

A Machine Learning Inspired Method Reveals the Mass of K2-167 b

Zoe L. de Beurs^{1,2}, Andrew Vanderburg^{1,2}, Christopher J. Shallue³, Joseph E. Rodriguez⁴, Sebastian Zieba^{5,10}, Annelies Mortier^{6,7}, Lars Buchhave⁸, Luca Malavolta⁹, HARPS-N Telescope Collaboration

1. Massachusetts Institute of Technology, Cambridge, MA
2. Department of Astronomy, University of Wisconsin-Madison, Madison, WI
3. Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA
4. Department of Physics and Astronomy, Michigan State University, East Lansing, MI
5. Max-Planck-Institut für Astronomie (MPIA), Heidelberg, Germany

6. Astrophysics Group, Cavendish Laboratory, Cambridge, UK
7. Kavli Institute for Cosmology, University of Cambridge, Cambridge, UK
8. DTU Space, National Space Institute, Technical University of Denmark, Lyngby, Denmark
9. Dipartimento di Fisica e Astronomia "Galileo Galilei", Università di Padova, Padova, Italy
10. Leiden Observatory, Leiden University, Leiden, The Netherlands



zdebeurs@mit.edu

Background

Discovering earth-mass exoplanets orbiting bright nearby stars requires precise radial velocity (RV) observations. Currently, we are limited by spurious RV signals introduced by the host star in the form of stellar activity (i.e. faculae, starspots). These stellar signals can mimic or hide exoplanet signals. Inspired by our successful implementation of a neural network stellar activity mitigation method on solar data (de Beurs et al. 2020), we designed a simplified machine learning (ML) method that can be applied to *extrasolar* stars. As a proof-of-concept, we tested this method on one star: K2-167.

K2-167 was first found to host a transiting super-Earth by the K2 mission (Vanderburg et al. 2016b; Mayo et al. 2018). Recently, the planetary signal was re-detected by TESS (Ikwa-Ukwa et al. 2020). K2-167 is an ideal test case for our method. It is the brightest star observed by K2 that hosts a validated planet, and it is a relatively active, yet slowly rotating F7. HARPS-N began observing K2-167 in late 2015, but stellar activity has prevented a confident detection of the planet's RV signal, even with the aid of state-of-the-art methods like GP regression.

Our strategy is to use a simplified ML method to identify and interpret the subtle changes to stellar spectra that are caused by stellar activity.

