The Influence of Stellar Contamination on the Interpretation of Near-infrared Transmission Spectra of Sub-Neptune Worlds around M-dwarfs

Aishwarya R. Iyer and Michael R. Line
School of Earth and Space Exploration Arizona State University 525 E. University Dr., Tempe AZ 85281, USA; aiyer13@asu.edu
Received 2019 October 10; revised 2019 December 6; accepted 2019 December 10; published 2020 January 28

Abstract

The impact of unocculted stellar surface heterogeneities in the form of cool spots and hot faculae on the spectrum of a transiting planet has been a daunting problem for the characterization of exoplanet atmospheres. The wavelength-dependent nature of stellar surface heterogeneities imprinting their signatures on planetary transmission spectra are of concern particularly for systems of sub-Neptunes orbiting M-dwarfs. Here we present a systematic exploration of the impact of this spot-contamination on simulated near-infrared transmission spectra of sub-Neptune planets. From our analysis, we find that improper correction for stellar surface heterogeneities on transmission spectra can lead to significant bias when inferring planetary atmospheric properties. However, this bias is negligible for lower fractions of heterogeneities ($<1\%$). Additionally, we find that acquiring a priori knowledge of stellar heterogeneities does not improve precision in constraining planetary parameters if the heterogeneities are appropriately marginalized within a retrieval; however, these are conditional on our confidence of stellar atmospheric models being accurate representations of the true photosphere. In sum, to acquire unbiased constraints when characterizing planetary atmospheres with the *James Webb Space Telescope*, we recommend performing joint retrievals of both the disk-integrated spectrum of the star and the stellar-contamination-corrected transmission spectrum.

Unified Astronomy Thesaurus concepts: Exoplanet atmospheres (487); M dwarf stars (982); Hot Neptunes (754); Spectroscopy (1558); Radiative transfer (1335); Bayesian statistics (1900); Exoplanet atmospheric composition (2021); Stellar activity (1580); Starspots (1572)

1. Introduction

Transit spectroscopy of planets around small, cool stars offers a promising opportunity to characterize small, cool worlds (e.g., Seager & Sasselov 2000; Swain et al. 2008; Huitson et al. 2012; Gillon et al. 2016, 2017; Luger et al. 2017). An important consideration in characterizing such planets is the nature of the host star (Apai et al. 2018). Transmission spectroscopy of hot Jupiters to terrestrial worlds alike have demonstrated evidence of stellar activity (Pont et al. 2007; Alonso et al. 2008; Zhang et al. 2018; Rackham et al. 2017). M-dwarf hosts, in particular, exhibit higher recorded detections of stellar activity in the form of flares (Hawley & Pettersen 1991; Hawley et al. 2014; Schmidt et al. 2014) and H$_\alpha$ chromospheric emissions (Delfosse et al. 1998) relative to F-, G-, and K-type host stars, indicative of frequent occurrence of photospheric heterogeneities such as stellar spots and faculae. Study of M-dwarf systems are extremely timely, especially with the recent launch of NASA's *Transiting Exoplanet Survey Satellite*, which will discover thousands of exoplanets around such stars (Muirhead & Veyette 2018; Ballard 2019), providing the need for follow up with future missions like the *James Webb Space Telescope* (*JWST*).

Stellar spots that are occulted by a transiting planet contribute to observable changes in the transit light curve. Techniques to model the occulting heterogeneity have been successful to some degree by either excluding the spot-crossing event from the transit fit (Pont et al. 2008; Carter et al. 2011; Narita et al. 2013) or by correcting for epoch-to-epoch variations in the star from relative offsets between observations, therefore accounting for any differences in the planetary spectrum (Zellem et al. 2017). These techniques, however, do not account for homogeneously scattered regions of

heterogeneities that are persistent with the entire data in all observations.

Unocculted stellar spots as well as homogeneously covered regions of smaller spots affect the transit baseline in addition to causing discrepancies in the transit depth measurements depending on the fractional coverage in the area of the spots and their temperature contrast relative to the photosphere (Pont et al. 2008; Silva-Valio 2008; Czesla et al. 2009; Wolter et al. 2009; Agol et al. 2010; Berta et al. 2011; Carter et al. 2011; Désert et al. 2011; Sing et al. 2011; Fraine et al. 2014; McCullough et al. 2014; Oshagh et al. 2014; Barstow et al. 2015b; Damasso et al. 2015; Zellem et al. 2015; Rackham et al. 2017). Attempts to understand the nature of unocculted spots have thus far been done with photometric monitoring of the star, i.e., tracking the rotational variability in the modulating brightness on the stellar photosphere translating onto the light curve (Pont et al. 2008, 2013; Berta et al. 2011; Désert et al. 2011; Sing et al. 2011; Knutson et al. 2012; Zellem et al. 2015). Despite reasonable success, efforts with photometric monitoring of unocculted stellar spots have been inadequate in teasing out underlying heterogeneities spread across the entire disk of the star, especially those that are not significant enough to show variability effects on the stellar light curve (Jackson & Jeffries 2012; Rackham et al. 2017, 2018). Studies by Chapman (1987) and Shapiro et al. (2014) have also shown that every differential region of the star's photosphere, whether the transit chord or the entire disk, has its own unique spectrum. Given these challenges, the effect of unocculted stellar spots remains a persistent problem despite corrections from photometric variability monitoring, which at best provide only a lower limit estimate for regions of heterogeneities on the photosphere (Rackham et al. 2018).

Several recent studies (Pont et al. 2013; McCullough et al. 2014; Oshagh et al. 2014; Barstow et al. 2015b, 2015c; Rackham et al. 2017, 2018; Barstow 2018; Bruno et al. 2018; Murgas et al. 2019; Pinhas et al. 2018; Zhang et al. 2018) have demonstrated that unoccolted as well as homogeneous spots and faculae on the host star present themselves in a wavelength-dependent manner in the planetary transit data. The wavelength dependence of stellar heterogeneity signatures is particularly influential for transit spectroscopy where these features can often be degenerate with planetary atmosphere features. This necessitates accounting for stellar heterogeneities in order to probe the true nature of a planetary atmosphere. Improper characterization of the effects of stellar surface heterogeneities on the transmission spectrum limits our ability to place precision constraints on basic planetary conditions like temperature structure, composition, and cloud properties. Therefore, it is vital to thoroughly understand the influence of stellar surface heterogeneities on our ability to infer basic planetary properties in the era of high-fidelity observations anticipated from the *JWST*.

To understand these degeneracies and the impact of not accounting for unoccolted star spots on atmospheric inference, we leverage the utility of atmospheric retrieval. In Section 2, we briefly explore the effect of stellar heterogeneities on the shape of a planetary transmission spectrum analytically, providing intuition for degeneracies that might occur in interpretations of transit spectra, followed by a description of our atmospheric retrieval model. In Section 3 we investigate the level of bias that arises while inferring planetary properties from *JWST* transmission spectra when ignoring corrections for stellar heterogeneities. Additionally, we address the question of whether a priori knowledge of stellar heterogeneity, given the current state of M-dwarf stellar models, could improve precision on the inferred planetary characteristics from *JWST* transmission spectra. In this section, we also investigate the potential biases arising directly from the stellar models when inferring planetary properties. Section 4 presents a discussion on various sources of bias that can affect appropriate characterization of the transmission spectrum in addition to the stellar contamination correction, followed by a brief summary of our findings. In sum, our work demonstrates that, regardless of the coverage area or temperature contrast of a given stellar heterogeneity feature, one can infer unbiased planetary properties retrieved from contaminated transmission spectra provided an appropriate correction for stellar heterogeneity has been incorporated; however, these are conditional on our confidence of stellar atmospheric models being accurate representations of the true photosphere.

2. Anticipated Degeneracies

Here we briefly build our intuition for how the effects of stellar contamination can alter the shape of the planetary transmission spectrum. The transmission spectrum corrected for stellar surface heterogeneities is given by

$$\alpha_{\lambda,c} = \epsilon_\lambda \alpha_\lambda \tag{1}$$

where α_λ is the native planetary transit spectrum and ϵ_λ is the contamination factor (McCullough et al. 2014; Zellem et al. 2017; Rackham et al. 2018, see their Equations (1) and (2))

given by

$$\epsilon_\lambda = \frac{1}{(1 - f_s(1 - Q_{\lambda,s}))} \tag{2}$$

where $Q_{\lambda,s}$ is the spot-to-photospheric stellar flux ratio $(F_{\lambda,s}/F_{\lambda,p})$ and f_s is the fractional coverage of spots. The shape of the transmission spectrum (when corrected for stellar heterogeneities) is given by the derivative of Equation (1) combined with the expressions given in Line & Parmentier (2016; see their Equations (6)–(9) for a gas+gray cloud system):

$$\frac{d\alpha_{\lambda,c}}{d\lambda} = \epsilon_\lambda \frac{2R_p}{R_*^2} H\left(\frac{1}{1+\gamma_\lambda}\right)\frac{d\ln(\sigma_\lambda)}{d\lambda} + \frac{d\epsilon_\lambda}{d\lambda}\left(\frac{Rp^2}{R_*^2} + \frac{2R_p z_\lambda}{R_*^2}\right) \tag{3}$$

where R_p and R_* are the radius of the planet and star respectively, z_λ is the wavelength-dependent sharp occulting disk radius, H is the planetary atmospheric scale height, σ_λ is the absorption cross-section for a given gas, and γ_λ is the gas-to-gray cloud opacity ratio.

The shape of the stellar contamination spectrum, which is simply the wavelength-dependent derivative of the contamination factor (Equation (2)), is given by

$$\frac{d\epsilon_\lambda}{d\lambda} = \frac{\left(f_s \frac{dQ_{\lambda,s}}{d\lambda}\right)}{(1 - f_s + f_s Q_{\lambda,s})^2}. \tag{4}$$

Simplistically assuming that the stellar fluxes behave like blackbodies and also that the temperature contrast between the photosphere and spot is small, consistent with Doppler imaging observations for M-dwarfs (Strassmeier 2010; Andersen & Korhonen 2015), i.e., $T_p \approx T_s$, therefore we can write

$$Q_{\lambda,s} \approx 1 + \frac{hc}{\lambda k_B}\left(\frac{1}{T_p} - \frac{1}{T_s}\right) \tag{5}$$

$$\frac{dQ_{\lambda,s}}{d\lambda} \approx -\frac{hc}{\lambda^2 k_B}\left(\frac{1}{T_p} - \frac{1}{T_s}\right). \tag{6}$$

At longer wavelengths $dQ_{\lambda,s}/d\lambda$ (as well as Equation (4)) tends to zero with ϵ_λ approaching a constant "stretching" term, thus nulling the second term in Equation (3). From the surviving first term, we can see clear degeneracies between the stellar contamination ϵ_λ and the scale height H which both act like "stretching" terms on the opacity structure; e.g., an unaccounted stretch due to the spot could be compensated with scale height, leading to a bias in the unknown terminator temperature and/or molecular weight.

