The Astronomical Journal, 163:260 (15pp), 2022 June https://doi.org/10.3847/1538-3881/ac609a

Into the Depths: A New Activity Metric for High-precision Radial Velocity Measurements Based on Line Depth Variations

Jared C. Siegel[1], Ryan A. Rubenzahl[2,4], Samuel Halverson[3], and Andrew W. Howard[2]
[1] Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA; siegeljc@uchicago.edu
[2] Department of Astronomy, California Institute of Technology, Pasadena, CA 91125, USA
[3] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
Received 2021 November 19; revised 2022 March 8; accepted 2022 March 22; published 2022 May 11

Abstract

The discovery and characterization of extrasolar planets using radial velocity (RV) measurements is limited by noise sources from the surfaces of host stars. Current techniques to suppress stellar magnetic activity rely on decorrelation using an activity indicator (e.g., strength of the Ca II lines, width of the cross-correlation function, broadband photometry) or measurement of the RVs using only a subset of spectral lines that have been shown to be insensitive to activity. Here, we combine the above techniques by constructing a high-signal-to-noise activity indicator, the depth metric $\mathcal{D}(t)$, from the most activity-sensitive spectral lines using the "line-by-line" method of Dumusque (2018). Analogous to photometric decorrelation of RVs or Gaussian progress regression modeling of activity indices, time series modeling of $\mathcal{D}(t)$ reduces the amplitude of magnetic activity in RV measurements; in an αCenB RV time series from HARPS, the RV rms was reduced from 2.67 to 1.02 m s^{-1}. $\mathcal{D}(t)$ modeling enabled us to characterize injected planetary signals as small as 1 m s^{-1}. In terms of noise reduction and injected signal recovery, $\mathcal{D}(t)$ modeling outperforms activity mitigation via the selection of activity-insensitive spectral lines. For Sun-like stars with activity signals on the m s^{-1} level, the depth metric independently tracks rotationally modulated and multiyear stellar activity with a level of quality similar to that of the FWHM of the CCF and $\log R'_{\rm HK}$. The depth metric and its elaborations will be a powerful tool in the mitigation of stellar magnetic activity, particularly as a means of connecting stellar activity to physical processes within host stars.

Unified Astronomy Thesaurus concepts: Exoplanet detection methods (489); Radial velocity (1332)

1. Introduction

The noise floor for the radial velocity (RV) detection of extrasolar planets is set by a combination of three sources: photon-limited uncertainties ($\sigma_{\rm photon}$), systematic errors in the instrument ($\sigma_{\rm sys}$), and stellar activity ($\sigma_{\rm activity}$). Unlike $\sigma_{\rm photon}$, which depends on the signal-to-noise ratio of the spectrum, $\sigma_{\rm sys}$ and $\sigma_{\rm activity}$ both can have structured temporal dependence. For an overview of the various physical processes that give rise to RV noise, we refer the interested reader to the introduction section of de Beurs et al. (2020). With improved resolution, calibration, and stabilization, next-generation instruments, such as the Keck Planet Finder (KPF; Gibson et al. 2016), ESPRESSO (Mégevand et al. 2014), EXPRES (Jurgenson et al. 2016), and NEID (Schwab et al. 2016), are significantly lowering the $\sigma_{\rm sys}$ noise floor; KPF, a high-precision Doppler spectrometer scheduled for first light in the summer of 2022 on the Keck I telescope, is expected to achieve $\sigma_{\rm sys} = 30$ cm s^{-1} and will achieve $\sigma_{\rm photon} = 30$ cm s^{-1} on a V=10.9 G2 star in only 30 min. While these advancements are significant, stellar activity on the m s^{-1} level dominates over these small errors and limits the RV detections of small planets. We investigated novel stellar activity mitigation methods, with the goal of lowering the $\sigma_{\rm activity}$ noise floor in anticipation of KPF.

RV studies typically employ one of three methods of RV measurement: cross-correlation using a synthetic template (Baranne et al. 1996), maximum likelihood estimation to a reference spectrum (Anglada-Escudé & Butler 2012), or forward modeling (Butler et al. 1996). However, some spectral lines are more sensitive to systematic instrumental errors (Dumusque et al. 2015) and others are more sensitive to stellar magnetic activity (Thompson et al. 2017; Dumusque 2018; Wise et al. 2018; Cretignier et al. 2020). These three methods introduce noise into the derived RVs by treating all lines equally.

Stellar activity and instrumental noise are often statistically accounted for through the addition of a time-independent white-noise jitter term to the RV uncertainties (e.g., Wright 2005), or detrended with an activity model, such as linear decorrelation and Gaussian process (GP) regression (Haywood et al. 2014; Rajpaul et al. 2015; Dai et al. 2017; Ahrer et al. 2021). Activity models, whether parametric or nonparametric, commonly rely on a handful of spectral lines, such as the Ca II H&K lines ($\log R'_{HK}$) and Hα, or diagnostics from the cross-correlation function (CCF), such as the full width at half maximum (FWHM) and the bisector inverse slope (BIS), to track stellar activity (Saar et al. 1998; Kürster et al. 2003; Queloz et al. 2009). Alternatively, Aigrain et al. (2012) used a spot model to track stellar activity via the star's photometric light curve. This method, known as FF′, has more recently been modified to use activity indicators in place of photometric flux (Giguere et al. 2016).

Dumusque (2018) developed a novel approach to mitigating stellar noise by considering the Doppler shift of each spectral line independently; for HARPS spectra, "line-by-line" RV measurement was shown to reach the same precision as standard techniques. Dumusque (2018) also introduced the "line selection method," in which only spectral lines shown to be activity-insensitive are considered for the RV measurement; the line selection method achieves a factor of ∼1.5 reduction in

[4] NSF Graduate Research Fellow.

the αCenB stellar noise. In a follow-up study, Cretignier et al. (2020) found that the effect of stellar activity on a given line is inversely proportional to the line's depth (a proxy for physical depth inside the stellar atmosphere). Cretignier et al. (2020) then demonstrated that using the difference between the RVs of deep and shallow spectral lines as an activity proxy (hereafter referred to as the "formation depth method") yields better than a factor-of-two reduction in αCenB's stellar activity amplitude.

In this paper, we develop a combination of the above line-by-line and detrending techniques into the depth metric method. Prior studies have demonstrated that, for a Sun-like star, more than 1000 spectral lines are highly correlated with stellar activity (Dumusque 2018; Cretignier et al. 2020). We combine a subpopulation of spectral lines shown to be activity-sensitive into a high-signal-to-noise activity metric. Motivated by Wise et al. (2018), we are particularly interested in the relationship between observation-to-observation variations in line depth and stellar activity; variable line depth has the benefit of being available at all wavelengths, is computationally inexpensive, is translation-invariant, and is a largely unexplored parameter space. Our depth metric can be modeled using time series activity mitigation techniques. The co-adding of orders of magnitude more lines than traditional indices has the potential to greatly improve modeling results (Giguere et al. 2016). To demonstrate the diagnostic power of the depth metric, we conducted a comparative study between linear decorrelation, FF′, and GP models based on either the depth metric, $\log R'_{HK}$, or the FWHM of the CCF; we also considered the line selection and formation depth methods.

This paper is organized as follows. In Section 2, we describe the HARPS observations used in our study. Section 3 summarizes our implementation of the line-by-line analysis technique of Dumusque (2018) and develops our activity index, the depth metric. Then in Section 4, we describe the activity mitigation methods considered in our study, including the line selection method of Dumusque (2018), the formation depth method of Cretignier et al. (2020), our modified FF′ model, and our GP models. We applied these activity models to HARPS observations of αCenB and HD 13808 in Section 5, and summarize our conclusions in Section 6.

2. Observations

We considered two stars observed with the HARPS spectrograph: αCenB (HD 128621) and HD 13808. With high observation cadence and significant activity signals, these targets offer a valuable test bed for stellar activity mitigation methods. αCenB presents a clear rotationally modulated stellar activity signal and lacks a significant planetary signal, although several candidates have been proposed (Dumusque et al. 2012; Demory et al. 2015). Between 2008 February and 2011 July, αCenB was observed with HARPS between 1 and 187 times a night (median of 35). This campaign was designed to capture both long- and short-term stellar activity variations in αCenB's RVs (Dumusque et al. 2011). As part of a sample of bright stars with relatively low levels of RV scatter, HD 13808 has been observed with HARPS since 2003 using exposure times greater than 15 minutes (Ahrer et al. 2021). HD 13808 exhibits a multiyear stellar magnetic activity cycle and hosts two recently confirmed Neptune-mass planets, allowing for a test of our mitigation methods' effects on planet characterization.

We correct each observation for known systematics following the prescriptions in Dumusque (2018) and Cretignier et al. (2020). Here, we briefly describe the corrections. We first divided the raw stellar spectrum by the blaze of the instrument. The resulting pixel fluxes were then continuum normalized via a rolling maximum, with a 10 Å window size. To account for spectra with outlier continua, for each order index between 30 and 71 we fit the continuum (in the cross-dispersion direction) with a linear function. For a given target, we then rejected all spectra that had a single order's slope $>3\sigma$ from the mean slope of that order. Each spectrum's wavelength solution was reconstructed from the 288 polynomial coefficients stored in the spectrum's header. The wavelength solution was then Doppler shifted to correct for the barycentric Earth RV, which we adopted from the HARPS data reduction software (DRS), and was also Doppler shifted in accordance with the DRS-reported instrumental drift.

For αCenB, we used the 2010 HARPS observations, over which the canonical activity indices ($\log R'_{\rm HK}$, Hα, FWHM, and BIS) all present quasi-sinusoidal oscillations with periods near the $\sim$37 day rotation period (DeWarf et al. 2010). After rejecting all observations with an airmass greater than 1.5, we fit the HARPS reported RVs (stored in the header of each spectrum) using a quadratic polynomial plus a sinusoid as a function of time; the sinusoid component of the RV model tracks stellar activity, while the quadratic component accounts for long-term variations induced by αCenA. We then performed 3σ clipping on the DRS RVs with respect to our polynomial–sinusoid model for five iterations. The filtering process retained 908/1014 observations spanning 80 days.

Following Cretignier et al. (2020), we corrected for the long-term RV signal induced by αCenB's binary companion by subtracting the quadratic component of the above model from the RVs. Via a least-squares regression, we found a sinusoid period of 35.9 days, nearly identical to that reported by Cretignier et al. (2020). Unless otherwise stated, these corrected RVs are the ones used in our analyses.

For HD 13808, we considered all 11 yr of HARPS observations prior to the fiber change on the 1st of June 2015. After filtering for outliers in the HARPS DRS RVs (5σ clipping for two iterations), filtering for continuum outliers, and filtering observations with an airmass greater than 1.5, we had 235/276 observations.

3. Line-by-line Analysis

As developed by Dumusque (2018), the line-by-line method measures the RV of each spectral line independently using template matching. By taking the weighted average over all lines, the precise RV measurement for the entire spectrum can be obtained. For a given star, the line-by-line method can be broken down into three primary steps: first, construct a reference spectrum; second, generate a spectral line list from the reference spectrum or adopt a line list from a synthetic template; third, conduct template matching for each spectral line in each spectrum. Our implementation of the Dumusque (2018) line-by-line method will be released for public use in J. Siegel et al., 2022, in preparation.

We generated the reference spectrum for a given star by co-adding the systematics-corrected spectra on a common wavelength solution; the common wavelength solution is taken to be that of the spectrum with the lowest barycentric Earth RV, and linear interpolation was used to evaluate all spectra on this common solution.

Next, we generated a catalog of absorption lines following the methods detailed in Appendix A of Cretignier et al. (2020).

