Jabble: RV Extraction, Stellar Templates, and Normalization

NEW YORK UNIVERSITY

Matthew Daunt

Center for Cosmology and Particle Physics, Department of Physics, New York University

Abstract

In this project, we deliver a data-driven method to fitting a static stellar template, and Doppler shifts (RVs), to a multi-epoch spectroscopic campaign. We use a spline-like model for the stellar spectrum, another spline-like model for the telluric absorption, and a lower-resolution model for pseudo-continuum normalization. With the construction of new spectrographs like NEID, and substantial archives from spectrographs like HARPS and Keck HIRES, there are abundant high signal-to-noise, multi-epoch spectral campaigns of relevance. Small improvements in data analysis can leverage improvements in hardware, which are continually being advanced. Here we implement a statistical (frequentist) method, `wobble`, that delivers appropriately co-added spectral data and maximum-likelihood RVs to optimally extract RV information from the spectra Bedell2019. This method takes as input un-normalized (or poorly normalized) heteroskedastic one-dimensional spectral data, even on heterogeneous wavelength grids, and outputs **precise relative RVs** at all epochs, a **high signal-to-noise tellurics-free** mean spectrum, and **structured residuals in the wavelength domain.**

Cardinal Spline Mixture Model

We chose to call these cardinal basis spline models, cardinal spline mixture models (CSM). CSM models are evenly spaced local basis functions built with piece-wise $p+1$ polynomials of degree p. We chose this model because these models guarantee continuity to the $p-1$ derivative [3, 2].

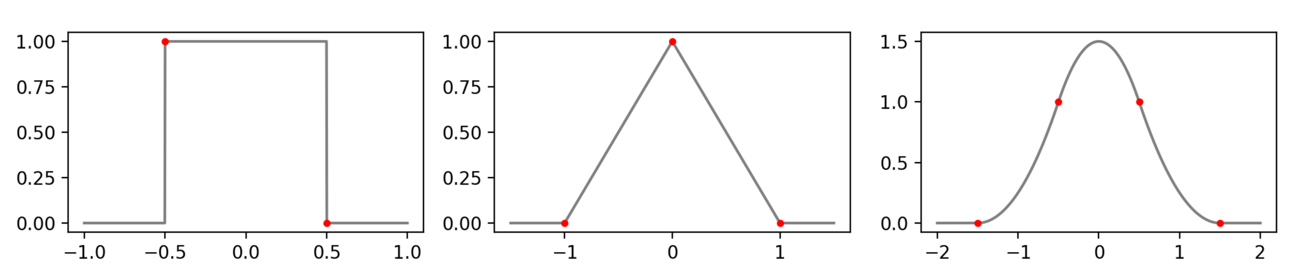


Figure 1. Kernel Functions for $p = 0, 1, 2, 3$ cardinal basis splines (top to bottom), where the kernels are located at each integer.

Understanding the Data

Transform Data to Log-Flux, Log-Wavelength space.

$$y_{ij} = \log f_{ij},\ x_{ij} = \log \lambda_{ij},\ \delta x_i = \log \sqrt{\frac{1+v_i/c}{1-v_i/c}},\ \sigma_{yij} = \frac{\sigma_{fij}}{f_{ij}} \tag{1}$$

The tellurics, stellar, and pseudo-normalization component models will all be Cardinal Spline Mixture Models (CSM) that sum together to get a prediction,

$$\hat{y}_{ij} = f(x_{ij} + \delta x_i|\theta_{\mathrm{s}}) + a_i f(x_{ij}|\theta_{\mathrm{t}}) + f(x|\theta_{\mathrm{N},i})\ . \tag{2}$$

where i is the index for epoch, and j is the index for wavelength. The differences are as follows:

- **Stellar** allowed to shift, δx_i at each epoch, i
- **Telluric** stretches with airmass, a_i
- **Normalization** low order in log-wavelength space, varies in each epoch, i

The prediction from this model is then iteratively fit to the data using L-BFGS-B method.

$$\arg\min_{\theta} \sum_{ij} \frac{(\hat{y}_{ij}(\theta) - y_{ij})^2}{\sigma_{ij}^2} = \hat{\theta}\ , \tag{3}$$

where θ is all unknown parameters from equation 2, including, θ_s, θ_t, $\theta_{N,i}$, and δx_i.

Results

Here we demonstrate the normalization model on some data pulled from the HARPS online archive from HD4307. HD4307 is a G-type star with no known companion, and is visible the whole year from the La Silla Observatory.