At shorter wavelengths, on the other hand, both terms in Equation (3) survive, resulting in a complex shape interaction. The scaling behavior from the first term persists with a power-law-like behavior from the second term, yielding additional degeneracies with z_λ—the wavelength-dependent radius alongside the factors that contribute to the shape of the stellar contamination spectrum, $\frac{dQ_\lambda}{d\lambda}$ i.e., the temperature contrast between T_p and T_s and fractional coverage in area of spots, f_s.
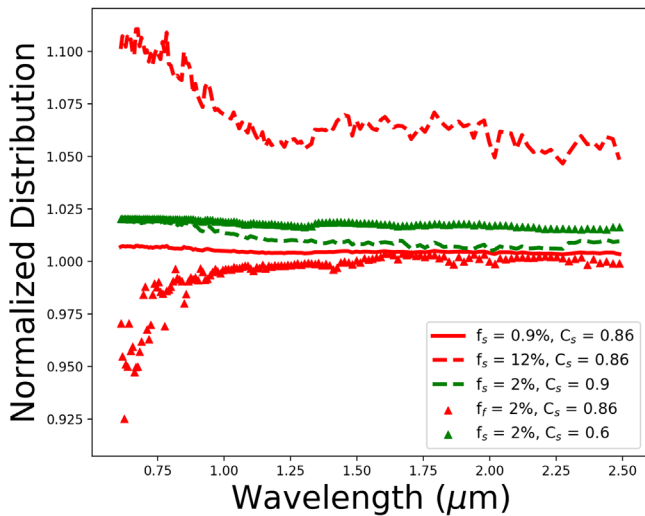
Figure 1. Stellar contamination spectra as a function of wavelength modeled (using PHOENIX models from Allard et al. 2003, 2007 and Allard & Freytag 2009) for an M-dwarf with photosphere temperature $T_p = 3300$ K with varying cases of surface heterogeneities (following Equation (2)). The contamination spectra in red are for varying fractional coverage (f_s) in the heterogeneity area on the photosphere at a fixed temperature contrast ($C_s = T_s/T_p$ of 0.86). The contamination spectra in green are for changing spot temperature contrast at a constant f_s (2%). Small values of f_s produce a contamination spectrum close to 1 (e.g., no contamination), whereas a large f_s results in a power-law-like effect at shorter wavelengths. Faculae cause a negative power-law-like behavior at shorter wavelengths. For a fixed f_s, increasing the temperature contrast alone causes an increased stretching effect at longer wavelengths.

Figure 1 shows example contamination spectra for both spots and faculae with various temperature contrasts and covering fractions (e.g., similar to Rackham et al. 2017 and Zhang et al. 2018) for realistic stellar spectra (using PHOENIX models from Allard et al. 2003, 2007 and Allard & Freytag 2009). Consistent with the preceding analytical result, with realistic spectra we also see that increasing the spot-covering fraction f_s from a small 0.9% to a large 12% value produces a scaling effect at longer wavelengths. Moreover, at shorter wavelengths, we see a power-law (or slope) effect, similar to what may be expected from scattering in the sub-Neptune atmosphere, as also demonstrated for the case of hot Jupiter HD 189733b (McCullough et al. 2014).

Now, armed with a first-order intuition for the effect of stellar contamination on the transmission spectrum, we can explore numerically, through retrievals, the influence of spot contamination.

3. Exploring the Impact of Unoccluded Star Spots on Retrieved Atmospheric Properties

In this section we aim to explore the role of spot contamination by addressing the following questions.

1. Will there be a bias in *JWST* results if we do not account for stellar contamination in atmospheric retrieval models? How can we rigorously explore this bias to decode degeneracies that arise in transmission spectra?
2. Does a priori knowledge of stellar contamination parameters reduce bias in our interpretations?

We first describe our retrieval setup, followed by a numerical exploration addressing each of the aforementioned questions.

3.1. Model Description

We use the planetary transmission/retrieval tool CHIMERA[1] (Line et al. 2013; Swain et al. 2014; Kreidberg et al. 2014, 2015; Mai & Line 2019) to simulate and retrieve upon *JWST*-like spectra from 0.6 to 5 μm (e.g., NIRISS, NIRCam, or NIRSpec). The specific model used here is from Mai & Line (2019) but with modifications to accommodate for the transit light source effect (Apai et al. 2018; Rackham et al. 2018) arising from unoccluded starspots (e.g., Equation (2)). The relevant model variables are given in Table 1. Briefly, the model[2] computes a limb transmission spectrum via the relations in Brown et al. (2001) and Tinetti et al. (2012). We include correlated-K opacities (derived from the line-by-line cross-sections of Freedman et al. 2008, 2014) for H_2–H_2/He collision-induced absorption, H_2O, CH_4, CO, CO_2, NH_3, Na, K, TiO, VO, C_2H_2, HCN, H_2S, and FeH. The retrieval uses the "chemically consistent" approach (Kreidberg et al. 2015, 2018) given a metallicity and C/O assuming equilibrium gas+condensate chemistry (computed using the NASA Chemical Equilibrium with Applications; Gordon & McBride 1996), computed along the three-parameter analytic T–P profile (Guillot 2010; Line et al. 2013). Hazes and clouds are parameterized with a simple power law (Etangs et al. 2008) and gray uniform opacity, respectively.

We use a grid of PHOENIX (Allard et al. 2003, 2007; Allard & Freytag 2009) stellar models (from the STScI pysynphot routine; STScI Development Team 2013) in effective temperature to construct the photospheric as well as spot/facular region spectrum (fixing the stellar metallicity to solar and log (g) to 5.0). Within the forward model/retrieval, the photosphere/spot spectra are interpolated to the appropriate effective temperature using the Scipy package griddata (Virtanen et al. 2019).

With this modeling framework, we investigate the influence of stellar contamination on a generic M-dwarf-orbiting sub-Neptune (a clear 40× solar, and a cloudy 300× solar atmosphere; Table 1) observed with a *JWST*-like platform from 0.6 to 5 μm (broken up between NIRISS-like: 0.6–2.5 μm and NIRCam/NIRSpec G395-like: 2.4–5.0 μm.). In all cases the resolving power (R) is assumed at 100 and, for simplicity, we assume a wavelength-independent uncertainty of 30 ppm per bin, loosely motivated by reasonable *JWST* performance expectations (e.g., Greene et al. 2016). We do not randomize the data-points so as to remove any bias due to arbitrary noise instantiations (e.g., Feng et al. 2018 and Mai & Line 2019). Parameter estimation and model selection (used to assess the need for spot-correction) are performed with the PyMultiNest (Feroz et al. 2009; Buchner et al. 2014) nested sampling routine (Skilling et al. 2006) with uniform priors, given in Table 1.

Figure 2 illustrates the influence of unoccluded spot-contamination on sub-Neptune transmission spectra. Consistent with recent studies (Pinhas et al. 2018; Rackham et al. 2018), there can be a significant shape change in the transmission spectrum, where the wavelength-dependent stellar contamination parameters could mimic atmospheric effects on the planet's spectrum (e.g., molecular features and haze-like power-law slopes).

Table 1
Model Parameters, Nominal "Truth" Values, and Uniform Prior Ranges

Parameter	Description	Initial Value	Status	Prior Range
$\log(\kappa_{ir})$[a]	Thermal profile gray IR opacity (cm^2 g^{-1})	0.306	Fixed	N/A
$\log(\gamma_v)$[a]	Thermal profile Vis/IR opacity	-2.02	Fixed	N/A
R_p[b]	Adopted 10 bar planetary radius (R_{jup})	0.239	Fixed	N/A
M_p[b]	Planetary mass (M_{jup})	0.0204	Fixed	N/A
R_*[b]	Stellar radius ($R_\odot$)	0.211	Fixed	N/A
T_{irr}[a]	Irradiation temperature (K)	794.0	Retrieved	[300, 3000]
[M/H]	Planetary metallicity	1.61 (40×) or 2.47 (300×)	Retrieved	$[-2, 3]$
$\log(C/O)$	C-to-O ratio	-0.72	Retrieved	$[-2, 2]$
×Rp	10 bar radius scaling	1.0	Retrieved	[0.5, 1.5]
$\log(\sigma_0)$[c]	Haze cross-section amplitude	-10.0	Retrieved	$[-15, 2]$
β[c]	Haze scattering slope	4.	Retrieved	[0, 6]
$\log(\kappa_c)$	Well-mixed gray cloud opacity	-35.5 (clear) or -29 (cloudy)	Retrieved	$[-40, -20]$
T_p	Stellar photosphere temperature (K)	3300	Retrieved	[2000, 4000]
T_s	Stellar spot temperature (K)	2838	Retrieved	[2000, 4000]
f_s	Fractional coverage of spots	Case specific[d]	Retrieved	[0, 1]
T_f	Stellar faculae temperature (K)	3400	Retrieved	[2000, 4000]
f_f	Fractional coverage of faculae	Case specific[d]	Retrieved	[0, 1]

Notes.
[a] Three-parameter model for thermal profile (Guillot 2010).
[b] GJ1214/b, from exoplanets.org (Han et al. 2014).
[c] From Etangs et al. (2008). Haze cross-section given by $\sigma_\lambda = \sigma_0(\lambda_0/\lambda)^\beta$ with $\lambda_0 = 0.43$ μm and σ_0 given in units relative to H$_2$ Rayleigh scattering (2×10^{-27} cm^2).
[d] Initial values are specific to cases of spot- and faculae-covering fractions, listed in Table 2, also given in Rackham et al. (2018) and Zhang et al. (2018).

3.2. To What Degree Is There Bias if Not Accounting for Unoccluded Star Spots?

3.2.1. JWST NIRISS-like Bandpass

We choose spot/faculae-covering fractions and contrasts to be consistent with 1% I-band rotational variability for mid-to-late M-dwarfs (Newton et al. 2016; Rackham et al. 2018). Specifically, We have assumed $T_s = 0.86 \times T_p$ K (or $C_s = 0.86$) and $T_f = T_p + 100$ K (or $C_f = (T_p+100)/T_p$) (Gondoin 2008; Afram & Berdyugina 2015; Rackham et al. 2018). We also explore a range of planetary atmospheric conditions (e.g., clear, cloudy, high metallicity (300× solar)/molecular weight) given in Table 2 and shown in Figure 2.

For each scenario in Table 2 (as observed from 0.6 to 2.5 μm, $R = 100$, with 30 ppm precision) we perform a series of retrievals: first, a control case that assumes a homogeneous stellar photosphere (no spot contamination correction, retrieves only the atmospheric properties T_{irr}, [M/H], $\log(C/O)$, $\log(\sigma_0)$, β, and $\log(\kappa_c)$) and, second, accounting for the spot contamination (e.g., additionally retrieves for T_p, T_s, f_s, T_f, and f_f). Two figures of merit are used when assessing the level of "bias" in the inferred planetary properties. The first (Figures 3 and 4) is a qualitative metric describing the bias by simply measuring the actual deviation of the posterior probability distributions (medians of the distributions) from the truth values. We only report the bias in terms of the width of the posterior probability distributions (i.e., shift of the histogram in units of σ width relative to the retrieved precision) for a select few cases with distributions lying fully within the chosen prior range. Instances where the posterior histograms converge partly on the edge of the priors skew the true level of bias by falsely portraying additional width to the histograms; therefore, this is purely a "qualitative metric" to measure the bias in the inferences (e.g., see Figure 3, right column for high metallicity and cloudy planetary atmosphere with $f_s = 12\%$ for the carbon-

to-oxygen ratio: $\log(C/O)$). The second metric (discussed below) is a quantitative measure of the level of bias in the inferred planetary parameters, accounted for by computing the Bayes factor (Table 2, last column) between the atmosphere-only and the atmosphere+spot model retrieval results for all cases with varying levels of stellar heterogeneities.

