

CROCODILE

Incorporating medium-resolution spectroscopy of close-in directly imaged exoplanets into atmospheric retrievals via cross-correlation

J. Hayoz¹ , G. Cugno^{1,2} , S. P. Quanz¹ , P. Patapis¹ , E. Alei^{1,3} , M. J. Bonse¹ , F. A. Dannert^{1,3} , E. O. Garvin¹,
T. D. Gebhard^{1,4}, B. S. Konrad^{1,3} , and L. F. Sartori¹

¹ ETH Zurich, Institute for Particle Physics and Astrophysics, Wolfgang-Pauli-Strasse 27, 8093 Zurich, Switzerland
e-mail: jhayoz@ethz.ch

² Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA

³ National Center of Competence in Research PlanetS, Switzerland

⁴ Max Planck Institute for Intelligent Systems, Max-Planck-Ring 4, 72076 Tübingen, Germany

Received 21 December 2022 / Accepted 28 August 2023

ABSTRACT

Context. The investigation of the atmospheres of closely separated, directly imaged gas giant exoplanets is challenging due to the presence of stellar speckles that pollute their spectrum. To remedy this, the analysis of medium- to high-resolution spectroscopic data via cross-correlation with spectral templates (cross-correlation spectroscopy) is emerging as a leading technique.

Aims. We aim to define a robust Bayesian framework combining, for the first time, three widespread direct-imaging techniques, namely photometry, low-resolution spectroscopy, and medium-resolution cross-correlation spectroscopy in order to derive the atmospheric properties of close-in directly imaged exoplanets. Current atmospheric characterisation frameworks are indeed either not compatible with all three observing techniques or they lack the commitment to efficient sampling strategies that allow high-dimensional forward models.

Methods. Our framework CROCODILE (cross-correlation retrievals of directly imaged self-luminous exoplanets) naturally combines the three techniques by adopting adequate likelihood functions. To validate our routine, we simulated observations of gas giants similar to the well-studied β Pictoris b planet and we explored the parameter space of their atmospheres to search for potential biases.

Results. We obtain more accurate measurements of atmospheric properties when combining photometry, low- and medium-resolution spectroscopy into atmospheric retrievals than when using the techniques separately as is usually done in the literature. Indeed, the combined fit is, on average, 20% more accurate than fitting only medium-resolution cross-correlation spectroscopy. We find that medium-resolution ($R \approx 4000$) K -band cross-correlation spectroscopy alone is not suitable to constrain the atmospheric properties of our synthetic datasets; however, this problem disappears when simultaneously fitting photometry throughout the Y and M bands and low-resolution ($R \approx 60$) spectroscopy between the Y and H bands. Our thorough testing demonstrates that free chemistry is a suitable forward model to retrieve the atmospheric thermal and chemical properties of cloudless gas giants at chemical equilibrium.

Conclusions. CROCODILE provides a robust statistical framework to interpret medium-resolution spectroscopic data of close-in directly imaged exoplanets, where speckles originating from stellar stray light render the extraction of the continuum difficult. Our framework allows the atmospheric characterisation of directly imaged exoplanets using the high-quality spectral data that will be provided by the new generation of instruments such as the Enhanced Resolution Imager and Spectrograph (ERIS) at the Very Large Telescope, the Mid-Infrared Instrument (MIRI) aboard the *James Webb* Space Telescope, and in the future the Mid-infrared ELT Imager and Spectrograph (METIS) at the Extremely Large Telescope.

Key words. planets and satellites: atmospheres – methods: data analysis – techniques: imaging spectroscopy – techniques: high angular resolution

1. Introduction

Determining the ratios of elemental abundances in the atmosphere of exoplanets is emerging as a key analysis to unwrap the history of their formation. Nascent planets do indeed inherit some of the chemical properties of the protoplanetary disk at the location of their formation: Öberg et al. (2011) showed that the molecular snowlines are expected to directly affect the atmospheric carbon-to-oxygen (C/O) number ratio. Further processes that modify the chemical composition of the disk can add nuance to this initial picture; for example, the inward drift of pebbles covered with CO ice might enhance the abundance of CO gas

within the CO iceline (Öberg & Bergin 2016), or dust transport processes such as grain growth, settling, and vertical mixing can affect the C/O ratio depending on the location of snowlines and pressure bumps in the disk (van der Marel et al. 2021). Furthermore, the chemical composition of a planet can drastically change after formation due to the later accretion of surrounding material while it migrates. For instance, the accretion of ice-rich planetesimals can lead to the enrichment of metals in the atmosphere of hot Jupiters (Mordasini et al. 2016), or affect the initial atmospheric C/O ratio inherited from core accretion or gravitational collapse and its initial location with respect to the water and CO₂ icelines (Nowak et al. 2020). Currently, the details of

gas giant formation are still mostly unconstrained. Many interconnected physical processes are believed to influence their final bulk and atmospheric composition. This creates model degeneracies that prevent us from unequivocally inverting their full formation history (Mollière et al. 2022). Nevertheless, on a case-by-case basis, measurements of atmospheric composition might be leveraged to identify specific aspects of formation processes; for example, a high N/O ratio may indicate a large orbital distance during formation (Piso et al. 2016; Ohno & Fortney 2022). Meanwhile population analyses might identify statistical trends to inform formation models (Hoch et al. 2022).

It is particularly important to develop robust statistical frameworks in light of the new generation of telescopes and instruments that have recently started operating or are coming up this decade: the *James Webb* Space Telescope (JWST) with its Mid-Infrared Instrument (MIRI; Wells et al. 2015), capable of integral field spectroscopy at a spectral resolution of $R \sim 1500\text{--}3500$ between 4.9 and 28.3 μm ; the new Enhanced Resolution Imager and Spectrograph (ERIS; Davies et al. 2018) at the Very Large Telescope (VLT), which contains the newly refurbished Integral Field Unit (IFU) SPIFFIER (George et al. 2016), upgraded from the SPectrometer for Infrared Faint Field Imaging (SPIFFI) within the old Spectrograph for INtegral Field Observations in the Near Infrared (SINFONI), with a spectral resolution of $R \approx 5000\text{--}10\,000$ in the J , H , and K bands; and finally the Mid-infrared ELT Imager and Spectrograph (METIS; Brandl et al. 2021) on the upcoming Extremely Large Telescope (ELT), which allows long-slit spectroscopy at low and medium resolution ($R \approx 400\text{--}1900$), as well as integral field spectroscopy at high resolution ($R \approx 100\,000$) in the L and M bands.

The first measurement of the atmospheric C/O ratio of an exoplanet was performed by Madhusudhan et al. (2011), who applied an atmospheric fitting code to *Spitzer* and ground-based photometric data of the transiting hot Jupiter WASP-12b. Since then, many atmospheric retrieval codes have been developed for photometric and spectroscopic data of both transiting and directly imaged exoplanets; for example, readers can refer to CHIMERA (Line et al. 2013), NEMESIS (Lee et al. 2012), Tau-REx (Waldmann et al. 2015), HELIOS-RETRIEVAL (Lavie et al. 2017), and petitRADTRANS (Mollière et al. 2019). On the direct-imaging side, the first proposed strategy to measure the atmospheric C/O ratio has been the combination of photometric and low-resolution ($R < 1000$) spectroscopic data: for example, Lavie et al. (2017) analysed HR 8799 data from Bonnefoy et al. (2016) and Zurlo et al. (2016) with the VLT Spectro-Polarimetric High-contrast Exoplanet REsearch instrument (SPHERE; Beuzit et al. 2019) and the Gemini Planet Imager (GPI; Macintosh et al. 2014); Mollière et al. (2020) studied HR 8799 e more closely with additional VLTI/GRAVITY spectro-interferometric data; and Nowak et al. (2020) observed β Pictoris b with VLTI/GRAVITY to measure its atmospheric C/O ratio. More recently, Ruffio et al. (2021) observed the HR 8799 system with the OH-Suppressing Infrared Integral Field Spectrograph (OSIRIS; Quirrenbach et al. 2003; Larkin et al. 2006) at the Keck Observatory in the K band at medium spectral resolution ($R \approx 4000$) and measured stellar C/O ratios for the b, c, and d planets, contradicting the previous findings of Lavie et al. (2017) of the supersolar C/O ratio for the b planet, hinting at a possible discrepancy between measurements at low versus medium spectral resolution. Finally, Xuan et al. (2022) simultaneously analysed photometric and low- and high-resolution spectroscopic data of the brown dwarf HD 4747 B to derive its C/O ratio, while Wang et al. (2023) additionally considered

medium-resolution spectroscopy while applying the same framework based on Wang et al. (2020) to constrain the atmospheric parameters of HR 8799 c.

Several studies have explored how to leverage medium-resolution spectroscopy to investigate the atmospheric composition of directly imaged gas giants, in particular when using cross-correlation with spectral templates – a method informally called cross-correlation spectroscopy (CCS). Relying on it, Hoeijmakers et al. (2018) introduced the technique of molecular mapping, which consists in applying a spectral cross-correlation to the spectrum in each spatial pixel (spaxel) of the datacube of an integral field unit (IFU), thereby unveiling the imprint of each molecule in the atmosphere of the observed exoplanet. Instead of filtering out the starlight as part of data post-processing, Ruffio et al. (2019) – later improved by Ruffio et al. (2021) – defined a joint likelihood function of the forward model for the planetary atmosphere and stellar point spread function (PSF), allowing to constrain the barycentric radial velocity of the exoplanet as well as the atmospheric parameters using a grid of BT-Settl models (Allard 2013) after marginalising over the stellar PSF model.

The molecular mapping technique resulted in the detection of water and carbon monoxide in the atmospheres of β Pic b (Hoeijmakers et al. 2018) and HIP 65426 b Petrus et al. (2021) using K -band medium-resolution ($R \approx 5000$) spectroscopic data obtained with VLT/SINFONI (Eisenhauer et al. 2003), as well as in the atmosphere of HR 8799 b (Petit dit de la Roche et al. 2018) using H - and K -band medium-resolution ($R \approx 4000$) spectroscopic data obtained with Keck/OSIRIS. Cugno et al. (2021) applied the technique to K -band spectroscopic data of the forming planet PDS 70 b taken with VLT/SINFONI, and quantified the extinction of the surrounding dusty environment required to explain the non-detection of H_2O and CO. Patapis et al. (2022) examined the feasibility and limiting factors of applying molecular mapping with the Mid-Infrared Instrument (MIRI; Wells et al. 2015) on the *James Webb* Space Telescope (JWST) using mock observations of the systems HR 8799 and GJ 504, and predicted the detectability of H_2O , CO, CH_4 , and NH_3 in the atmospheres of directly imaged gas giants. Using the joint modelling of the stellar PSF and exoplanet atmosphere, Ruffio et al. (2019) inferred the barycentric radial velocity (RV) of the planets HR 8799 b and c using H - and K -band medium-resolution ($R \approx 4000$) spectroscopic data from Keck/OSIRIS, while Ruffio et al. (2021) constrained the C/O ratios of the HR 8799 b, c, and d planets.

Arguably, the frameworks described in these studies are missing at least one of two key ingredients, namely 1. the simultaneous consideration of archival photometric and low-resolution spectroscopic data on top of the newly published medium-resolution spectroscopic data within the Bayesian framework, or 2. the commitment to a full-on atmospheric retrieval by exploring the whole parameter space using dedicated sampling algorithms to not only find the maximum likelihood estimator, but also derive the full posterior distribution.

With this study, we propose a new framework to combine low resolution spectrophotometric and cross-correlation spectroscopic data into a full Bayesian statistical framework to derive atmospheric properties of directly imaged exoplanets and to prepare for newly commissioned and future instruments. To that end, we built upon the work of Brogi & Line (2019), who defined a Bayesian statistical framework combining low- and high-resolution spectroscopy with the original purpose of deriving atmospheric properties of transiting exoplanets. We present our framework CROCODILE (CROss-CORrelation retrievals of

Directly Imaged self-Luminous Exoplanets)¹, which embeds the radiative transfer routine *petitRADTRANS* (Mollière et al. 2019) into the mathematical framework of Brogi & Line (2019).

In Sect. 2, we present the main parts of CROCODILE, namely the Bayesian estimator, the modification of the log-likelihood function to include the spectral cross-correlation between model and data, and our forward model for the atmosphere of gas giant exoplanets. In Sect. 3, we put CROCODILE to a full battery of tests, including a comparison between a regular atmospheric retrieval running on low-resolution spectroscopy (LRS) and photometry and a pure cross-correlation-based retrieval, as well as a review of the performance of CROCODILE when applied to a large portion of the parameter space of the atmospheres of gas giants. Finally, we discuss the implications of our findings as well as future opportunities for CROCODILE in Sect. 4, before summarising our work in Sect. 5.

2. CROCODILE: A Bayesian framework to combine all direct-imaging observing techniques into atmospheric retrievals

CROCODILE is composed of a Bayesian estimator coupled to a forward model for the emission spectra of gas giants. More precisely, the routine starts with the prior distributions of the model parameters, from which the Bayesian estimator samples a set of input parameters for the simulator. The forward model computes the spectrum from the sampled parameters, before formatting it to the observing techniques that make up the data, that is photometry, low-resolution spectroscopy, or continuum-removed medium-resolution spectroscopy. These synthetic model spectra are used together with the data to calculate the sum of the three log-likelihood functions associated to each observing technique. Finally, the log-prior is added to the log-likelihood to obtain the log-posterior which is then used to update the sampled points according to a nested sampling algorithm. The input of CROCODILE is the data, the forward model, the opacity database, and the priors. Its output is the posterior distribution of the forward model, from which the posterior thermal structure, chemistry, and spectrum can be computed. A conceptual sketch of CROCODILE is shown in Fig. 1, while we describe its implementation in the following sections.

2.1. The Bayesian estimator

Atmospheric retrievals are a well-established framework within the exoplanet community (for a review, readers can refer to Madhusudhan 2018). Fundamentally, they seek to invert the data, meaning that, for a given model describing the transfer of electromagnetic radiation through the atmosphere of exoplanets, they estimate how well each combination of parameters of said model explains the data. Essentially, they contain three components, namely 1. the data (which are usually obtained via photometry or spectroscopy), 2. the forward model (which maps the parameters of the exoplanetary atmosphere to an emission or transmission spectrum), and 3. a Bayesian inference algorithm that takes in the data as well as prior knowledge and calculates a posterior probability distribution over the parameter space of the model. Bayes' theorem (Bayes 1763) provides the mathematical framework by formulating the problem as the computation of the posterior

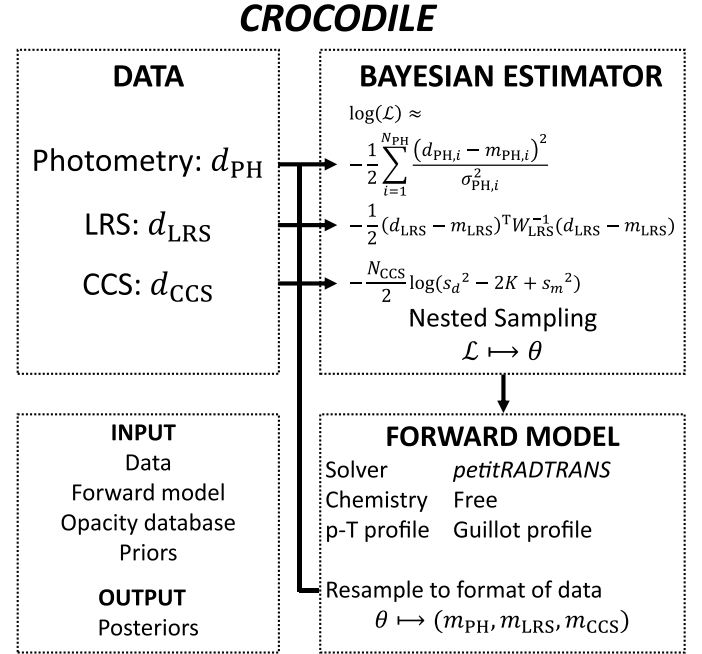


Fig. 1. Schematic of our atmospheric retrieval framework CROCODILE. The input of the routine is the data – which can be photometry, low-, or medium-resolution spectroscopy –, the forward model, the opacity database, and the priors. CROCODILE starts by sampling parameters from the prior, from which the forward model is evaluated. It is then compared to the data by computing the sum of the log-likelihood functions corresponding to the three observing techniques, from which new parameters are sampled and the loop restarts. The output is the posterior distribution of the forward model.

distribution:

$$p(\theta | d) = \frac{p(d | \theta)p(\theta)}{p(d)}, \quad (1)$$

where the vector θ denotes the parameters of the forward model (e.g. the equilibrium temperature, surface gravity, or chemical composition), d the data, $p(d | \theta) \equiv \mathcal{L}(\theta)$ the likelihood function, $p(\theta)$ the prior distribution, and $p(d) = \int_{\Theta} p(d | \theta)p(\theta)$ the evidence defined as the integral of the likelihood function multiplied by the prior over the parameter space Θ . Calculating the likelihood function is expensive since it requires to evaluate the forward model including the computationally heavy radiative transfer routine. Therefore, we need an efficient parameter sampling method to reduce the number of evaluations of the forward model. Our code CROCODILE makes use of the software *pymultinest* (Buchner et al. 2014), which is broadly used in the community of atmospheric retrievals (e.g. Lavie et al. 2017; Mollière et al. 2019, 2020; Konrad et al. 2022), a Python implementation of *MultiNest* (Feroz et al. 2009, 2019) which utilises the multimodal nested sampling algorithm from Skilling (2006).

