



SPORK That Spectrum: Increasing Detection Significances from High-resolution Exoplanet Spectroscopy with Novel Smoothing Algorithms

Kaitlin C. Rasmussen¹ , Matteo Brogi^{2,3,4} , Fahin Rahman⁵ , Hayley Beltz⁵ , Miles Currie¹ , Emily Rauscher⁵ , and Alexander P. Ji^{6,7,8}

¹ Department of Astronomy, University of Washington, Seattle, WA 98195, USA; kcrasmus@uw.edu

² Department of Physics, University of Warwick, Coventry CV4 7AL, UK

³ INAF-Osservatorio Astrofisico di Torino, Via Osservatorio 20, I-10025, Pino Torinese, Italy

⁴ Centre for Exoplanets and Habitability, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

⁵ Department of Astronomy and Astrophysics, University of Michigan, Ann Arbor, MI 48109, USA

⁶ Observatories of the Carnegie Institution for Science, 813 Santa Barbara Street, Pasadena CA 91101, USA

⁷ Department of Astronomy & Astrophysics, University of Chicago, 5640 S Ellis Avenue, Chicago, IL 60637, USA

⁸ Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA

Received 2021 August 26; revised 2022 April 13; published 2022 July 5

Abstract

Spectroscopic studies of planets outside of our own solar system provide some of the most crucial information about their formation, evolution, and atmospheric properties. In ground-based spectroscopy, the process of extracting the planets signal from the stellar and telluric signal has proven to be the most difficult barrier to accurate atmospheric information. However, with novel normalization and smoothing methods, this barrier can be minimized and the detection significance dramatically increased over existing methods. In this paper, we take two examples of CRIRES emission spectroscopy taken of HD 209458 b and HD 179949 b, and apply SPORK (SPectral cOntinuum Refinement for telluriKs) and iterative smoothing to boost the detection significance from 5.78σ to 9.71σ , and from 4.38σ to 6.89σ , respectively. These methods, which largely address systematic quirks introduced by imperfect detectors or reduction pipelines, can be employed in a wide variety of scenarios, from archival data sets to simulations of future spectrographs.

Unified Astronomy Thesaurus concepts: [Hot Jupiters \(753\)](#); [High resolution spectroscopy \(2096\)](#); [Exoplanet atmospheres \(487\)](#)

1. Introduction

The high-resolution ($R \gtrsim 25,000$) spectroscopic study of exoplanets, which has unlocked characterization of atmospheres at unprecedented scales, began in 1999 with Charbonneau et al. (1999), who used optical time-series spectra of the hot Jupiter τ Bootis b to claim a low albedo for the planet. Later that year, the first direct detection of reflected light from the same planet was made by Collier Cameron et al. (1999), and in the coming decade, more attempts would be made (Collier Cameron et al. 2002; Leigh et al. 2003; Rodler et al. 2008). A number of further attempts were performed to characterize the atomic and molecular makeup of τ Bootis b as well as other hot Jupiters known at the time (Wiedemann et al. 2001; Deming et al. 2005), but the first successful high-resolution measurements of an atomic species in an exoplanet atmosphere were Snellen et al. (2008) and Redfield et al. (2008), who both measured optical Na lines via transmission spectroscopy in the planets HD 209458 b and HD 189733 b, respectively. Later, Snellen would be the first to utilize cross-correlation as a detection method Snellen et al. (2010). High-resolution emission spectroscopy was introduced by Mandell et al. (2011), who used Kecks NIRSPEC to dispute a low-resolution claim of water in the L band of HD 189733 b (Barnes et al. 2010). Since then, the field has expanded considerably with the growing availability of high-resolution spectrographs, large

space- and ground-based telescopes, and increasingly effective statistical analysis methods. An excellent summary of the methodology of high-resolution exoplanet spectroscopy is presented in Birkby (2018).

Since its conception, however, the study of exoplanet atmospheres has been hindered by the difficulty of removing the stellar and telluric components of the spectra, which make up the vast majority of the signal. Stellar lines, to first order, do not vary over an observation.⁹ Tellurics, on the other hand, being a part of Earths ever-changing atmosphere, can vary strongly even from spectrum to spectrum depending on a number of factors such as water vapor density, season, cloud coverage, and zenith angle of the observation. Analysis of the exoplanet spectrum cannot begin until both of these components are removed.

Over the years, several measures have been adopted to remove stellar and telluric lines either separately or together. Snellen et al. (2008) as well as many later emission spectroscopy papers (Brogi et al. 2012, 2014, 2016), perform a relatively simple and fast removal by linearly fitting airmass trends and dividing out a reference (median) spectrum. This method works because, due to the rapid change in velocity that a close-in planet undergoes (especially as it approaches or recedes from eclipse or transit), the exoplanet spectrum moves over pixels, whereas the position and depth of the stellar lines, and the position (but not the depth) of the tellurics remains

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

⁹ It is known, but not well understood, how spot-crossing events play a role in changing the depth/shape of some stellar lines. Similarly, stellar flares may also impact the spectra in poorly characterized ways.

static. The results of this method can then be directly cross-correlated with the atmospheric model.

Another method for telluric removal is principal component analysis (PCA). PCA decomposes a set of vectors into a linear combination of eigenvectors and eigenvalues. In this case, the input vectors can be the individual spectra (i.e., PCA in the wavelength domain; see Lockwood et al. 2014; Zellem et al. 2014a; Piskorz et al. 2016, 2017; Giacobbe et al. 2021), or the individual spectral channels (PCA in the time domain; see de Kok et al. 2013). Regardless of the choice of domain, PCA identifies “components,” i.e., trends that are in common mode between all the input vectors. The user can select the minimum set of these trends sufficient to reproduce the main variations in the data, and divide their linear combination out to essentially normalize the data.

The program SYSREM (Tamuz et al. 2005) is a PCA routine which can account for error bars which are not static. This is crucial because the error bars on each point in a spectrum are correlated with where the point falls on the spectrograph; i.e., points which fall in the center pixels of the CCD have a higher signal-to-noise ratio (S/N) than pixels at the edge. SYSREM disassembles time-series spectra into principal components and removes them to an order chosen by the user. This method has been used successfully in emission spectra by Birkby et al. (2013), Nugroho et al. (2017), Alonso-Floriano et al. (2019), Sánchez-López et al. (2019), and many others.

In this paper, we focus on the former of these methods: airmass detrending. Airmass detrending is simple and fast, and unlike PCA and SYSREM, has a high degree of user control and flexibility, making it an ideal testing ground for our new methods.

2. Data and Models

We take two examples of hot Jupiters: the well-known HD 209458 b and the lesser-studied HD 179949 b, and detect CO in the former, and CO and H₂O in the latter.