In all cases (see Table 2)[3] with a JWST NIRISS-like transmission spectrum (covering 0.6–2.5 μm, with 30 ppm precision) of a generic sub-Neptune orbiting an M-dwarf, we demonstrate that there is significant "qualitative bias" in the retrieved atmospheric properties when not correcting for stellar contamination. For instance, in the case with a spot-covering fraction $f_s = 0.9\%$, an uncorrected transmission spectrum for a clear atmosphere sub-Neptune yields a bias in the retrieved planetary irradiation temperature (T_{irr}), bulk metallicity ([M/H]), and the bulk carbon-to-oxygen ratio ($\log(C/O)$) of 1.6σ, 13.7σ, and 95.5σ respectively (illustrated in Figures 3 and 4). Moreover, in Figure 4, a correlation between the planetary radius (fiducial radius at 10 bar × R_p) and the stellar spot-covering fraction (f_s) emerges from the retrieval 2D posterior histograms. Consistent with the previously illustrated analytical result (Equation (3)), the wavelength-dependent stellar contamination term is degenerate with both the planetary atmospheric scale height H as well as the wavelength-dependent sharp occulting disk radius z_λ, altogether affecting the shape of the transmission spectrum. We also see this effect in Figure 2, where increasing the level of stellar photospheric heterogeneity (a linear combination of the spot-covering fraction and the temperature contrast between the spot and

---

[3] As a sanity check, we also performed a similar exercise on an uncontaminated stellar photosphere to determine whether the introduction of the contamination parameterization within the retrieval would unduly influence the retrieved atmospheric properties. Unsurprisingly, we found that it did not, e.g., no "intrinsic" bias is introduced by adding in the contamination parameterization.
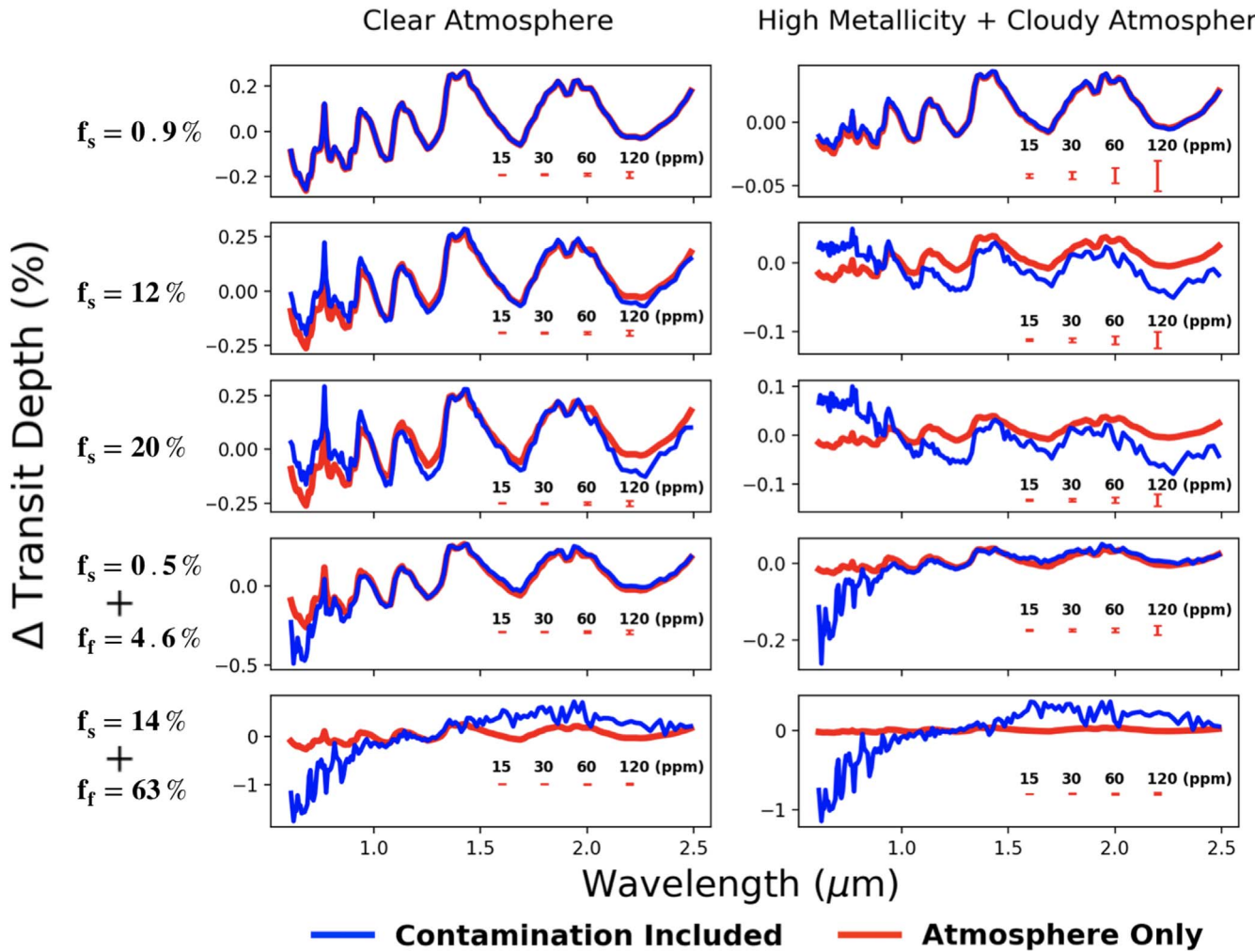
Figure 2. Transmission spectra resulting from including (blue) and ignoring (red) corrections for stellar photosphere with varying degrees of heterogeneities on the host M-dwarf (rows) under a solar composition atmosphere (left) and a cloudy, high-metallicity (300× solar) atmosphere (right). Representative spectral error bars (15–120 ppm) are shown in the corner of each panel. We see here that contamination covering fractions less than 1% do not produce a significant shape change in the transmission spectrum.

the immaculate photosphere) modifies the shape of the transmission spectrum and hence the derived planetary radius significantly, beyond anticipated spectral precisions (15–120 ppm). Focusing, however, only on 30 ppm precision as an example case, the trend of increasing bias with the first metric (deviance from truth) persists for a combination of stellar heterogeneities, ranging from 0.5% to 20% in spot-covering fraction and 4.6%–63% for faculae-covering fraction, as illustrated in Figure 3.

We also find that, for the smallest heterogeneity fractions, muting the transit spectral features with clouds and high metallicity reduces the effective "qualitative bias" as seen in the retrieved planetary parameters (see Figure 3, top right panel). For $f_s = 0.9\%$ in particular, we find that there is a bias of 0.48, 0.55, and 0.01σ for the inferred irradiation temperature, planetary metallicity, and carbon-to-oxygen ratio respectively, for an uncorrected transmission spectrum. With increasing stellar heterogeneity covering fractions, however (ranging from $f_s = 12\%$ to 20% alone and for $f_s = 0.5\%$ along with $f_f = 4.6\%$ up to 63%), the qualitative increase in the bias (deviation from the truth values) of the posterior probability distributions for cloudy and high-metallicity atmospheres (although slightly reduced) are consistent with those of a clear sub-Neptune atmosphere. Only for small spot-covering fractions do we find

that the level of bias falls well within 1σ of the true value, and this effect is persistent through a range of spectral precisions (15–120 ppm; see Figure 5, left column for $f_s = 0.6\%$). This is primarily because, as the transit feature signal-to-noise decreases (whether due to clouds and/or high mean molecular weight), thereby increasing degeneracy in the transmission spectrum, this results in larger parameter uncertainties. These uncertainties therefore seemingly reduce the effective bias under our first figure of merit, in turn motivating the question: at what degree of stellar heterogeneity in fractional coverage does the bias in retrievals matter?

To address this question, we start by qualifying our second figure of merit, the Bayes factor, which provides a quantitative measure of the need to include as "nuisance" parameters a stellar correction scheme (last column in Table 2, for 30 ppm precision). Specifically, we compute the log of the Bayes factor (ln B; Cornish & Littenberg 2007) which is a metric permitting quantitative comparison of two models with different complexity (e.g., with and without a spot-correction parameterization). For all cases in Table 2, we highlight two points illustrating the behavior of the Bayes factor: (1) ln B increases with an increase in contamination covering fraction and (2) it decreases with increasing cloudiness/molecular weight in the planetary

Table 2
Bayes Factor between the Contamination-included Model and the
Atmosphere-only Model

Stellar Contamination	Atmosphere Type	Bayes Factor (ln B)
NIRISS-like		
$f_s = 0.9\%$	Clear	10
	Cloudy	2.5
	High metallicity + cloudy	0.23
$f_s = 12\%$	Clear	3191
	Cloudy	2653
	High metallicity + cloudy	1036
$f_s = 20\%$	Clear	10,634
	Cloudy	6085
	High metallicity + cloudy	2479
$f_s = 0.5\%, f_f = 4.6\%$	Clear	14,634
	Cloudy	14,628
	High metallicity + cloudy	6389
$f_s = 14\%, f_f = 63\%$	Clear	1,575,792
	Cloudy	1,543,800
	High metallicity + cloudy	66,478
NIRCam-like		
$f_s = 14\%, f_f = 63\%$	Cloudy	63,295

Notes. Bayes Factor between an atmosphere-only model and a model that accounts for the spot contamination under the several atmosphere scenarios (Table 1) and a NIRISS-like (30 ppm, $R = 100$, 0.6–2.5 μm) and NIRCam-like (2.4–5 μm) observational setup.

transmission spectrum. The first point is consistent for all planet atmosphere scenarios, as seen previously with our first figure of merit (Figure 3). For instance, a clear atmosphere scenario with increasing spot-covering fraction f_s from 0.9 to 20% increases ln B from 10 to 10,634, with the entire range demonstrating strong evidence for preferring the model including stellar contamination as per the Jeffrey's scale (Trotta 2008). The increasing trend in ln B is similar through all atmospheric scenarios including for cases with cloudy and high-metallicity atmospheres, where the ln B ranges from 0.23 to 2479 for f_s increasing from 0.9% to 20% (see Table 2). This pattern is consistent for combinations of spot- and faculae-covering fractions both for clear as well as cloudy and high-metallicity atmospheres, thereby quantitatively indicating the degree to which the model including stellar contamination is preferred. The second point shows a decrease with increasing feature muting (due to clouds/metallicity). For the entire range of stellar contamination cases in Table 2 we see this behavior, especially for the case of a small spot-covering fraction of 0.9% (under the cloudy and high-metallicity atmosphere), which yields a ln $B = 0.23$, demonstrating inconclusive evidence favoring either model (true for all ln B values on the Jeffrey's scale <1, Trotta 2008). Despite the decreasing trend of ln B with increasing atmospheric degeneracies, for other spot- and spot+faculae-covering fraction combinations, the Bayes factors strongly prefer the model including stellar corrections to prevent bias in the inferences, i.e., ln $B > 5$. From this trend, we can say that, regardless of the planetary atmosphere properties explored here, a system with a host star of spot-covering fraction $>1\%$ must include a parameterization for stellar activity to mitigate the bias atmospheric inferences.