The choice of the likelihood function $\mathcal{L}(\theta)$ depends on the noise statistics of the data, which depend on the observing and post-processing techniques used. Typically, photometric data d_{PH} obtained by different instruments or in different observing conditions are statistically independent of each other, allowing the decomposition of the likelihood function into a product of likelihood functions for each datapoint. Furthermore, the noise in high-contrast imaging data is usually assumed to be independent identically distributed Gaussian, leading to the following log-likelihood function (ignoring constant factors that only depend

¹ CROCODILE is available as a Python package at <https://github.com/JHayoz/CROCODILE>

on the data):

$$\log \mathcal{L}_{\text{PH}}(\theta) \approx -\frac{1}{2} \sum_{i=1}^{N_{\text{PH}}} \frac{(d_{\text{PH},i} - m_{\text{PH},i}(\theta))^2}{\sigma_{\text{PH},i}^2}, \quad (2)$$

where $m_{\text{PH}}(\theta)$ is the synthetic photometric data obtained by evaluating the forward model at the parameter θ , and σ_{PH} is the uncertainty associated to the data d_{PH} .

For low-resolution spectroscopic data d_{LRS} , the noise statistics can be more intricate, since different wavelength channels of a spectrum might be correlated (e.g. [Todorov et al. 2016](#); [Madhusudhan 2018](#); [Mollière et al. 2019](#); [Nowak et al. 2020](#)). In this case, the likelihood function is often chosen as a multivariate normal distribution (again ignoring constant factors):

$$\log \mathcal{L}_{\text{LRS}}(\theta) \approx -\frac{1}{2} (d_{\text{LRS}} - m_{\text{LRS}}(\theta))^T W_{\text{LRS}}^{-1} (d_{\text{LRS}} - m_{\text{LRS}}(\theta)), \quad (3)$$

where $m_{\text{LRS}}(\theta)$ is the synthetic low-resolution spectrum that is obtained by evaluating the forward model at the parameter θ ([Greco & Brandt 2016](#)). The covariance matrix W_{LRS} describes the correlation of the data across spectral channels and needs to be computed for each observation (e.g. [Nowak et al. 2020](#)).

2.2. Incorporating cross-correlation spectroscopy into atmospheric retrievals

Cross-correlation spectroscopy (CCS; [Sparks & Ford 2002](#)) was introduced to the field of exoplanet direct-imaging to leverage the full spectral resolution of spectrographs. With the exception of M stars, the high temperatures found in the photosphere of stars prevent the atoms from bonding into molecules. For that reason, stellar spectra only show atomic absorption lines in the infrared range, which differ significantly from the molecular absorption lines found in colder objects such as brown dwarfs and gas giant exoplanets. Therefore, it is possible to filter out the stellar spectrum and cross-correlate the residuals with molecular spectral templates, thereby picking up the signal of molecules present in the atmosphere of the low-mass companion, for example using high-resolution spectral differential imaging (HRSDI; e.g. [Hoeijmakers et al. 2018](#); [Haffert et al. 2019](#); [Cugno et al. 2021](#)).

To incorporate HRSDI-treated IFU data into atmospheric retrievals, we followed the framework of [Brogi & Line \(2019\)](#), who expanded the conventional Gaussian likelihood function from Eq. (2) by assuming constant noise across all spectroscopic channels and replacing the standard deviation by its maximum likelihood estimator, thereby making the cross-covariance K between model and data appear in the log-likelihood function:

$$\log \mathcal{L}_{\text{CCS}}(\theta) \approx -\frac{N_{\text{CCS}}}{2} \log(s_d^2 - 2K(\theta) + s_m^2(\theta)), \quad (4)$$

with the following definitions

$$s_d^2 = \frac{1}{N_{\text{CCS}}} \sum_{i=1}^{N_{\text{CCS}}} d_{\text{CCS},i}^2 \quad (5)$$

$$s_m^2(\theta) = \frac{1}{N_{\text{CCS}}} \sum_{i=1}^{N_{\text{CCS}}} m_{\text{CCS},i}^2(\theta) \quad (6)$$

$$K(\theta) = \frac{1}{N_{\text{CCS}}} \sum_{i=1}^{N_{\text{CCS}}} d_{\text{CCS},i} m_{\text{CCS},i}(\theta). \quad (7)$$

In Eq. (7), the cross-covariance K is linked to the cross-correlation C over the relation:

$$K \equiv C \sqrt{s_d^2 s_m^2}, \quad (8)$$

meaning that either the cross-correlation C or the cross-covariance K can be computed as long as the right variable is used in Eq. (4). For the sake of completeness, we provide the derivation of Eq. (4) used by [Brogi & Line \(2019\)](#) in Appendix A. So far, we have kept the explicit dependency of the model M on the parameter θ for the sake of mathematical rigorosity; however, we have dropped the dependency of the model spectrum on the frame of reference. Whereas the effect is insignificant for photometry and low-resolution spectroscopy of close-by stellar systems, the Doppler shift created by the radial velocity v_R of the target exoplanet with respect to the observatory indeed becomes significant at medium spectral resolution. More precisely, for a spectrum at spectral resolution R and at wavelength λ , and therefore with channel spacing $\Delta\lambda = \frac{\lambda}{R}$, the Doppler shift $\frac{\Delta\lambda}{\lambda} = \frac{v_R}{c}$ (with c the speed of light in vacuum) resulting from the radial velocity v_R is equal to the wavelength spacing of the spectrum when $v_R = \frac{c}{R}$. At medium spectral resolution ($R \approx 5000$), this effect occurs with a radial velocity of $v_R \approx 60 \text{ km s}^{-1}$, whereas it requires a much higher radial velocity ($v_R > 600 \text{ km s}^{-1}$) at low spectral resolution ($R < 500$). Hence, the forward model for medium- and high-resolution spectroscopy needs to be Doppler-shifted to the reference frame of the observed spectrum. Instead of computing it once and for all before the retrieval or adding it in as a parameter of the model, we incorporated it into the algorithm by taking the maximum of the cross-covariance K over a range of radial velocity

$$K(\theta) = \max_{v_0 \leq v_R \leq v_1} K(\theta, v_R), \quad (9)$$

where $K(\theta, v_R)$ is the cross-covariance where the model is Doppler shifted by the radial velocity v_R , and v_0 and v_1 are the lower and higher bounds of the range of radial velocities considered. To Doppler shift the forward model spectrum, we used the methods `dopplerShift` and `crosscorrRV` from the Python package `PyAstronomy` ([Czesla et al. 2019](#)), and we selected radial velocities between $v_0 = -400$ and $v_1 = 400 \text{ km s}^{-1}$ in steps of 0.5 km s^{-1} , which corresponds to 0.8% of the wavelength spacing at medium spectral resolution $R = 5000$. The size of the steps have conservatively been chosen small because Doppler-shifting the spectra is a cheap computational operation relative to the cost of computing the forward model.

These algorithms use interpolation to Doppler shift spectra, which can lead to non-conservation of the total flux. To limit this effect, we applied the Doppler shift before down-sampling the spectra to a lower spectral resolution when simulating data. On the other hand, the routine `crosscorrRV` interpolates the already down-sampled forward model to the Doppler shifted wavelength before computing the correlation. To understand how the use of interpolation might negatively affect the cross-correlation, we implemented and tested an algorithm which first applies the Doppler shift before down-sampling and computing the correlation. To quantify the improvement, we simulated three spectra using our forward model from Sect. 3.3: one data spectrum with added random scatter, and two template spectra

with different model parameters. We then computed the likelihood ratio (Eq. (4)) between the two template spectra using both algorithms. The new algorithm did not significantly change the results: the likelihood ratios differed by 0.03%. However, the new algorithm was slower by a factor of 1000 due to the extra down-sampling step, which would become a bottleneck during retrievals. Hence, we did not use our new algorithm in this work and continued evaluating Eq. (4) using the routine `crosscorrRV`. We note that the issue discussed here might yield a different result with another combination of spectral resolution and wavelength range. Therefore, we recommend caution when using an interpolation-based Doppler shifting algorithm; in particular, the test described above should be repeated when working with alternate datasets.

Finally, it is worth stating explicitly that the spectral data d_{CCS} are assumed to have been reduced using HRSDI, and thus its continuum has been removed. Therefore, the model also needs to be high-pass filtered with the same cutoff frequency as the data. We describe our high-pass filter in Sect. 3.2.

2.3. Combining data from different instruments

In the two previous sections, we motivated our choice of log-likelihood functions associated to photometric, low-resolution spectroscopic, and cross-correlation spectroscopic data. To integrate all three techniques into one framework, we need to compute their joint likelihood function:

$$\mathcal{L}(\theta) = p(d_{\text{PH}}, d_{\text{LRS}}, d_{\text{CCS}} | \theta). \quad (10)$$

In general, the joint likelihood functions of two datasets d_1 and d_2 depends on their conditional probability $p(d_1, d_2 | \theta) = p(d_1 | d_2, \theta)p(d_2 | \theta)$. In this work, we assumed that all datasets are independent of each other, that is $p(d_1 | d_2, \theta) = p(d_1 | \theta)$. This assumption should hold whenever two datasets come from different telescopes or different instruments, or if they were taken on different nights or under different meteorological conditions. Therefore, we can apply this argument recursively to the joint likelihood function given above to obtain

$$p(d_{\text{PH}}, d_{\text{LRS}}, d_{\text{CCS}} | \theta) = p(d_{\text{PH}} | \theta)p(d_{\text{LRS}} | \theta)p(d_{\text{CCS}} | \theta), \quad (11)$$

which translates to the following log-likelihood function for CROCODILE:

$$\begin{aligned} \log \mathcal{L}(\theta) &= \log \mathcal{L}_{\text{PH}}(\theta) + \log \mathcal{L}_{\text{LRS}}(\theta) + \log \mathcal{L}_{\text{CCS}}(\theta) \\ &\approx -\frac{1}{2} \sum_{i=1}^{N_{\text{PH}}} \frac{(d_{\text{PH},i} - m_{\text{PH},i}(\theta))^2}{\sigma_{\text{PH},i}^2} \\ &\quad -\frac{1}{2} (d_{\text{LRS}} - m_{\text{LRS}}(\theta))^T W_{\text{LRS}}^{-1} (d_{\text{LRS}} - m_{\text{LRS}}(\theta)) \\ &\quad -\frac{N_{\text{CCS}}}{2} \log(s_d^2 - 2K(\theta) + s_m^2(\theta)). \end{aligned} \quad (12)$$

2.4. The forward model

The forward model of an atmospheric retrieval is composed of two parts, namely the theoretical model describing the chemical and thermal properties of the planetary atmosphere, and a radiative transfer routine that calculates the spectrum escaping from it. For the latter, we utilised `petitRADTRANS` (Mollière et al. 2019), which solves the radiative transfer equation on a discrete 1D plane-parallel atmosphere. It can be used in two spectral modes, namely the high-resolution line-by-line

mode (lbl; $\lambda/\Delta\lambda = 10^6$) or the medium-resolution correlated-k (c-k) approximation ($\lambda/\Delta\lambda = 10^3$). We used the lbl mode when calculating spectra of instruments with a spectral resolution higher than 1000, and the other for lower resolution or photometry. We set the number of atmospheric layers on which the radiative transfer equation is solved to 100 between a top and bottom pressure of 10^{-6} and 10^2 bar.

For the physical model of the atmosphere, two main approaches are used in the literature, namely self-consistent (Seager et al. 2005; Burrows et al. 2008; Fortney et al. 2008; Moses et al. 2011; Gandhi & Madhusudhan 2017) and free models (Todorov et al. 2016; Lavie et al. 2017; Nowak et al. 2020; Mollière et al. 2020). Self-consistent models aim to account for the known relevant physical processes that might take place in a planetary atmosphere and therefore implement physical constraints such as hydrodynamical, thermodynamical, and chemical equilibria, and sometimes even disequilibria due to dynamical interactions such as, for example, winds (Gandhi & Madhusudhan 2018) or photochemistry (Moses et al. 2011). Evidently, the risk of this approach is to falsely account for the relevant physical processes, for example by wrongly assuming chemical equilibrium (Madhusudhan 2018). The solution to that problem is to leave out physical consistency and adopt a parameterised model that is flexible enough to simultaneously account for both well-known and unknown physical processes. The disadvantage of that approach is that it often requires a higher-dimensional model and that it might return physically impossible solutions. In this work, we selected a mixed approach with a physically motivated analytical thermal structure combined with free chemistry. Indeed, we are mainly interested in constraining the abundances of specific molecules, therefore requiring to fit for each of them independently, whereas investigating the diversity of thermal structures that can be constrained by atmospheric retrievals is beyond the scope of this work.

Specifically, our free chemistry model was parameterised by the mass fractions X_i of the molecular abundances which are assumed to be constant vertically across all atmospheric layers (e.g. Todorov et al. 2016; Line et al. 2017). The rest of the atmospheric mass was made of 75% molecular hydrogen and 25% helium (Line et al. 2012). We considered the line opacities of H_2O , CO , CO_2 , CH_4 , H_2S , FeH , TiO , K , and VO , at low (c-k) and high resolution (lbl) depending on the spectral resolution of the instrument for which the spectrum was calculated. We included Rayleigh scattering of molecular hydrogen and helium together with the Collision-Induced Absorption (CIA) cross-sections of H_2 - H_2 and H_2 - He . The line lists and opacities used in this study were downloaded at the online documentation of `petitRADTRANS`² and are listed in Table 1 for reference.

We parameterised the thermal structure of the atmosphere with the analytic semi-grey model derived by Guillot (2010). It is parameterised by six quantities, of which the equilibrium temperature T_{equ} and surface gravity g are free parameters of our forward model, whereas the other four are fixed: the average opacity in the infrared $\kappa_{\text{IR}} = 0.01 \text{ cm}^2/\text{g}$, the ratio of the average opacity in the optical and infrared $\gamma = \kappa_{\text{V}}/\kappa_{\text{IR}} = 0.4$, the interior temperature $T_{\text{int}} = 200 \text{ K}$, and the bottom pressure $P_0 = 100 \text{ bar}$.

The last parameter was the planet radius R which, together with the distance from the Earth d , served to scale the atmospheric spectrum down by a multiplicative factor of R^2/d^2 , since

² The documentation of `petitRADTRANS` is available at https://petitRADTRANS.readthedocs.io/en/latest/content/available_opacities.html

Table 1. Line lists and opacities included.

Opacity	c-k	lbl
H2	...	HITRAN
H ₂ O	HITEMP	HITEMP
CO	HITEMP	HITEMP
CH ₄	ExoMolOP	YT14
CO ₂	ExoMolOP	HITEMP
H ₂ S	ExoMolOP	HITRAN
FeH	ExoMolOP	YT14
TiO	ExoMolOP	ExoMolOP
K	VALD	VALD
VO	ExoMolOP	ExoMolOP
Rayleigh		
H2	DW62	DW62
He	CD65	CD65
Quasi-continuum		
H2-H2 CIA	BR, RG12	BR, RG12
H2-He CIA	BR, RG12	BR, RG12

Notes. The acronyms given above stand for the following references: ExoMolOP: Chubb et al. (2021), HITEMP: Rothman et al. (2010), HITRAN: Rothman et al. (2013), YT14 = Yurchenko & Tennyson (2014), VALD: Piskunov et al. (1995), DW62: Dalgarno & Williams (1962), CD65: Chan & Dalgarno (1965), BR: Borysow et al. (1988, 1989); Borysow & Frommhold (1989); Borysow et al. (2001); Borysow (2002), RG12: Richard et al. (2012).

petitRADTRANS calculates the flux escaping the top of the atmosphere of the planet per unit surface.

In this study, we purposefully ignored clouds, which are known to be present in the atmospheres of gas giant exoplanets (Nowak et al. 2020; Mollière et al. 2020). However, considering clouds would add too much complexity to our model and would dilute our message. Indeed, the modelling of clouds is an issue that all atmospheric retrieval frameworks face and therefore need to be investigated separately.

Finally, we ignored rotational broadening and limb darkening. Indeed, Palma-Bifani et al. (2023) included both parameters into their retrieval on VLT/SINFONI *J*-, *H*-, and *K*-band spectroscopic observations of AB Pic b and reported an erratic behaviour of their posteriors, which was probably due to the insufficient spectral resolution of VLT/SINFONI. Furthermore, there was no noticeable correlation between $v \sin(i)$ and any other parameter in their published 2-D corner-plots (figures 7 and 8). Similarly, Bryan et al. (2020) measured the rotational broadening of the planetary-mass companion 2MASS J01225093–2439505 using high-resolution ($R \approx 25\,000$) spectroscopic observations taken with Keck/NIRSPEC and reported no significant dependence on the C/O ratio of the atmospheric models used to cross-correlate with the data.

3. Tests on simulated data

In the following, we aim to demonstrate the capability of CROCODILE to recover the thermal and chemical properties of the atmospheres of gas giants using a common combination of ground-based techniques in the near and mid infrared such as photometry, low-resolution spectroscopy (LRS), and medium-resolution cross-correlation spectroscopy (MRCCS).

We therefore submit it to a full series of tests that we describe and justify in the following.

As a preliminary test of CROCODILE, we checked our statistical framework in Sect. 3.3 for potential mistakes or biases. For this, we controlled our implementation of the cross-correlation-based likelihood given by Eq. (4) by running a retrieval with MRCCS data alone. Then, we tested our implementation of the regular retrieval framework on LRS and photometric (LRSP) data. Finally, we tested the implementation of CROCODILE on the combination of the three techniques.

