Bachelor Project in Physics

Jakob Harteg, wmc573

June 2022, University of Copenhagen

todo: Make cover page

todo: Add abstract

1 Note about the project

In the beginning of February 2022, Troels Petersen (KU) and Lars Buchhave (DTU) started working together to improve computational LFC calibration methods, and Troels invited me and a few other students to come along for the ride. During the project I've been exploring real data and developing a crude method for performing LFC calibrations and subsequent radial velocity extractions. With this project report it is my intent 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.

2 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 exoplanets [5]. With new extreme-precision radial velocity (EPRV) spectrographs such as the EXtreme PREcision Spectrograph (EXPRES), data from which this project is based on, 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), 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've ignored many aspects of it during this project. I've worked directly on the LFC calibration and RV extractions and tried to orient myself about the most important corrections, thus that is what I will describe in this report. The full pipeline is described in detail in [6].

3 Theory

3.1 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, compared to the star, the larger this movement will be. For now, it is not possible to directly observe exoplanets. One way of indirect detection however is thus measuring the movement of the star. We can measure the relative movement of a star through the doppler effect; the electromagnetic spectrum of the star observed on Earth will be blue shifted when the star is moving toward us and red shifted when moving away. The potential indirect signal of a planet will thus consist of a periodic doppler shift in the spectrum of the host star. For this method in general, several years of data collection is necessary. One should at least have one full orbit of observational data, and, to decrease statistical uncertainties, several orbits is preferential.

A large planet like Jupiter induces a radial velocity (RV) in the Sun of about 12.7 m/s when observed in its plane of orbit. While a small one like Earth only induces an RV of about 9 cm/s. (p. 29, [5]).

todo: Possibly describe the RV calculation in more detail and compute Earth K.

3.1.1 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 betweeen the source and an observer at zero gravitatonal potential is given by

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

which accounts for both special relativistic effects and gravitatonal doppler shift described by general relativity. Where λ is the observed wavelength, λ_0 is the emitted wavelength, Φ is the Newtonian gravitational potential at the source ($\Phi = GM/r$ at a distance r of a spherically symmetric mass M), $\mathbf{k}$ is the unit vector pointing from the observer to the source, $\mathbf{v}$ is the velocity of the source relative to the observer and c is the speed of light [4].

Special relativistic effects we can safely ignore, as we are dealing with velocity shifts on the order of meters or centimeters per second, and thereby cross out the third term in the denominator. The remaining terms $(1 - \Phi/c^2)$ evaluted for HD 34411 ($M = (1.08 \pm 0.14)M_\odot$, $R = (1.28 \pm 0.04)R_\odot$ [2]) is around 0.999998. Adding this term to the computation of RV does not change my results, so we can neglect the denominator completely.

todo: How to check analytically though?

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 inclication angle, a possible simplification is to omit $\mathbf{k}$ is treat the resulting $\mathbf{v}$ as a minimum radial velocity.

Thus we are left with

$$\lambda = \lambda_0 \times \left(1 + \frac{v}{c}\right), \tag{2}$$

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.

3.2 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.

Like in 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 is EXPRES an STA1600LN CCD backside illuminated image sensor with a $10,560 \times 10,560$ array containing 9µm×9µm pixels, designed to with a wavelength range of $3800-7800$Å. 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 wrap 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 in 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 can otherwise cause the spectra to change position on the CCD and thus lead to errors in the RV mesaurements.

Calibration device: Wavelength calibrations are performed with the use of a Laser Frequency Comb (LFC), produced by Menlo Systems, which is a laser source whose spectrum consists 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 spectral resolution of $R = 150'000$, where R is defined as $R = \lambda/\Delta\lambda$. Inverting this we get what's called the resolution element of the Instrumental Spread Function (ISF) (or Line Spread Function (LSF)), $\Delta\lambda = \lambda/R$. For a wavelength of say of $\lambda = 5000$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 Gausssian function with FWHM $= \Delta\lambda$. Specifically for EXPRES, a super-gauss is however a better fit. 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 peaks are seperated by about 10 pixels to remain distinct after this blurring. For star-spectra however some emission and absorption lines will be too close together and will appear "blended" on the detector.

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 [3].

3.3 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 Zhao and is by no means raw data, but data that has already gone through a lot of processing.

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

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

The observations for HD34411 are plotted in figure 1. Most days have 3-4 observations, and there are significant gaps in the data as well.

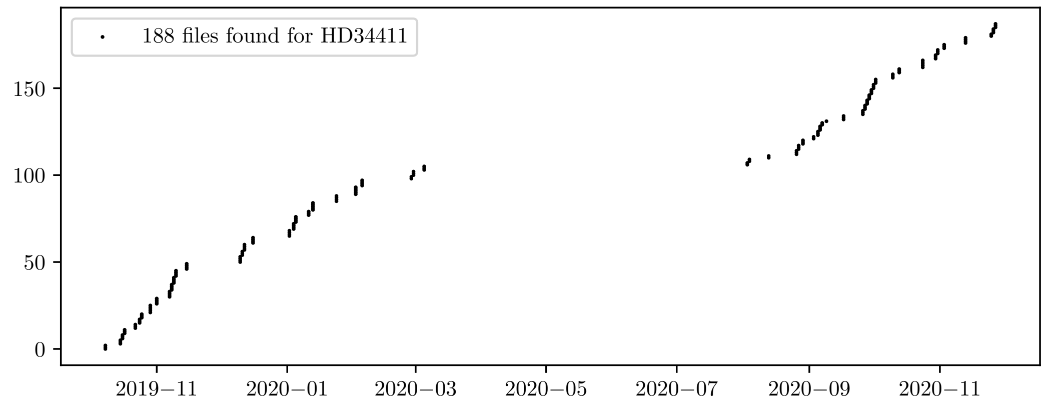

Figure 1: EXPRES observations of HD 34411

todo: Specify which data was used for LFC calibration method development

3.3.1 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 mesaurements for each pixel on the CCD. The rows of the CCD data are referred to as orders. There are 86 orders each of which has 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 elonganted dimensions of 86×7920, but as mentioned earlier, the CCD is actually square. The orders however 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 [6].

