





Retrieving Temperatures and Abundances of Exoplanet Atmospheres with High-resolution Cross-correlation Spectroscopy

Matteo Brogi^{1,2,3}  and Michael R. Line⁴ ¹ Department of Physics, University of Warwick, Coventry CV4 7AL, UK; m.brogi@warwick.ac.uk² INAF—Osservatorio Astrofisico di Torino, Via Osservatorio 20, I-10025, Pino Torinese, Italy³ Centre for Exoplanets and Habitability, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK⁴ School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287, USA

Received 2018 November 5; revised 2019 January 4; accepted 2019 January 16; published 2019 February 13

Abstract

High-resolution spectroscopy ($R \geq 25,000$) has recently emerged as one of the leading methods for detecting atomic and molecular species in the atmospheres of exoplanets. However, it has so far been lacking a robust method for extracting quantitative constraints on the temperature structure and molecular/atomic abundances. In this work, we present a novel Bayesian atmospheric retrieval framework applicable to high-resolution cross-correlation spectroscopy (HRCCS) that relies on the cross-correlation between data and models to extract the planetary spectral signal. We successfully test the framework on simulated data and show that it can correctly determine Bayesian credibility intervals on atmospheric temperatures and abundances, allowing for a quantitative exploration of the inherent degeneracies. Furthermore, our new framework permits us to trivially combine and explore the synergies between HRCCS and low-resolution spectroscopy to maximally leverage the information contained within each. This framework also allows us to quantitatively assess the impact of molecular line opacities at high resolution. We apply the framework to VLT CRIRES *K*-band spectra of HD 209458 b and HD 189733 b and retrieve abundant carbon monoxide but subsolar abundances for water, which are largely invariant under different model assumptions. This confirms previous analysis of these data sets, but is possibly at odds with detections of H₂O at different wavelengths and spectral resolutions. The framework presented here is the first step toward a true synergy between space observatories and ground-based high-resolution observations.

Key words: methods: data analysis – planets and satellites: atmospheres – techniques: spectroscopic

1. Introduction

The field of exoplanet characterization has matured to the point where we can begin to answer fundamental questions regarding planetary climate, composition, and formation, and provide context for understanding our own solar system planets (Bailey 2014; Burrows 2014; Madhusudhan et al. 2014; Crossfield 2015). The community has leveraged the power of ground- and space-based observatories to find and characterize a diverse range of planets, ranging from hot giants to terrestrial-sized, potentially habitable worlds. Atmospheric characterization has emerged as a key area of intense focus as of late because the atmosphere is the most readily accessible part of a planet via remote observations.

The most scientifically valuable measurements of exoplanet atmospheres are those constraining their composition and temperature changes with altitude, ideally as a function of orbital phase. The most stringent constraints so far come from observations with low-resolution spectroscopy (hereafter LRS; at a resolving power $R = \lambda/\Delta\lambda \leq 200$), primarily with the *Hubble Space Telescope* (*HST*; Deming et al. 2013; Mandell et al. 2013; Knutson et al. 2014; Kreidberg et al. 2014a; Stevenson et al. 2014; Haynes et al. 2015; Evans et al. 2016) and *Spitzer Space Telescope* (Grillmair et al. 2008). The instruments on board *HST* permit near-continuous coverage over a broad wavelength range spanning ~ 0.3 – $1.7 \mu\text{m}$ (e.g., Sing et al. 2016) split across three passbands. In addition, *Spitzer* provides complementary broadband photometry from 3.5 to $5.4 \mu\text{m}$ (and historically out to $8 \mu\text{m}$). Although the coverage is broad, the resolution is very coarse, typically $R \sim 30$ – 200 . *HST* observations in both emission and

transmission generally permit order-of-magnitude constraints on the molecular abundance of water (e.g., Kreidberg et al. 2014b; Line et al. 2016; Wakeford et al. 2017; Tsiaras et al. 2018). The low-resolution and limited near-IR coverage, however, has precluded our ability, for most objects, to sufficiently break degeneracies to constrain (beyond upper limits) the abundances of other key diagnostic molecules like methane, carbon dioxide, carbon monoxide, ammonia, hydrogen cyanide, and acetylene. The *James Webb Space Telescope* (*JWST*), anticipated to launch within the next few years, will significantly improve precisions on the aforementioned quantities (Greene et al. 2016), due to its extremely broad, continuous (0.6 – $12 \mu\text{m}$), low-to-moderate resolution spectroscopy ($R \sim 100$ – 3500). However, *JWST* is going to be a limited resource, and it is therefore crucial to pair it with independent but complementary observations.

A powerful emerging approach for characterizing exoplanet atmospheres is high-resolution cross-correlation spectroscopy (HRCCS). It leverages the ability to resolve molecular bands into individual lines and to detect the planet’s Doppler shift directly at the $\sim \text{km s}^{-1}$ level. In addition, it benefits from the large collective area of ground-based telescopes. Spectral information is extracted through cross-correlation with model templates, which acts as a robust signal filtering technique to recognize the peculiar fingerprint of each species. HRCCS is the best technique so far to unambiguously identify molecules, and it is the only technique to have reliably detected carbon-based molecules in the atmospheres of transiting and non-transiting exoplanets, starting with the pioneering detection of carbon monoxide by Snellen et al. (2010).

Despite this potential, little work has focused on rigorously determining the abundances of molecules, the vertical temperature structure, and other fundamental atmospheric properties from HRCCS data. One of the primary challenges in doing so is placing HRCCS within a robust atmospheric retrieval framework. In contrast to LRS data, where spectra are either calibrated in flux or measured in comparison to a reference star, HRCCS data are “self-calibrated,” meaning that the broadband information and the time variations of flux at each wavelength are divided out of the data. With such normalization, the small variations due to the planet atmosphere are all measured relative to the stellar flux. This peculiarity, together with the small residual broadband structure due to imperfect normalization, results in the loss of a reliable continuum, which inhibits the use of the standard (data spectrum – model spectrum) residual vector used in typical LRS parameter estimation via chi-square. Cross-correlating with model spectra, besides providing a matching filter for robust identification of species, is also insensitive to any residual broadband variations.

The challenge with retrieving atmospheric parameters from high-resolution spectra is converting the cross-correlation values into a goodness-of-fit estimator. Brogi et al. (2012) developed a statistical test to assess the significance of cross-correlation signals by comparing the distribution of cross-correlation values around the planet radial velocity (typically labeled as the “in-trail” sample) to the values away from it (the “out-of-trail” sample). This strategy has been widely adopted in the literature since then (Birkby et al. 2013, 2017; Brogi et al. 2013, 2014, 2018; Nugroho et al. 2017; Hawker et al. 2018). Due to the necessity to compute a statistically significant number of cross-correlation values (typically thousands of values mapping the planet’s orbital and systemic velocity), its application is limited to the evaluation of a relatively small (~ 100) number of models. Furthermore, this test is somewhat sensitive to the range around the planet radial velocity chosen for building the in-trail distribution.

Brogi et al. (2017) introduced a different method for overcoming some of the above limitations. They estimated as accurately as possible the model cross-correlation function and compared it to the *observed* cross-correlation function via chi-square fitting. To replicate the astrophysical, instrumental, and data analysis effects as closely as possible, each model is added to the real data, albeit at a sufficiently low level so that the noise properties and data analysis are not altered. This alternative strategy still requires significant computational resources and is therefore limited to the evaluation of only a few thousands model HRCCS spectra sampled from a pre-existing LRS posterior. Such sparse sampling is only sufficient to constrain confidence intervals within the 3σ level, and on a limited portion of the parameter space. As a consequence, full independent retrievals on HRCCS data cannot be performed. Conditioning the HRCCS retrieval on the LRS retrieval implicitly assumes that the two data sets contain the same amount of evidence (or weighting), which is not necessarily the case given the different spectral range and signal to noise of the observations. Lastly, because neighboring cross-correlation values can be correlated (depending on the sampling in velocity and the instrumental resolution), chi-square is not necessarily the correct statistic to compare cross-correlation functions.

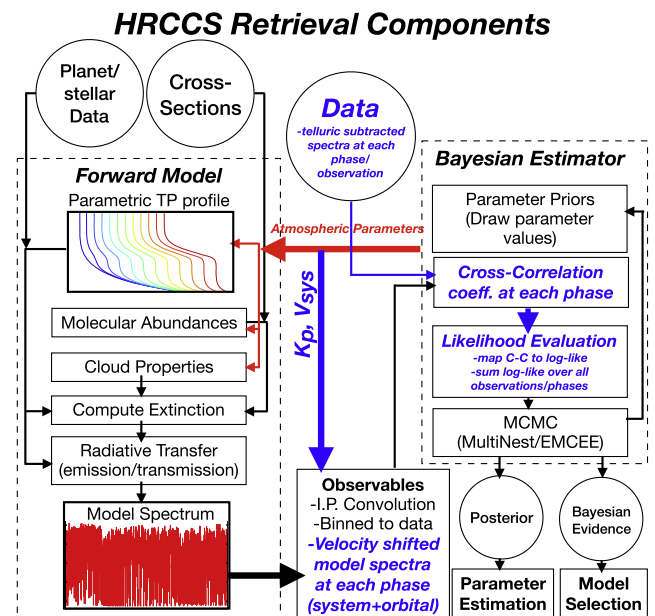


Figure 1. Illustration of the key components of the HRCCS retrieval algorithm. Many components are the same as in a classic “low-resolution” retrieval: the forward model, the data, and the Bayesian estimator. The key novelties (indicated in blue) required to perform retrievals in cross-correlation space are the inclusion of the radial velocities (system+orbital), a telluric subtracted “data cube” of normalized spectrum vs. time/phase, and a mathematical mapping from the correlation coefficient to a likelihood function that can be used inside a parameter estimation package such as MCMC (figure modified from MacDonald & Madhusudhan 2017 with permission).

The primary goal of this paper is to introduce a robust and unbiased framework to perform Bayesian retrieval analyses on HRCCS data, free from the limitations of previous approaches. In Section 2, we define a likelihood function for high-resolution spectra, and we describe its implementation into a nested-sampling Bayesian estimator. In Section 3, we describe the setup used to test our new Bayesian framework, including a simulated data set replicating previous work on dayside spectroscopy of HD 209458 b. We present the analysis of the simulated data set (Section 3.2), the excellent match between the retrieved and the modeled atmospheric parameters (Section 3.3), the effects of uncertainties in the line list of water vapor (Section 3.5), and the increased precision when combining space and ground observations (Section 3.6). We then apply the framework to real observations of two exoplanets, HD 189733 b and HD 209458 b, in Section 4. We conclude in Section 5 by highlighting future applications of this framework, in particular coordinated observations with space and ground-based telescopes.

2. A Novel HRCCS Atmospheric Retrieval Framework

At its core, an HRCCS retrieval is no different from the typical LRS retrievals applied in numerous previous works (see, e.g., the review by Madhusudhan 2018). The key components of any retrieval algorithm are the data or observable, the forward model that maps the quantities of interest onto the data/observable, and the Bayesian estimator that optimizes the forward model parameters (or the range thereof) given a likelihood function (Figure 1). Below we describe each of these key components and how they are adapted within our HRCCS retrieval framework.

2.1. The Forward Model

A typical spectral retrieval forward model is a radiative transfer routine that takes in gas abundances, cloud properties, vertical thermal structure information, and/or geometric information to produce a transmission or emission spectrum. In this work, we leverage the CHIMERA forward model (Line et al. 2013, 2014, 2015, 2016, 2017; Line & Yung 2013). The CHIMERA⁵ forward model offers the flexibility to handle a variety of atmospheric assumptions from simple 1D non-self-consistent (Line et al. 2013, 2014, 2016) to full 1D self-consistent thermochemical radiative-convective equilibrium (Arcangeli et al. 2018; Kreidberg et al. 2018; Mansfield et al. 2018) to quasi-2D/3D parameterizations and observing geometries (Feng et al. 2016; Line & Parmentier 2016) as well as flexible treatments of opacities (e.g., line by line, line sampling, and correlated- k) based upon a pre-tabulated line-by-line ($<0.01 \text{ cm}^{-1}$) absorption cross-section database (Freedman et al. 2008, 2014).

Specifically, here, we explore the classic free and chemically consistent versions described in Kreidberg et al. (2015) but in the emission geometry. In the free forward model, we include as free parameters the H₂O and CO mixing ratios (constant with pressure), H₂-H₂/He collision-induced absorption (CIA), and the three-parameter⁶ disk-integrated temperature-pressure profile (T - p profile) parameterization of Guillot (2010).⁷ In the chemically consistent forward model, we utilize the same T - p profile parameterization but parameterize the composition with a metallicity ($[M/H]$, where solar is 0) and log-carbon-to-oxygen ratio ($\log(C/O)$, where solar is -0.26) under the assumption of pure thermochemical equilibrium computed using the NASA CEA2 routine (Gordon & McBride 1994; Kreidberg et al. 2015) by scaling the solar elemental abundances of Lodders et al. (2009). In the forward modeling used in this work, we include H₂O, CH₄, CO, CO₂, NH₃, HCN, and the H₂-H₂/He CIA as opacities; however, for nearly solar abundances, only H₂O and CO present themselves over the K -band spectrum. The output from the forward model is a line-by-line top-of-atmosphere flux over the K -band (2.26 – $2.35 \mu\text{m}$), which is then normalized to the stellar spectrum and planet-to-star area ratio.