We now focus on quantifying the bias as a function of spectral precision, under the more likely scenario of a cloudy, high-metallicity atmosphere (Fortney et al. 2013). We do this by generating a sparse grid of retrievals—varying spot-covering fractions on the first dimension, ranging from 0.6% to 12% and spectral precision on the second dimension ranging from 15 to 120 ppm, consistent with expectations for the *JWST* based on achieved *Hubble Space Telescope*/Wide Field Camera 3 precisions (e.g., down to 15 ppm; Line et al. 2016). For each point on the grid, we perform two sets of retrievals: the first with the atmosphere-only model and the second with stellar corrections included. Our main findings from this exercise are: for the entire range of precisions, an increase in fractional coverage in area of spots (f_s) on the stellar photosphere increases the significance with which the model with stellar corrections is favored (ln B). Particularly for spot-covering fractions $f_s > 1\%$, we find that the Bayes factor ranges from moderately favoring the model with stellar contamination included (e.g., case of 120 ppm precision at $f_s = 3\%$ in Figure 6 is in the moderate evidence regime as per the Jeffrey's scale) to strongly favoring the model for all other sensitivities. This trend provides strong evidence for the spectral presence of stellar heterogeneities in the transmission spectrum throughout these levels of precision.

For cases of $f_s < 1\%$ and for all sensitivities >30 ppm on our grid (see Figure 6), we only see weak to moderate evidence for stellar contamination. This is consistent with the reduction in our "qualitative bias" metric (see Figure 3, top panel, right column and Figure 5, left column) as demonstrated previously, due to increasing the level of degeneracies in the spectrum for cloudy/high-metallicity scenarios, especially at small spot-covering fractions. For the case of 15 ppm precision, however, the uncertainties on the transmission spectrum are significantly small enough that even a region of heterogeneity as small as $f_s = 0.6\%$ shows a very high ln B, providing strong spectral evidence for the contamination, although this is not apparent from our "qualitative" bias metric as seen in the top left panel of Figure 5 for $f_s = 0.6\%$. Overall, we see that for spot-covering fractions $>1\%$, for our defined grid of precision levels, increasing f_s leads to an increase in the "quantitative bias" (ln B), providing strong evidence for stellar contamination; however, this increase is penalized by the level of uncertainties on the *JWST* transmission spectrum, thereby reducing the apparent "qualitative bias" (deviance from the truth) as we see in the right column of Figure 5.

3.2.2. JWST NIRCam/NIRSPEC G395-like Bandpass

In Section 2 we showed that stellar contamination has a stretching effect on the shape of the transmission spectrum at longer wavelengths, and was therefore degenerate with the scale height. In this section we explore the role of stellar contamination over the *JWST* NIRCam/NIRSPEC G395-like bandpass (2.4–5.0 μm, and again assuming a constant 30 ppm precision) for a cloudy sub-Neptune atmosphere. The 3–5 μm region is critical for determining the carbon inventory in planetary atmospheres since large CO, CO_2, and CH_4 features are present throughout. We explore a scenario with a 14% spot and 63% faculae fraction, under the same temperature contrast as above.

We find that when not including the contamination model parameters there is a deviation of 21.7, 197, and 16.9σ respectively from the truth values, for the metallicity, C/O, and irradiation temperature along with a log Bayes factor value of 63,295, which falls under the regime of strong evidence (Trotta 2008) for stellar contamination (Table 2, Figures 7 and 8). From the 2D posterior
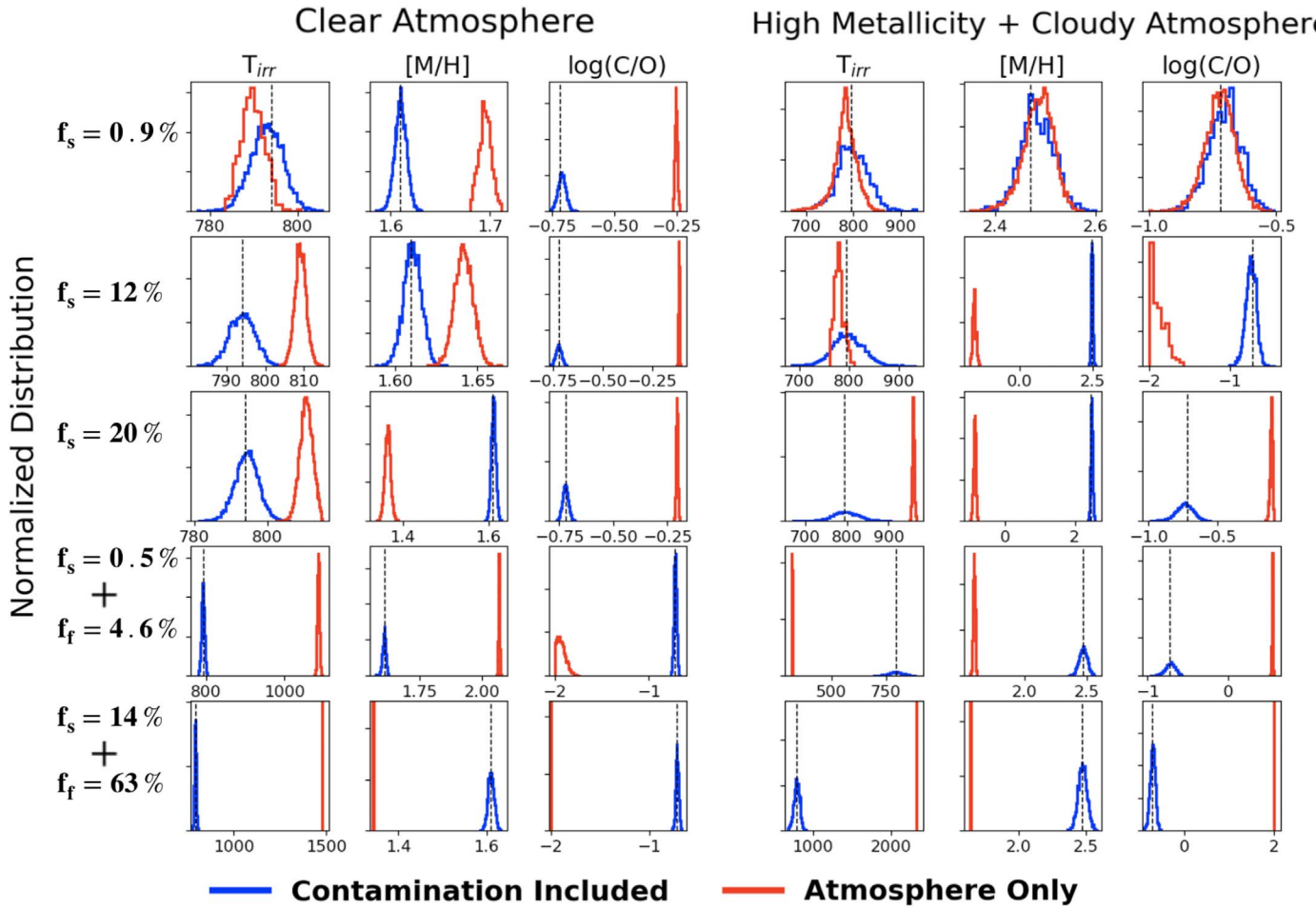
Figure 3. Constraints and biases on T_{irr}, [M/H], and log(C/O) retrieved from the transmission spectra (0.6–2.5 μm, $R = 100$, 30 ppm precision) of a clear (left) and cloudy, high-metallicity (right) sub-Neptune both accounting for (blue) and not accounting for (red) spot contamination for a spectrum constructed with contamination. Each row is labeled with the contamination properties used to construct the "true" model. The vertical dashed line indicates the "true" parameter values. An increasing degree of bias in the model that does not account for contamination (red) is seen as the spot/faculae-covering fraction increases (top-to-bottom). A significant bias is present in all clear-atmosphere scenarios, though covering fractions less than 1% result in little bias for a cloudy/high-metallicity scenario (right).

probability distributions (Figure 8), we also find noteable correlations that emerge between the spot-covering fraction (f_s) and the gray cloud opacity (log(κ_c)) as well as a correlation between log(C/O) and the spot-covering fraction. These degeneracies are consistent with our qualitative findings in Section 2 (Equation (3)), between cloud opacity and metallicity. We also note in Figure 7 that the stellar contamination spectrum is a major contributor to the variations in the shape of the transmission spectrum within this bandpass. These results suggest that stellar contamination can have a similarly large influence on longer wavelengths due to the presence of stellar molecular spectral features (our Figure 8; see also Rackham et al. 2017).

3.3. How Precisely Do We Need to Know the Stellar/Spot Properties to Remove Any Bias?

The natural next issue to arise in light of the influence of spot contamination on retrieved atmospheric properties is determining the level of required a priori spot contamination knowledge to mitigate the bias. In this section, we investigate whether having any prior knowledge of the level of stellar contamination on the host star's photosphere would mitigate atmospheric retrieval bias.

