A Review of RV Pipelines

From Raw Spectra to Extreme Precisions

Arpita Roy

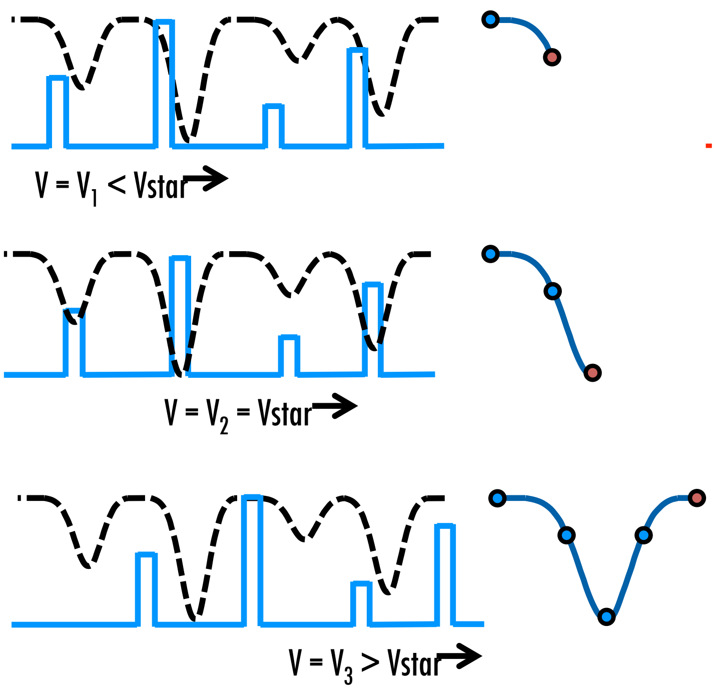
CALIFORNIA INSTITUTE OF TECHNOLOGY 1891

Radial Velocity Measurement Technique

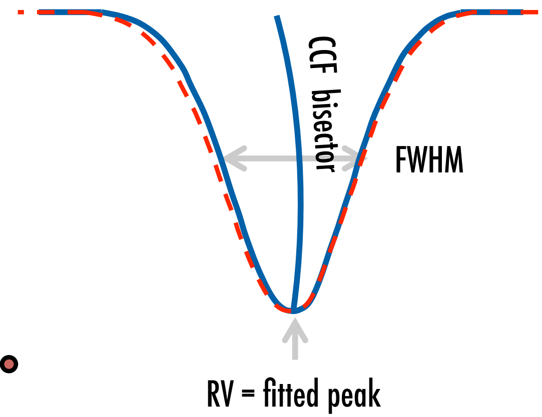
Cross Correlation with Mask or Template

True bulk motion causes all stellar lines to move in an identical way [without changing shape]

Cross-correlation with Weighted Mask



Measuring Radial Velocity
Via Gaussian Fitting



Pick best areas, rich in RV information and clean from tellurics

Works well for FGK stars with well defined lines and continuum

Mask can to evolve over time as the data stream lengthens

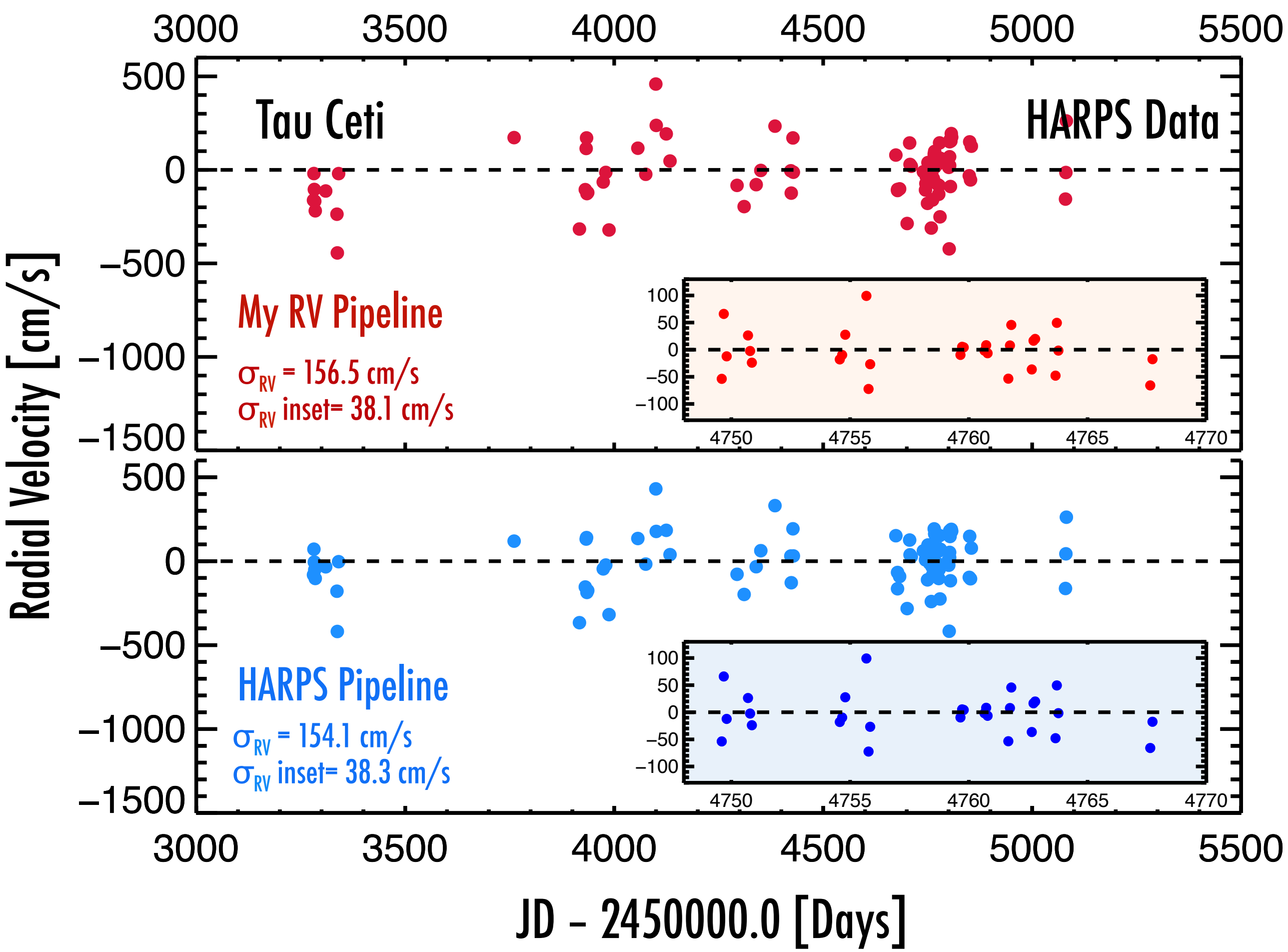
Very hard for M dwarfs: lack of clean lines + poor synthetic models

Through the scary parts of this talk remember

[Instruments +] Pipelines Do Work!