To test CROCODILE in a meaningful way, we considered a realistic study case for ground-based instruments. This had the following two consequences, namely first that we chose a well-known target to simulate for which some atmospheric properties have been derived already, and second we assumed a realistic list of instruments to simulate spectroscopic and photometric observations with. To that end, we selected β Pictoris b (Lagrange et al. 2010) as a bright and young exoplanet that was observed by many instruments and was the object of several atmospheric studies (e.g. Hoeijmakers et al. 2018; Stolker et al. 2020; Nowak et al. 2020). We describe this target and the used instruments in detail in Sects. 3.1 and 3.2, and show the results of that test in Sect. 3.4.

Furthermore, we show that combining MRCCS with LRSP generally allows to better constrain the forward model as compared to using only LRSP data as it has been done until now, by applying our framework to different values of the parameter space of the atmospheres of gas giants (readers can refer to Sect. 3.1 for details). The results of that test are presented in Sect. 3.5.

3.1. Simulated targets

To test our method, we need to know the ground truth values of the parameters of an exoplanetary atmosphere and see if we can retrieve them truthfully. However, it would not be realistic (namely: too optimistic) to use the same chemical model to simulate such an atmosphere that we use as our forward model. In particular, the underlying chemical model of our target should have a more complex chemistry than our simple forward model. Therefore, whereas our forward model assumed free chemistry as we described in Sect. 2.4, for our simulated targets we chose the chemical equilibrium model that was computed by easyCHEM from Mollière et al. (2017)³ as a first approximation of the chemistry of an exoplanetary atmosphere, not accounting for clouds (Burrows & Sharp 1999), vertical mixing (Griffith & Yelle 2000), or other processes causing disequilibrium chemistry. Concretely, it was calculated in the following way: we set the pressure-temperature profile as well as the carbon-to-oxygen (C/O) ratio and metallicity (Fe/H) and the code calculated the resulting steady-state molecular abundances at chemical equilibrium in each layer of the atmosphere by minimising the Gibbs free energy. We considered the opacities of H₂O, CO, CH₄, CO₂, H₂S, NH₃, FeH, HCN, TiO, PH₃, K, VO, and Na (i.e. HCN in addition to the opacities included in Nowak et al. 2020). We included more opacities here than in our forward model described in Sect. 2.4. Moreover, we also included Rayleigh scattering of molecular hydrogen and helium together with the Collision-Induced Absorption (CIA) cross-sections of H₂-H₂ and H₂-He. Finally, we used the same p-T profile as our forward

³ A grid of molecular abundances pre-computed by easyCHEM can be found at https://petitradtrans.readthedocs.io/en/latest/content/notebooks/poor_man.html

Table 2. Parameter values assumed for the simulated datasets.

ID	T_{equ} (K)	C/O	Fe/H
1	1742	0.49	0.61
2	1400	0.53	0.59
3	2000	0.47	0.62
4	1742	0.32	0.75
5	1742	1.04	0.38
6	1742	0.46	-1.02
7	1742	0.52	1.76

Notes. Each row corresponds to a simulated atmosphere used to compute synthetic MRCCS, LRS, and photometric measurements as described in Sect. 3.2.

model (Guillot 2010). The spectra of our synthetic targets were also generated with `petitRADTRANS` (Mollière et al. 2019).

As mentioned above, we aim to test CROCODILE on a meaningful subspace of exoplanetary atmospheres. In particular, we considered the three parameters that control the chemical equilibrium model, namely the equilibrium temperature, the carbon-to-oxygen ratio, and the metallicity, and we selected a baseline equal to the values derived for β Pic b by Nowak et al. (2020), namely $T_{\text{equ}} = 1742$ K, $\text{C/O} = 0.44$, and $\text{Fe/H} = 0.66$ dex, and subsequently varied each of these parameters. Table 2 summarises the full list of exoplanetary atmospheres simulated. We point out that the values given in the table did not exactly match the input values used for `easyCHEM` (which were 0.3, 0.44, and 1 for C/O and -1, 0.66, and 2 for Fe/H), since we only considered the gaseous content of the atmosphere and not the liquid and solid reactants available with `easyCHEM`. The values given in Table 2 are the mean C/O ratio and metallicity along the vertical extent of the simulated atmospheres after removing all opacities from non-gaseous reactants.

Additionally, our simulated targets had a planetary radius $R_p = 1.36 R_J$ (Nowak et al. 2020) and an arbitrary radial velocity of $v_R = 31 \text{ km s}^{-1}$, the former being a parameter of the forward model while the latter only plays a role at medium (or higher) spectral resolution and gets determined at every computation of the likelihood function using Eq. (9). Finally, our model of the targets implicitly included six more parameters which were kept fixed between the data and the forward model. Five of these describe the p-T profile and are the same as given in Sect. 2.4. The remaining parameter was the distance $d = 19.75 \text{ pc}$ to β Pictoris b as measured by Gaia Collaboration (2018).

3.2. Synthetic data

For each of the seven synthetic targets described above, as well as for each evaluation of our forward model, we recreated synthetic photometry, LRS, and MRCCS by matching the instrumental characteristics of the archival datasets of β Pic b listed in Table 3. Considering real observations of β Pic b enabled us to simulate realistic measurement uncertainties and random noise for each instrument in a straightforward way, thereby creating a realistic synthetic dataset. To simulate the format recorded by each of these instruments, we started by computing a spectrum with `petitRADTRANS` using the atmospheric model described in Sect. 3.1 in either the full spectral resolution line-by-line mode ($R \approx 10^6$) for VLT/SINFONI and VLT/GRAVITY or the correlated-k approximation ($R \approx 1000$) for VLT/SPHERE and photometry.

Photometry was calculated by integrating the spectrum multiplied by the normalised filter transmission function, which we downloaded with the Python package `species` (Stolker et al. 2020)⁴. We estimated the uncertainty of the photometric flux by taking the same relative error as the real measurement, and added normally distributed random scatter with standard deviation equal to the estimated uncertainty.

For the VLT/GRAVITY instrument, the resampling from the `petitRADTRANS` to the observation bins was done using the Python package `spectRES` (Carnall 2017) which preserves the total flux. We then estimated the uncertainty of the synthetic fluxes by scaling the covariance matrix of the real measurement to the simulated spectrum. To create realistic flux measurements, we added a random (multivariate normally distributed) scatter to each flux bin with a covariance matrix equal to the simulated covariance matrix. For VLT/SPHERE, where we had no access to a spectrum of β Pic b, we chose a simpler approach by setting the uncertainty to 10% (which is an optimistic estimation of the errorbars found in Fig. 11 of Langlois et al. 2021) and added normally distributed random scatter in the same way as for photometry.

Concerning the VLT/SINFONI MRCCS data, we first applied a Doppler shift with a radial velocity of 31 km s^{-1} to the spectrum before resampling the spectrum using `spectRES`. To match the treatment of HRSDI, we applied a low-pass filter consisting of a Gaussian filter with standard deviation of 4.9 nm to compute the continuum, which we subsequently subtracted from the spectrum. We found the value of 4.9 nm by repeating our preliminary test from Sect. 3.3 for different filter sizes (readers can refer to Appendix C for details); however, we note that the optimal filter size should be investigated separately when using a real observation. To add random scatter to our MRCCS data, we aimed to reproduce the signal-to-noise ratio (S/N) of the real observations that we considered. Therefore, we generated a continuum-removed spectrum using the values retrieved by Nowak et al. (2020) for β Pic b and cross-correlated it with the SINFONI spectrum from Hoeijmakers et al. (2018). We then took the ratio of the peak of the cross-correlation function to its standard deviation 200 km s^{-1} away from the peak (similar to Cugno et al. 2021) and obtained a value of 14.4. Thus, when we simulated SINFONI spectra, we added Gaussian noise with a standard deviation such that the CCF between noisy and clean spectra also yielded the same ratio of the peak to standard deviation. This made the synthetic SINFONI planet spectra similar (in terms of S/N) to the β Pic b data presented in Hoeijmakers et al. (2018).

3.3. Verifying the framework of CROCODILE

To verify the statistical framework used in CROCODILE, we submitted it to a preliminary test in an idealised setup. To that end, we created yet another synthetic target with the same free chemistry as in our forward model. More precisely, we used the same p-T profile as in our simulated atmosphere 1 from Table 2 but changed the chemical composition to only contain H_2O and CO with vertically constant mass fractions equal to -1.8 dex and -1.6 dex respectively. Moreover, we did not add any random scatter contained in our low-resolution datasets to ensure that, if the retrieval did not converge to the input values, it was not because of the noise but rather because of a mistake in our framework. Finally, we reduced the amount of random scatter by a factor of three in the MRCCS data with respect to the synthetic

⁴ <https://github.com/tomasstolker/species>

Table 3. List of observations of β Pic b used for our simulations.

Instrument	Band	Technique	Wavelength (μm)	R	Reference
VLT/SPHERE/IFS	$Y-H$	LRS	0.96–1.63	30	...
VLT/GRAVITY	K	LRS	1.97–2.49	500	Nowak et al. (2020)
VLT/SINFONI	K	MRCCS	2.09–2.45	4000	Hoeijmakers et al. (2018)
LCO/VisAO	Y_s	PHOT	0.99	...	Males et al. (2014)
VLT/NACO	J	PHOT	1.28	...	Currie et al. (2013)
Gemini/NICI	ED286	PHOT	1.59	...	Males et al. (2014)
VLT/NACO	H	PHOT	1.67	...	Currie et al. (2013)
VLT/SPHERE/IRDIS	K12.1	PHOT	2.11	...	Lagrange et al. (2020)
VLT/NACO	K_s	PHOT	2.15	...	Bonnefoy et al. (2011)
VLT/SPHERE/IRDIS	K12.2	PHOT	2.26	...	Lagrange et al. (2020)
VLT/NACO	NB374	PHOT	3.74	...	Stolker et al. (2020)
VLT/NACO	Lp	PHOT	3.81	...	Stolker et al. (2019)
VLT/NACO	NB405	PHOT	4.07	...	Stolker et al. (2020)
VLT/NACO	Mp	PHOT	4.79	...	Stolker et al. (2019)

Notes. The VLT/SPHERE/IFS spectrum is not public and therefore we do not put a reference next to it, however we do refer to Langlois et al. (2021) to estimate the errorbars of our synthetic spectra.

observations introduced in Sect. 3.2 instead of removing it completely because otherwise the argument of the logarithm could become zero if the forward model matches the simulated data.

We first ran CROCODILE on MRCCS data alone to test the implementation of the likelihood function based on the cross-correlation function (Eq. (4)). We then tested the implementation of the regular likelihood function based on the residuals between data and model (Eq. (3)) using only LRSP. Finally, we checked that the combination of MRCCS together with LRSP via the addition of their log-likelihood functions was also valid. We show the resulting 2D posterior distributions of these three evaluations of CROCODILE in Fig. 2.

The three tests yielded educative results: the retrieval on LRSP data alone yielded perfectly centred posteriors with small uncertainty as can be expected from the lack of random scatter in the data. In particular, the abundances of water and carbon monoxide were well constrained due to the good coverage of their spectral features by the VLT/SPHERE/IFS/ $Y-H$ and VLT/GRAVITY/ K low-resolution spectroscopic data. On the other hand, the retrieval on MRCCS data yielded a systematic overestimation of the temperature, surface gravity, radius, and H_2O abundance, while the abundance of CO was constrained very accurately. Furthermore, the radius seemed correlated with the surface gravity and the abundances of H_2O and CO, and to a lesser extent to T_{equ} .

These correlations can be tentatively explained by the removal of the continuum in the treatment of the MRCCS data in the following way. Without continuum, the radius only has the role of scaling the absorption lines, which is counteracted on the one hand by the surface gravity, which in the p-T model from Guillot (2010) brings the same temperatures at higher pressures, which reduces the thermal gradient between the top of the atmosphere and the region at optical depth of unity and hence decreases the strength of the line. The molecular abundances also counteract the scaling effect of the radius by increasing the height at which the optical depth reaches unity, thereby also reducing the thermal gradient between the top of the atmosphere and the emitting layer. Interestingly, if the radius is already known and fixed to a previously measured value during the retrieval, then the results improve greatly as can be seen in Fig. D.1. Indeed, the abundances of H_2O and CO and the surface

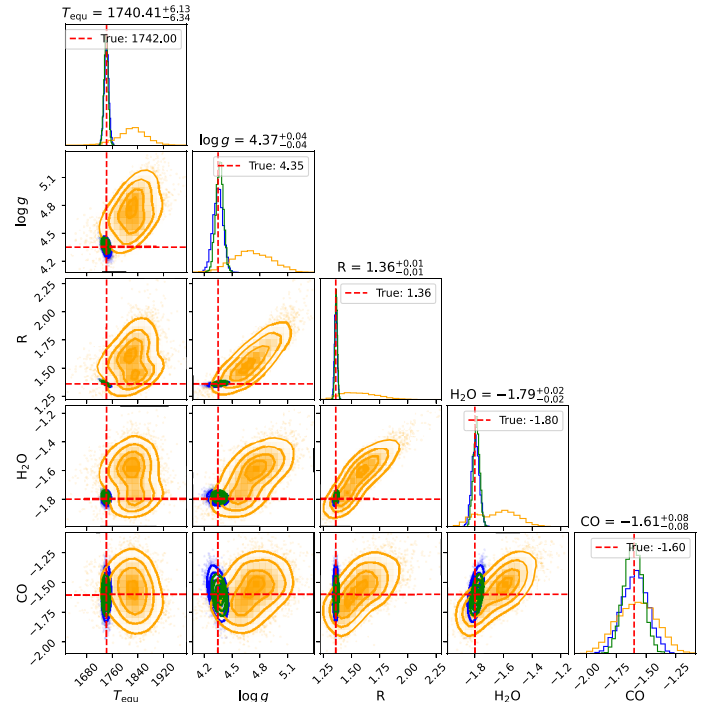


Fig. 2. Resulting posterior distributions for our preliminary test that controls our implementation of CROCODILE using the same chemistry as input as our forward model, namely vertically constant abundances of H_2O and CO, and reduced random scatter. The orange posterior shows the results of the retrieval using only MRCCS data to test the implementation of the cross-correlation likelihood function, the blue posterior is for the regular framework using only LRSP, and the green posterior shows the result for the combination of MRCCS and LRSP data. The dashed red line shows the ground truth used as input.

gravity were all retrieved within 1σ with a smaller uncertainty; however, the temperature was still equally overestimated. This test indirectly hints that if the radius – or any other flux scaling parameter – is set as fixed parameter to a wrong value, then there is a high risk to bias the results when using MRCCS data only.

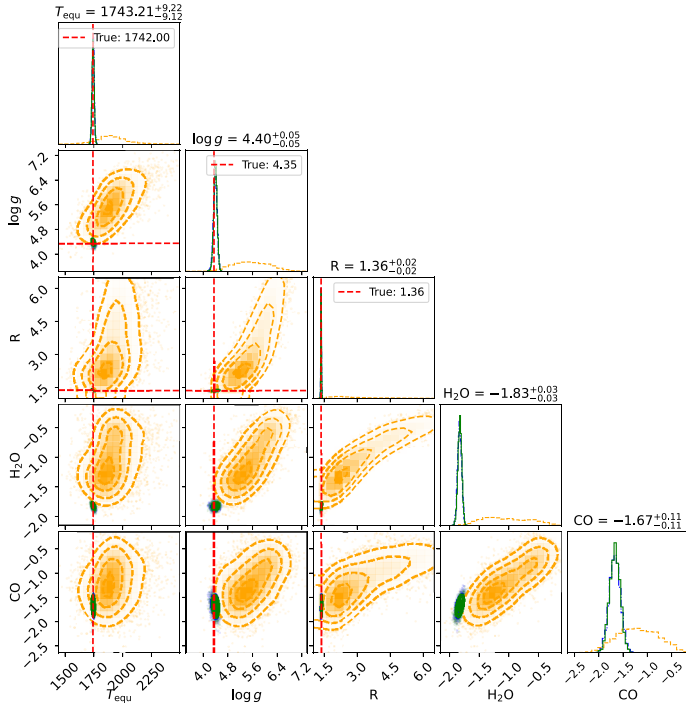


Fig. 3. Results of the three retrievals of the same simulated atmosphere using different combinations of the synthetic spectra of β Pic b: the SIN-FONI MRCCS-only retrieval in orange, the LRSP-only retrieval in blue, and the retrieval using all techniques in green. The off-axis panels show the projected 2D posterior distributions of the main model parameters (readers can refer to the description in Sect. 2.4). The indicated values above each histogram on the diagonal refer to the median and 1-sigma uncertainty for CROCODILE. For the molecular abundances, we omit the logarithm in front of the names of the molecules for the sake of readability. The dashed red line represents the true parameter value, and it is not shown for the abundances given the different chemical treatment between simulated observations and forward model.

Finally, when combining LRSP and MRCCS data, the results were slightly more accurate than when used on LRSP data alone. In particular the abundance of CO was retrieved with a 25% smaller 1σ spread. In summary, this test verified the statistical framework of CROCODILE and already showed the systematics that appear when using only MRCCS data.

3.4. Case study: A synthetic β Pictoris b

After validating our statistical framework, we investigated possible biases by applying it to the synthetic targets described in Sect. 3.1, which span across a portion of the parameter space around the gas giant β Pic b. First, we give a more in-depth report for our baseline in this subsection while we describe all our results in the next. Figure 3 compares the posterior distributions obtained with MRCCS data only, with LRSP, or with MRCCS and LRSP simultaneously; the posterior distributions of the molecular abundances are shown in Fig. 4; and the retrieved thermal profiles, C/O ratio, and metallicity Fe/H are presented in Fig. 5. Our retrieval setup and priors are described in Appendix B.