First, we again simulate a cloudy, high-metallicity sub-Neptune atmosphere with a 12% region of unoccluted spot as observed under the *JWST*–NIRISS-like bandpass with 30 ppm precision. We choose this spot-covering fraction as it falls under the regime that is sensitive to detect stellar contamination bias ($f_s > 1\%$) as demonstrated in Section 3.2 and Figure 6. For this exercise, we perform retrievals using two modeling approaches: (1) joint fit model, which simultaneously fits the stellar spectrum (assuming a linear combination of the the photosphere and spot spectra) and transmission spectrum for the contamination parameters under the assumption that the out-of-transit stellar photosphere has similar properties as the in-transit photosphere (e.g., Zhang et al. 2018), and (2) corrected transmission spectrum model, which only fits for the contamination parameters from the transmission spectrum itself (as in the above sections). We stress that we implicitly operate under the assumption that the stellar model is an accurate representation of an actual stellar spectrum, e.g., we assume no systematic "bias" in stellar model fits (however, we explore this assumption in Section 3.4 and Figure 10). We also assume that the absolute flux calibration uncertainty of a stellar spectrum is incorporated into our assumed photometric precision, which may not necessarily be true.
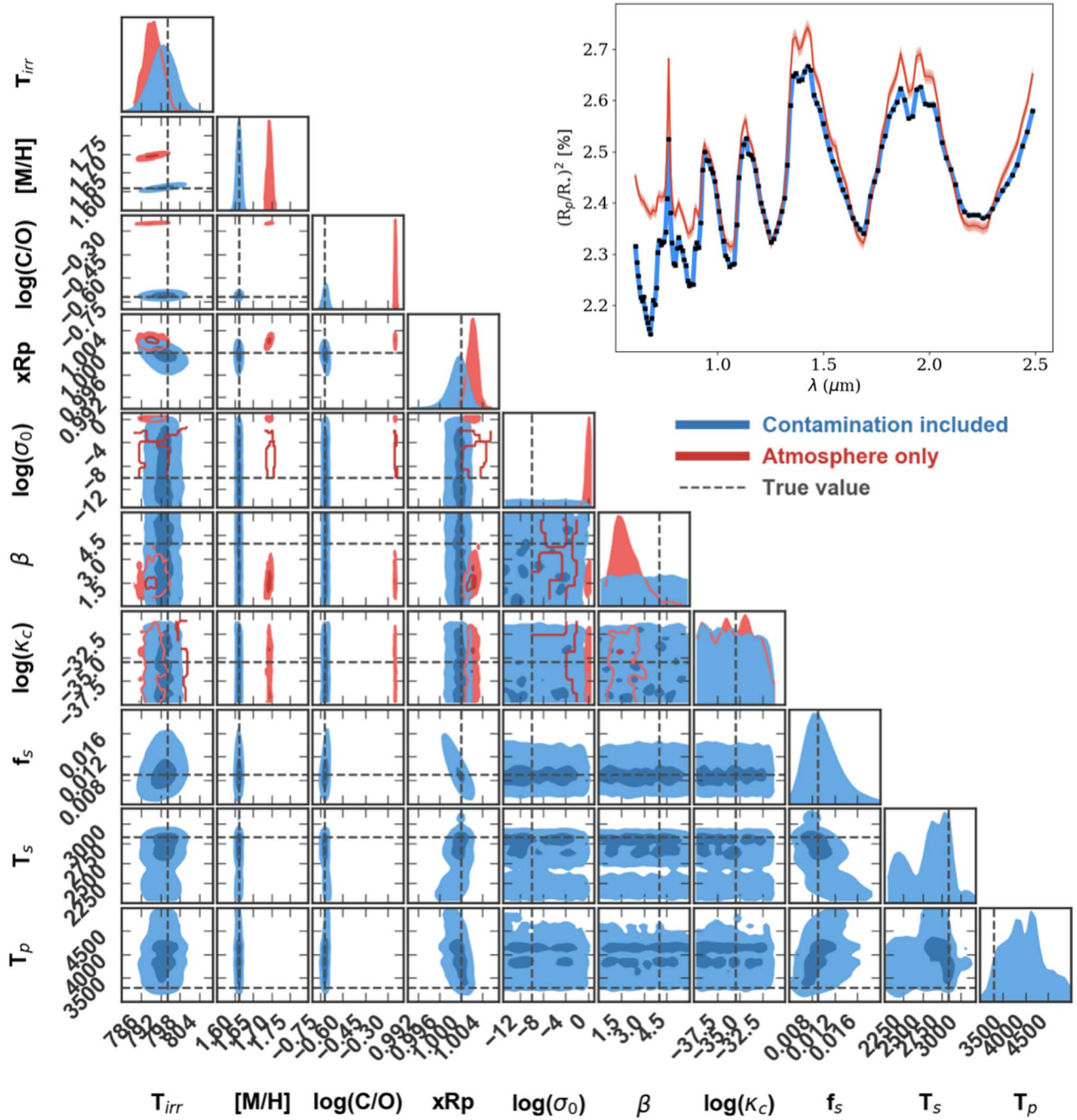
Figure 4. Corner plot summarizing the constraints and degeneracies from a NIRISS-like transmission spectrum (30 ppm sensitivity) for a clear, 40×solar composition sub-Neptune orbiting an M-dwarf with a 0.9% coverage (no faculae) when accounting for spot contamination (blue) and not accounting for it (red). There is significant bias in the inferred atmospheric properties (red vs. blue) as well as mismatch in the spectral fit when excluding the spot parameterization.

The posterior probability distributions of f_s (fractional coverage of the spot region on the photosphere), T_s (temperature of the spot region), and T_p (temperature of the photosphere) retrieved from the disk-integrated spectrum under both approaches are presented in Figure 9, top panel. We see here that the precision in the inferred stellar contamination parameters improves by several orders of magnitude for f_s, T_s, and T_p.

However, utilizing the stellar spectrum itself (the joint fit model) shows no statistically significant improvement in the atmospheric parameter precisions (i.e., we cannot reject the null hypothesis that both samples are drawn from the same population by performing a two-sample Kolmogorov–Smirnov test (Smirnov 1939); see Figure 9, bottom panel). This important result suggests, at least in this example, that simply including a parameterization for the contamination spectrum, and marginalizing over said parameters within a retrieval, is enough to "remedy" the stellar inhomogeneity problem; no need for a priori stellar contamination knowledge (e.g., as in Pinhas et al. 2018). This result, however, may only be true if the stellar models are accurate representations of the true photosphere, which we explore in the next section.
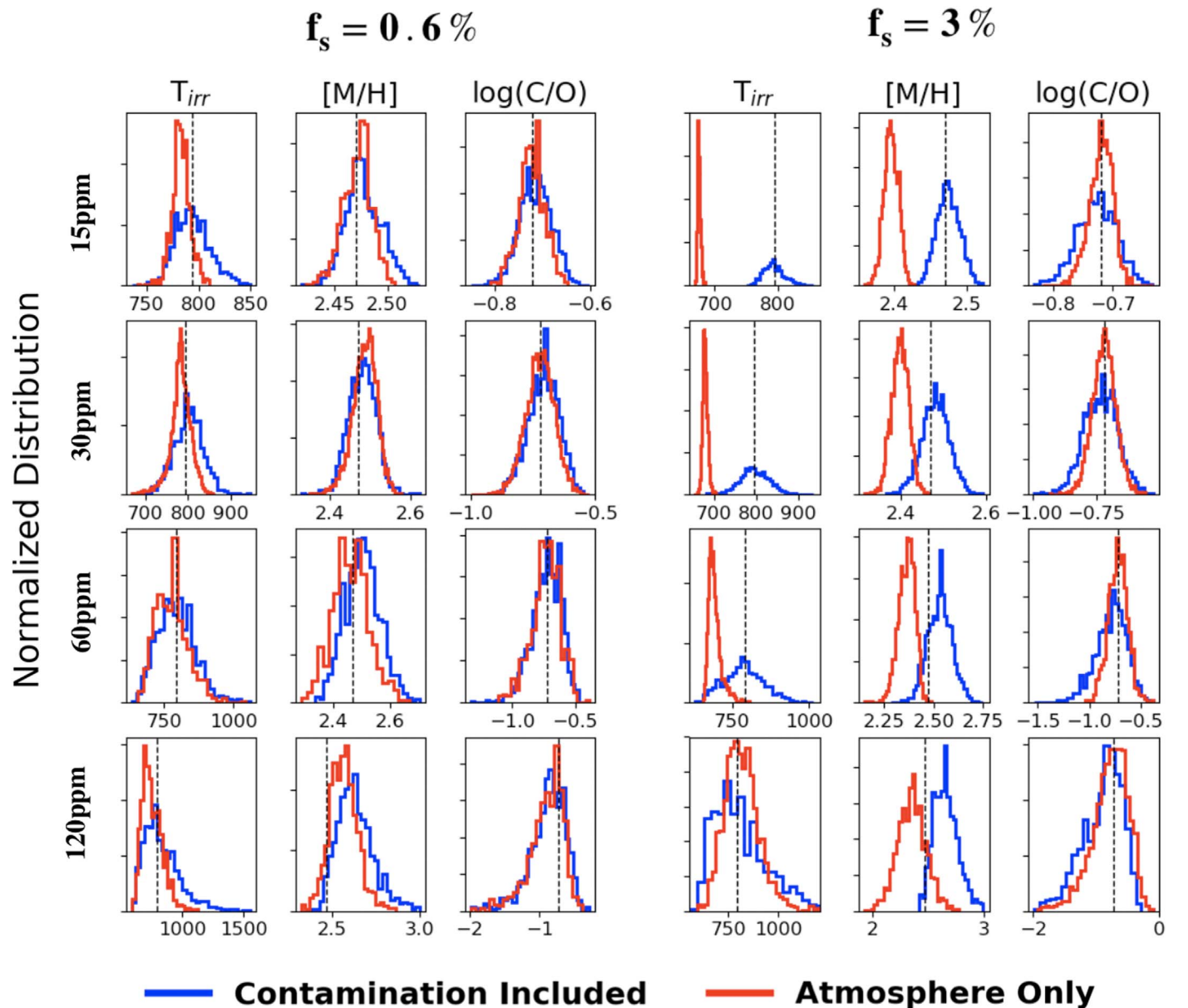
Figure 5. Constraints/bias on T_{irr}, [M/H], and log(C/O) as a function of spectral precision (15–120 ppm, rows) and two spot-covering fractions (0.6% left, 3% right) under the high-metallicity cloudy atmosphere setup, as observed with a NIRISS-like instrument. The red curves are from the atmosphere-only model (no spot correction) and the blue for the spot-retrieved model. Generally, increasing the spot-covering fraction increases the level of bias but overall bias decreases with larger spectrophotometric uncertainties on the transmission spectrum.

3.4. What if the Stellar Models are Inaccurate?

In the above analyses we assumed that the stellar models (for both the photosphere and spots) were accurate representations of a true contaminated stellar spectrum. However, model stellar spectra are not necessarily perfect in all situations (Veyette et al. 2017). Here, we explore the impact of an inaccurate stellar model on our ability to remedy spot contamination. We repeat the exercise of retrieving for planetary atmospheric properties of irradiation temperature, metallicity, and C/O from a contamination-corrected transmission spectrum with the same modeling approach as described in Section 3; however, we use a different stellar model to incorporate stellar contamination in our simulated data.

We again simulate a *JWST*–NIRISS-like transmission spectrum for a cloudy, high-metallicity sub-Neptune and

"mimic" potential stellar model deviations relative to a true stellar spectrum using a different stellar model. For the "true" stellar spectrum, we use the PHOENIX-ACES model (Husser et al. 2013, with $T_p = 3300$ K, $T_s = 2838$ K, $f_s = 12\%$, log(g) = 5.0, and solar metallicity) and for the retrievals, we fit with the older PHOENIX models (Allard et al. 2003, 2007; Allard & Freytag 2009) as illustrated above (see Figure 1). We are assuming that generational model differences are representative of model–stellar data differences (see Figure 10, top left panel). For this exercise, we assume a spot-covering fraction of 12% as this falls within the regime sensitive to detect bias in the planetary parameter retrievals ($f_s > 1\%$) as we illustrated in Figure 6.