HD 209458 b: HD 209458 b is a canonical example of a hot Jupiter, with a mass and radius of $0.67 M_{\text{Jupiter}}$ and $1.38 R_{\text{Jupiter}}$, respectively, and a period of 3.52 days (Southworth 2010). Spitzer phase curves of thermal emission from the planet show a dayside brightness temperature of 1499 ± 15 K and 972 ± 44 K on the nightside (Zellem et al. 2014b). Many species have been detected in its atmosphere, including He (Alonso-Floriano et al. 2019), water vapor (Deming et al. 2013; Sánchez-López et al. 2019), and several detections of CO (Gandhi et al. 2019; Beltz et al. 2021; Giacobbe et al. 2021), which we focus on in this work.

We use simulated spectra generated from a 3D General Circulation Model (GCM) of HD 209458 b, postprocessed with a line-by-line radiative transfer routine that only includes CO opacities, taken from the HITEMP database (Rothman et al. 2010), as an opacity source. Details on the GCM can be found in Beltz et al. (2021), and more information about the postprocessing routine used can be found in Zhang et al. (2017). Due to the strong influence the underlying chemistry assumptions on the detection significance of the models containing water in Beltz et al. (2021), we chose to use the CO-only model spectra. Specifically, Beltz et al. (2021) found that assumptions regarding the volume mixing ratios of water in the atmosphere had a significant impact ($\sim 2.5\sigma$ between the range of models tested) on the resulting detection significance. The models containing CO only on the other hand only showed

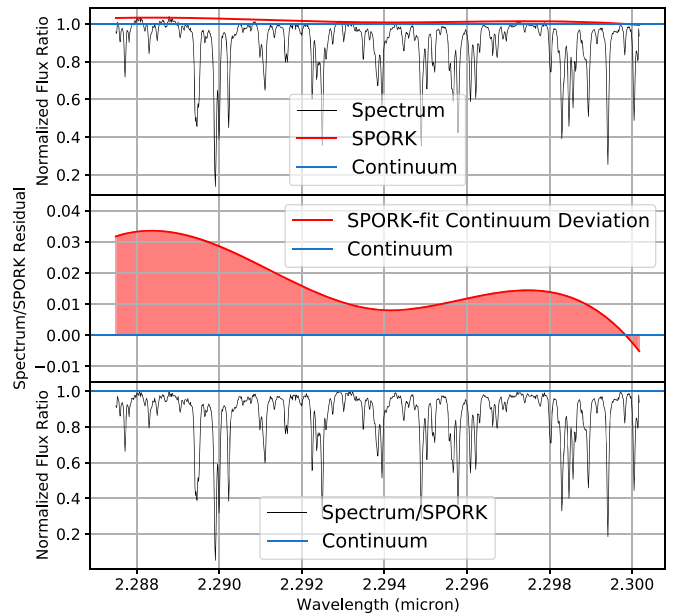


Figure 1. SPORK applied to CRILES spectra of HD 209458 b from the ESO reduction pipeline prior to telluric removal. Although the blaze function has been removed within the pipeline, residual wiggles in the continuum can be seen, especially between 2.288 and 2.292 μm , where the unnormalized continuum rises $\sim 3\%$ above the “true” continuum. If left uncorrected, these wiggles will persist throughout the telluric removal process and hinder the cross-correlation of the data against a perfectly flat model spectrum.

marginal ($\sim 0.3\sigma$) differences across the models tested. To avoid having our returned significance strongly influenced by assumptions regarding volume mixing ratios of water, we chose to use the more robust CO-only spectra.

The spectroscopic data was originally published in Schwarz et al. (2015), and detailed information about the observations (taken with the 8.2 m Very Large Telescope’s CRILES instrument; 2.285–2.348 micron, $R \sim 100,000$) can be found there. The spectra were optimally extracted using the ESO pipeline (Freudling et al. 2013) and wavelength-calibrated using the known positions of telluric lines.

HD 179949 b: HD 179949 b is a nontransiting hot Jupiter similar to HD 209458 b with a mass of $0.92 M_{\text{Jupiter}}$, a radius of $1.05 R_{\text{Jupiter}}$ (Wang & Ford 2011), a period of 3.09 days (Wittenmyer et al. 2007), and an equilibrium temperature of about 1950 K (Webb et al. 2020). A Spitzer phase curve of this planet implied fairly inefficient transport of heat to the planet’s nightside (Cowan et al. 2007). H₂O and CO have been detected in its atmosphere by Brogi et al. (2014) and Webb et al. (2020).

The atmospheric model used to cross-correlate against the spectra of HD 179949 was chosen from a large grid of models described in Brogi et al. (2014). A model with $\text{VMR}(\text{CO}) = \text{VMR}(\text{H}_2\text{O}) = 10^{-4.5}$, $\text{VMR}(\text{CH}_4) = 10^{-9.5}$, and a steep lapse rate of $\frac{dT}{d \log(p)} \sim 330$ K per pressure decade was the best fit to the data in that work, and thus is our choice for this analysis.

The spectroscopic data, taken at the same spectrograph, resolution, and wavelength range as HD 209458 b (CRILES; 2.285–2.348 micron, $R \sim 100,000$) was originally published in Brogi et al. (2014), which thoroughly details its observation and calibration. The spectra we use in this analysis have been reduced and wavelength-corrected using the same methods as HD 209458 b. As in Schwarz et al. (2015), the fourth detector has been discarded due to known odd–even pixel effects.