The results obtained with MRCCS data only (in orange in Fig. 3) showed that the surface gravity, planet radius, and abundances of water and carbon monoxide all yielded Gaussian marginal distributions with the simulated values outside of 1σ of the mean. The equilibrium temperature was overestimated by 1σ resulting in a p-T profile that was about 150 K too hot at

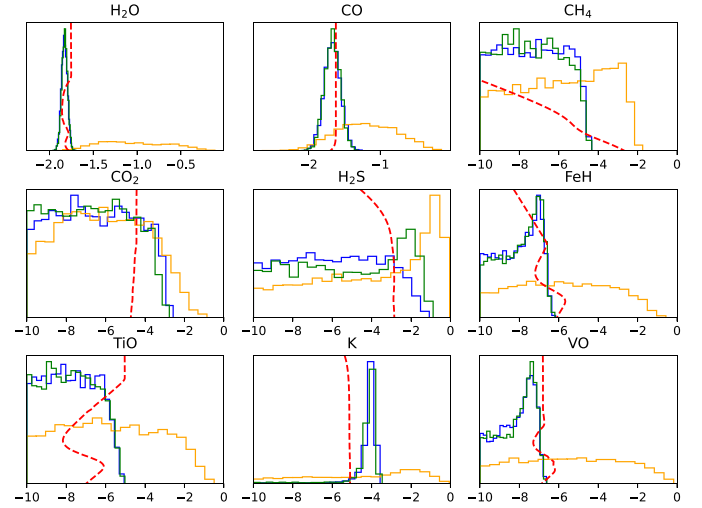


Fig. 4. Retrieved posterior distributions of all the molecular abundances included in the three retrievals of the experiment described in 3.4. The solid red lines represent the profile (in log scale) of each molecular abundance along the vertical extent of the atmosphere between 100 bar at the bottom and 10^{-6} bar at the top.

all altitudes relative to the ground truth, similarly to the planetary radius which was also overestimated. This behaviour is identical to what we observed in the previous section, with similar correlations between the same parameters, once again demonstrating the effects of the removal of the continuum. The posterior distributions of the molecular abundances resulted in a good constraint of CO with a large standard deviation of 0.4 dex. The abundance of H_2O was overestimated by 1 dex, while we obtained correct upper bounds (i.e. the true value was indeed lower than the bound) for CH_4 and CO_2 , and uniform posteriors for FeH, TiO, H_2S , K, and VO, which is expected from the weak spectral signatures of these molecules in the K band. The C/O ratio and metallicity did not get retrieved within 1σ , despite large standard deviations of 0.14 and 0.5 dex respectively.

When we applied CROCODILE to LRSP data only (in blue in Fig. 3), the equilibrium temperature, surface gravity, and planetary radius were retrieved within 1σ with a precision of 10 K, 0.05 dex, and $0.015 R_J$ respectively. The p-T profile was well constrained, with the true profile less than 20 K away from the posterior median. Contrary to the abundances of H_2O and CO which were well constrained with Gaussian posterior distributions, the abundances of FeH, K, and VO resulted in an upper bound, where the posterior maximum correctly identified an upper bound and lower values were not totally ruled out.

This can be attributed to the strong features produced by those molecules at short wavelengths, identified by the low-resolution SPHERE/IFS data. For the other molecules, we obtained correct upper bounds with the true abundances at lower values. The C/O ratio and metallicity were accurately retrieved within 1σ with a spread of 0.06 and 0.09 dex respectively.

Finally, when we applied CROCODILE to all three techniques (in green in Fig. 3), we obtained equivalent constraints on all parameters, with a negligible improvement on the spread of the C/O ratio and the accuracy of the metallicity. Interestingly, the lower C/O ratio constrained by MRCCS relative to LRSP is similar to the inconsistent C/O ratios that have been reported for the HR 8799 b planet: Lavie et al. (2017) used photometric and low-resolution H- and K-band spectroscopic data from KECK/OSIRIS (binned down to a resolution of 60

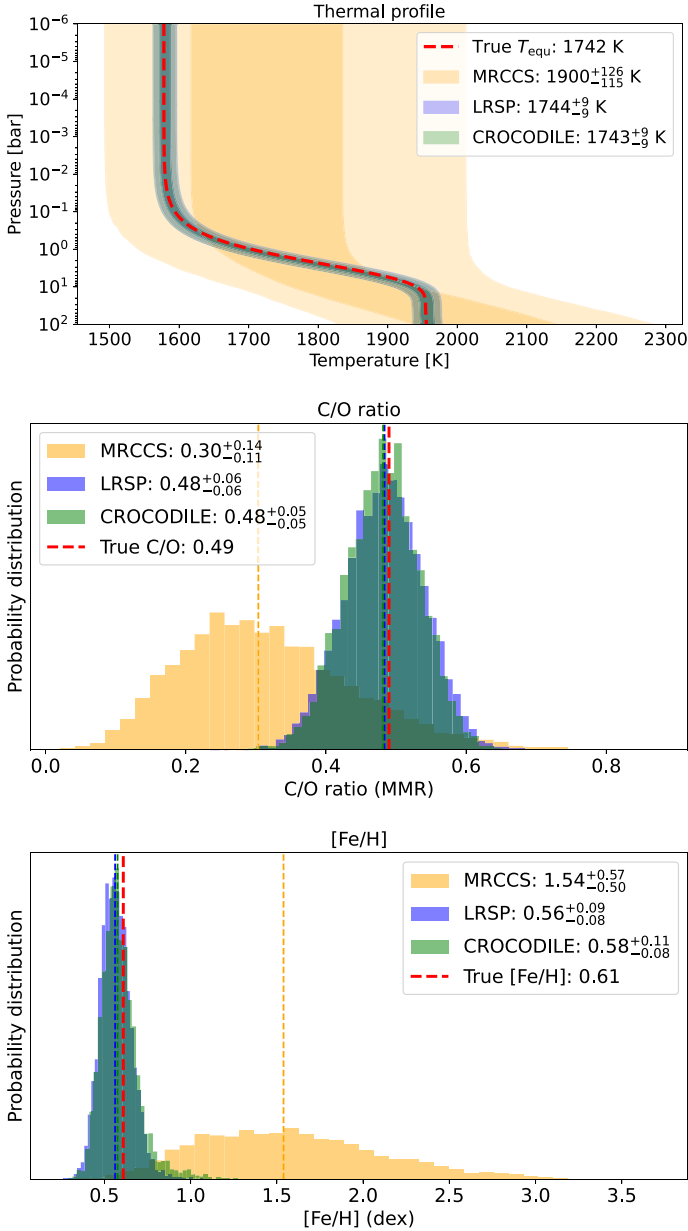


Fig. 5. Results of the experiment described in 3.4. Top panel: thermal profiles retrieved for each technique (MRCCS: orange, LRSP: blue, CROCODILE: green). The two levels of opacity correspond to the 16 to 84 and the 2 to 98 percentile envelopes of the thermal profile. The legend reports the median and 64 % confidence interval retrieved for the equilibrium temperature. Middle and bottom panel: posterior distributions of the C/O ratio and metallicity Fe/H inferred from the retrieved abundances for each technique. The dashed lines represent the posterior median. The dashed red line shows the ground truth in every panel.

to increase the S/N) and obtained a value close to unity, while Ruffio et al. (2021) derived a much lower value of 0.578 on medium-resolution ($R \approx 4000$) H and K band cross-correlation spectroscopic observations taken with the same instrument but processed differently.

In summary, our test seemed to further confirm the systematics of the retrieval on K -band MRCCS data observed in the previous section, while these systematics disappear when combined with LRSP data. In the next section, we investigate whether that is still the case when changing the input atmosphere.

3.5. Exploration of the parameter space

Since our results – derived for a simple synthetic model of β Pic b – might depend on the input atmosphere, we next explored how our framework fares in the parameter space around it as described in 3.1. The model setup and priors used for our retrievals were the same as in the previous section (readers can refer to Appendix B for details). Figure 6 reports the scores of three metrics applied to our retrievals in order to compare the results of the three techniques. Figure 7 focuses on the fifth atmospheric model as it is discussed below, while the fitted SEDs and posterior p-T profiles are shown in Figs. 8 and 9 respectively. Figure E.1 summarises all the resulting posterior distributions of the seven synthetic targets defined in Sect. 3.1, while the tables F.2, F.3, and F.4 report the numerical values.

In order to quantitatively compare the results obtained by LRSP, MRCCS, or LRSP+MRCCS, we computed a slightly modified Mahalanobis distance d_θ (Eq. (13)), the relative error ϵ_θ (Eq. (14)), and the mean squared error MSE_θ (Eq. (15)) from the marginalised posterior distribution of each parameter θ and for each simulated atmosphere:

$$d_\theta = (\hat{\theta} - \theta_T) / \sigma_\theta \quad (13)$$

$$\epsilon_\theta = (\hat{\theta} - \theta_T) / \theta_T \quad (14)$$

$$\text{MSE}_\theta = \frac{1}{N_S} \sum_{i=1}^{N_S} (\theta_i - \theta_T)^2. \quad (15)$$

In the equations above, $\hat{\theta}$ is the posterior median, θ_T is the ground truth of parameter θ , σ_θ is the distance of the median from the closest bound of the 68th percentile confidence interval (i.e. if $\theta < \hat{\theta}$, then σ_θ is equal to the difference between the median and the 16th percentile, and the difference between the median and the 84th percentile otherwise), and θ_i denotes the N_S points sampled by the nested algorithm. We choose these three metrics for the following reasons. First, the Mahalanobis metric measures the distance from the ground truth in terms of standard deviations (more precisely here in terms of 68th percentile confidence interval), meaning that a bad score is obtained when the uncertainty is underestimated. Second, the relative error measures the accuracy with which the parameters are retrieved. Finally, the mean squared error measures the quality of the median posterior as an estimator for the ground truth as it incorporates both its variance and its bias. Since the true molecular abundances have non-constant vertical profiles, θ_T represents the point of the true profile closest to the retrieved value, that is $(\hat{\theta} - \theta_T)$ is zero if there is an atmospheric layer where the retrieved abundance matches the true value. Figure 6 summarises the results of these three metrics for each retrieval case and for each technique (MRCCS, LRSP, or MRCCS+LRSP, i.e. CROCODILE), and by averaging over three groups of parameters: the molecular abundances, the physical parameters T_{equ} , $\log(g)$, and R , and the C/O ratio and metallicity.

The retrieval using MRCCS data alone resulted in the worst overall score with respect to every metric and for every subgroup of parameters, except for the molecular abundances with the Mahalanobis metric. This can be explained by the broad posterior distributions obtained with the technique, meaning that the ground truth was more often contained within the 68th percentile confidence interval although the retrieved value was significantly far from the true value, especially for T_{equ} , $\log(g)$, and R . Combining MRCCS with LRSP data lead to the lowest relative error and mean squared error for all groups of parameters, and yielded approximately equal Mahalanobis distance than the retrievals on

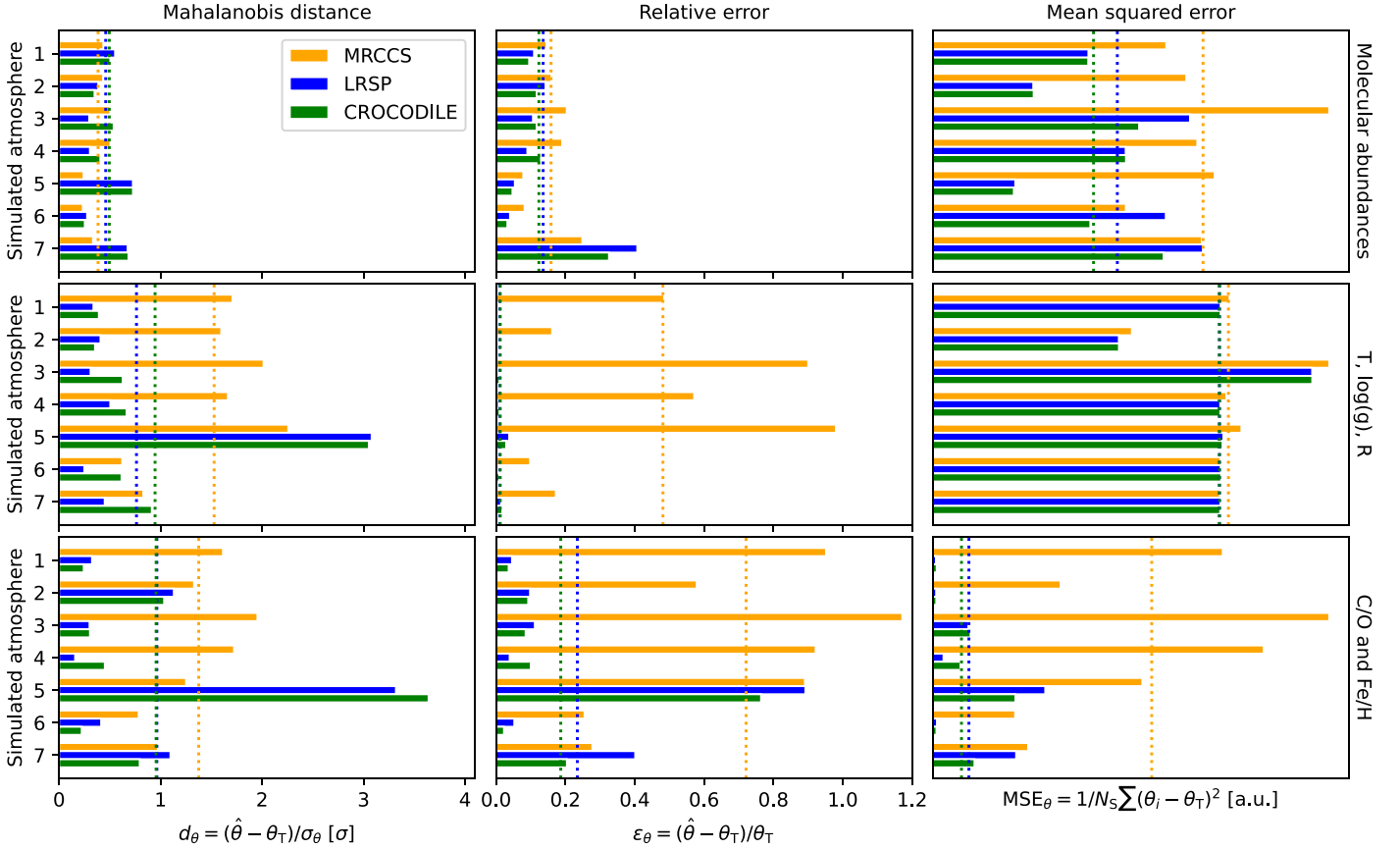


Fig. 6. Scores of the three metrics summarising the results of our exploration of the parameter space of gas giants using LRSP, MRCCS, or LRSP+MRCCS (CROCODILE). Each column of panels represents a different metric, and the three rows refer to different parameters grouped together. Within each panel, the score is shown for each technique and for each simulated atmosphere (cf. Table 2 for the input parameters), and averaged over the group of parameters. The dotted lines show the average score over the seven cases.

LRSP-only data. Overall, that is across all the cases, the combined fit on LRSP and MRCCS data achieved to retrieve 68% of the parameters within 0.7σ , against 0.6σ for the LRSP-only fit and 1.1σ for the MRCCS-only fit, while the same proportion of the parameters was retrieved with a relative error of less than 6% for LRSP+MRCCS, against 8% for LRSP and 28% for MRCCS. Hence, the combined fit on MRCCS and LRSP data was overall slightly more accurate and resulted in less underestimation of the uncertainty than the fit on LRSP data only, while the fit on MRCCS data only was largely inaccurate.

It is worth noting which molecule could be constrained and in what scenario. While the abundances of H_2O and CO resulted in narrow posterior distributions, those of CO_2 , H_2S , TiO, and K almost were never retrieved accurately, with at most upper bounds and only a good constraint of K at high C/O ratio. This can be explained by the strong absorption features of H_2O and CO included in the spectral range covered by our synthetic data, while this is not the case for CO_2 , H_2S , TiO, and K. This can be seen in Fig. 10, where we illustrate the spectral signatures of the molecules included in our retrieval by computing the absolute difference in flux between the full chemical model at equilibrium and when each molecule is removed successively. The abundance of CH_4 seemed to be retrievable when it was significantly higher in the input atmosphere, namely at low temperature and at high C/O ratio. Finally, FeH, and VO obtained similar results with mostly correct upper bounds and a few good constraints of FeH at high C/O ratio and low metallicity when LRSP data were included, which can be explained by the spectral features of FeH

and VO that are covered by the VLT/SPHERE/IFS $Y - H$ -band LRS data but were not included in the MRCCS data.

Looking at Fig. 6 case by case, it seems that there is one case that yielded relatively worst scores than the others when including LRSP data (i.e. LRSP alone or combined with MRCCS), namely at high C/O ratio (case 5). Indeed, all three subgroups of parameters resulted in relatively higher Mahalanobis distance and relative error than the other cases. This can be understood better in Fig. E.1: the retrieved surface gravity, the abundances of CO and K as well as the metallicity all overestimated the ground truth by a few σ . However, this is more due to the small spread of the posterior distributions (as is the case for the surface gravity), quickly resulting in large Mahalanobis distance, or to the fact that the ground truth itself is a small number (for example the metallicity), also resulting in a large relative error. Looking at the mean squared error, the fifth case actually obtained fairly similar scores to the others.