Legacy Results

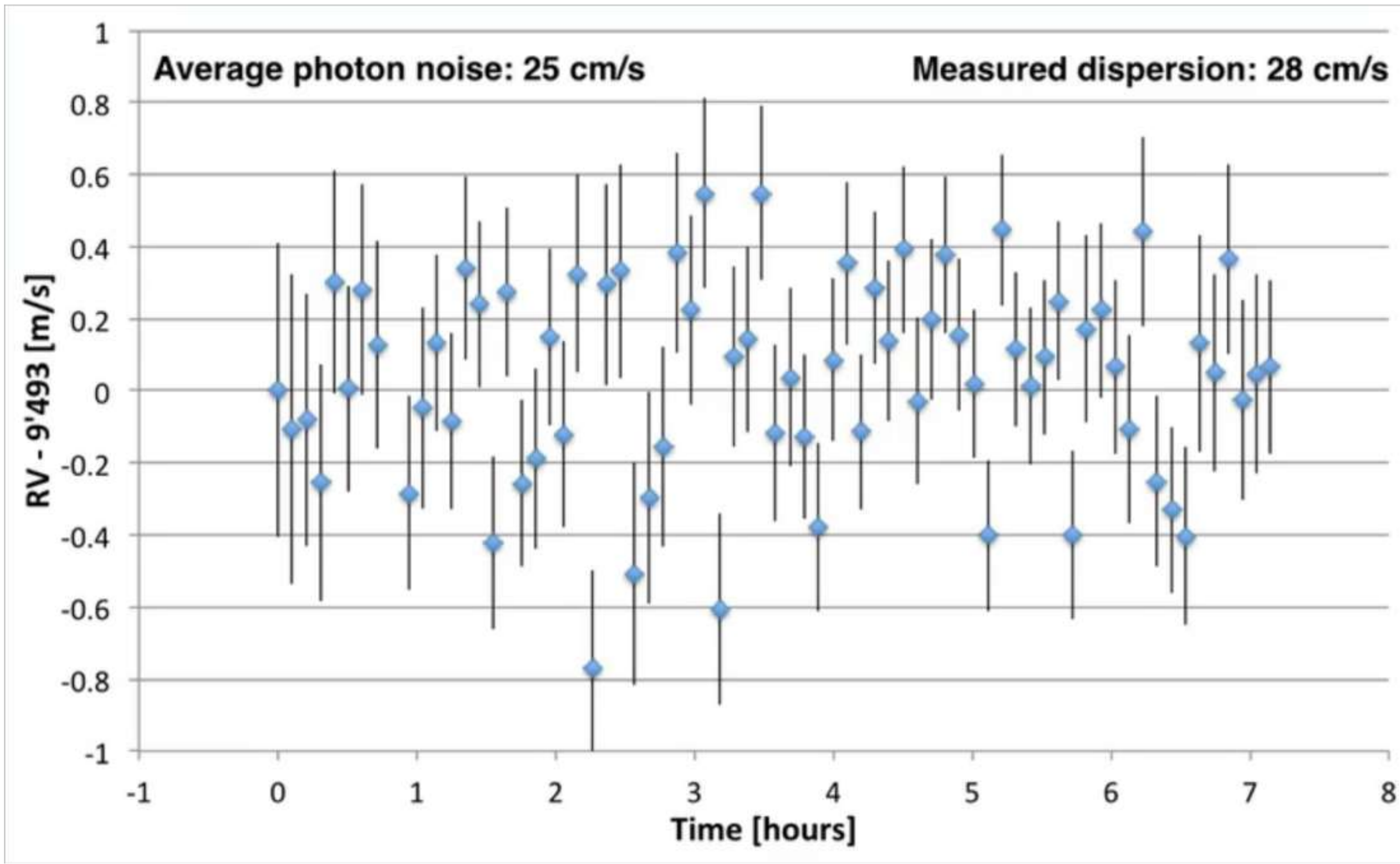
From HARPS Pipeline ≅ PARAS Pipeline



Pipelines can already go sub-m/s on bright, quiet stars

Awesome New Results

From ESPRESSO

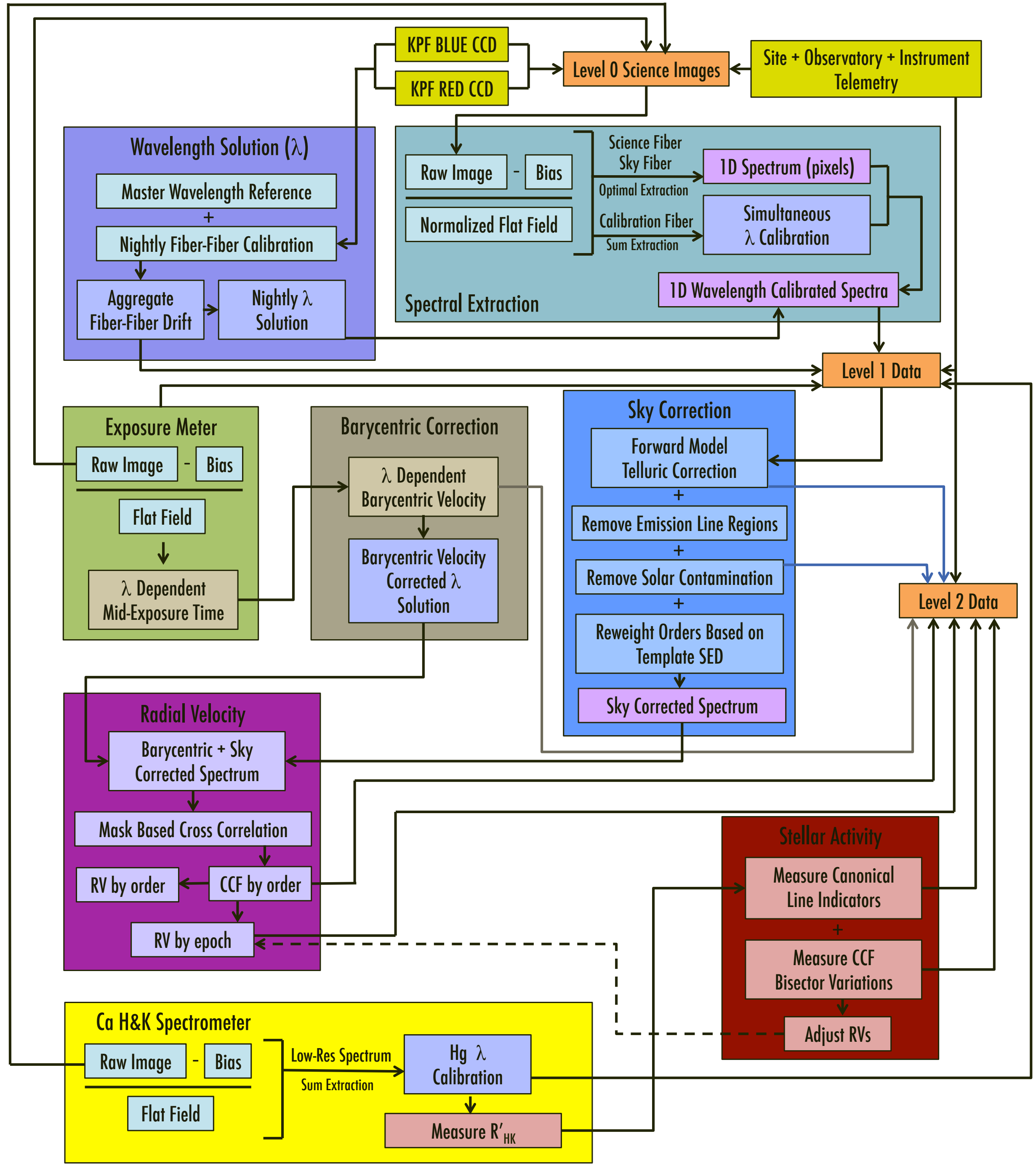


Pipelines can already go sub-m/s on bright, quiet stars

ESPRESSO early commissioning results from SPIE 2018

Data Pipeline

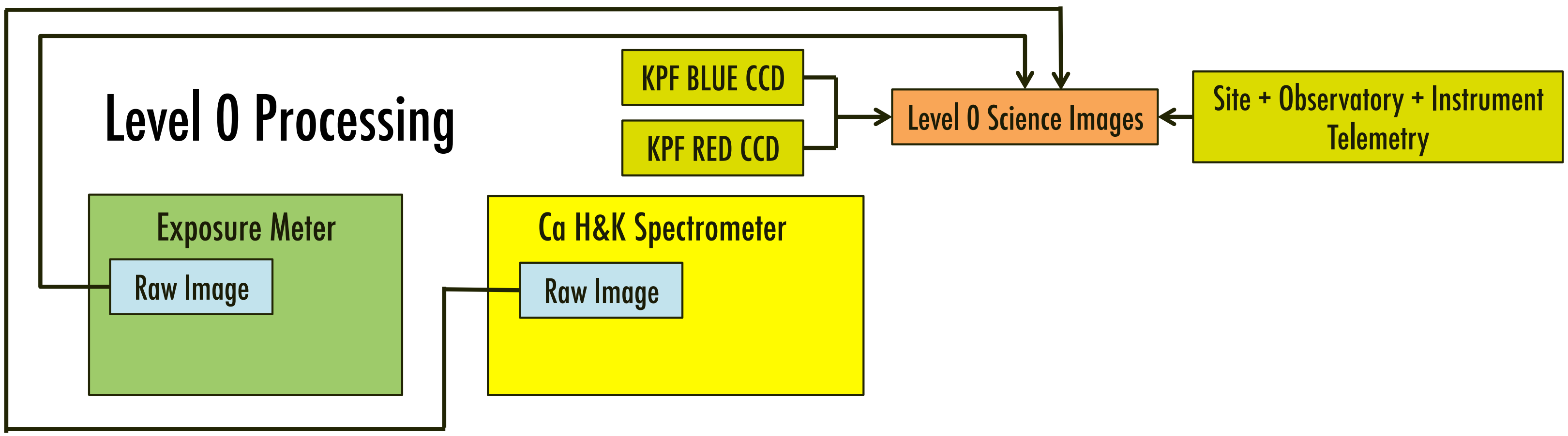
Need extreme precision analysis to produce extreme precision RVs!



Let's Get To Know +
[♡] Our Data

At the Start of Analysis

Gather All Raw Data Relevant to One Epoch*

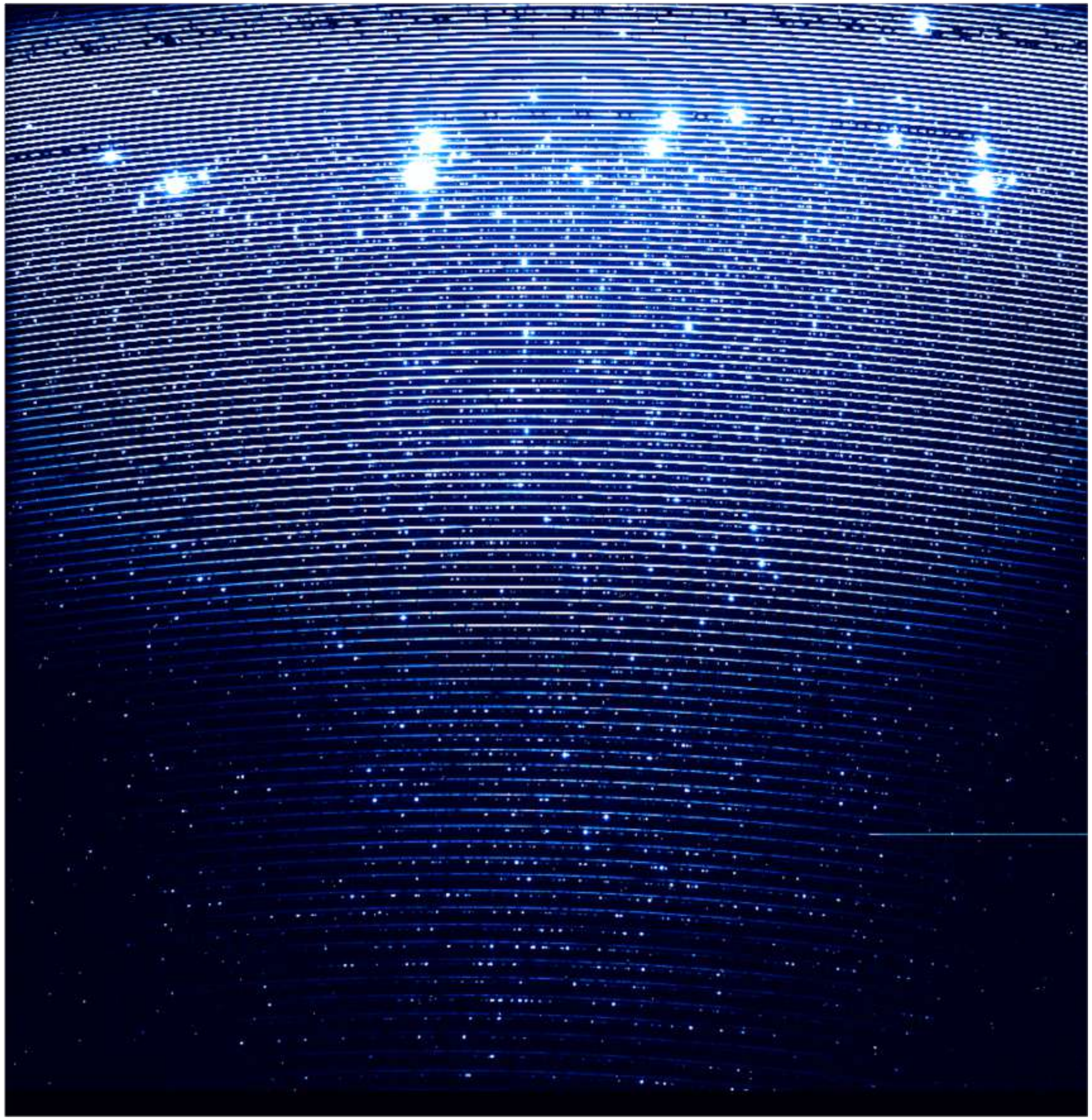


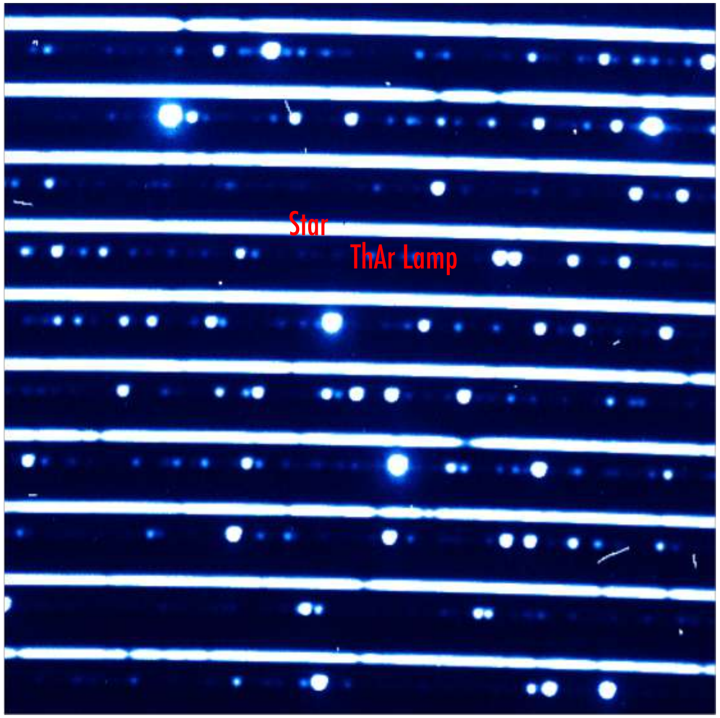
Gather all relevant raw data and telemetry necessary for running the DRP, and for look-back diagnosis of issues that arise in the field

*We will handle the complete history of the instrument outside single epoch data bundling

At the Start of Analysis

Beautiful (albeit complex) CCD images





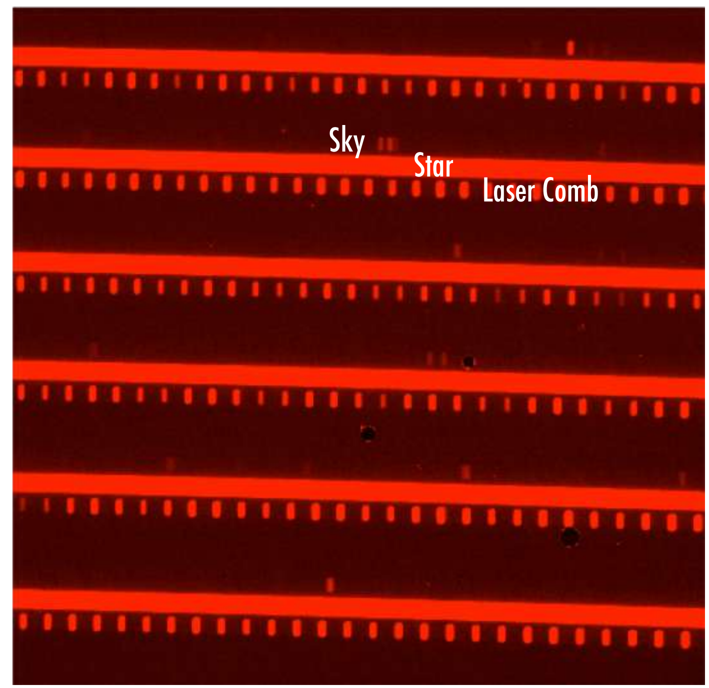
PARAS data on Tau Ceti, 4k x 4k e2v CCD, approximately 380 – 960 nm, very similar to NEID

At the Start of Analysis

Beautiful (albeit complex) CCD images

All your careful instrument design effort is already evident in these images

- Well focused spots
- Minimal scattered light
- Low aberrations
- No bright ghosts
- Minimal slit tilt
- Many more…

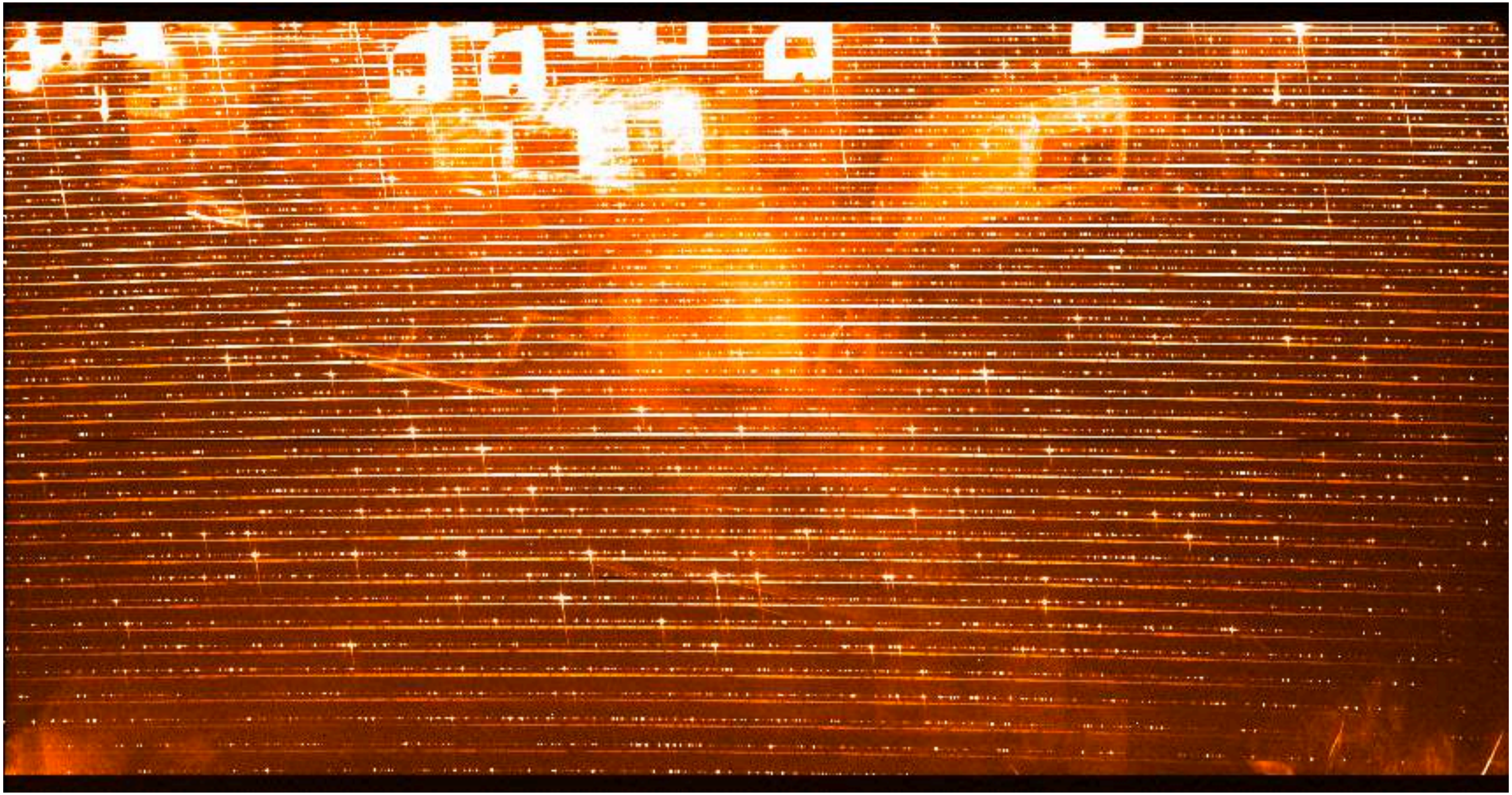


Habitable Zone Planet Finder: Raw frame of Barnard's Star with laser frequency comb calibration

At the Start of Analysis

Beautiful (albeit complex) CCD images

All your careful instrument design effort is already evident in these images – OR NOT!