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 which shows which pixels are should be used.

To perform the calibration the following variables are used: `spectrum`, `uncertainty`, `wavelength` and `continuum`. To perform the RV extraction on the pre-calibrated data the following variables are used: `bary_excalbur`, `excalibur_mask`, `spectrum`, `uncertainty` and `continuum`.

3.3.2 Noise

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 emperically, details of which will not be discussed, but is assumed to be consistent throughout each night of observation.

3.3.3 Correction: Scatter / blaze

Although manufactures have tried their best to limit it, the CCD still gets hits by scattering light, being the strongest in the center of the detector. It is modeled and subtracted by measureing the photon count in between orders.

3.3.4 Correction: Tellurics

Tellurics (general definition being originating from the Earth) in this context refers to the contamination that ground based spectrographs have to deal with, which occurs as the light passes through the Earth's atmosphere, encountering molecules such as oxygen and water vapor. On EXPRES the technique used is called SELENITE [https://arxiv.org/pdf/1903.08350.pdf].

3.3.5 Correction: Barycentric corrections

todo:

4 Data Analysis

4.1 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. To do this, we need a light source with known frequencies, preferably many discrete peaks. EXPRES uses a Thorium Argon lamp for an initial trial wavelength solution and then a laser frequency comb (LFC) for an more precise solution.

The Thorium Argon lamp produces 4,000 lines across 82 orders, which can be identified and mapped to a wavelength through a *line atlas*. An intial wavelength solution for all pixels is then produced by linear interpolation. (In this project I have not done this calibration).

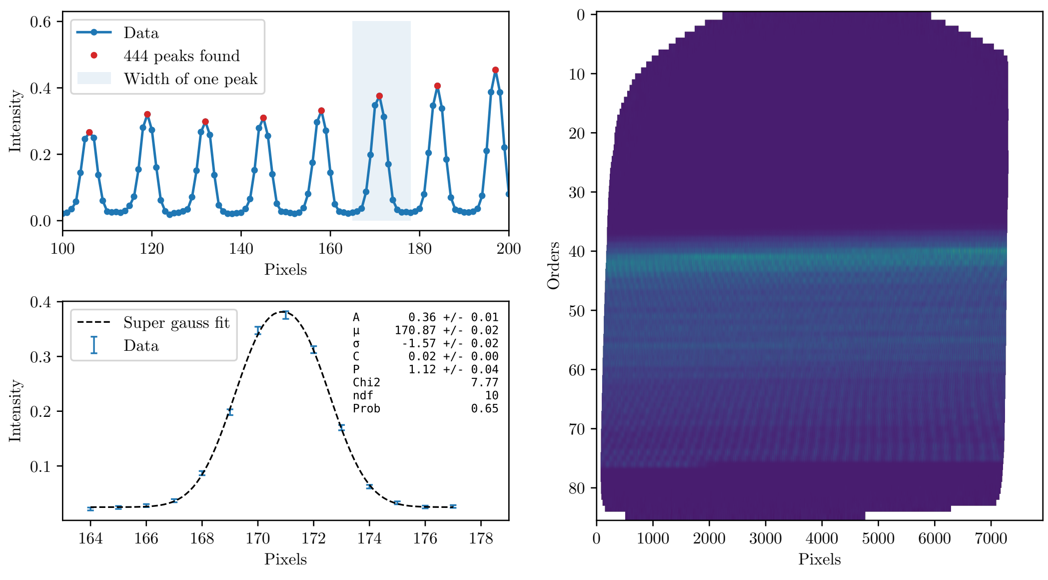

Figure 2: Right: Measured intensities for the LFC across the CCD (unitless). Upper left: illustration of a few LFC peaks in order 65. Peaks are identified with scipy peak finder. Lower left: each peak is fitted with a super gauss to find the exact top of the peak with uncertainties.

The LFC generates a series of equadistant (evenly spaced) spectral lines, typically 20,000 lines across 50 orders. The range of the LFC is thus shorter, and for this reason the ThAr exposures can also be used for a rough calibration outside the LFC range. The frequencies of the LFC peaks are given by the relation

$$v_n = v_{\text{rep}} \times n + v_{\text{offset}} \tag{3}$$

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 1 s. (p. 8, [6]). The values I have used in the calibration, $v_{\text{rep}} = 14e9$ and $v_{\text{offset}} = 6.19e9$, were provided by Lars Buchhave, but may be outdated. See figure 2 right side for a plot of the intensities measured across the CCD.

The following procedure is followed to determine the location of the LFC peaks on the CCD: 1) Find peaks using scipy peak finding algorithm 2) make data slices around each peak with the size of the average distance between peaks, 3) using iminuit do a χ^2 minimisation fit to each peak with a super-gauss plus a linear background. See figure 2 left side.

