Bachelor Project in Physics

Jakob Harteg, wmc573

June 2022, University of Copenhagen

1 Introduction

To discover Earth analogs around other stars, next generation spectrographs must measure radial velocity (RV) with 10 cm/s precision.

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 [3]. With new extreme-precision spectrographs such as EXPRES, we are slowly approaching the precision necessary for the discovery of Earth-sized planets around Sun-like stars. To achieve such precision however, it is necessary to understand and mitigate many effects. This includes the movement of the Earth relative to the center of mass of the solar system (bary-centric corrections), light scattering in the Earth's atmosphere (tellurics), light scattering inside the spectrograph (blaze). Furthermore a general wavelength calibration of the spectrograph data is needed, the quality of which of course directly influences the precision of the final radial velocities that can be obtained. To perform this calibration on the EXPRES spectrograph a laser frequency comb (LFC) is used. The full procedure from raw data to results, also spoken of as the pipeline, is extensive and described in more detail in [6]. This project mainly revolves around performing the calibration and subsequent computation of the radial velocities, and will ignore remaining parts of the pipeline by using already corrected data. The aim of this project is method exploration and development, and not exoplanet detection.

todo: Write that the EXPRES pipeline is very long and complicated and in this paper I will describe a few basic things in the pibeline along with the few things that I have actually worked on.

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, 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, [3]). In the hunt for detecting Earth-sized planets around Sun-like stars, a new generation of spectrographs are appearing called extreme-precision radial velocity (EPRV) spectrographs. The EXtreme PREcision Spectrograph (EXPRES) is one such spectrograph, data from which this project is based on. **todo: move this to the introduction**

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

2.2 Description of the instrument

The EXtreme PREcision Spectrograph or EXPRES is situated at the Lowell Observatory's 4.3m Lowell Discovery Telescope (LDT) near Flagstaff, Arizona, U.S.A. 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 suply 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. [4]. todo: maybe add illustration

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.

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.

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 [2].

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 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)

todo: For development of a LFC calibration method data from what was used

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 mesaurements for each pixel on the CCD. The rows of the CCD data are refered 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 unuseful. 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`.

2.3.2 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 devitation is there teh square root of the 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.

2.3.3 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 modelled and subtracted by measureing the photon count in between orders.

2.3.4 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].

2.3.5 Correction: Bary centric corrections

todo: not sure what to write here, if anything

3 Data Analysis

3.1 Calibration

todo: add error analysis for the calib data

The calibration is needed to map each pixel on the CCD to a specific wavelength. Such a map is refered to as a wavelength solution. To do this, we need a light source with known frequencies, preferably many discret peaks. EXPRES uses a Thorium Argon lamp for an initial trial wavelength solution and 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. (I this project I have not done this calibration).

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

$$v_n = v_{\text{rep}} \times n + v_{\text{offset}} \tag{1}$$

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.

The location of the LFC peaks (also refered to as modes) I determine first with a peak finding algorithm from scipy and then by fitting a super-gauss with a linear background to each peak:

$$A \exp\left(-\left(\frac{(x-\mu)^2}{2\sigma_x^2}\right)^P\right) + B(x-\mu) + C \tag{2}$$

A super-gauss is a regular gaussian but with an extra parameter, here denoted P, that allows the top of the peak to be flattened. To map the LFC peaks with the frequencies given by equation (1)

I can then use the intial wavelength solution (provided in the data that I have used) to map the LFC peaks to the frequencies given by equation (1).

From here, I have explored two approaches to produce a wavelength solution for all pixels: cubic interpolation and polynomial fit. I can evaluate the quality of the interpolation calibration by choosing to omit every second peak from the interpolation and then computing the residuals between the omitted peaks and the resulting interpolation function. For the polynomial, I can compute residuals simply by subtracting the location peaks from the fit function. Residuals from the two methods are compared in figure 1.
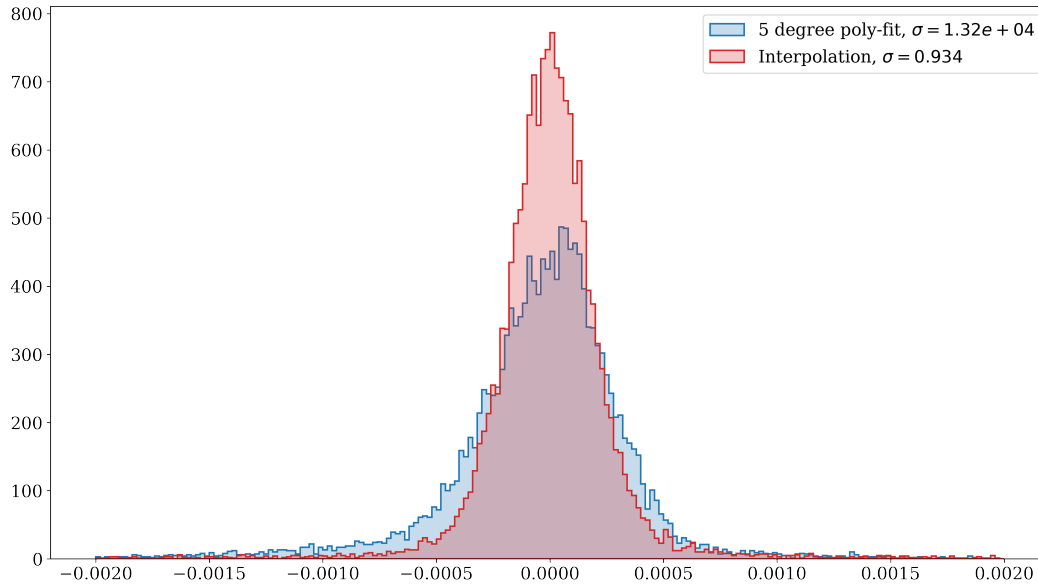
Figure 1: Residuals from calibrations performed through poly-fit and interpolation. Both results contain approximately the same amount of points, but the poly-fit has a much larger spread and therefore appears smaller. **todo: polyfit has 18373 lines, interp has 16888. Why?** **todo: find out x units**

