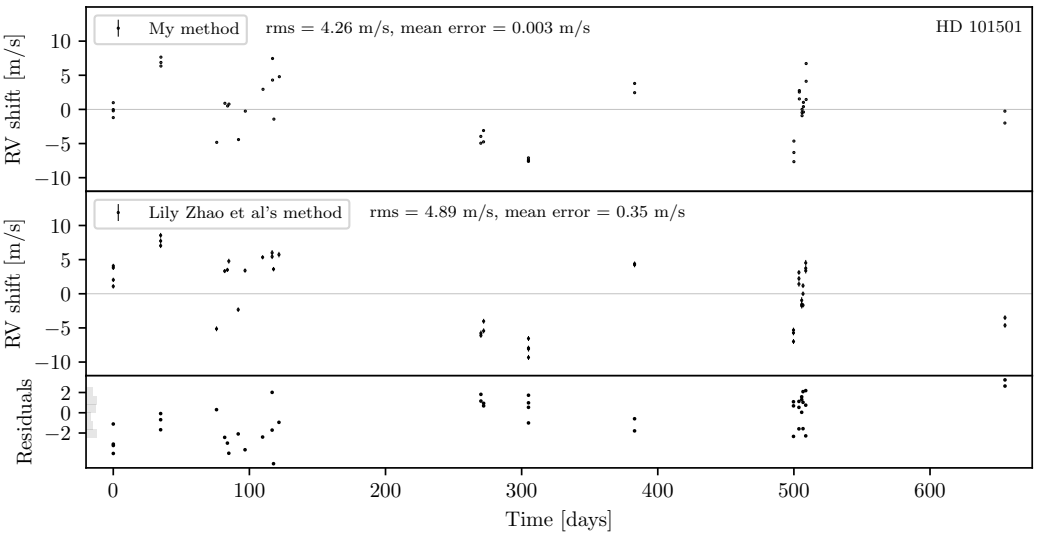

Figure 15: HD 101501
45 observations made between 2019-2-10 and 2020-11-26.

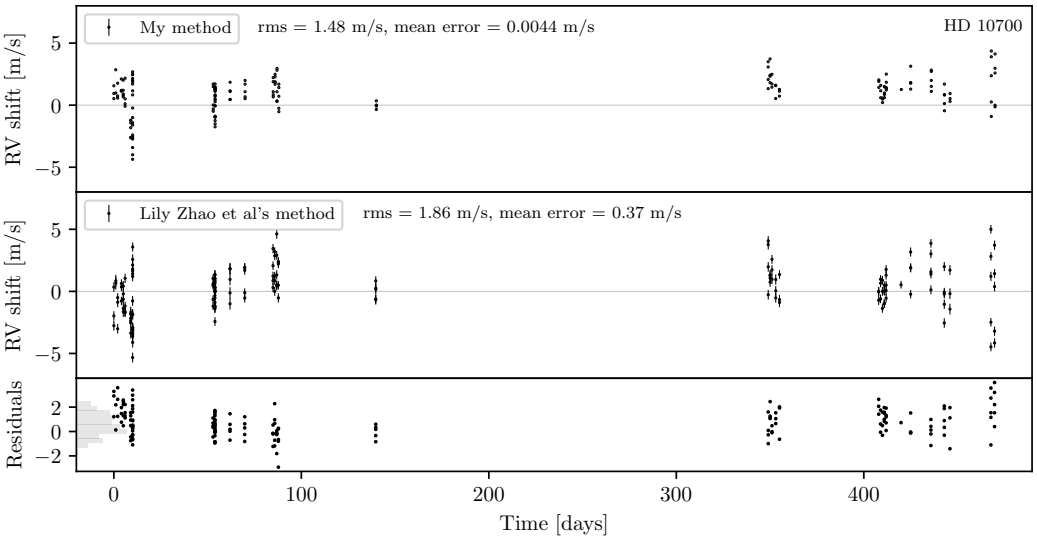

Figure 16: HD 10700
174 observations made between 2019-8-15 and 2020-11-27.

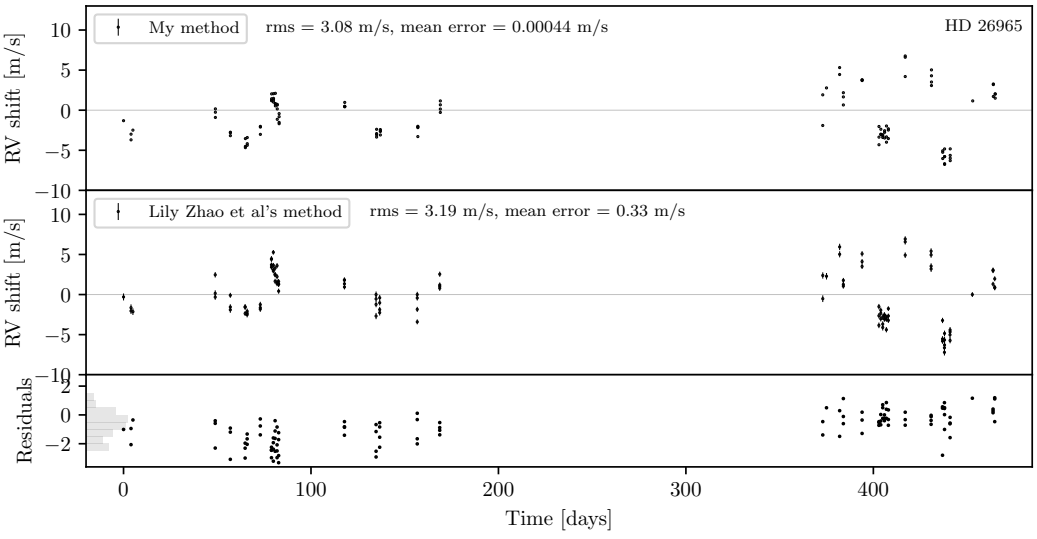

Figure 17: HD 26965
114 observations made between 2019-8-20 and 2020-11-27.