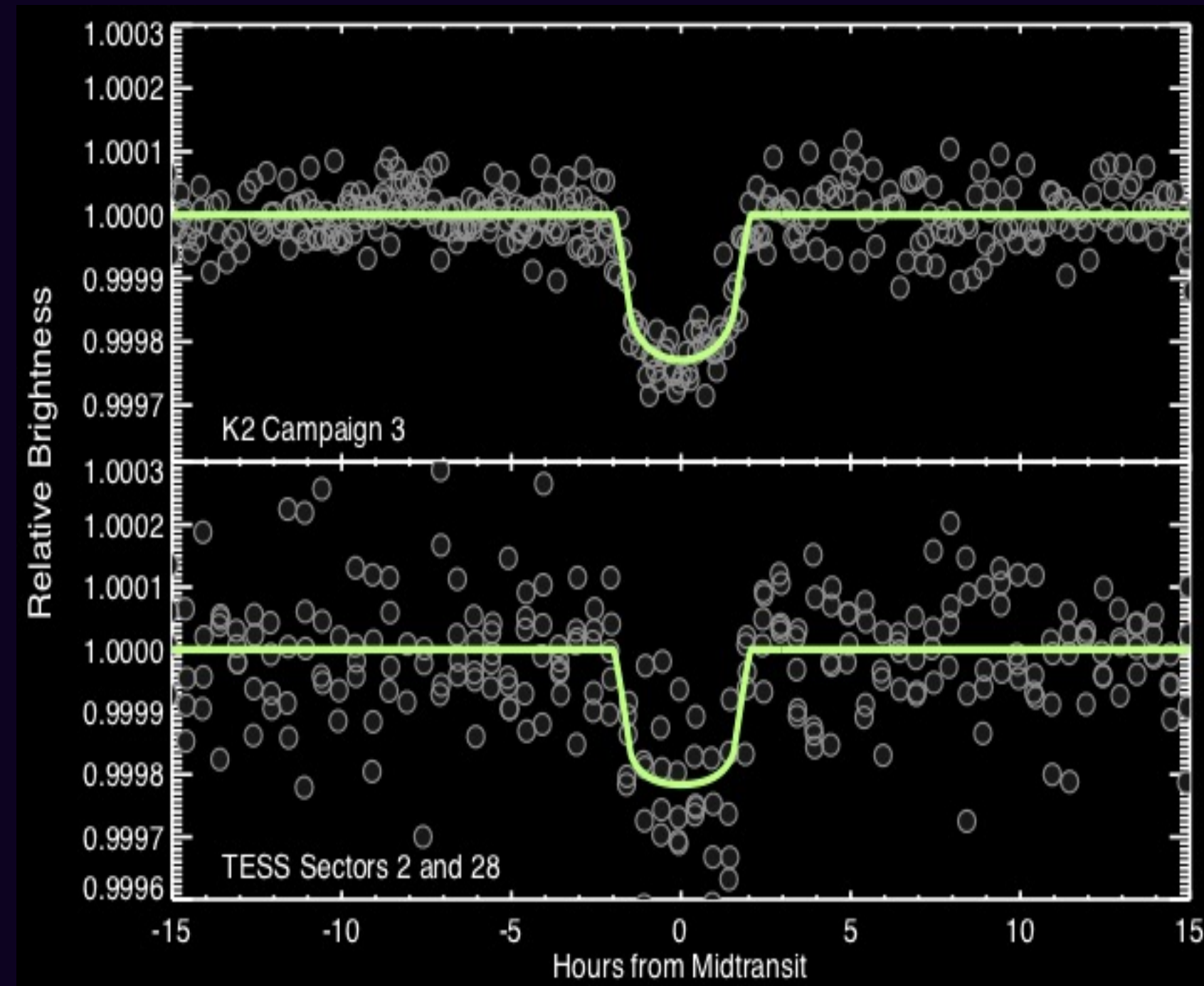


Figure 1: K2 (top) and TESS (bottom) lightcurves. The green lines show the best fitting transit models.

Data

K2 – K2-167 was observed during K2 Campaign 3 from 2014 November 17 to 2015 January 23. After extraction, the lightcurves were reprocessed by using the methods described in Vanderburg & Johnson (2014) and Vanderburg et al. (2016a). TESS – K2-167 was observed by Camera 1 during TESS Sector 2 from 2018 Aug22 to Sept 20. The lightcurves were generated using NASA's SPOC pipeline. HARPS-N – HARPS-N is a vacuum-enclosed cross-dispersed echelle spectrograph designed to achieve high-precision RVs (Cosentino et al. 2012). We collected 76 precise RV observations of K2-167 between Aug 2015 and Oct 2016.

CCF Input Representation – For our ML model, we have to pre-process our data products into a uniform format such that the model becomes sensitive to shape changes due to stellar activity, not translational shifts caused by Keplerian shifts. We describe this input representation in detail in de Beurs et al. 2020.

Simplified ML Method

To extend our previous ML method for solar data (de Beurs et al. 2020) to extrasolar observations, a new simplified method was required for two reasons:

1. RV datasets for extrasolar stars rarely have as many observations as the solar dataset from the HARPS-N Solar Telescope (~600 days of observations)
2. We lack a firm "ground-truth" for the stellar activity signals of extrasolar stars.

To overcome these issues, we first greatly simplified our ML model by using a linear network rather than a convolutional neural network. Second, we also simultaneously fit the stellar activity signals *and* the planetary signals since we cannot remove all the planets *a priori*. In detail, our model is given by:

$$RV = w_1 \cdot CCF_1 + w_2 \cdot CCF_2 + \dots + w_m \cdot CCF_m + b_1 \cdot Kepler_1 + b_2 \cdot Kepler_2 + \dots + b_n \cdot Kepler_n$$
$$= RV_{activity} + RV_{Keplerian}$$

High precision Spectra

Input: Δ Cross-Correlation Function

Linear Regression Model

MCMC fit: disentangled RV signals:
(i) Stellar activity RV.
(ii) Planet signal(s)