A super-gauss, defined in eq. (4), 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.

$$f(x; A, B, C, P, \mu, \sigma) = A \exp\left(-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)^P\right) + B(x-\mu) + C \tag{4}$$

The fit then is a minimisation of

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - f(x; A, B, C, P, \mu, \sigma)}{\sigma_i}\right]^2 \tag{5}$$

Where N is the number of data points, x is pixel-space, y_i and σ_i is the measured photon count and uncertainty respectively. The fit returns the values and uncertainties for the parameters A, B, C, P, μ, σ when the value of χ^2 is minimized.

We are most interested in μ, which gives the position of the LFC peak on the CCD (in pixel-space). With the intial rough wavelength solution dervied from the ThAr lamp (pre-calculated in the data set that I've used) 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. 3. And we now have a map of 20,000 points on the CCD with a good wavelength solution.

Of course we need to have a wavelength solution for all points on the CCD and to do that I have explored two approaches: polynomial fitting and cubic interpolation.

4.1.1 Poly-fit calibration

Since the LFC peak positions, as seen in the right plot in figure 2, appear to exhibit a certain periodic behavour, my initial approach to compute a wavelength solution across the whole CCD was to fit the LFC peak positions with a polynomial. Looking at the residuals of fitting the LFC peaks with polynomial of increasing degree revealed smaller and smaller periodic variations, until reaching 5th degree, see figure 3.

But, as the orders are far from idential, this turns out not to work very well across the CCD, see figure 4. So I decided to turn to interpolation.

4.1.2 Interpolation calibration

A cubic interpolation would force all the residuals to be zero, so in order to evaluate the quality of the method, we can for instance omit every second peak from the interpolation and then compute the residual between the omitted peaks and the resulting interpolation function. We can additionally flip the data set and interpolate the peaks we left out before and compute residuals for the rest, thus ending up with an array of residuals equal to the length of the results from the poly-fit method, allowing us to compare the two, as is done in figure 5.

The RMS of the residuals from the interpolation come out much smaller than that of the polyfit (values specified in figure 5), in this example, suggesting that the interpolation method is superior. 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.

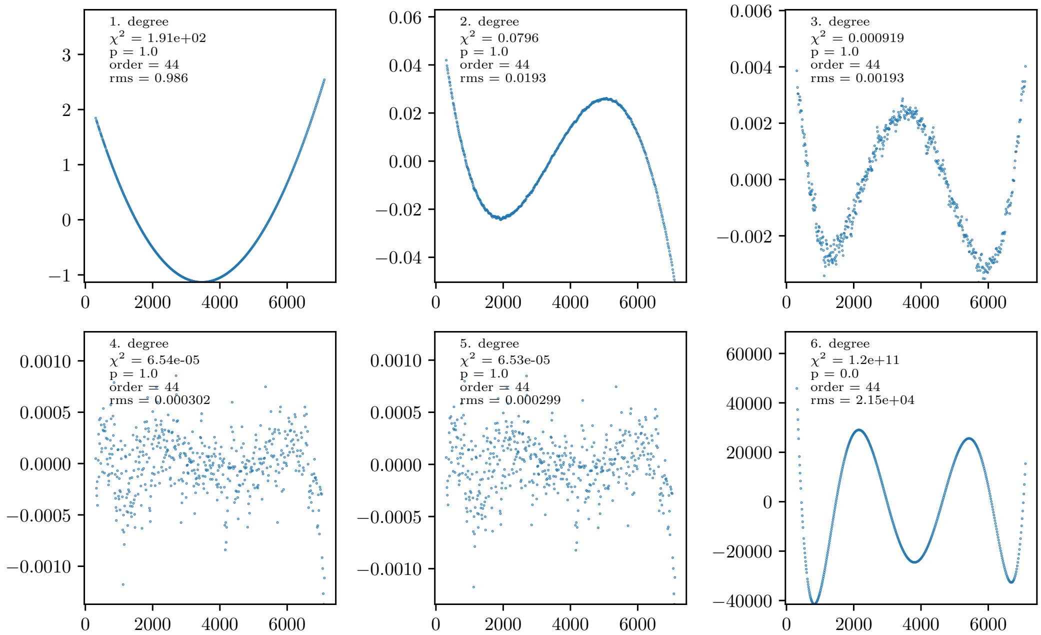

Figure 3: Residuals from fitting LFC peak positions with polynomials of increasing degree. Pixels on the x-axis and residuals in ångstrom on the y-axis

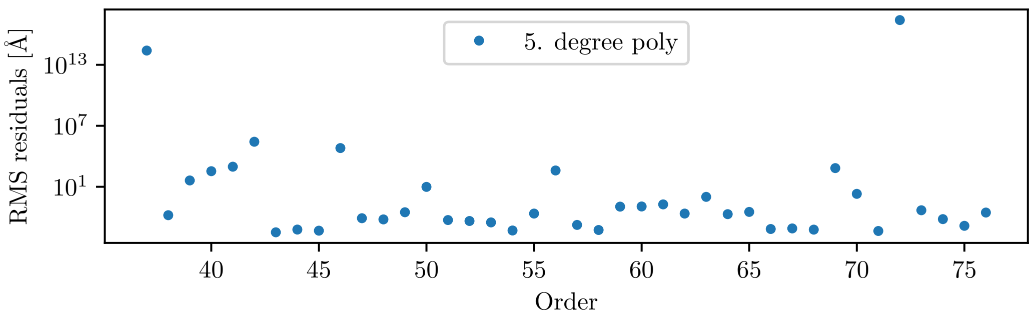

Figure 4: RMS of residuals from fitting LFC peak positions with 5. degree polynomial. Pixels on the x-axis and residuals in ångstrom on the y-axis