The standard deviation of the residuals from the interpolation come out much smaller than that of the polyfit (values specified in figure 1), 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, it will be even better when performed on all data points. A similar comparison was done to determine that the polynomial fit got better with increasing degrees until 5th. **todo: specify what file?**

todo: add graph comparing residuals using gauss vs super gauss

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

todo: specify run times for calibration using poly fit and interp

3.1.1 Errors in the calibration data

todo: Rewrite this:

The chi2 values should be equal to the number of parameters in the fit. We have 5 parameters (although Troels said something about 12 or 14 .... why?). So we can see that the initial errors give a chi2 distribution that is waaay to flat. If we multiply by $\sqrt{3}$ we get something better. $\sqrt{10}$ looks perhaps even better! But, the p-value distribution should be flat (TODO: ask Troels why or look it up), so we can see that the original errors are probably too small and the ones multiplied by $\sqrt{10}$ are too high, because we get a spike at 1 — i.e. the fit is too good. The spike at 0 is okay, those are just all the fits that sucked and should perhaps be removed.

The Chi2 values are about a factor 3 too larger, thus the uncertainties are about a factor sqrt(3) too small. My evaluation is based on the notion, that the Prob-distribution should be about flat, when it is right (with peaks at the ends). We then seem to have a good fraction (10

With this correction the standard deviation of the residuals of the interpolation decreases from 2.34 to 0.934.

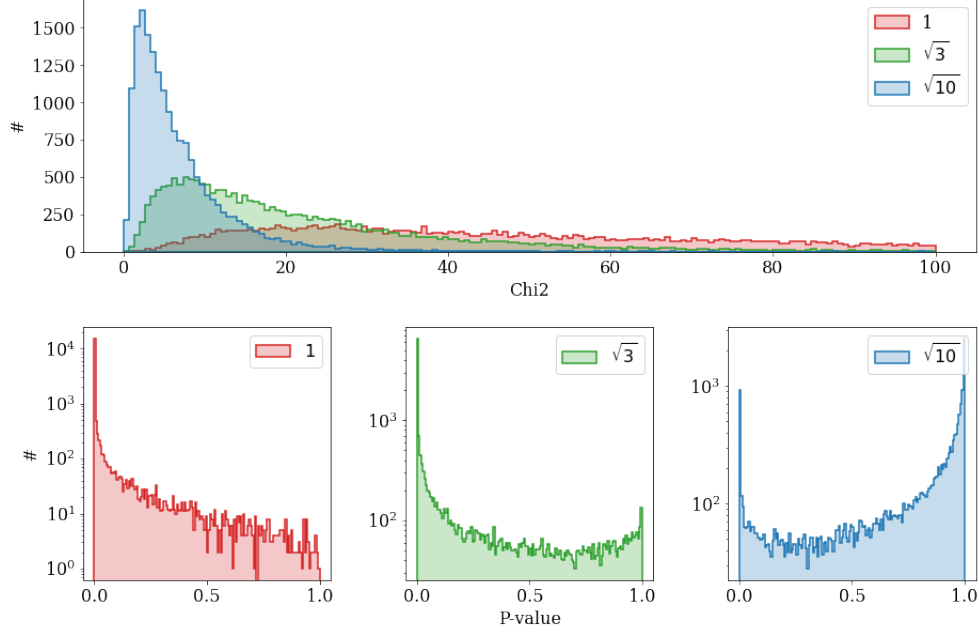**todo: Check the resulting sigma for different error scale factors**

Figure 2: 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 $\sqrt{10}$). As the distribution of P-values generally should be flat, the graphs suggest that the given errors (scale-factor 1) are too small. Multiplying errors by $\sqrt{3}$ is a conservative correction.

3.2 RV extraction

- Hist fit / cross-correlation
  - Order approach
  - Feature approach
  - (Run times)

- Extracting relative radial velocities from over constraint system (Matrix reduction to circumvent correlations)

- Future ideas
  - Auto encoder

4 (Results)

5 Discusion

6 Conclusion

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] 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.

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

[4] n. a. What is a ccd detector?, n.d.

[5] Malte Olsen. Vejer man mindre ved ækvator?, 2017.

[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.