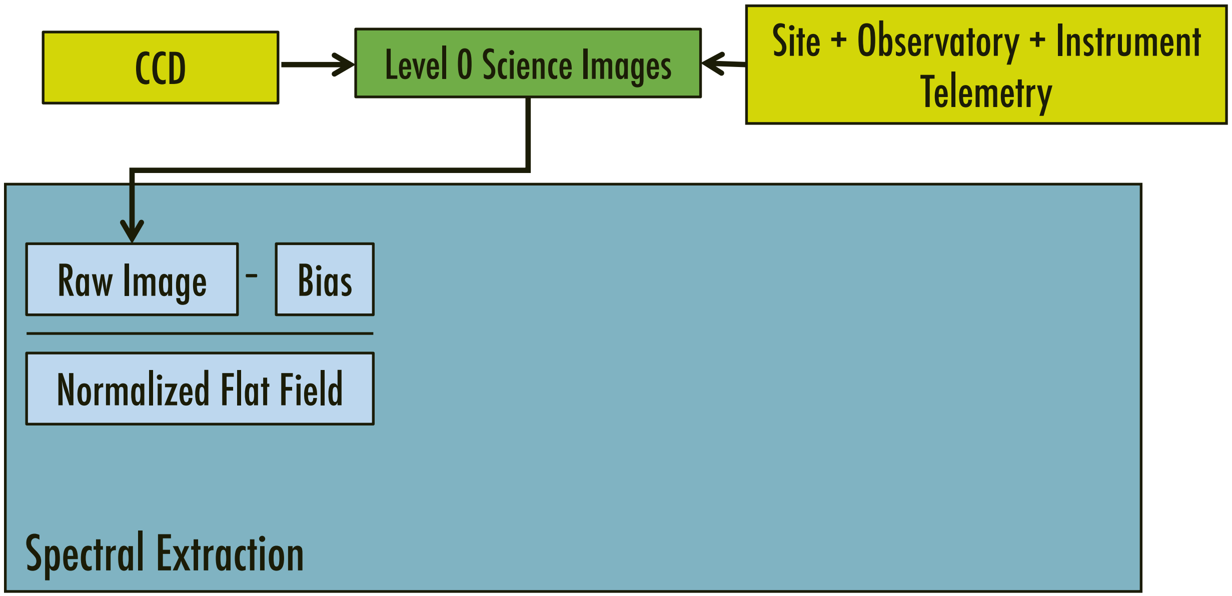
SOPHIE early data

Basic++

Image Processing

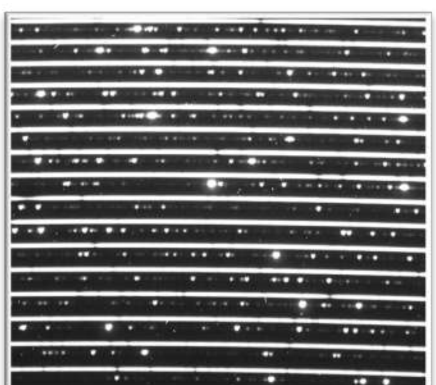
Overview of RV Data Analysis

Image Processing

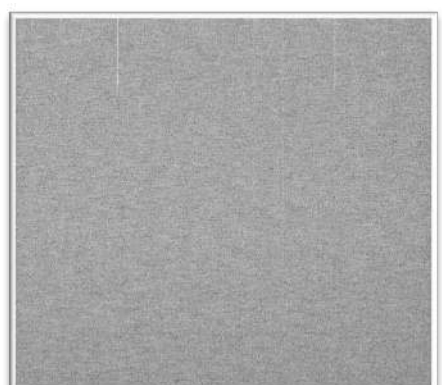


Things to be careful about:

- Bad pixels (change over time)
- Flat edges
- Pixel-to-pixel variation
- Sub-pixel shape



Raw Image



Bias

Flat

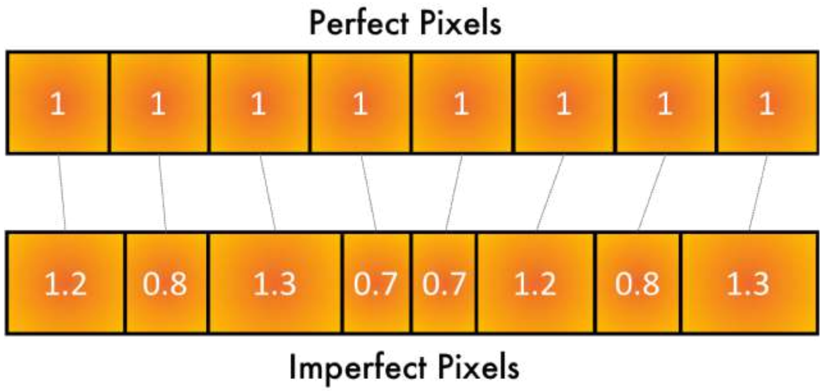
Overview of RV Data Analysis

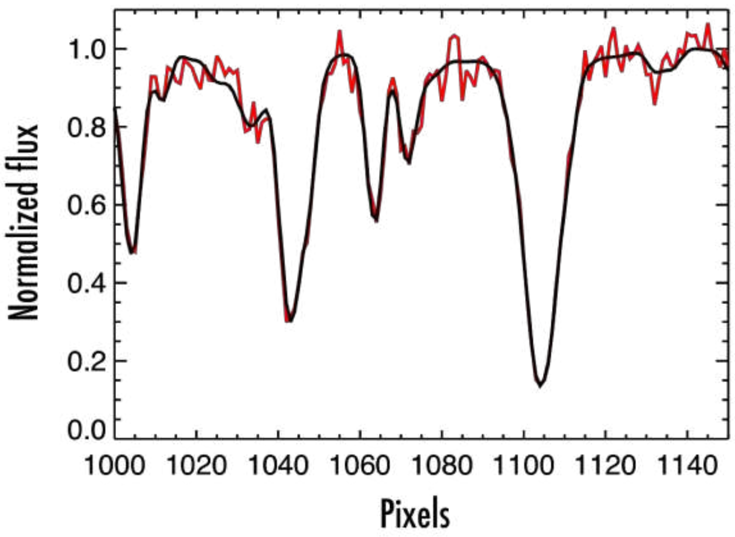
Image Processing

Aside: Detectors are not perfect!

Example A: Pixel Size Inhomogeneity

Adds noise to the spectrum and causes errors in wavelength calibration





Correction: Flux effect should be removable with flat fielding (assuming intra-pixel uniformity)
Wavelength error might be possible to correct if every sub-pixel shape mapped out in laboratory

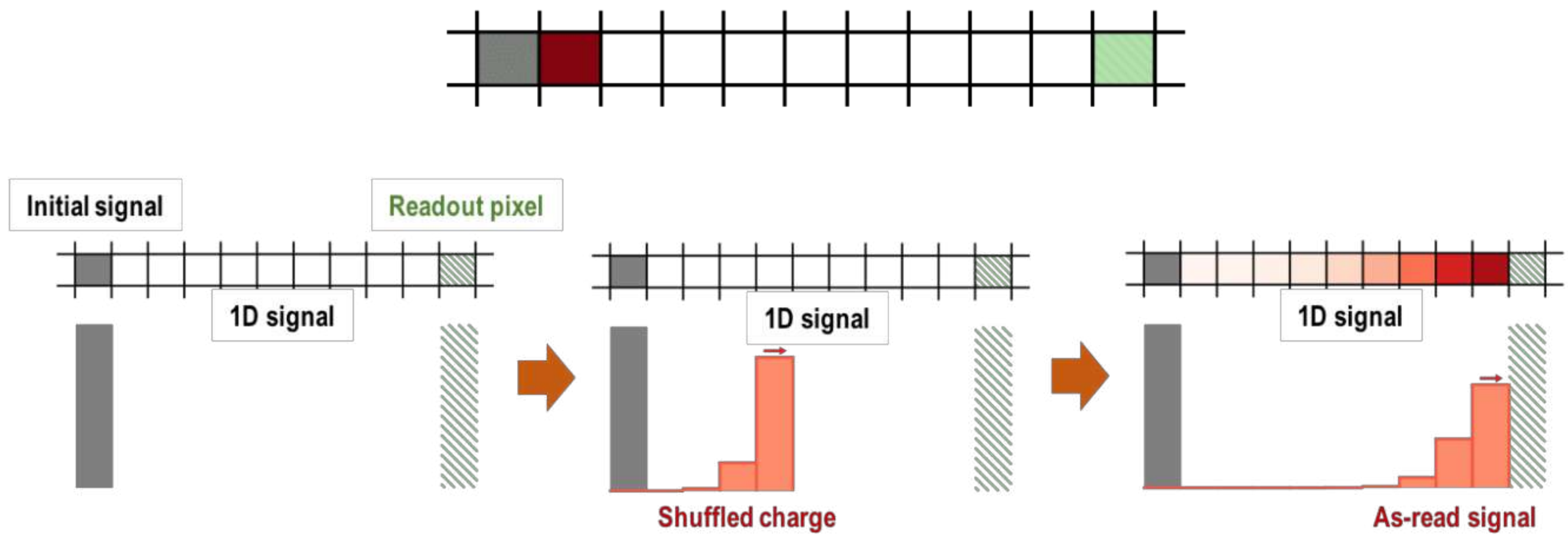
Overview of RV Data Analysis

Image Processing

Aside: Detectors are not perfect!

Example B: Charge Transfer Inefficiency

Changes continuum level and skews line shape in direction of readout



Correction: Measure CTI for your CCD, consider correcting in pipeline
Difficult because flux dependent - need library of flats at different SNR
(or demonstrate insensitivity with testing on sky like ESPRESSO)

Courtesy: Sam Halverson

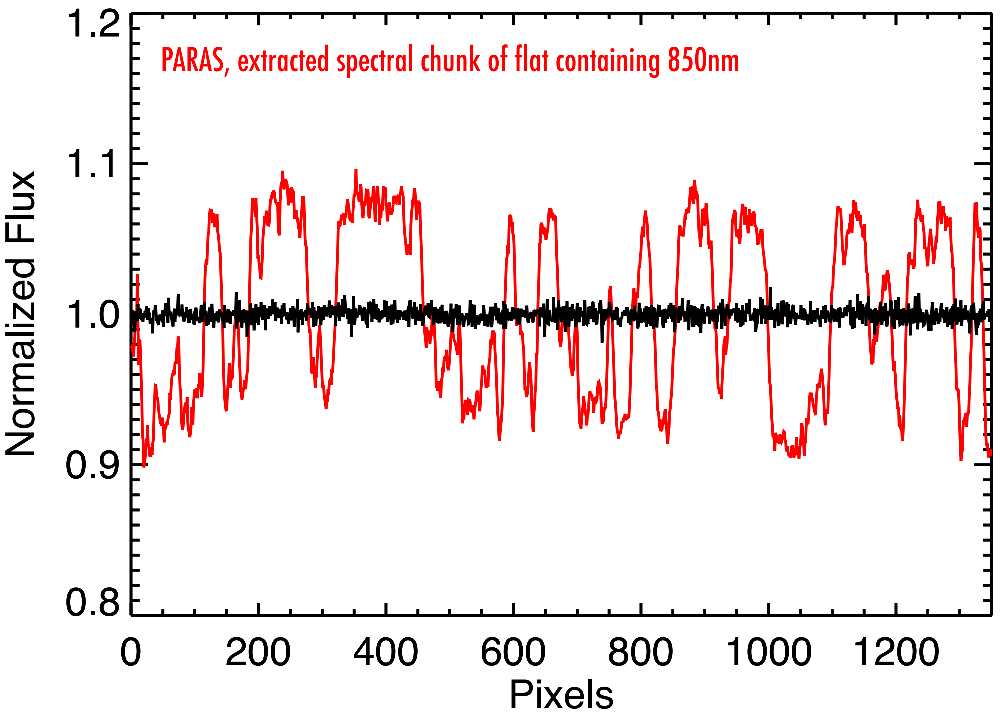
Overview of RV Data Analysis

Image Processing

Aside: Detectors are not perfect!