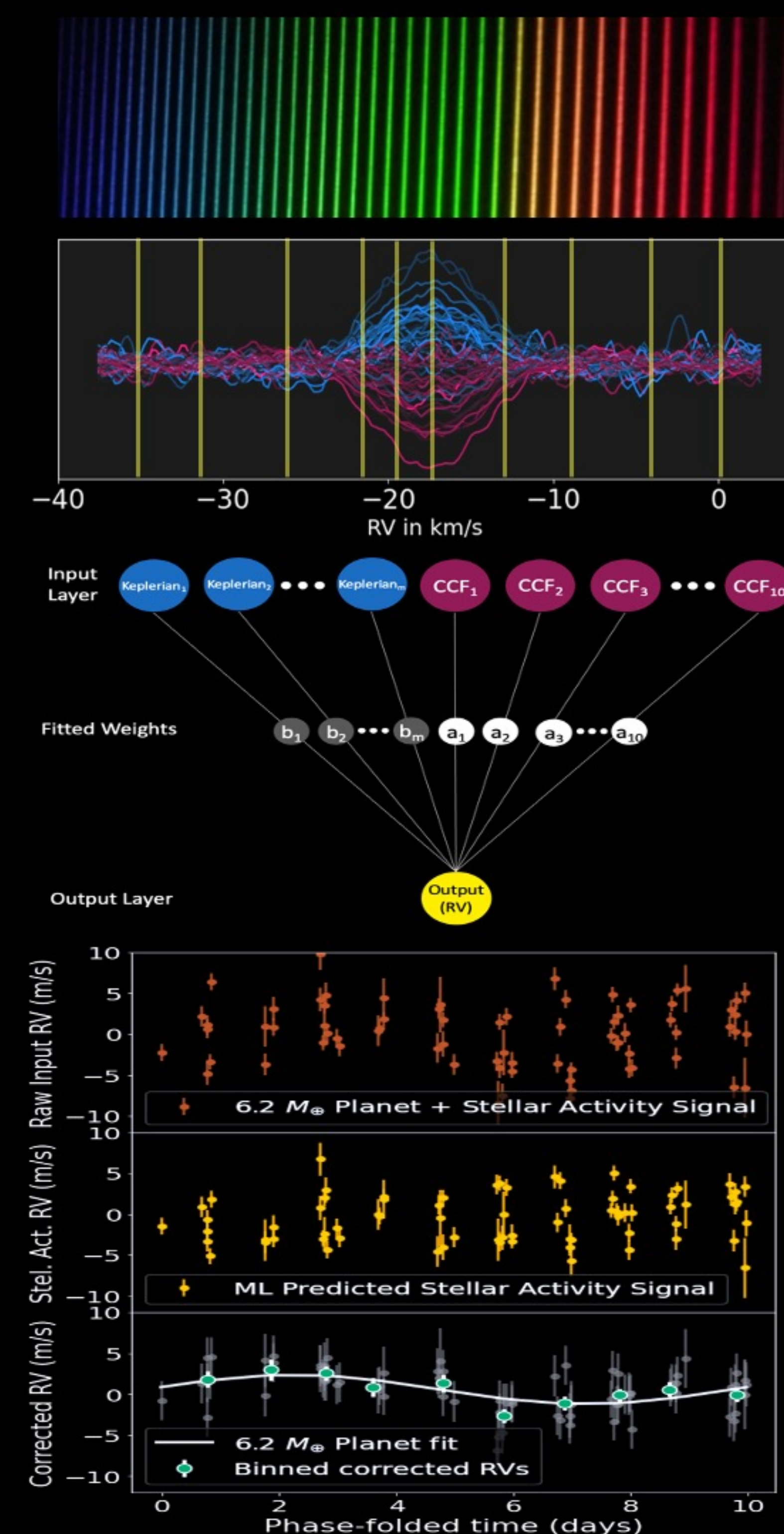


Figure 2: Schematic of our simplified ML model for extrasolar stars. We use the changing shape of spectral features (measured by the change in the cross-correlation function or CCF) to predict and remove stellar activity signals. We use a linear regression model with only about 10 points from the CCF residuals used as inputs. We also simultaneously fitting Keplerians along with the activity corrections to predict the overall RV signal.