The forward model spectra are then Doppler-shifted by spline interpolation⁸ (Brogi et al. 2014) based on the semi-amplitude of the planet radial velocity (K_p), the systemic velocity (V_{sys}), and the barycentric velocity of the observer V_{bary} , according to

$$V_p(t) = V_{\text{sys}} + V_{\text{bary}}(t) + K_p \sin[2\pi\varphi(t)], \quad (1)$$

where we have ignored cross-terms in velocity due to their small impact (on the order of m s^{-1}). The Doppler-shifted model spectra are finally convolved with the appropriate instrumental profile (IP). Due to the fact that velocities in Equation (1) are not perfectly known, we introduce two additional key parameters: a differential systemic velocity (dV_{sys}) and a differential planet radial velocity (dK_p). With

⁵ A version of the transmission spectrum code is publicly available through STScI's ExoCTK package: <https://github.com/ExoCTK/chimera>.

⁶ $\log(\kappa_{\text{ir}}) = \log$ of the Planck mean infrared opacity; $\log(\gamma) = \log$ of the visible-to-infrared opacity; $T_{\text{irr}} = \text{top-of-atmosphere irradiation temperature}$.

⁷ Different temperature profile parameterizations can be explored in a future investigation.

⁸ With Python's `scipy.interpolation` package.

typical uncertainties of a few km s^{-1} , detectable at the spectral resolution and signal-to-noise ratio (S/N) of HRCCS observations, ignoring these differential velocities would lead to the incorrect localization of the planet's signal.

2.2. The Bayesian Estimator: Cross-correlation to Log-likelihood Mapping

Bayesian estimation necessarily requires a well-defined likelihood function (Gregory 2005; Sivia & Skilling 2006; Feroz et al. 2009; Foreman-Mackey et al. 2013). The standard likelihood function in LRS analysis is a Gaussian, or similarly, a quadratic or chi-square log-likelihood. The challenge with HRCCS data is to exploit the power of the large number of spectral lines and relatively well-known orbital velocity via the cross-correlation between a template model and the phase-dependent data. Once a mapping from the readily calculable cross-correlation coefficient to log-likelihood is found, one can employ the standard suite of Bayesian analysis tools including parameter estimation, prior inclusion, and model selection. We utilize the powerful `pymultinest` tool (Buchner et al. 2014), a Python wrapper to `multinest` (Feroz et al. 2009), to perform all of the parameter estimation within our HRCCS retrieval framework.

In building this HRCCS likelihood function, we want to exploit some unique characteristics of the analysis of high-resolution spectra. First, the continuum is divided out of the data to allow for their self-calibration (i.e., to eliminate the necessity for a reference star). Second, the planetary signal is extracted via cross-correlation with model spectra, so our metric for the goodness of fit must incorporate the cross-correlation function or a closely related quantity. Third, the sign of the correlation coefficient matters, as it allows us to discriminate between emission lines incorrectly fitted with an absorption spectrum and vice versa. Lastly, although the cross-correlation function is by definition normalized and thus insensitive to scaling of the model and/or the data, observations have a finite S/N. This means that two nearly identical model spectra in terms of line-to-line ratios, but differing by orders of magnitude in the overall line strength, should yield a different likelihood.

2.2.1. Previous Cross-correlation to Log-likelihood Mappings

The idea of mapping the correlation coefficient to a log-likelihood (hereafter CC-to-log L) is not new. One of the earlier applications of such a mapping in the exoplanet community was in the context of precise stellar radial velocities (Zucker 2003). The aim of Zucker (2003) was to derive a formalism that facilitated combining radial velocity measurements obtained via cross-correlation from multiple spectral regions and at varying S/Ns. Starting from the definition of χ^2 , Zucker (2003) derived the following CC-to-log L mapping:

$$\log(L) = -\frac{N}{2} \log(1 - C^2). \quad (2)$$

The key feature of this mapping is that the log L depends on the square of the correlation coefficient (C). For our application, the use of the square of the cross-correlation function limits the sensitivity to inversion layers. Although the actual line shape in dayside spectra can be quite complex and it is not a trivial “sign flip” when going from non-inverted to inverted (e.g., Schwarz et al. 2015), spectra with emission lines cross-correlated with

absorption models (and vice versa) produce some anticorrelation that would not be discriminated due to the insensitivity to the sign of C . In addition, the log-likelihood of Zucker (2003) weights the spectra based on their cross-correlation value, assuming therefore that each spectrum delivers a peak significantly above the level of the noise. Although this is appropriate for stellar radial velocities, it is not the case for planetary radial velocities, and we thus expect this likelihood to struggle at low-S/N levels. This mapping should work well for high-S/N spectra of isolated objects like brown dwarfs and directly imaged planets, as demonstrated in Bowler et al. (2017).

Lockwood et al. (2014) were the first to apply a CC-to-log L mapping to high-resolution exoplanet spectroscopy of the non-transiting planet τ Boötis b. Their analysis comprised two steps. First, the formalism of Zucker (2003) as implemented in the two-dimensional routine TODCOR is used to combine all of the cross-correlation functions taken at a certain epoch (e.g., during one night of observations) into a maximum-likelihood estimator, i.e., an effective cross-correlation value. Subsequently, a CC-to-log L mapping is derived as

$$\log(L) = C \quad (3)$$

and used to combine cross-correlation functions taken at different epochs. An important underlying hypothesis of their formalism, which is instead violated by our CRIFES observations, is that the change in planet radial velocity during one set of observations is negligible with respect to the instrumental resolution. This is ensured by typically taking NIRSPEC spectra when the planet is in quadrature. An additional substantial difference with our analysis is that the cross-correlation function contains both the stellar and the planet spectrum, i.e., it is a two-dimensional cross-correlation with the stellar coefficients dominating the planet coefficients by orders of magnitude.

The mapping of Lockwood et al. (2014) was also recently adopted in Piskorz et al. (2018) to combine Keck NIRSPEC K -band data with *Spitzer* for the transiting hot-Jupiter Kelt-2 Ab. However, we again stress that such formalism cannot be applied to our data analysis where we make use of the change in planet radial velocity with time. We show indeed in Section 3.4 that if we incorrectly apply Equation (3) to our data, the resulting $\log L$ is not distributed as a χ^2 (Figure 5) as Wilks' Theorem demands (Wilks 1938).

2.2.2. A New Mapping

In this section, we derive a new CC-to-log L mapping that leverages all of the aforementioned key components of a HRCCS observation. The starting point for building our mapping closely matches the derivation in Section 2 of Zucker (2003). We denote a single observed spectrum by $f(n)$, where n is the bin number, or spectral channel.

We compute a template spectrum $g(n)$ in the same reference frame as the data. We assume that the model describing the data is

$$f(n) = a g(n - s) + d_n, \quad (4)$$

where a is a scaling factor, s a bin/wavelength shift, and d_n the noise at bin n .

It is important that $f(n)$ and $g(n)$ are continuum subtracted. In our current analysis and following common implementation of numerical cross-correlation routines, we achieve this by subtracting the mean from each of the vectors prior to cross-correlation. Under these assumptions, we have that $\sum_n f(n) = 0$ and $\sum_n g(n) = 0$.

We assume that the noise is Gaussian distributed at each pixel with standard deviation σ . The likelihood function L of our model is

$$\begin{aligned} L &= \prod_n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{[f(n) - a g(n - s)]^2}{2\sigma^2}\right\} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left\{-\sum_n \frac{[f(n) - a g(n - s)]^2}{2\sigma^2}\right\}, \end{aligned} \quad (5)$$

where N is the total number of spectral channels. This is typically the number of pixels per spectrum (or per detector, or per spectral order). The log-likelihood can be derived from the above:

$$\log(L) = -N \log \sigma - \frac{1}{2\sigma^2} \sum_n [f(n) - a g(n - s)]^2, \quad (6)$$

after ignoring constant additive terms. We note that throughout this section, the function “log” will indicate the natural logarithm.

At this stage, our analysis diverges from Zucker (2003). We impose that the scaling factor is unity ($a = 1$). Physically, this means that we want the overall strength of spectral lines in the model to match the observed data. Practically, because our data $f(n)$ is normalized and telluric corrected, any residual variations are relative to the stellar continuum. We therefore scale $g(n)$ by the stellar flux and planet-to-stellar area ratio (Equation (12) below). We note that the formalism in Zucker (2003) does not specify whether the cross-correlation is performed with a model spectrum, a binary mask, or generically a template. Therefore, a does not carry any physical meaning in that context, as the units of the template are arbitrary. This is why in their work a is substituted with its maximum-likelihood estimator \hat{a} .

Equation (6) contains another variable, σ . We do not fix this, as it is difficult to estimate the exact level of noise in each spectral channel, especially under imperfect removal of the telluric spectrum. We instead compute the maximum-likelihood estimator $\hat{\sigma}$ by nulling the partial derivative of the log-likelihood:

$$\begin{aligned} 0 &= \frac{\partial \log(L)}{\partial \sigma} \\ \frac{N}{\hat{\sigma}} &= \frac{1}{\hat{\sigma}^3} \sum_n [f(n) - g(n - s)]^2 \\ \hat{\sigma}^2 &= \frac{1}{N} \sum_n [f(n) - g(n - s)]^2. \end{aligned} \quad (7)$$

Substituting $\hat{\sigma}$ into $\log(L)$ and ignoring constant additive terms, we get

$$\begin{aligned} \log(L) &= -N \log \left\{ \sqrt{\frac{1}{N} \sum_n [f(n) - g(n - s)]^2} \right\} - \frac{N}{2} \\ &= -\frac{N}{2} \log \left\{ \frac{1}{N} \sum_n [f(n) - g(n - s)]^2 \right\} \\ &= -\frac{N}{2} \log \left\{ \frac{1}{N} \sum_n [f(n)^2 - 2f(n)g(n - s) + g(n - s)^2] \right\}. \end{aligned} \quad (8)$$

We can now write the formulas for the variance of the data (s_f^2), the variance of the model (s_g^2), and the cross-covariance $R(s)$:

$$\begin{aligned} s_f^2 &= \frac{1}{N} \sum_n f^2(n) \\ s_g^2 &= \frac{1}{N} \sum_n g^2(n-s) \\ R(s) &= \frac{1}{N} \sum_n f(n)g(n-s). \end{aligned}$$

Substituting them into $\log(L)$ leads us to

$$\log(L) = -\frac{N}{2} \log [s_f^2 - 2R(s) + s_g^2]. \quad (9)$$

Factorizing out the product $s_f s_g$, we can make the cross-correlation appear

$$\begin{aligned} \log(L) &= -\frac{N}{2} \left\{ \log(s_f s_g) + \log \left[\frac{s_f}{s_g} + \frac{s_g}{s_f} - 2 \frac{R(s)}{\sqrt{s_f^2 s_g^2}} \right] \right\} \\ &= -\frac{N}{2} \left\{ \log(s_f s_g) + \log \left[\frac{s_f}{s_g} + \frac{s_g}{s_f} - 2C(s) \right] \right\}, \end{aligned} \quad (10)$$

with the correlation coefficient

$$C(s) = \frac{R(s)}{\sqrt{s_f^2 s_g^2}}. \quad (11)$$

For practical applications, Equation (9) is slightly faster to compute than Equation (10) and is the preferred choice for our numerical implementation. It is important to note that the data variance (s_f^2) only needs to be computed once at the end stage of the data analysis (step 7 in Section 3.2 below). However, s_g^2 will change as a function of the model tested, and also to a lesser extent as a function of the Doppler shift tested. Therefore, in our analysis we will recompute s_g^2 every time a model is evaluated, and for each of the spectra in the time sequence.

Equation (9) preserves the sign of the cross-covariance and will therefore discriminate between correlation and anticorrelation. This is a direct consequence of imposing $a = 1$. In addition, when the variance of the data and the variance of the (scaled) model differ significantly, the likelihood decreases accordingly. This incorporates a metric for comparing the average line depth to the S/N of the data.

It is important to realize that if we carried out the mathematical calculations with the scaling factor a as an explicit variable and then imposed $\partial \log(L) / \partial a = 0$ at the stage of Equation (9) (had we kept an a and a^2 multiplier in front of the $R(s)$ and s_g^2 terms, respectively), we would have obtained as solution $a = 1$. This means that our physically motivated choice of $a = 1$ also corresponds to choosing the maximum-likelihood estimator for this variable.

All of the quantities listed in Equation (9) are obtained as byproducts of the current analysis techniques of high-resolution spectra. In Section 3.2, we discuss additional details of the data analysis important for the application of this formalism.

3. Tests on Simulated Data

In this section, we demonstrate, on a simulated emission spectrum data set, the feasibility and utility of our novel HRCCS retrieval framework and the CC-to- $\log L$ mapping presented in Section 2. We start by describing the construction of the simulated data set in Section 3.1 and its analysis in Section 3.2. We present the “fiducial” retrieved constraints in Section 3.3, compare the constraints derived from the different CC-to- $\log L$ mappings in Section 3.4, explore the impact of different water line lists in Section 3.5, and finally combine in a coherent way the high-resolution spectra with a simulated *HST* Wide Field Camera 3 (*HST* WFC3) LRS data set in Section 3.6.

3.1. Construction of the Simulated Data Set

One half night of data is simulated based on real CRIRES observations of HD 209458 b (Schwarz et al. 2015; Brogi et al. 2017). The synthetic data set incorporates the photon noise from the star, variations in the Earth’s transmission spectrum with airmass, variable detector efficiency, the phase-dependent Doppler shift of the planet, and the time-dependent IP. This simulated data set constitutes the basis for testing the retrieval framework presented in the previous sections, as it incorporates all the major sources of uncertainties in the analysis of HRCCS data.