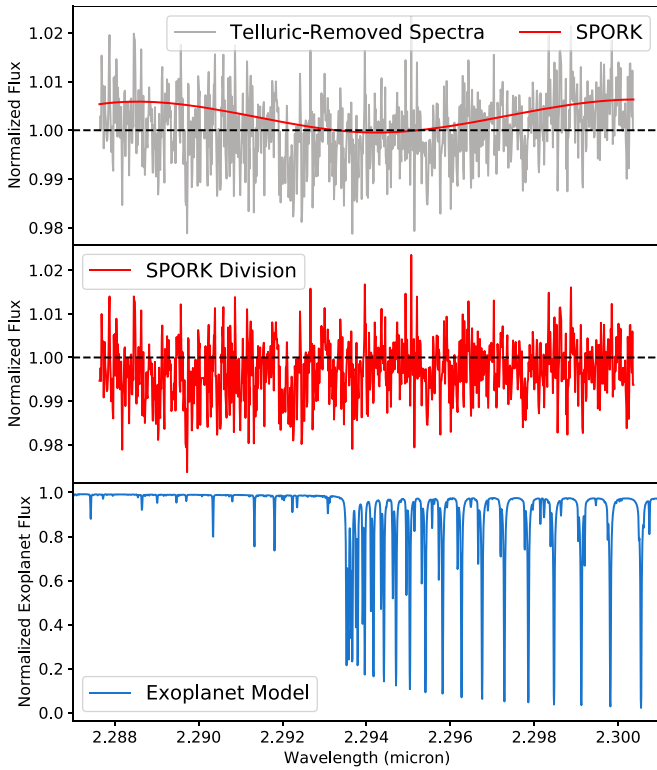


Figure 2. SPORK applied to prenormalized CRILES spectra of HD 179949 b from the ESO reduction pipeline *after* the airmass detrending, described in Section 3.1, has been performed. Top: a telluric-removed spectrum (gray) displays a strong offset from the continuum, and is fit with SPORK (red). Middle: the SPORK fit (knot spacing $M/2-1$ where M is the length of the wavelength array, $\sigma_{\text{upper}} = 3$, $\sigma_{\text{lower}} = 3$) is divided out, leaving behind a flat telluric-removed spectrum. Bottom: the normalized model exoplanet spectrum which corresponds to the telluric-removed spectral region is shown. Systematic offsets in the normalization of the HD 179949 b data cannot be removed by performing SPORK on the individual spectra, as they persist in the median spectrum as well. In this case, it is better to perform telluric removal first and then detrend the residual spectra. These spectra will perform better in the cross-correlation routine against the perfectly normalized exoplanet spectrum.

3. Methods

3.1. Airmass Detrending

As aforementioned, the key to extracting the exoplanet signal from the star’s signal, and often the Earth’s atmosphere’s signal, is the fact that the unwanted features remain stationary while the fast-moving exoplanet spectrum shifts across wavelength elements. For short-period, close-in planets, the timescale for this motion is 5–10 minutes, or roughly the length of a single exposure. This means that at exposure number 1 of N , a line in the exoplanet spectrum may occupy pixel number A ; but, by the time the time series of exposure is over, the same line now occupies pixel number $A+N$.

In airmass detrending, a median spectrum is generated from the sequential series of N exposures. A second-order polynomial is then fit between each spectrum and the median spectrum, and divided out, removing stationary features like the star. Because tellurics may change in depth throughout an observation for a variety of reasons (the largest of which is airmass variation), some will remain after this step. For this reason, a second median “spectrum,” this time in channel space (N elements long) is generated to reduce the impact of the change in airmass over time. This is also fit with a second-order polynomial that is divided out of each of the M different

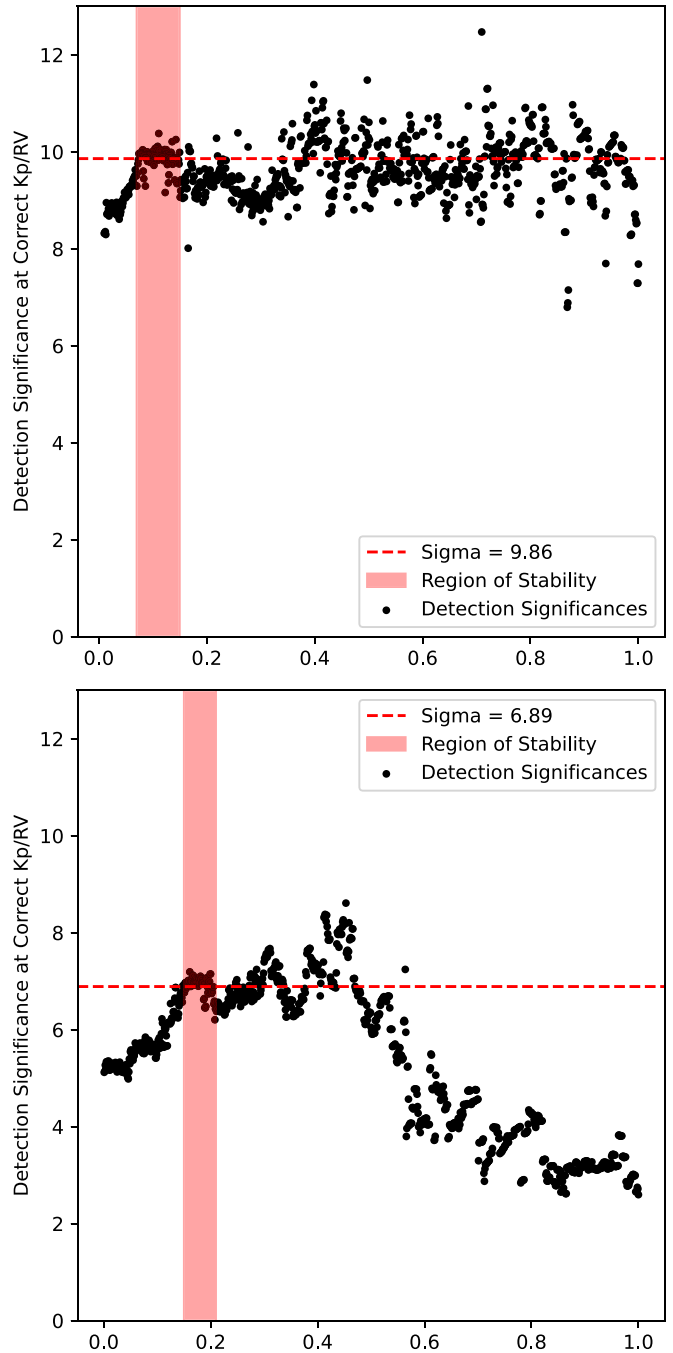


Figure 3. Detection significance vs. smoothing factor for HD 209458 b and HD 179949 b. Top: the full smoothing factor range from 0 to 1 is shown, with the region of stability highlighted in red. This region is chosen to avoid over-smoothing of the spectrum (see Figure 4, blue line), which begins to occur at 0.40 and produces high, but unstable detection significances. We select a smoothing factor from the region of stability of 0.12 to use in our analysis. Bottom: the same smoothing factor range is shown for HD 179949 b, with the region of stability highlighted. As with the case of HD 209458 b, the standard deviation from the running mean increases after 0.40. From this region of stability we choose the value of 0.194 for our analysis.

N -element channels, where M is the number of pixels in the wavelength regime. In this way, both stellar and telluric lines can be quickly and efficiently removed from a spectrum.

