Discovering Exoplanets With Hermite-Gaussian Linear Regression

Yale

Parker Holzer,
Department of Statistics & Data Science, Yale University

09/11/2020

Work done in collaboration with…

Jessi Cisewski-Kehe[*]

Debra Fischer[†]

Lily Zhao[†]

[*]Department of Statistics, University of Wisconsin - Madison

[†]Department of Astronomy, Yale University

Introduction: The Radial Velocity Method

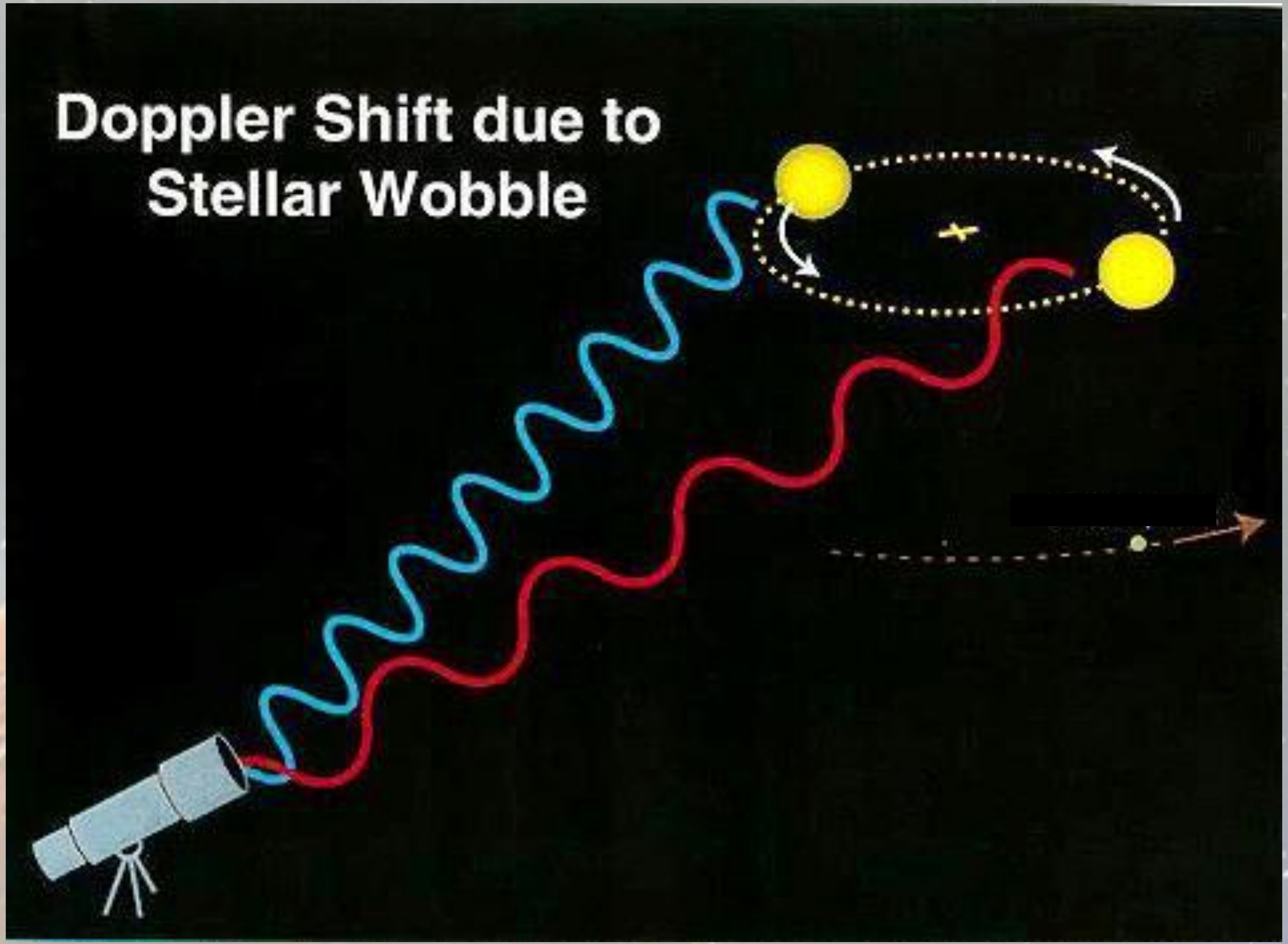
Christian Doppler
(1803-1853)

v_r : radial velocity

λ : wavelength of light

$$\gamma_v = 1 + \frac{v_r}{c}$$

$$\lambda_v = \gamma_v \cdot \lambda$$

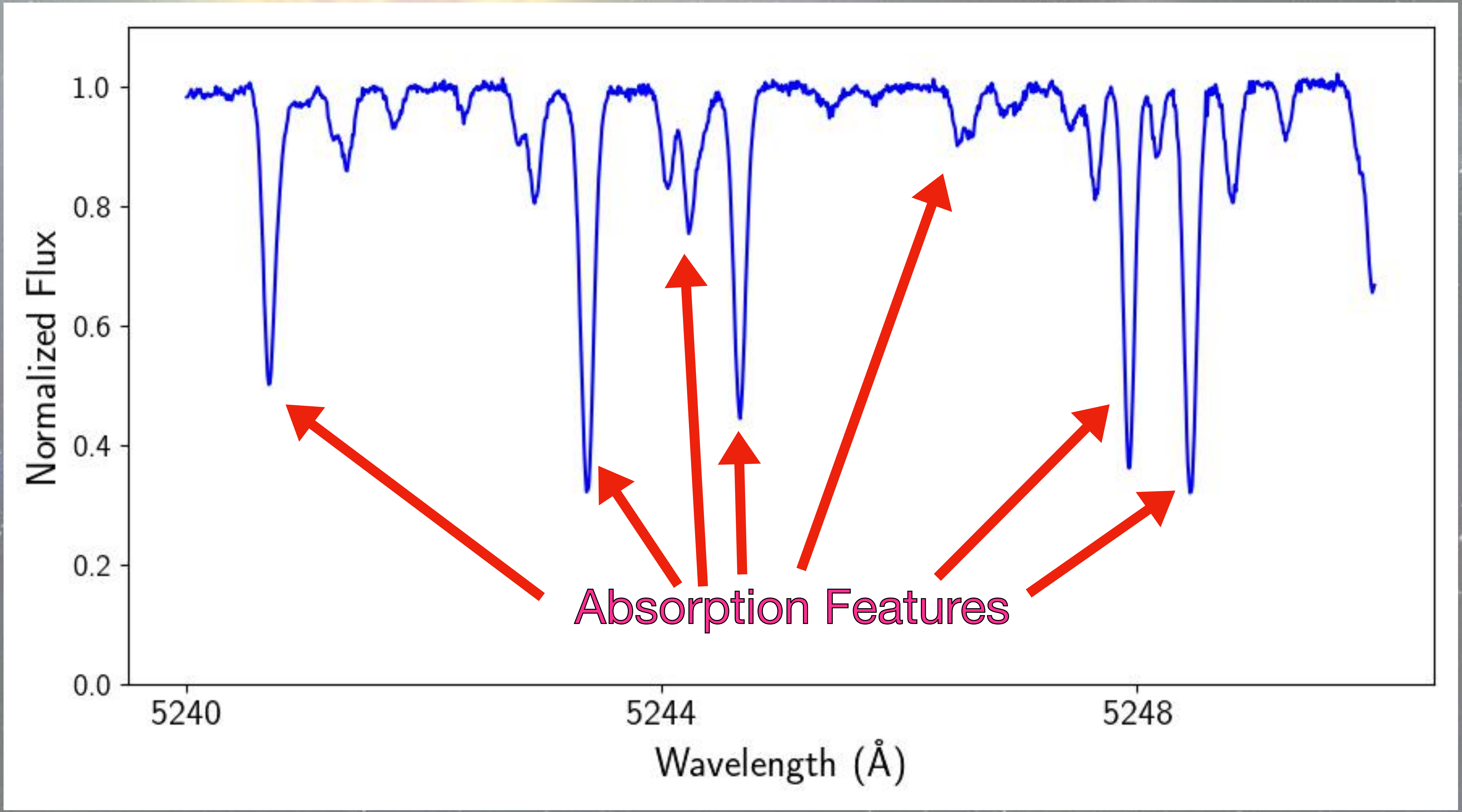


Two steps:

1. At each observed time, t, detect a Doppler-shift and infer a $v_r(t)$.

2. Fit a “RV Curve” over time to $v_r(t)$.

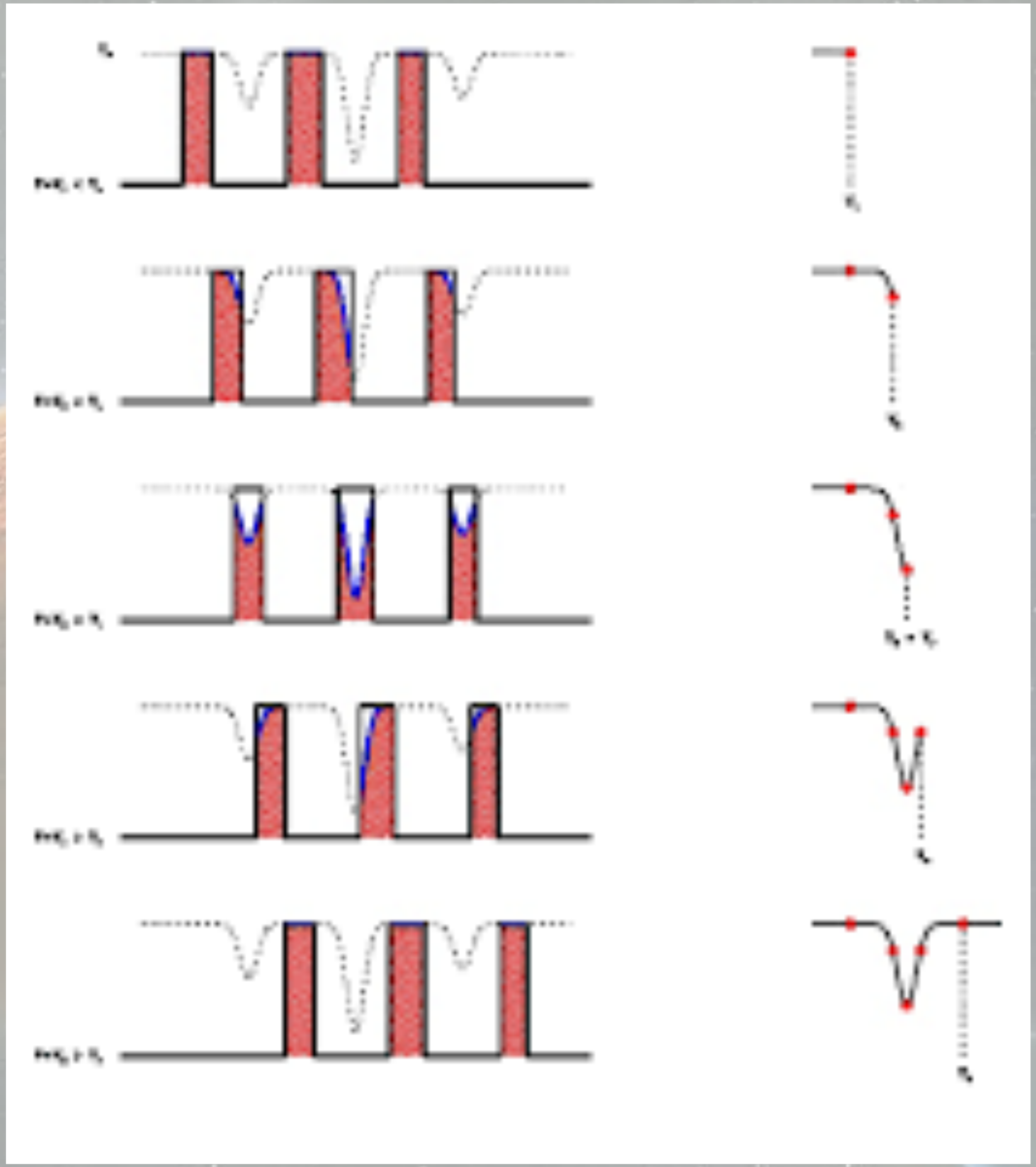
Introduction: Stellar Spectra



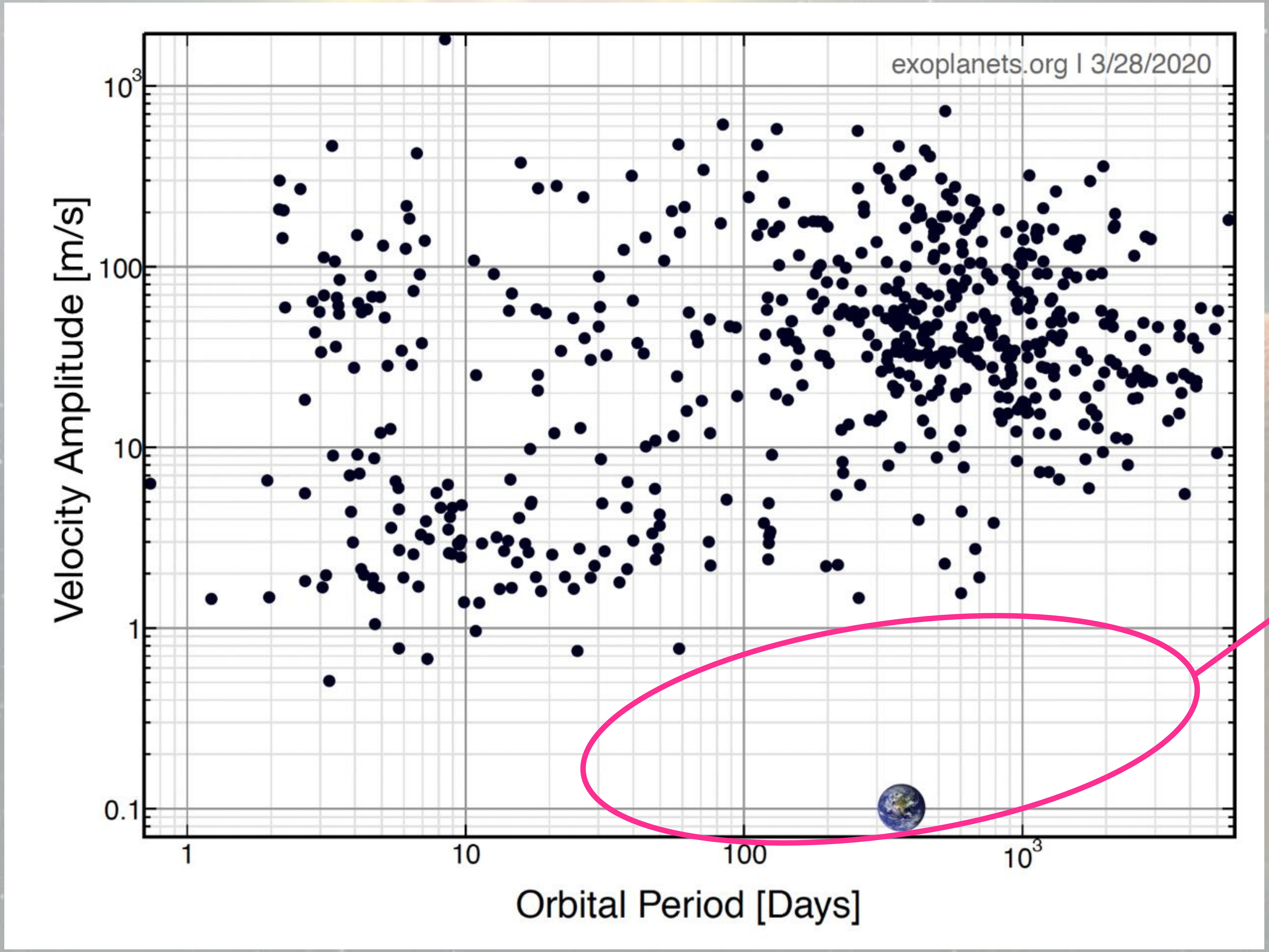
51 Pegasi spectra collected by EXPRES [(Petersburg et al. 2020)]()

RV Method Techniques

- Cross-Correlation Function (CCF)
    - (e.g. Pepe et a. (2002))
- Other Techniques
    - Template Matching (e.g. Astudillo-Defru et al. (2015))
    - Forward-Modeling (e.g. Petersburg et al. (2020))
    - etc.