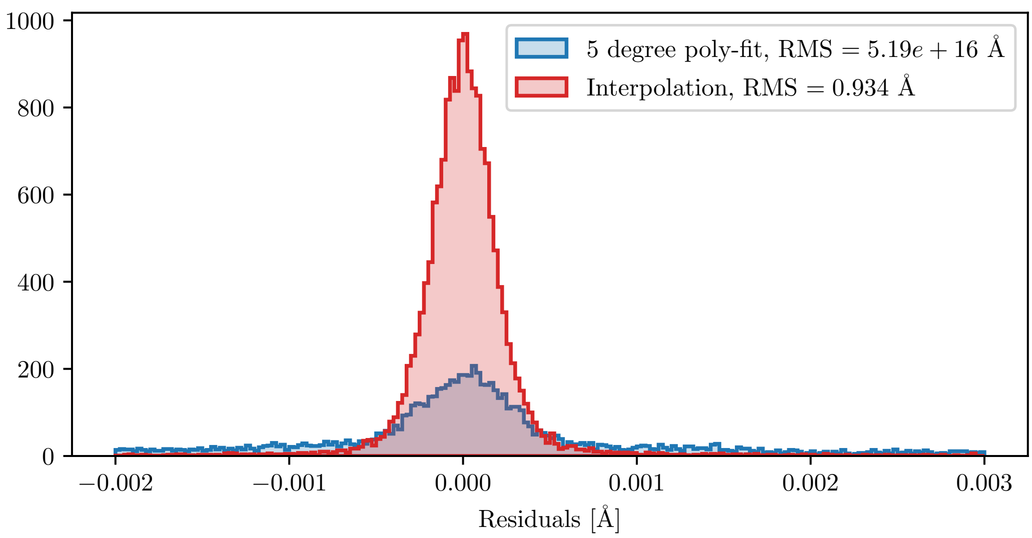

Figure 5: Residuals from calibrations performed through a 5th degree poly-fit and a cubic interpolation. Both results contain approximately the same amount of points.

todo: Can we convert this RMS to an RV error ?

todo: add graph comparing residuals using gauss vs super gauss

todo: perhaps add plot of changes in parameters across the CCD

4.1.3 Errors in the calibration data

The LFC fits files come with an uncertainty on the intensity. It appears however that this uncertainty might be a bit underestimated. We can see this by plotting the χ^2- and P-values for the LFC peak super-gauss fits, as done in figure 6. The χ^2 value should be roughly equal to the number of degrees of freedom in the fit (eq. 4), which is:

$$N_{\text{dof}} = N_{\text{data-points}} - N_{\text{fit-parameters}} = 13 - 5 = 8, \tag{6}$$

as I use about 13 data points in each fit. This is however not the case for the original data (red). Looking at the bottom left plot (red) of the p-values, we can also see that the majority of the fits are very bad. The red chi2 values peak, although only slightly, at around 25, which suggests that the errors are about a factor $\sqrt{3}$ too small (square-root because χ^2 of course is squared and $25/3 \sim 8$). Scaling the errors up by $\sqrt{3}$ yields a chi2 peak at 7.5 and a more flat p-value distribution (green). Scaling by $\sqrt{10}$ is too much (blue). More scale-factors are presented in figure 13 in appendix A.

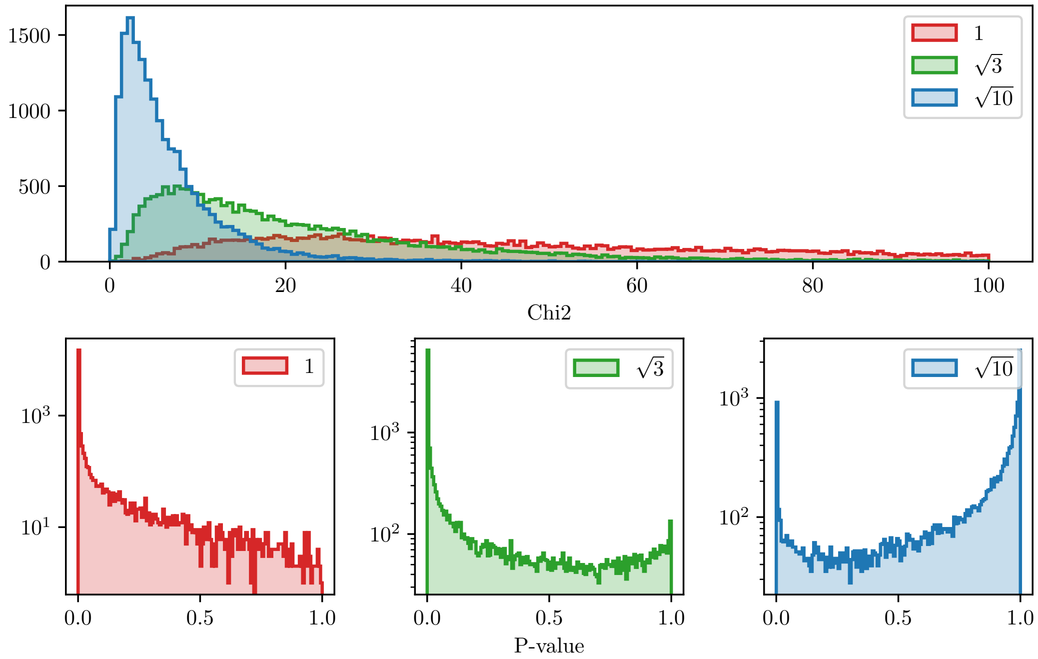

Figure 6: Chi2-values and p-values from individual LFC peak super-gauss fits with photon count (spectrum) errors multiplied by different scale-factors (1, $\sqrt{3}$ and $\sqrt{10}$). See text for more details.