Using a given star's reference spectrum, absorption lines were identified in terms of local minima and local maxima. For two neighboring local maxima at $\lambda_{i,\mathrm{left}}$ and $\lambda_{i,\mathrm{right}}$, with a local minima between them at λ_i, we treated λ_i as the line's center and adopted a window width of $\min(\lambda_i - \lambda_{i,\mathrm{left}}, \lambda_{i,\mathrm{right}} - \lambda_i)$. We then rejected all lines where $\lambda_{i,\mathrm{left}}$ and $\lambda_{i,\mathrm{right}}$ were within 10 pixels of each other and demanded that the line's depth (defined below in Equation (4)) be greater than 0.05. For αCenB and HD 13808, we identified a total of ∼7500 and ∼6500 lines, respectively, which reduced to ∼5500 and ∼5000 when overlap of spectral orders was taken into account. For lines found in multiple spectral orders, we considered the spectral order for which the line had the lowest mean RV uncertainty, $\langle\sigma_{\mathrm{RV}_{i,j}}\rangle$. Given the similar spectral types of αCenB and HD 13808, we attribute the difference in the number of detected lines for each star to the higher average signal-to-noise ratio of αCenB's spectra; for order 50, αCenB has an average signal-to-noise ratio of 320, while HD 13808 has an average of 110. Consistent with this explanation, we found more shallow spectral lines for αCenB than HD 13808; for αCenB, we identified ∼200 spectral lines with $d_{i,\mathrm{reference}} < 0.1$, while for HD 13808, we identified only ∼100 such spectral lines.

The radial velocity $\mathrm{RV}_{i,j}$ was measured by numerically solving Equation (2) of Bouchy et al. (2001). For the ith line of the jth spectrum,

$$S_{i,j}(\lambda) = A\left[S_{i,\mathrm{reference}}(\lambda) + \frac{\partial S_{i,\mathrm{reference}}(\lambda)}{\partial\lambda}\delta\lambda\right], \quad (1)$$

where $S_{i,j}$ are the continuum normalized pixel flux values within the line window, and $S_{i,\mathrm{reference}}$ are the reference spectrum flux values over the same interval. The reference spectrum was evaluated on the wavelength solution of the jth spectrum via cubic spline interpolation. $\mathrm{RV}_{i,j}$ was then given by

$$\mathrm{RV}_{i,j} = \frac{c}{\lambda}\delta\lambda. \quad (2)$$

Uncertainty in $\mathrm{RV}_{i,j}$ was inferred by propagating the uncertainties in $A\delta\lambda$ and A from the least-squares fit,

$$\sigma_{\mathrm{RV}_{i,j}} = \frac{c}{\lambda}\delta\lambda\sqrt{\left(\frac{\sigma_{A\delta\lambda}}{A\delta\lambda}\right)^2 + \left(\frac{\sigma_A}{A}\right)^2}. \quad (3)$$

We also extracted the depth of each spectral line, defined as

$$d_{i,j} = 1 - \frac{\min(S_{i,j})}{C_{i,j}}, \quad (4)$$

where $C_{i,j}$ is the continuum flux, defined as the average of the leftmost and rightmost pixel fluxes in the line's window; Equation (4) describes the depth of a spectral line relative to the local continuum, which is equivalent to the depth relative to the continuum spectrum for isolated lines. For high-signal-to-noise lines, greater precision may be achieved by inferring the minimum flux by cubic spline interpolation; however, we found native pixel sampling yields results nearly indistinguishable from those of cubic spline interpolation and avoids potential instabilities for low-signal lines. Finally, we defined the normalized depth parameter,

$$\tilde{d}_{i,j} = d_{i,j}/d_{i,\mathrm{reference}}, \quad (5)$$

where $d_{i,\mathrm{reference}}$ is the line's depth in the reference spectrum (effectively the line's mean depth). Uncertainty in the depth parameter was calculated via propagation of the flux uncertainties.

Following our initial line-by-line analysis, we proceeded to clean the data set of outlier measurements. Similar to the sigma-clipping procedures of Dumusque (2018), for each observation we performed a 4σ clipping procedure (with two iterations) on $\mathrm{RV}_{i,j}$, $\sigma_{\mathrm{RV}_{i,j}}$, $\chi^2_{i,j}$ of the $A\delta\lambda$ and A fit, $\tilde{d}_{i,j}$, and $\sigma_{\tilde{d}_{i,j}}$. We then rejected all lines for which more than 1% of the measurements were rejected by the above filters, and we removed all lines within 48 pixels of a telluric line. Telluric lines were identified using a sample `TAPAS` model spectrum (Bertaux et al. 2014) for Keck Observatory (30° zenith). Spectral regions in the vicinity of TAPAS features with a normalized contamination depth of 1% or greater were rejected; we found that adopting a 0.1% depth threshold or filtering all lines in our line list with a central wavelength >5000 Å did not yield RVs significantly different from our baseline 1% threshold. Following Table C.1 of Cretignier et al. (2020), we also rejected asymmetric absorption lines, because such lines have been shown to be anticorrelated with stellar activity. After applying these filters, we retained ∼1500 and ∼1100 lines for αCenB and HD 13808, respectively. Greater than 400 lines appear in both star's line lists.

Using the cleaned data set, we next combined the line-by-line information into bulk properties. For each observation, we conducted a weighted average to derive the cumulative RV:

$$\mathrm{RV}_j = \sum_i \frac{\mathrm{RV}_{i,j}}{\sigma^2_{\mathrm{RV}_{i,j}}} \Big/ \sum_i \frac{1}{\sigma^2_{\mathrm{RV}_{i,j}}}. \quad (6)$$

3.1. The Depth Metric

Wise et al. (2018) demonstrated that, for a subpopulation of visually identified spectral lines, there is a strong correlation between line depth and stellar activity as traced by the Ca II H&K index. Here, we conducted a data-driven search for such spectral lines. For a given stellar target, we calculated the Pearson correlation coefficient $\mathcal{R}$ between the bulk RV, RV_j, and line depth $d_{i,j}$. As seen in Figure 1, both αCenB and HD 13808 possess substantial subpopulations of activity-sensitive lines. We used RV_j to calculate each line's activity sensitivity, because standard activity indices (e.g., $\log R'_{\mathrm{HK}}$ and FWHM of the CCF) have been shown to lag in time relative to the RV signal (Collier Cameron et al. 2019). However, planetary signals dilute the correlation between depth and the activity contribution to measured RVs, potentially limiting the scope of this approach; we discuss methods of selecting activity-sensitive lines for stars with strong planetary signals in Section 6.

To reflect our uncertainty in $d_{i,j}$ and RV_j, we conducted 1000 Monte Carlo trials for each line. For a given trial, we drew an independent sample of depth and RV values, assuming Gaussian uncertainties. A given line's Pearson correlation coefficient $\mathcal{R}$ was taken as the average over all 1000 trials.

For αCenB and HD 13808, the subpopulation of activity-sensitive lines was defined as all lines satisfying $\mathcal{R} < 5$ th percentile of each star's $\mathcal{R}$ distribution, i.e., only the most activity-sensitive lines were used. As shown in Figure 2, these adopted thresholds maximize each star's stellar activity signal, while still maintaining a significant population of spectral lines; >60 lines for αCenB and >50 lines for HD 13808.

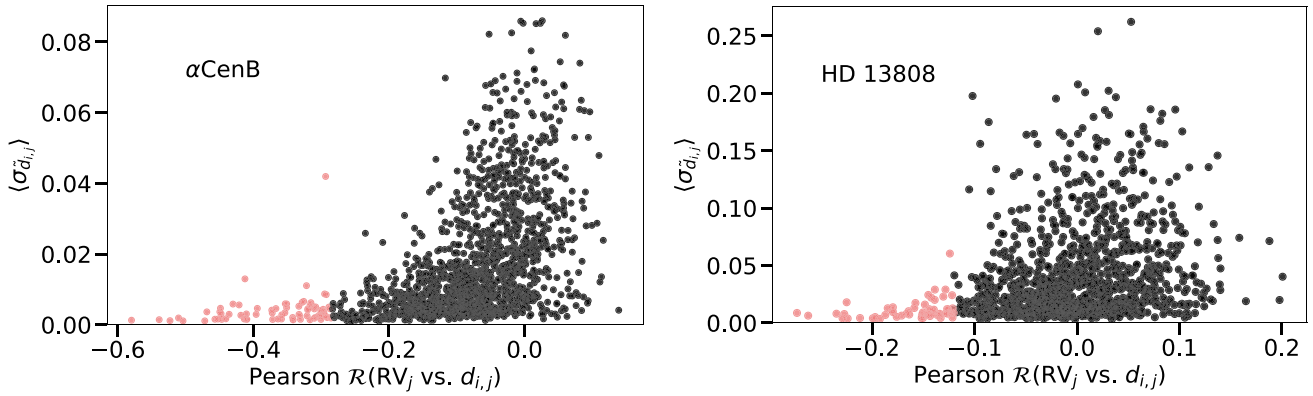

Figure 1. Both αCenB (left) and HD 13808 (right) possess clear subpopulations of activity-sensitive absorption lines (lines with $\mathcal{R}(\mathrm{RV}_j$ vs. $d_{i,j}) <$ 5th percentile of each star's $\mathcal{R}(\mathrm{RV}_j$ vs. $d_{i,j})$ distribution are highlighted in red). For each star, we present the Pearson correlation coefficient $\mathcal{R}$ between RV_j and line depth $d_{i,j}$ against each line's mean uncertainty in scaled depth.

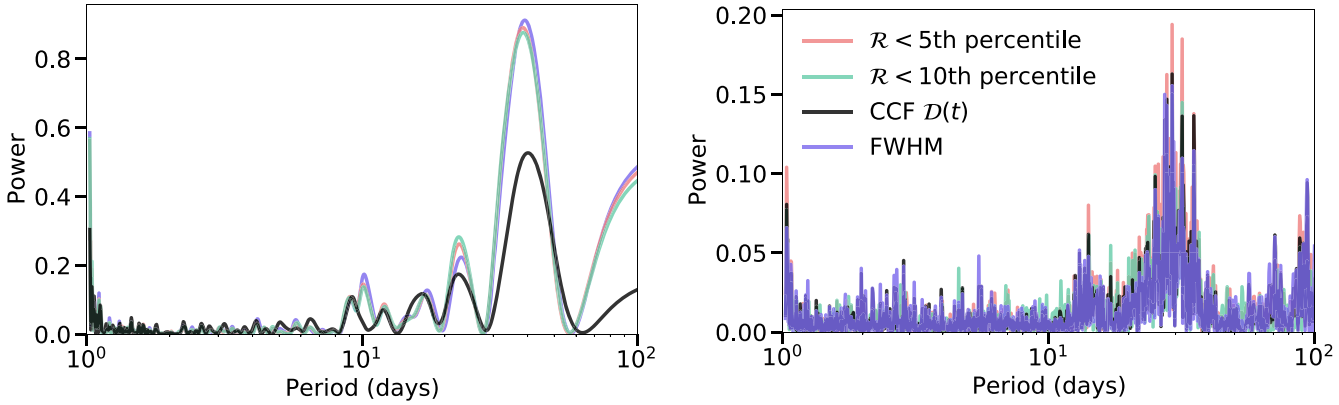

Figure 2. For both αCenB and HD 13808, selecting a subpopulation of activity-sensitive absorption lines better isolates the star's stellar activity signal. We present Lomb–Scargle periodograms of the depth metric $\mathcal{D}$(t) for αCenB (left) and HD 13808 (right) for two different spectral line subpopulations—$\mathcal{R}(\mathrm{RV}_j$ vs. $d_{i,j}) <$ 5th and 10th percentile of each star's $\mathcal{R}(\mathrm{RV}_j$ vs. $d_{i,j})$ distribution—as well as each star's CCF-based depth metric CCF $\mathcal{D}(t)$ and the FWHM of the CCF.

Although line depth variations are reflected in each observation's computed CCF, which includes thousands of lines, we clearly see in Figure 2 that the signal is greatly amplified by selecting an activity-sensitive subpopulation. The CCF-based depth metric (CCF $\mathcal{D}(t)$) was obtained by fitting each HARPS DRS CCF to a Gaussian model with a constant continuum and feeding the resulting depth time series into Equation (9) (described below); inferring the depth of the CCF with cubic spline interpolation, instead of the Gaussian model, yielded negligible differences in CCF $\mathcal{D}(t)$ for both αCenB and HD 13808.