Finally, the quality of an atmospheric retrieval is usually estimated from the fit to the input spectrum, which we provide in Fig. 8. Remembering that all our datasets contained random scatter corresponding to their errorbar, the spectral fit – computed using the median of the posterior distributions of each parameter – showed no significant deviation from the LRSP data when it was included in the fit. Non-surprisingly, the retrieval using only MRCCS data did not match the LRSP data due to its inability to constrain the continuum and thus the thermal profile of the atmosphere. Indeed, as shown in Fig. 9, the temperature-pressure profile was largely unconstrained by the MRCCS data with a 1σ

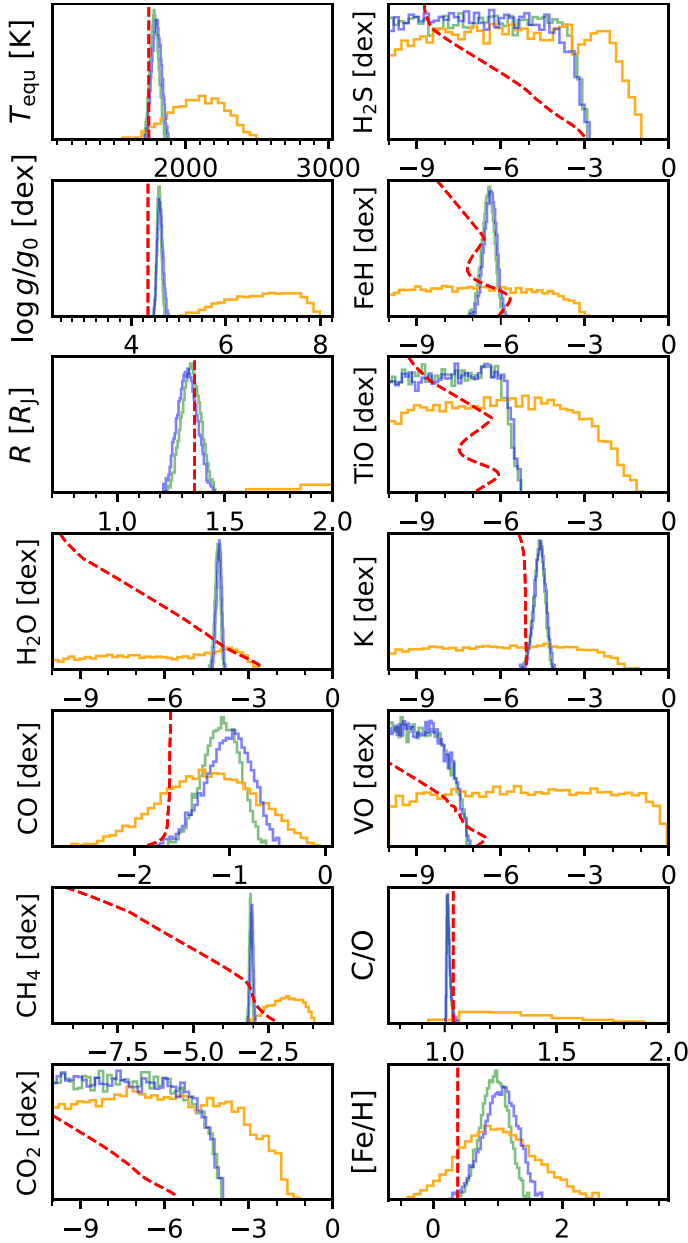


Fig. 7. Posterior distributions obtained for the retrievals of our simulated atmosphere 5 (cf. Table 2) using CROCODILE on MRCCS data alone (orange), on LRSP data (blue), and on both MRCCS+LRSP combined (green). The input values are shown as dashed red lines.

envelope of 200 K mostly overestimating the input profile except in the cases 6 and 7. Both the fit on LRSP data only and the combined fit were able to precisely constrain the thermal profile.

4. Discussion

4.1. Applicability of CROCODILE

The results presented in the last section show the robustness of the framework of CROCODILE to retrieve the parameters of a diverse population of gas giant atmospheres across a range of temperatures and chemical composition. However, our analysis also showed that the addition of *K*-band SINFONI MRCCS data along the list of LRSP data in the retrieval did not significantly improve our results: the posterior distributions mostly matched,

with a slight gain of accuracy and better estimation of the uncertainty as reported in Sect. 3.5. We argue that this is because of the high-quality *K*-band VLT/GRAVITY spectrum at a spectral resolution of $R \approx 500$ which can already constrain the abundance of CO on its own, and hence the medium-resolution spectrum provided by VLT/SINFONI does not add much information that is not already there. To demonstrate this point, we repeated the retrievals on case 1, however this time we excluded the GRAVITY spectrum from the fit. The results of this final test are presented in Fig. 11 and in Table F.5. Indeed, the low-resolution *Y*- to *H*-band VLT/SPHERE spectrum together with the photometric points were not enough on their own to constrain the molecular abundance of CO, leading to a mostly undetermined C/O ratio and a poorly constrained metallicity. When the *K*-band SINFONI data were added to the fit, the difference was significant: the abundance of CO became well constrained, leading to a good estimation of the C/O ratio ($C/O = 0.41^{+0.17}_{-0.16}$ against the input value of 0.49) and accurate constraint of the metallicity ($Fe/H = 0.61^{+0.30}_{-0.23}$ against the input value of 0.61).

While Sect. 3.5 demonstrated the robustness of CROCODILE against different input atmospheres, this last test showed its actual strength. Namely, it allows the combination of low signal-to-noise cross-correlation spectroscopic data together with a few photometric points and low-resolution spectroscopic data to measure the molecular abundances of closely separated directly imaged low-mass companions, without the need for high-quality spectroscopic data such as measured by VLT/GRAVITY.

4.2. Future opportunities

The major disadvantage of removing the continuum to perform cross-correlation with spectral templates is made obvious in Fig. E.1: the information carrying the temperature is mostly lost with the continuum, rendering its determination almost impossible using only cross-correlation and further degrading the determination of the other parameters such as surface gravity, C/O ratio, and metallicity. It therefore seems that the continuum should not be removed. If it is true that molecular lines present in substellar companions differ enough from the atomic lines found in stellar spectra, then this fact stays true if the continuum of the planet can be retrieved. Several studies are indeed able to retrieve the continuum of directly imaged low-mass companions using medium-resolution spectroscopy: for example Daemgen et al. (2017) and Petrus et al. (2021) were able to measure the spectrum of the planetary-mass objects HD 106906 b and HIP 65426 b using VLT/SINFONI, while most recently Miles et al. (2022) released the spectrum of the young brown dwarf VHS 1256 b taken with the NIRSpec IFU and MIRI MRS onboard JWST, the highest fidelity spectrum to date of a planetary-mass object, which will undoubtedly yield the most robust derivation of the atmospheric properties of a planetary-mass object to date.

The point of our study is not that the continuum of those spectra could have been removed with similar constraints for the atmosphere of the exoplanet, but rather to enable the study of close-in low-mass companions where the continuum cannot be extracted at all (or at least not within reasonable integration times) due to contamination by stellar speckles, as is the case in Hoeijmakers et al. (2018), Cugno et al. (2021), Ruffio et al. (2021), and Wang et al. (2021). Indeed, future direct-imaging searches for exoplanets will focus more and more on the innermost regions of planetary systems, where radial velocity searches indicate that the occurrence rate of planets is higher (Cumming et al. 2008; Fulton et al. 2021). However, at such close separations, the location of a planet is only a few resolution elements

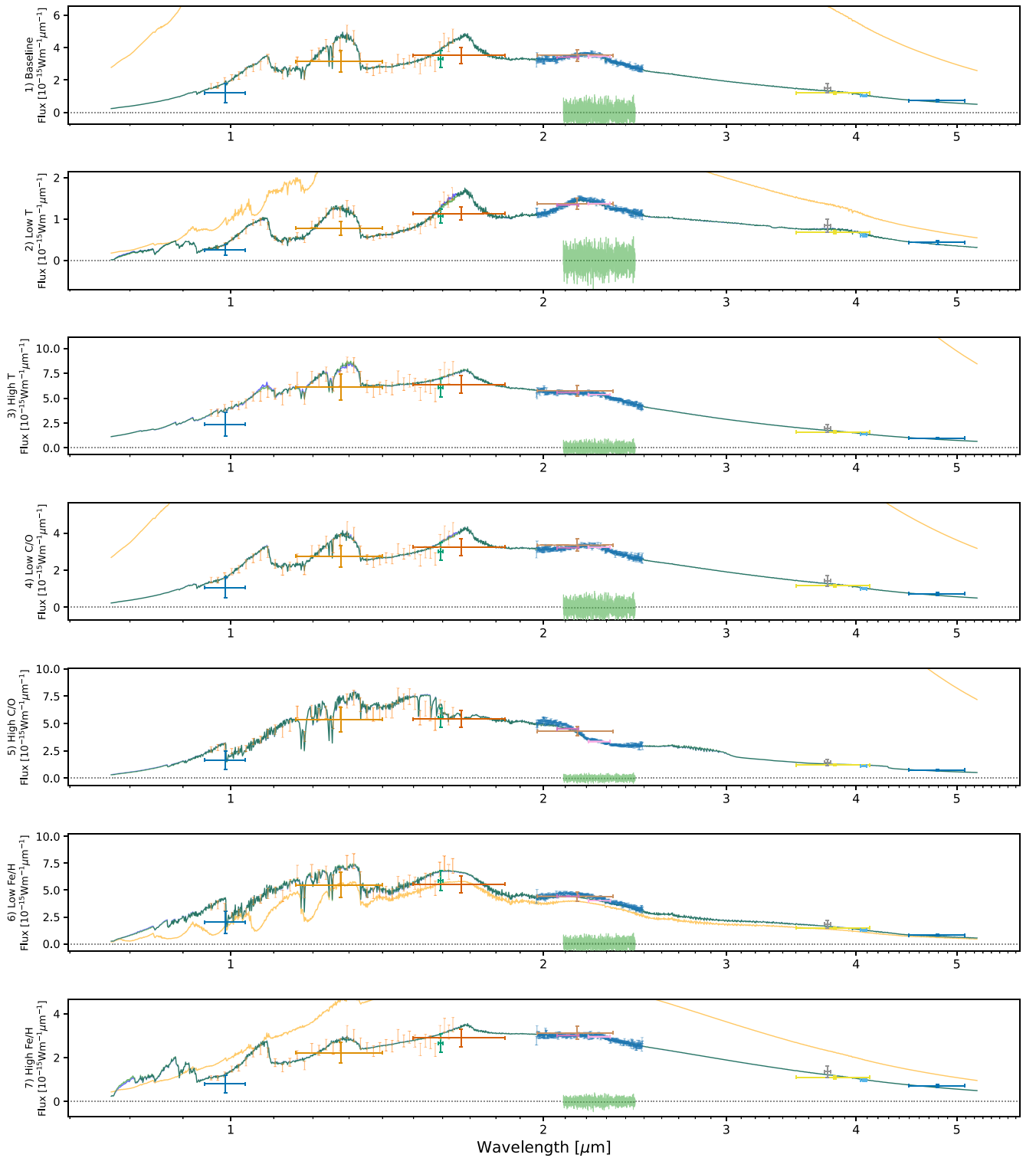


Fig. 8. Comparison of the fitted spectra to our simulated datasets 1 to 7 (cf. Table 2) using CROCODILE on MRCCS data alone (orange), on LRSP data (blue), and on both MRCCS+LRSP combined (green). The fitted spectra were calculated using our forward model from the posterior median of the retrieved distributions shown in Fig. E.1. The 11 colourful crosses represent the simulated photometric values and their corresponding uncertainty and equivalent width; the orange and blue error bars show the simulated *YH*-band VLT/SPHERE/IFS and *K*-band VLT/IRISA spectra and corresponding uncertainties; and the continuum-subtracted light green spectrum shows the simulated VLT/SINFONI MRCCS data.

from the central star. There, the stellar PSF is still so intense that extracting medium and high spectral resolution spectra for such planets will come at the cost of the continuum – at least initially. That is when CROCODILE – and similar statistical frameworks

based on the cross-correlation of continuum-removed spectra with spectral models (e.g. Ruffio et al. 2021) – comes into play by enabling a full Bayesian analysis to characterise more and more challenging directly imaged exoplanets.

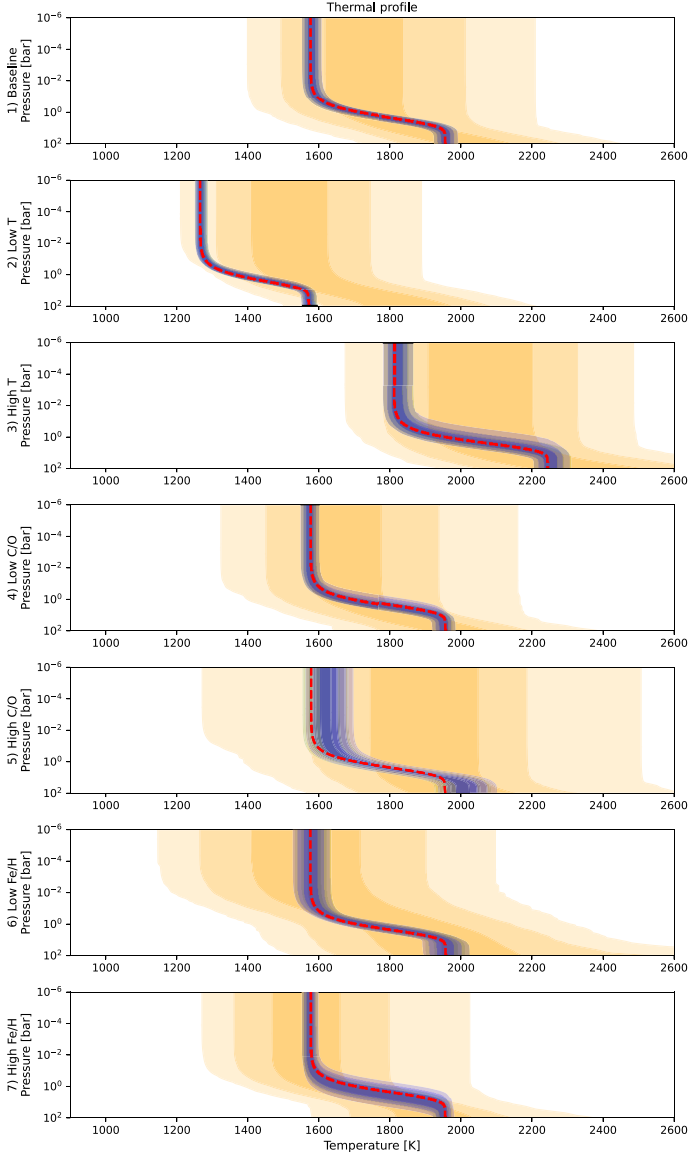


Fig. 9. Posterior thermal profiles obtained for the retrievals of our simulated atmospheres 1 to 7 (cf. Table 2) using CROCODILE on MRCCS data alone (orange), on LRSP data (blue), and on both MRCCS+LRSP combined (green). The different envelopes of each colour represent the 68th, 95th, and 99th percentile credible intervals – corresponding to 1σ , 2σ , and 3σ of a normal distribution – of the posterior distributions of the retrieved thermal profiles. The input p-T profiles are shown as dashed red lines.

5. Summary and conclusion

The results of this work are summarised in the following:

- We adapted the statistical framework of Brogi & Line (2019) from exoplanet transit spectroscopy to the different observational techniques of directly imaged exoplanets, including photometry, low-resolution spectroscopy, and medium (and higher) resolution cross-correlation spectroscopy, which can constrain the atmospheric properties of gas giant exoplanets such as temperature-pressure profile and chemical composition. Our Python routine, called CROCODILE, is freely available for download on GitHub⁵.

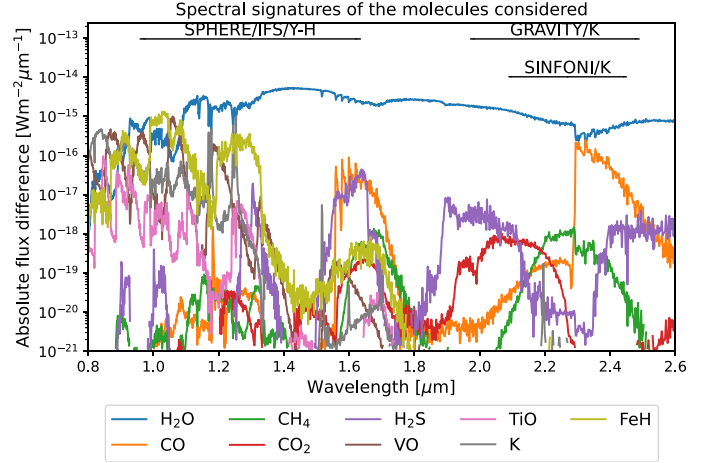


Fig. 10. Absolute difference between the spectrum calculated with the full chemical model used for our simulated atmosphere 1 (as described in Table 2) and spectra where each molecule included in our framework was removed one after the other. For example water shows a large difference due to its absorption features throughout the whole range. The black lines indicate the spectral coverage of the different spectroscopic instruments included in our tests.