Results

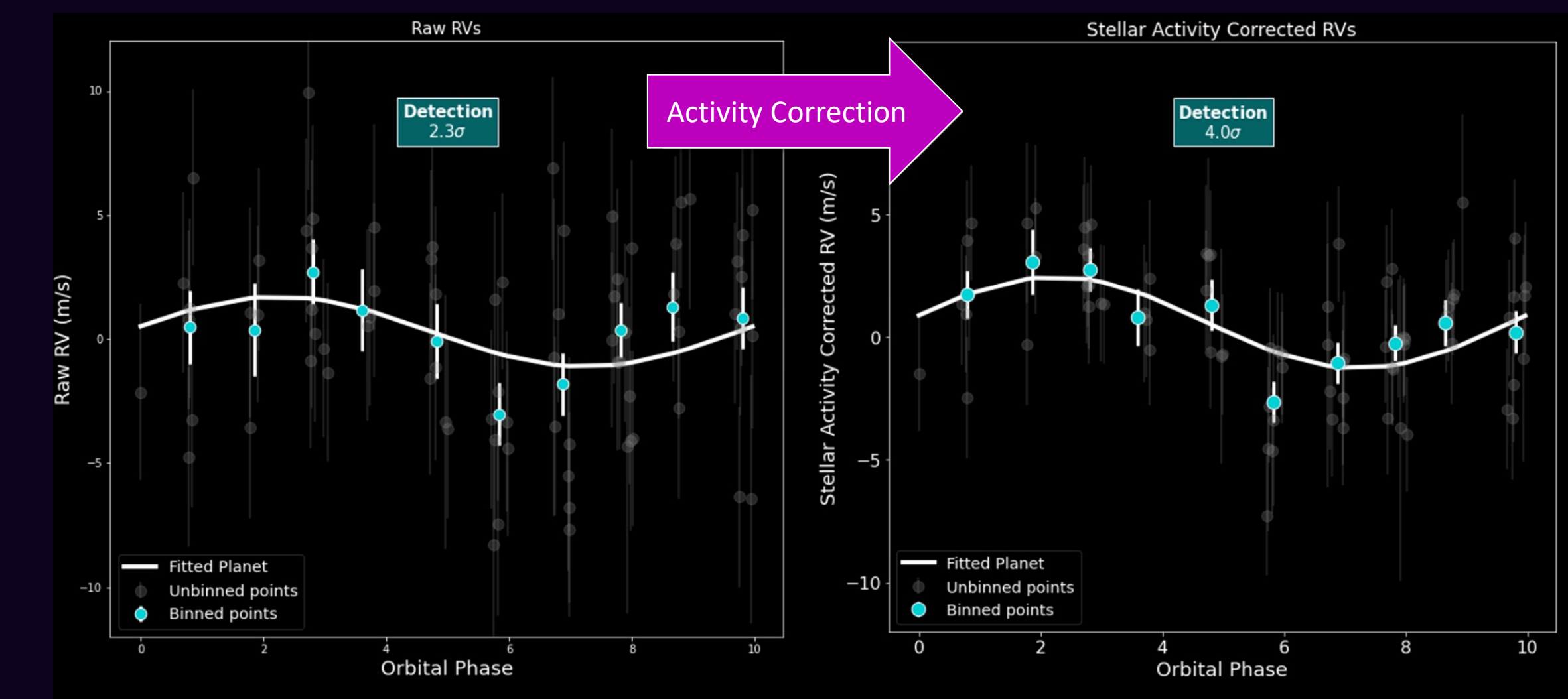


Figure 3: Phase-folded RVs before (a) and after (b) applying our activity correction model. Before our activity correction, we detect the planet with only 2.3σ confidence. (b) After applying our model, the detection significance increases to 4.0σ .

- We find that our simplified approach to stellar activity corrections yields a significant improvement in both the RV scatter (by a factor of ~ 2) and the precision with which we recover the mass of the transiting planet (improving a 2.3σ significance detection into a 4.0σ significance detection; Figure 3). Our MCMC fit found that the mass of $6.12^{+1.61}_{-1.56} M_E$, the orbital period and phase of the detection agrees with the known 9.97857 day period from transit observations
- We investigated the behavior of the raw and stellar activity corrected RVs in Fourier domain to see which signals are being removed to achieve this reduction in scatter (Fig. 4).