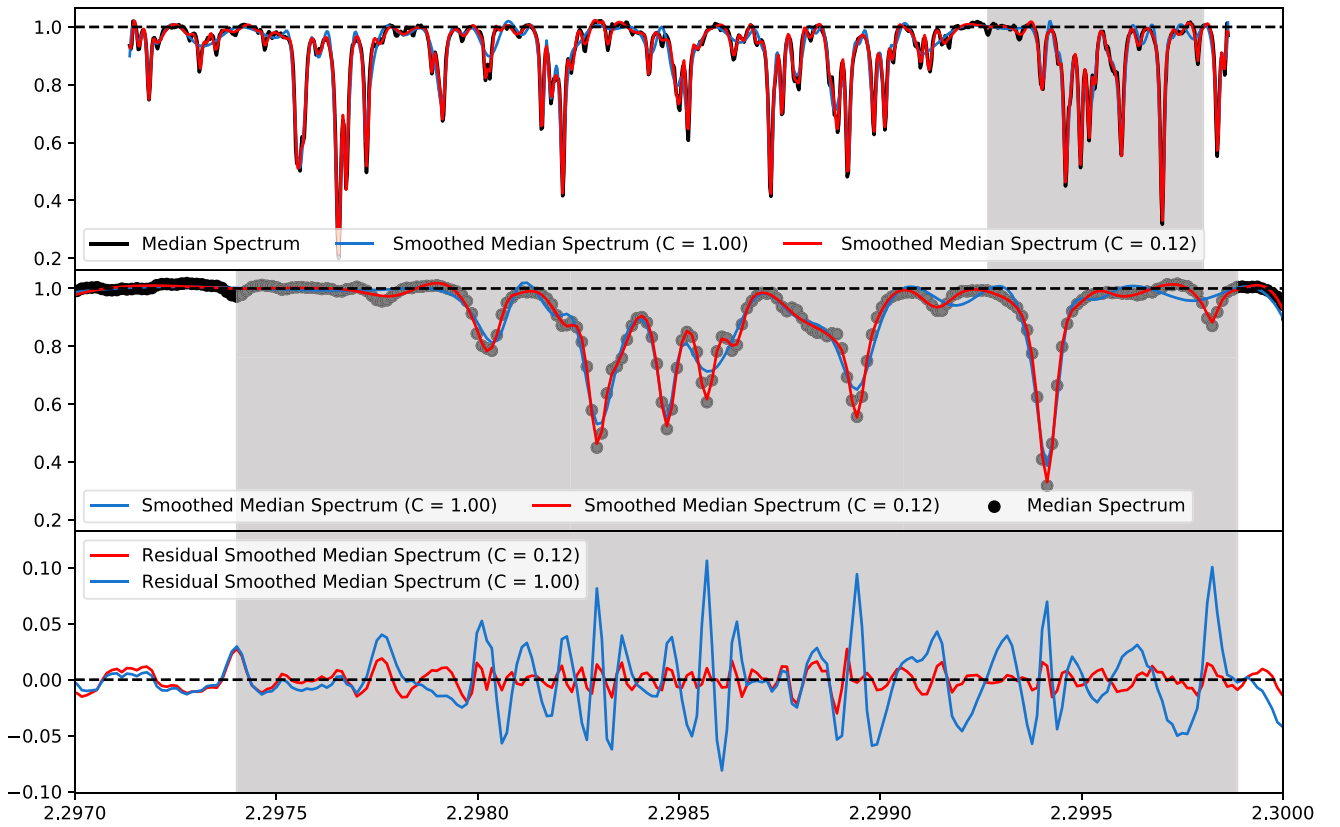


Figure 4. A median CRIRES spectrum of HD 209458 b and its optimally smoothed version. The gray region is highlighted to show the resolution of the CRIRES instrument and to highlight the closeness of the fit of the chosen factor. In iterative smoothing, the smoothing factor C in the python function `splrep` is increased by 0.001 over a range of 0–1, and a value is chosen from the region of stability to be the final smoothing factor. The red line represents the optimal smoothing factor, while the blue line represents an overfit.

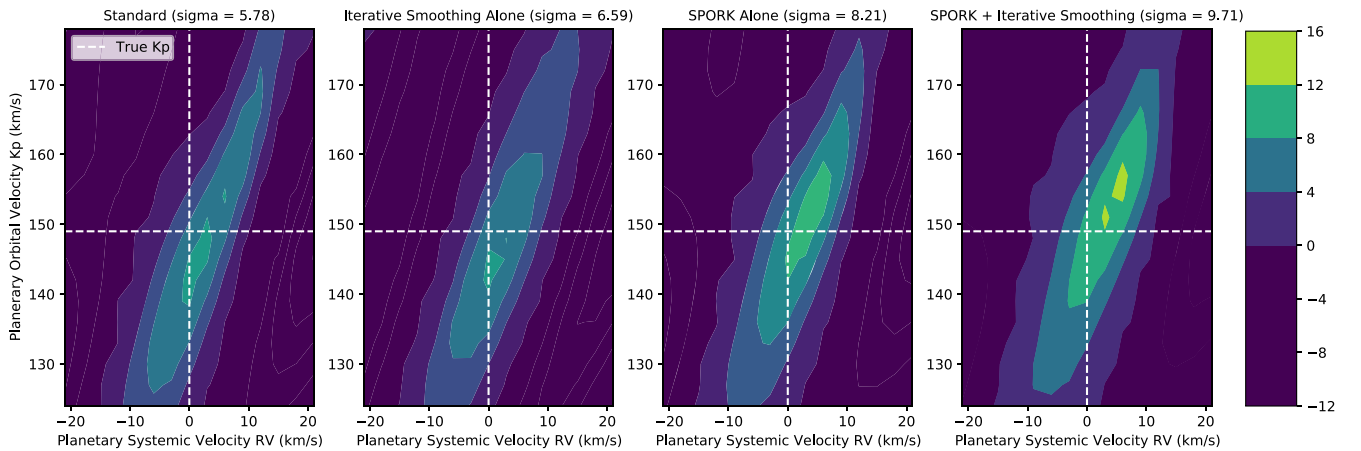


Figure 5. Results of cross-correlation on telluric-removed spectra for HD 209458 b. Left: the standard telluric routine leads to a detection of $\sigma = 5.78$ at $RV = 0$, $K_p = 149$, as is reported in Beltz et al. (2021). (Middle) When SPORK is applied before the telluric removal process, the significance of the detection increases to $\sigma = 8.21$. Right: when SPORK is applied, then iterative smoothing is run, this factor increases to $\sigma = 9.71$.

3.2. SPORK and Iterative Smoothing

In a typical high-resolution observation of an exoplanet, the information is extracted by cross correlating a normalized model spectrum against the telluric-removed spectra. Because the exoplanet’s signal is only $\sim 10^{-3}$ of the total normalized signal, slight deviations in the stellar continuum can mean major deviations in the exoplanet continuum (Snellen et al. 2010; Hoeijmakers et al. 2018; Brogi & Line 2019). Significant departure of the real exoplanet’s continuum from the model exoplanet’s continuum can produce low significance values

even when the model is a good match. Thus we show that correcting even slight deviations in the spectrum normalization can greatly increase the significance of molecular species detections. It should be noted that regardless of the technique used to isolate the exoplanet spectrum—even if it is not airmass detrending—as long as it is within the framework of high-resolution cross-correlation spectroscopy, SPORK in particular can be utilized to improve detection significance.