To generate this data set, we compute a solar-composition planet atmosphere with parameters listed in Table 1 and using the modeling tools described in Section 2.1. We run the computations over the wavelength range 2267–2350 nm (matching the CRIRES setup of the real observations) at a resolving power of $R \sim 440,000$ and scale the model to the stellar flux of HD 209458 via

$$F_{\text{scaled}}(\lambda) = \frac{F}{B(\lambda, T_{\text{eff}})} \left(\frac{R_p}{R_*} \right)^2, \quad (12)$$

where F is the model flux in $\text{W m}^{-2} \text{m}^{-1}$, R_p and R_* are the stellar and planet radii, respectively, and B the Planck function at the stellar effective temperature T_{eff} approximating the stellar spectrum (Table 2). The top panel (Step 1) of Figure 2 shows a small portion of this spectrum in the wavelength range corresponding to detector 1 of CRIRES.

We adopt a Keplerian circular velocity of 145.9 km s^{-1} , i.e., the literature value for HD 209458 b, and a combination of systemic and barycentric velocities to match the actual observations of night 1 in Brogi et al. (2017; Table 2). The scaled model F_{scaled} is Doppler-shifted according to the radial velocity at each epoch of observations computed via Equation (1) and saved in a matrix $F^i(\lambda, t)$. In this test case, the observations span 1024 pixels/wavelength channel and 59 separate integrations (spectra) covering phases 0.506–0.577 resulting in a $\sim 75 \text{ km s}^{-1}$ change in Doppler shift throughout the sequence (Figure 2, Step 2).

The wavelength- and time-dependent transparency $T(\lambda, t)$ of Earth’s atmosphere (the telluric spectrum) is computed via the ESO Skycalc command-line tool, which is based on the Cerro Paranal Sky Model (Noll et al. 2012; Jones et al. 2013). The model takes into account the sky position of the target at the time of the observation and meteorological data, except for the precipitable water vapor (PWV) that needs to be adjusted manually. We find a good match to the HD 209458 data set by adopting the average PWV of 2.5 mm for Cerro Paranal.

Table 1
Forward Model Parameters and Uniform Prior Ranges

| Parameter | Symbol | Uniform Prior Range | Model Input “Truth” |
|---|----------------------------|----------------------------|---------------------|
| <i>Doppler/Scale Parameters</i> | | | |
| Relative Systemic Velocity | dV_{sys} | $-50-50 \text{ km s}^{-1}$ | 0 |
| Relative Planet Radial Velocity | dK_{p} | $-50-50 \text{ km s}^{-1}$ | 0 |
| Model Scale Factor | $\log(a)$ | $-2-2$ | 0.0 |
| <i>Temperature–Pressure Profile Parameters</i> | | | |
| Planck Mean IR Opacity | $\log(\kappa_{\text{ir}})$ | $-3-1$ | -1 |
| Visible to IR Opacity | $\log(\gamma)$ | $-4-2$ | -1.5 |
| Irradiation Temperature | T_{irr} | 300–2800 K | 1400 |
| <i>Free Retrieval Abundance Parameters</i> | | | |
| H ₂ O Mixing Ratio | $\log(\text{H}_2\text{O})$ | $-12-0$ | -3.4 |
| CO Mixing Ratio | $\log(\text{CO})$ | $-12-0$ | -3.22 |
| <i>Chemically Consistent Abundance Parameters</i> | | | |
| Metallicity | $[M/H]$ | $-2-2$ | 0.0 |
| Carbon-to-oxygen Ratio | $\log(\text{C/O})$ | $-2-1$ | -0.26 |

Note. The first three parameters are required to Doppler-shift and scale the template model spectrum. The second three control the temperature–pressure profile. The planet metallicity and carbon-to-oxygen ratio replace the H₂O and CO mixing ratios when using the chemically consistent model. In most tests, the free model includes only H₂O, CO, and H₂/He CIA as opacity sources, whereas the chemically consistent model includes opacities from H₂O, CH₄, CO, CO₂, NH₃, HCN, and the H₂–H₂/He CIA. Each model (free or chemically consistent) has eight total free parameters unless specified otherwise.

We measure the average flux levels in the observed CRILES spectra by taking the median of their brightest pixels, and we compute the number of recorded photoelectrons by multiplying by the exposure time and detector gain of those observations. This gives us a vector $\epsilon(t)$, where we ignore any wavelength dependence of the instrumental throughput. Such dependence certainly occurs even over the small wavelength range of one CRILES detector (10–15 nm) and indeed measurable trends in the continuum (typically a slope) are on the order of $\sim 1\%$. However, owing to the good level of thermal stability of CRILES (0.01 K over a half night), these effects are not time-dependent and are therefore divided out during telluric removal.

The noiseless modeled data set is obtained by combining all the above sources:

$$F_{\text{mod}}(\lambda, t_i) = [1 + F'(\lambda, t_i)] T(\lambda, t_i) \epsilon(t_i), \quad (13)$$

where the product is scalar, i.e., computed element by element, and adding 1 accounts for the normalized stellar spectrum.

In the bright source limit, the noise budget is completely dominated by the stellar photons, so the noise matrix is governed by Poisson statistics. This leads to the following noise matrix:

$$N_{\text{mod}}(\lambda, t) = \mathcal{N}(0, 1) \sqrt{F_{\text{mod}}(\lambda, t)}, \quad (14)$$

where $\mathcal{N}(0, 1)$ is a normally distributed random variable. The simulated spectrum is the sum of the recorded photons and the noise matrix, i.e.,

$$F_{\text{sim}}(\lambda, t) = F_{\text{mod}}(\lambda, t) + N_{\text{mod}}(\lambda, t). \quad (15)$$

The panel labeled with step 3 in Figure 2 shows the final simulated “raw data” product. The obvious features seen in the final data matrix are the stationary telluric lines (dark vertical stripes) and the airmass-throughput dependence (horizontal stripes). This synthetic data set is then processed using the procedure outlined below.

3.2. Analysis of the Simulated Data Set

The analysis devised for processing data within this new data set is largely adapted from previous literature; however, there are some caveats related to the nature of our cross-correlation-to-likelihood mapping that require particular care. In Figure 2, this analysis is labeled steps 4–7, which we detail below:

1. Step 4: each spectrum is calibrated in wavelength by comparing the pixel position of telluric lines in the observed spectra to their theoretical wavelength obtained from a telluric model. As in Flowers et al. (2018), a common wavelength solution with constant space in velocity (constant $d\lambda/\lambda$) is computed and data are re-gridded to this solution via spline interpolation. Each spectrum is normalized by the median of its brightest 300 pixels to correct for throughput variations. Re-gridding is necessary for estimating the IP of CRILES, which is done at this stage through the procedure described in Rucinski (1999) and implemented in Brogi et al. (2016) and Flowers et al. (2018). The IP is used to convolve the model spectra at a later stage.
2. Step 5: the spectra are averaged in time and the mean spectrum is fitted to each observed spectrum with a second-order polynomial. This procedure removes the main variability in the depth of methane telluric lines; however, residuals are still visible at the position of water vapor telluric lines, which behave differently from methane due to the different scale height in the Earth’s atmosphere and temporal changes in humidity.
3. Step 6: these extra changes in water telluric lines are corrected by modeling the flux in each spectral channel as a function of time with a second-order polynomial and dividing out the fit. Since the planet’s orbital motion produces a time-varying Doppler shift of the spectrum, spectral lines from the exoplanet’s atmosphere will shift across adjacent spectral channels and will be nearly unaffected by the above correction.

Table 2
Relevant Parameters for the Systems HD 209458 and HD 189733 Used Throughout the Paper

| Parameter | Symbol | HD 209458 | | HD 189733 | |
|--|------------------------------------|---------------------|-----------|----------------------|------------|
| | | Value | Reference | Value | References |
| Stellar radius (R_{Sun}) | R_* | 1.162 | S10 | 0.756 | T08 |
| Effective temperature (K) | T_{eff} | 6065 | T08 | 5040 | T08 |
| Planet gravity (\log_{10} , cgs) | $\log(g_p)$ | 2.96 | S10 | 3.28 | ... |
| Planet radius (R_p) | R_p | 1.38 | S10 | 1.178 | T09 |
| Radial-velocity amplitude (km s^{-1}) | K_p | 145.9 | B17 | 152.5 | B16 |
| Phase range (# spectra) | ϕ | [0.506, 0.578] (59) | K07 | [0.383, 0.479] (110) | A10 |
| | | [0.557, 0.622] (54) | K07 | ... | ... |
| Radial velocity (km s^{-1}) | $V_{\text{sys}} + V_{\text{bary}}$ | [-26.92, -26.25] | K07+T08 | [-9.40, -8.84] | T09+A10 |
| | | [-13.44, -12.81] | K07+T08 | ... | ... |

Note. Parameter ranges for the phase range and radial velocity for HD 209458 system are given for both nights. References are K07 = Knutson et al. (2007), T08 = Torres et al. (2008), T09 = Triaud et al. (2009), A10 = Agol et al. (2010), S10 = Southworth (2010) B16 = Brogi et al. (2016), B17 = Brogi et al. (2017).

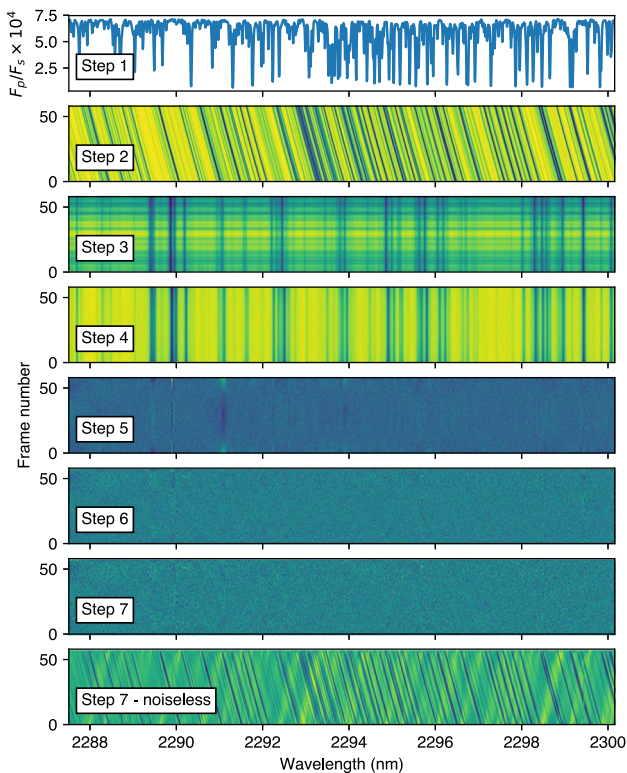


Figure 2. Step-by-step visualization of the process used to construct a simulated data set (steps 1–3; detailed in Section 3.1) and analyze it (steps 4–7; detailed in Section 3.2). This sequence is based on real on-sky performances of the infrared spectrograph CRILES while observing the exoplanet HD 209458 b. Shown is the spectral sequence computed for the first detector of the spectrograph (out of four detectors), incorporating all the major astrophysical and instrumental sources of noise. This simulated data set is used to test the Bayesian framework explained in Section 2. The analysis at steps 4–7 is also performed on the observed data sets, as explained in Section 4.

- Step 7: any further alteration of the noise properties in the data must be avoided, because they would alter the data variance s_f^2 in Equation (9). Consequently, the common practice of weighting down noisy spectral channels by dividing them through their variance in time must be avoided. In this revised analysis, noisy columns above $3 \times$ the standard deviation of the matrix are masked and

not used at the cross-correlation stage. The total number of channels N is modified accordingly for subsequent use in Equation (9).

As in previous work, the above analysis exploits the fact that the planet spectrum is subject to a variable Doppler shift during a few hours of observations, whereas the contaminant signals (telluric and stellar) are stationary or quasi-stationary. However, we note that steps 5 and 6 are achieved with a variety of methods in the literature, either by fitting airmass dependence and then sampling time variations in common between spectral channels directly from the data (Brogi et al. 2012, 2013, 2014, 2016), or by applying blind detrending algorithms such as principal component analysis (de Kok et al. 2013; Piskorz et al. 2016, 2017) or Sysrem (Birkby et al. 2013, 2017). All these approaches rely on the assumption that a certain spectral line from the exoplanet spectrum sweeps several spectral channels (several columns in our data matrix) during one night of observations, thus not influencing the detrending process significantly. In reality, whatever algorithm is applied to the data, the planet signal is stretched and scaled in the process. We show an example of this alteration in the bottom panel (step 7 - noiseless) of Figure 2, which is the end point of the data analysis applied to a noiseless data set. When compared to the initial spectrum at step 2, the end result clearly shows artifacts and scaling effects due to telluric removal. If unaccounted for, these artifacts can bias the retrieved planet parameters (velocities, abundances, and thermal vertical structure).

As in previous work, the data at step 7 is cross-correlated with models, and each cross-correlation value is converted into the $\log(L)$ value through Equation (9). To account for the stretching of the planet signal, we repeat steps 1–7 on each of the tested model spectra, but without adding the noise matrix N_{mod} . This mimics the effects of the data analysis on the model, and eliminates the biases, at a small computational cost.