In Figure 3, we present the distributions of reference spectrum line depth $d_{i,\mathrm{reference}}$, central wavelength λ_i, and symmetry score for the activity-sensitive lines alongside the general population. For a given line, the symmetry score is defined as the absolute difference between the leftmost and the rightmost pixel fluxes in the line's window, divided by the mean of the leftmost and the rightmost pixel fluxes; a highly symmetric line has a symmetry score near zero. For both αCenB and HD 13808, the activity-sensitive population is comprised mostly of deeper lines at longer wavelengths, with a slight preference for symmetric lines. Since $\mathcal{R}$ is highly dependent on $\langle\sigma_{\tilde{d}_{i,j}}\rangle$ (see Figure 1), we also isolated the subpopulation of lines with low $\langle\sigma_{\tilde{d}_{i,j}}\rangle$ for each star. Although the preference for deep lines is consistent with the preference for small uncertainties in depth, the low $\langle\sigma_{\tilde{d}_{i,j}}\rangle$ populations do not significantly differ from the general population with regard to central wavelength or symmetry. This preference for longer wavelengths may reflect the higher density of lines at shorter wavelengths (see Figure 4); line blending makes it difficult to measure a given line's activity signal.

Since a single spectral line possesses only a small amount of information, we combined the activity-sensitive lines into the depth metric. Adapting the formalism of Aigrain et al. (2012), for a given star we calculated each spectrum's weighted average line depth over our selected subpopulation and rescaled the resulting time series to introduce the depth metric $\mathcal{D}(t)$:

$$\tilde{d}(t) = \sum_i \frac{\tilde{d}_{i,j}}{\sigma^2_{\tilde{d}_{i,j}}} \Bigg/ \sum_i \frac{1}{\sigma^2_{\tilde{d}_{i,j}}}, \tag{7}$$

$$\tilde{d}_0 = \tilde{d}_{\max} + \sigma_{\tilde{d}}, \tag{8}$$

$$\mathcal{D}(t) = 1 - \frac{\tilde{d}(t)}{\tilde{d}_0}, \tag{9}$$

where $\tilde{d}_{\max}$ is the maximum depth metric of the $\tilde{d}(t)$ time series and $\sigma_{\tilde{d}}$ is the standard deviation of the time series. As seen in Figure 5, the depth metric captures the stellar activity cycles of αCenB and HD 13808. For comparison, we also present the $\log R'_{\mathrm{HK}}$ and CCF FWHM activity indices for each star; $\log R'_{\mathrm{HK}}$ was obtained by first calculating the $\mathcal{S}_{\mathrm{index}}$ via

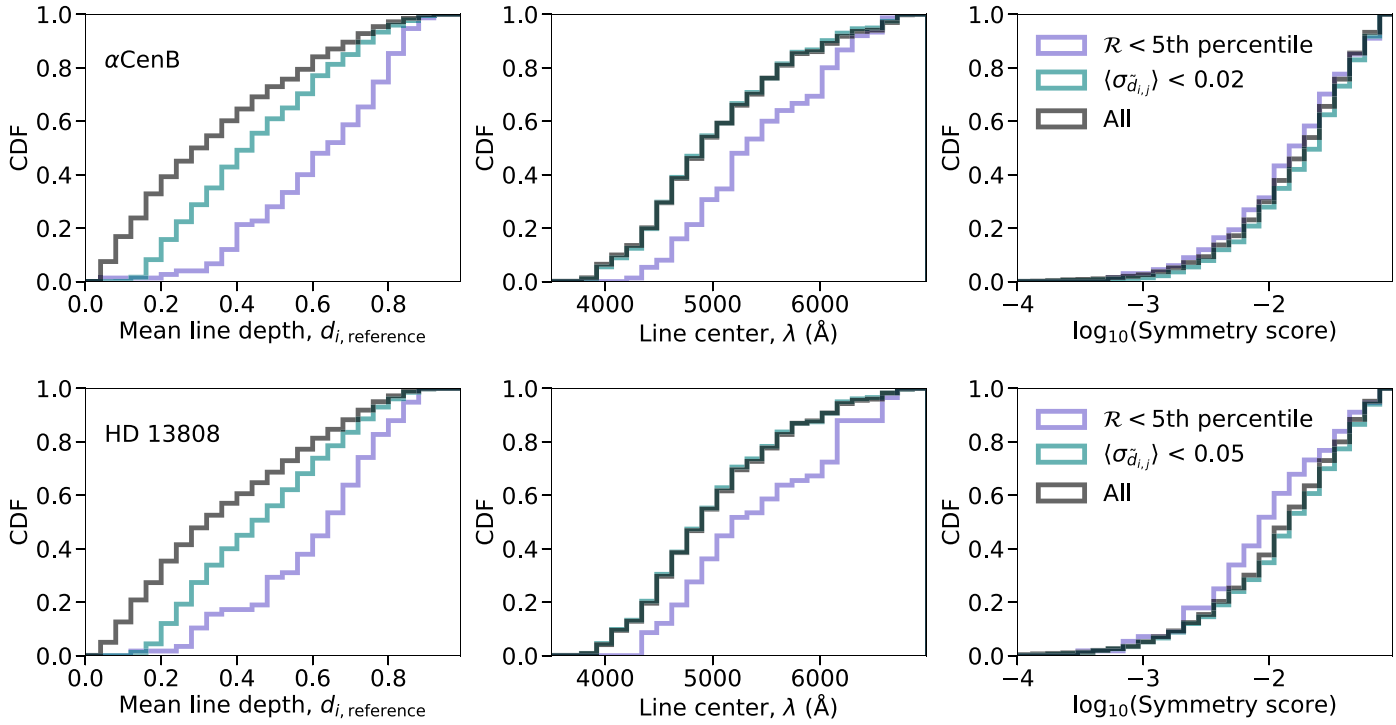

Figure 3. The spectral lines shown to be sensitive to stellar activity tend to be deeper and favor longer wavelengths. We present the cumulative distribution functions of reference spectrum line depth ($\tilde{d}_{i,\mathrm{reference}}$), central wavelength ($\lambda_i$), and symmetry for all lines and two subpopulations—activity-sensitive and low mean depth uncertainty—for αCenB (top row) and HD 13808 (bottom row).

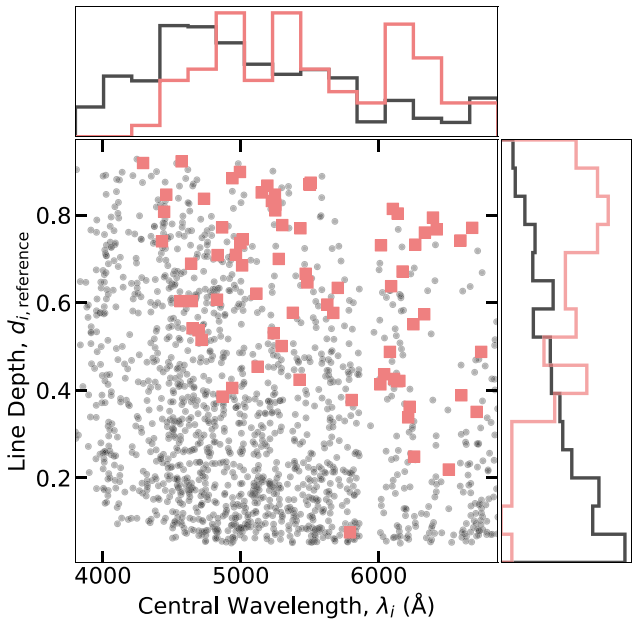

Figure 4. The subpopulation of activity-sensitive spectral lines in comparison with the entire population of spectral lines, for αCenB. The activity-sensitive population is defined as all lines where $\mathcal{R}(\mathrm{RV}_j$ vs. $d_{i,j})$ is less than the 5th percentile of αCenB's $\mathcal{R}(\mathrm{RV}_j$ vs. $d_{i,j})$ distribution. The activity-sensitive lines are shown as red squares, while all other lines are shown as black circles.

Gomes da Silva et al. (2018) and rescaling the $\mathcal{S}_{\mathrm{index}}$ by stellar color following Noyes et al. (1984).

Compared to standard activity indices, $\mathcal{D}(t)$ has several advantages. Unlike $\log R'_{\mathrm{HK}}$ and Hα, the depth metric covers a wide range of wavelength space and should be accessible to spectrometers spanning different wavelength regions. By likely probing temperature changes on the stellar surface through Zeeman splitting (Wise et al. 2018), the depth metric more closely tracks the stellar surface than chromospheric indices. Unlike diagnostics of the CCF, namely FWHM and BIS, the depth metric only considers lines shown to be activity-sensitive, more clearly isolating the stellar activity signal (see Table 1). Table 1 reports the Pearson correlation coefficient between the observed RVs and each activity index for αCenB and HD 13808; of the activity indices we considered, the depth metric has the strongest correlation with the measured RVs, for both stars.

However, linear correlation is a useful, but imperfect, standard for evaluating the quality of an activity index. The diagnostic power of an activity index should be weighted by considering both its application to RV detrending models and its accessibility (e.g., photometry facilitates high-quality activity tracking but is unavailable for the majority of RV measurements). To determine the quality of the depth metric, we conducted a comparative study of several activity indices (FWHM of the CCF, $\log R'_{\mathrm{HK}}$, and the depth metric) for a variety of detrending models; see Section 5. In Section 6, we discuss the accessibility of the depth metric and outline potential use cases.

4. Stellar Activity Mitigation Methods

As instruments approach the precision level necessary to detect an Earth twin, stellar activity can still generate m s^{-1} level noise and induce planet-like RV signatures. The line-by-line technique has the potential to greatly reduce such noise, since each spectral line responds differently to stellar activity depending on the details of the line transition. Here, we describe four stellar activity corrections: line selection (Dumusque 2018), formation depth (Cretignier et al. 2020),