Effects on calibration:
The errors used during the production of the calibration residuals shown in figure 5 I have already multiplied by $\sqrt{3}$. This gave a $\sigma = 0.934$, without this correction I got $\sigma = 2.34$.

4.2 RV 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've 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've worked on RV extractions using already calibrated data provided by Lily Zhao. This data has been calibrated using a technique called excalibur [7].

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. On the right side is plotted all continuum normalized data within the EXCALIBUR mask, i.e. data marked as having a proper calibration.

4.2.1 Finding and matching features across observations

In order to measure how much individual absorption features move in between observations the first challenge is to find the "same" features in both observations. What follows is a

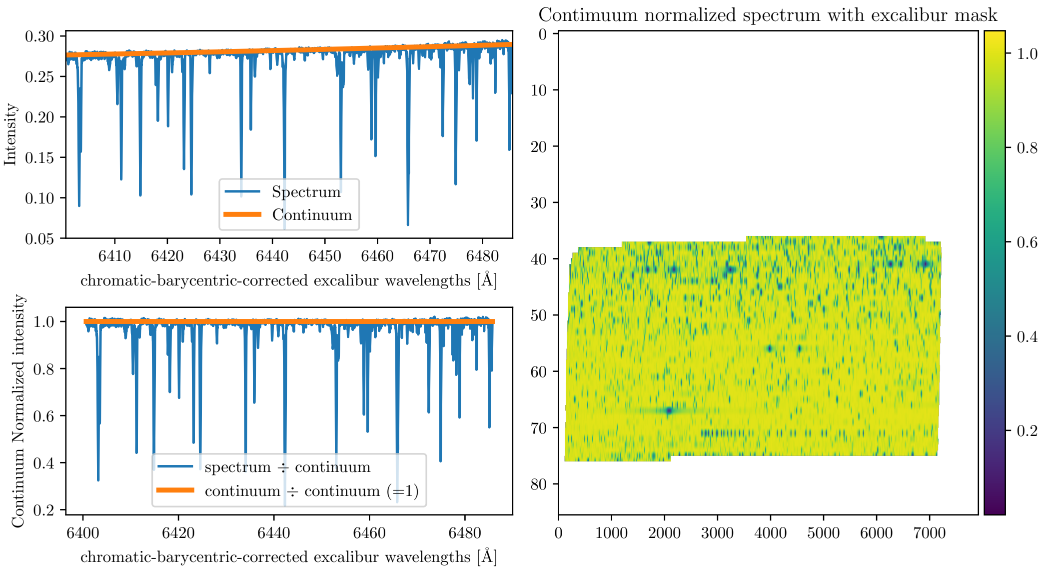

Figure 7: Overview of excalibur calibrated data from an observation of HD 34411. Upper left: extract of wavelength solution vs. intensity. Lower left: continuum normalized spectrum. Right: all continuum normalized data within the excalibur mask.

quick rundown of the procedure I've devised:

- 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:

- Iterate through the peaks of observations1 and 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 sometime 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 equivalant to a radial velocity greater then 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 actually 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, equivalant to about 20-30 km/s depending on the wavelength, filters out the few very bad matches, but leaves the rest.
  - Maximum difference between the area under the graph of two peaks: i.e. the sum of the intensity values. Peaks with similar shapes will give a low difference. Some bad matches can be avoided with this filter, but it also excludes a lot of good matches. Ideally I would keep the filter as strict as possible while maintaining a generous amount of matches. In figure 8 is plotted the number of matches found between all files as a function of the maximum area difference. From that I conclude that 0.2 is a good option, as the number of matches doesn't rise much with a looser restriction.

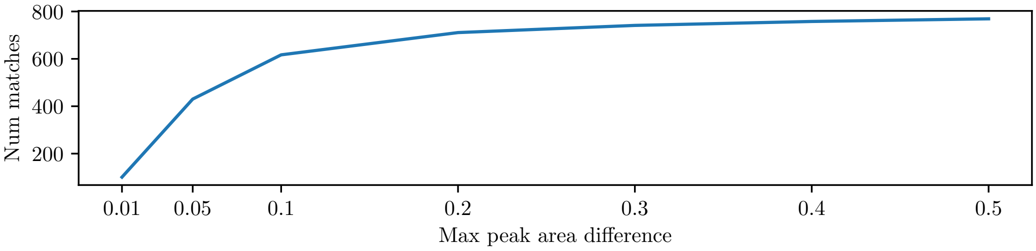

Figure 8: Average number of matches found as a function of max peak area difference for all features in observations for HD 34411.

4.2.2 Cross correlation

At this point, I have a list of matching features in different observations. The cross-correlations will be performed as a chi2 minimization fit to find the radial velocity that for each match correctly shifts one feature onto the other one.

Before moving on though, we have to cubic-spline interpolate the spectra data. I do this 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) in order 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 equation 2. 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:

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - f(x;v)}{\sigma_i} \right]^2 \tag{7}$$

where y_i are the unshifted interpolated intensity from the first observation, σ_i are 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 equation 2, evaluated on the wavelength range common to both features:

$$f(x;v) = \mathbf{interp}[x \times (1 + v/c), y]\,(x_{\text{common}}) \tag{8}$$

I then compute the cross-correlation and obtain the radial velocity, v, as a minimization of eq (7) using iminuit.

todo: Talk about the mean/median etc

4.2.3 Extracting relative shifts from over an overconstrained system

With the devised method we can now compute the relative readial 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 observation 1 and 10 will depend on all the observations in bewteen, and if there is one bad one, this will affect all the rest. To circumvent this, we can compute the relative shift bewteen 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 9 for an example.