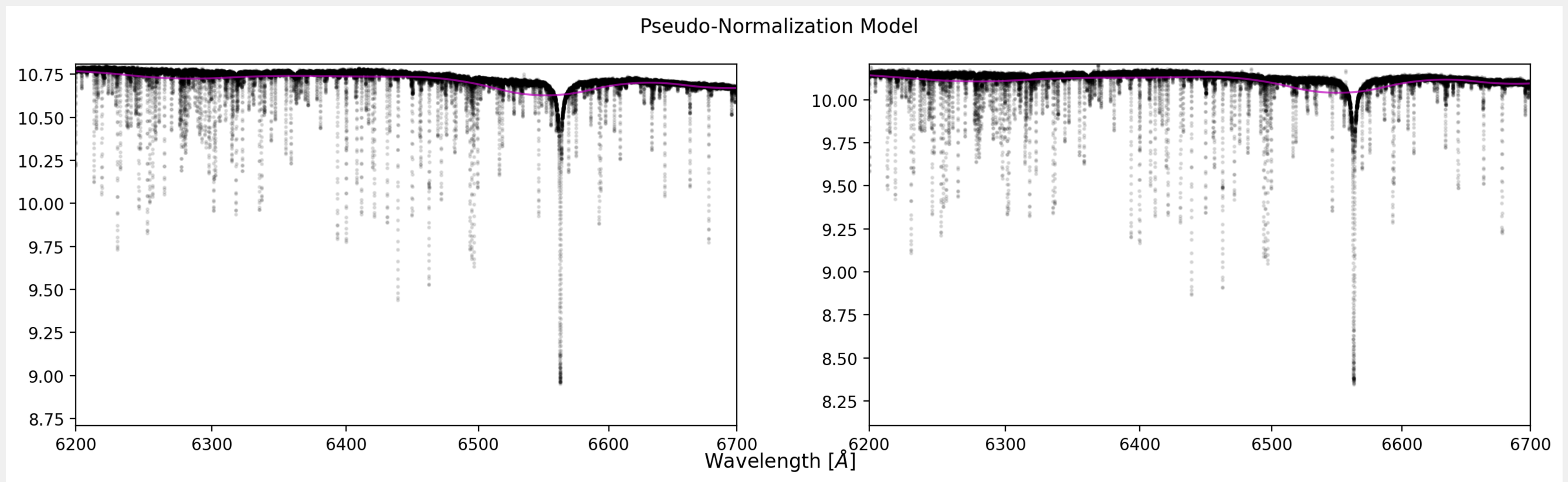


Figure 2. In magenta, the widely spaced in log-wavelength normalization model with unnormalized data in black.