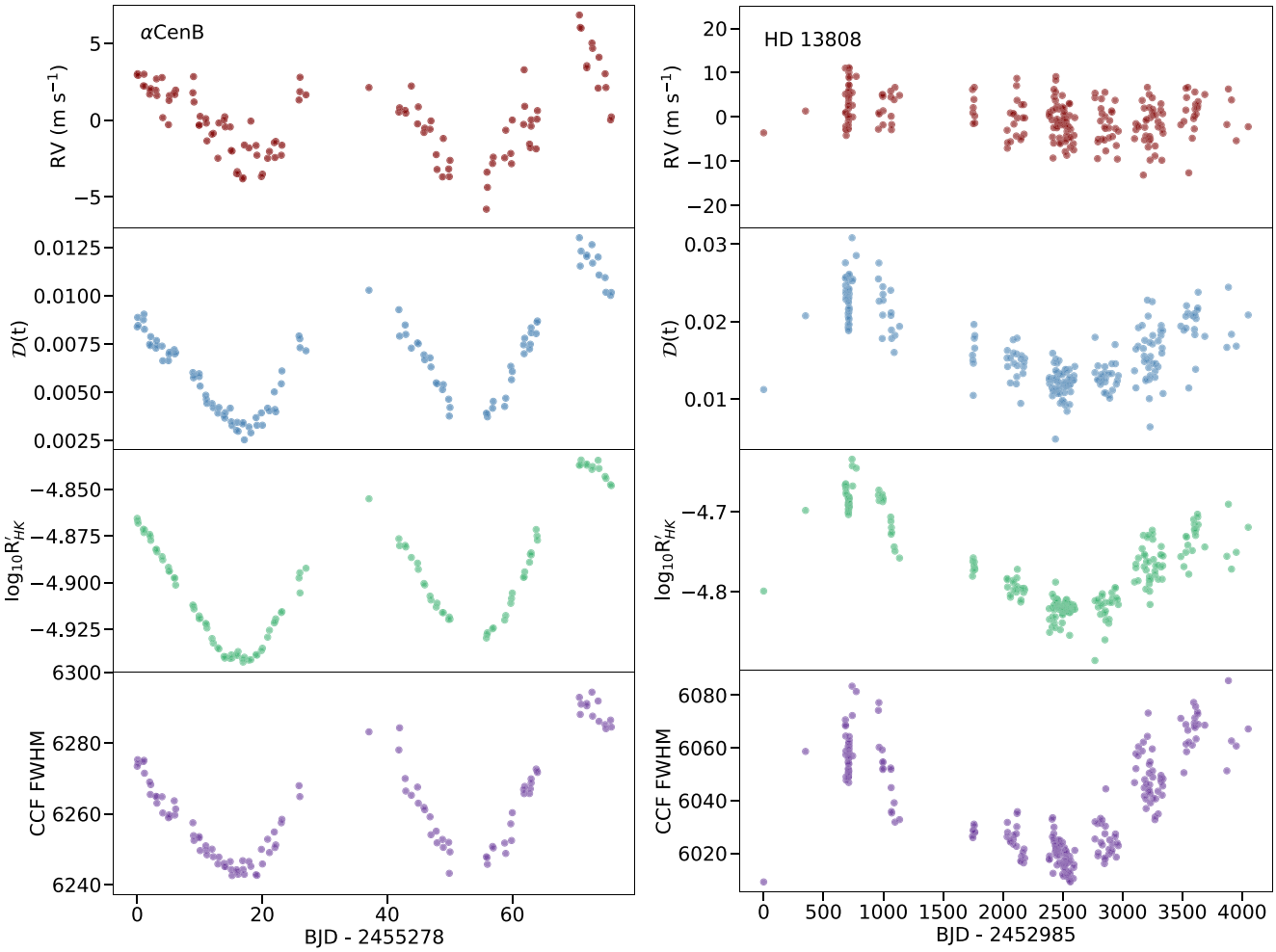

Figure 5. For both αCenB and HD 13808, the depth metric closely tracks the stellar activity signal. Here, we present time series of RV_j compared with the depth metric $\mathcal{D}(t)$, $\log R'_{HK}$, and FWHM of the CCF for αCenB (*left*) and HD 13808 (*right*). Due to the high cadence of observations, the αCenB data were binned to one tenth of a day.

Table 1
Activity Index Correlation Comparison

		RV	$\log_{10}R'_{HK}$	FWHM	$\mathcal{D}(t)$
αCenB					
	RV	⋯	0.776	0.770	0.813
	$\log_{10}R'_{HK}$	0.776	⋯	0.974	0.979
	FWHM	0.770	0.974	⋯	0.981
	$\mathcal{D}(t)$	0.813	0.979	0.981	⋯
HD 13808					
	RV	⋯	0.369	0.310	0.446
	$\log_{10}R'_{HK}$	0.369	⋯	0.835	0.836
	FWHM	0.310	0.835	⋯	0.681
	$\mathcal{D}(t)$	0.446	0.836	0.681	⋯

Note. For each star, we report the Pearson correlation coefficient $\mathcal{R}$ between each activity index, as well as the correlation coefficient between each activity index and the radial velocity measurements.

FF′ (Aigrain et al. 2012), and Gaussian progress (GP) regression (Haywood et al. 2014). In Section 5, we conduct a thorough comparison of these mitigation methods using αCenB and HD 13808 HARPS observations.

4.1. Line Selection Method

Dumusque (2018) first demonstrated that, by selecting a population of activity-insensitive lines, the RV scatter can be reduced; in the case of αCenB, Dumusque (2018) achieved a factor of $\sim$1.5 scatter reduction. For this initial investigation, the subpopulation of activity-insensitive lines was selected based on each line's mean uncertainty in $RV_{i,j}$ and each line's Pearson correlation coefficient between $RV_{i,j}$ and RV_j.

We selected our activity-insensitive set of lines through a similar filter. For a given star, we calculated each line's mean correlation coefficient between $RV_{i,j}$ and RV_j from 1000 Monte Carlo trials. We then defined the activity-insensitive population as all lines satisfying $|\mathcal{R}| < 0.1$ and $\langle \sigma_{RV_{i,j}} \rangle < 50$ m s^{-1}. Activity-corrected RVs were determined by recalculating RV_j using only these activity-insensitive lines.

4.2. Formation Depth Method

Utilizing the relationship between a line's formation depth and the magnitude of convective blueshift, Cretignier et al. (2020) introduced a stellar activity correction based on formation depth. For a population of activity-sensitive lines, Cretignier et al. (2020) first demonstrated that the magnitude of a given line's sensitivity to activity is inversely proportional to

that line's depth in the reference spectrum (effectively the line's mean depth). A subpopulation of shallow activity-sensitive lines thus show a greater RV scatter due to stellar activity than a subpopulation of deep activity-sensitive lines. As a result, the difference between the weighted average RV of the shallow subpopulation ($RV_{j,shallow}$) and the deep subpopulation ($RV_{j,deep}$) acts as an activity proxy. The stellar activity-corrected RVs are then $RV_j - (RV_{j,shallow} - RV_{j,deep})$.

Similar to the line selection method, there is considerable flexibility, and likely great room for improvement, in defining these subpopulations of lines. For αCenB, we defined the population of active lines as those where the correlation coefficient between RV_j and $RV_{i,j}$ is greater than 0.2, and for HD 13808 we adopted a threshold of 0.1. From this population of active lines, we defined the shallow population as all lines satisfying $0.1 < d_{i,\mathrm{reference}} < 0.4$ and the deep population as those satisfying $d_{i,\mathrm{reference}} > 0.6$; these depth thresholds were applied to both αCenB and HD 13808.

4.3. FF′ Method

The FF′ model, first described by Aigrain et al. (2012), uses photometric flux and a spot model to predict the RV signal due to stellar activity. For a normalized flux time series $F(t)$, the FF′ model predicts the effects of both convective blueshift ΔRV_c and the rotation of a single spot ΔRV_r,

$$\Delta \mathrm{RV}_c = F^2(t)\delta V \kappa / f, \tag{10}$$

$$\Delta \mathrm{RV}_r = -F(t)\dot{F}(t)R_\star / f, \tag{11}$$

where $R_\star$ is the stellar radius, f is the relative flux drop for a spot at the disk center, δV is the difference between the convective blueshift in the unspotted photosphere and that within the magnetized area, and κ is the ratio of the magnetized area to the area of the spot. The combined prediction for the stellar RV signal is

$$\mathrm{FF}'(t) = \Delta \mathrm{RV}_c + \Delta \mathrm{RV}_r. \tag{12}$$

Giguere et al. (2016) introduced the HH′ method, which uses the Hα index as a photometric proxy. Using the same model as Aigrain et al. (2012), with the addition of two free parameters to linearly scale Hα to flux and a smoothing term to compensate for low cadence, Giguere et al. (2016) found the HH′ method outperformed linear decorrelation but yielded a factor of two more scatter than the FF′ model. Giguere et al. (2016) postulated that the inclusion of more activity indicators could improve the HH′ method.

Using a modified FF′ framework, we applied the methods of Aigrain et al. (2012) to the depth metric $\mathcal{D}(t)$, the FWHM of the CCF, and $\log R'_{\mathrm{HK}}$. Treating the relative amplitudes of the spot rotation and convective blueshift effects as a free parameter and allowing for a linear relationship between $\mathcal{A}(t)$ and flux, our modified FF′ RV prediction is

$$\begin{aligned}\mathrm{FF}'(t) = &-\alpha(\mathcal{A}(t) + \beta)\dot{\mathcal{A}}(t)/f \\ &+ \gamma(\mathcal{A}(t) + \beta)^2/f \\ &+ \mathcal{C}_2 t + \mathcal{C}_1,\end{aligned} \tag{13}$$

where $\mathcal{A}(t)$ is a normalized activity index time series (defined analogously to Equation (9)), f is included to conveniently scale α and γ, β is the zero point of the assumed linear relationship between $\mathcal{A}(t)$ and photometric flux, and C_1 and C_2 are an arbitrary zero point and linear drift, respectively; following Aigrain et al. (2012), we assumed f is given by

$$f = \frac{\mathcal{A}_0 - \min(\mathcal{A}(t))}{\mathcal{A}_0}, \tag{14}$$

where $\mathcal{A}_0$ is defined analogously to Equation (8).

To estimate $\dot{\mathcal{A}}(t)$ for low-cadence targets, we smoothed $\mathcal{A}(t)$ using a smoothing parameter σ_t. For the jth observation, the smoothed $\mathcal{A}_j$ is the weighted average of the entire activity metric time series, where the weights are assigned via a Gaussian centered at t_j (the time of the jth observation) with a standard deviation of σ_t. The smoothed activity metric is fit with a cubic spline and $\dot{\mathcal{A}}(t)$ is determined analytically. Our modified FF′ model has six free parameters: α, β, γ, $\mathcal{C}_1$, $\mathcal{C}_2$, and σ_t.

4.4. Gaussian Process Regression

Gaussian processes (GPs) model a stochastic process using a parametric form for the covariance matrix. Prior studies have modeled structured stellar noise in RVs using a quasi-periodic kernel (Haywood et al. 2014; Dai et al. 2017; Ahrer et al. 2021). Following Haywood et al. (2014), we define a given element of the covariance matrix as

$$\begin{aligned}C_{n,m} = \eta_1^2 \exp\left[-\frac{(t_n - t_m)^2}{\eta_2^2} - \frac{\sin^2(\pi|t_n - t_m|/\eta_3)}{2\eta_4^2}\right] \\ + \delta_{n,m}[\sigma^2_{\mathrm{RV}_n} + \sigma^2_{\mathrm{jitter}}],\end{aligned} \tag{15}$$

where n and m index over the observations, η_1 is the activity amplitude, η_2 is the correlation length scale, η_3 is the period of the covariance, η_4 is the relative significance of the squared exponential and periodic components, $\delta_{n,m}$ is the Kronecker delta function, and σ_{jitter} is a white-noise jitter term.

The logarithmic likelihood function is

$$\ln \mathcal{L} = -\frac{N}{2}\log 2\pi - \frac{1}{2}\log|\boldsymbol{C}| - \frac{1}{2}\boldsymbol{r}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{r}, \tag{16}$$

where N is the number of observations, $\boldsymbol{C}$ is the covariance matrix, and $\boldsymbol{r}$ is the residual between the observed RVs and the model RVs.

For a given star and activity index, we modeled the structured stellar noise with a GP model conditioned on the activity index. We first modeled the activity index with the quasi-periodic GP kernel. For all activity indices, we adopted Jeffreys priors on η_1, η_2, and σ_{jitter}. For η_4, we adopted a Gaussian prior of 0.5 ± 0.05 (Kosiarek & Crossfield 2020), and for η_3 we adopted a Gaussian prior centered on the star's rotation period. Through Markov Chain Monte Carlo (MCMC) sampling, we generated posterior distributions for η_1, η_2, η_3, η_4, and σ_{jitter}. Using the same GP kernel, we then modeled the observed RVs. The priors for η_2, η_3, and η_4 were set by the posteriors of the activity index GP model. η_1 and σ_{jitter} again had Jeffreys priors. GP models were implemented via `radvel` (Fulton et al. 2018) and `emcee` (Foreman-Mackey et al. 2013).

5. Model Comparison

As described in Section 4, we considered four primary methods of mitigating stellar activity: line selection, formation depth, FF′, and GP regression. We conducted a systematic exploration of these four models using HARPS observations of

αCenB and HD 13808 and include linear decorrelation models for comparison.

5.1. αCenB

We applied a series of activity mitigation methods to αCenB HARPS observations from 2010. We constructed linear decorrelation, FF′, and GP models using the depth metric, the FWHM of the CCF, and $\log R'_{\rm HK}$. We also implemented two line-by-line based activity mitigation methods: line selection and formation depth.

To ensure our mitigation methods are suppressing stellar activity while preserving Keplerian signals, we considered the unmodified RV time series (for which there is no significant planetary signal) and time series with injected planetary signals. Signals were injected by Doppler shifting the wavelength solutions of the stellar spectra in accordance with the injected planet's RV signature. We considered a five-day planet on a circular orbit for three injection scenarios: $K = 0.75$, 1, and 2 m s^{-1}.