Example C: Fringing

Makes it difficult to use redder wavelengths on detector – important for M dwarfs on CCD instruments



Correction: Flats correct for effective QE of pixel with uniform illumination, not valid when spectrum has slopes (pixel QE is function of beam angle)

Ideally solve this in hardware with optimized CCD coatings

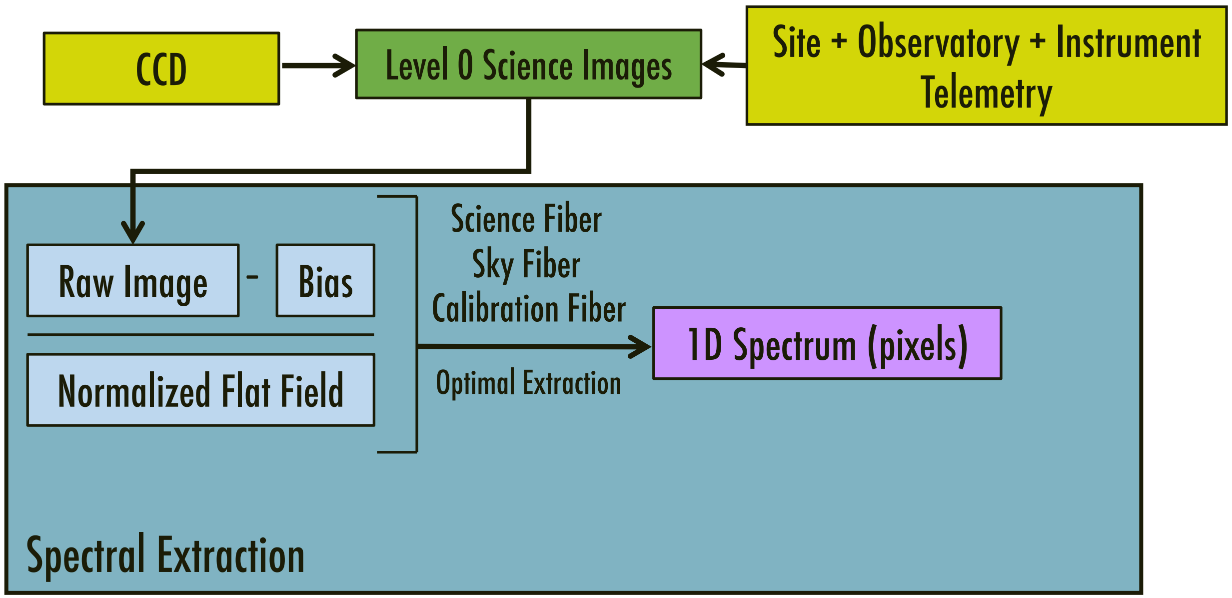
Roy et al. 2016

Basic++

Data Reduction

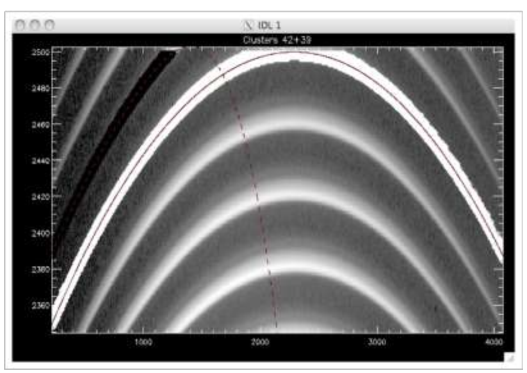
Overview of RV Data Analysis

Spectral Extraction



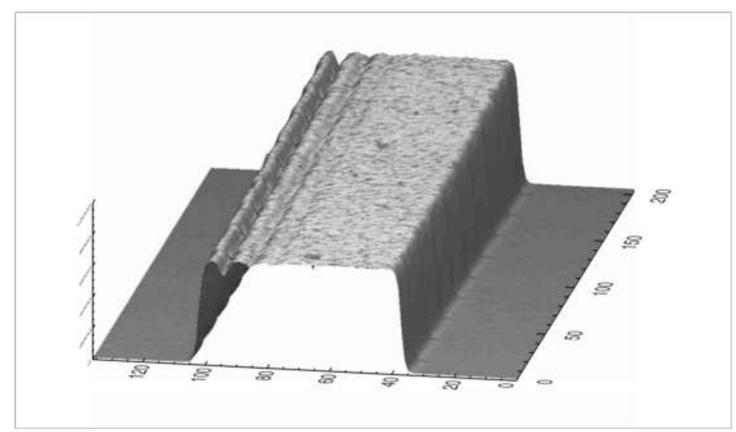
Things to be careful about:

- Curved order rectification
- Slit tilt
- Cross-talk between orders
- Scattered light/ghosts
- Stitching boundaries



Order Trace

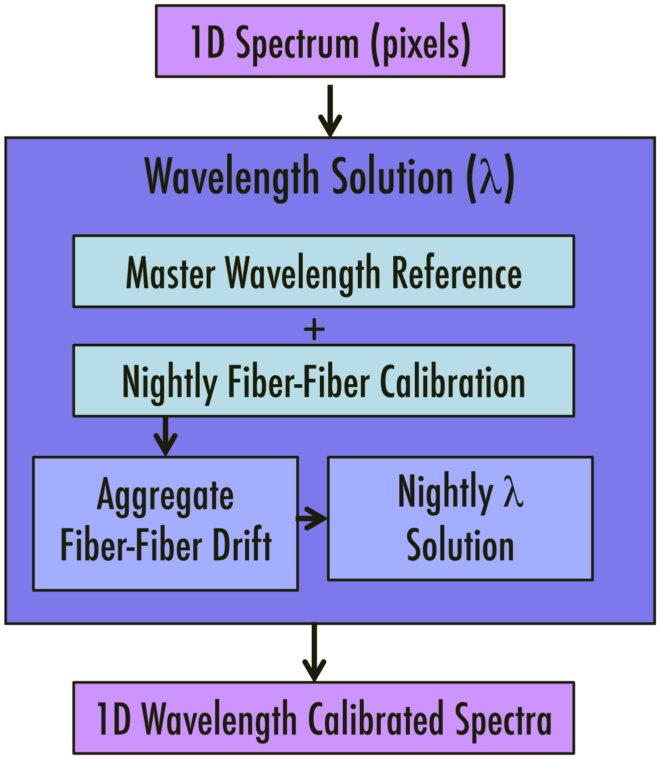
Remove cosmic rays and bleeding from contiguous orders

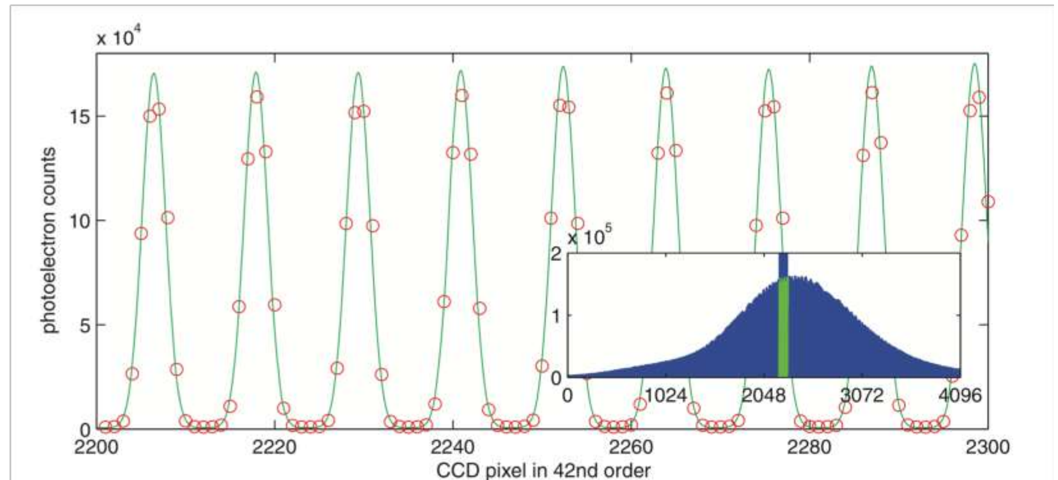


Optimally Extract Spectra

Overview of RV Data Analysis

Wavelength Calibration





Wilken+2010

Things to be careful about:

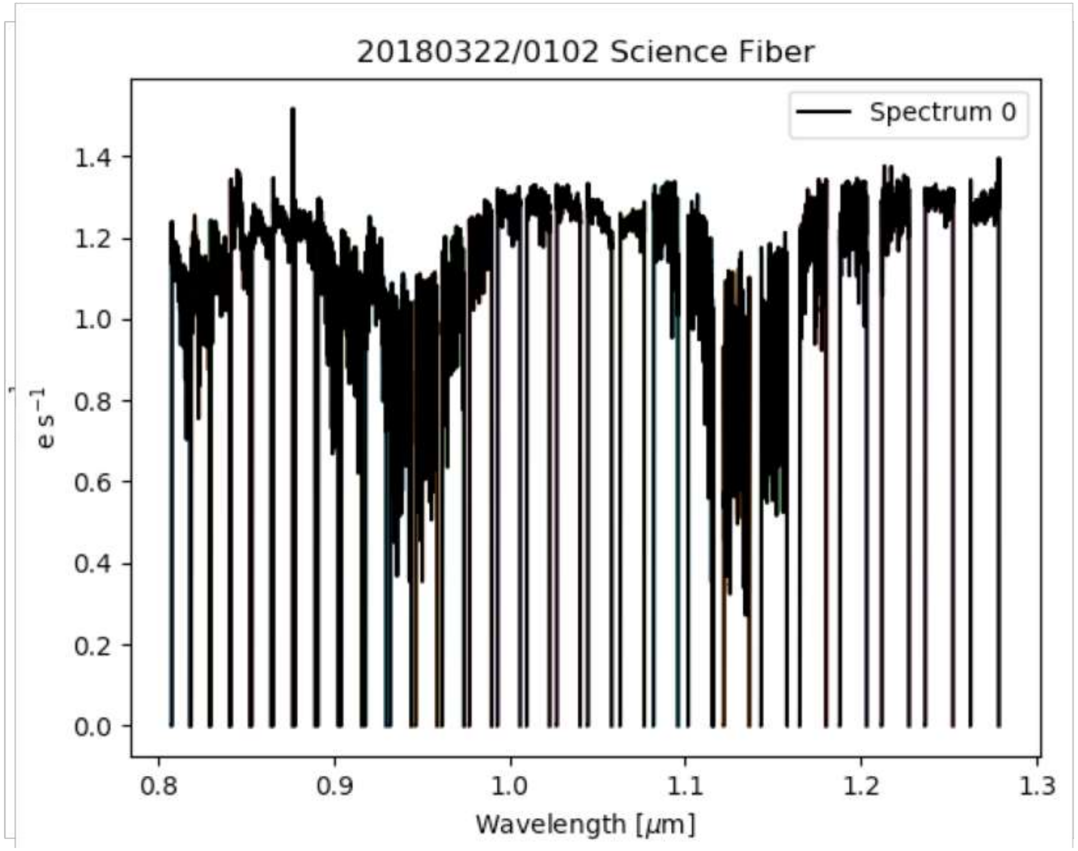
- Emission lamps have differential drift between species, aging effects, contaminants
- Fabry-Perot etalons must be evaluated for drift (including chromatic drift)
- Laser frequency combs are so nice and narrow that they can cause sampling problems

Instrument-Based

Corrections

Overview of RV Data Analysis

Flux Calibration

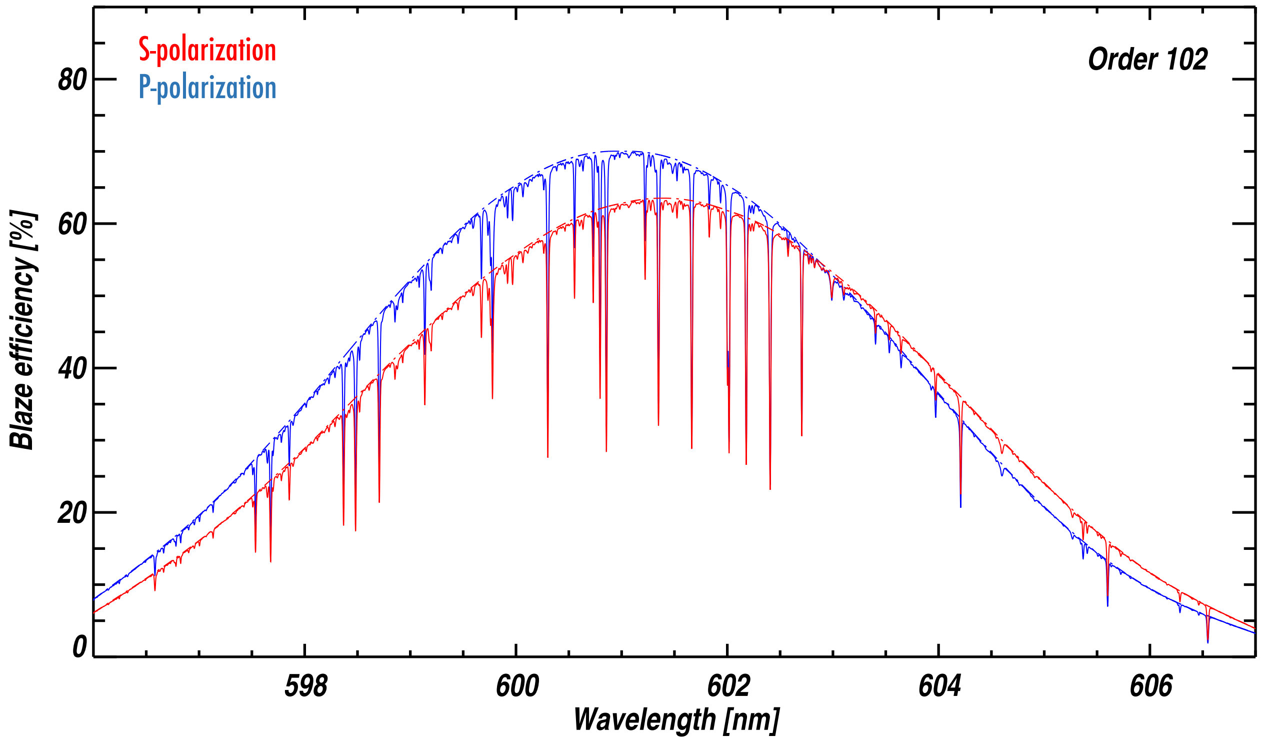


Things to be careful about:

- Overall continuum fluctuates a lot based on clouds/seeing/guiding
    - Scrambling cannot restore light outside fiber
    - Light loss is chromatic (based on level of atmospheric dispersion correction)

Overview of RV Data Analysis

Flux Calibration



Things to be careful about:

- Blaze curve within an order varies depending on echelle illumination (variable with guiding, seeing etc.)
    - Reweights lines and can skew line shapes
    - Want to remove variability but not inherent photon noise information