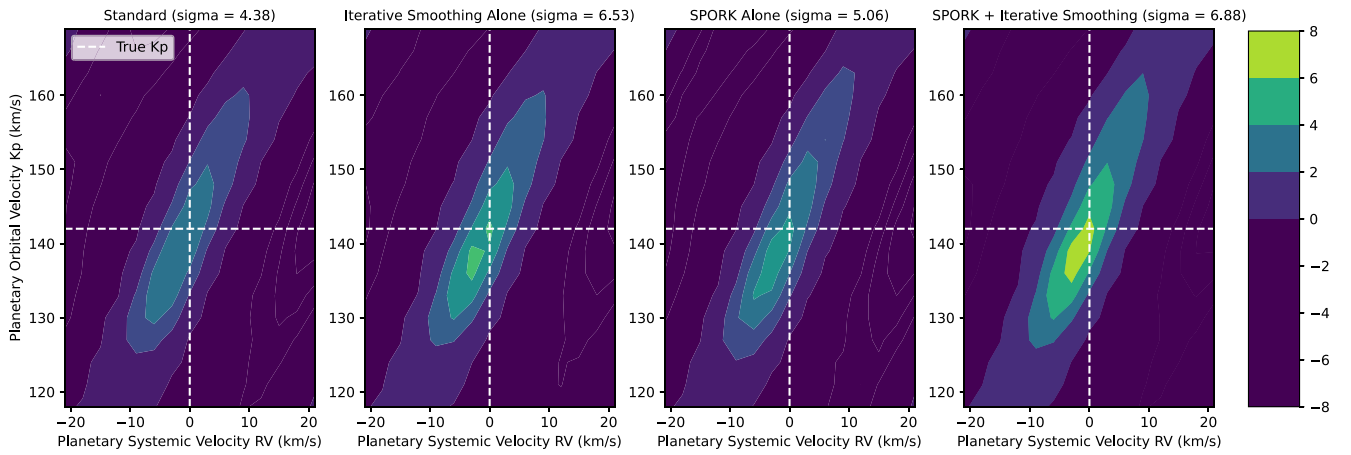


Figure 6. Results of cross-correlation on telluric-removed spectra for HD 179949 b. Left: the standard telluric routine leads to a detection of $\sigma = 4.38$. Middle left: the inclusion of iterative smoothing on its own, in this case, results in an increase of 2.15σ . Middle right: when SPORK alone is applied after the telluric removal process, the significance of the detection stays largely the same ($\sigma = 5.06$). Right: when SPORK is applied, then iterative smoothing is run, this factor increases to $\sigma = 6.88$ for a total increase of 2.50σ .

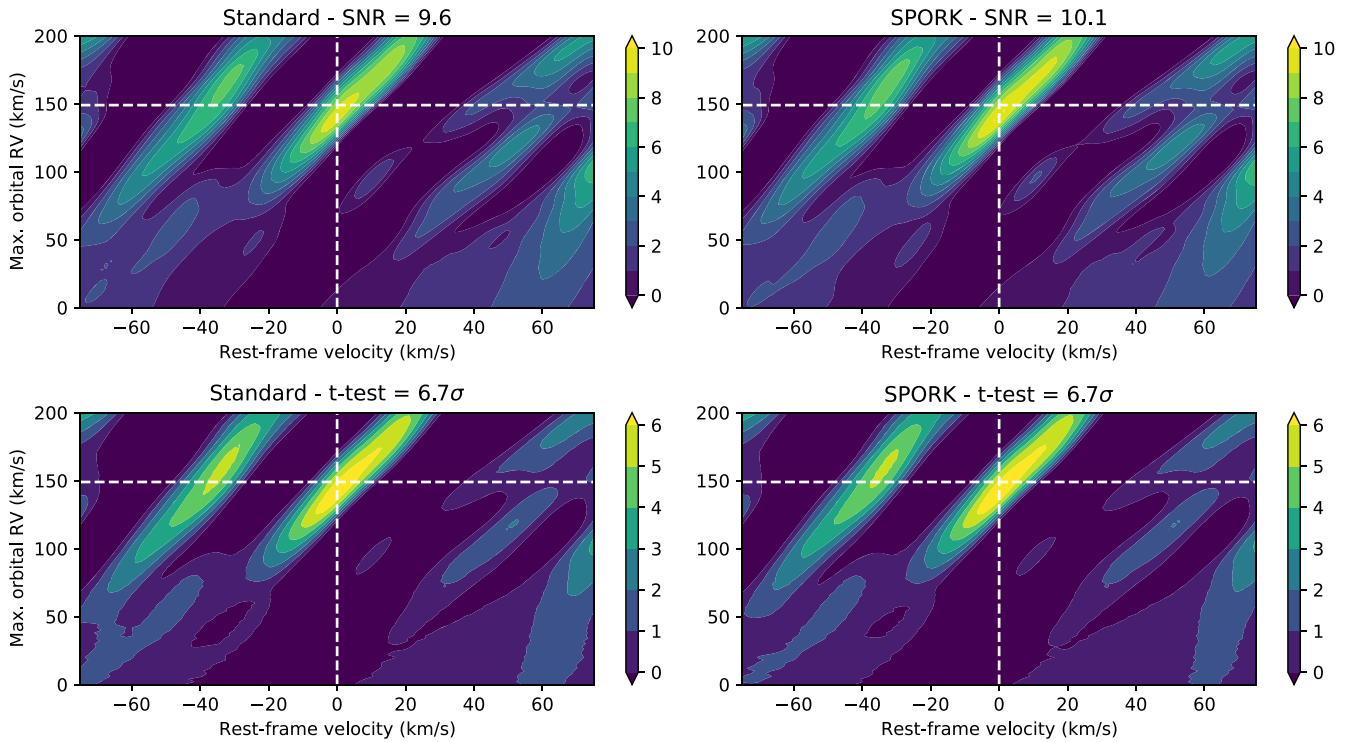


Figure 7. Results of S/R and t-tests on spectral with and without SPORK/iterative smoothing performed for HD 209458 b. Top panels: the implementation of SPORK/iterative smoothing in S/R tests result in a 0.5 planetary S/N increase. Bottom panel: in the t-test application, the implementation of SPORK/iterative smoothing does not have an impact.