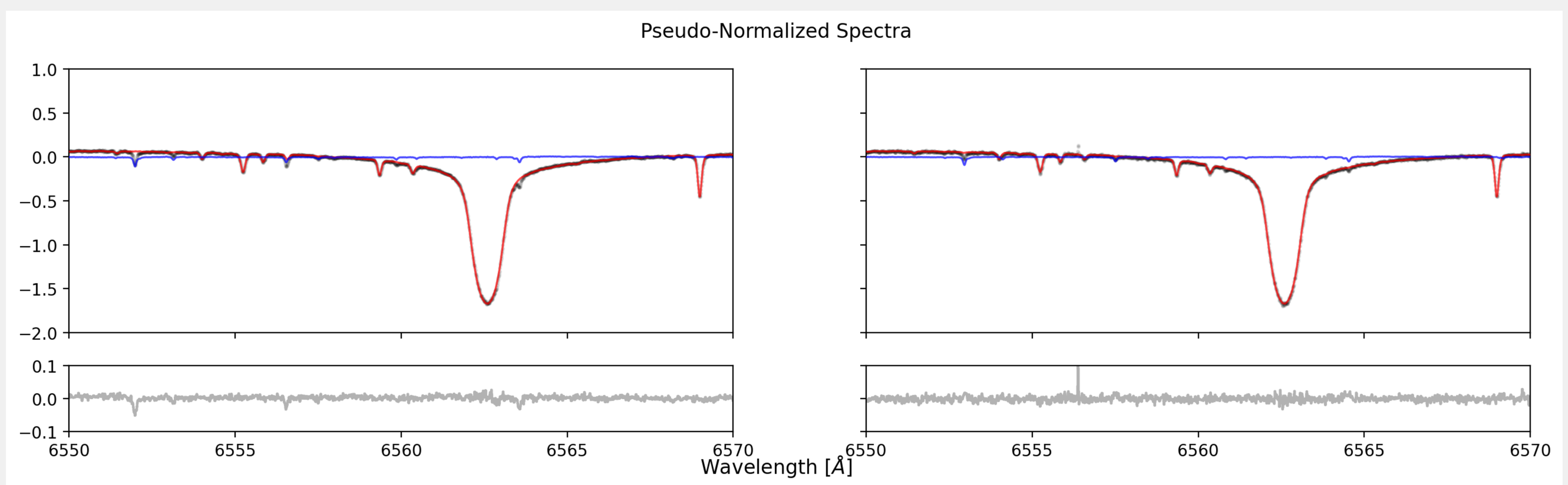
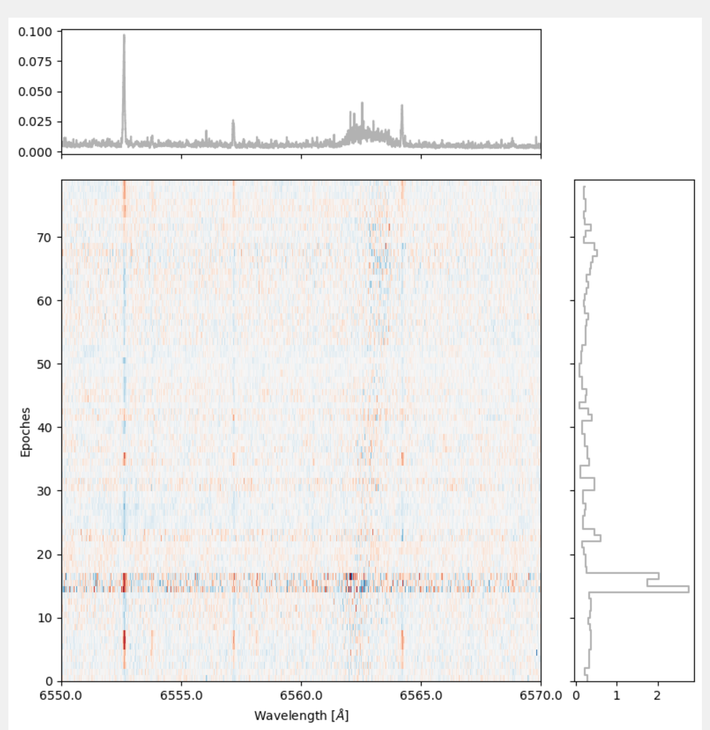
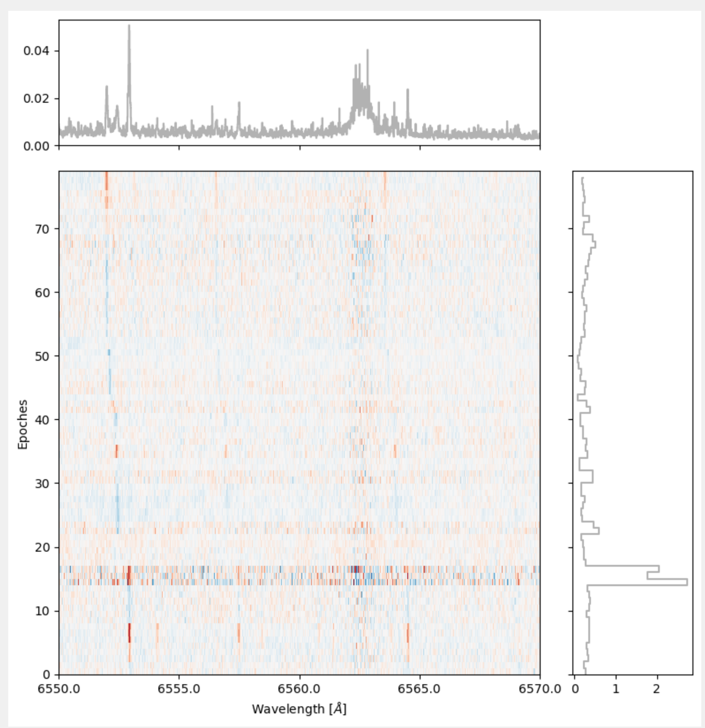


Figure 3. In black, the data minus the normalization model. In blue, the tellurics model, and red, the stellar model. Note that the data has been transformed into the barycentric rest frame as part of the HARPS pipeline.



(a) The sharp spike on the left on the upper plot appears to indicate a damaged pixel in the spectrograph.



(b) The gaussian about the H-alpha line tightens when transformed into the star reference frame indicating star variability.

Figure 4. Above are two figures of residuals of HD4307, figure 4a in the reference frame of earth, and figure 4b in the reference frame of the star. The side plots are the sum of squares of residuals along those columns or rows.