With the exception of the GP models, we considered the joint stellar activity and Keplerian model logarithmic likelihood function,

$$\ln \mathcal{L}(\mathrm{RV}_j|\theta) = -\sum_{j=1}^{N} \ln \sqrt{2\pi(\sigma^2_{\mathrm{RV}_j} + \sigma^2_{\mathrm{jitter}})} - \sum_{j=1}^{N} \frac{1}{2}\left[\frac{(\mathrm{RV}_j - \mathrm{RV}(t_j, \theta))^2}{\sigma^2_{\mathrm{RV}_j} + \sigma^2_{\mathrm{jitter}}}\right]. \quad (17)$$

RV (t_j, θ) are the joint activity and Keplerian RV predictions, RV_j are the observed values, σ_{RV_j} are the observational uncertainties, and σ_{jitter} is an additive "jitter" noise term. For the GP kernels, we modeled the RVs using a joint activity–Keplerian GP model conditioned on an activity index (see Section 4.4).

The Keplerian parameters of a given planet were K (RV semi-amplitude), P (orbital period), e (eccentricity), ω (argument of periastron), and T_0 (time of transit). We adopted uniform priors of $0 < K < 10$ m s^{-1}, $0.95 \cdot P_0 < P < 1.05 \cdot P_0$ days (for some central orbital period P_0), and $0 < \mathrm{T}_0 < P$. We fixed $e = 0$ and set $P_0 = 5$ days; for the null-injection case ($K = 0$ m s^{-1}), we froze the Keplerian component to $K = 0$ m s^{-1}.

For the stellar activity model components, both the line selection and formation depth methods had no free parameters, the FF′ models each had six, the GP models had five, and the linear decorrelation models had two. For the line selection and formation depth methods, we adopted the spectral line subpopulations outlined in Sections 4.1 and 4.2, respectively. In principle, the thresholds used for defining these subpopulations could be treated as free parameters, but we leave such analysis to later work.

The joint activity–Keplerian models were optimized via affine-invariant MCMC sampling using `emcee` (Foreman-Mackey et al. 2013). For the FF′ and linear decorrelation models, we adopted uniform priors for the activity model parameters and employed 90 walkers for 10,000 iterations each, where the first 2000 iterations were rejected as burn-in. For the GP models, we adopted the priors described in Section 4.4 and employed 50 walkers for 10,000 iterations each.

For the null-injection case, the corrected RVs for a selection of mitigation methods are shown in Figure 6. Each method shows a significant reduction in RV scatter and reduces the amplitude of the stellar activity signal. In terms of rms, the FF′ and GP models perform the best. For all methods, the resulting rms is considerably greater than the RV uncertainties. The bulk line-by-line RV measurements have a mean binned (to one tenth of a day) RV uncertainty of 0.18 m s^{-1}. The line selection and formation depth methods have mean binned uncertainties of 0.33 m s^{-1} and 0.68 m s^{-1}, respectively. These uncertainties only consider the uncertainties in $A\delta\lambda$ and A from the least-squares fits to each line. The DRS RVs have a mean binned RV uncertainty of 0.18 m s^{-1}.

To gauge the mitigation methods' planet detection capabilities, in Figure 7 we present the Lomb–Scargle periodograms of the activity-corrected RVs for a selection of mitigation methods under each injection case. To mimic a blind planet search, we omit the Keplerian component and model the RVs using pure activity models. Each correction method successfully recovers the $K = 2$ m s^{-1} planet at a 1% false-alarm probability level, a helpful but nonrigorous standard. However, only the FF′ models significantly recover the $K = 1$ m s^{-1} signal. While the depth metric based FF′ model successfully recovers the $K = 0.75$ m s^{-1} signal at a 1% false-alarm probability level, the detection is only marginally better than the FWHM FF′ model. Without a Keplerian component, the GP-based models had a strong tendency to absorb the injected planetary signals into the activity component of the model. This is less conducive to planet discovery than the other methods, which enhance the injected planetary signal after removing activity (Figure 7).

Table 2 shows the residual rms and the recovered planet parameters for each combination of mitigation method, activity index, and injection scenario. In Figure 8, we compare the recovered RV semi-amplitude for each mitigation method. With the exception of the line selection and formation depth methods, we find every mitigation method recovers the semi-amplitude within approximately 1σ for each injection case and constrains the orbital period to within 0.1 days. Between the GP and FF′ models, the FF′ models consistently yield lower uncertainties.

The FF′ models perform quite similarly across the three input activity indices. The depth metric FF′ model yields marginally lower uncertainties, σ_{jitter}, and rms than the FWHM- and $\log R'_{\rm HK}$-based FF′ models. For the GP models, we find no significant dependence on the input activity index. Unlike the FF′ and linear decorrelation methods, the GP models do not directly detrend the RVs using an activity index. Instead, the GP models are conditioned on the activity index by adopting the posteriors of the activity index GP model as the priors of the RV model.

With varying degrees of success, all mitigation methods considered here reduce the rotationally modulated activity signal in the αCenB time series. Evaluated on residual scatter and signal recovery, we favor the FF′ and GP models over the line selection, formation depth, and linear decorrelation methods. For all activity models, the depth metric performs as well as, or marginally better than, the FWHM of the CCF and $\log R'_{\rm HK}$.

5.2. Planet Characterization: HD 13808

Having validated that our models successfully track rotationally modulated stellar activity, we next applied these

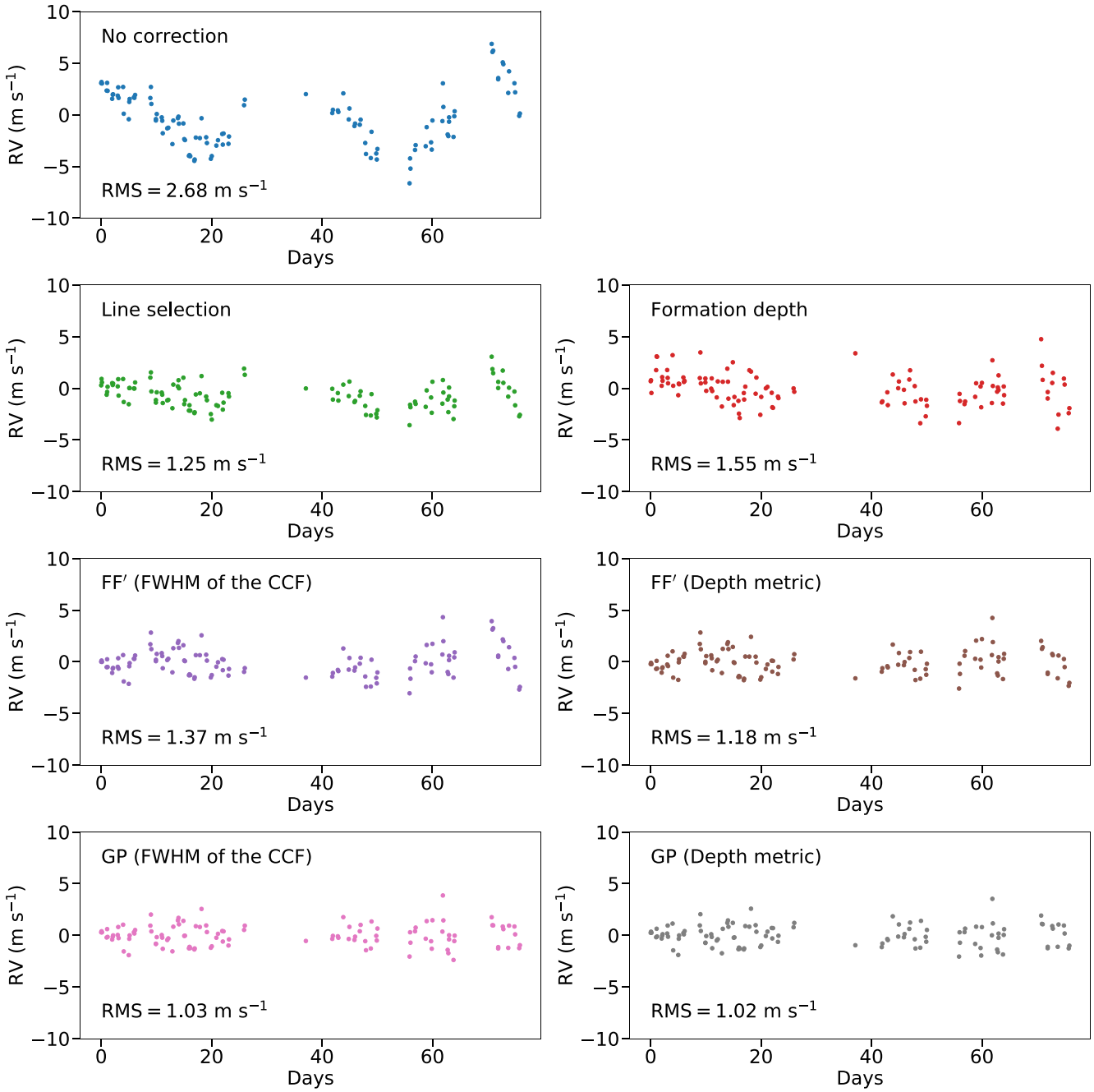

Figure 6. Each method successfully reduces the amplitude of αCenB's stellar activity signal, with the depth metric based methods performing at least as well as the FWHM of the CCF-based methods. We present the RVs of the 2010 HARPS αCenB observations under six different stellar activity mitigation methods, as well as the uncorrected case. The mitigation methods are described in Section 4.

models to a system with a multiyear magnetic activity cycle. For this exploration, we considered HD 13808, which was recently studied with several stellar activity mitigation methods, including linear decorrelation, CCF FWHM-based FF′, harmonic activity modeling, and Gaussian processes regression (Ahrer et al. 2021). That study confirmed the presence of two exoplanets, HD 13808b and HD 13808c, with orbital periods of 14.2 and 53.8 days and minimum masses of 11 and 10 $M_\oplus$.

We employed joint activity–Keplerian FF′, line selection, and formation depth models. These models only differ from those considered in Section 5.1 by the addition of a second Keplerian component. Free parameters in the models were inferred via MCMC sampling with 90 walkers for 100,000 iterations each, where the first 20,000 iterations were rejected as burn-in. We adopted uniform priors of $2 < K < 10$ m s^{-1}, $0.95 \cdot P_0 < P < 1.05 \cdot P_0$ days (for some central orbital period P_0), $-3 < \log_{10}(e) < 0$, $0 < \omega < 2\pi$, and $0 < T_0 < P$. For HD 13808b and HD 13808c, we set P_0 following the periods reported by Ahrer et al. (2021).