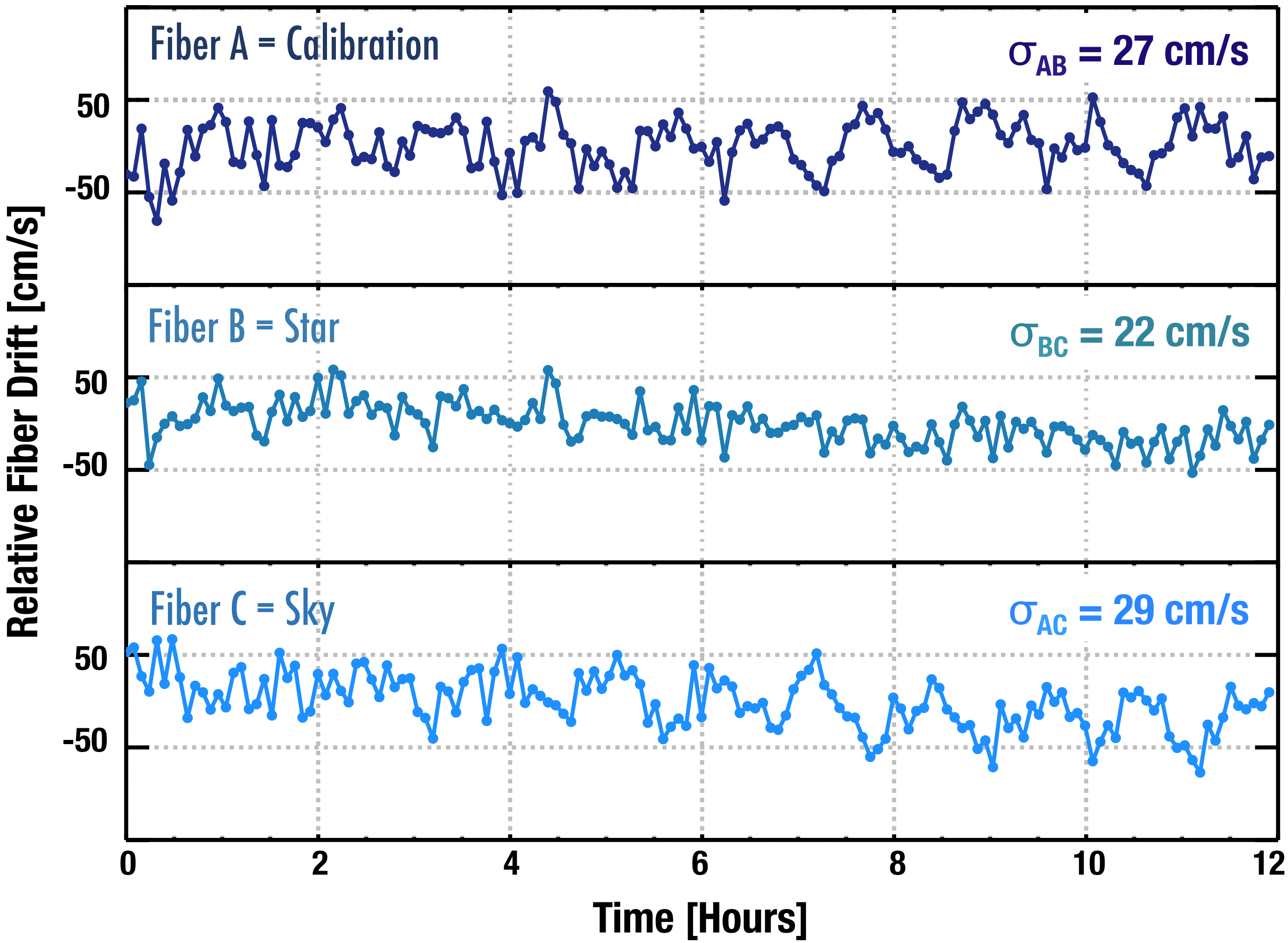
*Illustration of extreme case for SMF spectrographs

Halverson, Roy et al. 2015

Overview of RV Data Analysis

Instrument Drift Correction

Fiber-fiber drift is our main diagnostic of instrument stability



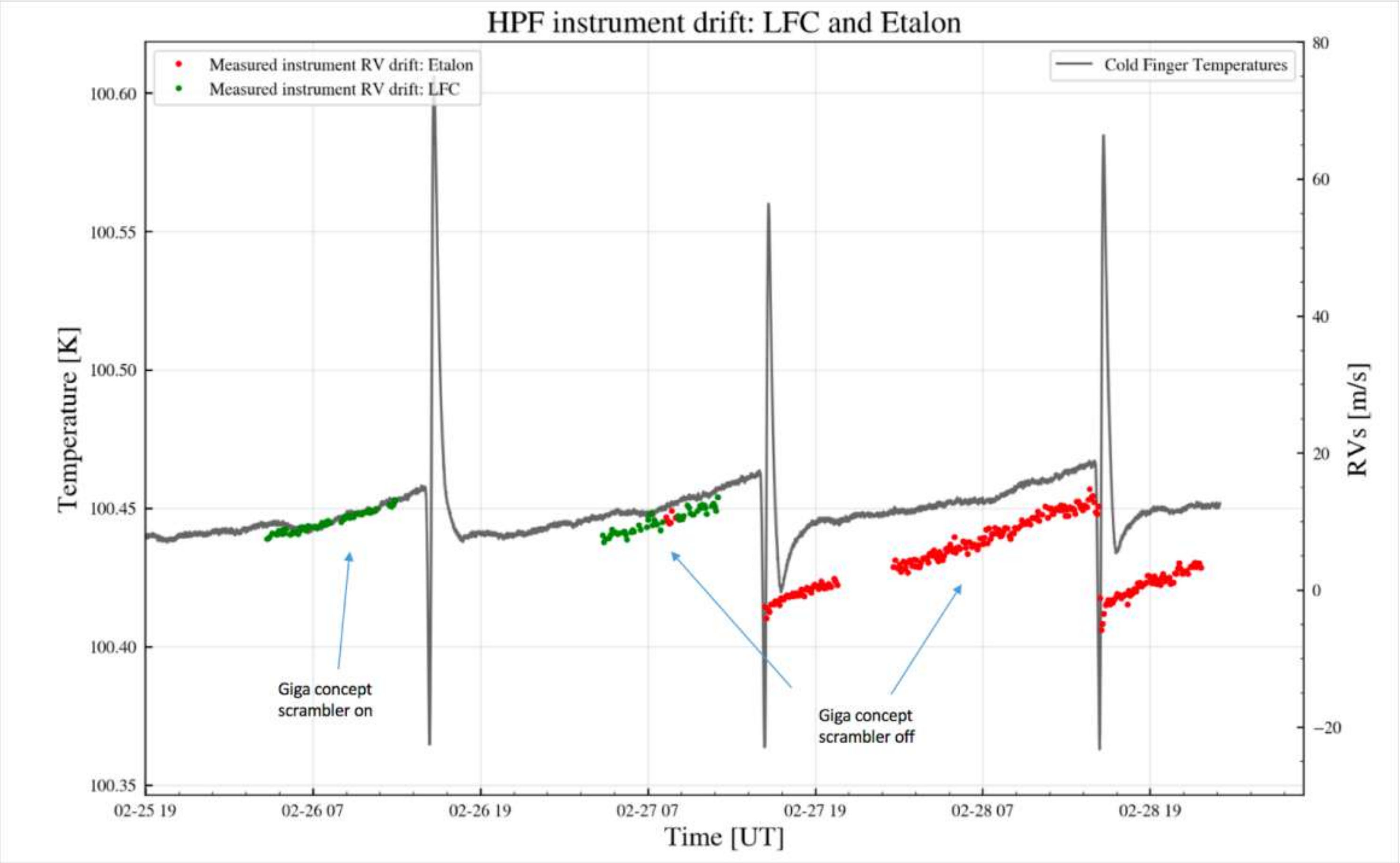
Habitable Zone Planet Finder Early Commissioning Data

Overview of RV Data Analysis

Instrument Drift Correction

Keep an eye on absolute fiber drift for systematic changes

Especially daily, seasonal, or yearly cycles



Habitable Zone Planet Finder Early Commissioning Data

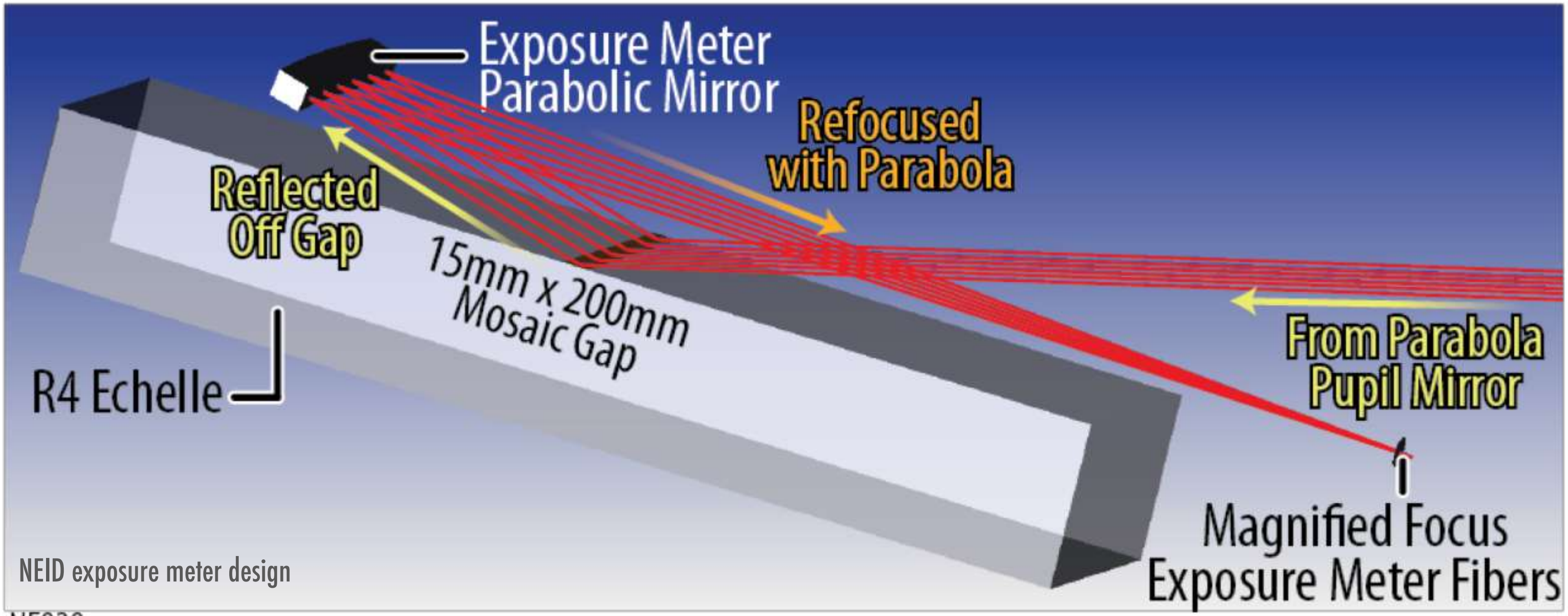
Courtesy: Joe Ninan

Earth-Based

Corrections

Introduction to RV Data Analysis

Barycentric Correction



NE030

Need to calculate barycentric motion at flux-weighted mid-point of observation

[Cannot assume exposure was uniformly illuminated – variables include guiding, seeing, airmass, clouds etc.]

Need chromatic barycentric correction

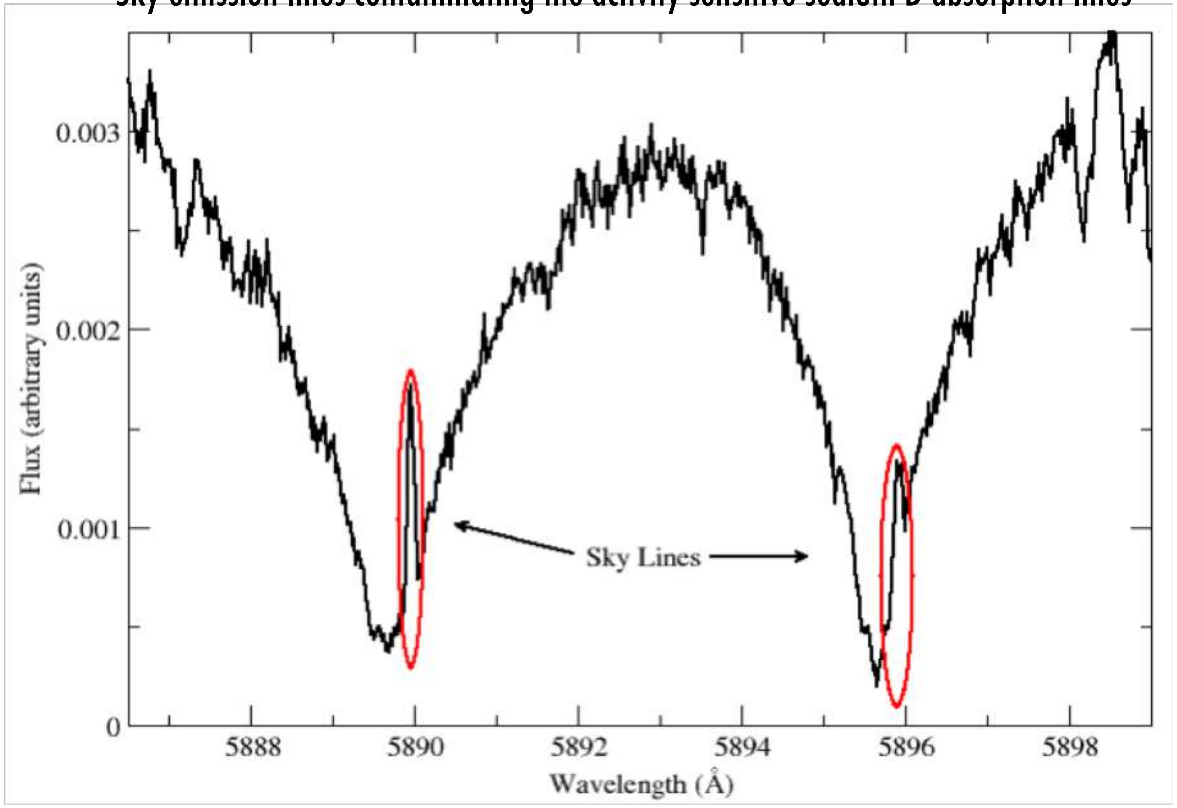
[Different color photons arrive at different rates]