Current State of Discovered Exoplanets

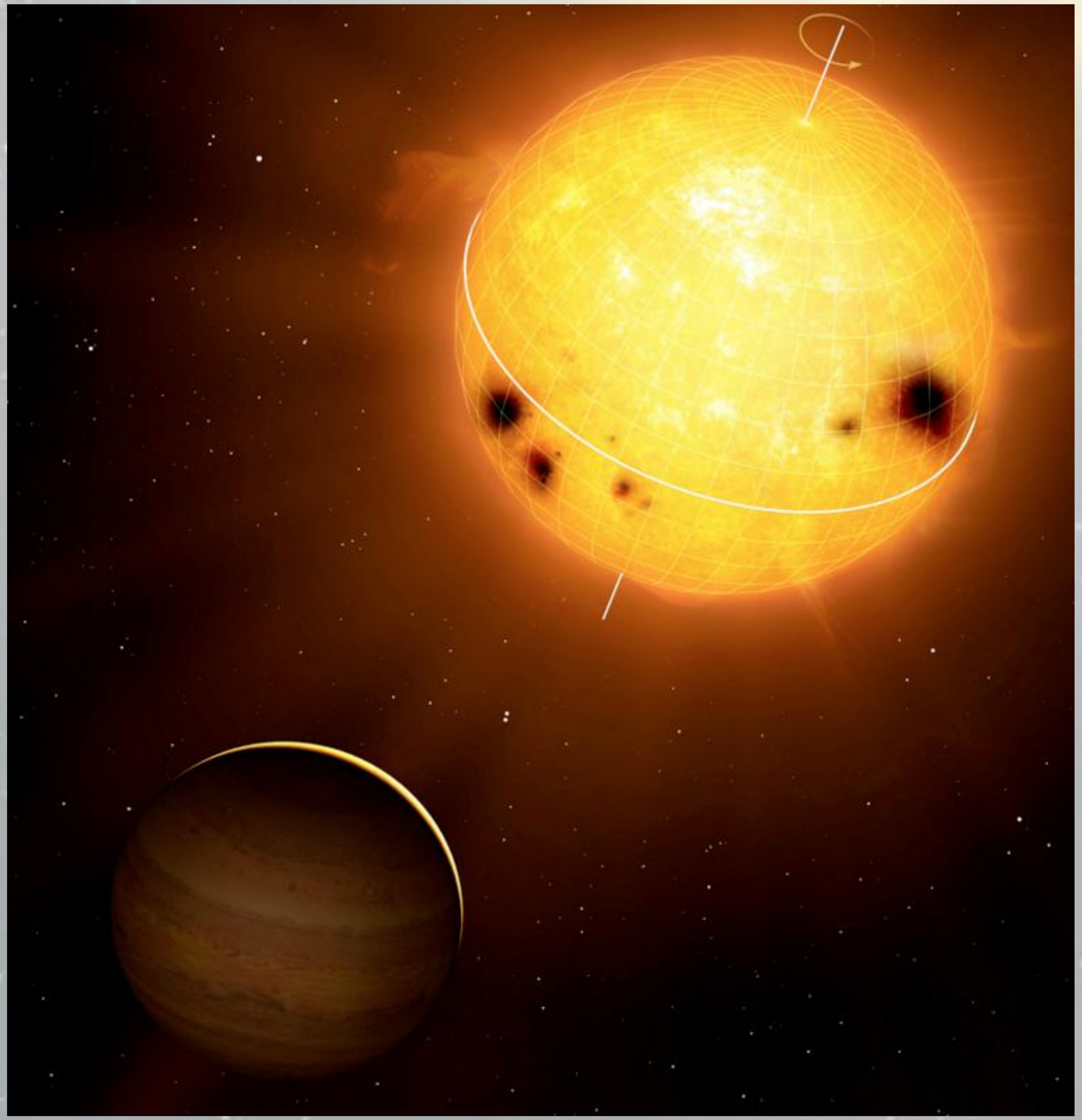


Objective:

Discover more Earth-like, Sun-like systems!

Exoplanets discovered via RV Method

Underlying Issue: Stellar Activity



Stellar activity (starspots, faculae, granulation, etc.) can mimic the signal of exoplanets!

The effects of stellar activity do not currently have a well-known (and accurate) physical model.

There does exist a detectable difference between signals of stellar activity and a Doppler shift (Davis et al. (2017)).

Attempts to Disentangle Stellar Activity

RV Estimation

- Detect, and remove, absorption features that are sensitive to stellar activity. (Dumusque (2018); Ning et al. (2019); Petersburg et al. (2020))
    - Drawback: at least 89% of absorption features are sensitive to stellar activity Cretignier et al. (2019)

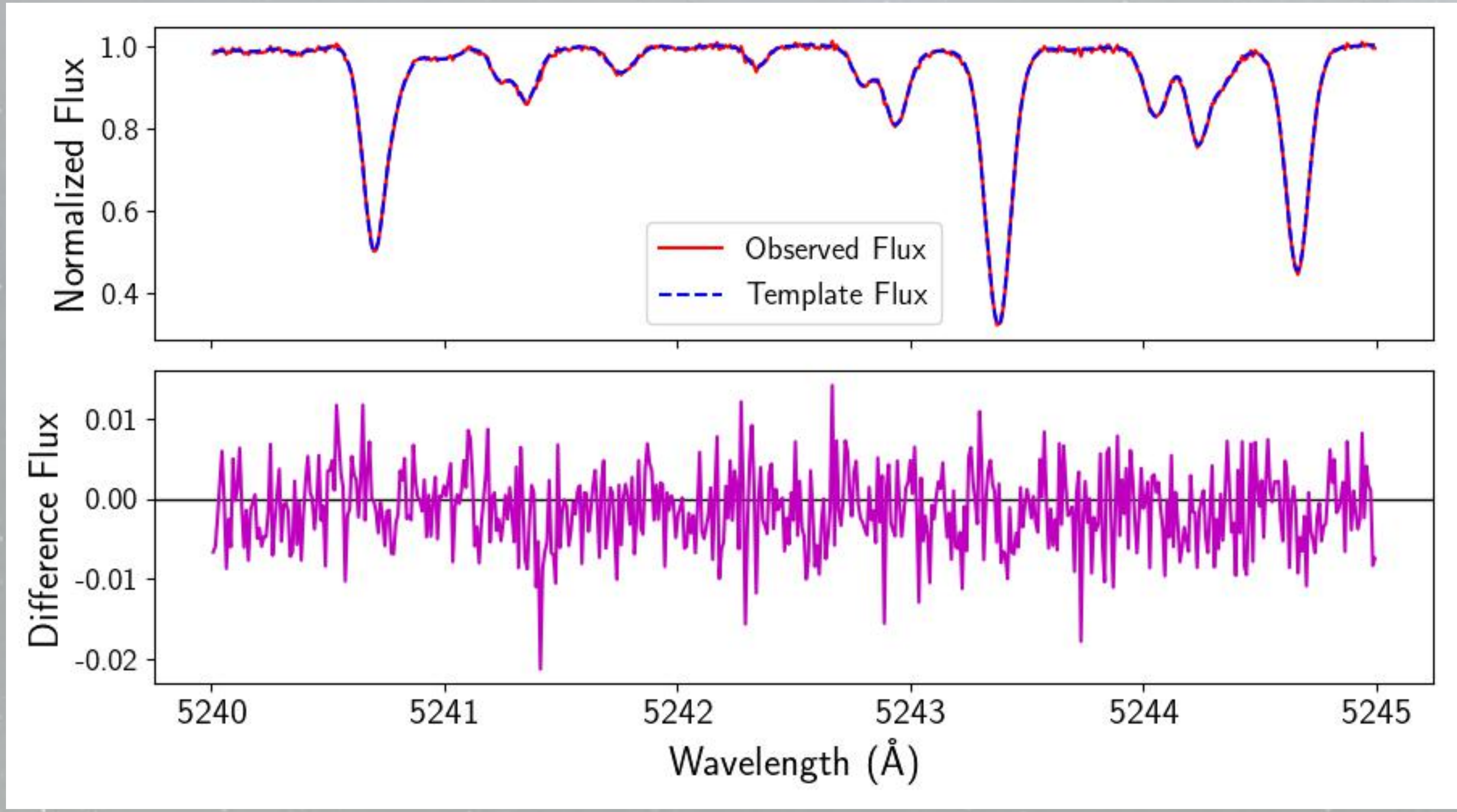
- **This Work:** Reformulate the RV estimation in terms of an easily extendable statistical method.