Figure 9 shows the posterior distributions of the line selection, formation depth, and depth metric based FF′ two-planet models. The phase-folded RVs of HD 13808b and HD 13808c for the depth metric FF′ two-planet model are shown in Figure 10. The 16th, 50th, and 84th percentiles for each model's posteriors are reported in Table 3. Both the line

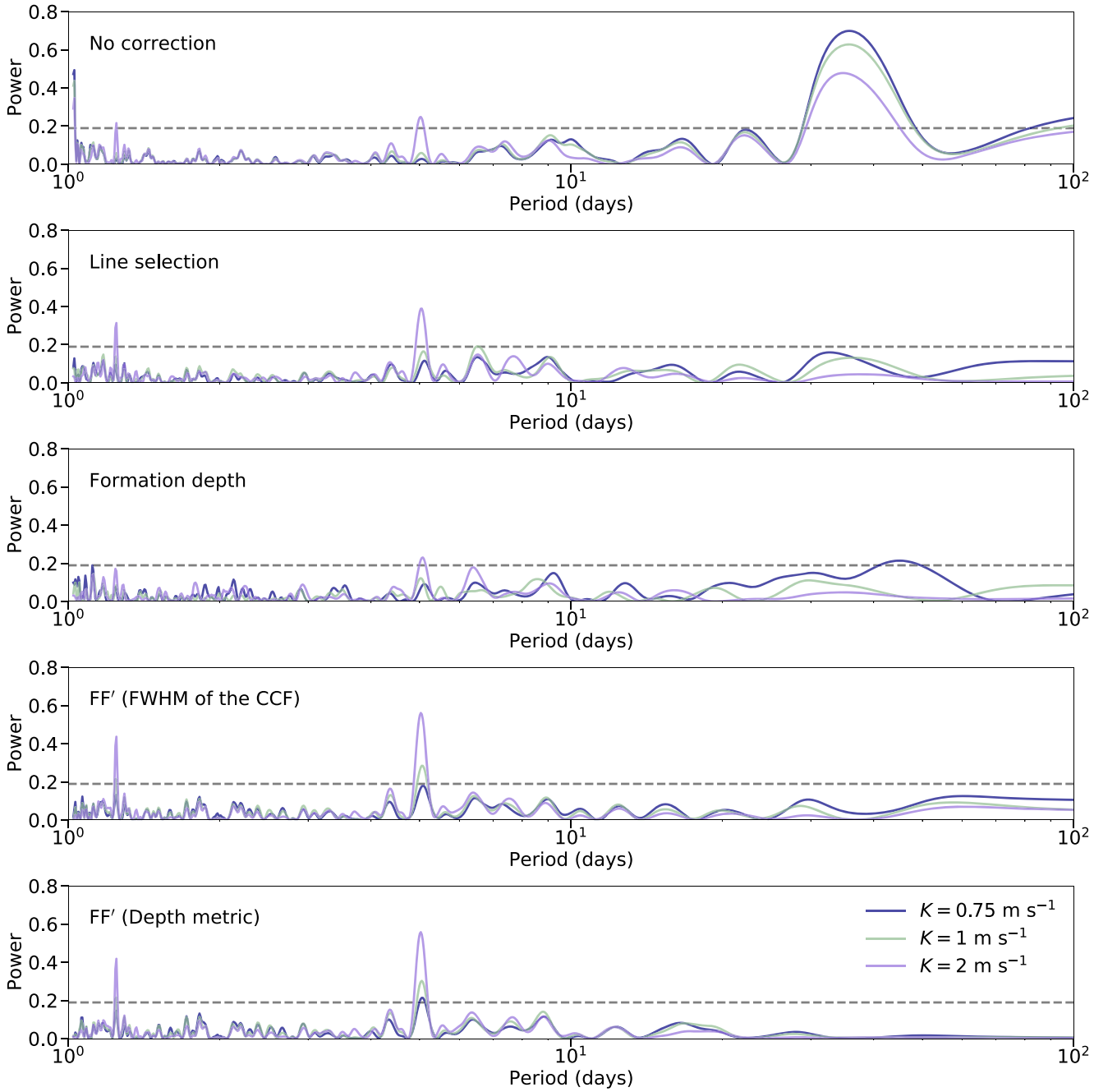

Figure 7. In terms of signal recovery, the FF′ models outperform the line selection and formation depth methods. We present the Lomb–Scargle periodograms for the HARPS αCenB observations under four stellar activity mitigation methods, as well as the uncorrected case; to mimic a blind planet search, we omit the Keplerian component in the RV models. Each color corresponds to a different RV semi-amplitude of the injected five-day planet. The horizontal dashed line demarcates the 1% false-alarm-probability level. RVs were binned to one tenth of a day.

selection and FF′ corrections achieve well-constrained posteriors and yield median parameter values in agreement with the preferred model of Ahrer et al. (2021). On the other hand, the formation depth method struggles to characterize the HD 13808 system. In addition to favoring high stellar jitter, the formation depth model yields wide posteriors for the RV semi-amplitudes and favors significantly higher eccentricities.

Although the line selection and depth metric based FF′ models both yield well-constrained posteriors, there are notable differences between the two. Compared to the line selection model, the FF′ method yields considerably narrower eccentricity posteriors; the line selection model has a significant high eccentricity skew for both HD 13808b and HD 13808c. Moreover, the two models are rather discrepant for the RV semi-amplitude of HD 13808b and HD 13808c, with the line selection model favoring lower amplitude solutions. A similar trend was seen in Section 5.1.

For a quantitative assessment of each models' relative evidence, we considered the deviance information criterion (DIC), the Bayesian information criterion (BIC), and the Watanabe–Akaike information criterion (WAIC). Following Gelman et al. (2013), we found information criterions relative to the depth metric based FF′ two-planet model of $\Delta\mathrm{DIC} = 90$ and 490, $\Delta\mathrm{BIC} = 40$ and 440, and $\Delta\mathrm{WAIC} = 70$ and 460 for

Table 2
αCenB Activity Mitigation and Signal Recovery

Correction Method	rms ($m\ s^{-1}$)	Injection Case	K ($m\ s^{-1}$)	P (days)	$\sigma_{\rm jitter}$ ($m\ s^{-1}$)
No correction (line-by-line RVs)	2.67	…	…	…	…
No correction (DRS RVs)	2.52	…	…	…	…
Line selection	1.29	…	…	…	…
	1.25	$K = 0.75$ m s^{-1}	$0.495^{+0.199}_{-0.206}$	$5.1084^{+0.0527}_{-0.0619}$	$1.375^{+0.103}_{-0.094}$
	1.26	$K = 1$ m s^{-1}	$0.677^{+0.204}_{-0.207}$	$5.1083^{+0.0430}_{-0.0492}$	$1.437^{+0.110}_{-0.098}$
	1.16	$K = 2$ m s^{-1}	$1.281^{+0.177}_{-0.179}$	$5.0390^{+0.0237}_{-0.0235}$	$1.285^{+0.095}_{-0.091}$
Formation depth	1.59	…	…	…	…
	1.78	$K = 0.75$ m s^{-1}	$0.648^{+0.258}_{-0.267}$	$5.1025^{+0.0489}_{-0.0539}$	$1.654^{+0.132}_{-0.126}$
	1.63	$K = 1$ m s^{-1}	$0.854^{+0.229}_{-0.228}$	$5.0347^{+0.0385}_{-0.0337}$	$1.450^{+0.129}_{-0.117}$
	1.64	$K = 2$ m s^{-1}	$1.374^{+0.230}_{-0.232}$	$5.0637^{+0.0299}_{-0.0293}$	$1.500^{+0.132}_{-0.116}$
Linear decorrelation ($\log R'_{\rm HK}$)	1.67	…	…	…	…
	1.64	$K = 0.75$ m s^{-1}	$0.888^{+0.231}_{-0.235}$	$5.0727^{+0.0472}_{-0.0425}$	$1.682^{+0.121}_{-0.109}$
	1.60	$K = 1$ m s^{-1}	$1.071^{+0.217}_{-0.225}$	$5.0562^{+0.0370}_{-0.0351}$	$1.637^{+0.115}_{-0.104}$
	1.61	$K = 2$ m s^{-1}	$2.081^{+0.228}_{-0.223}$	$5.0250^{+0.0181}_{-0.0174}$	$1.649^{+0.118}_{-0.105}$
Linear decorrelation (FWHM of the CCF)	1.89	…	…	…	…
	1.82	$K = 0.75$ m s^{-1}	$0.970^{+0.251}_{-0.248}$	$5.0505^{+0.0442}_{-0.0393}$	$1.838^{+0.098}_{-0.110}$
	1.69	$K = 1$ m s^{-1}	$1.165^{+0.236}_{-0.235}$	$5.0410^{+0.0351}_{-0.0309}$	$1.725^{+0.123}_{-0.111}$
	1.68	$K = 2$ m s^{-1}	$2.175^{+0.222}_{-0.231}$	$5.0202^{+0.0177}_{-0.0172}$	$1.716^{+0.115}_{-0.110}$
Linear decorrelation (depth metric)	1.54	…	…	…	…
	1.49	$K = 0.75$ m s^{-1}	$0.922^{+0.203}_{-0.206}$	$5.0529^{+0.0415}_{-0.0372}$	$1.520^{+0.113}_{-0.099}$
	1.47	$K = 1$ m s^{-1}	$1.123^{+0.202}_{-0.207}$	$5.0438^{+0.0304}_{-0.0293}$	$1.498^{+0.108}_{-0.098}$
	1.53	$K = 2$ m s^{-1}	$2.075^{+0.212}_{-0.212}$	$5.0193^{+0.0174}_{-0.0169}$	$1.560^{+0.114}_{-0.099}$
FF′ ($\log R'_{\rm HK}$)	1.40	…	…	…	…
	1.38	$K = 0.75$ m s^{-1}	$0.859^{+0.213}_{-0.196}$	$5.0900^{+0.0419}_{-0.0429}$	$1.419^{+0.109}_{-0.097}$
	1.41	$K = 1$ m s^{-1}	$1.034^{+0.214}_{-0.201}$	$5.0697^{+0.0342}_{-0.0332}$	$1.458^{+0.111}_{-0.099}$
	1.37	$K = 2$ m s^{-1}	$2.025^{+0.197}_{-0.191}$	$5.0340^{+0.0162}_{-0.0168}$	$1.414^{+0.106}_{-0.092}$
FF′ (FWHM of the CCF)	1.37	…	…	…	…
	1.29	$K = 0.75$ m s^{-1}	$0.845^{+0.184}_{-0.186}$	$5.0784^{+0.0417}_{-0.0416}$	$1.333^{+0.101}_{-0.096}$
	1.28	$K = 1$ m s^{-1}	$1.110^{+0.193}_{-0.206}$	$5.0596^{+0.0286}_{-0.0294}$	$1.324^{+0.103}_{-0.087}$
	1.28	$K = 2$ m s^{-1}	$2.064^{+0.183}_{-0.187}$	$5.0288^{+0.0162}_{-0.0150}$	$1.324^{+0.102}_{-0.088}$
FF′ (depth metric)	1.18	…	…	…	…
	1.20	$K = 0.75$ m s^{-1}	$0.848^{+0.167}_{-0.169}$	$5.0521^{+0.0367}_{-0.0330}$	$1.227^{+0.091}_{-0.082}$
	1.22	$K = 1$ m s^{-1}	$1.044^{+0.168}_{-0.166}$	$5.0467^{+0.0279}_{-0.0275}$	$1.247^{+0.091}_{-0.082}$
	1.23	$K = 2$ m s^{-1}	$1.996^{+0.172}_{-0.181}$	$5.0216^{+0.0147}_{-0.0142}$	$1.264^{+0.097}_{-0.084}$
GP ($\log R'_{\rm HK}$)	1.02	…	…	…	…
	1.40	$K = 0.75$ m s^{-1}	$0.898^{+0.183}_{-0.204}$	$5.0468^{+0.0002}_{-0.0007}$	$1.048^{+0.113}_{-0.089}$
	1.14	$K = 1$ m s^{-1}	$1.052^{+0.188}_{-0.183}$	$5.0468^{+0.0001}_{-0.0003}$	$1.048^{+0.115}_{-0.095}$
	1.61	$K = 2$ m s^{-1}	$1.521^{+0.364}_{-0.346}$	$5.0172^{+0.0001}_{-0.0001}$	$1.074^{+0.124}_{-0.106}$
GP (FWHM of the CCF)	1.03	…	…	…	…
	1.39	$K = 0.75$ m s^{-1}	$0.900^{+0.152}_{-0.182}$	$5.0487^{+0.0003}_{-0.0005}$	$1.053^{+0.103}_{-0.099}$
	1.06	$K = 1$ m s^{-1}	$1.037^{+0.185}_{-0.168}$	$5.0486^{+0.0001}_{-0.0004}$	$1.044^{+0.106}_{-0.081}$
	2.22	$K = 2$ m s^{-1}	$1.693^{+0.318}_{-0.433}$	$5.0172^{+0.0001}_{-0.0001}$	$1.065^{+0.120}_{-0.108}$
GP (depth metric)	1.02	…	…	…	…
	1.23	$K = 0.75$ m s^{-1}	$0.891^{+0.181}_{-0.214}$	$5.0468^{+0.0004}_{-0.0006}$	$1.072^{+0.109}_{-0.106}$
	0.97	$K = 1$ m s^{-1}	$1.076^{+0.176}_{-0.184}$	$5.0468^{+0.0002}_{-0.0002}$	$1.038^{+0.105}_{-0.092}$
	1.45	$K = 2$ m s^{-1}	$1.654^{+0.357}_{-0.489}$	$5.0171^{+0.0001}_{-0.0001}$	$1.089^{+0.132}_{-0.121}$

Note. We report the 16th, 50th, and 84th percentiles of the posteriors.

the line selection and formation depth two-planet models, respectively. We thus found strong evidence that the formation depth two-planet model is insufficient, and therefore we favor the FF′ two-planet model over the line selection model. However, greater exploration of all three methods is warranted, as there is room for optimization of the line selection and formation depth methods.