The standard approach to interpret HRCCS observations would be to store the cross-correlation coefficients for each velocity, each spectrum, each detector, and each night, and to determine the “detection significance” of the planetary signal in the planetary–systemic velocity plane via the total cross-correlation coefficient summed over all observed orbital phases. This sum could be weighted to account for the different S/Ns of each spectrum or variable telluric or planet signal of CRILES detectors, introducing

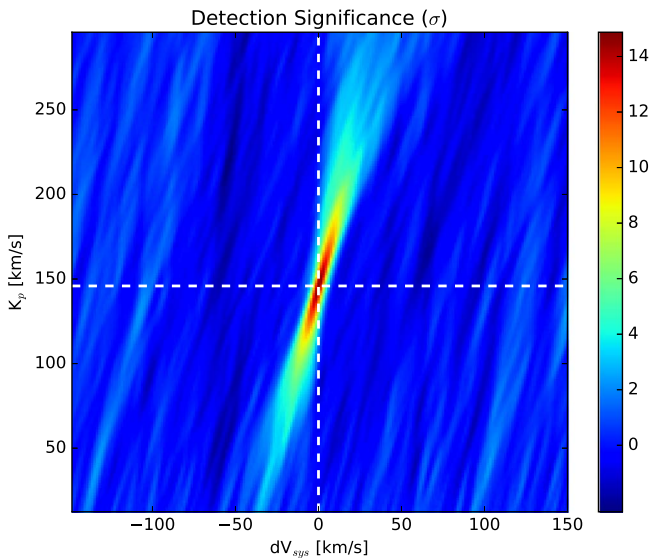


Figure 3. Planetary signal detected in the K_p - dV_{sys} plane. The x -axis is the differential system velocity (0 km s^{-1} represents no deviation from the known time-dependent system velocity) and the y -axis is the maximum radial (Keplerian) velocity of the planet. The figure is generated by summing the cross-correlation coefficient at all phases for each combination of K_p and dV_{sys} , and normalizing by the “noise,” which is an average of the cross-correlation coefficients over a portion of velocity space far from the planetary signal. This noise/planetary parameter setup results in a $\sim 12\sigma$ detection of the planetary atmosphere at the “known” velocities (white dashed lines).

some sort of subjectivity to the process. The significance of the detection is then measured either by taking the ratio between the peak value of the total cross-correlation and the standard deviation of the cross-correlation coefficients around it, or by applying a more sophisticated t-test on the distributions of the cross-correlation coefficients around and far from the planet radial velocity. If a high significance (usually larger than $\sim 4\sigma$) is obtained at the expected velocity pair, then the planetary atmosphere is said to be detected (Snellen et al. 2010; Brogi et al. 2012; Birkby et al. 2013, 2017; de Kok et al. 2013; Schwarz et al. 2015). Figure 3 shows the detection significance for the simulated planet signal (with the underlying CO and H₂O abundances given in Table 1), which should aid in the qualitative mapping between standard methods and the constraints obtained through our retrieval method. Given our simulated planetary/stellar/noise properties, we would obtain a $\sim 14\sigma$ detection of the planetary signal in the planetary–systemic velocity plane at the input velocities. This fairly optimistic signal compared to the majority of past CRIFES detections is a combination of a model particularly rich in spectral lines, and of the omitted effects of detector readout, thermal noise, additional photon noise from sky emission, and effects of damaged pixels and regions on the detector.

In this new framework, to obtain the total signal from the planet we just need to co-add all the log-likelihood values as a function of time, detector, and/or night of observation. Contrary to previous studies, there is no longer any need to weight cross-correlation functions, because our likelihood contains the data and model variances (s_f^2 and s_g^2 , respectively), hence it intrinsically incorporates the variable S/N of the observations. This is another significant advancement of our framework, and it adds objectivity to the retrieval process. In the following sections, we perform a series of exploratory

experiments with our novel HRCCS retrieval approach on this processed simulated data set.

3.3. A Simple HRCCS “Free” Retrieval Example

Figure 4 summarizes the retrieved properties on the simulated HRCCS data set (under the free retrieval assumption parameterized with the variables given in Table 1). These constraints are quite remarkable—on the order of 0.5 dex for CO and H₂O, despite the inclusion of realistic noise sources and common atmospheric retrieval parameterizations utilized in LRS data interpretations. It is encouraging to note that there is no bias in the retrieved parameters, including the scale factor a ,⁹ beyond what is expected due to the random noise instantiation ($N_{\text{mod}}(\lambda, t)$ in Equation (15)).

We notice some curious, but expected, degeneracies. First, the water and CO abundances are strongly correlated. Increasing the water abundance would require an increase in the CO abundance to maintain an acceptable log-likelihood. The reason for this degeneracy is that the retrieval tries to preserve the ratio between the CO and the water lines. Increasing both together, over some range, preserves this ratio. This suggests that the HRCCS data is highly sensitive to the abundance ratios. In fact, the precision on $\log(\text{CO})-\log(\text{H}_2\text{O})$, a good proxy for C/O, is ± 0.18 dex, about a factor of 2 better than for absolute abundances. Another noteworthy, but unsurprising, degeneracy is between the two velocities. This is simply a reproduction of the degeneracy between K_p and V_{sys} seen in the total cross-correlation signal/detection significance (Figure 3) that is easily lifted by repeating observations at different phase ranges. At least in this example, there do not appear to be any degeneracies between the abundances and the velocities. Additional degeneracies appear among the three parameters describing the T - p profile (not shown) that reflect the retrieval’s desire to maintain the temperature gradient over the 1 bar to 10 mbar region of the atmosphere.

It is worth noting that the tightest constraint on the water abundance through *HST* WFC3 emission spectroscopy (1.1–1.7 μm) is ± 0.6 dex (WASP-43 b, Kreidberg et al. 2014b). The high-resolution data, in particular this narrow slice of K -band spectra from 2.29 to 2.34 μm , show potential to constrain not only the water abundance to a higher precision, but also the CO abundance unobtainable with WFC3 alone.

The sensitivity to absolute abundances is perhaps the most unexpected outcome of this framework. One would expect that normalizing the data as described in Section 3.2 removed any sensitivity to absolute fluxes, hence absolute abundances. However, this sensitivity is partially recovered by choosing $a = 1$ in Section 2.2.2. This still does not set the absolute continuum; however, it sets the absolute planet line depth compared to the continuum of the star. Conditional to a proper normalization of the model (in planet/star units), this choice is completely consistent with the analysis of transit or eclipse spectroscopy at low spectral resolution. Furthermore, if significantly uncertain, the normalization parameters (stellar temperature, planet/star radii) can be inserted as free parameters with appropriate priors into the framework at nearly no computational cost.

We acknowledge that large uncertainties in the absolute scaling (or stretch) in the line-to-continuum contrast, due to

⁹ Leaving a as a free parameter allows us to verify that $a = 1$ is an unbiased choice.

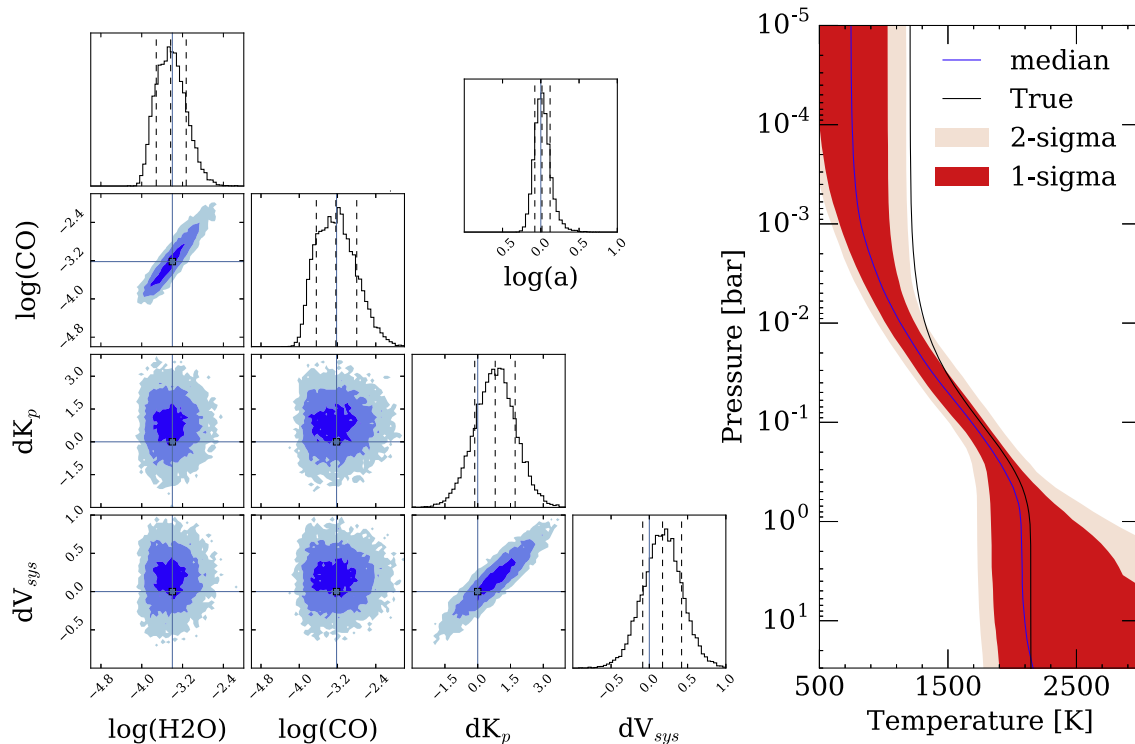


Figure 4. Retrieved constraints from the simulated HDS data set. Constraints are summarized with a corner diagram for the abundances and velocities, and the T - p profile spread reconstructed from the parameterization, as is common in the LRS retrieval literature (e.g., Line et al. 2013). We also show the marginalized constraint on the retrieved scale factor, a , to show that we indeed recover the expected maximum-likelihood estimator ($\hat{a} = 1$ or $\log \hat{a} = 0$). The water abundance is constrained to ± 0.3 dex, CO to ± 0.43 dex, and the velocities to ± 0.93 and ± 0.26 km s $^{-1}$, respectively. The difference between $\log(\text{CO})$ and $\log(\text{H}_2\text{O})$, a proxy for the C-to-O ratio (not shown), is constrained to ± 0.18 dex. The input values, or “truths,” are indicated with the vertical light blue lines and box in the corner plot. All corner plots in the remainder of the manuscript were generated with a modified version of the `corner.py` routine.

either uncertain planet/star properties or inaccurate telluric removal, would seemingly inhibit absolute abundance determinations. However, simply scaling the line-to-continuum ratio is not the same as increasing or decreasing the absolute abundances. Changing absolute abundances will not affect all lines equally due to their different relative positions on the curve of growth. Strong lines may saturate whereas weak lines will continue to increase their contrast relative to the continuum. It is this relative line depth (and shape) behavior of lines of the same gas that permits the absolute abundance constraints. Certainly there will be regimes where this degeneracy is prohibitive, such as in nearly isothermal atmospheres or ultra-low abundances of all gases over a given passband. However, such scenarios are likely to be rare.

It is also important to assess the statistical validity of the cross-correlation to log-likelihood mapping. Wilks’ theorem (Wilks 1938) suggests that the distribution of $\Delta \log(L)$ differenced from the maximum $\log(L)$ over a posterior probability distribution should follow a chi-square distribution with number of degrees of freedom equal to the number M of parameters—in this example, $M = 8$ free parameters (Table 1). The left panel in Figure 5 shows the distribution of $-2\Delta \log(L)$ drawn from the posterior probability distribution of our test case under the CC-to- $\log L$ mapping described in Section 2.2.2. The histogram of $-2\Delta \log(L)$ correctly follows a χ^2 distribution with eight degrees of freedom, suggesting an appropriate mapping.

As an additional check, though not shown, we performed 100 independent HRCCS retrievals on 100 separate photon noise instances ($N_{\text{mod}}(\lambda, t)$, but under the same telluric and planetary model, $F_{\text{mod}}(\lambda, t)$). From this experiment, we found

that the distribution of parameter means (due to random noise scatter) agreed with what was expected from the uncertainties derived on individual retrievals. In addition, the deviations from the truth values are random in the parameter space, i.e., they occur in random directions according to the particular noise instance. This fact strongly points to the absence of biases. The success of these robustness tests should not be surprising as the $\log(L)$ given by Equation (9) derives directly from inserting the relation between data and model in Equation (6) and carrying out the algebraic passages without approximations.