RV Curve Fitting

- Fit the estimated RV's simultaneously with a time series of stellar activity indicators using Gaussian Processes. (Rajpaul et al. (2015); Jones et al. (2017))

Hermite-Gaussian Radial Velocity (HGRV) method at a glance

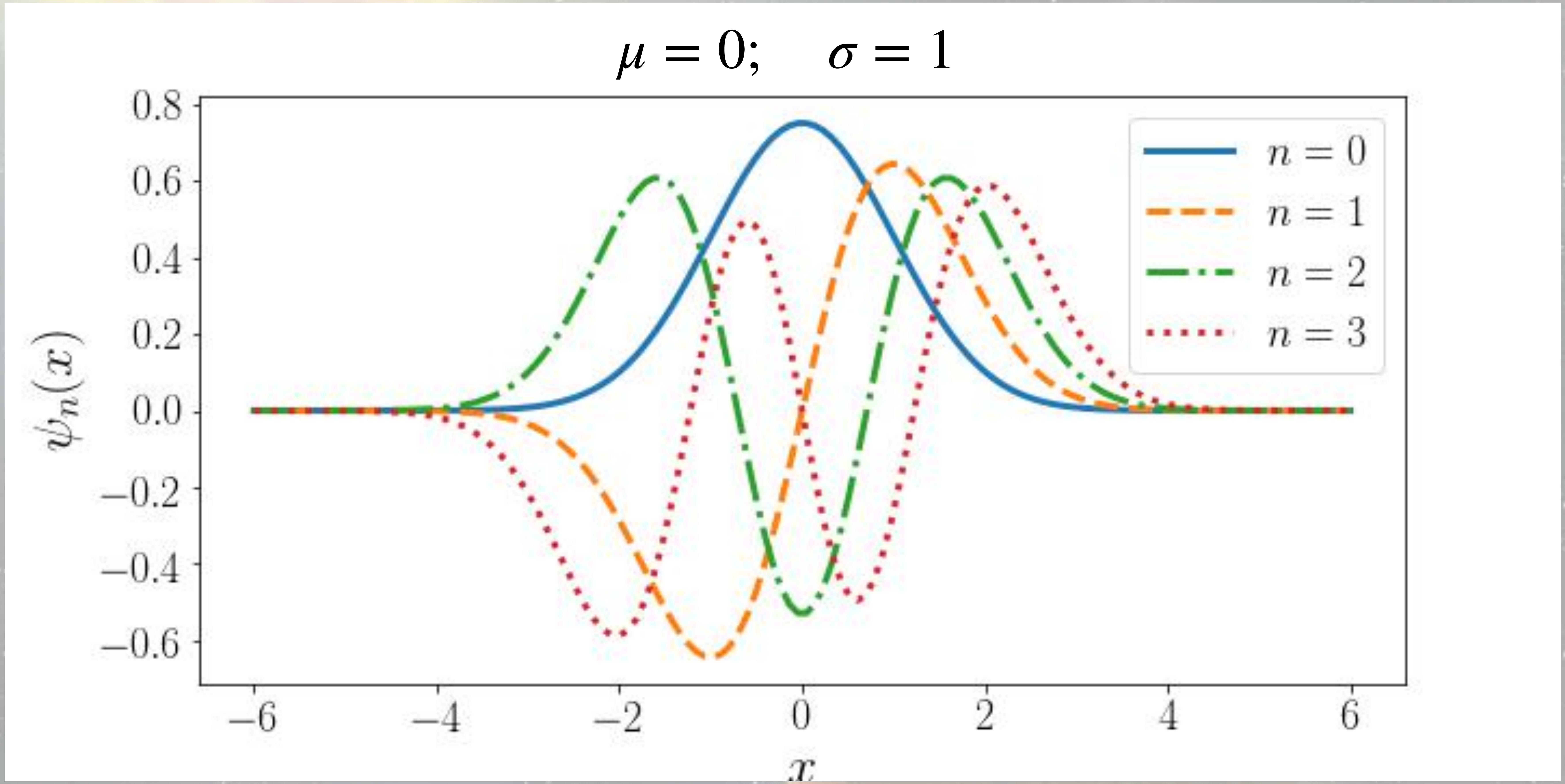
Assume the true velocity is small (i.e. due to an exoplanet) and estimate the velocity with (weighted) simple linear regression of the difference flux on a sum of first-degree Hermite-Gaussian functions.



difference flux: $\text{flux}_{\text{observed}} - \text{flux}_{\text{template}}$

"quiet" spectrum estimated through nonparametric smoothing

Definitions



$\psi_n(x; \mu, \sigma)$: n^{th} degree generalized Hermite – Gaussian function

$$\psi_n(x; \mu, \sigma) = \frac{1}{\sqrt{\sigma 2^n n! \sqrt{\pi}}} H_n\left(\frac{x-\mu}{\sigma}\right) e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

H_n : n'th degree (physicist's) Hermite polynomial

$D(\phi || \varphi)$: standardized approximation error for approximating φ with ϕ

$$D(\phi || \varphi) = \frac{\int_{-\infty}^{\infty} (\varphi(x) - \phi(x))^2 dx}{\int_{-\infty}^{\infty} \varphi(x)^2 dx}$$

Theorem: *For any* $\sigma > 0$ *and any* μ, ξ *and* $g(x;\xi) = e^{-\frac{(x-\mu)^2}{2\sigma^2}} - e^{-\frac{(\xi x-\mu)^2}{2\sigma^2}}$ *decomposed in the Hermite-Gaussian basis as* $g(x;\xi) = \sum_{n=0}^{\infty} c_n(\xi)\psi_n(x;\mu,\sigma)$, $\lim_{\xi \to 1} D(c_1(\xi)\psi_1(x;\mu,\sigma) || g(x;\xi)) = \frac{1}{1+\frac{2\mu^2}{3\sigma^2}}$.

At small radial velocities, the difference between a Gaussian and a Doppler-shift of it can be well approximated as a constant multiple of the first-degree Hermite-Gaussian function.

HGRV Linear Model

$$y_i = v_r \sum_{j=1}^{m} \frac{\sqrt{\sqrt{\pi}\, d_j \mu_j}}{c\sqrt{2\sigma_j}} \psi_1\left(x_i; \mu_j, \sigma_j\right) + \varepsilon_i \quad , \quad \varepsilon_i \sim N(0,\varrho_i)$$

y_i : difference flux of pixel i

x_i : wavelength of pixel i

m : number of absorption features

v_r : unknown velocity coefficient

ψ_1 : first – degree Hermite – Gaussian function

c : speed of light

$\boxed{d_j,\ \mu_j,\ \sigma_j,\ \varrho_i}$: additional parameters

easily estimated separately
with template spectrum

HGRV Linear Model

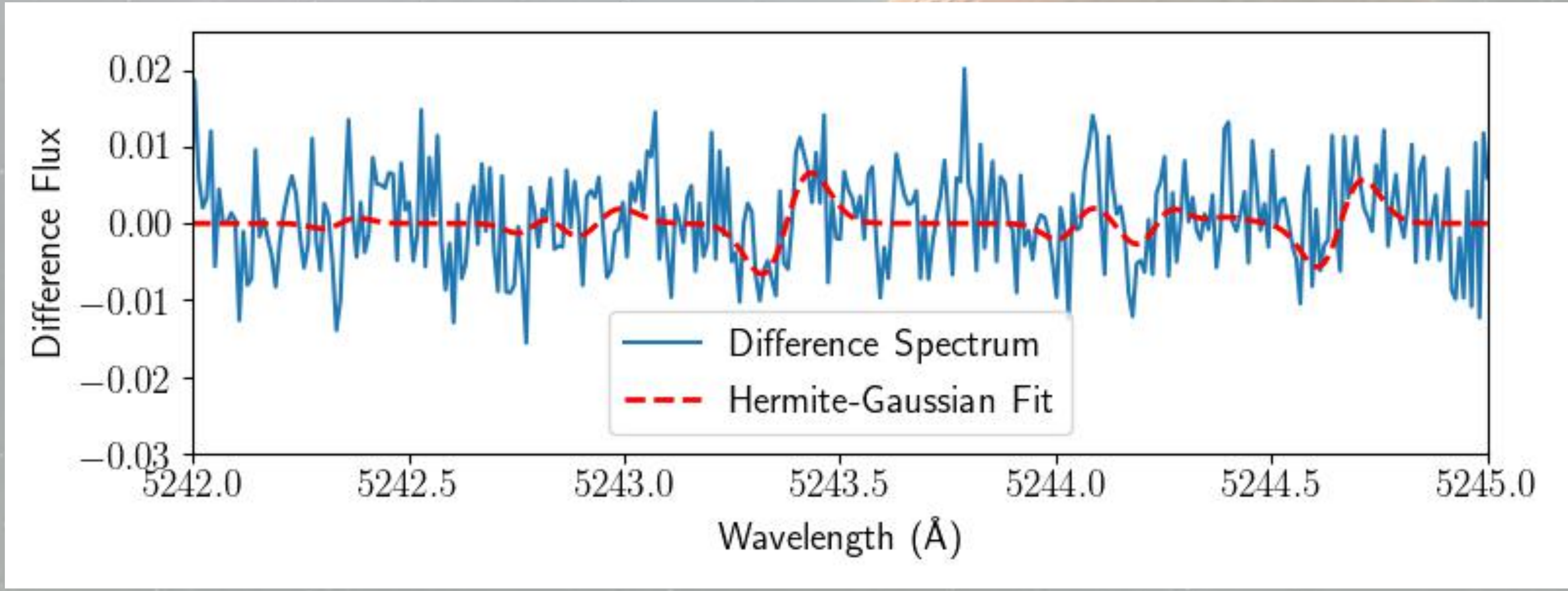
$$y_i = v_r \sum_{j=1}^{m} \frac{\sqrt{\sqrt{\pi}\, d_j \mu_j}}{c\sqrt{2\sigma_j}} \psi_1\left(x_i; \mu_j, \sigma_j\right) + \varepsilon_i \quad , \quad \varepsilon_i \sim N(0, \varrho_i)$$