6. Conclusions

We introduced a high-signal-to-noise stellar activity indicator, the depth metric $\mathcal{D}(t)$, which co-adds depth variations across a set of spectral lines shown to be activity-sensitive. Through line-by-line radial velocity analysis of HARPS spectra, we identified >60 such lines for αCenB and >50

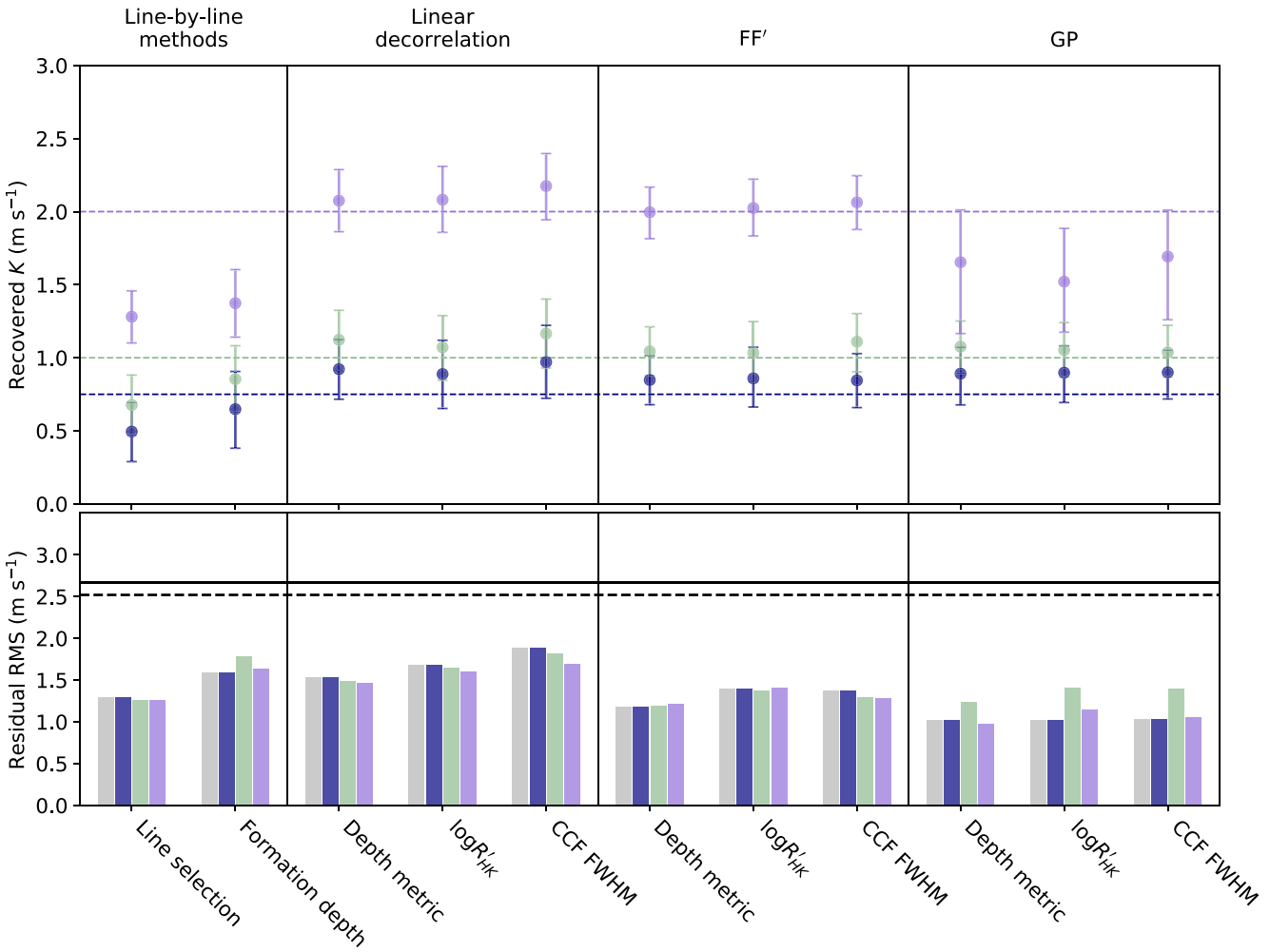

Figure 8. For a range of activity mitigation models, the depth metric performs at least as well as FWHM of the CCF and $\log R'_{\rm HK}$. We present the recovered RV semi-amplitude K (m s^{-1}) and residual rms (m s^{-1}) for each combination of mitigation method, activity index, and injection scenario. Top panel: the recovered RV semi-amplitude for the $K = 0.75$, 1, and 2 m s^{-1} injection cases are shown in blue, green, and purple, respectively. The injected RV semi-amplitudes are shown as dashed horizontal lines. Bottom panel: the residual rms for each injection case. The rms for the null-injection case is shown in gray. The rms of the uncorrected line-by-line RVs and the DRS RVs are shown as solid and dashed horizontal lines, respectively.

lines for HD 13808. This novel activity indicator has the benefit of being self-contained, computationally inexpensive, and translation-invariant. Unlike chromospheric indices, such as $\log R'_{\rm HK}$, this metric does not require coverage of a particular part of the spectrum. Wise et al. (2018) also postulated that spectral line depth variations are tied to Zeeman splitting from star spots, potentially making the depth metric a more direct probe of the stellar surface; however, the connection between spectral line distortions and physical processes within a host star remains poorly understood and warrants further study.

We then conducted a comparative activity mitigation study. We considered linear decorrelation, FF′, and GP regression models using three separate activity indices (the depth metric $\mathcal{D}(t)$, FWHM of the CCF, and $\log R'_{HK}$), as well as two line-by-line activity mitigation techniques, namely line selection (Dumusque 2018) and formation depth (Cretignier et al. 2020). We found the $\mathcal{D}(t)$-based FF′ and GP models outperformed the linear decorrelation and line-by-line methods and reached the quality of the FWHM- and $\log R'_{\rm HK}$-based FF′ and GP models. The depth metric based FF′ model successfully reduced the αCenB RV rms from 2.67 to 1.18 m s^{-1} and accurately characterized a five-day $K = 1$ m s^{-1} injected signal; the depth metric based GP model yielded a rms of 1.02 m s^{-1}.

Having proven it to be a powerful tool for mitigating the rotationally modulated activity signal of αCenB, we next applied the depth metric to HARPS observations of HD 13808. This system presents a clear multiyear stellar activity cycle and hosts two confirmed exoplanets. Between the line selection, formation depth, and $\mathcal{D}(t)$ FF′ models, only the line selection and FF′ two-planet models yielded well-constrained posteriors for HD 13808b and HD 13808c; these models generally agree with the preferred model of Ahrer et al. (2021). The deviance information criterion, the BIC, and the Watanabe–Akaike information criterion all significantly favor the depth metric based FF′ model.

We anticipate the depth metric will facilitate high-precision stellar activity tracking for a variety of applications. In the future, the depth metric has the potential for an exciting range of applications and elaborations. Below, we briefly describe several potential avenues for future research.

For Sun-like stars with activity signals on the m s^{-1} level, the depth metric has been shown to successfully track stellar activity with a level of quality similar to that provided by the

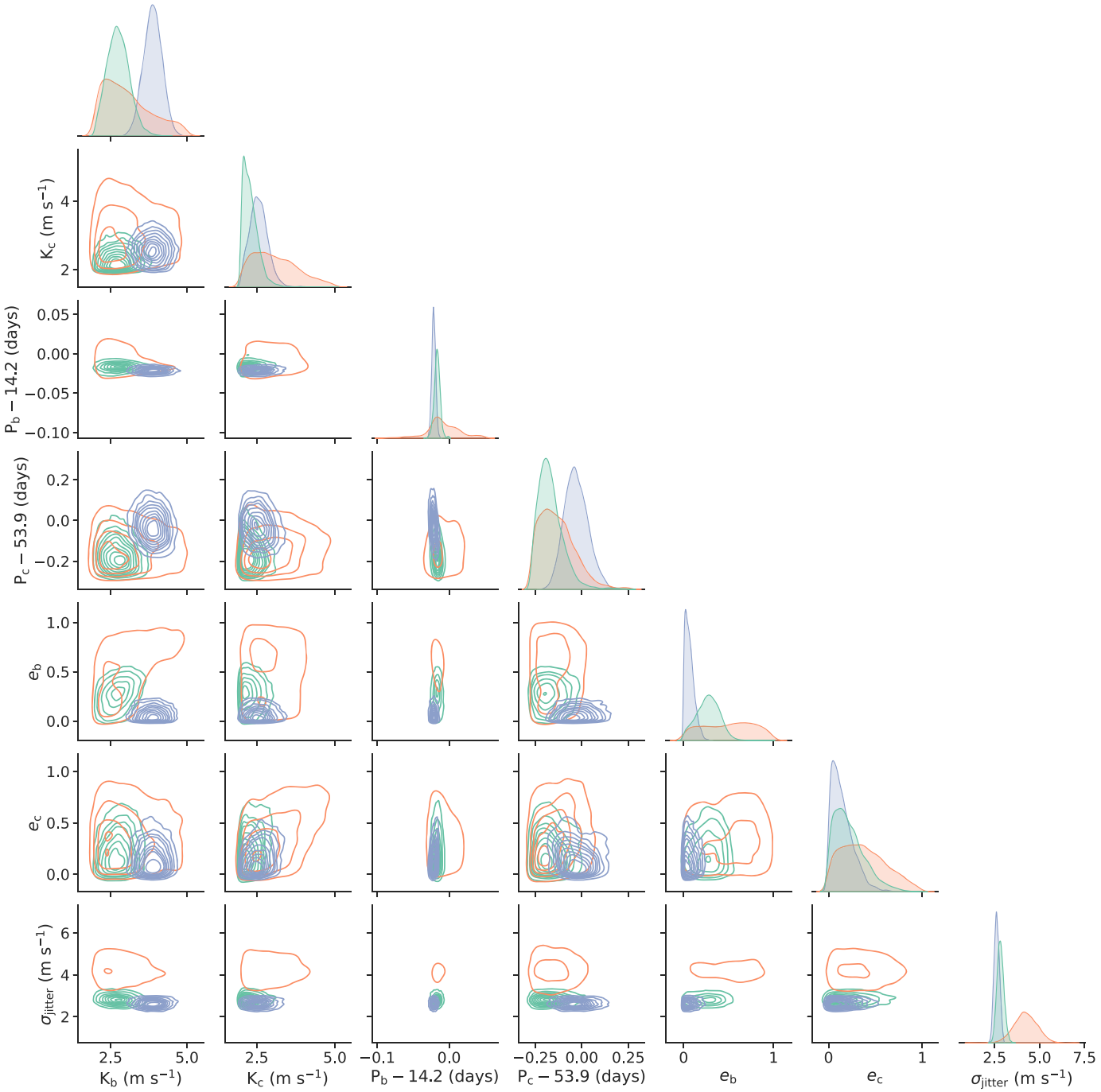

Figure 9. The depth metric based FF′ model yields well-constrained posteriors for HD 13808b and HD 13808c. We present the corner plot of the activity-corrected two-planet model posteriors for RV semi-amplitudes (K), orbital periods (P), eccentricities (e), and "jitter" ($\sigma_{\rm jitter}$). The posteriors for the line selection model are shown in green, the formation depth posteriors are shown in orange, and the depth metric based FF′ posteriors are shown in blue.