3.4. Comparison of CC-to- $\log L$ Mappings

Figure 6 compares the constraints obtained under the three different mappings. The mappings derived through Equation (9) in this work and Equation (2) in Zucker (2003) provide comparable parameter constraints. Nevertheless, the Zucker (2003) mapping results in a bias in the medians of the retrieved H $_2$ O and CO abundances. These differences arise because the two mappings respond differently to a particular noise instantiation due to the inclusion (or lack thereof) of the s_f^2 and s_g^2 terms. As with our mapping, we find that the distribution of $-2\Delta \log(L)$ derived from the Zucker (2003) mapping agrees well with a chi-square distribution of eight degrees of freedom, as required by Wilks’ theorem (Figure 5, left and middle panels). As an example of the importance of using the correct CC-to- $\log L$ mapping, we also show the consequence of applying Equation (3) incorrectly to our data. In Section 2.2.1, we explained that the mapping described in Lockwood et al. (2014) is only valid for stationary planet

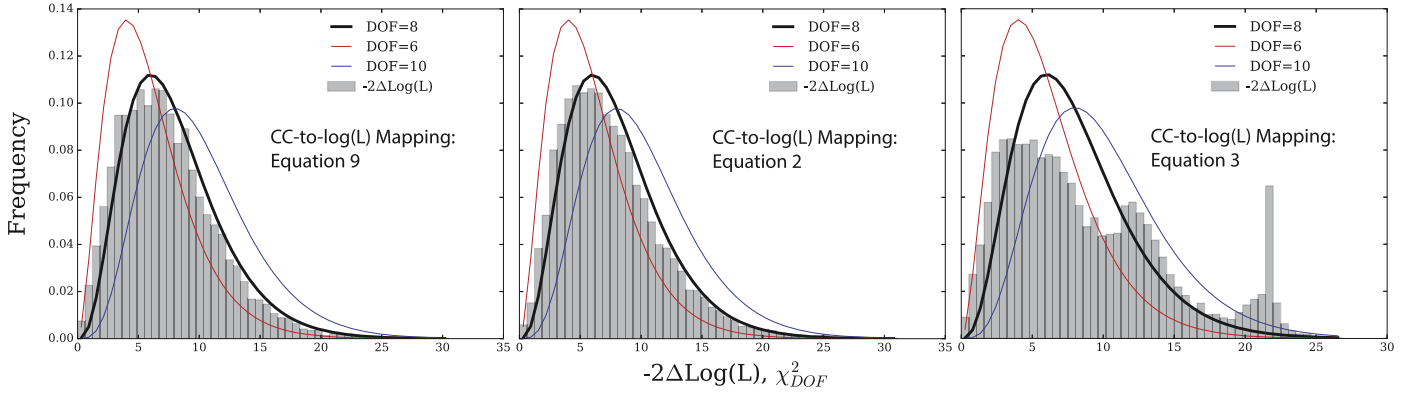


Figure 5. Test of the robustness of our cross-CC-to-log(L) mapping. Wilks’ theorem states that the test statistic $-2\Delta\log(L)$ for an M -parameter (in this case $M = 8$) estimator should follow a chi-square distribution with M degrees of freedom (DOF). Here, $\Delta\log(L)$ is the difference in the log-likelihood between the maximum likelihood and all other likelihoods within the posterior probability distribution. Histograms of this test statistic computed with our CC-to-log L mapping (Equation (9), left panel) or with the mapping of Zucker (2003) (Equation (2), middle panel) closely follow a χ^2_8 distribution as expected. By contrast, if we incorrectly apply the mapping of Lockwood et al. (2014) to our data (Equation (3) and Section 2.2.1), the resulting histogram does not appear to follow any χ^2_{DOF} distribution.

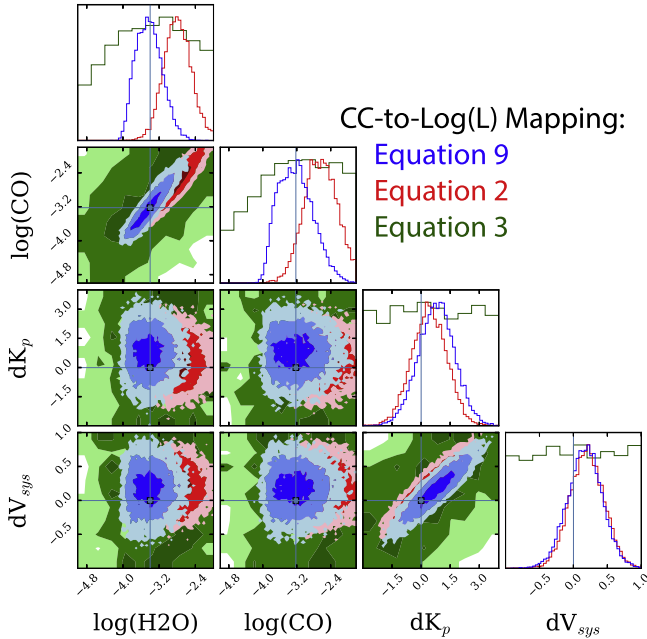


Figure 6. Comparison of constraints obtained under different CC-to-log L mappings on the same simulated data set. The mapping derived in this work (Equation (9)) is shown in blue (same as Figure 4), Zucker (2003) in red, and Lockwood et al. (2014) in green. Our mapping and that of Zucker (2003) result in similar constraints, albeit with a parameter bias when using Zucker (2003). The Lockwood et al. (2014) mapping provides virtually no constraint on our data set, which is expected because such a formalism is not applicable to our CRIRES data (see Section 2.2.1). Note that the parameter prior ranges are broader than the plot axes.

signals, and indeed forcing such mapping on our data analysis results in very broad, virtually non-existent constraints on all of the parameters (Figure 6, green contours). Furthermore, the resulting distribution of $-2\Delta\log(L)$ clearly does not follow a chi-square distribution (Figure 5, right panel), confirming that this mapping is not appropriate for spectral time sequences where the planet radial velocity changes rapidly.

3.5. Impact of H_2O Cross-sections

Brogi et al. (2017) explored the potential impact of incomplete or incorrect line lists on the cross-correlation

signal. That work concluded that these uncertainties were not important mainly because water vapor was not detected in the high-resolution spectra. Having tested simulated spectra with both CO and H_2O in this work allows us to reassess the importance of cross-sections in a completely controlled environment, and we present in this section the main results.

Our simulated spectral sequence is the same as in Section 3.3, and it uses H_2O cross-sections calculated from Freedman et al. (2014) based upon the Partridge & Schwenke (1997) line list. We run two separate retrievals on this spectral sequence. The former uses the same absorption cross-sections and should therefore result in unbiased parameters. The latter uses absorption cross-sections generated from HITEMP (Rothman et al. 2010) via the HITRAN HAPI routine (Kochanov et al. 2016) and is based on the Barber et al. (2006) line list. These cross-sections are computed at 0.003 cm^{-1} sampling resolution, 100 cm^{-1} Voigt wing cutoff. They assume “air” broadening over a well-sampled pressure and temperature grid (Freedman et al. 2014, 2008). Although there are subtle differences¹⁰ in the line profile assumptions between the two sets of absorption cross-sections, the main impact on HRCCS is the variable line position arising from different choices of line lists. We note that Shabram et al. (2011) explored the differences in these two line lists at low resolution and found a negligible difference.

Figure 7 shows the impact of the line-list/cross-section assumptions. By eye (top panel), it is easily seen that the two spectra are not perfectly overlaid on each other at the resolutions attainable by CRIRES. The line position differences are not uniform in wavelength, suggesting that these differences cannot be compensated by a single velocity offset. These inconsistencies, when combined over the entire CRIRES K -band, result in substantial biases in the retrieved parameter distributions, which are offset by many sigma from their true state (Figure 7). The constraints on the abundances are also much tighter when using the “incorrect” line list. It is a reasonable question, in the case of actual data, to ask which line list is correct. Unlike LRS data, with HRCCS data it is difficult to obtain a “visual” model fit to make such assessments. We therefore rely upon the Bayesian evidence to guide us (a natural output of the nested-sampling algorithm used in this work). The

¹⁰ The difference in line widths due to air versus H_2/He pressure broadening is well below the CRIRES instrumental resolution.

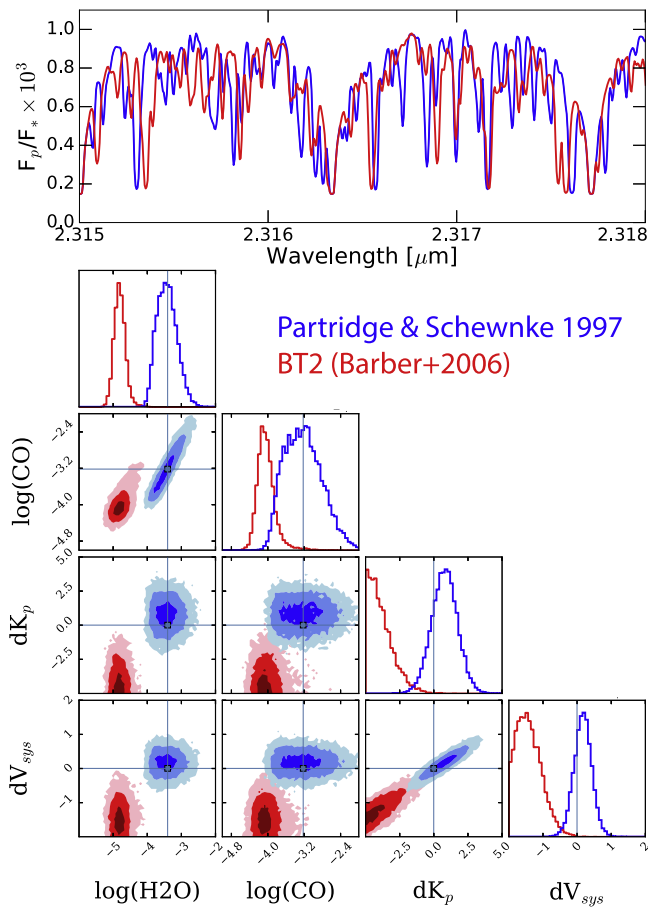


Figure 7. Impact of the line-list choice on the retrieved parameters. We retrieve on the same synthetic data set under two different line-list assumptions: in blue, Partridge & Schwenke (1997) implemented with the Freedman et al. (2014) cross-sections (used to generate the “true” model) and HITEMP based upon the BT2 (Barber et al. 2006) line list (red). The top panel compares spectra (using the same inputs given in Table 1) generated with the different absorption cross-section databases over a small portion of the K -band. The lines are clearly haphazardly shifted. The posterior probability distribution summarized in blue is the same as in Figure 4. It is clear that there are significant parameter biases and uncertainty differences when retrieving with different sets of water absorption cross-sections utilizing different underlying line lists.

log-Bayes factor between the model using the “correct” line list (blue in Figure 7) and the model using the “incorrect” line list (red in Figure 7) is 147.4. Note that there is no change in the number of parameters between the two models. Since a log-Bayes factor larger than 5.0 on the Jefferey’s scale is considered significant, this extreme difference overwhelmingly favors the model utilizing the “correct” line list. This conclusion would not be apparent from the posterior alone, hence it is always important to compare the model evidences.

Performing the same experiment within the chemically consistent model framework results in a similar degree of bias (in metallicity and carbon-to-oxygen ratio) and large Bayesian evidence differences. This suggests that line-list biases may exist regardless of the model parameterization.

It is important to note that if our simulated data set was calculated with cross-sections from Rothman et al. (2010), the results would have been the opposite, i.e., the retrieval with Freedman et al. (2014) cross-sections would have been biased. This is an important point to make, because although there is no extensive testing in the literature regarding the choice of line

lists at high spectral resolution, all the past H_2O detections with CRILES were achieved with the line lists from Rothman et al. (2010). When CO and H_2O were detected simultaneously, measured radial velocities (K_p and V_{sys}) were consistent between species, which suggests unbiased water determinations with this database. The only explicit test of different line lists in the literature, besides that of Brogi et al. (2017), is mentioned in Flowers et al. (2018). In the latter paper, water opacities from Lupu et al. (2014; which are similar to those in Freedman et al. 2014) did not produce any correlation signal with CRILES transmission spectra of HD 189733 b, in contrast to the $>5\sigma$ detection of Brogi et al. (2016) obtained with the HITEMP database.

The results of this section are rather alarming, and we should take this as a further warning that all HRCCS interpretations are going to be strongly dependent on the choice of line lists used in the model templates. From a purist retrieval modeling perspective, line-list properties (positions, broadening, strengths) should be parameterized so as to appropriately marginalize over them within an HRCCS retrieval. However, this is extraordinarily unwieldy as this would slow down a forward model computation to the point of being unusable within a retrieval framework. Furthermore, a proper way of readily parameterizing these effects other than the standard brute-force, line-by-line computation does not yet exist. Such an approach, to be feasible, would likely have to make assumptions and approximations that would introduce additional uncertainties to the point of obviating its purpose. Instead, we strongly advocate for further laboratory, astrophysical, and ab initio efforts to determine accurate line positions, intensities, and broadening for exo-atmosphere relevant molecules and conditions (e.g., Fortney et al. 2016). We would anticipate that as we push toward cooler temperatures or more Earth-like conditions, we will have the opportunity to better quantify and reduce line-list uncertainties by validating the output of radiative transfer calculations on spectra of our own planets or other solar system planets. More accurate line lists will likely increase the overall level of correlation with terrestrial exoplanet spectra and minimize biases in retrieved parameters, which will be crucial for the robust assessment of habitability and biosignatures.

3.6. Combined LRS and HRCCS

Once a statistically robust CC-to-log L mapping is achieved, one can trivially combine information from various data sets. One such potentially useful combination is that between low-resolution spectroscopy with *HST* WFC3 and high-resolution spectroscopy with VLT CRILES or Keck NIRSPEC. In this section, we combine simulated *HST* WFC3 data (loosely based upon the typical emission observations—30 ppm/channel at $0.035 \mu\text{m}$ bins) with the above CRILES K -band data within the chemically consistent modeling framework (whereby metallicity and C/O are retrieved under the assumption of equilibrium chemistry).