Introduction to RV Data Analysis

Sky Correction

a. Telluric Emission

Sky emission lines contaminating the activity-sensitive sodium D absorption lines



Solution: Subtract sky fiber spectrum (might need to be scaled for slightly different fiber)

Even simple subtraction can make a big improvement

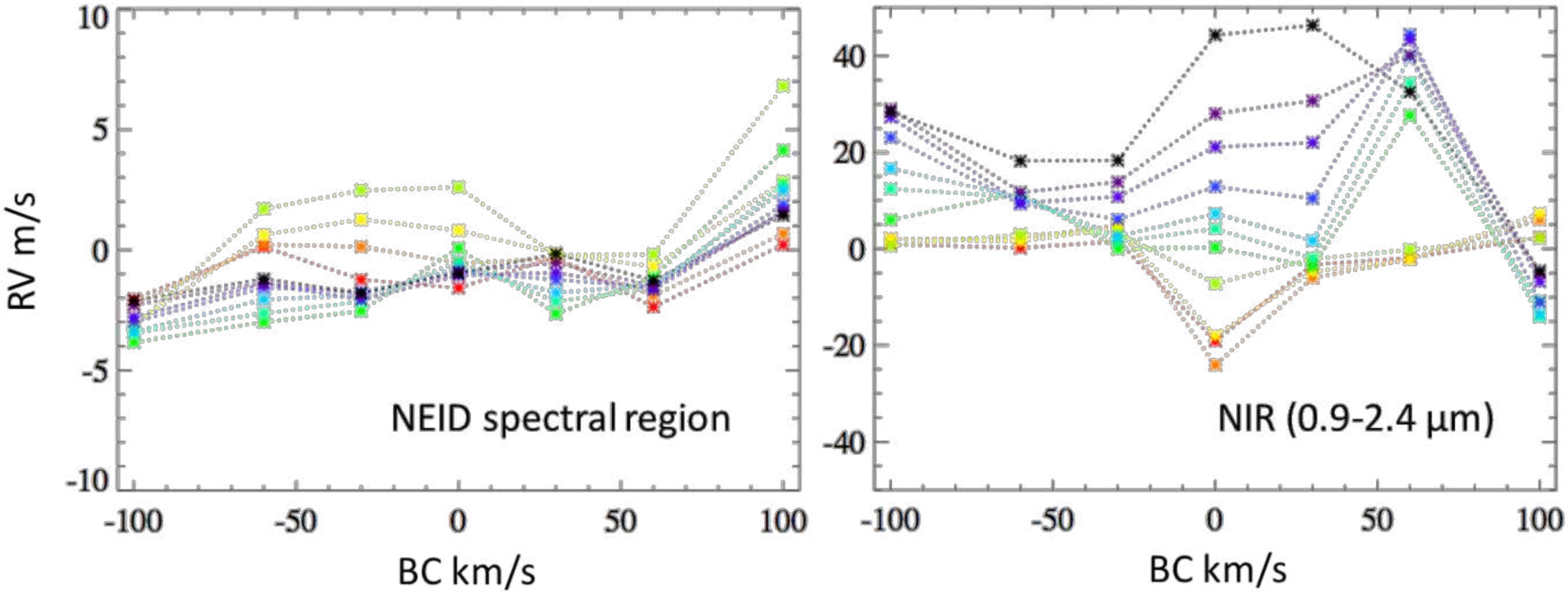
Courtesy: Paul Robertson

Introduction to RV Data Analysis

Sky Correction

b. Telluric Absorption

Left untreated, telluric lines will have very large impacts on RV precision.

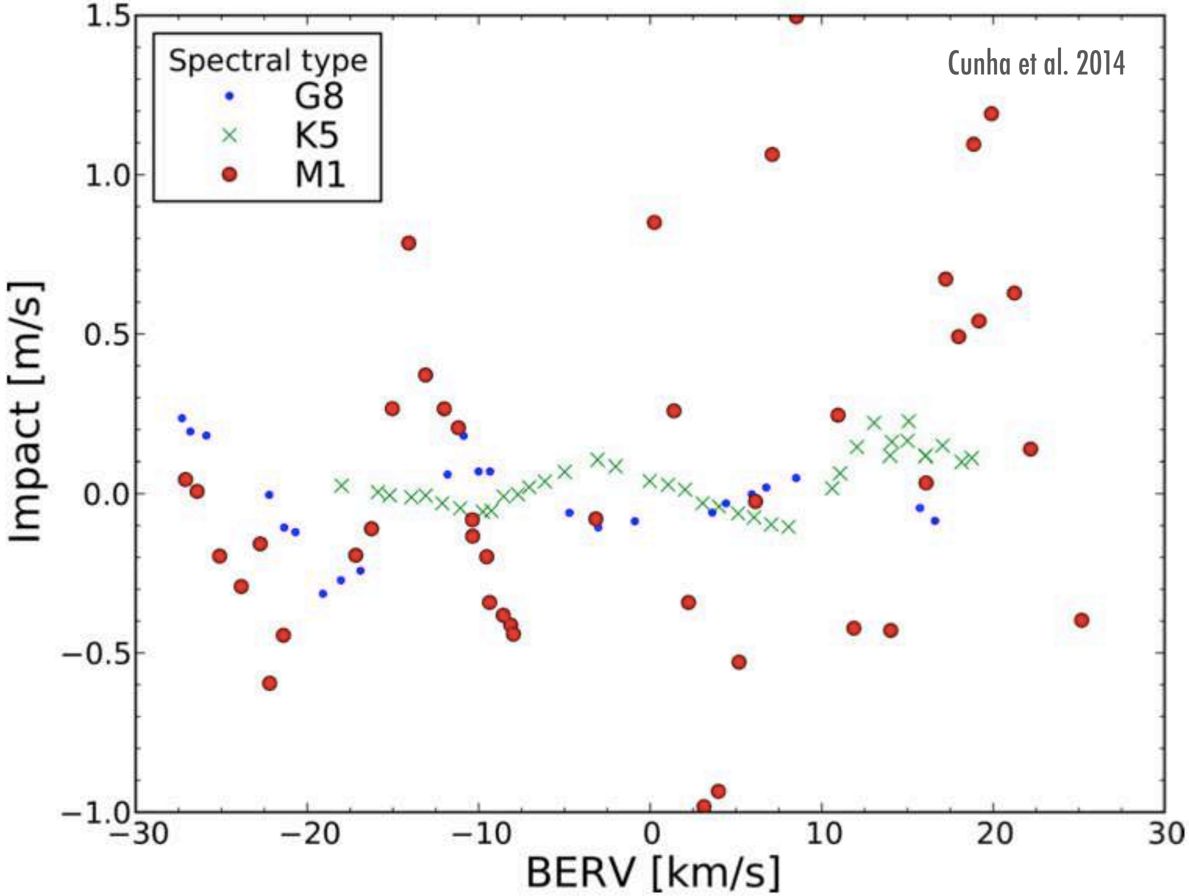


Sithajan, Ge, & Wang, AAS 2016

Introduction to RV Data Analysis

Sky Correction

b. Telluric Absorption



Micro-telluric lines (depth of <1%) are particularly insidious because they are everywhere in the spectrum and cannot be effectively excluded from the CCF mask

Cunha et al. (2014) showed they can impact RVs at the 20 cm/s level or greater if not carefully treated.

Solution: Forward modeling based on sky fiber + measured telluric spectrum *at telescope site* over different seasons

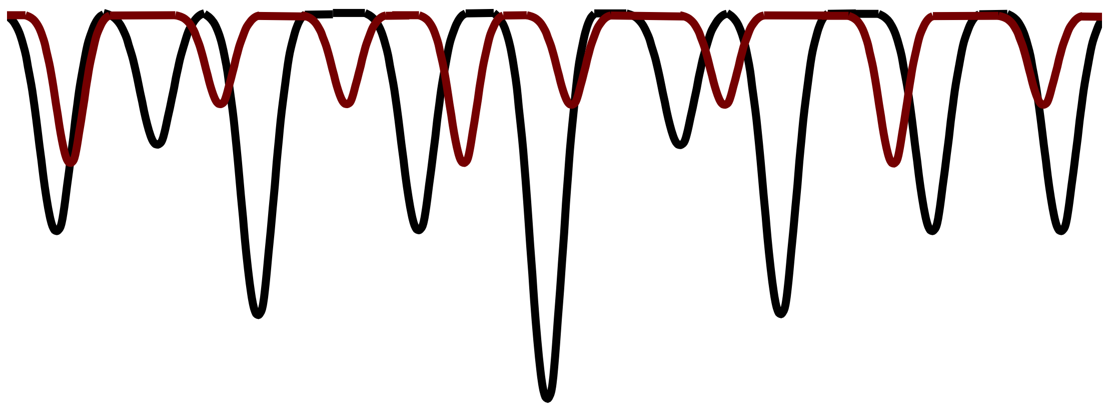
Introduction to RV Data Analysis

Sky Correction

c. Scattered Sunlight Contamination

In general, spectral contamination is a concern

Temporal variability is usually present due to barycentric + stellar motion



Possible Culprits Include:

Other stars, both bound and background

Faint close-by objects – level of contamination is seeing + orientation dependent

Moonlight

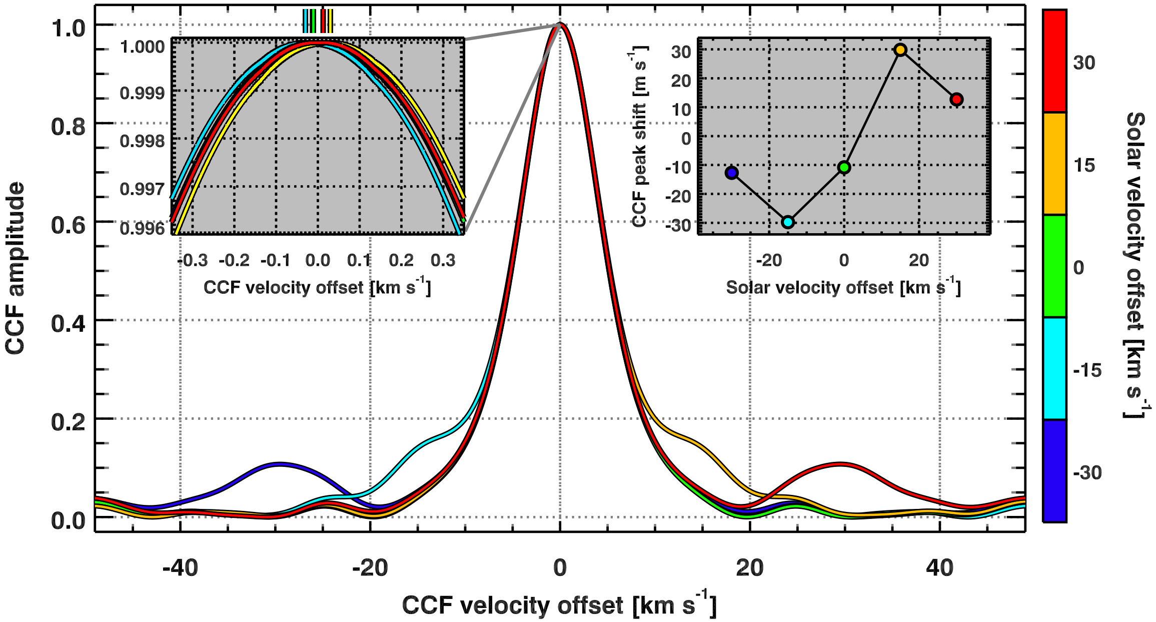
Zodiacal Light

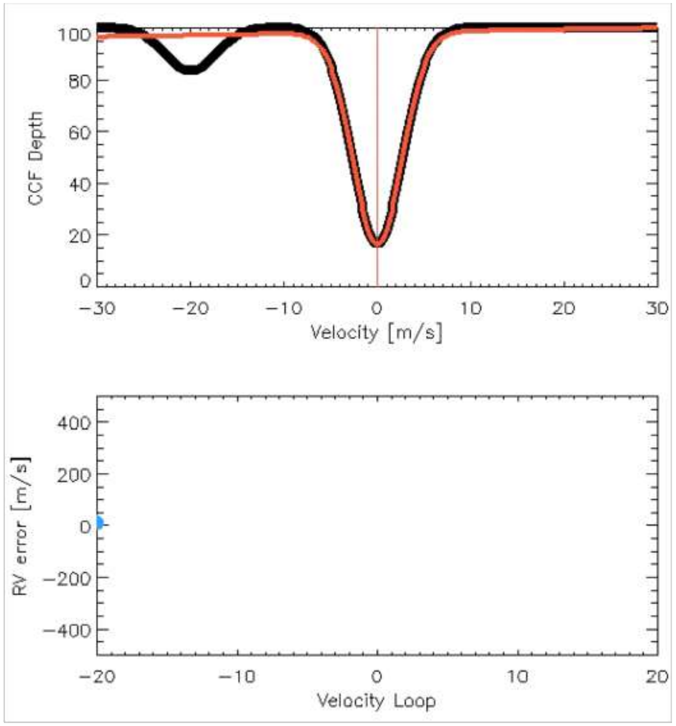
Unexpected sources must be diagnosed directly from the data [eg. Wright, Roy et al. 2013]

Introduction to RV Data Analysis

Sky Correction

c. Scattered Sunlight Contamination





Sunlight is particularly insidious when observing Sun-like stars since the contaminating spectrum is a good match to the mask lines

Solution: Sky fiber is essential
Correct directly from sky fiber in CCF space, or
Model sky spectrum and correct in spectral space
Can also replace sky fiber with broadband proxy for faint sky