- We thoroughly tested our framework, first on a synthetic atmosphere created with the same input as our forward model to verify the framework, and then on a grid of atmospheres at chemical equilibrium to test the ability of our free forward model to robustly constrain atmospheric properties in a realistic scenario.
- We find that using medium-resolution cross-correlation spectroscopy alone leads to a generally poor constraint of the atmospheric parameters due to the loss of the continuum. The results improved a lot when fixing the planetary radius to its ground truth. However, we noticed strong correlations between the radius and the surface gravity and the molecular abundances of H_2O and CO . This means that fixing the radius to a wrong value might strongly bias the retrieved values.
- We further find that combining medium-resolution cross-correlation spectroscopy to photometry and low-resolution spectroscopy leads to an accurate determination of the atmospheric parameters. Indeed, 68% of the parameters were retrieved within 0.9σ of the ground truth, with no systematical bias despite the different chemical model used in the forward model and in the simulated data.
- Our last test shows that low-quality (i.e. high noise) medium-resolution ($R \approx 4000$) K -band cross-correlation spectroscopy, when combined with photometry, leads to an equivalently accurate constraint of the abundance of CO than high-quality (i.e. low noise) low-resolution ($R \approx 500$) K -band spectroscopy. To put into perspective, the actual observations that our synthetic data are based on was acquired in 2.5 h of integration time for VLT/SINFONI on one telescope (Hoeijmakers et al. 2018) versus 1.4 hr for VLT/GRAVITY on four telescopes (Nowak et al. 2020), that is the medium-resolution cross-correlation spectroscopy required less than half of the total telescope time that was required for the low-resolution spectro-interferometry.
- Our work shows that medium-resolution cross-correlation spectroscopy in the K band is best interpreted alongside $Y - M$ -band photometry and, if available, low-resolution spectroscopy in the $Y - H$ bands, to robustly infer the atmospheric properties of gas giant exoplanets.

⁵ <https://github.com/JHayoz/CROCODILE>

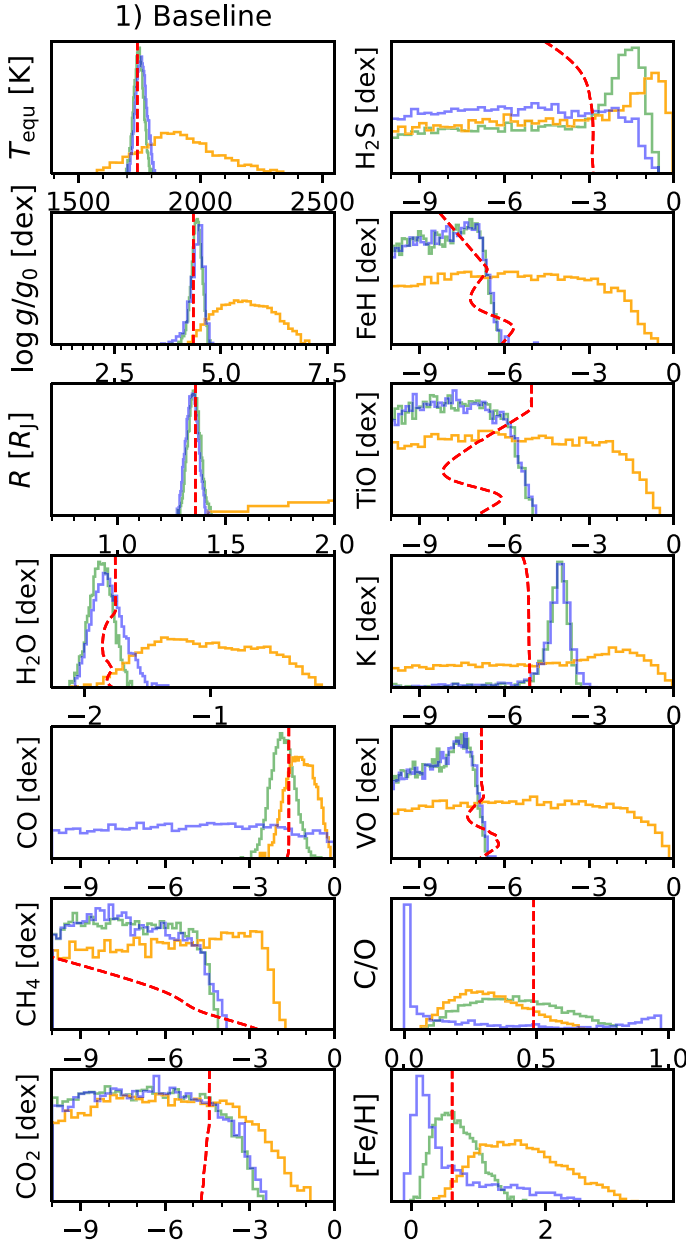


Fig. 11. Posterior distributions obtained when removing the synthetic VLT/IRAF data from the atmospheric fit. The input atmosphere is the same as case 1 (cf. Table 2), and as in Fig. 7 we also compare the results when applying CROCODILE to MRCCS data alone (orange), LRSP data (blue), and on both MRCCS+LRSP combined (green). However, this time the low-resolution spectroscopy only included synthetic VLT/SPHERE data in the $Y-H$ bands. The input values are shown as dashed red lines.

- In summary, CROCODILE provides the statistical framework to interpret medium- (and possibly higher-) resolution spectroscopic data alongside photometric and low-resolution spectroscopic data of directly imaged gas giant exoplanets at close separations to their host star, where the continuum of their spectrum can only be extracted reliably at low spectral resolution due to the pollution by the stellar PSF. CROCODILE allows the measurement of the atmospheric properties of gas giants such as the pressure-temperature profile and the abundances of individual molecules. In the future, we plan to add the option to choose between different thermal, chemical,

and cloud models, in order to allow model selection via computation of Bayes' factor between competing models.

With the arrival of the next generation of exoplanet direct-imaging instruments such as ERIS at the VLT, MIRI onboard JWST, and the upcoming METIS at the ELT, it is crucial to define robust statistical frameworks that take into account the intricacies of the different observational and post-processing techniques, while acknowledging the risk of model degeneracies. In particular, such frameworks should be validated on a representative subset of the parameter space of exoplanetary atmospheres to identify the limits of applicability of the forward model before being carried out on real data.

Acknowledgements. The authors thank the anonymous referee for their constructive insights which helped to improve the quality of this work. JH and SPQ gratefully acknowledge the financial support from the Swiss National Science Foundation (SNSF) under project grant number 200020_200399. GC, PP, EOG thank the SNSF for financial support under grant numbers P500PT_206785 and 200020_200399. MJB acknowledges the financial support from ETH Zurich. This work has been carried out within the framework of the NCCR PlanetS supported by the Swiss National Science Foundation under grants 51NF40_182901 and 51NF40_205606. TDG acknowledges funding from the Max Planck ETH Center for Learning Systems. Some of our plots were made using the Python package `corner.py` developed by Foreman-Mackey (2016). Contributions: the authors listed in this article are ordered according to the following: JH built the framework, did the analysis, and wrote the paper; GC, SPQ, and PP provided feedback and supervision during the whole project; while the rest of the co-authors, who are ordered alphabetically, provided feedback in later stages of the project.

References

- Allard, F. 2013, *Proc. Int. Astron.*, **8**, 271
 Bayes, T. 1763, *Phil. Trans. R. Soc.*, **53**, 370
 Beuzit, J.-L., Vigan, A., Mouillet, D., et al. 2019, *A&A*, **631**, A155
 Bonnefoy, M., Lagrange, A. M., Boccaletti, A., et al. 2011, *A&A*, **528**, L15
 Bonnefoy, M., Zurlo, A., Baudino, J. L., et al. 2016, *A&A*, **587**, A58
 Borysow, A. 2002, *A&A*, **390**, 779
 Borysow, A., & Frommhold, L. 1989, *ApJ*, **341**, 549
 Borysow, J., Frommhold, L., & Birnbaum, G. 1988, *ApJ*, **326**, 509
 Borysow, A., Frommhold, L., & Moraldi, M. 1989, *ApJ*, **336**, 495
 Borysow, A., Jørgensen, U. G., & Fu, Y. 2001, *J. Quant. Spec. Radiat. Transf.*, **68**, 235
 Brandl, B., Bettonvil, F., Van Boekel, R., et al. 2021, *The Messenger*, **182**, 22
 Brogi, M., & Line, M. R. 2019, *ApJ*, **157**, 114
 Bryan, M. L., Chiang, E., Bowler, B. P., et al. 2020, *AJ*, **159**, 181
 Buchner, J., Georgakakos, A., Nandra, K., et al. 2014, *A&A*, **564**, A125
 Burrows, A., & Sharp, C. M. 1999, *ApJ*, **512**, 843
 Burrows, A., Budaj, J., & Hubeny, I. 2008, *ApJ*, **678**, 1436
 Carnall, A. C. 2017, arXiv e-prints [arXiv:1705.05165]
 Chan, Y. M., & Dalgarno, A. 1965, *Proc. Phys. Soc.*, **85**, 227
 Chubb, K. L., Rocchetto, M., Yurchenko, S. N., et al. 2021, *A&A*, **646**, A21
 Cugno, G., Patapis, P., Stolker, T., et al. 2021, *A&A*, **653**, A12
 Cumming, A., Butler, R. P., Marcy, G. W., et al. 2008, *PASP*, **120**, 531
 Currie, T., Burrows, A., Madhusudhan, N., et al. 2013, *ApJ*, **776**, 15
 Czesla, S., Schröter, S., Schneider, C. P., et al. 2019, *PyA: Python astronomy-related packages*, Astrophysics Source Code Library [record ascl:1906.010]
 Daemgen, S., Todorov, K., Quanz, S. P., et al. 2017, *A&A*, **608**, A71
 Dalgarno, A., & Williams, D. A. 1962, *ApJ*, **136**, 690
 Davies, R., Esposito, S., Schmid, H. M., et al. 2018, in *Ground-based and Airborne Instrumentation for Astronomy VII* (SPIE)
 Eisenhauer, F., Abuter, R., Bickert, K., et al. 2003, in *SPIE Proceedings*, **4841: Instrument Design and Performance for Optical/Infrared Ground-based Telescopes (SPIE)
 Feroz, F., Hobson, M. P., & Bridges, M. 2009, *MNRAS*, **398**, 1601
 Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2019, *OJAp*, **2**, 10
 Foreman-Mackey, D. 2016, *JOSS*, **1**, 24
 Fortney, J. J., Lodders, K., Marley, M. S., et al. 2008, *ApJ*, **678**, 1419
 Fulton, B. J., Rosenthal, L. J., Hirsch, L. A., et al. 2021, *ApJS*, **255**, 14
 Gaia Collaboration 2018, *VizieR Online Data Catalog*: I/345
 Gandhi, S., & Madhusudhan, N. 2017, *MNRAS*, **472**, 2334
 Gandhi, S., & Madhusudhan, N. 2018, *MNRAS*, **474**, 271**

- George, E. M., Gräff, D., Feuchtgruber, H., et al. 2016, in [SPIE Proceedings, 9908: Ground-based and Airborne Instrumentation for Astronomy VI](#) (SPIE)
- Greco, J. P., & Brandt, T. D. 2016, [ApJ](#), **833**, 134
- Griffith, C. A., & Yelle, R. V. 2000, [ApJ](#), **532**, L59
- Guillot, T. 2010, [A&A](#), **520**, A27
- Haffert, S. Y., Bohn, A. J., de Boer, J., et al. 2019, [Nat. Astron.](#), **3**
- Hoch, K. K. W., Konopacky, Q. M., Theissen, C. A., et al. 2022, 2023, [AJ](#), **166**, 85
- Hoeijmakers, H. J., Schwarz, H., Snellen, I. A. G., et al. 2018, [A&A](#), **617**, A144
- Konrad, B. S., Alei, E., Quanz, S. P., et al. 2022, [A&A](#), **664**, A23
- Lagrange, A.-M., Bonnefoy, M., Chauvin, G., et al. 2010, [Science](#), **329**, 57
- Lagrange, A.-M., Rubini, P., Nowak, M., et al. 2020, [A&A](#), **642**, A18
- Langlois, M., Gratton, R., Lagrange, A.-M., et al. 2021, [A&A](#), **651**, A71
- Larkin, J., Barczys, M., Krabbe, A., et al. 2006, in [SPIE Proc., 6269: Ground-based and Airborne Instrumentation for Astronomy](#) (SPIE)
- Lavie, B., Mendonça, J. M., Mordasini, C., et al. 2017, [ApJ](#), **154**, 91
- Lee, J.-M., Fletcher, L. N., & Irwin, P. G. J. 2012, [MNRAS](#), **420**, 170
- Line, M. R., Zhang, X., Vasisht, G., et al. 2012, [ApJ](#), **749**, 93
- Line, M. R., Wolf, A. S., Zhang, X., et al. 2013, [ApJ](#), **775**, 137
- Line, M. R., Marley, M. S., Liu, M. C., et al. 2017, [ApJ](#), **848**, 83
- Macintosh, B., Graham, J. R., Ingraham, P., et al. 2014, [PNAS](#), **111**, 12661
- Madhusudhan, N. 2018, in [Handbook of Exoplanets](#) (Springer International Publishing), 2153
- Madhusudhan, N., Harrington, J., Stevenson, K. B., et al. 2011, [Nature](#), **469**, 64
- Males, J. R., Close, L. M., Morzinski, K. M., et al. 2014, [ApJ](#), **786**
- Miles, B. E., Biller, B. A., Patapis, P., et al. 2022, 2023, [ApJ](#), **946**, L6
- Mollière, P., Van Boekel, R., Bouwman, J., et al. 2017, [A&A](#), **600**, A10
- Mollière, P., Wardenier, J. P., Van Boekel, R., et al. 2019, [A&A](#), **627**, A67
- Mollière, P., Stolker, T., Lacour, S., et al. 2020, [A&A](#), **640**, 131
- Mollière, P., Molyarova, T., Bitsch, B., et al. 2022, [ApJ](#), **934**, 74
- Mordasini, C., van Boekel, R., Mollière, P., et al. 2016, [ApJ](#), **832**, 41
- Moses, J. I., Visscher, C., Fortney, J. J., et al. 2011, [ApJ](#), **737**, 15
- Nowak, M., Lacour, S., Mollière, P., et al. 2020, [A&A](#), **633**, A110
- Öberg, K. I., & Bergin, E. A. 2016, [ApJ](#), **831**, L19
- Öberg, K. I., Murray-Clay, R., & Bergin, E. A. 2011, [ApJ](#), **743**, L16
- Ohno, K., & Fortney, J. J. 2022, AAS J., submitted, [arXiv:2211.16877]
- Palma-Bifani, P., Chauvin, G., Bonnefoy, M., et al. 2023, [A&A](#), **670**, A90
- Patapis, P., Nasedkin, E., Cugno, G., et al. 2022, [A&A](#), **658**, A72
- Petit dit de la Roche, D. J. M., Hoeijmakers, H. J., & Snellen, I. A. G. 2018, [A&A](#), **616**, A146
- Petrus, S., Bonnefoy, M., Chauvin, G., et al. 2021, [A&A](#), **648**, A59
- Piskunov, N. E., Kupka, F., Ryabchikova, T. A., Weiss, W. W., & Jeffery, C. S. 1995, [A&AS](#), **112**, 525
- Piso, A.-M. A., Pegues, J., & Öberg, K. I. 2016, [ApJ](#), **833**, 203
- Quirrenbach, A., Larkin, J. E., Krabbe, A., Barczys, M., & LaFreniere, D. 2003, in [SPIE Proc., 4841: Instrument Design and Performance for Optical/Infrared Ground-based Telescopes](#), eds. M. Iye, & A. F. M. Moorwood, 1493
- Richard, C., Gordon, I., Rothman, L., et al. 2012, [J. Quant. Spec. Radiat. Transf.](#), **113**, 1276
- Rothman, L., Gordon, I., Barber, R., et al. 2010, [J. Quant. Spec. Radiat. Transf.](#), **111**, 2139
- Rothman, L., Gordon, I., Babikov, Y., et al. 2013, [J. Quant. Spec. Radiat. Transf.](#), **130**, 4
- Ruffio, J.-B., Macintosh, B., Konopacky, Q. M., et al. 2019, [AJ](#), **158**, 200
- Ruffio, J.-B., Konopacky, Q. M., Barman, T., et al. 2021, [AJ](#), **162**, 290
- Seager, S., Richardson, L. J., Hansen, B. M. S., et al. 2005, [ApJ](#), **632**, 1122
- Skilling, J. 2006, [Bayesian Anal.](#), **1**, 833
- Sparks, W. B., & Ford, H. C. 2002, [ApJ](#), **578**, 543
- Stolker, T., Bonse, M. J., Quanz, S. P., et al. 2019, [A&A](#), **621**, A59
- Stolker, T., Quanz, S. P., Todorov, K. O., et al. 2020, [A&A](#), **635**, A182
- Todorov, K. O., Line, M. R., Pineda, J. E., et al. 2016, [ApJ](#), **823**, 14
- van der Marel, N., Bosman, A. D., Krijt, S., et al. 2021, [A&A](#), **653**, A9
- Waldmann, I. P., Tinetti, G., Rocchetto, M., et al. 2015, [ApJ](#), **802**, 107
- Wang, J., Wang, J. J., Ma, B., et al. 2020, [AJ](#), **160**, 150
- Wang, J. J., Ruffio, J.-B., Morris, E., et al. 2021, [AJ](#), **162**, 148
- Wang, J., Wang, J. J., Ruffio, J.-B., et al. 2023, [AJ](#), **165**, 4
- Wells, M., Pel, J.-W., Glaspe, A., et al. 2015, [PASP](#), **127**, 646
- Xuan, J. W., Wang, J., Ruffio, J.-B., et al. 2022, [ApJ](#), **937**, 54
- Yurchenko, S. N., & Tennyson, J. 2014, [MNRAS](#), **440**, 1649
- Zurlo, A., Vigan, A., Galicher, R., et al. 2016, [A&A](#), **587**, A57

Appendix A: Derivation of the log-likelihood function for cross-correlation spectroscopy

In the following, we provide the derivation of the log-likelihood function $\log \mathcal{L}_{\text{CCS}}$ given in Eq. 4 originally developed by Brogi & Line (2019), but adapted to our notation. Let $d = (d_1, \dots, d_N)$ be the measured continuum-removed spectrum of an observed exoplanet with N spectral channels and $m(\theta, v_R) =: m = (m_1, \dots, m_N)$ be the associated forward model. For the sake of readability, we drop the dependency on the model parameters θ and the radial velocity V_R in the following. We assume that the data are of the form $d = am + n$, with a scaling factor a and noise $n = (n_1, \dots, n_N)$, where the components obey a normal distribution with a common standard deviation σ , that is $n_i \sim N(0, \sigma)$. Brogi & Line (2019) choose $a = 1$ and scale their model spectrum by the stellar flux and the area ratio between the planet and the star. In our study, the data and the forward model were created with the same scaling factor, therefore $a = 1$ also holds here. However, if the observed spectrum is provided in arbitrary units, then one could replace the multiplicative factor R^2/d^2 used in our forward model by a and treat it as a nuisance parameter.