To reduce the matrix to one array, we can perform another chi2 minimization fit, defined bellow in equation 9, in which we fit an array of parameters we can call V_r^i of length N (the number of observations), simply 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}

[1] 1000 steps appeared as the best balance between run time and resulting uncertainty. See figure 15 in appendix B

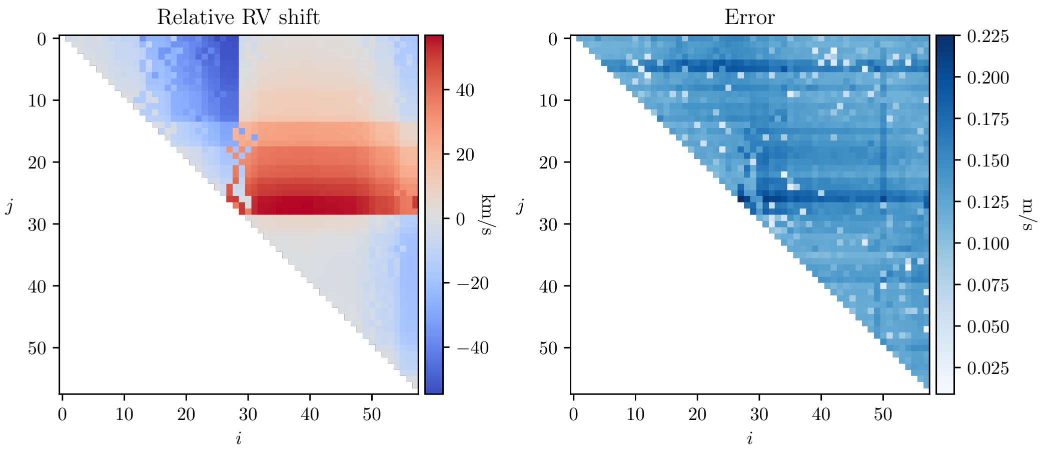

Figure 9: RV shifts matrix computed for HD34411 using non-calibrated, non-barycentric-corrected data (column `wavelength`). Each cell shows the median relative radial velocity across all features found between observations i and j. The diagonal should be zero and has been omitted for the sake of computational speed. I only used one out of the several exposures from each day.

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.

$$\chi^2 = \sum_{i,j=0}^{N} \left[\frac{\Delta V_r^{ij} - (V_r^i - V_r^j)}{\sigma(\Delta V_r^{ij})} \right]^2 \quad : \quad i < j. \tag{9}$$

For the sake of illustrating the method, I've analysed non-barycentric corrected data, in which we should be able to see the movement of the Earth around the center of mass of the solar system. That is the data plotted in the matrix in figure 9 and the resulting extracted relative radial velocities (the fit parameters V_r^i) are plotted in figure 10 (black), where we see a clear signal of Earth's movement. I've fitted the data with a periodic function and found that the period is $(367.522 \pm 5\text{e}{-5})$ days. Then there are three things to notice: 1) the error does not cover the discrepancy with the acrtual 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, from which I confirm that my method in general is working. Comparing with the direct differences between observations plotted in blue in figure 10, it is also clear that the last step of computing the relative shift between all observatons is vital.

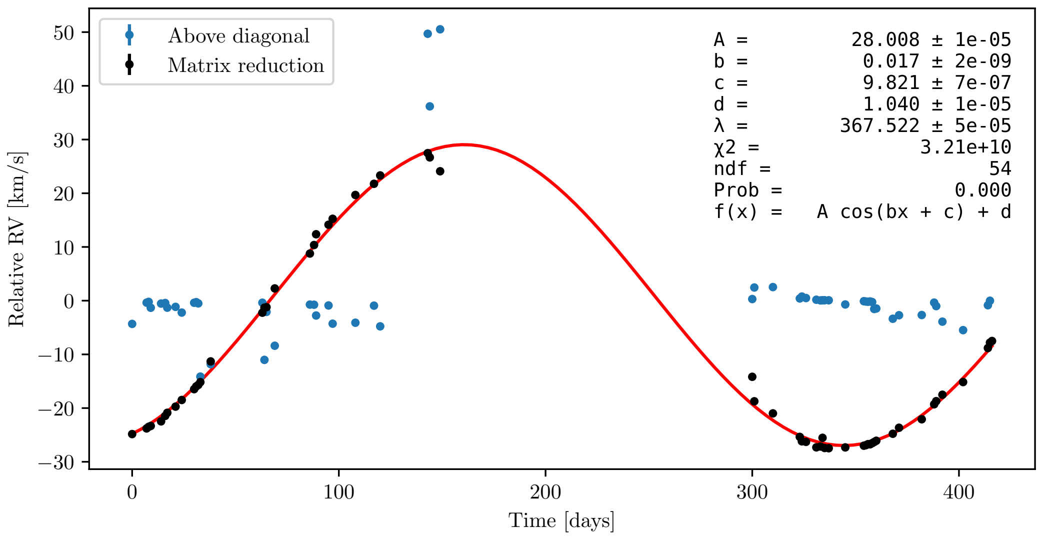

Figure 10: Computed relative wavelength shifts for HD34411 using non-calibrated, non-barycentric-corrected data (column `wavelength`). Blue: the shifts from day to day. Black: values computed through the chi2 minimization.

4.2.4 Computational run times

todo: add number of function calls for the fitting todo: add number of peak analyzed and so on

todo: add run times

5 Results

In figure 11 is plotted the ΔV_r^{ij} matrix for barycentric-corrected excalibur calibrated data for HD 34411. We are now on the order of m/s instead of km/s.