Coefficient to estimate

Sum across m absorption features

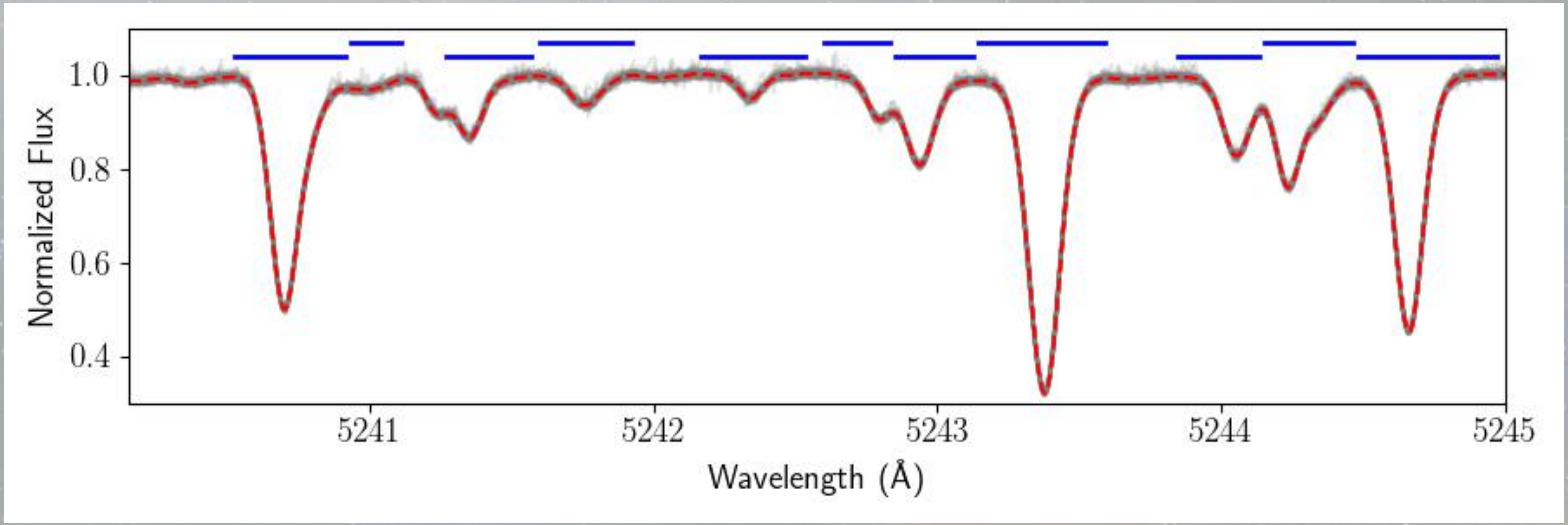
Relative amplitude of absorption feature j

First-degree Hermite-Gaussian function centered at μ_j



Application to 51 Pegasi

Step 1: Stack all observed spectra and estimate the template through local quadratic regression

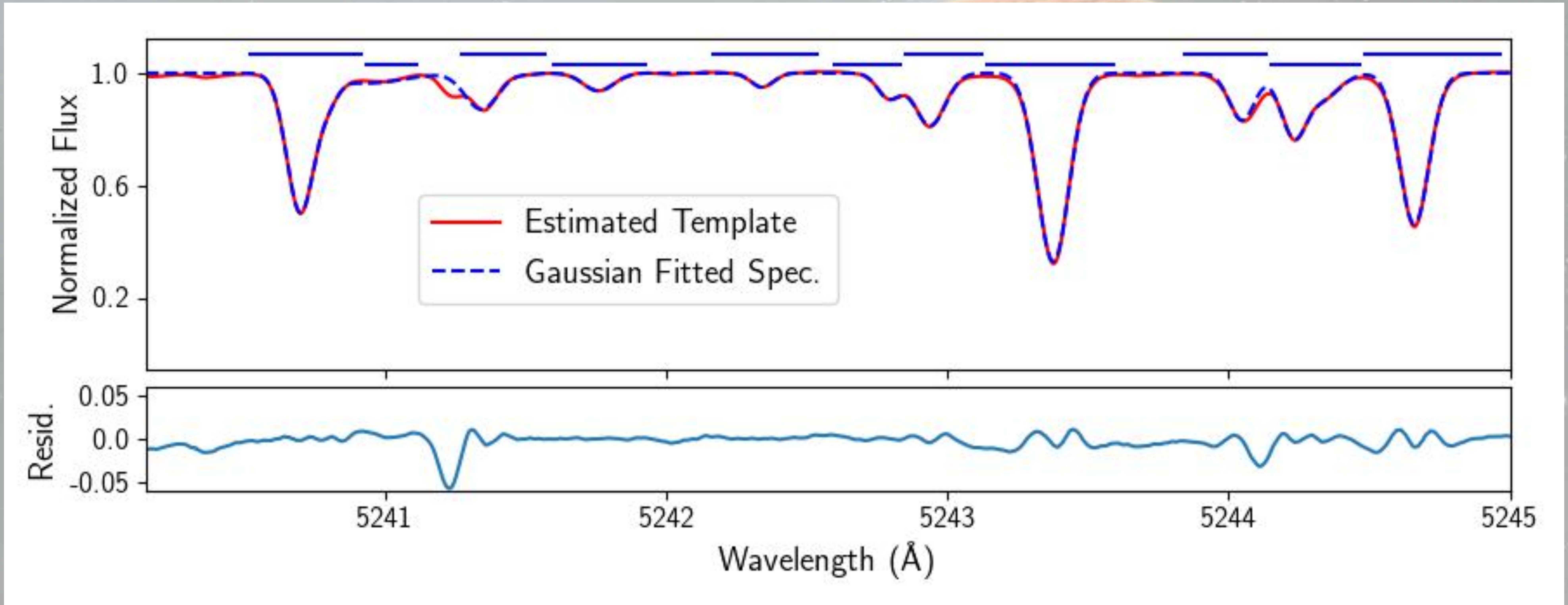


56 observations of 51 Pegasi collected by EXPRES (Petersburg et al. 2020)