The disk-integrated spectra from both our "true" star as well as the model star are presented in the top left panel of Figure 10. We also see a significant difference in the shape of
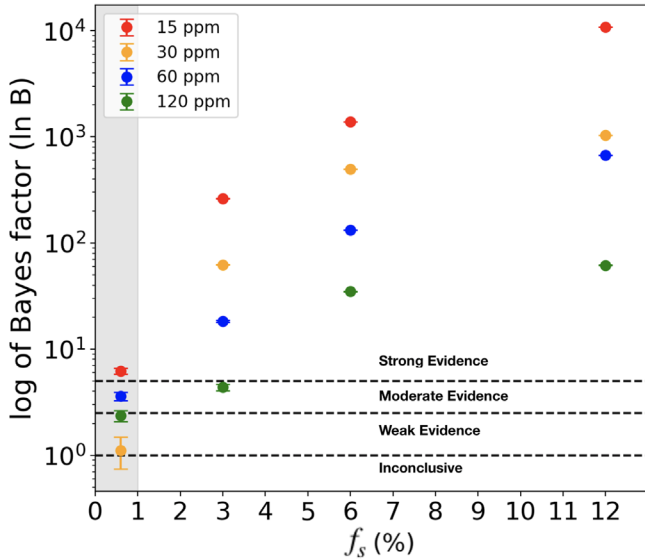
Figure 6. Bayes factor behavior between the atmosphere-only and atmosphere +spot correction model for a spot-contaminated cloudy, high-metallicity atmosphere as a function of spot-covering fraction (horizontal axis) and spectrophotometric precision over a NIRISS-like passband (colored dots). The horizontal black dotted lines indicate the degree to which the contamination-included model is preferred over the atmosphere-only model (Trotta 2008). Above 1%, the corrected model is preferred mostly with strong evidence for all precision levels with the exception of the 120 ppm error bar case for 3% coverage. Below 1% coverage the bias is not significant except for the 15 ppm precision, where the stellar correction model is always highly preferred. Note that the awkward crossing of the 30 ppm point below 1% is due to the inaccuracies in computing Bayes factors in low-evidence regimes (e.g., Lupu et al. 2016).
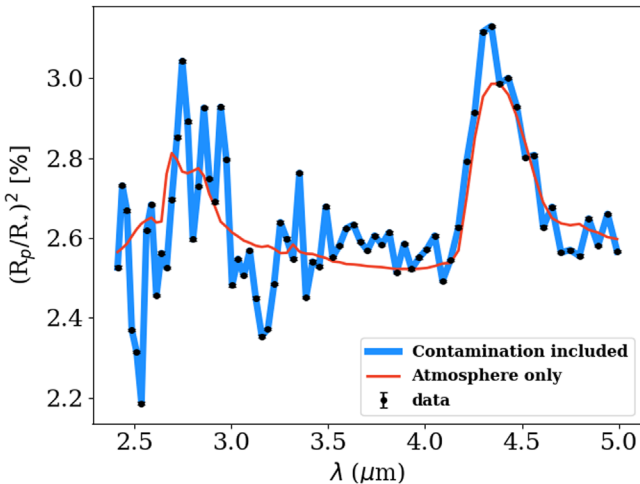
Figure 7. Spot-contamination influence over a NIRCam/NIRSpec G395-like bandpass with 30 ppm precision for a cloudy sub-Neptune atmosphere with spot- and faculae-covering fractions of 14% and 63%, respectively. The true spectrum is shown as the black points (errors are small), the best-fit model that includes the contamination parameterization is shown in blue, and the model that fails to account for contamination (atmosphere only) in red. Most of the "high-frequency" spectral features seen in this wavelength range under this particular spot/faculae setup are due to the stellar contamination. The strong 4.25 μm CO_2 feature at high metallicity stands above the stellar contamination "noise."

the same planetary transmission spectrum as expected from the differences translating from their respective contamination spectra (see Figure 10). In gray, we show the "best fit" to our simulated *JWST*–NIRISS transmission spectrum that includes correction for heterogeneity using the "true" stellar spectrum.

Overall, we see that the fit is quite reasonable in explaining the shape of the transmission spectrum; however, since the contamination retrieval model is missing the "true" stellar information (due to model versus measured stellar data differences), it appears to compensate for it by inducing incorrect variations in the planetary spectrum, indicative of false opacity signatures and incorrect terminator temperature estimates for the atmosphere. This is also evident from the posterior probability distributions on the right panel of Figure 10, where the irradiation temperature, metallicity, and C/O are significantly biased from the truth values of the planetary atmosphere.

With this important result we demonstrate that, while we can correct for stellar heterogeneity by simply parameterizing for the contamination spectra and marginalizing over them within a retrieval, additional uncertainties must be accounted for based on our confidence in the accuracy of stellar atmosphere models themselves.

4. Discussions and Conclusions

We have shown that stellar spot/faculae contamination will strongly influence our ability to retrieve atmospheric properties from *JWST* transmission spectra of sub-Neptune-like planets orbiting M-dwarfs. We find that, for a range of plausible near-infrared spectrophotometric precisions, there is significant bias when marginalizing over planetary parameters from the transmission spectrum if not correcting for these contamination sources appropriately. The bias appears to be less significant with decreased transmission feature signal-to-noise (e.g., muting due to clouds/metallicity). We also demonstrate that our ability to adequately marginalize over the stellar contamination parameters is only as good as the degree to which stellar photosphere models match reality.

Within the context of our work, there is very little knowledge of physically reasonable coverage fractions of stellar heterogeneities as well as their distributions and occurrence frequencies throughout the photosphere of an M-dwarf in the current literature. Priors for spot-covering fractions we used were derived from *I*-band variability of 1% (Newton et al. 2016; Rackham et al. 2018) or an order of 0.05 or less from variability in *G* and *R* photometric bands (Rockenfeller et al. 2006a, 2006b). Spectroscopic methods to measure TiO lines emerging from spot regions of M-dwarfs have produced dramatically different estimates for spot-covering fractions when compared to methods such as Doppler imaging, varying from ~50% to ~10% (O'Neal et al. 1998; Barnes et al. 2011). A significant source of uncertainty with photometric variability measurements is also the assumption that starspot patterns vary as a function of rotation (Browning et al. 2010; Barnes et al. 2011); however, there is only a weak correlation between rotation and activity for M-dwarfs according to Reiners & Basri (2010). Moreover, uncertainties associated with the frequency of the occurrence of homogeneously scattered stellar spots provide difficulties in performing multi-wavelength stitching of non-simultaneous observations (e.g., Barstow et al. 2015a and Bruno et al. 2018) as we could potentially expect changes in the stellar contamination spectra in a time-dependent manner.

Several uncertainties arise in determining accurate temperature contrast ratios between stellar photospheric and heterogeneity regions, as these have previously only been modeled assuming single spots and not for uniformly covered spot regions (Barnes et al. 2011; Afram & Berdyugina 2015).
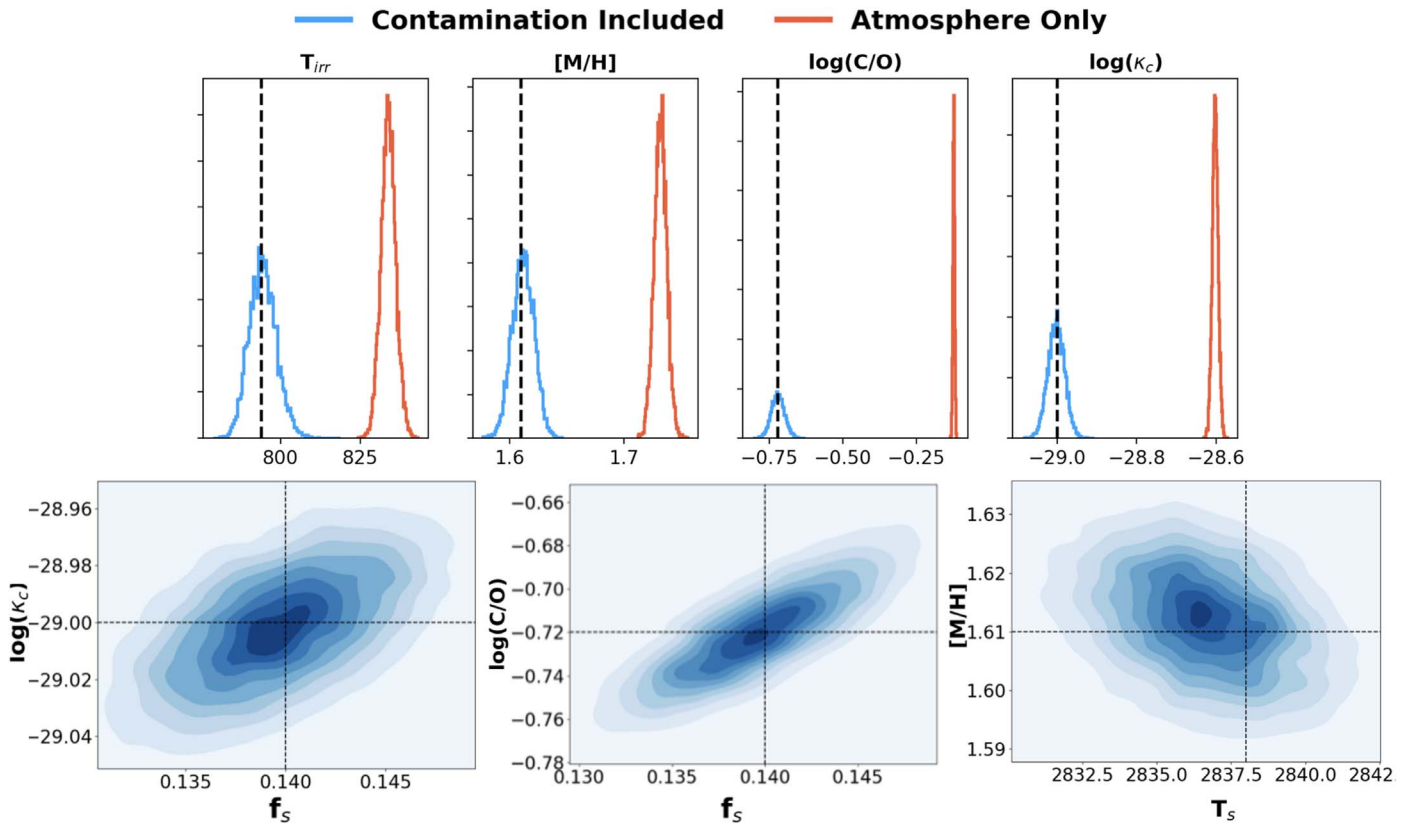
Figure 8. Retrieval results/biases for the irradiation temperature (T_{irr}), metallicity ([M/H]), carbon-to-oxygen ratio (log(C/O)), and cloud opacity (log(κ_c)) over the NIRCam/NIRSpec G395-like wavelength range (and 30 ppm precision) for a cloudy sub-Neptune orbiting a 14% spot, 63% faculae-contaminated M-dwarf under the atmosphere-only scenario. The top row illustrates the constraints and biases on the key atmospheric parameters under the atmosphere-only (red) and the contamination-correction-included scenario (blue). The bottom row shows the 2D histograms illustrating the prominent degeneracies between the spot parameters f_s and T_s with the atmospheric parameters.

However, Barnes et al. (2011) show a clear relation between stellar activity induced jitter in radial velocity observations as a function of wavelength, especially at low-temperature contrasts between the stellar photosphere and spot regions. Leveraging this relation could potentially provide one avenue for acquiring observational constraints on temperature contrasts between the photosphere and heterogeneity regions of M-dwarfs.

Numerous methods exist to estimate M-dwarf photospheric inhomogeneaities; however, it remains a challenging problem. The aim of this work is not to debate the presence of stellar contamination but to approach the problem from a "marginalization-in-retrievals" perspective. By doing so we found the following.

1. There are degeneracies between photospheric contamination properties and transiting planet atmosphere properties within the transmission spectra.

   In Section 2 we gained a qualitative intuition for the expected degeneracies. From the relations in Section 2 it is apparent that stellar contamination ϵ_λ and planetary scale height H are degenerate in that both result in a "stretching" effect in the spectrum at longer wavelengths. Ignoring stellar contamination therefore will lead to incorrect estimates of the inferred terminator temperature or molecular weight of the planetary atmosphere. At shorter wavelengths, in particular, additional degeneracies emerge between the stellar contamination temperature contrast C_s, covering fraction in area f_s, and the wavelength-dependent radius z_λ. Therefore, stellar

contamination parameters can alter the apparent opacity structure of the planetary spectrum in a wavelength-dependent manner.