Trivially, to combine the inference from different sets of data, we just sum their log-likelihoods:

$$\log(L_{\text{tot}}) = \log(L_{\text{HDS}}) + \log(L_{\text{LDS}}), \quad (16)$$

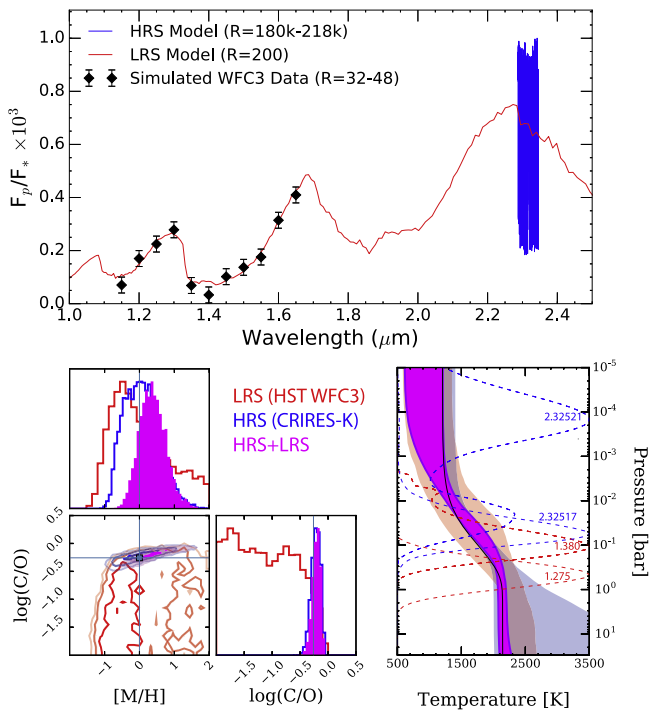


Figure 8. Simulated combined observations with *HST*/WFC3 (black dots, top panel) and VLT/CRILES around $2.3 \mu\text{m}$ (blue lines, top panel). The bottom panels show the posterior distributions for planet metallicity ($[M/H]$), carbon-to-oxygen ratio (C/O), and T - p profile (bottom-right panel; summarized with the 68% confidence intervals) obtained by running our framework on the *HST* data alone (red curves), VLT data alone (blue curves), and on the combined data set (magenta curves). The dashed curves (blue = CRILES, red = WFC3) are the temperature Jacobians at the indicated wavelengths (on and off band/line). In general *HST* WFC3 probes a relatively deep and narrow region. In contrast, the high dynamic range in the CRILES spectrum permits broad altitude coverage. Combining low- and high-resolution spectra leads to a substantial improvement in the precision of these measurements.

where $\log(L_{\text{HDS}})$ is given by Equation (9) and

$$\log(L_{\text{LDS}}) = -\frac{1}{2}\chi^2. \quad (17)$$

The top panel of Figure 8 shows the spectral regions covered by these simulated data and is representative of observations that would have been already feasible five years ago. The bottom panels show the retrieved planet properties by running our framework on the two data sets separately (WFC3 only in red, CRILES K -band in blue, and on their combination in magenta derived through the combined log-likelihood function in Equation (16)) under the chemically consistent model parameterization (Table 1).

The simulated *HST* WFC3 data alone are able to constrain the planet metallicity ($[M/H]$) to within 1.7 dex (68% confidence), slightly better than expectations from published results (Line et al. 2016). This uncertainty is primarily driven by the degeneracy between the C/O and metallicity in producing the same water abundance (H_2O is the main measurable gas over the WFC3 passband). In contrast, the K -band CRILES observations alone result in a 1.0 dex (68% confidence) constraint, nearly a factor of 2 better than that obtained with WFC3. However, the power in these particular HRCCS observations is in their ability to constrain the carbon-to-oxygen ratio to within 0.2 dex (68% confidence), driven by the enhanced sensitivity to CO -to- H_2O line ratios at high-resolution. WFC3 observations alone naturally struggle to

constrain the C/O (upper limits only), due to the lack of presence of carbon-bearing species in this wavelength range. In this particular setup, however, combining these two data sets through a joint likelihood does little to improve beyond the HRCCS constraints alone. In fact, the abundant parameter space mass present at higher metallicities in the WFC3 only scenario tends to pull the joint constraint toward these higher metallicities. The *HST* observations, however, help in constraining the T - p profile as they are weighted toward deeper layers of the atmosphere. Certainly this is only a single example of the combination of these two data sets—perhaps unfairly at the detriment to the *HST* observations. We imagine that perhaps a more synergistic setup might be with LRS mid-infrared ($>5 \mu\text{m}$), which are best obtained from space (e.g., *JWST* MIRI), combined with ground-based near-IR HRCCS observations. This would provide maximum leverage of the strengths of each approach.

Since we are testing a data set based on a planet without a thermal inversion, in terms of the T - p profile, the two data sets probe reasonably similar regions of the planet photosphere (Figure 4, bottom-right panel), with only the CO line cores absorbing at significantly lower pressures (up to 10^{-4} bar). Nevertheless, the combined data set allows us to precisely determine both the atmospheric lapse rate and the photospheric temperatures.¹¹ In the case of high-altitude thermal inversion layers (temperature increasing with altitude), we anticipate that the complementarity between low- and high-resolution spectra would be even more evident, with the two constraining mostly the lower and the upper atmospheres, respectively.

3.7. Influence of Missing Gases

As with LRS retrieval analyses, failure to include all of the key relevant opacity sources in an HRCCS model could result in significant biases in inferred atmospheric properties. We perform a simple experiment to explore the impact that unaccounted gases can have on the retrieved atmospheric properties.

We utilize the same setup as in Section 3.3 (the “free retrieval”) but include in the simulated “true” spectrum additional molecular opacities due to CH_4 , NH_3 , CO_2 , and HCN ($\log(\text{mixing ratio}) = -5.0, -5.0, -4.5,$ and -9 , respectively), though the latter two gases have little impact. Figure 9 shows the abundance-weighted contributions of the dominant gases to the spectrum.

We then perform three different retrievals (summarized in Figure 10). The first scenario (red posterior distribution in Figure 10) zeros out the abundances of the other gases within the retrieval and fits for the standard “free retrieval” parameters, as in Section 3.3. In other words, this scenario fails to account for all of the opacity sources. The resulting constraints are biased and produce large uncertainties. The water abundance is poorly constrained. The second scenario (blue) retrieves for all six gases. That is, the wide uncertainties reflect the full marginalization over all of the included gases, but there is no bias. This is the “most correct” of the three scenarios. Finally, the last scenario (green) retrieves again the “default” set of parameters, but we assume we have perfect a priori knowledge of the other four gases of which are fixed to their true input

¹¹ We are aware that these may be overly optimistic constraints due to the particular choice of T - p profile parameterization. Certainly the temperature is not constrained this precisely over the entire atmospheric column shown.

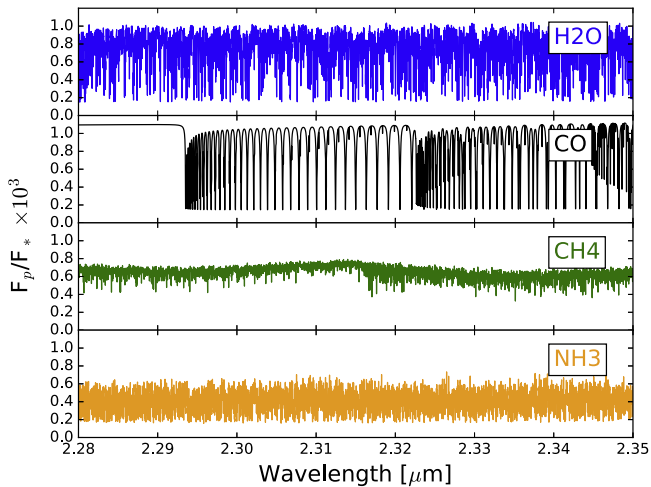


Figure 9. Spectral components of the important absorbers over the CRILES K -band. H_2O and CO have the strongest influence, whereas CH_4 and NH_3 contribute mostly to continuum absorption due to their reduced line-to-continuum contrast. These spectra are at the native line-by-line cross-section resolution.

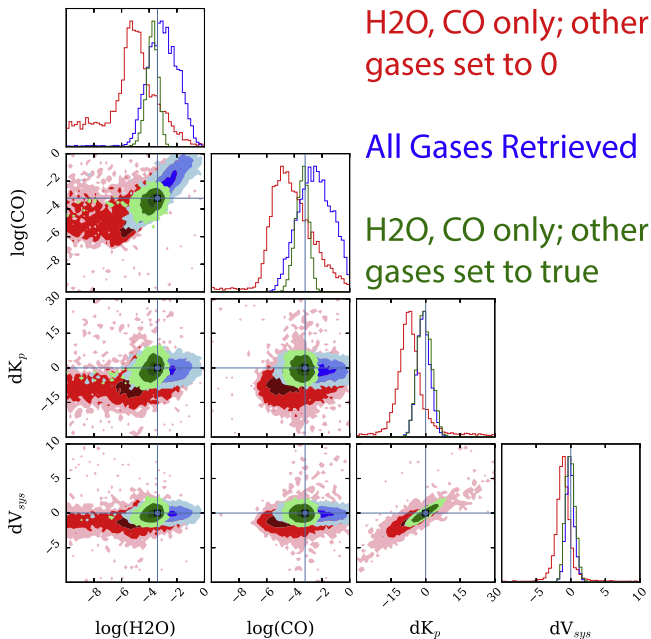


Figure 10. Influence of unaccounted-for absorbers on the water, CO, and velocity constraints. The underlying true model includes opacities from H_2O , CO , CH_4 , CO_2 , NH_3 , and HCN in near-solar proportions, though only H_2O , CO , CH_4 , and NH_3 noticeably contribute. The posterior distributions summarized in red represent the scenario for which only H_2O and CO are retrieved, and the other opacities are turned off. This results in both large uncertainties and biases in the constraints. Blue represents a full marginalization over all six gases. The uncertainties are larger, but there is no bias. Finally, green represents a scenario where the other gases are fixed to their true input values but only H_2O and CO are retrieved. This produces the most precise constraints.

values. Of course, this would rarely be the case. This results in artificially tight constraints, unsurprisingly, as there are fewer “free” gases to confuse water and CO. The Bayesian evidence overwhelmingly favors, obviously, the latter scenario (log-Bayes factor of 22.8 between scenario 3 and 1, and 34.3 between 3 and 2). However, perhaps unfortunately, the first scenario (H_2O , CO only; other gases set to 0) is favored (11.5) over the second (all gases retrieved). Relying upon the Bayes

factor alone, in this situation, would lead one to conclude that the $\text{H}_2\text{O}/\text{CO}$ only model is the correct one, when in fact it is not. This is likely because of the unnecessarily large prior volume due to the inclusion of CO_2 and HCN in the retrieval despite their negligible impact on the spectrum (due to their low abundances). In practice, as is now routinely done in LRS modeling analysis (e.g., Kreidberg et al. 2015), nested model comparison with sequential removal of gases relative to some parent model should be performed to determine the “optimal” number of gases to include.

In summary, it is extremely important to be cognizant of all of the potential sources of opacity present over a given high-resolution bandpass and how their lack of inclusion could bias atmospheric inferences. This will indeed become increasingly important with broader spectral range instruments slated to come online in the not too distant future.

4. Application to CRILES HD 209458 b and HD 189733 b K -band Data

4.1. Reduction Process and Setup

We applied the framework described in Sections 2 and 3 to real K -band dayside spectra of exoplanets obtained with CRILES at the VLT. We reanalyze the half night of observations of HD 189733 b published in de Kok et al. (2013) and the two half nights of HD 209458 b presented in Schwarz et al. (2015) and Brogi et al. (2017), to which we point the reader for more information. Here we recall that these data sets were observed around the strong 2–0 rovibrational band of carbon monoxide at $2.3 \mu\text{m}$, but also that they should contain additional opacity from water vapor. Although H_2O was indeed detected in the same CRILES range for other exoplanets observed in dayside (Brogi et al. 2013, 2014), it was not detected for these two planets. In the case of Schwarz et al. (2015), only a tentative detection of CO was presented, whereas a more advanced weighting of the CRILES detectors allowed Brogi et al. (2017) to recover the signal of CO at an $\text{S/N} = 5$, but again no water. Recently, Hawker et al. (2018) revisited the data set with a different detrending algorithm for telluric lines and confirmed the detections of CO and H_2O .

Since we have substantial literature to support the potential of these data, in this work we apply the most objective analysis process by matching steps 4–7 described in Section 3.2 and visualized in Figure 2. In contrast to past work, this analysis does not require the optimization of detrending parameters and is therefore ideal to apply our framework to data as uniformly as possible.