Application to 51 Pegasi

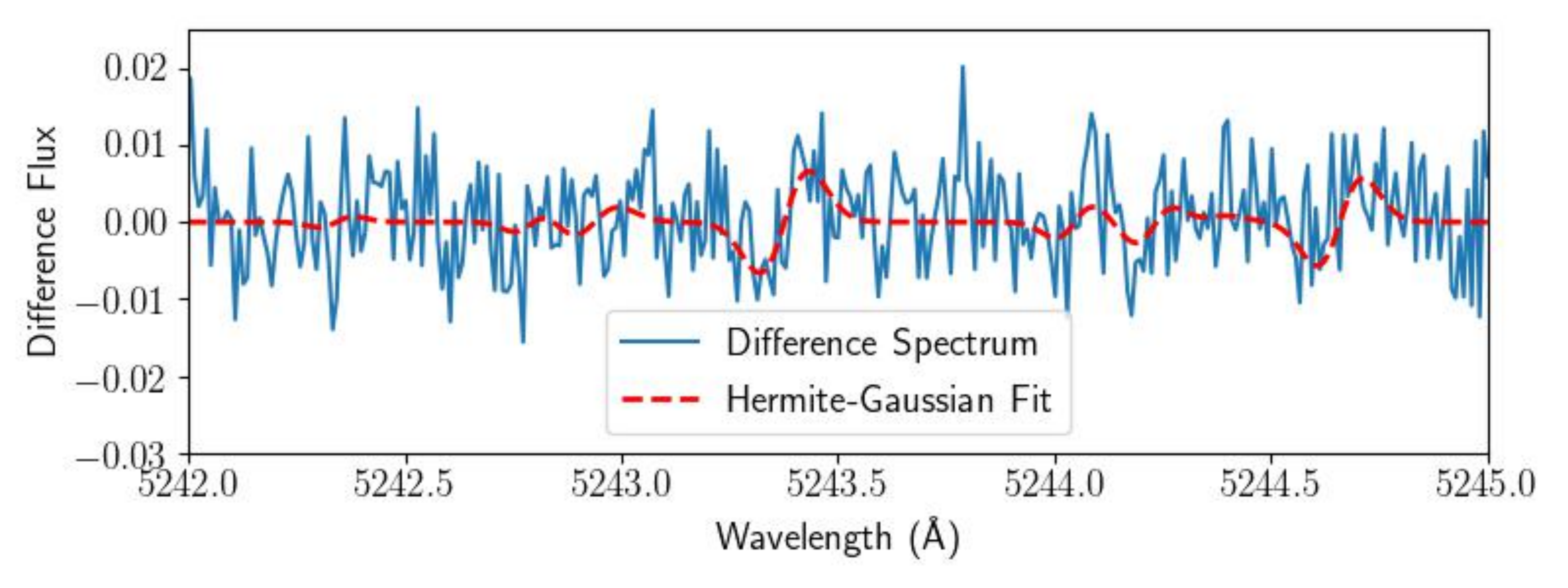
Step 2: Identify wavelength intervals of absorption features in the estimated template (using Absorption Feature Finder algorithm in Holzer et al. (in review)).

Step 3: Estimate d_j, μ_j, and σ_j for each j by fitting Gaussians to all absorption features in the estimated template.



Application to 51 Pegasi

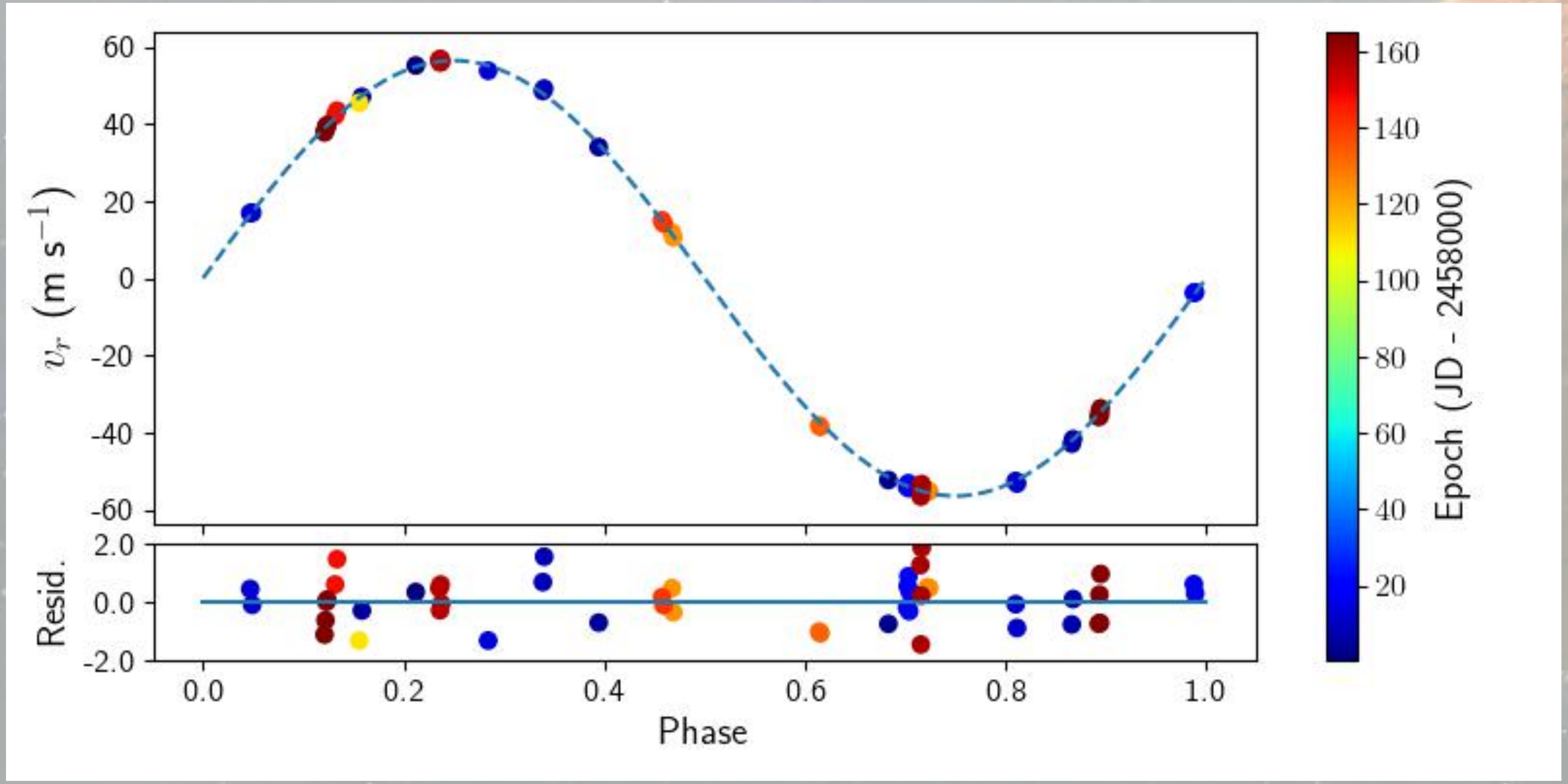
Step 4: Estimate v_r, and its uncertainty, using simple linear regression (without an intercept).



$$y_i = v_r \sum_{j=1}^{m} \frac{\sqrt{\sqrt{\pi}\, d_j \mu_j}}{c\sqrt{2\sigma_j}} \psi_1\left(x_i; \mu_j, \sigma_j\right) + \varepsilon_i \quad , \quad \varepsilon_i \sim N(0, \varrho_i)$$

Application to 51 Pegasi

	HGRV	CCF	FM
$\hat{K}$	56.48 ± 0.16 m s^{-1}	56.20 ± 0.19 m s^{-1}	56.17 ± 0.18 m s^{-1}
$\hat{P}$	4.2308 ± 0.0001 days	4.2304 ± 0.0002 days	4.2306 ± 0.0002
$\hat{\phi}$	-1.333 ± 0.006	-1.326 ± 0.007	-1.331 ± 0.007
RMS	0.774 m s^{-1}	0.936 m s^{-1}	0.902 m s^{-1}