Results Cont'd

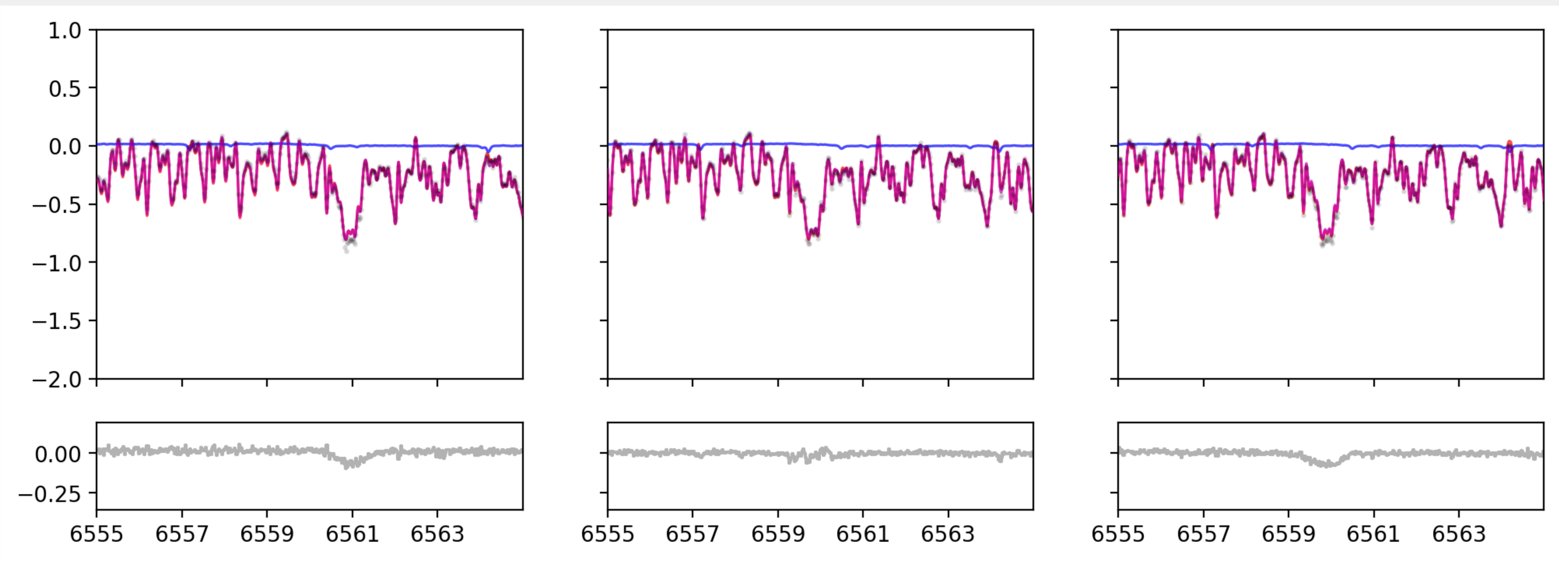


Figure 5. Above is the extracted static stellar template from Barnard's Star, an M dwarf

Discussion

Drawbacks

- **Stellar Variability** The largest source of error to center of mass radial velocity prediction has yet to be tackled
- **Relative Velocity** Without information about the rest frame wavelength positions of spectral features, this method only predicts the change in velocity of the star with respect to itself, and cannot find drift velocities.
- **Full Coverage** In order to decouple the stationary telluric model from the moving stellar model, there needs to be large changes in the stars velocity. Thus, it must be observed at many points throughout the year.

Future Work

- **Co-adding APOGEE** model APOGEE subframes to produce RV measurements at higher temporal cadence.
- **Wavelength Calibration** model EXPRES spectra with iodine cell, to solve for wavelength solution within `jabble`
- **Multi-Star Model** Fit for spectra across multiple stars simultaneously

References

[1] Megan Bedell, David W. Hogg, Daniel Foreman-Mackey, Benjamin T. Montet, and Rodrigo Luger. ttWOBBLE/tt: A data-driven analysis technique for time-series stellar spectra. *The Astronomical Journal*, 158(4):164, September 2019.

[2] Philip Hall. The distribution of means for samples of size n drawn from a population in which the variate takes values between 0 and 1, all such values being equally probable. *Biometrika*, 19(3/4):240, December 1927.

[3] J. O. IRWIN. ON THE FREQUENCY DISTRIBUTION OF THE MEANS OF SAMPLES FROM a POPULATION HAVING ANY LAW OF FREQUENCY WITH FINITE MOMENTS, WITH SPECIAL REFERENCE TO PEARSON's TYPE II. *Biometrika*, 19(3-4):225–239, 1927.