2. There can be significant biases in the near-infrared transmission spectra for sub-Neptune planets orbiting M-dwarfs for a range of stellar contamination scenarios.

   We showed that biases can exist in the atmospheric properties of interest, here the metallicity, C/O, and scale height temperature (see Figures 3 and 4). Failure to account for contamination within a retrieval would result in incorrect atmospheric parameter estimates, by several sigma, depending on the degree of the heterogeneities and the planetary feature signal-to-noise (e.g., muted versus non-muted atmospheres). We also showed through model comparison (the Bayes factor) that the evidence favoring the inclusion of a contamination parameterization increased with increasing spectrophotometric precision and contamination fractions larger than 1%. We also identified over the 2.5–5 μm spectral window notable degeneracies between the spot-covering fraction and the atmospheric C/O as well as gray cloud opacity, and also a degeneracy between the stellar spot temperature and the atmospheric metallicity.

3. There is no substantial improvement in the atmospheric parameter constraints if we have a priori knowledge of the spot contamination characteristics.

   Assuming the spot parameterization and stellar models are accurate representations of the true stellar photosphere, simultaneously retrieving the contamination
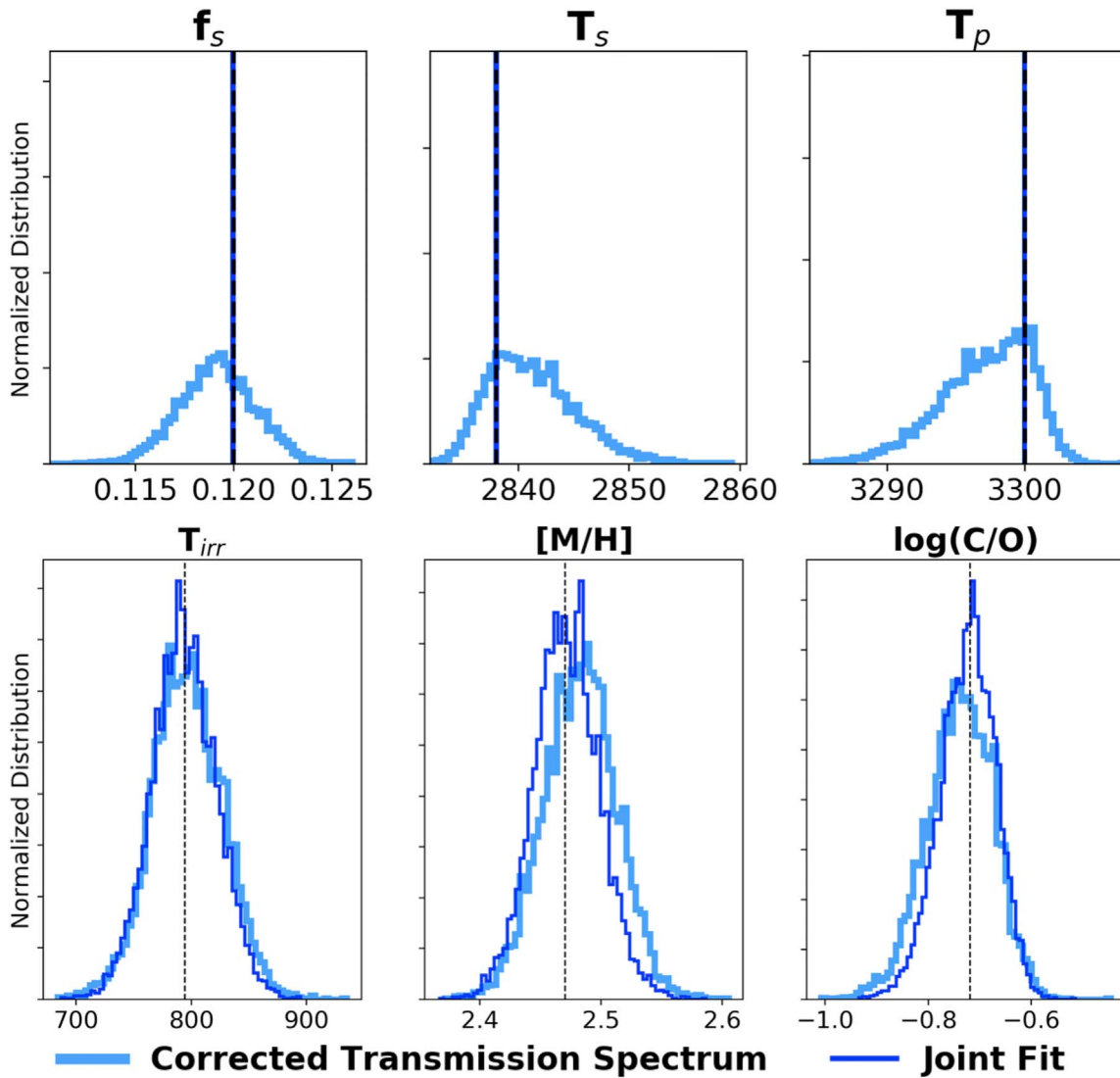
Figure 9. Impact of simultaneously retrieving on the stellar spectrum (joint fit model) to determine more precisely the spot properties, under the cloudy, high-metallicity scenario with a 12% spot-covering fraction (3300 K photosphere and 2838 K spot) as observed under the *JWST*-NIRISS-like bandpass with 30 ppm precision. The top panel compares the constraints on the stellar contamination parameters when utilizing only the transmission spectrum (light blue) and including both the stellar spectrum and transmission spectrum in the retrieval (dark blue). The bottom panel illustrates the behavior of the atmospheric parameter constraints when including the stellar spectrum in the retrieval. Despite the very precise inferences of stellar contamination parameters from the joint fit model, there is no significant improvement in atmospheric parameter constraints beyond simply including and marginalizing over the spot parameters on the transmission spectrum alone (corrected transmission spectrum model).

parameters from the stellar spectrum directly (e.g., the joint fit model) does not provide a statistically significant improvement in the precision on atmospheric parameters. Simply including stellar contamination parameterization as a set of nuisance parameters within the transmission spectrum retrieval alone is sufficient.

4. An incorrect parameterization/representation of the contaminated stellar photosphere can result in a substantial bias on the atmospheric properties.

Mimicking the differences between model and measured stellar spectra, we find that, despite including corrections for stellar contamination (see Figure 10), retrievals would result in significantly biased atmospheric properties. This is evident for the regime that is sensitive to detecting bias within the retrieval framework for planetary parameters, i.e., $f_s > 1\%$. Therefore, our ability to correct for stellar contamination on the transmission spectrum is

only as good as the accuracy of the stellar models. Understanding gaps within the stellar atmosphere models would be an important step toward alleviating this issue. Characterization of exoplanet atmospheres is a rewarding but challenging venture. Upcoming precision facilities/ instrumentation will greatly improve our understanding of exoplanet atmospheres, but will also bring numerous new challenges. This work has only focused on the impact of a single "degeneracy" on our ability to infer atmospheric properties (namely the transit light source effect, e.g., Apai et al. 2018; Rackham et al. 2018). This is but one in a large realm of possible retrieval degeneracies/complications of varying degrees that include, but are not limited to, the spectral resolution linked bias (Deming & Sheppard 2017), treatment of opacities (Baudino et al. 2017), cloud model assumptions (e.g., Helling et al. 2008a, 2008b, 2016; Mai & Line 2019), choice of abundance parameterizations
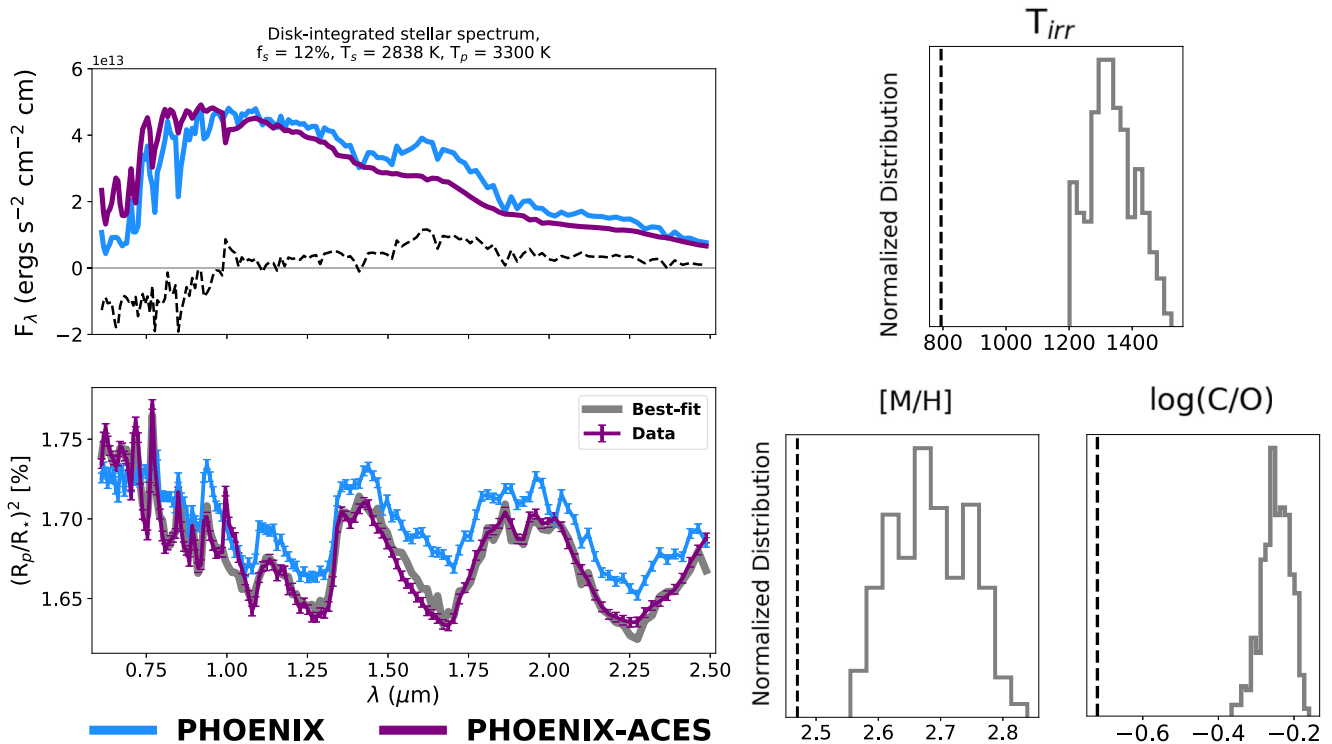
Figure 10. Demonstration of the impact of the choice of stellar model atmosphere used to correct the transmission spectrum, effectively mimicking anticipated deviations of a stellar model with the true stellar spectrum. This is under the cloudy, high-metallicity scenario orbiting an M-dwarf with 12% spot as observed with a 30 ppm precision over a NIRISS-like passband. We emulate the "true" stellar spectrum (top left) and contaminated transmission spectrum (bottom left) with the newer PHOENIX-ACES models (Husser et al. 2013) (in purple) but apply the correction in the transmission retrieval using the older PHOENIX models (Allard et al. 2003, 2007; Allard & Freytag 2009) (light blue). In the transmission spectrum panel (bottom left) we compare the "true spectrum" (e.g., "data") produced with the PHOENIX-ACES stellar model contamination to the incorrect contamination model produced with the PHOENIX stellar model, along with the "best" fit PHOENIX corrected stellar model (gray). The histograms on the right show the strong biases on the atmospheric properties that can occur when an inadequate stellar photosphere model is used for the spot correction (truth values as vertical dashed lines)

(Kreidberg et al. 2015; Kirkpatrick et al. 2016), and three-dimensional effects (Feng et al. 2016; Line & Parmentier 2016; Blecic et al. 2017). Diagnosing the impact and improving our understanding of these biases/degeneracies is critical to our determination of the true nature of extrasolar planetary atmospheres.