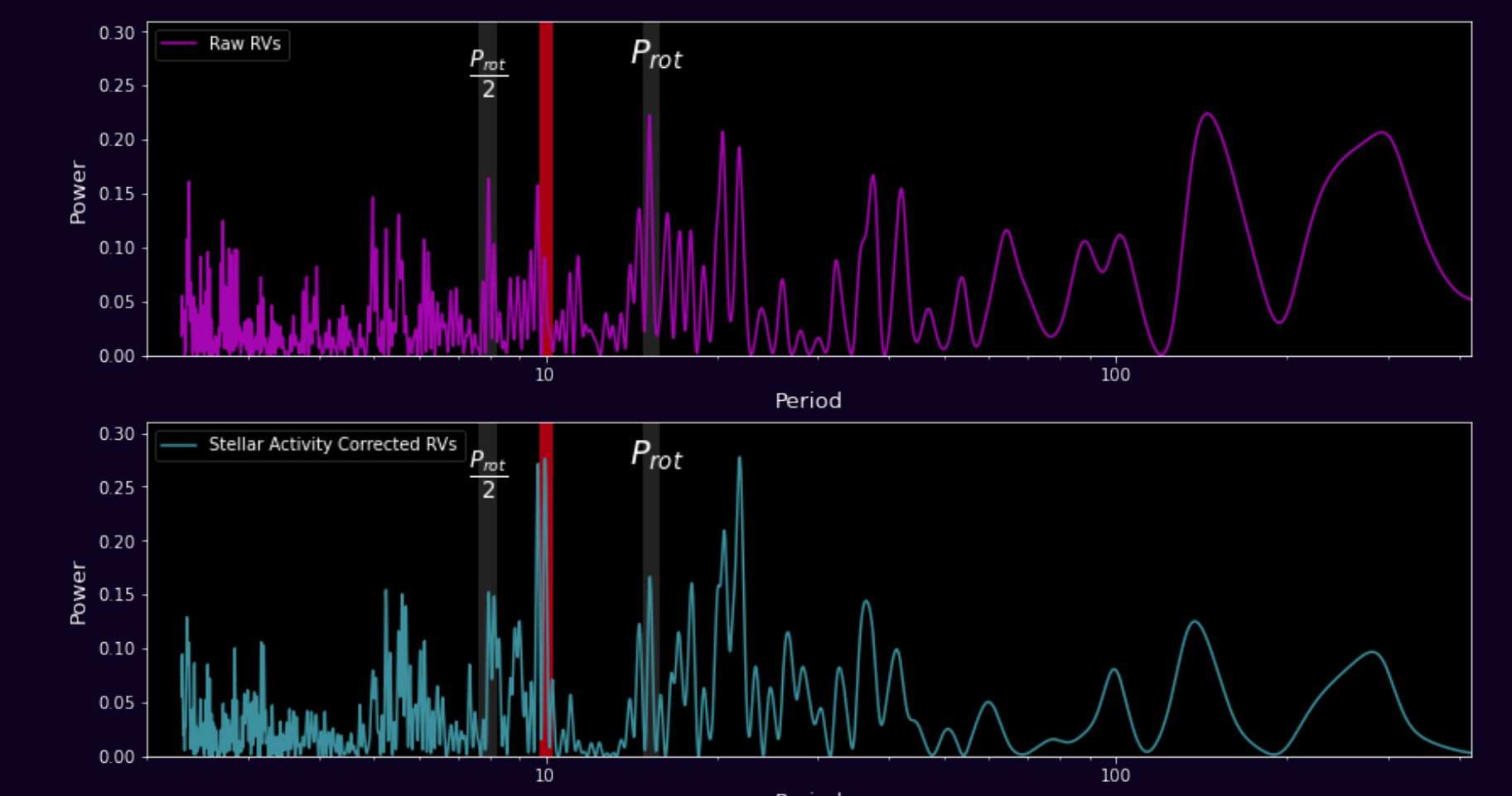


Figure 4: Periodogram: HARPS-N Raw (a) and Corrected (b) RVs in Fourier space. The peaks in the top panel that correspond to stellar activity signals decrease in magnitude in the bottom panel after applying our method and a planet signal emerges at 9.97857 days.