SPORK (SPECTral cONTinuum Refinement for telluriKs)¹⁰ is a spectrum normalization routine adapted from the stellar abundance determination software Spectroscopy Made Hard.¹¹ The iterative smoothing function is based on the python scipy package `interpolate`.

3.2.1. Better Spectrum Normalization with SPORK

SPORK is a normalization tool optimized specifically to locate the continuum of the host star even in the presence of many large dips, such as the absorption features one encounters in stellar spectra. The tool iteratively fits a univariate natural cubic spline and identifies outlier pixels to be sigma clipped. Each order of an echelle spectrum has highly varying signal-to-noise as a function of wavelength, so the spectrum uncertainties reported by the pipeline are used to perform sigma clipping (rescaling by the standard deviation of the error-normalized deviations). Spline knots can be placed arbitrarily, but we maximize the number of knots as $N/2 - 1$ (where N is length of the wavelength array). This is because, in the case of

¹⁰ <https://github.com/fahnr/SPORK-Iterative-Smoothing>

¹¹ The most recent version of SMH, Spectroscopy Made Hard(er), can be found at <https://github.com/andycasey/smhr> SMH was first described in Casey (2014).

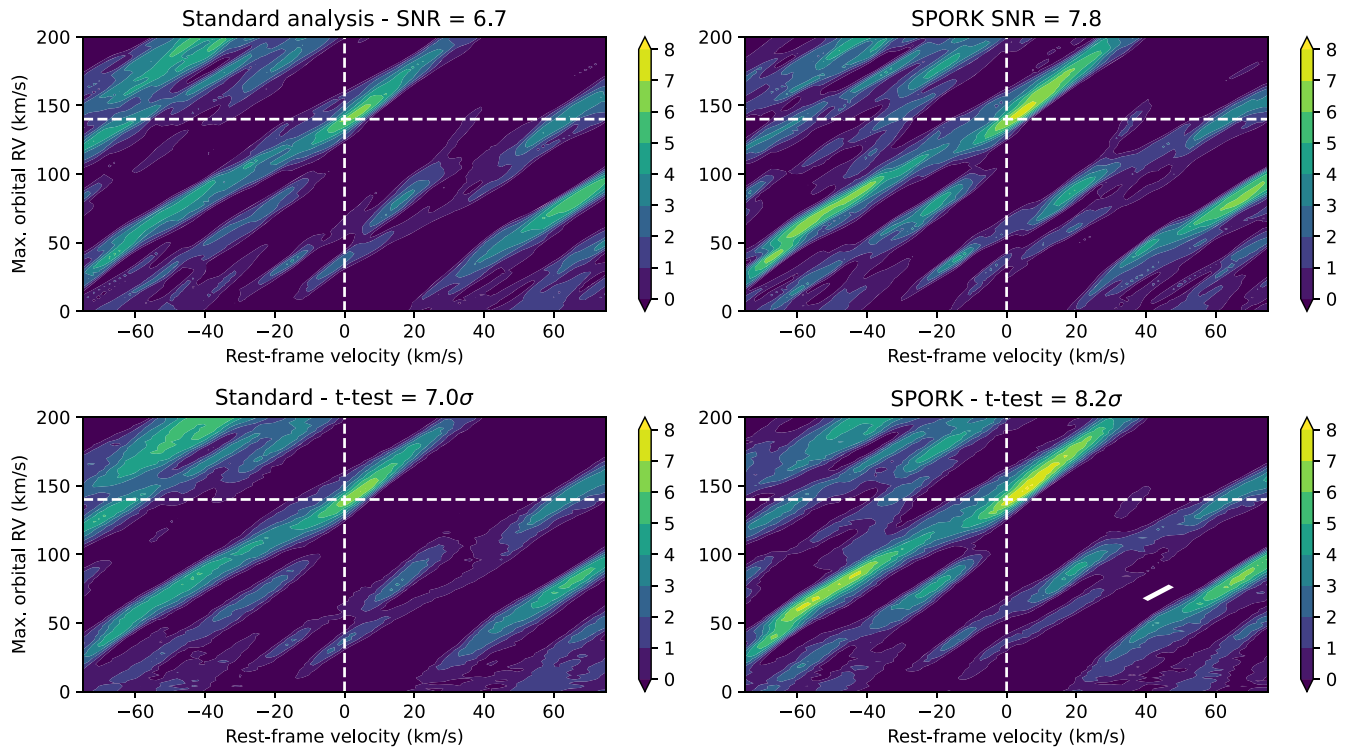


Figure 8. Results of S/N and t-tests on spectra with and without SPORK/iterative smoothing performed for HD 179949 b. Top panels: the implementation of SPORK/iterative smoothing in S/N tests result in a 1.1 planetary S/N increase. Bottom panel: in the t-test application, the implementation of SPORK/iterative smoothing results in a 1.2σ increase.

absorption spectra (as stellar spectra tend to be), we wish to achieve the “loosest” fit of the spline function: lowering the knot spacing leads to the fitting of individual spectral lines, which in this scenario, is not desirable. The lower sigma clipping threshold is set to 1.0 (points which deviate below the running mean by 1.0σ or more will be masked), as points which lie far below the continuum belong to stellar or telluric lines and should not be fit. The upper sigma clipping threshold is set to 5.0 (points that deviate above the running mean by 5.0σ or more will be masked), as features which rise sharply above a stellar spectrum are typically cosmic rays which should also be ignored.

An example of one usage of SPORK, where the algorithm is applied before telluric removal, is shown in Figure 1. In this particular case, each spectrum shows slight deviations from the continuum which are averaged out, and thus SPORK should be applied *before* telluric removal. However, this is not always the case; sometimes deviations in the spectra are systematic across the entire set. In this instance, it is more appropriate to apply the algorithm *after* tellurics have been removed, such as in Figure 2.

3.2.2. Better Preservation of the Exoplanet Signal with Iterative Smoothing

Airmass detrending, the process by which stationary features of a spectrum are divided out via a median spectrum (leaving behind noise, telluric residuals, and the exoplanet spectrum), is not a perfect process. Often, systematic pixel-to-pixel noise is preserved as well, muddying the buried exoplanet spectrum. One way to circumvent this is to apply a small degree of smoothing to the median spectrum to denoise it, reducing the

overall amount of noise that goes into the telluric removal process.