Star-Based

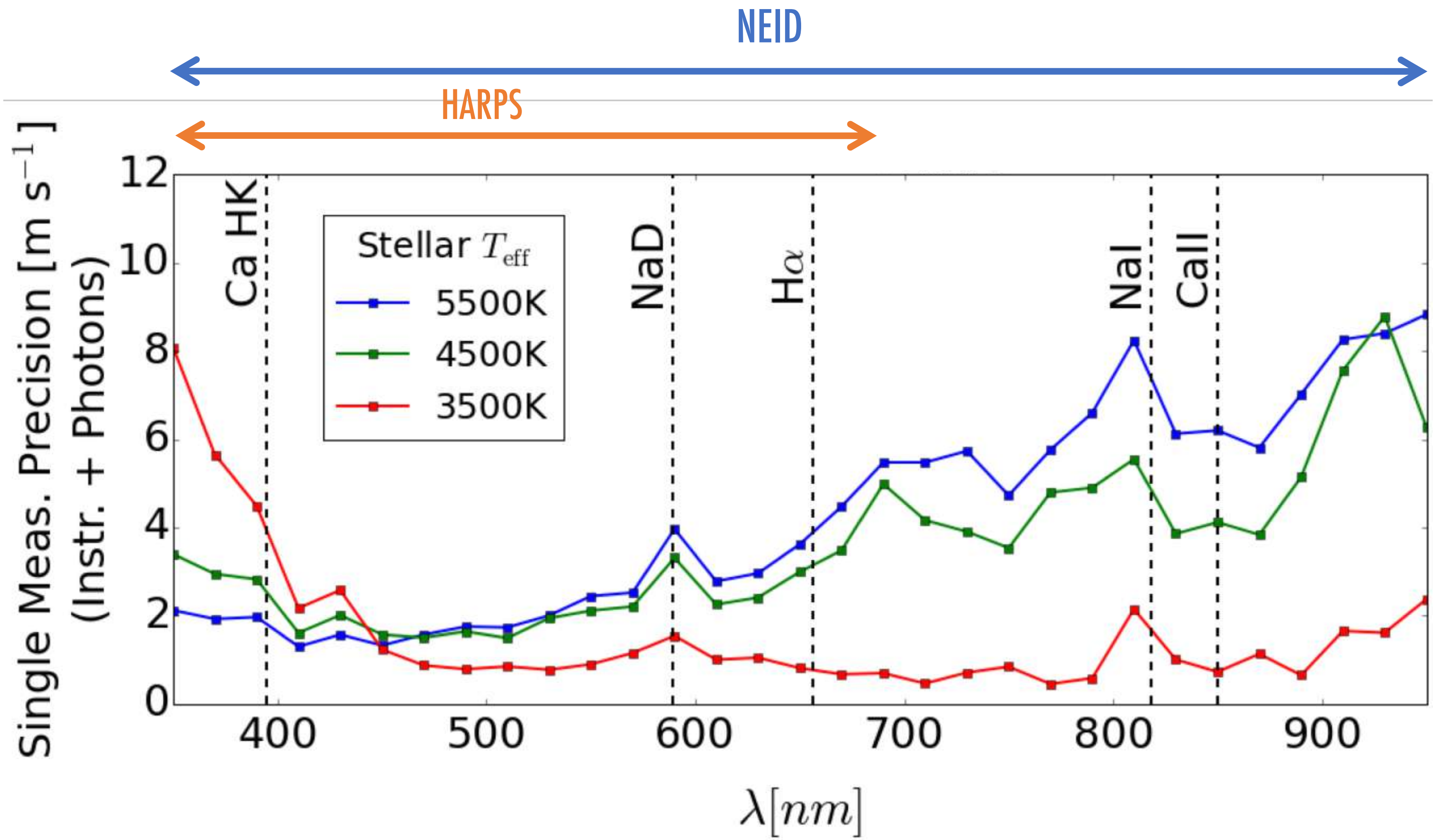
Corrections

Stellar Activity

How Do We Deal With It?

Capture as many activity indicators in bandpass as possible

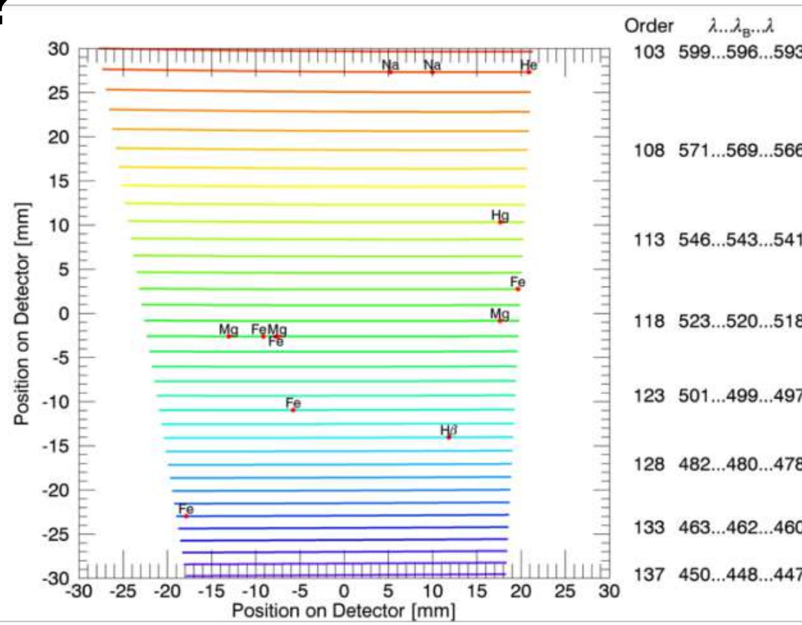
This also maximizes Doppler information content



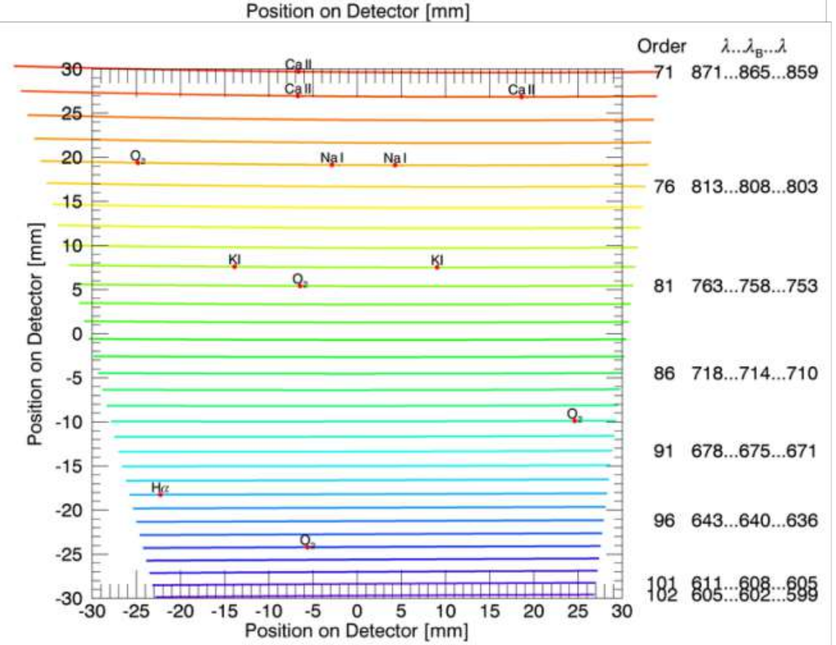
Stellar Activity

How Do We Deal With It?

KPF Green Channel



KPF Red Channel

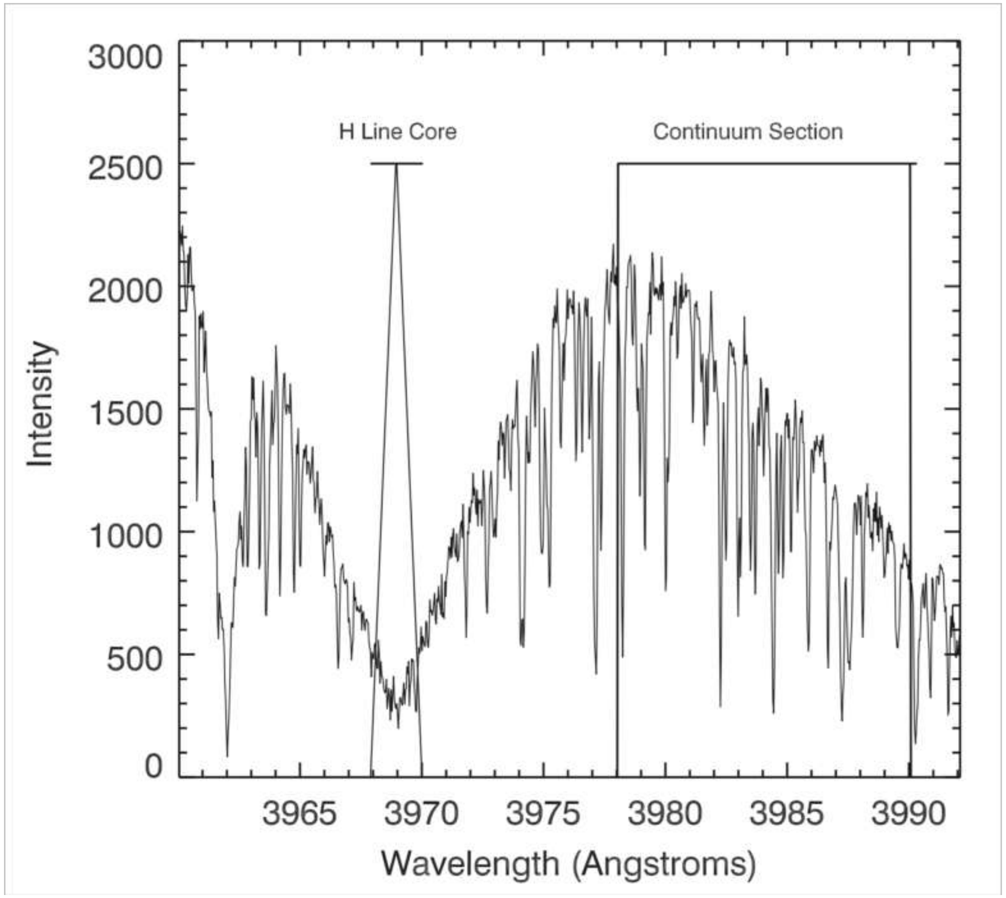


+ Separate Ca H&K spectrophotometer

Stellar Activity

How Do We Deal With It?

Measure Known Activity-Sensitive Lines



Start with Canonical Activity Indicators

Eg. Ca H&K from Isaacson & Fischer 2010

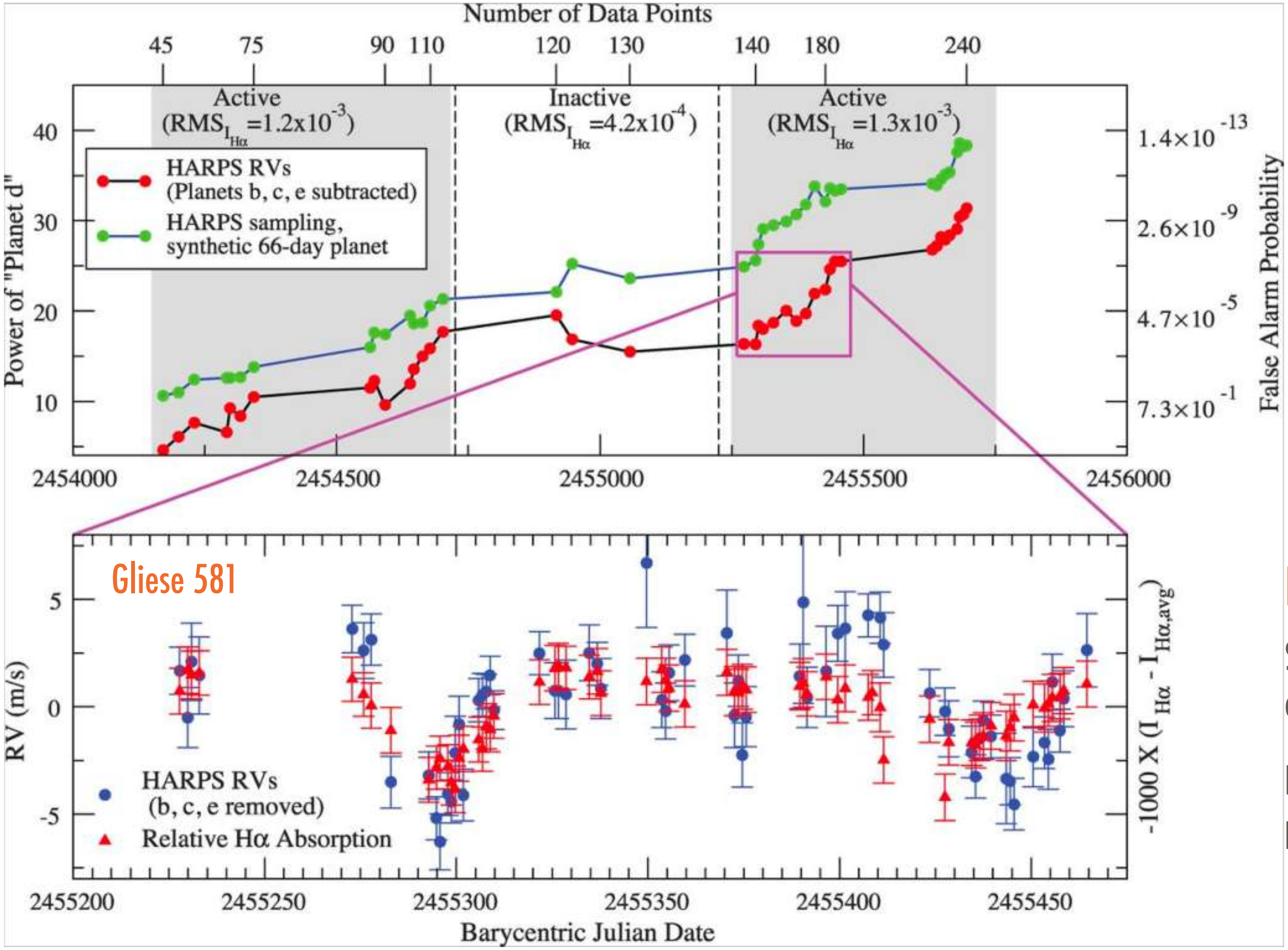
Mainly interested in relative variability but good to have handle on absolute activity levels as well

Stellar Activity

How Do We Deal With It?