There is only one extra step required in the analysis of HD 189733 b data. With a spectral type of K1V, the parent star shows strong CO absorption lines in the CRILES spectral range. These lines are not completely stationary in wavelength in the observer reference frame, because the barycentric velocity changes by about 0.5 km s^{-1} during the 5 hr of observations. This is due to the changing orbital and rotational radial velocity of Earth compared to the center of mass of the solar system. Fortunately, we have devoted abundant work in the past to the correction of stellar CO lines from HD 189733. In this context, we apply the state-of-the-art three-dimensional modeling of the stellar photosphere described in Magic et al. (2013) and Chiavassa et al. (2018), and implemented as in Flowers et al. (2018). We divide out this modeled stellar spectrum between steps 4 and 5 of the analysis (see Figure 2), i.e., just before the removal of telluric lines. We note that this

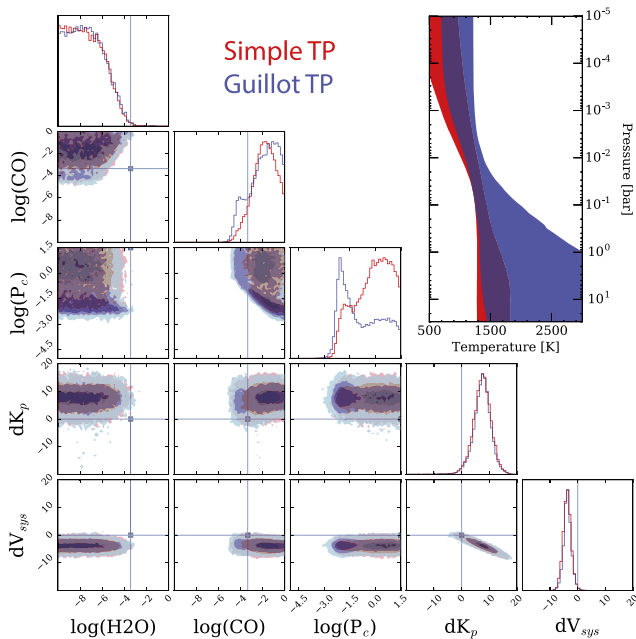


Figure 11. Summary of the HRCCS retrieval results for the CRILES K -band dayside emission spectrum of HD 189733 b. The blue histograms summarize the posterior under the “default” Guillot (2010) T - p profile parameterization, and red the “simple” T - p profile parameterization described in Line et al. (2016). The light blue lines/box within the 1D/2D histograms indicate the approximate mixing ratios predicted by thermochemical equilibrium at solar abundance and the zero-offset velocities. Only an upper bound on water and a lower bound on CO are retrieved, with the cloud and T - p profile parameterization having a negligible impact.

modeling is not parametric. Being completely self-consistent, our stellar 3D models only assume an initial metallicity for the star, which is well constrained in the literature and has been also verified by inspecting the shape and depth of stellar CO lines in the CRILES data. Therefore, there is no subjectivity in the correction of the stellar spectrum, as no extra fitting or optimization is required at this stage.

One final caveat for the analysis of real spectra is that we do not have an accurate model for the temporal variations of the telluric spectrum, which is needed to replicate the stretching and scaling of planetary signals due to telluric removal. We thus store the fitted telluric absorption spectrum obtained through steps 4–6 of the analysis and use it to process each of the tested models, following the same prescriptions as for the simulated data set (Section 3.2).

The forward model used for both objects deviates slightly from that used in the simulated case (Section 3.3) in that we add an optically thick gray cloud parameterized by a cloud-top pressure (CTP, $\log(P_c)$) and explore an additional “simple” T - p profile parameterization similar to that described in Line et al. (2016). We also adjusted the prior upper bound on the irradiation temperature (T_{irr} in Table 1) for each object to prevent unphysically hot temperatures.

4.2. HD 189733 b Retrieval Results

Figure 11 summarizes the HRCCS retrieval results within our framework. We find only an upper limit on the water abundance, which is significantly lower than expected for solar elemental ratios, and a lower limit on the CO abundance but consistent with solar expectations. The lower pressure limit on

the CTP is consistent with a cloud-free dayside; however, there is a notable degeneracy with the CO abundance whereby decreasing the CTP (higher altitude) results in an increase in CO abundance. This is simply understood as the competition between the muting of the line-to-continuum ratio by the cloud and increasing line-to-continuum ratio with increasing CO abundance. However, not including the cloud has little impact on the CO abundance. We also find that the abundances are largely insensitive to our choice of T - p profile parameterization, even if the retrieved T - p profiles themselves are different.

The planet velocities are shifted by a few km s^{-1} from their nominal values. We emphasize that it is important to marginalize over the velocities as uncertainties in orbital properties, especially years after their publication, can result in artificial velocity shifts. In fact, even for well-known exoplanets such as HD 189733 b, errors in the quantities defined in Equation (1), in particular K_p and φ , can lead to radial velocity uncertainties of a few km s^{-1} , well above the sensitivity of these observations. A more subtle aspect is that many of the fundamental orbital parameters such as semimajor axis and orbital period can be correlated, hence using the reported error bars in the literature might lead to lower limits only on the final uncertainties.

In order to assess the uncertainties in planet velocities, we compute the error bars on K_p from the stellar radial velocity amplitude and planet/star mass ratio, and the error bars on the orbital phase φ from the time of midtransit and the orbital period, all in the hypothesis of circular orbits. In the case of HD 189733 b, we obtain negligible errors in the orbital phase, but a ΔK_p of about 6 km s^{-1} , which is sufficient to bring the retrieved K_p within the 1σ uncertainty.

The obvious finding from the retrieval of HD 189733 b data is the strong detection of CO (lower abundance limit 4.67σ , according to the nested-sampling-derived Bayesian evidence ratio) and non-detection of water (upper abundance limit). Although this is surprising from solar elemental ratios and thermochemical arguments, it is not in the context of previous analysis of this data set. Using a completely independent analysis on the same $2.3 \mu\text{m}$ data, de Kok et al. (2013) reported a detection of carbon monoxide at 5σ and no detection of water vapor. In contrast, the L -band data presented in Birkby et al. (2013) showed a clear detection of water vapor at 4.8σ . The spectra are taken with the same instrument (CRILES), but at wavelengths with radically different water opacity. It is possible that a moderately low abundance of water vapor (e.g., $\text{VMR} \sim 10^{-5}$) produces a water spectrum too weak to be detected at $2.3 \mu\text{m}$ but sufficiently strong to dominate the spectrum at $3.2 \mu\text{m}$. Through transmission spectroscopy, water vapor is detected both at $2.3 \mu\text{m}$ (Brogi et al. 2016) and over the entire NIR (Brogi et al. 2018), though the abundance of H_2O is unreported in those works. We note that a recent independent reanalysis of the CRILES spectra above (Cabot et al. 2019) confirms the detection of CO at $2.3 \mu\text{m}$ and H_2O at $3.2 \mu\text{m}$. Additionally, HCN absorption is found in the $3.2 \mu\text{m}$ data.

Our retrieved T - p profile is qualitatively consistent with that of de Kok et al. (2013) and Birkby et al. (2013), who both rule out a strong inversion due to the poor correlation with atmospheric models containing emission lines.

Finally, as motivated by our findings in Section 3.5, we again explored the impact of the line-list choice. Unsurprisingly, because

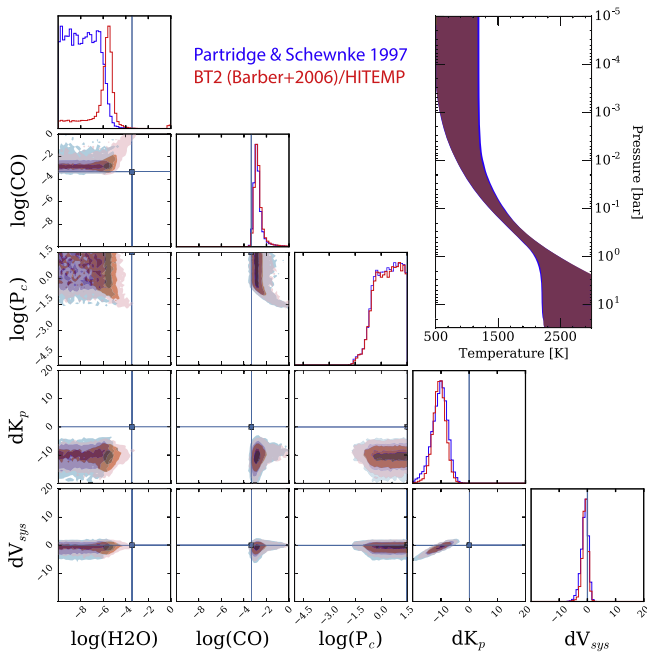


Figure 12. Summary of the HRCCS retrieval results for the CRRES K -band dayside emission spectrum of HD 209458 b. The blue histograms summarize the posterior under the “default” water line list (Partridge & Schwenke 1997; Freedman et al. 2014), and red using HITEMP (Rothman et al. 2010; Barber et al. 2006). Both scenarios assume the Guillot (2010) T - p profile parameterization. The light blue lines/box within the 1D/2D histograms indicate the approximate mixing ratios predicted by thermochemical equilibrium at solar abundance and the zero-offset velocities. An upper limit on the water abundance is obtained under both line-list scenarios, but the CO abundance is fairly tightly constrained at values slightly higher than expected under solar composition, thermochemical equilibrium at these temperatures.

there is a lack of detection of water, we found virtually no difference in the posterior probability distribution.

4.3. HD 209458 b Retrieval Results

Figure 12 summarizes the posterior probability distribution under the assumption of the default T - p profile parameterization from Section 3.3. As with HD 189733 b, we again find only a subsolar upper limit on the water abundance. However, we obtain a rather stringent constraint (± 0.3 dex, resulting in a 7.44σ evidence-based detection) on the CO abundance, which is only marginally supersolar (0.5 dex higher, or just over 1σ). We again find a lower pressure limit on the CTP with a similar, albeit with a less pronounced, degeneracy with CO.

Similar to HD 189733 b, the dayside spectrum of HD 209458 b was also observed with CRRES at both 2.3 and 3.2 μm . The 2.3 μm data were originally published by Schwarz et al. (2015) and resulted in a marginal detection of CO absorption, no detection of water, and rejection of strong inversion layers. A reanalysis by Brogi et al. (2017) found that the planet signal was very unequally distributed across the four detectors of CRRES and could recover CO absorption signal at $S/N = 5$ by unequally weighting the data.

These data along with previously unpublished 3.2 μm data were recently analyzed by an independent team (Hawker et al. 2018). They reported the detection of both H_2O and CO absorption in K -band data, and in addition, the L -band data reveal clear absorption from HCN but not from H_2O . Therefore, opposite to HD 189733 b, water is detected at 2.3 μm but undetected in the L -band.

In contrast to HD 189733 b, though, the CRRES transmission spectrum of HD 209458 b at 2.3 μm does not show any water absorption lines. As this was the first successful detection published with HRCCS (Snellen et al. 2010) and the analysis was optimized for CO lines, it is possible that the sensitivity of the data was not sufficient to detect H_2O .

We again explore the impact of the water line list, as with HD 189733 b. In general we do not find that the choice of water opacity influences the primary conclusions but it does change the shape of the marginalized water histogram toward a more concentrated solution near 1 ppm. However, there still exists a non-negligible tail extending to lower abundances, suggesting again that this is still just an upper limit, consistent with a non-detection (according to the Bayesian evidence ratios).

4.4. Implications

It is worth briefly discussing how these results compare to LRS retrievals from emission spectroscopy with *HST* and *Spitzer*, as both objects have been thoroughly investigated with these instruments. The dayside emission spectrum of HD 189733 b is by far the most complete in terms of wavelength coverage and “spectral density” composed of *HST* WFC3, NICMOS, and *Spitzer* IRAC/MIPS/IR (see Lee et al. 2012; Line et al. 2014, for a summary), though with questionable reliability of some of the data sets (Gibson et al. 2011; Hansen et al. 2014). Retrievals on this emission data set have suggested water and CO abundances that are broadly consistent with solar thermochemical expectations (Lee et al. 2012; Line et al. 2014). Our retrieved abundances for CO are rather consistent with those previous findings (e.g., $\sim 10^{-5}$ – 10^{-2} in Line et al. 2014) but our retrieved water abundance upper limit ($\sim 10^{-4}$, Figure 11 top histogram) skirts the lower bounds of previous findings (see, e.g., Table 3 in Line et al. 2014). The most recent analysis of the fairly complete dayside emission spectrum of HD 209458 b (Line et al. 2016) includes a strong water vapor absorption feature over the *HST* WFC3 bandpass suggestive of water mixing ratios no lower than 10^{-5} , right near the upper limit of what we retrieve (Figure 12, top histogram). Brogi et al. (2017) investigated both the LRS (Line et al. 2016) and HRS data sets (Schwarz et al. 2015) and found that the water abundances are largely inconsistent, due to the fairly stringent lower bound from the LRS data and lack of water detection in the HRS data.

Both objects also have fairly precise *HST*/WFC3 (Deming et al. 2013; McCullough et al. 2014) and *HST*/STIS (Sing et al. 2011, 2016) transmission spectra, with multiple independent analyses suggesting water abundances that span a broad range from $\sim 10^{-6}$ to 10^{-2} depending on the specific analysis and data sets used (Madhusudhan et al. 2014; Line & Parmentier 2016; MacDonald & Madhusudhan 2017; Tsiaras et al. 2018), but no constraints on CO due to the limited wavelength coverage.

The reason for the HRS–LRS water inconsistency is not immediately clear. Whereas one would be tempted to blame it on water line-list uncertainties, our tests reveal no impact of the different line lists on the analysis of HD 189733 b data, and only a marginal impact on the analysis of HD 209458 b. One additional speculation relates to the different atmospheric pressures at which the core of H_2O and CO lines are formed. As low- and high-altitude wind patterns can differ by many km s^{-1} in hot Jupiters, it

is possible that CO and H₂O lines track slightly different Doppler velocities. As water has the biggest dynamic range in terms of weak/strong lines, it is the most affected by vertical wind shears, to the point where the cross-correlation signal is smeared below detectability. Only a comparison with three-dimensional general circulation models as recently shown by Flowers et al. (2018) will help us determine if atmospheric circulation has a detectable impact on the dayside spectra of hot Jupiters observed through HRCCS. In general, it would not be surprising that the 3D nature of a planet can and will play a role in interpreting HRS data as these data are observed over a range of orbital phases with the potential for spatially variable temperatures, composition, clouds, and winds.