The likelihood function associated to the noise statistics is given by

$$\mathcal{L} = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^N \exp \left(- \sum_{i=1}^N \frac{(d_i - m_i)^2}{2\sigma^2} \right), \quad (\text{A.1})$$

which leads to the following log-likelihood function (ignoring constant additive terms):

$$\log \mathcal{L} \approx -N \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^N (d_i - m_i)^2. \quad (\text{A.2})$$

The variable σ represents the uncertainty associated to each spectral channel and is assumed to be constant for all spectral channels. It is challenging to compute due to the pollution by telluric absorption lines or stellar light. Therefore, we replace it by its maximum likelihood estimator $\hat{\sigma}$:

$$0 = \frac{\partial \log \mathcal{L}}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^N (d_i - m_i)^2$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (d_i - m_i)^2. \quad (\text{A.3})$$

By inserting $\hat{\sigma}$ back into Eq. A.2 and expanding the product $(d_i - m_i)^2 = d_i^2 - 2d_i m_i + m_i^2$, we finally obtain:

$$\log \mathcal{L} \approx -\frac{N}{2} \log \left(\frac{1}{N} \sum_{i=1}^N (d_i^2 - 2d_i m_i + m_i^2) \right) - \frac{N}{2}$$

$$\approx -\frac{N}{2} \log (s_d^2 - 2K + s_m^2), \quad (\text{A.4})$$

with the definitions

$$s_d^2 = \frac{1}{N} \sum_{i=1}^N d_i^2 \quad (\text{A.5})$$

$$s_m^2 = \frac{1}{N} \sum_{i=1}^N m_i^2 \quad (\text{A.6})$$

$$K = \frac{1}{N} \sum_{i=1}^N d_i m_i, \quad (\text{A.7})$$

which corresponds to Eq. 4. For the sake of clarity, we note here again that instead of the cross-covariance K , one can use the normalised cross-correlation C by using Eq. 8 and inserting it into the log-likelihood function:

$$\log \mathcal{L} \approx -\frac{N}{2} \log (s_d^2 - 2K + s_m^2) \quad (\text{A.8})$$

$$\approx -\frac{N}{2} \log (s_d^2 - 2C \sqrt{s_d^2 s_m^2} + s_m^2). \quad (\text{A.9})$$

Appendix B: The retrieval setup

Our atmospheric retrieval was setup in the following configuration. We set the equilibrium temperature T_{equ} and logarithm of the surface gravity $\log g$ as free retrievable parameters governing the thermal structure. Implicitly, the thermal structure from Guillot (2010) contains four more parameters which were kept fixed in the simulated datasets and in the forward model: the average opacity in the infrared $\kappa_{\text{IR}} = 0.01 \text{ cm}^2/\text{g}$, the ratio of the average opacity in the optical and infrared $\gamma = \kappa_V/\kappa_{\text{IR}} = 0.4$, the interior temperature $T_{\text{int}} = 200 \text{ K}$, and the bottom pressure $P_0 = 100 \text{ bar}$. For the free parameters governing the chemical composition, we included the logarithms of the mass fractions (assumed to be vertically constant) of the following molecules: H_2O , CO , CO_2 , CH_4 , H_2S , FeH , TiO , K , and VO . Finally, we set the abundances of H_2 and He equal to 75 % and 25 % of the remaining mass fraction in each layer. The last free parameter of the model was the radius of the planet R . The priors used in this study are given in Table B.1.

Table B.1: Priors used.

Parameter	Prior	Parameter	Prior
T_{equ}	$\mathcal{U}(0, 5000)$	$\log(\text{CO}_2)$	$\mathcal{U}(-10, 0)$
$\log(g/g_0)$	$\mathcal{U}(1, 8)$	$\log(\text{H}_2\text{S})$	$\mathcal{U}(-10, 0)$
R	$\mathcal{U}(0.1, 10)$	$\log(\text{FeH})$	$\mathcal{U}(-10, 0)$
$\log(\text{H}_2\text{O})$	$\mathcal{U}(-10, 0)$	$\log(\text{TiO})$	$\mathcal{U}(-10, 0)$
$\log(\text{CO})$	$\mathcal{U}(-10, 0)$	$\log(\text{K})$	$\mathcal{U}(-10, 0)$
$\log(\text{CH}_4)$	$\mathcal{U}(-10, 0)$	$\log(\text{VO})$	$\mathcal{U}(-10, 0)$

Appendix C: Derivation of the optimal size of the Gaussian filter

We derived the optimal size of the Gaussian filter used to remove the continuum of our synthetic medium-resolution spectroscopic data with the following test. We repeated the setup of Sect. 3.3 on MRCCS data alone for filter sizes of 1.2, 2.4, 4.9, 7.4, 14.7, 22.1, 29.4, 36.8, and 49 nm (corresponding to 5, 10, 20, 30, 60, 90, 120, 150, and 200 spectral bins at the spectral resolution of VLT/SINFONI), and computed the same three metrics as in Sect. 3.5. To obtain an overall score for each of the three metrics, we multiplied the results for each parameter together, and divided by the worst score to give the results as percentages. The result of this analysis is shown in Fig. C.1. The filter size at 4.9 nm outperformed all other filters with respect to the Mahalanobis and relative distance, and was a close second in terms of mean squared error (7.0% against 5.4%).

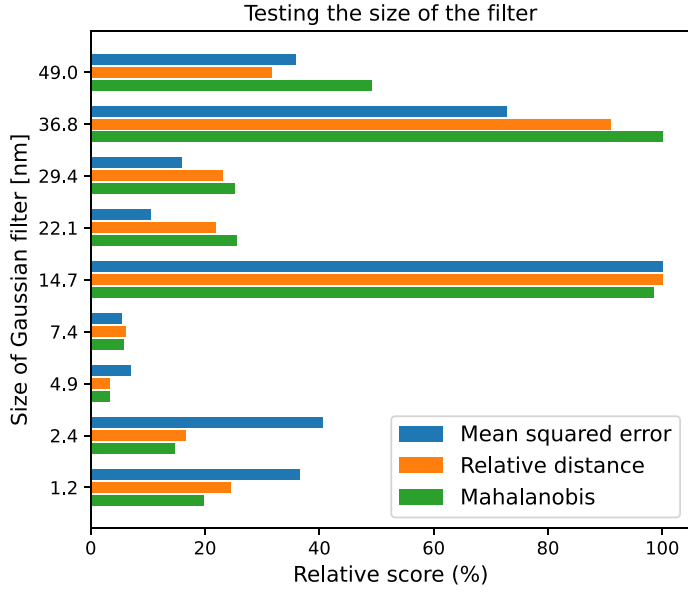


Fig. C.1: Scores for different sizes of the Gaussian filter used to remove the continuum of the synthetic medium-resolution spectroscopic data used in this work.

Appendix D: Results: Verification of the framework

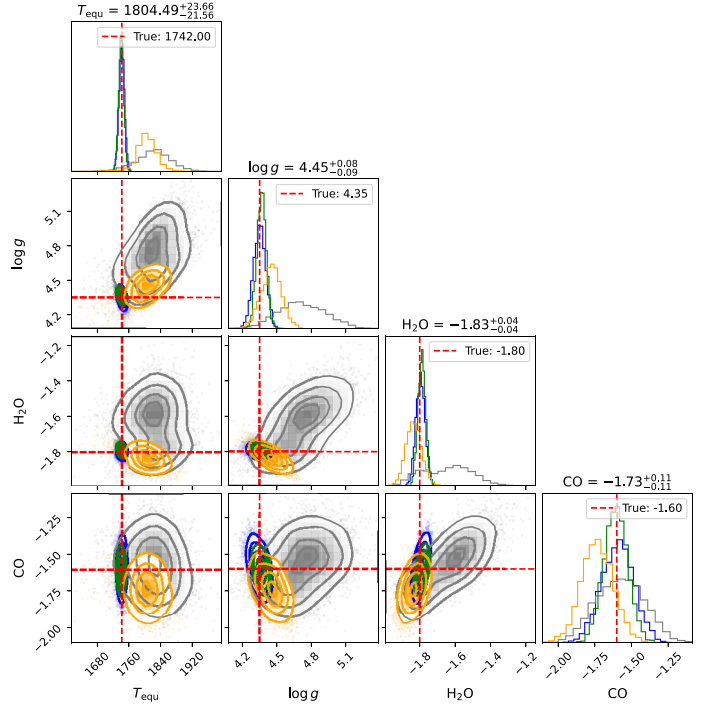


Fig. D.1: Resulting posterior distributions of our preliminary test on MRCCS data when fixing the radius to its simulated value (orange), compared to the previous results with the radius included as free parameter (grey: old results on MRCCS data, blue: only LRSP, green: MRCCS and LRSP). The dashed red line shows the ground truth used as input.

Appendix E: Results: Exploration of the parameter space

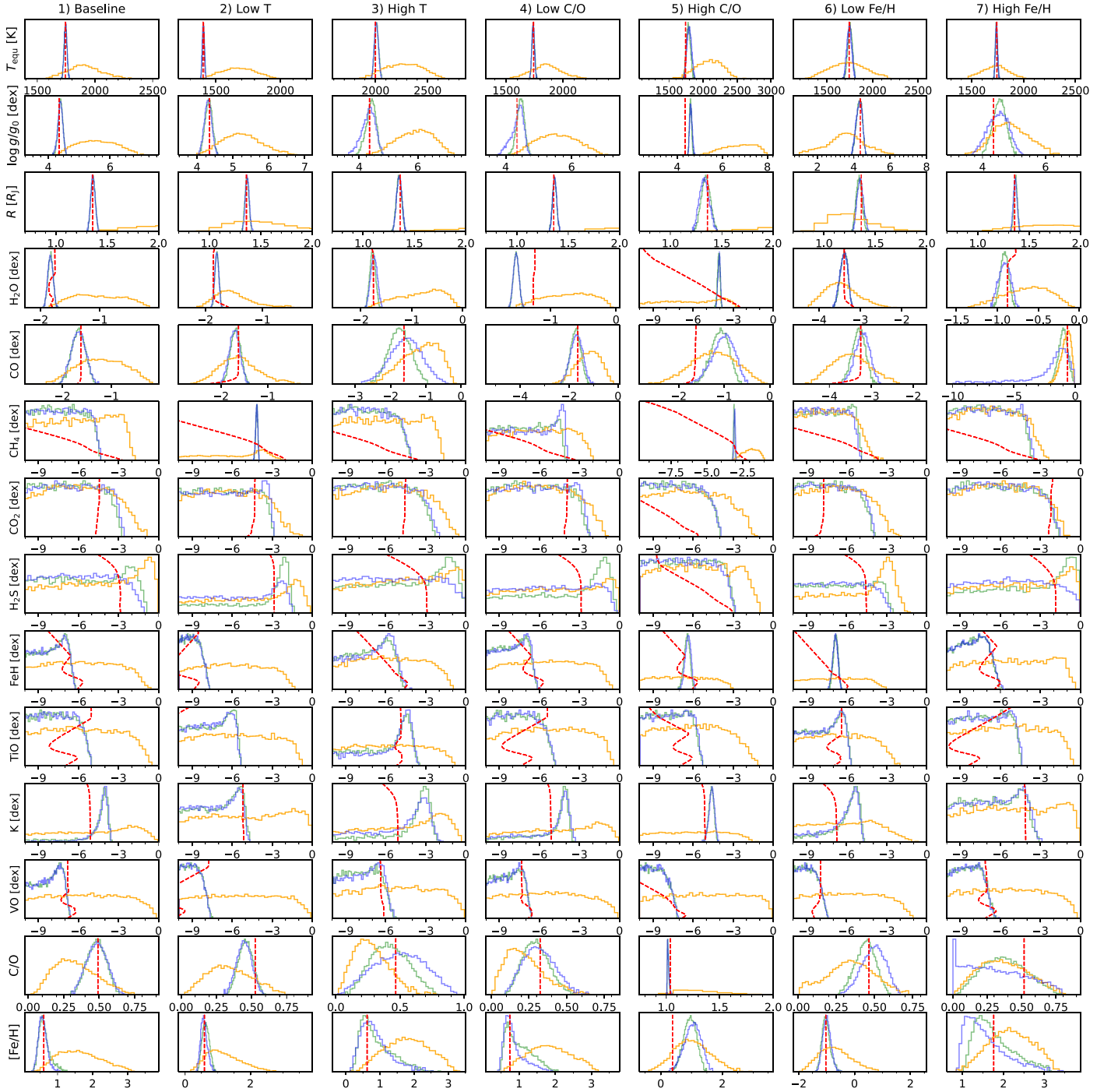


Fig. E.1: Comparison of the posterior distributions obtained for the retrievals of our simulated atmospheres 1 to 7 (cf. Table 2) using CROCODILE on MRCCS data alone (orange), on LRSP data (blue), and on both MRCCS+LRSP combined (green). The posterior distributions are shown as histograms of the sampled parameters and are to be read row by row for each model parameter and column by column to compare to the other input atmospheres. The input values are shown as dashed red lines. The input molecular abundances – which were calculated at chemical equilibrium and therefore vary vertically with the pressure and the temperature – are shown as functions of the atmospheric layers between 10^2 and 10^{-6} bar in logarithmic scale. The posterior distributions of the C/O ratio and metallicity Fe/H are inferred from the posterior distributions of the molecular abundances.

Appendix F: Detailed numerical results of our study

Table F.1: List of posterior median and 68th percentile credible interval for our verification of the statistical framework of CROCODILE (cf. Sect. 3.3).

Parameter		Truth	MRCCS	LRSP	MRCCS+LRSP	MRCCS w/o R
T_{equ}	[K]	1742	1823^{+41}_{-46}	1742^{+7}_{-7}	1740^{+6}_{-6}	1804^{+24}_{-22}
$\log(g/g_0)$	[dex]	4.35	$4.71^{+0.22}_{-0.21}$	$4.35^{+0.05}_{-0.05}$	$4.37^{+0.04}_{-0.04}$	$4.45^{+0.08}_{-0.09}$
R	[R_J]	1.36	$1.56^{+0.18}_{-0.15}$	$1.36^{+0.01}_{-0.01}$	$1.36^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$
H_2O	[dex]	-1.8	$-1.64^{+0.13}_{-0.17}$	$-1.80^{+0.03}_{-0.03}$	$-1.79^{+0.02}_{-0.02}$	$-1.83^{+0.05}_{-0.04}$
CO	[dex]	-1.6	$-1.57^{+0.17}_{-0.17}$	$-1.59^{+0.11}_{-0.11}$	$-1.61^{+0.08}_{-0.08}$	$-1.73^{+0.11}_{-0.11}$

Table F.2: List of posterior median and 68th percentile credible interval for the retrievals based only on MRCCS data.