$$v_r(t) = K \cdot sin\left(\frac{2\pi}{P}t + \phi\right) + b$$

K : velocity semi – amplitude

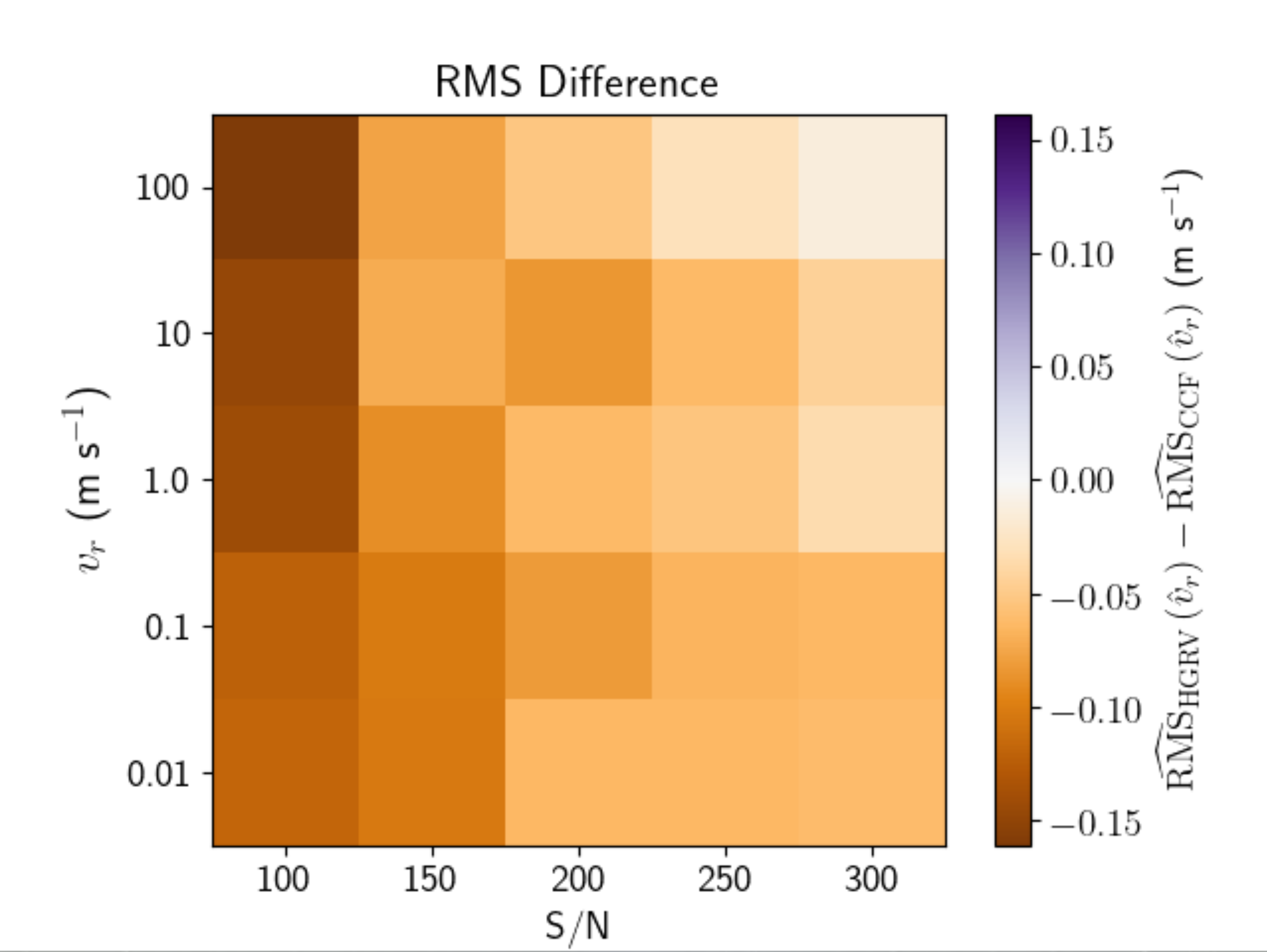
P : orbital period

ϕ : phase offset

b : velocity offset

CCF and FM (Forward-modeling) velocities provided by Petersburg et al. (2020)

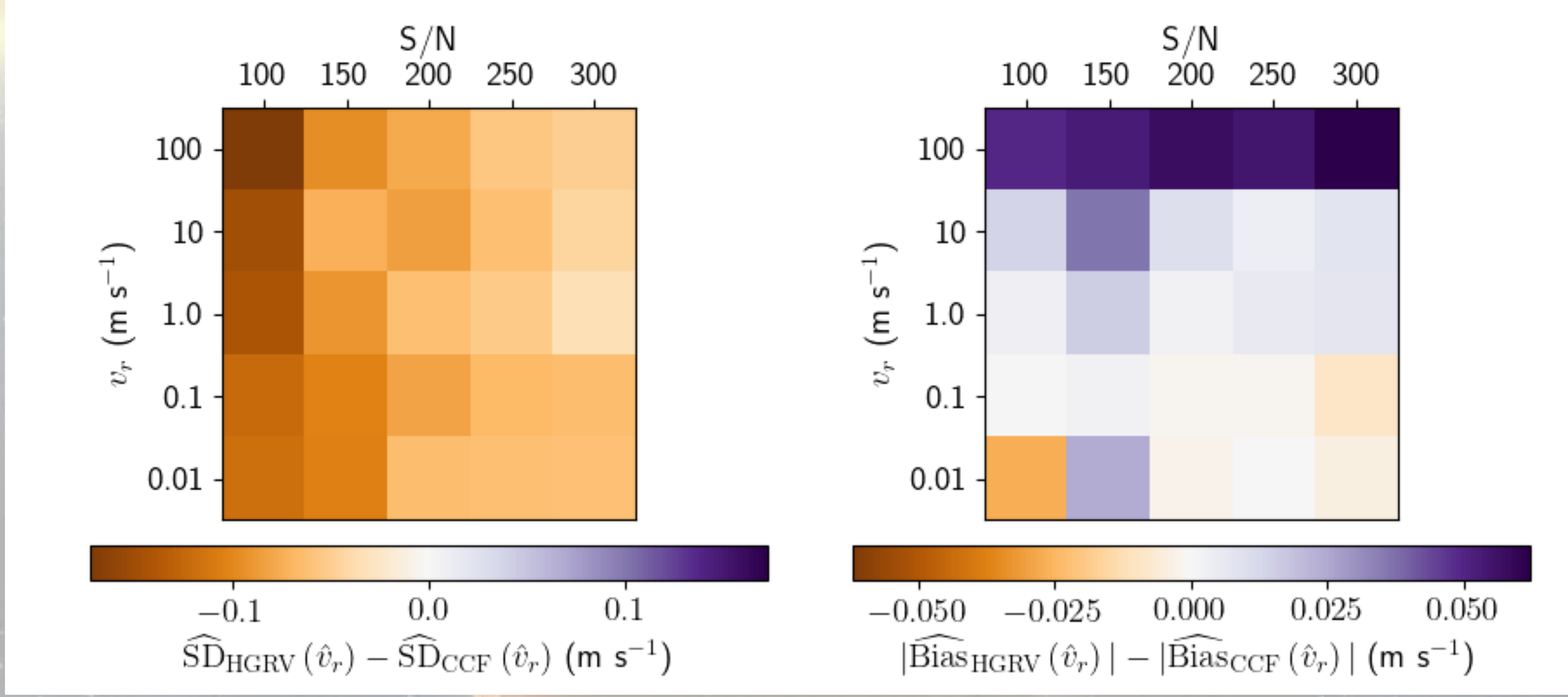
Simulation Studies:
Comparison to the classical Cross-Correlation Function (CCF) approach



The HGRV approach outperforms the traditional CCF approach.

Simulation Studies:

Comparison to the classical Cross-Correlation Function (CCF) approach

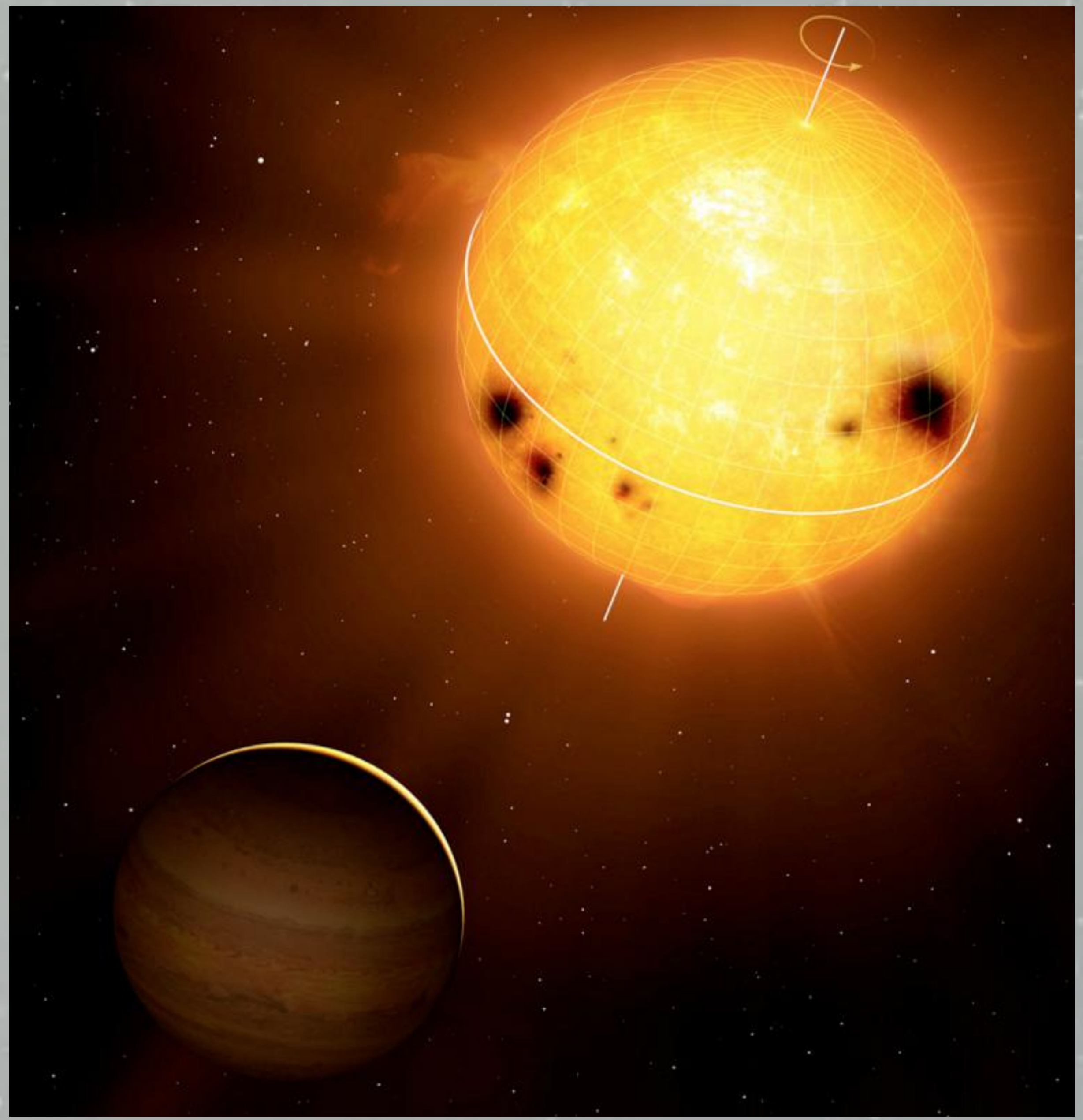


The HGRV is an example of reducing the overall MSE by adding in a small amount of bias.

Sources of bias:

- Assuming absorption features are Gaussian shaped
- Treating a multiplicative shift (Doppler shift) as an additive shift

Extending the HGRV Method



1. Indicate the presence of stellar activity
    - HGRV Extension: F-test with additional variables composed of higher-degree Hermite-Gaussian functions. (Holzer et al. (in prep)
2. Model out stellar activity
    - HGRV Extension: estimate stellar activity signal using variables built to be orthogonal to HGRV variable. (Holzer et al. (in prep)

Conclusions

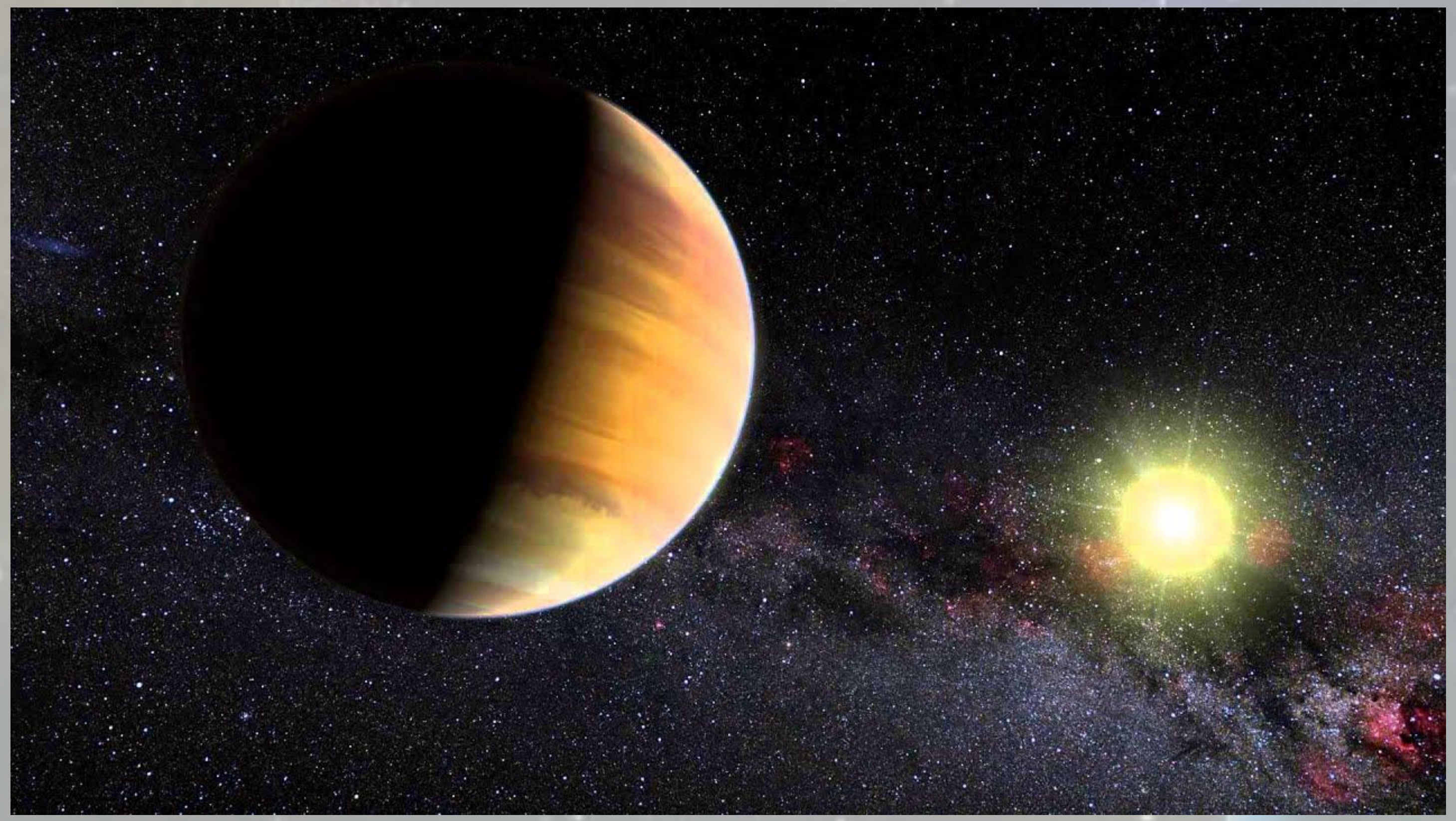
- The Radial Velocity method for finding exoplanets can be formulated as simple (weighted) linear regression using generalized Hermite-Gaussian functions.

- The Hermite-Gaussian Radial Velocity (HGRV) approach outperforms the traditional CCF approach in the simulation study and 51 Pegasi data considered.

Based on the paper *Holzer et al. (in review)* arxiv 2005.14083

Methodology implemented in:

- **R** package *rvmethod* (publicly available through the CRAN)
- **Python** available at *https://github.com/parkerholzer/hgrv_method*



References:

Pepe, F., et al. "The CORALIE survey for southern extra-solar planets VII-Two short-period Saturnian companions to HD 108147 and HD 168746." *Astronomy & Astrophysics* 388.2 (2002): 632-638.

Astudillo-Defru, N., et al. "The HARPS search for southern extra-solar planets-XXXVI. Planetary systems and stellar activity of the M dwarfs GJ 3293, GJ 3341, and GJ 3543." *Astronomy & Astrophysics* 575 (2015): A119.

Petersburg, Ryan R., et al. "An Extreme-precision Radial-velocity Pipeline: First Radial Velocities from EXPRES." *The Astronomical Journal* 159.5 (2020): 187.

Rajpaul, Vinesh, et al. "A Gaussian process framework for modelling stellar activity signals in radial velocity data." *Monthly Notices of the Royal Astronomical Society* 452.3 (2015): 2269-2291.

Holzer, Parker, et al. "A Hermite-Gaussian Based Radial Velocity Estimation Method." *arXiv preprint arXiv:2005.14083* (2020).

Davis, Allen, et al. “Insights on the Spectral Signatures of Stellar Activity and Planets from PCA" *arXiv:1708.00491* (2017).

References:

Cretignier, M., et al. "Measuring precise radial velocities on individual spectral lines-II. Dependance of stellar activity signal on line depth." *Astronomy & Astrophysics* 633 (2020): A76.

Dumusque, Xavier. "Measuring precise radial velocities on individual spectral lines-I. Validation of the method and application to mitigate stellar activity." *Astronomy & Astrophysics* 620 (2018): A47.

Jones, David E., et al. "Improving Exoplanet Detection Power: Multivariate Gaussian Process Models for Stellar Activity." *arXiv preprint arXiv:1711.01318* (2017).

Ning, Bo, et al. "Identifying Activity-sensitive Spectral Lines: A Bayesian Variable Selection Approach." *The Astronomical Journal* 158.5 (2019): 210.