Measure Known Activity-Sensitive Lines

Check for correlations with RV



Important Lessons

Stellar activity not coherent over long timescales

Need tight cadence over a period of time to see this

Robertson, Mahadevan, Endl, Roy, Science 2014

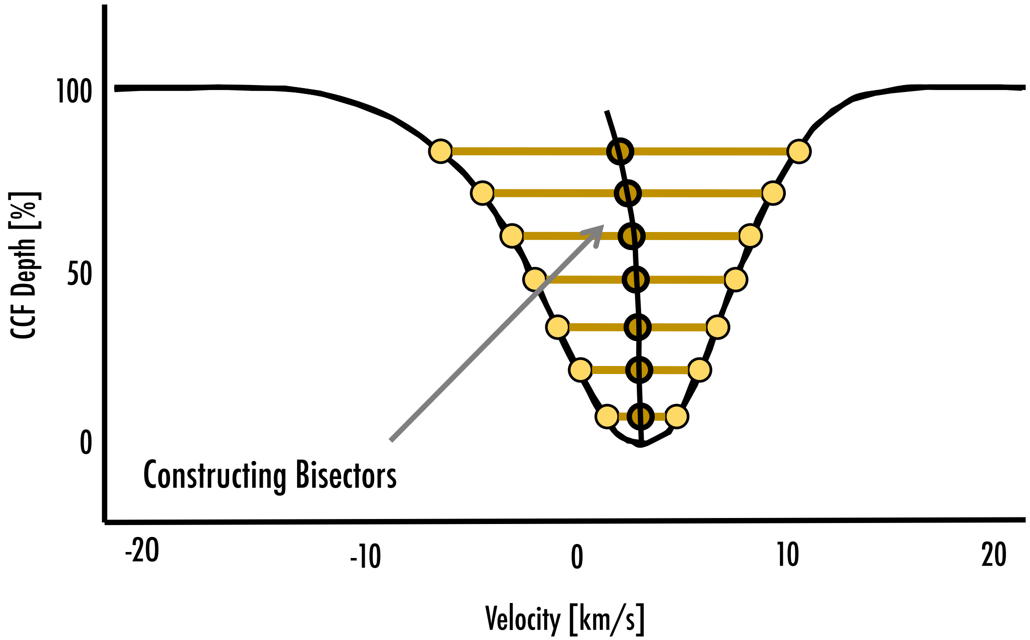
Stellar Activity

How Do We Deal With It?

Scrutinize Line Profiles: Bisectors & FWHM

If the instrument profile is stable then any line variations can be ascribed an astrophysical origin

Bisectors have long been used to distinguish bulk motion of the star from spectral line asymmetries due to activity or contamination

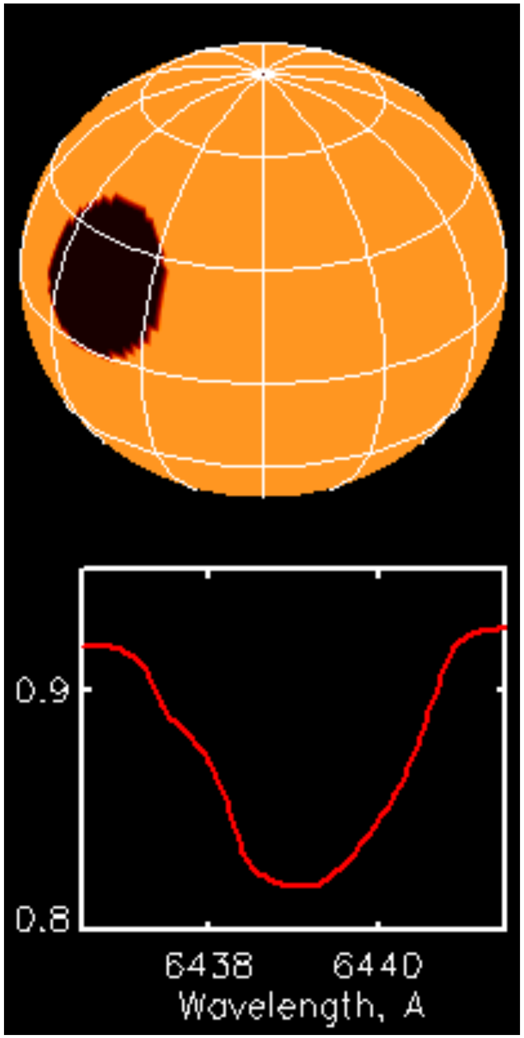


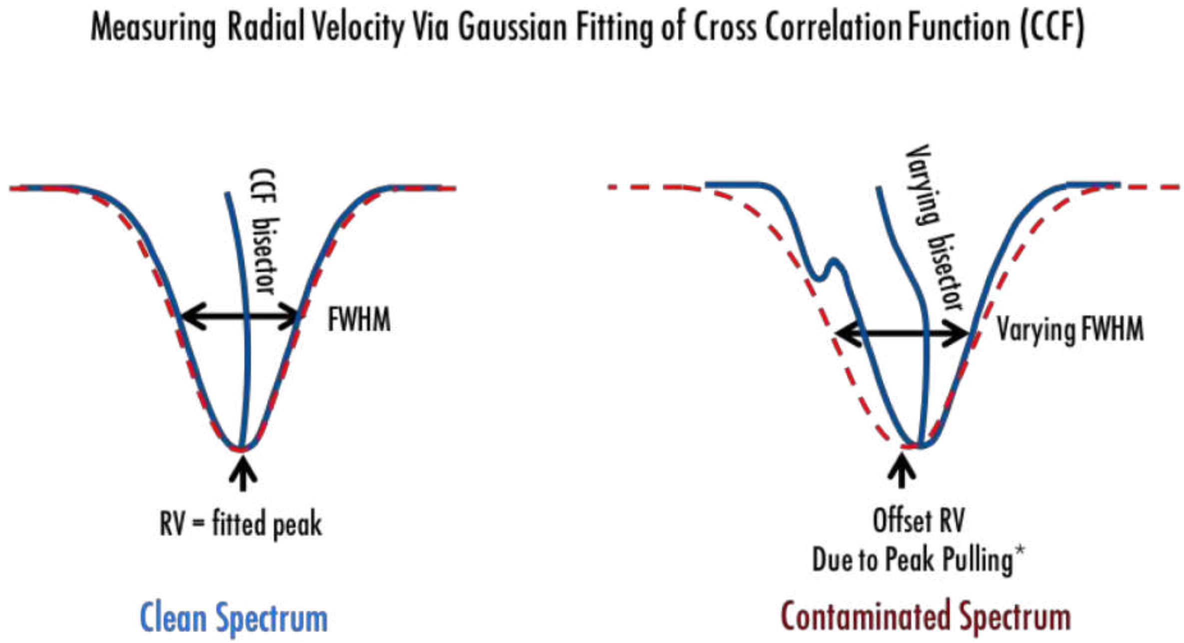
- True Doppler motion should not affect line shapes
- Single lines have low SNR in high resolution spectroscopy
- Use the cross correlation function as an "aggregate line"

Stellar Activity

How Do We Deal With It?

Scrutinize Line Profiles: Bisectors & FWHM





Bisector variation is caused by both stellar activity and spectral contamination

Need to be very careful which lines to "average"

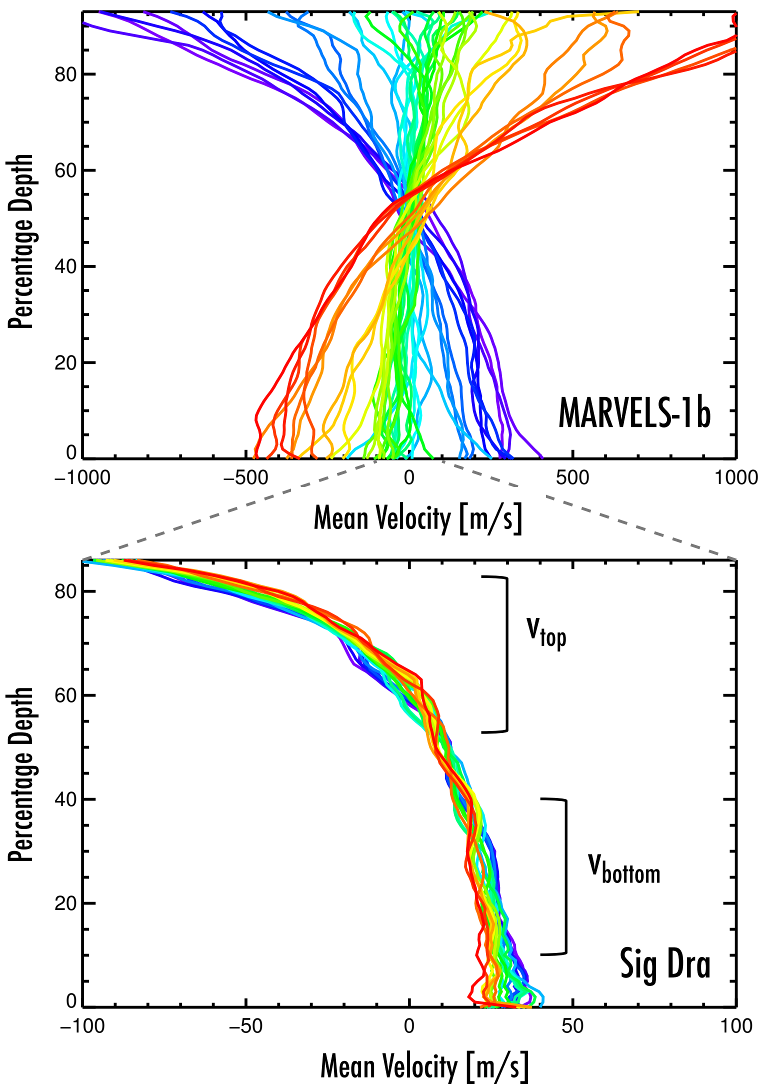
Need to be very careful what wavelength range to use

[otherwise dilute chromatic effect]

Stellar Activity

How Do We Deal With It?

Scrutinize Line Profiles: Bisectors & FWHM



Example: Marvels-1b

Suspicious planet signal

Investigation revealed extremely variable bisectors

[Took a while to see this because of lack of stabilized instruments]

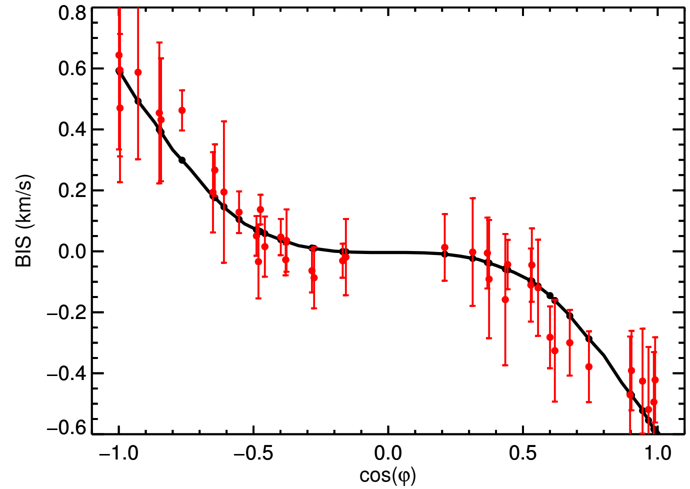
Bisector Inverse Slope [BIS] = v_{top} - v_{bottom}

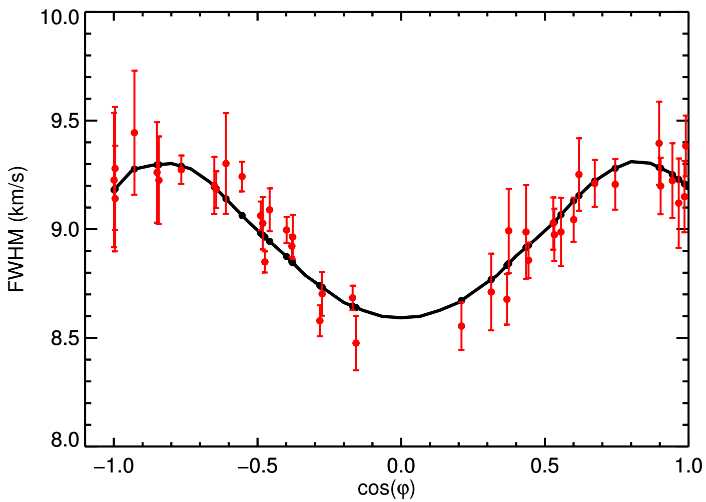
Wright, Roy et al. 2013

Stellar Activity

How Do We Deal With It?

Scrutinize Line Profiles: Bisectors & FWHM





Example: Marvels-1b

Clear correlation between RV and both BIS and FWHM

Modeling showed it was a rare face-on binary masquerading as a double planet system

Unfortunately stellar activity signal not so coherent or clean

Wright, Roy et al. 2013

Pipeline Development

Advancing as a Community

KPF Pipeline Development

Investing in Building A Community Resource

Overarching Philosophy

- Best practices in software engineering
- Written in Python with small subset of specific widely-used libraries
- Other languages if necessary (IDL for heritage, C for speed)
- Object-oriented
- Highly modular
- Version controlled in GitHub
- Unit/regression testing implemented on HARPS/PARAS spectra and KPF simulator data

DRP Management

- Substantial subsystem (both in terms of budget and schedule)
- Full project management under PM and DRP Lead
- Considering separate reviews for DRP

DRP Distribution

- Pipeline will be public, open source, and benefit from community improvement
- Acutely aware of long-term arc of DRP and RV precision optimization