Parameter		1) Baseline	2) Low T	3) High T	4) Low C/O	5) High C/O	6) Low Fe/H	7) High Fe/H
T_{equ}	[K]	1900^{+126}_{-115}	1675^{+120}_{-119}	2272^{+159}_{-167}	1850^{+110}_{-111}	2102^{+160}_{-177}	1727^{+168}_{-171}	1732^{+100}_{-112}
$\log(g/g_0)$	[dex]	$5.55^{+0.62}_{-0.58}$	$5.20^{+0.46}_{-0.43}$	$5.98^{+0.49}_{-0.58}$	$5.53^{+0.57}_{-0.52}$	$6.85^{+0.63}_{-0.81}$	$3.59^{+0.73}_{-0.77}$	$4.83^{+0.51}_{-0.41}$
R	[R_J]	$2.84^{+2.10}_{-0.87}$	$1.49^{+0.41}_{-0.25}$	$4.34^{+1.91}_{-1.86}$	$3.24^{+1.63}_{-1.07}$	$4.30^{+2.76}_{-1.76}$	$1.21^{+0.20}_{-0.14}$	$1.90^{+0.59}_{-0.45}$
H_2O	[dex]	$-1.07^{+0.49}_{-0.38}$	$-1.58^{+0.28}_{-0.21}$	$-0.71^{+0.32}_{-0.52}$	$-0.69^{+0.32}_{-0.40}$	$-5.84^{+2.13}_{-2.59}$	$-3.50^{+0.33}_{-0.29}$	$-0.60^{+0.22}_{-0.28}$
CO	[dex]	$-1.21^{+0.45}_{-0.42}$	$-1.64^{+0.32}_{-0.31}$	$-1.02^{+0.38}_{-0.53}$	$-1.13^{+0.37}_{-0.44}$	$-1.22^{+0.40}_{-0.41}$	$-3.51^{+0.37}_{-0.36}$	$-0.67^{+0.27}_{-0.35}$
CH_4	[dex]	$-5.74^{+2.44}_{-2.83}$	$-5.02^{+1.49}_{-3.37}$	$-5.96^{+2.77}_{-2.66}$	$-6.06^{+2.34}_{-2.49}$	$-1.89^{+0.48}_{-0.58}$	$-7.41^{+1.66}_{-1.69}$	$-6.73^{+2.07}_{-2.04}$
CO_2	[dex]	$-6.17^{+2.48}_{-2.33}$	$-6.31^{+2.38}_{-2.35}$	$-5.90^{+2.55}_{-2.58}$	$-6.17^{+2.54}_{-2.43}$	$-6.38^{+2.53}_{-2.35}$	$-6.81^{+2.06}_{-2.04}$	$-5.93^{+2.48}_{-2.54}$
H_2S	[dex]	$-3.37^{+2.57}_{-4.21}$	$-2.68^{+1.60}_{-4.68}$	$-4.01^{+2.97}_{-3.67}$	$-4.27^{+3.28}_{-3.58}$	$-5.74^{+2.80}_{-2.69}$	$-3.44^{+0.77}_{-3.85}$	$-3.96^{+3.04}_{-3.80}$
FeH	[dex]	$-5.81^{+2.72}_{-2.64}$	$-6.06^{+2.63}_{-2.44}$	$-5.88^{+2.72}_{-2.48}$	$-5.41^{+2.75}_{-2.88}$	$-6.84^{+2.16}_{-2.07}$	$-6.72^{+2.07}_{-2.10}$	$-5.07^{+3.03}_{-3.08}$
TiO	[dex]	$-5.87^{+2.96}_{-2.64}$	$-5.73^{+2.91}_{-2.71}$	$-5.71^{+2.75}_{-2.71}$	$-5.39^{+2.79}_{-2.80}$	$-6.05^{+2.38}_{-2.54}$	$-6.59^{+2.28}_{-2.20}$	$-5.40^{+2.78}_{-2.89}$
K	[dex]	$-3.90^{+2.29}_{-3.92}$	$-4.54^{+3.14}_{-3.55}$	$-3.11^{+1.55}_{-4.02}$	$-3.61^{+2.29}_{-3.78}$	$-5.88^{+2.49}_{-2.62}$	$-6.03^{+2.25}_{-2.63}$	$-4.85^{+3.37}_{-3.20}$
VO	[dex]	$-5.41^{+2.96}_{-2.98}$	$-5.34^{+3.02}_{-3.03}$	$-5.21^{+2.95}_{-2.97}$	$-5.45^{+2.86}_{-2.91}$	$-5.05^{+3.11}_{-3.15}$	$-5.97^{+2.62}_{-2.58}$	$-5.45^{+2.91}_{-2.96}$
C/O	[dex]	$0.30^{+0.14}_{-0.11}$	$0.32^{+0.15}_{-0.12}$	$0.25^{+0.12}_{-0.09}$	$0.19^{+0.11}_{-0.08}$	$1.29^{+0.41}_{-0.22}$	$0.35^{+0.12}_{-0.12}$	$0.35^{+0.17}_{-0.14}$
Fe/H	[dex]	$1.54^{+0.57}_{-0.50}$	$1.05^{+0.52}_{-0.35}$	$1.79^{+0.57}_{-0.57}$	$1.80^{+0.54}_{-0.48}$	$0.97^{+0.48}_{-0.44}$	$-0.76^{+0.50}_{-0.44}$	$2.17^{+0.48}_{-0.43}$

Table F.3: List of posterior median and 68th percentile credible interval for the retrievals based only on LRSP data.

Parameter		1) Baseline	2) Low T	3) High T	4) Low C/O	5) High C/O	6) Low Fe/H	7) High Fe/H
T_{equ}	[K]	1744^{+9}_{-9}	1400^{+6}_{-6}	2007^{+15}_{-14}	1738^{+9}_{-9}	1796^{+26}_{-25}	1742^{+19}_{-20}	1739^{+8}_{-8}
$\log(g/g_0)$	[dex]	$4.39^{+0.06}_{-0.06}$	$4.31^{+0.07}_{-0.08}$	$4.36^{+0.15}_{-0.19}$	$4.41^{+0.10}_{-0.15}$	$4.60^{+0.05}_{-0.04}$	$4.30^{+0.13}_{-0.15}$	$4.52^{+0.26}_{-0.28}$
R	[R_J]	$1.36^{+0.02}_{-0.02}$	$1.37^{+0.01}_{-0.01}$	$1.35^{+0.02}_{-0.02}$	$1.37^{+0.01}_{-0.01}$	$1.33^{+0.04}_{-0.04}$	$1.35^{+0.02}_{-0.02}$	$1.36^{+0.01}_{-0.01}$
H_2O	[dex]	$-1.83^{+0.03}_{-0.03}$	$-1.80^{+0.03}_{-0.03}$	$-1.74^{+0.05}_{-0.05}$	$-1.52^{+0.03}_{-0.03}$	$-4.05^{+0.06}_{-0.07}$	$-3.39^{+0.07}_{-0.08}$	$-0.91^{+0.06}_{-0.06}$
CO	[dex]	$-1.67^{+0.11}_{-0.11}$	$-1.70^{+0.08}_{-0.08}$	$-1.54^{+0.34}_{-0.36}$	$-1.68^{+0.19}_{-0.19}$	$-0.99^{+0.17}_{-0.20}$	$-3.21^{+0.13}_{-0.14}$	$-1.24^{+0.55}_{-0.96}$
CH_4	[dex]	$-7.32^{+1.78}_{-1.81}$	$-4.16^{+0.05}_{-0.06}$	$-7.41^{+1.78}_{-1.72}$	$-6.39^{+1.98}_{-2.47}$	$-3.05^{+0.04}_{-0.04}$	$-7.71^{+1.61}_{-1.56}$	$-6.98^{+2.14}_{-1.98}$
CO_2	[dex]	$-6.57^{+2.36}_{-2.25}$	$-6.34^{+2.44}_{-2.51}$	$-6.54^{+2.45}_{-2.29}$	$-6.22^{+2.50}_{-2.49}$	$-7.41^{+1.92}_{-1.77}$	$-7.31^{+1.92}_{-1.80}$	$-6.24^{+2.63}_{-2.42}$
H_2S	[dex]	$-5.92^{+2.70}_{-2.71}$	$-4.60^{+2.28}_{-3.73}$	$-5.34^{+3.20}_{-3.13}$	$-5.03^{+3.14}_{-3.29}$	$-6.81^{+2.21}_{-2.13}$	$-6.70^{+2.22}_{-2.25}$	$-5.30^{+2.97}_{-3.05}$
FeH	[dex]	$-7.95^{+0.97}_{-1.40}$	$-9.09^{+0.62}_{-0.60}$	$-6.66^{+1.09}_{-2.16}$	$-7.90^{+1.00}_{-1.40}$	$-6.36^{+0.16}_{-0.17}$	$-6.85^{+0.14}_{-0.15}$	$-8.04^{+0.99}_{-1.23}$
TiO	[dex]	$-7.82^{+1.53}_{-1.47}$	$-7.41^{+1.30}_{-1.70}$	$-4.65^{+0.43}_{-2.25}$	$-7.67^{+1.41}_{-1.52}$	$-7.72^{+1.40}_{-1.55}$	$-7.34^{+0.99}_{-1.80}$	$-7.20^{+1.81}_{-1.84}$
K	[dex]	$-4.06^{+0.18}_{-0.38}$	$-7.02^{+1.51}_{-2.04}$	$-3.48^{+0.59}_{-2.47}$	$-4.17^{+0.28}_{-1.24}$	$-4.58^{+0.15}_{-0.15}$	$-5.61^{+0.38}_{-1.10}$	$-6.34^{+1.97}_{-2.40}$
VO	[dex]	$-8.04^{+0.77}_{-1.25}$	$-9.08^{+0.67}_{-0.61}$	$-7.56^{+1.15}_{-1.59}$	$-8.16^{+0.83}_{-1.18}$	$-8.88^{+0.78}_{-0.76}$	$-8.98^{+0.70}_{-0.66}$	$-8.44^{+0.95}_{-1.00}$
C/O	[dex]	$0.48^{+0.06}_{-0.06}$	$0.45^{+0.04}_{-0.04}$	$0.50^{+0.17}_{-0.18}$	$0.31^{+0.09}_{-0.08}$	$1.01^{+0.01}_{-0.00}$	$0.50^{+0.06}_{-0.06}$	$0.23^{+0.26}_{-0.19}$
Fe/H	[dex]	$0.56^{+0.09}_{-0.08}$	$0.56^{+0.08}_{-0.07}$	$0.71^{+0.32}_{-0.23}$	$0.72^{+0.16}_{-0.11}$	$1.06^{+0.20}_{-0.21}$	$-1.00^{+0.11}_{-0.11}$	$1.32^{+0.41}_{-0.24}$

Table F.4: List of posterior median and 68th percentile credible interval for the retrievals based on both MRCCS and LRSP data.

Parameter		1) Baseline	2) Low T	3) High T	4) Low C/O	5) High C/O	6) Low Fe/H	7) High Fe/H
T_{equ}	[K]	1743^{+9}_{-9}	1403^{+6}_{-6}	2010^{+15}_{-14}	1739^{+9}_{-9}	1781^{+23}_{-23}	1753^{+19}_{-18}	1736^{+7}_{-8}
$\log(g/g_0)$	[dex]	$4.40^{+0.05}_{-0.05}$	$4.33^{+0.07}_{-0.08}$	$4.42^{+0.11}_{-0.13}$	$4.46^{+0.08}_{-0.09}$	$4.58^{+0.03}_{-0.03}$	$4.32^{+0.11}_{-0.13}$	$4.54^{+0.15}_{-0.16}$
R	[R_J]	$1.36^{+0.02}_{-0.02}$	$1.36^{+0.01}_{-0.01}$	$1.35^{+0.02}_{-0.02}$	$1.37^{+0.01}_{-0.01}$	$1.34^{+0.03}_{-0.03}$	$1.34^{+0.02}_{-0.02}$	$1.37^{+0.01}_{-0.01}$
H_2O	[dex]	$-1.83^{+0.03}_{-0.03}$	$-1.80^{+0.03}_{-0.03}$	$-1.77^{+0.04}_{-0.04}$	$-1.52^{+0.03}_{-0.03}$	$-4.10^{+0.07}_{-0.07}$	$-3.38^{+0.07}_{-0.07}$	$-0.91^{+0.04}_{-0.04}$
CO	[dex]	$-1.67^{+0.11}_{-0.11}$	$-1.68^{+0.08}_{-0.08}$	$-1.74^{+0.25}_{-0.24}$	$-1.73^{+0.15}_{-0.17}$	$-1.08^{+0.14}_{-0.16}$	$-3.29^{+0.12}_{-0.12}$	$-0.97^{+0.27}_{-0.29}$
CH_4	[dex]	$-7.47^{+1.79}_{-1.67}$	$-4.14^{+0.06}_{-0.06}$	$-7.49^{+1.84}_{-1.66}$	$-6.89^{+1.94}_{-2.08}$	$-3.09^{+0.03}_{-0.03}$	$-7.46^{+1.67}_{-1.71}$	$-7.12^{+2.01}_{-1.90}$
CO_2	[dex]	$-6.72^{+2.22}_{-2.25}$	$-6.66^{+2.33}_{-2.21}$	$-6.55^{+2.25}_{-2.31}$	$-6.56^{+2.41}_{-2.22}$	$-7.37^{+1.82}_{-1.80}$	$-7.41^{+1.85}_{-1.72}$	$-6.29^{+2.62}_{-2.46}$
H_2S	[dex]	$-5.26^{+3.03}_{-3.26}$	$-2.28^{+0.37}_{-2.97}$	$-3.33^{+2.20}_{-4.38}$	$-1.92^{+0.94}_{-4.97}$	$-6.61^{+2.24}_{-2.35}$	$-6.34^{+2.46}_{-2.48}$	$-2.44^{+1.78}_{-4.69}$
FeH	[dex]	$-7.86^{+0.92}_{-1.41}$	$-9.09^{+0.64}_{-0.60}$	$-7.00^{+1.27}_{-1.92}$	$-8.03^{+1.03}_{-1.30}$	$-6.43^{+0.16}_{-0.17}$	$-6.80^{+0.13}_{-0.14}$	$-8.07^{+0.98}_{-1.20}$
TiO	[dex]	$-7.80^{+1.50}_{-1.47}$	$-7.30^{+1.32}_{-1.84}$	$-4.80^{+0.54}_{-2.92}$	$-7.74^{+1.56}_{-1.51}$	$-7.83^{+1.44}_{-1.46}$	$-7.38^{+1.06}_{-1.79}$	$-7.25^{+1.73}_{-1.82}$
K	[dex]	$-4.08^{+0.18}_{-0.41}$	$-6.71^{+1.30}_{-2.19}$	$-3.20^{+0.37}_{-0.77}$	$-4.17^{+0.22}_{-0.82}$	$-4.61^{+0.14}_{-0.15}$	$-5.51^{+0.35}_{-0.85}$	$-6.55^{+2.06}_{-2.34}$
VO	[dex]	$-7.97^{+0.72}_{-1.35}$	$-9.05^{+0.67}_{-0.64}$	$-7.88^{+1.28}_{-1.37}$	$-8.28^{+0.92}_{-1.15}$	$-8.90^{+0.80}_{-0.76}$	$-9.01^{+0.72}_{-0.67}$	$-8.42^{+0.95}_{-1.02}$
C/O	[dex]	$0.48^{+0.05}_{-0.05}$	$0.46^{+0.04}_{-0.04}$	$0.41^{+0.13}_{-0.11}$	$0.29^{+0.07}_{-0.07}$	$1.02^{+0.01}_{-0.00}$	$0.45^{+0.05}_{-0.05}$	$0.36^{+0.14}_{-0.13}$
Fe/H	[dex]	$0.58^{+0.11}_{-0.08}$	$0.63^{+0.09}_{-0.08}$	$0.65^{+0.45}_{-0.19}$	$0.81^{+0.42}_{-0.17}$	$0.96^{+0.16}_{-0.17}$	$-1.03^{+0.11}_{-0.10}$	$1.57^{+0.40}_{-0.24}$

Table F.5: List of posterior median and 68th percentile credible interval for our retrieval excluding the synthetic VLT/GRAVITY data (cf. Sect. 4.1).

Parameter		MRCCS	LRSP	MRCCS+LRSP
T_{equ}	[K]	1900^{+126}_{-115}	1757^{+17}_{-17}	1748^{+15}_{-15}
$\log(g/g_0)$	[dex]	$5.55^{+0.62}_{-0.58}$	$4.47^{+0.09}_{-0.13}$	$4.43^{+0.09}_{-0.09}$
R	[R _J]	$2.84^{+2.10}_{-0.87}$	$1.34^{+0.02}_{-0.02}$	$1.35^{+0.02}_{-0.02}$
H ₂ O	[dex]	$-1.07^{+0.49}_{-0.38}$	$-1.83^{+0.10}_{-0.09}$	$-1.87^{+0.07}_{-0.07}$
CO	[dex]	$-1.21^{+0.45}_{-0.42}$	$-4.99^{+2.93}_{-3.19}$	$-1.85^{+0.32}_{-0.31}$
CH ₄	[dex]	$-5.74^{+2.44}_{-2.83}$	$-7.53^{+1.93}_{-1.56}$	$-7.37^{+1.84}_{-1.76}$
CO ₂	[dex]	$-6.17^{+2.48}_{-2.33}$	$-6.70^{+2.10}_{-2.13}$	$-6.79^{+2.15}_{-2.11}$
H ₂ S	[dex]	$-3.37^{+2.57}_{-4.21}$	$-5.58^{+2.91}_{-2.86}$	$-2.46^{+1.09}_{-4.72}$
FeH	[dex]	$-5.81^{+2.72}_{-2.64}$	$-8.10^{+1.03}_{-1.22}$	$-8.14^{+1.06}_{-1.23}$
TiO	[dex]	$-5.87^{+2.96}_{-2.64}$	$-7.80^{+1.45}_{-1.42}$	$-7.81^{+1.46}_{-1.43}$
K	[dex]	$-3.90^{+2.29}_{-3.92}$	$-4.03^{+0.23}_{-0.28}$	$-4.05^{+0.21}_{-0.28}$
VO	[dex]	$-5.41^{+2.96}_{-2.98}$	$-8.13^{+0.81}_{-1.17}$	$-8.19^{+0.84}_{-1.15}$
C/O	[dex]	$0.30^{+0.14}_{-0.11}$	$0.00^{+0.26}_{-0.00}$	$0.41^{+0.17}_{-0.16}$
Fe/H	[dex]	$1.54^{+0.57}_{-0.50}$	$0.20^{+0.37}_{-0.12}$	$0.61^{+0.30}_{-0.23}$