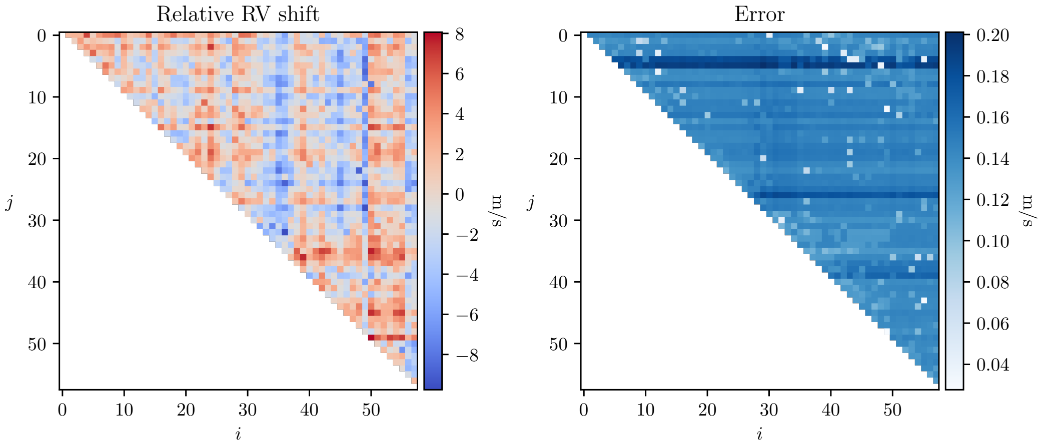

Figure 11: RV shifts matrix computed for HD34411 using excalibur calibrated, barycentric-corrected data (column `bary_excalibur`). Each cell shows the median relative radial velocity across all features found between observations i and j. The diagonal should be zero and has been ommited for the sake of computational speed.

Performing the matrix reduction chi2 fit yields the results plotted in figure

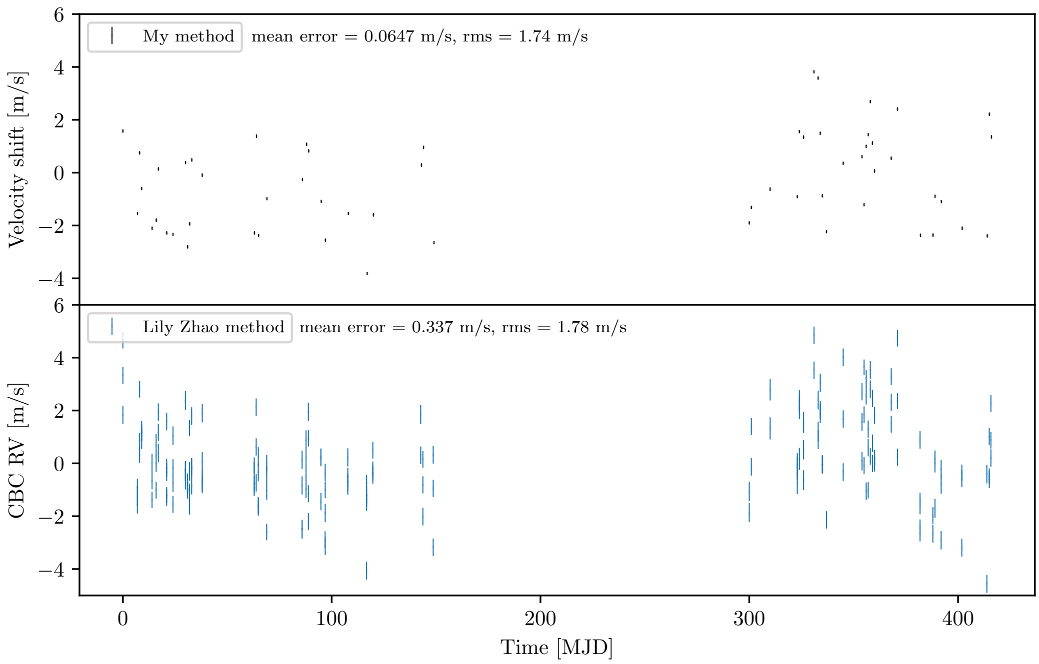

Figure 12: Computed relative wavelength shifts for HD34411 using excalibur calibrated, barycentric-corrected data (column `bary_excalibur`). Top: my results **todo: Errors are too small**. Bottom: Lily Zhao's results using a chunk-by-chunk analysis. **todo: Cite?**

todo: Add plot with all files **todo: Add plots for the rest of the stars**

6 Discussion

Future ideas:

- Errors
- Errors from the interpolation, the more points we sample the smaller the error??
- Filters: chauvenet's, rv cut, z-score cut

Future ideas:

- Auto encoder
- Lunch LFC into orbit around earth to have a base truth
- Better understand the stellar activities

7 Conclusion

8 Acknowledgements

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] 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.

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

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

[6] 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.

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

A LFC photon count errors

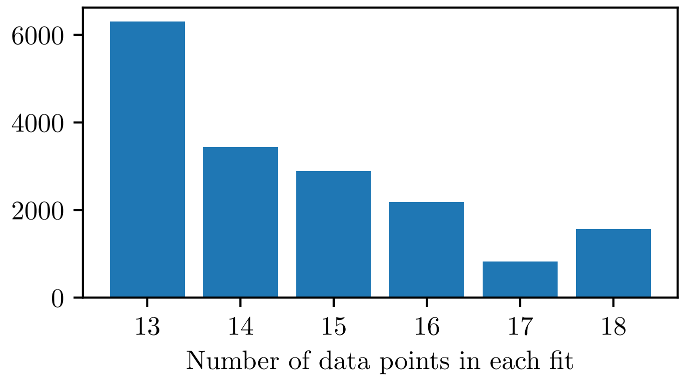

Figure 13: Number of data points in each LFC peak fit, determined by the average distance between peaks in each order.