Supplementary information: Figures showing posterior probability distributions of all simulation scenarios discussed in this paper are available to the reader here: doi:10.5281/zenodo.3605726.

Facility: JWST.

Software: PyMultiNest (Buchner et al. 2014), pygtc (Bocquet & Carter 2016), pysynphot (STScI Development Team 2013).

ORCID iDs

Michael R. Line https://orcid.org/0000-0002-2338-476X

References

Afram, N., & Berdyugina, S. V. 2015, A&A, 576, A34
Agol, E., Cowan, N. B., Knutson, H. A., et al. 2010, ApJ, 721, 1861
Allard, F., Allard, N. F., Homeier, D., et al. 2007, A&A, 474, L21
Allard, F., & Freytag, B. 2009, Proceedings of the International Astronomical Union, 5, 756
Allard, F., Guillot, T., Ludwig, H.-G., et al. 2003, in IAU Symp., 211, Brown Dwarfs, ed. E. Martín (Cambridge: Cambridge Univ. Press), 325
Alonso, R., Auvergne, M., Baglin, A., et al. 2008, A&A, 482, L21
Andersen, J. M., & Korhonen, H. 2015, MNRAS, 448, 3053
Apai, D., Rackham, B. V., Giampapa, M. S., et al. 2018, arXiv:1803.08708
Ballard, S. 2019, AJ, 157, 113
Barnes, J., Jeffers, S., & Jones, H. 2011, MNRAS, 412, 1599
Barstow, J. K. 2018, NatAs, 2, 188
Barstow, J. K., Aigrain, S., Irwin, P. G., Kendrew, S., & Fletcher, L. N. 2015a, arXiv:1501.06349
Barstow, J. K., Aigrain, S., Irwin, P. G. J., Kendrew, S., & Fletcher, L. N. 2015b, MNRAS, 448, 2546
Barstow, J. K., Aigrain, S., Irwin, P. G. J., Kendrew, S., & Fletcher, L. N. 2015c, MNRAS, 451, 1306
Baudino, J.-L., Mollière, P., Venot, O., et al. 2017, ApJ, 850, 150

Berta, Z. K., Charbonneau, D., Bean, J., et al. 2011, ApJ, 736, 12
Blecic, J., Dobbs-Dixon, I., & Greene, T. 2017, ApJ, 848, 127
Bocquet, S., & Carter, F. W. 2016, JOSS, 1, 46
Brown, T. M., Charbonneau, D., Gilliland, R. L., Noyes, R. W., & Burrows, A. 2001, ApJ, 552, 699
Browning, M. K., Basri, G., Marcy, G. W., West, A. A., & Zhang, J. 2010, AJ, 139, 504
Bruno, G., Lewis, N. K., Stevenson, K. B., et al. 2018, AJ, 156, 124
Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125
Carter, J. A., Winn, J. N., Holman, M. J., et al. 2011, ApJ, 730, 82
Chapman, G. A. 1987, ARA&A, 25, 633
Cornish, N. J., & Littenberg, T. B. 2007, PhRvD, 76, 083006
Czesla, S., Huber, K. F., Wolter, U., Schröter, S., & Schmitt, J. H. M. M. 2009, A&A, 505, 1277
Damasso, M., Biazzo, K., Bonomo, A., et al. 2015, A&A, 575, A111
Delfosse, X., Forveille, T., Perrier, C., & Mayor, M. 1998, A&A, 331, 581
Deming, D., & Sheppard, K. 2017, ApJL, 841, L3
Désert, J.-M., Charbonneau, D., Demory, B.-O., et al. 2011, ApJS, 197, 14
Feng, Y. K., Line, M. R., Fortney, J. J., et al. 2016, ApJ, 829, 52
Feng, Y. K., Robinson, T. D., Fortney, J. J., et al. 2018, AJ, 155, 200
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Fortney, J. J., Mordasini, C., Nettelmann, N., et al. 2013, ApJ, 775, 80
Fraine, J., Deming, D., Benneke, B., et al. 2014, Natur, 513, 526
Freedman, R. S., Lustig-Yaeger, J., Fortney, J. J., et al. 2014, ApJS, 214, 25
Freedman, R. S., Marley, M. S., & Lodders, K. 2008, ApJS, 174, 504
Gillon, M., Jehin, E., Lederer, S. M., et al. 2016, Natur, 533, 221
Gillon, M., Triaud, A. H. M. J., Demory, B.-O., et al. 2017, Natur, 542, 456
Gondoin, P. 2008, A&A, 478, 883
Gordon, S., & McBride, B. 1996, Computer Program for Calculation of Complex Chemical Equilibrium Compositions and Applications II. Users Manual and Program Description NASA-RP-1311, NASA
Greene, T. P., Line, M. R., Montero, C., et al. 2016, ApJ, 817, 17
Guillot, T. 2010, A&A, 520, A27
Han, E., Wang, S. X., Wright, J. T., et al. 2014, PASP, 126, 827
Hawley, S. L., Davenport, J. R. A., Kowalski, A. F., et al. 2014, ApJ, 797, 121
Hawley, S. L., & Pettersen, B. R. 1991, ApJ, 378, 725
Helling, C., Dehn, M., Woitke, P., & Hauschildt, P. H. 2008a, ApJL, 675, L105
Helling, C., Lee, G., Dobbs-Dixon, I., et al. 2016, MNRAS, 460, 855
Helling, C., Woitke, P., & Thi, W.-F. 2008b, A&A, 485, 547
Huitson, C. M., Sing, D. K., Vidal-Madjar, A., et al. 2012, MNRAS, 422, 2477
Husser, T. O., Wende-von Berg, S., Dreizler, S., et al. 2013, A&A, 553, A6
Jackson, R. J., & Jeffries, R. D. 2012, MNRAS, 423, 2966
Kirkpatrick, J. D., Kellogg, K., Schneider, A. C., et al. 2016, ApJS, 224, 36
Knutson, H. A., Lewis, N., Fortney, J. J., et al. 2012, ApJ, 754, 22
Kreidberg, L., Bean, J. L., Désert, J.-M., et al. 2014, ApJL, 793, L27
Kreidberg, L., Line, M. R., Bean, J. L., et al. 2015, ApJ, 814, 66
Kreidberg, L., Line, M. R., Parmentier, V., et al. 2018, AJ, 156, 17
Lecavelier Des Etangs, A., Pont, F., Vidal-Madjar, A., & Sing, D. 2008, A&A, 481, L83
Line, M. R., & Parmentier, V. 2016, ApJ, 820, 78

Line, M. R., Stevenson, K. B., Bean, J., et al. 2016, AJ, 152, 203
Line, M. R., Wolf, A. S., Zhang, X., et al. 2013, ApJ, 775, 137
Luger, R., Sestovic, M., Kruse, E., et al. 2017, NatAs, 1, 0129
Lupu, R. E., Marley, M. S., Lewis, N., et al. 2016, AJ, 152, 217
Mai, C., & Line, M. R. 2019, ApJ, 883, 144
McCullough, P. R., Crouzet, N., Deming, D., & Madhusudhan, N. 2014, ApJ, 791, 55
Muirhead, P. S., & Veyette, M. 2018, AAS Meeting Abstracts, 231, 120.04
Murgas, F., Chen, G., Pallé, E., Nortmann, L., & Nowak, G. 2019, A&A, 622, A172
Narita, N., Fukui, A., Ikoma, M., et al. 2013, ApJ, 773, 144
Newton, E. R., Irwin, J., Charbonneau, D., Berta-Thompson, Z. K., & Dittmann, J. A. 2016, ApJL, 821, L19
O'Neal, D., Neff, J. E., & Saar, S. H. 1998, ApJ, 507, 919
Oshagh, M., Santos, N. C., Ehrenreich, D., et al. 2014, A&A, 568, A99
Pinhas, A., Rackham, B. V., Madhusudhan, N., & Apai, D. 2018, MNRAS, 480, 5314
Pont, F., Gilliland, R. L., Moutou, C., et al. 2007, A&A, 476, 1347
Pont, F., Knutson, H., Gilliland, R. L., Moutou, C., & Charbonneau, D. 2008, MNRAS, 385, 109
Pont, F., Sing, D. K., Gibson, N. P., et al. 2013, MNRAS, 432, 2917
Rackham, B., Espinoza, N., Apai, D., et al. 2017, ApJ, 834, 151
Rackham, B. V., Apai, D., & Giampapa, M. S. 2018, ApJ, 853, 122
Reiners, A., & Basri, G. 2010, ApJ, 710, 924
Rockenfeller, B., Bailer-Jones, C. A., & Mundt, R. 2006a, A&A, 448, 1111
Rockenfeller, B., Bailer-Jones, C. A., Mundt, R., & Ibrahimov, M. A. 2006b, MNRAS, 367, 407
Schmidt, S. J., Prieto, J. L., Stanek, K. Z., et al. 2014, ApJL, 781, L24
Seager, S., & Sasselov, D. D. 2000, ApJ, 537, 916
Shapiro, A. I., Solanki, S. K., Krivova, N. A., et al. 2014, A&A, 569, A38
Silva-Valio, A. 2008, ApJL, 683, L179
Sing, D. K., Pont, F., Aigrain, S., et al. 2011, MNRAS, 416, 1443
Skilling, J., et al. 2006, BayAn, 1, 833
Smirnov, N. V. 1939, Bulletin Moscow University, 2, 3
Strassmeier, K. G. 2010, Proceedings of the International Astronomical Union, 6, 174
STScI Development Team 2013, pysynphot: Synthetic Photometry Software Package, pysynphot Version 0.9.12, Astrophysics Source Code Library, ascl:1303.023
Swain, M. R., Line, M. R., & Deroo, P. 2014, ApJ, 784, 133
Swain, M. R., Vasisht, G., & Tinetti, G. 2008, Natur, 452, 329
Tinetti, G., Tennyson, J., Griffith, C. A., & Waldmann, I. 2012, RSPTA, 370, 2749
Trotta, R. 2008, ConPh, 49, 71
Veyette, M. J., Muirhead, P. S., Mann, A. W., et al. 2017, ApJ, 851, 26
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2019, arXiv:1907.10121
Wolter, U., Schmitt, J., Huber, K., et al. 2009, A&A, 504, 561
Zellem, R. T., Griffith, C. A., Pearson, K. A., et al. 2015, ApJ, 810, 11
Zellem, R. T., Swain, M. R., Roudier, G., et al. 2017, ApJ, 844, 27
Zhang, Z., Zhou, Y., Rackham, B. V., & Apai, D. 2018, AJ, 156, 178