In order to determine the best factor of smoothing, we first set the smoothing factor to zero. The python module `scipy.interpolate` (specifically, `splrep` and `splev`) is used to construct a spline fit to the median spectrum. We then apply the smoothing factor to the spline fit of the median spectrum (in the first iteration, this will not affect the median spectrum). This way, instead of performing a second-order polynomial fit (polyfitting) of each individual spectrum to the median spectrum, each one is polyfit to the smoothed median spectrum instead. The temporal polyfitting is performed, and the entire model-injection cross-correlation routine used to detect the planet’s signal in the data (details can be found in Beltz et al. (2021); essentially, the detection significance is calculated by comparing the “true” cross-correlation function (CCF) generated by the cross correlating the telluric-removed spectra against the exoplanet model against a “model” CCF generated by injecting the telluric-removed spectra with the exoplanet model) is run at the literature orbital velocity (K_p) and the total radial velocity of the stellar system (RV_{rest}) position to determine the significance of the detection. We then increase the smoothing factor by 0.001, and the entire telluric removal and cross-correlation routine is performed again. This process is repeated 1,000 times with the smoothing factor ranging from 0 to 1. In this method, the smoothing factor function must reach stability (Figure 3; the region of stability will vary from instrument to instrument—here we define it as standard deviation $\lesssim 0.25$), at which point any factor in the stable region can be chosen as the set factor.

3.3. Telluric Removal

3.3.1. Telluric Removal with SPORK

Outermost pixels are masked and the median spectrum is constructed. However, before a polynomial fit is derived, the spectrum undergoes a round of SPORK normalization which removes any residual wiggles left over from the ESO CRRES reduction pipeline. After the continuum is located and the SPORK array is divided out, the wavelength and time polyfitting is carried out as in the standard process described in Section 3.2.1.

3.3.2. Telluric Removal with SPORK and Iterative Smoothing

After the masking of outermost pixels and generation of the median spectrum, one round of SPORK is applied to remove residual continuum wiggles. Then the median spectrum is subjected to an iterative degree of smoothing, and a smoothing factor is chosen from the region of stability—any value can be chosen from this region, as the final detection significance will remain largely the same. An example of the smoothed median spectrum using a smoothing factor from the region of stability as well as one which demonstrates an “overfit” is shown in Figure 4.

3.3.3. Usage

Any spectra, from any spectrograph and using any telluric removal method should be first visually inspected to determine whether deviations from the continuum are occurring spectrum-by-spectrum or whether the entire set deviates together. If the former, SPORK should be applied *before* telluric removal; if the latter, it should be applied *after*. For particularly messy spectra, SPORK may be used both before and after. We recommend testing SPORK as it is employed before, after, and both in order to obtain the best result. Iterative smoothing works solely within the framework of airmass detrending, and should be tested as such.

4. Results

The implementation of SPORK alone results in detection significance increases of 2.43 and 0.68σ at $RV=0$, and the literature K_p for HD209458 b and HD 179949 b, respectively. Iterative smoothing alone results in detection significance increases of 1.50 and 2.15. In total, we find that applying these new methods raised the detection significance of HD 209458 b and HD 179949 b from 5.78σ to 9.71σ and from 4.38σ to 6.88σ , respectively. These results can be seen clearly in Figures 5 and 6.

In the case of HD 209458 b, the application of SPORK to the spectra which come out of the ESO reduction pipeline has the stronger effect of the two methods. Not only does the signal increase dramatically, but the location of the “smear” of significance becomes more reflective of the literature K_p value. Applying an additional round of post-telluric-removal SPORK to the HD 209458 b analysis did not result in any further increase in signal. For HD 179949 b, the location of the “smear” of significance does not change—this is perhaps due to the different SPORK treatments between the two planets (for HD 209458 b, SPORK was applied *before* airmass detrending, for HD 179949 b it was applied *after*), as well as the fact that HD 209458 b’s atmosphere model was generated with a 3D GCM, while HD 179949 b’s model was not. Previous work

(Flowers et al. 2019; Beltz et al. 2021) has shown that postprocessing 3D GCMs to produce high-resolution spectra can return higher detection significance than suites of 1D models. Using a GCM as the starting point for the spectra of HD 179949b would have likely resulted in an increase in detection compared to the 1D models here but a new GCM is beyond the scope of this work.

4.1. Standard S/N and t-Tests

We test SPORK and iterative smoothing in two different significance determination frameworks: S/N calculation and t-test. For HD 209458 b and HD 179949 b, the measured planetary S/N increases by 0.5 and 1.1, respectively (Figure 7). When t-tests are performed, the significance either does not change (as in the case of HD 209458 b) or increases by 0.8σ (Figure 8).

4.2. Discussion and Future Applications

With the advent of large ground- and space-based observatories, and the coming era of terrestrial-sized exoplanet observations, SPORK and iterative smoothing will be essential to detecting and quantifying ever smaller exoplanet signals. They are simple and highly accessible techniques that can be applied to any set of spectra from any instrument in any wavelength range. Every spectrograph/reduction pipeline combination has systematic quirks that must be corrected for, and SPORK, in particular, can be applied not only to spectra that have already been reduced and normalized but even to reduced spectra from which the blaze function has not been removed (such as the output of the IGRINS reduction pipeline). For exoplanet science, this method can be utilized at any point before or during the telluric removal process, whether that process is airmass detrending, PCA, SYSREM (the PCA routine which accounts for non-uniform error bars), or model fitting.

The flexible SPORK can also be used to smooth, detrend, and normalize nonspectra data sets. In its current state, the knot spacing and the upper/lower sigma values are optimized for absorption spectra, but can also be customized for other one-dimensional data such as light curves. SPORK’s one drawback is that it is not a particularly fast algorithm (expect analysis times to be 2–4 times slower than standard airmass detrending), and thus significant overheads are expected when analyzing large sets of data (modern spectrographs can produce data sets as big as 10^7 data-points) or comparing them to large grids of models (10^5 – 10^6), including the future integration into Bayesian retrievals.

Iterative smoothing is best applied within the airmass detrending framework. As it requires the running of the full cross-correlation code for many values, it is also quite slow (in its current iteration, 4–6 hr are required for 1000 smoothing factors—although it is only run one time) and should be used on a case-by-case basis. Future investigations, including efforts to parallelize this method, could yield lower computing times.


One caveat of the current implementation of the analysis is that we are optimizing the smoothing parameters by maximizing the detection with a selected model. This could potentially lead to a model-dependent optimization. Therefore, we recommend to first run the classic telluric removal without smoothing, and select a reasonable family of models that leads to a detection of the planet’s atmosphere. The smoothing can

then be optimized by using this subset of models, which would avoid cases in which a nondetection could be potentially overly optimized to produce a detection.