Future Directions

- In the future, these or similar techniques could be widely applied to solar-type (FGK) stars, help measure masses of planets from TESS to fulfill the level 1 science requirement, and eventually help detect habitable-zone Earth-mass exoplanets.
- We plan to further validate our method with HARPS-N Rocky Planet Targets and young planets orbiting young active stars

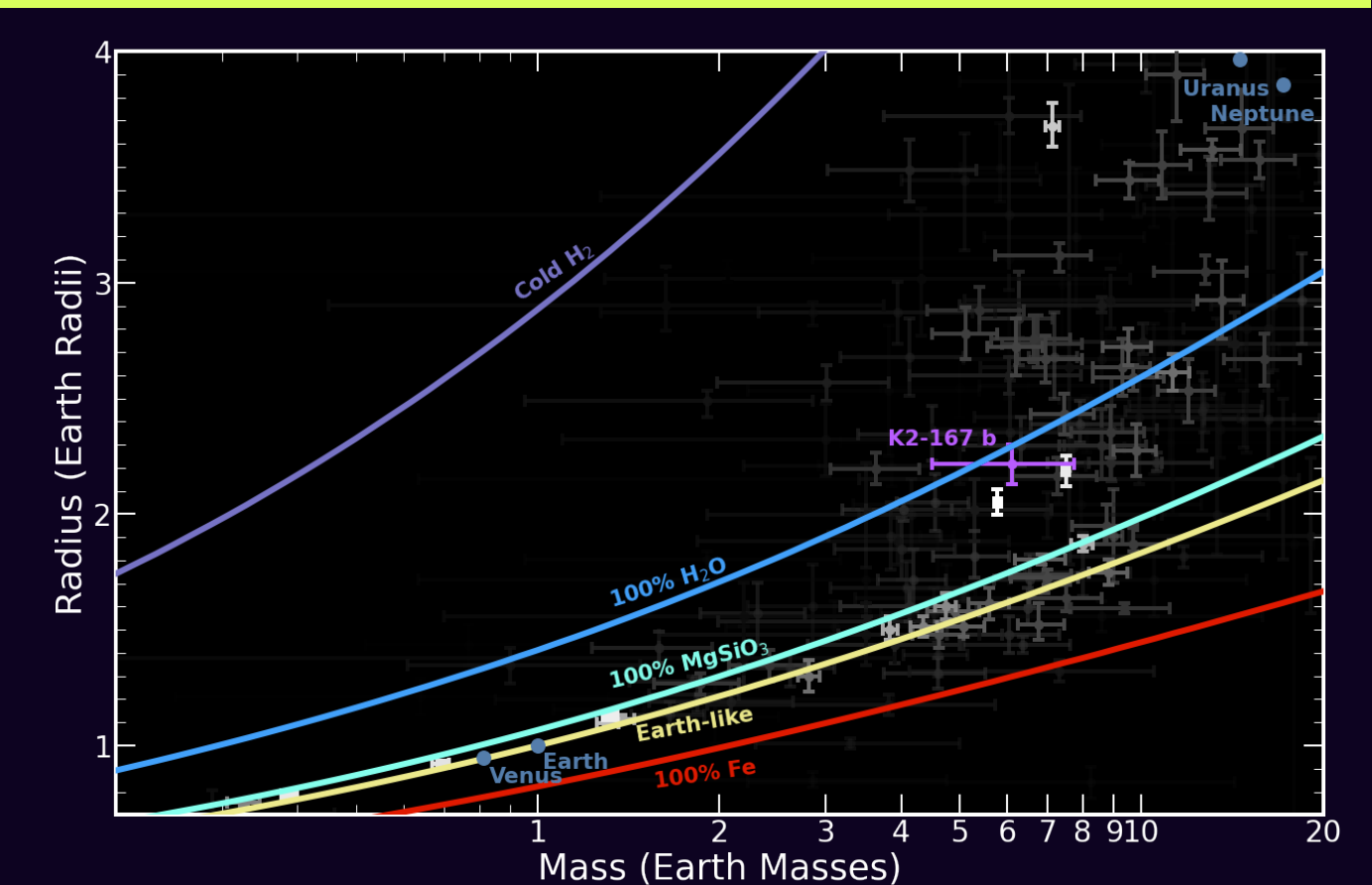


Figure 5: The mass/radius diagram for small exoplanets. K2-167b is plotted in purple.

Acknowledgments

We thank the Office of Undergraduate Research, TIDES Advanced Research Fellowship, Dean's Scholars, and the Junior Fellows Honors Program for their generous support.

References

Dumusque et al. 2014; de Beurs et al. 2020; Haywood PhD Thesis 2015; Vanderburg & Johnson 2014; Vanderburg et al. 2016a. Vanderburg et al. 2016b; Mayo et al. 2018; Shallue & Vanderburg, AJ 155, 94, 2018; Cosentino et al. 2012; Ikwa-Ukwa et al. 2020; Akeson et al. 2013