5. Conclusions

This work demonstrates that an appropriate mapping of the cross-correlation coefficient to likelihood function allows us to coherently merge high-resolution cross-correlation spectroscopy with powerful atmospheric retrieval techniques. This approach opens up a new avenue in interpreting HRCCS data and fully exploits the information buried within it. Below we summarize our primary developments and findings:

1. Developed (Section 2.2.2) a novel mapping from cross-correlation to log-likelihood and demonstrated that it is statistically appropriate, permitting accurate parameter confidence intervals (Figure 5).
2. Formalized data analysis techniques, in particular the removal of telluric lines, and made them applicable to this new formalism without artificially scaling or biasing the planetary signal (Section 3.2).
3. Explored the potential for this novel framework to constrain fundamental atmospheric properties like temperatures and abundances on a realistically simulated data set (Figure 4, Section 3.3). Despite being more degenerate than in LRS data, water abundance constraints are comparable to those obtained with *HST* WFC3. In addition, due to tight constraints on the relative molecular abundances of CO and H₂O, HRCCS + retrievals permit precise C-to-O ratio determinations.
4. Provided a comprehensive comparison and discussion of the strengths and weaknesses of other correlation coefficient to log-likelihood mappings (Sections 2.2.1, 3.4, Figures 5 and 6).
5. Determined the significance of the impact that water opacities can have on the results. Precise knowledge of exo-atmosphere relevant opacities is required at high resolution for these approaches to work (Section 3.5, Figure 7).
6. Explored the impact of missing gases on the retrieved constraints. Failure to include gases could result in biases and/or artificially broadened constraints (Section 3.7, Figure 10).
7. Provided a simple framework for combining HRCCS data and LRS data within a unified likelihood function (Section 3.6, Figure 8). Such an approach leverages the strengths of both types of data in a way that is analogous to combining radial velocity and transit data.
8. Applied the framework to existing dayside observations of the transiting hot Jupiters HD 189733 b and HD 209458 b, obtained with CRIRES at the Very Large

Telescope. In spite of a clear signature of CO absorption, H₂O is not clearly detected, regardless of the line list used, T - p profiles implemented, or assumptions on the presence of a thick cloud deck. We also rule out confidently the presence of inversion layers in the atmospheres of these two planets.

Currently, there is a shortage of observations of transiting exoplanets with both *HST*/WFC3 and high-resolution spectrographs from the ground. The main reason for this is that CRIRES and NIRSPEC, the two most active instruments to provide HRCCS observations in the past 5–10 yr, have relatively poor throughput and spectral range. They are thus limited to observing the brightest exoplanets in the sky, which are mostly non-transiting. However, modern spectrographs have drastically superior simultaneous spectral range and equal or better throughput. This increase in sensitivity can be used to move HRCCS observations to smaller telescope facilities (Brogi et al. 2018), to enable observations of fainter planets, or both, as in the case of the highest performance instruments such as CARMENES or SPIRou. This technological evolution timely matches the future availability of *JWST* observations, which will also have increased sensitivity, spectral range, and in some cases spectral resolution. It is also well timed with the recent launch of the *TESS* satellite, which will find a significant fraction of exoplanets orbiting bright-enough stars to be followed up for atmospheric characterization. Lastly, the age of the next-generation large telescopes is just a few years away. With the construction of the Giant Magellan Telescope and the Extremely Large Telescope already ongoing, and first-light high-resolution instrumentation approved for both telescopes, it is crucial to develop techniques to pair the most exquisite *JWST* observations to the enormous collective power of the GMT and ELT telescopes.

This paper sets the foundation for rigorous analysis of HRCCS data sets and their coherent combination with LRS data. Future work will focus on three main directions:

1. Implementing the retrieval on transmission spectroscopy data and jointly analyzing all of the currently available LRS and HRCCS data sets of HD 189733 b and HD 209458 b as well as those of other planets.
2. Implementing in the algorithm other telluric removal algorithms such as PCA and Sysrem, which seem to perform excellently at wavelengths affected by strong telluric bands. Currently, the main restriction in implementing these algorithm is purely computational, as the framework forces us to repeat on each tested model spectrum the same exact analysis as on the real data, which comes at a significant computational cost for PCA/Sysrem algorithms.
3. Exploiting the strong predictive power of the framework, in particular for simulating future joint *JWST* and ELT/GMT/TMT observations of temperate terrestrial worlds. These simulations will identify the optimal wavelength range, spectral resolution, and exposure times to maximize the science return from top-class space and ground-based observatories.

We thank R. de Kok, I. Snellen, and J. Birkby for the initial discussion on combining high- and low-resolution spectroscopy. We are indebted to J. Bean and J.-M. Désert for in-depth discussions leading to the development of this framework. We also

thank Ryan MacDonald for permission on adapting his lovely retrieval flow figure for this work. We would like to thank L. Pino, P. Mollière, J. Fortney, and E. Gharib-Nezhad for their unrelenting enthusiasm for us to get this out the door. We would also like to thank Richard Freedman, Roxana Lupu, Mark Marley, and Jim Lyons for invaluable discussions regarding the intricacies of line lists as well as the incalculable people-hours spent by groups such as HITRAN, ExoMol, and dedicated laboratory scientists on producing top-tier line lists. We also thank Geoff Blake for taking the time to clarify their team's KECK NIRSpec analysis methods. Finally, M.R.L. acknowledges support from the NASA Exoplanet Research Program award NNX17AB56G and Arizona State University Start Up funds.

Facilities: VLT (CRIRES), *HST* (WFC3), *Spitzer* (IRAC).

Software: python2.7, matplotlib, corner.py, pymultinest/multinest.

ORCID iDs

Matteo Brogi  <https://orcid.org/0000-0002-7704-0153>

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

References

- Agol, E., Cowan, N. B., Knutson, H. A., et al. 2010, *ApJ*, 721, 1861
- Arcangeli, J., Désert, J.-M., Line, M. R., et al. 2018, *ApJL*, 855, L30
- Bailey, J. 2014, *PASA*, 31, e043
- Barber, R. J., Tennyson, J., Harris, G. J., & Tolchenov, R. N. 2006, *MNRAS*, 368, 1087
- Birkby, J. L., de Kok, R. J., Brogi, M., et al. 2013, *MNRAS*, 436, L35
- Birkby, J. L., de Kok, R. J., Brogi, M., Schwarz, H., & Snellen, I. A. G. 2017, *AJ*, 153, 138
- Bowler, B. P., Kraus, A. L., Bryan, M. L., et al. 2017, *AJ*, 154, 165
- 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., Giacobbe, P., Guilluy, G., et al. 2018, *A&A*, 615, A16
- Brogi, M., Line, M., Bean, J., Désert, J.-M., & Schwarz, H. 2017, *ApJL*, 839, L2
- Brogi, M., Snellen, I. A. G., de Kok, R. J., et al. 2012, *Natur*, 486, 502
- Brogi, M., Snellen, I. A. G., de Kok, R. J., et al. 2013, *ApJ*, 767, 27
- Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, *A&A*, 564, A125
- Burrows, A. S. 2014, *Natur*, 513, 345
- Cabot, S. H. C., Madhusudhan, N., Hawker, G. A., & Gandhi, S. 2019, *MNRAS*, 482, 4422
- Chiavassa, A., Casagrande, L., Collet, R., et al. 2018, *A&A*, 611, A11
- Crossfield, I. J. M. 2015, *PASP*, 127, 941
- de Kok, R. J., Brogi, M., Snellen, I. A. G., et al. 2013, *A&A*, 554, A82
- Deming, D., Wilkins, A., McCullough, P., et al. 2013, *ApJ*, 774, 95
- Evans, T. M., Sing, D. K., Wakeford, H. R., et al. 2016, *ApJL*, 822, L4
- Feng, Y. K., Line, M. R., Fortney, J. J., et al. 2016, *ApJ*, 829, 52
- Feroz, F., Hobson, M. P., & Bridges, M. 2009, *MNRAS*, 398, 1601
- Flowers, E., Brogi, M., Rauscher, E., Kempton, E., & Chiavassa, A. M.-R. 2018, arXiv:1810.06099
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Fortney, J. J., Robinson, T. D., Domagal-Goldman, S., et al. 2016, arXiv:1602.06305
- 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
- Gibson, N. P., Pont, F., & Aigrain, S. 2011, *MNRAS*, 411, 2199
- Gordon, S., & McBride, B. J. 1994, Computer Program for Calculation of Complex Chemical Equilibrium Compositions and Applications Tech. Rep. 1311, NASA
- Greene, T. P., Line, M. R., Montero, C., et al. 2016, *ApJ*, 817, 17
- Gregory, P. 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica® Support (Cambridge: Cambridge Univ. Press)
- Grillmair, C. J., Burrows, A., Charbonneau, D., et al. 2008, *Natur*, 456, 767
- Guillot, T. 2010, *A&A*, 520, A27
- Hansen, C. J., Schwartz, J. C., & Cowan, N. B. 2014, *MNRAS*, 444, 3632
- Hawker, G. A., Madhusudhan, N., Cabot, S. H. C., & Gandhi, S. 2018, *ApJL*, 863, L11
- Haynes, K., Mandell, A. M., Madhusudhan, N., Deming, D., & Knutson, H. 2015, *ApJ*, 806, 146
- Jones, A., Noll, S., Kausch, W., Szyszka, C., & Kimeswenger, S. 2013, *A&A*, 560, A91
- Knutson, H. A., Benneke, B., Deming, D., & Homeier, D. 2014, *Natur*, 505, 66
- Knutson, H. A., Charbonneau, D., Noyes, R. W., Brown, T. M., & Gilliland, R. L. 2007, *ApJ*, 655, 564
- Kochanov, R. V., Gordon, I., Rothman, L., et al. 2016, *JQSRT*, 177, 15
- Kreidberg, L., Bean, J. L., Désert, J.-M., et al. 2014a, *Natur*, 505, 69
- Kreidberg, L., Bean, J. L., Désert, J.-M., et al. 2014b, *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
- Lee, J.-M., Fletcher, L. N., & Irwin, P. G. J. 2012, *MNRAS*, 420, 170
- Line, M. R., Knutson, H., Wolf, A. S., & Yung, Y. L. 2014, *ApJ*, 783, 70
- Line, M. R., Marley, M. S., Liu, M. C., et al. 2017, *ApJ*, 848, 83
- 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., Teske, J., Burningham, B., Fortney, J. J., & Marley, M. S. 2015, *ApJ*, 807, 183
- Line, M. R., Wolf, A. S., Zhang, X., et al. 2013, *ApJ*, 775, 137
- Line, M. R., & Yung, Y. L. 2013, *ApJ*, 779, 3
- Lockwood, A. C., Johnson, J. A., Bender, C. F., et al. 2014, *ApJL*, 783, L29
- Lodders, K., Palme, H., & Gail, H.-P. 2009, *LanB*, 4B, 712
- Lupu, R. E., Zahnle, K., Marley, M. S., et al. 2014, *ApJ*, 784, 27
- MacDonald, R. J., & Madhusudhan, N. 2017, *MNRAS*, 469, 1979
- Madhusudhan, N. 2018, in Handbook of Exoplanets, ed. H. Deeg & J. A. Belmonte (Berlin: Springer), 104
- Madhusudhan, N., Crouzet, N., McCullough, P. R., Deming, D., & Hedges, C. 2014, *ApJL*, 791, L9
- Mandell, A. M., Haynes, K., Sinukoff, E., et al. 2013, *ApJ*, 779, 128
- Magic, Z., Collet, R., Asplund, M., et al. 2013, *A&A*, 557, A26
- Mansfield, M., Bean, J. L., Line, M. R., et al. 2018, *AJ*, 156, 10
- McCullough, P. R., Crouzet, N., Deming, D., & Madhusudhan, N. 2014, *ApJ*, 791, 55
- Noll, S., Kausch, W., Barden, M., et al. 2012, *A&A*, 543, A92
- Nugroho, S. K., Kawahara, H., Masuda, K., et al. 2017, *AJ*, 154, 221
- Partridge, H., & Schwenke, D. W. 1997, *JChPh*, 106, 4618
- 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
- Piskorz, D., Buzard, C., Line, M. R., et al. 2018, *AJ*, 156, 133
- Rothman, L. S., Gordon, I. E., Barber, R. J., et al. 2010, *JQSRT*, 111, 2139
- Rucinski, S. 1999, in ASP Conf. Ser. 185, Precise Stellar Radial Velocities, ed. J. B. Hearnshaw & C. D. Scarfe (San Francisco, CA: ASP), 82
- Schwarz, H., Brogi, M., de Kok, R., Birkby, J., & Snellen, I. 2015, *A&A*, 576, A111
- Shabram, M., Fortney, J. J., Greene, T. P., & Freedman, R. S. 2011, *ApJ*, 727, 65
- Sing, D. K., Fortney, J. J., Nikolov, N., et al. 2016, *Natur*, 529, 59
- Sing, D. K., Pont, F., Aigrain, S., et al. 2011, *MNRAS*, 416, 1443
- Sivia, D., & Skilling, J. 2006, Data Analysis: A Bayesian Tutorial (Oxford: Oxford Univ. Press)
- 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
- Stevenson, K. B., Désert, J.-M., Line, M. R., et al. 2014, *Sci*, 346, 838
- Torres, G., Winn, J. N., & Holman, M. J. 2008, *ApJ*, 677, 1324
- Triaud, A. H. M. J., Queloz, D., Bouchy, F., et al. 2009, *A&A*, 506, 377
- Tsias, A., Waldmann, I. P., Zingales, T., et al. 2018, *AJ*, 155, 156
- Wakeford, H. R., Sing, D. K., Kataria, T., et al. 2017, *Sci*, 356, 628
- Wilks, S. S. 1938, *The Annals of Mathematical Statistics*, 9, 60
- Zucker, S. 2003, *MNRAS*, 342, 1291