Another important caveat is that in this work we explore the use of cross-correlation techniques to “detect” a species, i.e., to maximize the information coming from matching the position and depth of a set of spectral lines. This application is crucial, e.g., to compile the chemical inventory of exoplanets, even when additional properties of the atmosphere are unknown. Recently, high-resolution spectroscopy has been extended to extract information about abundances and temperature, e.g., in Brogi & Line (2019). In this latter case, the unavoidable alteration of the “true” exoplanet spectrum via the application of telluric removal needs to be thoroughly simulated and replicated on each model to avoid biases. Such level of detail is arguably beyond the scope of this work, and we defer to a follow-up study on the exploration of the use of SPORK within Bayesian retrievals of exoplanet atmospheres.

This research was supported by grant No. 2019-1403 from the Heising–Simons Foundation. K.C.R. would like to thank Andy Casey for the work in developing the software Spectroscopy Made Hard(er) from which SPORK is derived, and Isaac Malsky for insightful comments. M.B. acknowledges support from the STFC research grant ST/T000406/1.

ORCID iDs

Kaitlin C. Rasmussen  <https://orcid.org/0000-0002-0470-0800>
 Matteo Brogi  <https://orcid.org/0000-0002-7704-0153>
 Fahin Rahman  <https://orcid.org/0000-0001-7340-5912>
 Hayley Beltz  <https://orcid.org/0000-0002-6980-052X>
 Miles Currie  <https://orcid.org/0000-0003-3429-4142>
 Emily Rauscher  <https://orcid.org/0000-0003-3963-9672>
 Alexander P. Ji  <https://orcid.org/0000-0002-4863-8842>

References

- Alonso-Floriano, F. J., Sánchez-López, A., Snellen, I. A. G., et al. 2019, *A&A*, **621**, A74
 Barnes, J. R., Barman, T. S., Jones, H. R. A., et al. 2010, *MNRAS*, **401**, 445
 Beltz, H., Rauscher, E., Brogi, M., & Kempton, E. M. R. 2021, *AJ*, **161**, 1
 Birkby, J. L. 2018, arXiv:1806.04617
 Birkby, J. L., de Kok, R. J., Brogi, M., et al. 2013, *MNRAS*, **436**, L35
 Brogi, M., de Kok, R. J., Albrecht, S., et al. 2016, *ApJ*, **817**, 106
 Brogi, M., de Kok, R. J., Birkby, J. L., Schwarz, H., & Snellen, I. A. G. 2014, *A&A*, **565**, A124
 Brogi, M., & Line, M. R. 2019, *AJ*, **157**, 114
 Brogi, M., Snellen, I. A. G., de Kok, R. J., et al. 2012, *Natur*, **486**, 502
 Casey, A. R. 2014, PhD thesis, Australian National Univ.
 Charbonneau, D., Noyes, R. W., Korzennik, S. G., et al. 1999, *ApJL*, **522**, L145
 Collier Cameron, A., Horne, K., Penny, A., & James, D. 1999, *Natur*, **402**, 751
 Collier Cameron, Andrew, Horne, Keith, Penny, Alan, & Leigh, Christopher 2002, *MNRAS*, **330**, 187
 Cowan, N. B., Agol, E., & Charbonneau, D. 2007, *MNRAS*, **379**, 641
 de Kok, R. J., Brogi, M., Snellen, I. A. G., et al. 2013, *A&A*, **554**, A82
 Deming, D., Brown, T. M., Charbonneau, D., Harrington, J., & Richardson, L. J. 2005, *ApJ*, **622**, 1149
 Deming, D., Wilkins, A., McCullough, P., et al. 2013, *ApJ*, **774**, 95
 Flowers, E., Brogi, M., Rauscher, E., Kempton, E. M.-R., & Chiavassa, A. 2019, *AJ*, **157**, 209
 Freudling, W., Romaniello, M., Bramich, D. M., et al. 2013, *A&A*, **559**, A96
 Gandhi, S., Madhusudhan, N., Hawker, G., & Piette, A. 2019, *AJ*, **158**, 228
 Giacobbe, P., Brogi, M., Gandhi, S., et al. 2021, *Natur*, **592**, 205
 Hoeijmakers, H. J., Ehrenreich, D., Heng, K., et al. 2018, *Natur*, **560**, 453
 Leigh, C., Collier, A. C., Horne, K., Penny, A., & James, D. 2003, *MNRAS*, **344**, 1271
 Lockwood, A. C., Johnson, J. A., Bender, C. F., et al. 2014, *ApJL*, **783**, L29
 Mandell, A. M., Drake Deming, L., Blake, G. A., et al. 2011, *ApJ*, **728**, 18
 Nugroho, S. K., Kawahara, H., Masuda, K., et al. 2017, *AJ*, **154**, 221
 Piskorz, D., Benneke, B., Crockett, N. R., et al. 2016, *ApJ*, **832**, 131
 Piskorz, D., Benneke, B., Crockett, N. R., et al. 2017, *AJ*, **154**, 78
 Redfield, S., Endl, M., Cochran, W. D., & Koesterke, L. 2008, *ApJL*, **673**, L87
 Rodler, F., Kürster, M., & Henning, T. 2008, *A&A*, **485**, 859
 Rothman, L. S., Gordon, I. E., Barber, R. J., et al. 2010, *J. Quant. Spec. Radiat. Transf.*, **111**, 2139
 Sánchez-López, A., Alonso-Floriano, F. J., López-Puertas, M., et al. 2019, *A&A*, **630**, A53
 Schwarz, H., Brogi, M., de Kok, R., Birkby, J., & Snellen, I. 2015, *A&A*, **576**, A111
 Snellen, I. A. G., Albrecht, S., de Mooij, E. J. W., & Le Poole, R. S. 2008, *A&A*, **487**, 357
 Snellen, I. A. G., de Kok, R. J., de Mooij, E. J. W., & Albrecht, S. 2010, *Natur*, **465**, 1049
 Southworth, J. 2010, *MNRAS*, **408**, 1689
 Tamuz, O., Mazeh, T., & Zucker, S. 2005, *MNRAS*, **356**, 1466
 Wang, Ji, & Ford, Eric B. 2011, *MNRAS*, **418**, 1822
 Webb, R. K., Brogi, M., Gandhi, S., et al. 2020, *MNRAS*, **494**, 108
 Wiedemann, G., Deming, D., & Bjoraker, G. 2001, *ApJ*, **546**, 1068
 Wittenmyer, R. A., Endl, M., Cochran, W. D., & Levison, H. F. 2007, *AJ*, **134**, 1276
 Zellem, R. T., Griffith, C. A., Deroo, P., Swain, M. R., & Waldmann, I. P. 2014a, *ApJ*, **796**, 48
 Zellem, R. T., Lewis, N. K., Knutson, H. A., et al. 2014b, *ApJ*, **790**, 53
 Zhang, J., Kempton, E. M. R., & Rauscher, E. 2017, *ApJ*, **851**, 84