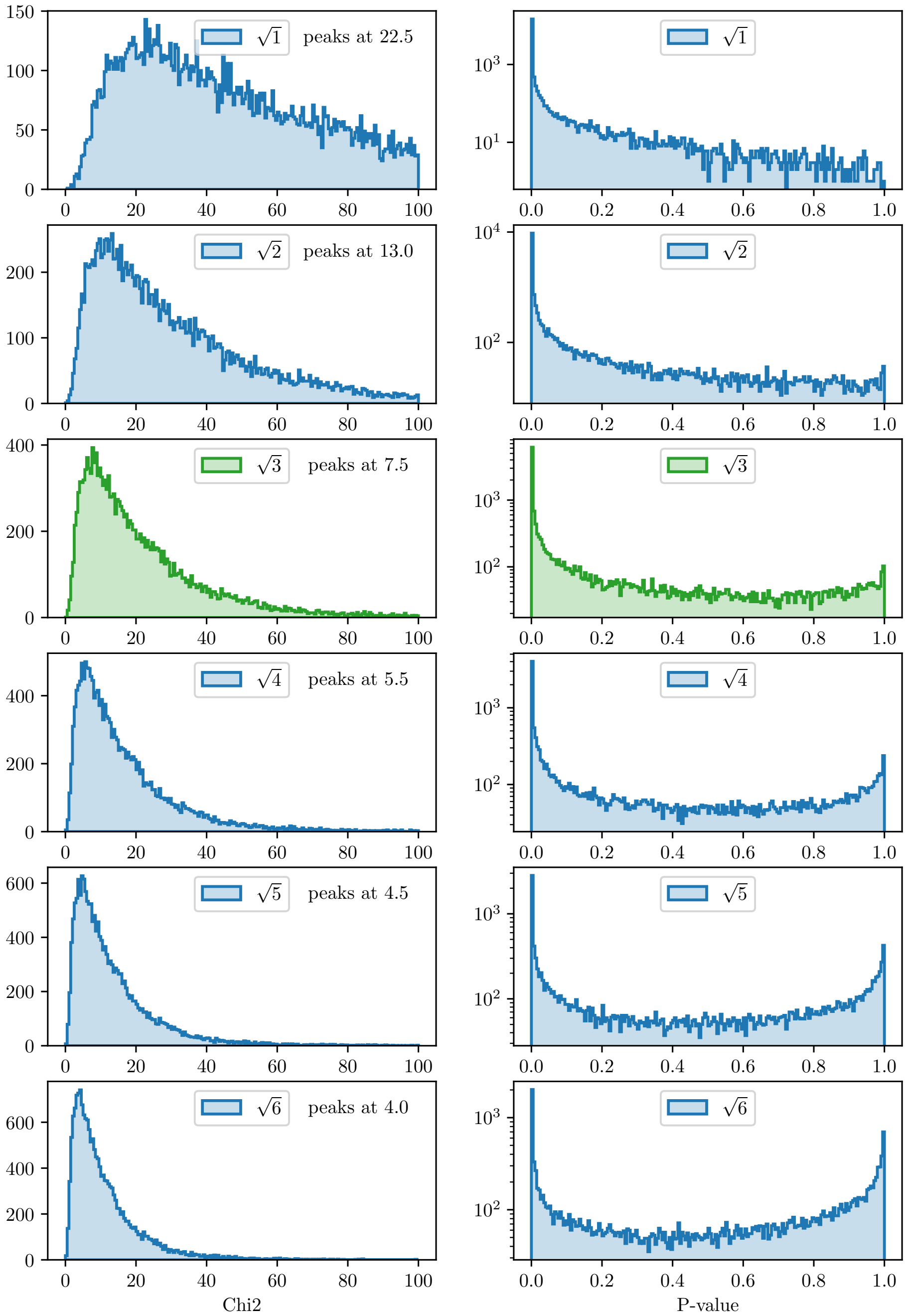

Figure 14: Chi2-values and P-values from individual LFC peak super-gauss fits with photon count (spectrum) errors multiplied by different scale-factors.

B RV extractions

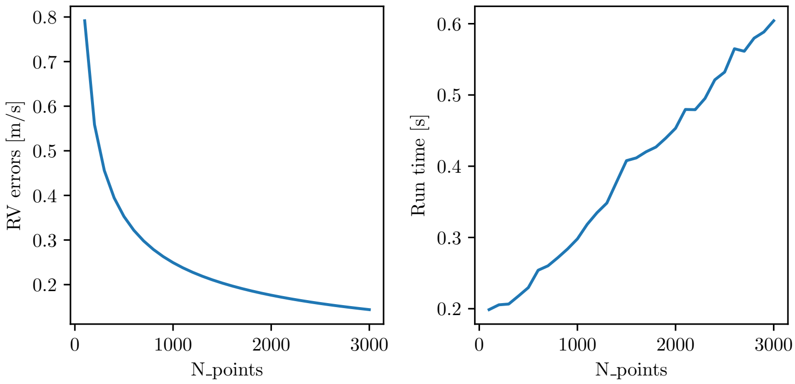

Figure 15: Number of data points in each LFC peak fit, determined by the average distance between peaks in each order.