FWHM of the CCF and $\log R'_{\rm HK}$. Since the depth metric can be derived independently from these commonly used indices, it offers an additional dimension for detrending models and provides a valuable cross-check in cases of contentious detections; applying the depth metric alongside existing activity indices should yield more robust activity tracking and false-positive flagging. For stars with relatively weak activity signals, the co-adding of activity-sensitive spectral lines should instill the depth metric with higher signal-to-noise than existing indices, which may be dominated by white noise for low-activity stars; however, applying the depth metric to such stars may require modifications to the methods presented in Section 3; see below for discussion.

Presently, the depth metric method relies on the identification of activity-sensitive spectral lines through Pearson correlation coefficient analysis. However, this analysis relies on a long observing baseline and a moderate- to high-amplitude stellar activity signal; this is particularly true when planets are present. Since these restrictions do not apply to most commonly used activity indicators (e.g., FWHM of the CCF and $\log R'_{\rm HK}$), addressing these limitations is key to making the depth metric a scalable and accessible activity indicator. To

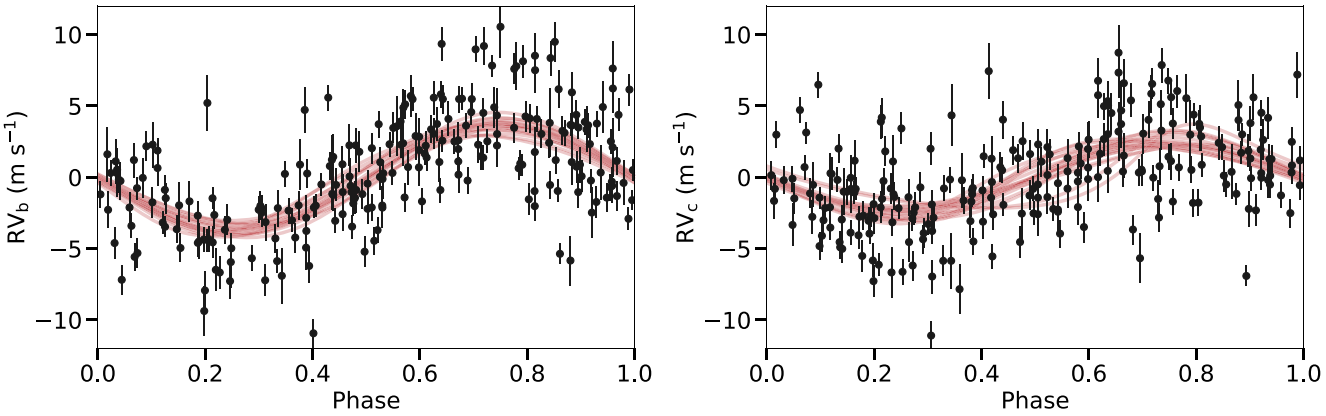

Figure 10. Phase-folded RVs of HD 13808b (left) and HD 13808c (right) for the depth metric FF′ two-planet model. For each planet, we present the observed RVs minus the stellar-activity correction and the RV model for the other planet (black circles). Random posterior samples for each planet's RV model are shown in red.

Table 3
HD 13808 Two-planet Model Posteriors

Correction Method		HD 13808b	HD 13808c
Line Selection			
	K (m s^{-1})	$2.700^{+0.384}_{-0.370}$	$2.265^{+0.299}_{-0.187}$
	P (days)	$14.184^{+0.006}_{-0.004}$	$53.724^{+0.069}_{-0.053}$
	e	$0.292^{+0.135}_{-0.141}$	$0.192^{+0.198}_{-0.134}$
	ω (rad)	$2.898^{+0.454}_{-0.522}$	$4.296^{+1.018}_{-1.732}$
	T_0 (days)	$4.673^{+1.852}_{-1.005}$	$40.434^{+3.604}_{-4.450}$
	$M \sin(i)$ ($M_\oplus$)	$8.155^{+1.108}_{-1.151}$	$10.850^{+1.332}_{-0.866}$
	$\sigma_{\rm jitter}$ (m s^{-1})	$2.845^{+0.208}_{-0.19}$	…
Formation Depth			
	K (m s^{-1})	$2.906^{+1.050}_{-0.637}$	$3.003^{+0.880}_{-0.650}$
	P (days)	$14.190^{+0.023}_{-0.014}$	$53.762^{+0.101}_{-0.086}$
	e	$0.510^{+0.295}_{-0.341}$	$0.358^{+0.290}_{-0.237}$
	ω (rad)	$3.499^{+1.294}_{-1.954}$	$4.278^{+0.823}_{-1.576}$
	T_0 (days)	$7.456^{+3.568}_{-4.569}$	$47.358^{+4.377}_{-13.427}$
	$M \sin(i)$ ($M_\oplus$)	$7.531^{+2.117}_{-1.763}$	$13.224^{+3.318}_{-2.798}$
	$\sigma_{\rm jitter}$ (m s^{-1})	$4.229^{+0.633}_{-0.636}$	…
FF′ (depth metric)			
	K (m s^{-1})	$3.872^{+0.301}_{-0.298}$	$2.552^{+0.293}_{-0.268}$
	P (days)	$14.179^{+0.003}_{-0.003}$	$53.869^{+0.066}_{-0.059}$
	e	$0.057^{+0.061}_{-0.040}$	$0.122^{+0.138}_{-0.086}$
	ω (rad)	$4.159^{+1.320}_{-2.090}$	$2.120^{+2.000}_{-1.263}$
	T_0 (days)	$6.718^{+0.573}_{-0.573}$	$33.345^{+3.371}_{-3.670}$
	$M \sin(i)$ ($M_\oplus$)	$12.288^{+0.958}_{-0.953}$	$12.468^{+1.372}_{-1.292}$
	$\sigma_{\rm jitter}$ (m s^{-1})	$2.619^{+0.154}_{-0.143}$	…
Ahrer et al. (2021)			
	K (m s^{-1})	3.67 ± 0.22	$2.18^{+0.22}_{-0.20}$
	P (days)	14.1815 ± 0.0015	$53.753^{+0.050}_{-0.082}$
	e	$0.071^{+0.027}_{-0.047}$	$0.156^{+0.050}_{-0.061}$
	$M \sin(i)$ ($M_\oplus$)	$11.2^{+1.2}_{-0.66}$	$9.96^{+1.8}_{-0.96}$

Note. We report the 16th, 50th, and 84th percentiles of the posteriors. T_0 values are relative to BJD = 2452985.6179. For comparison, we also present the maximum a posteriori (MAP) values from the preferred model of Ahrer et al. (2021).

facilitate the application of the depth metric method to all stars, regardless of observational cadence or activity level, a catalog of lines shown to be sensitive to stellar activity for a given stellar type should be generated. Such a catalog could also extend into the far red using data from MAROON-X, NEID, ESPRESSO, and KPF. For a given star, a set of activity-sensitive spectral lines could be interpolated from such a catalog. Alternatively, if a star hosts a planetary system, activity-sensitive spectral lines could be identified by constraining the activity sensitivity of each spectral line simultaneously with the planets' parameters.

The depth metric method and the broader Dumusque (2018) and Cretignier et al. (2020) methods only consider isolated spectral lines. This poses a considerable challenge when applying the line-by-line technique to stellar types dominated by blended lines (e.g., M dwarfs). Future modifications to the line-by-line method that consider blended lines would expand the stellar types accessible to line-by-line analysis.

Finally, solar line-by-line measurements, specifically $\mathcal{D}(t)$, should be mapped to independently derived solar observables that directly trace activity processes (Haywood et al. 2020). Such studies would help constrain the relation between magnetic fields and granulation on the stellar surface to RV variations.

We thank Jacob Bean and Andreas Seifahrt for helpful conversations and for spectra used for experimentation with this technique. We acknowledge NSF grant AST-2034278. R.A.R. is supported by an NSF Graduate Research Fellowship, grant No. DGE 1745301.

ORCID iDs

Jared C. Siegel https://orcid.org/0000-0002-9337-0902
Ryan A. Rubenzahl https://orcid.org/0000-0003-3856-3143
Samuel Halverson https://orcid.org/0000-0003-1312-9391
Andrew W. Howard https://orcid.org/0000-0001-8638-0320

References

Ahrer, E., Queloz, D., Rajpaul, V. M., et al. 2021, MNRAS, 503, 1248
Aigrain, S., Pont, F., & Zucker, S. 2012, MNRAS, 419, 3147
Anglada-Escudé, G., & Butler, R. P. 2012, ApJS, 200, 15
Baranne, A., Queloz, D., Mayor, M., et al. 1996, A&AS, 119, 373
Bertaux, J. L., Lallement, R., Ferron, S., Boonne, C., & Bodichon, R. 2014, A&A, 564, A46
Bouchy, F., Pepe, F., & Queloz, D. 2001, A&A, 374, 733
Butler, R. P., Marcy, G. W., Williams, E., et al. 1996, PASP, 108, 500
Collier Cameron, A., Mortier, A., Phillips, D., et al. 2019, MNRAS, 487, 1082
Cretignier, M., Dumusque, X., Allart, R., Pepe, F., & Lovis, C. 2020, A&A, 633, A76
Dai, F., Winn, J. N., Gandolfi, D., et al. 2017, AJ, 154, 226

de Beurs, Z. L., Vanderburg, A., Shallue, C. J., et al. 2020, arXiv:2011.00003
Demory, B.-O., Ehrenreich, D., Queloz, D., et al. 2015, MNRAS, 450, 2043
DeWarf, L. E., Datin, K. M., & Guinan, E. F. 2010, ApJ, 722, 343
Dumusque, X. 2018, A&A, 620, A47
Dumusque, X., Lovis, C., Ségransan, D., et al. 2011, A&A, 535, A55
Dumusque, X., Pepe, F., Lovis, C., et al. 2012, Natur, 491, 207
Dumusque, X., Pepe, F., Lovis, C., & Latham, D. W. 2015, ApJ, 808, 171
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Fulton, B. J., Petigura, E. A., Blunt, S., & Sinukoff, E. 2018, PASP, 130, 044504
Gelman, A., Hwang, J., & Vehtari, A. 2013, arXiv:1307.5928
Gibson, S. R., Howard, A. W., Marcy, G. W., et al. 2016, Proc. SPIE, 9908, 990870
Giguere, M. J., Fischer, D. A., Zhang, C. X. Y., et al. 2016, ApJ, 824, 150
Gomes da Silva, J., Figueira, P., Santos, N., & Faria, J. 2018, JOSS, 3, 667
Haywood, R. D., Collier Cameron, A., Queloz, D., et al. 2014, MNRAS, 443, 2517
Haywood, R. D., Milbourne, T. W., Saar, S. H., et al. 2020, arXiv:2005.13386
Jurgenson, C., Fischer, D., McCracken, T., et al. 2016, Proc. SPIE, 9908, 99086T
Kosiarek, M. R., & Crossfield, I. J. M. 2020, AJ, 159, 271
Kürster, M., Endl, M., Rouesnel, F., et al. 2003, A&A, 403, 1077
Mégevand, D., Zerbi, F. M., Di Marcantonio, P., et al. 2014, Proc. SPIE, 9147, 91471H
Noyes, R. W., Hartmann, L. W., Baliunas, S. L., Duncan, D. K., & Vaughan, A. H. 1984, ApJ, 279, 763
Queloz, D., Bouchy, F., Moutou, C., et al. 2009, A&A, 506, 303
Rajpaul, V., Aigrain, S., Osborne, M. A., Reece, S., & Roberts, S. 2015, MNRAS, 452, 2269
Saar, S. H., Butler, R. P., & Marcy, G. W. 1998, ApJL, 498, L153
Schwab, C., Rakich, A., Gong, Q., et al. 2016, Proc. SPIE, 9908, 99087H
Thompson, A. P. G., Watson, C. A., de Mooij, E. J. W., & Jess, D. B. 2017, MNRAS, 468, L16
Wise, A. W., Dodson-Robinson, S. E., Bevenour, K., & Provini, A. 2018, AJ, 156, 180
Wright, J. T. 2005, PASP, 117, 657