# ATMOSPHERIX: II- Characterising exoplanet atmospheres through transmission spectroscopy with SPIRou

Florian Debras, Baptiste Klein, Jean-François Donati, Thea Hood, Claire Moutou, Andres Carmona, Benjamin Charnay, Bruno Bézard, Pascal Fouqué, Adrien Masson, et al.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **527,** 566–582 (2024) https://doi.org/10.1093/mnras/stad2608
Advance Access publication 2023 September 14

# ATMOSPHERIX: II- Characterizing exoplanet atmospheres through transmission spectroscopy with SPIRou

Florian Debras [1]★ Baptiste Klein [1,2]★ Jean-François Donati [1] Thea Hood [1] Claire Moutou,[1] Andres Carmona,[3] Benjamin Charnay,[4] Bruno Bézard,[4] Pascal Fouqué,[1] Adrien Masson,[4] Sandrine Vinatier,[4] Clément Baruteau [1] Isabelle Boisse,[5] Xavier Bonfils,[3] Andrea Chiavassa,[6] Xavier Delfosse,[3] Guillaume Hebrard,[7,8] Jérémy Leconte,[9] Eder Martioli,[7,10] Merwan Ould-elkhim,[1] Vivien Parmentier,[2,6] Pascal Petit,[1] William Pluriel,[11] Franck Selsis,[9] Lucas Teinturier,[4,12] Pascal Tremblin [12] Martin Turbet,[9,13] Olivia Venot [14] and Aurélien Wyttenbach[3]

[1]*IRAP, Université de Toulouse, CNRS, UPS, F-31400 Toulouse, France*
[2]*Department of Physics, University of Oxford, OX13RH Oxford, UK*
[3]*Université Grenoble Alpes, CNRS, IPAG, F-38000 Grenoble, France*
[4]*LESIA, Observatoire de Paris, Université PSL, Sorbonne Université, Université Paris Cité, CNRS, 5 place Jules Janssen, F-92195 Meudon, France*
[5]*Aix Marseille Universite, CNRS, Laboratoire d'Astrophysique de Marseille UMR 7326, F-13388 Marseille, France*
[6]*Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Lagrange, CS 34229, F-06000 Nice, France*
[7]*Institut d'astrophysique de Paris, UMR7095 CNRS, Université Pierre & Marie Curie, 98bis boulevard Arago, F-75014 Paris, France*
[8]*Observatoire de Haute-Provence, CNRS, Université d'Aix-Marseille, F-04870 Saint-Michel-l'Observatoire, France*
[9]*Laboratoire d'astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, allée Geoffroy Saint-Hilaire, F-33615 Pessac, France*
[10]*Laboratório Nacional de Astrofísica, Rua EstadosUnidos 154, 37504-364 Itajubá - MG, Brazil*
[11]*Observatoire astronomique de l'Université de Genève, chemin Pegasi 51, CH-1290 Versoix, Switzerland*
[12]*Universite Paris-Saclay, UVSQ, CNRS, CEA, Maison de la Simulation, F-91191 Gif-sur-Yvette, France*
[13]*Laboratoire de Météorologie Dynamique/IPSL, CNRS, Sorbonne Université, Ecole Normale Supérieure, Université PSL, Ecole Polytechnique, Institut Polytechnique de Paris, F-75005 Paris, France*
[14]*Université de Paris Cité and Univ Paris Est Creteil, CNRS, LISA, F-75013 Paris, France*

Accepted 2023 July 27. Received 2023 July 27; in original form 2023 March 17

## ABSTRACT

In a companion paper, we introduced a publicly available pipeline to characterize exoplanet atmospheres through high-resolution spectroscopy. In this paper, we use this pipeline to study the biases and degeneracies that arise in atmospheric characterization of exoplanets in near-infrared ground-based transmission spectroscopy. We inject synthetic planetary transits into sequences of SPIRou spectra of the well known M dwarf star Gl 15 A, and study the effects of different assumptions on the retrieval. We focus on (i) mass and radius uncertainties, (ii) non-isothermal vertical profiles, and (iii) identification and retrieval of multiple species. We show that the uncertainties on mass and radius should be accounted for in retrievals and that depth-dependent temperature information can be derived from high-resolution transmission spectroscopy data. Finally, we discuss the impact of selecting wavelength orders in the retrieval and the issues that arise when trying to identify a single species in a multispecies atmospheric model. This analysis allows us to understand better the results obtained through transmission spectroscopy and their limitations in preparation to the analysis of actual SPIRou data.

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

## 1 INTRODUCTION

In a companion paper (Klein et al., submitted to MNRAS, hereafter named paper I), we have introduced our publicly available pipeline for the analysis of high-resolution spectroscopy (HRS) data of exoplanet atmospheres. This pipeline was developed in the framework of the ATMOSPHERIX consortium, a gathering of observers and theoreticians created to optimize the study of ground-based HRS for exoplanet atmospheres at the French level. We have shown the validity and robustness of this pipeline for single-component isothermal planetary atmospheres. However, we know this is a crude simplification as more and more molecular species are discovered in exoplanet atmospheres (see review in Guillot et al. 2022) and departures from vertically isothermal atmospheres are also commonly found thanks to stronger temperature constraints (e.g. Haynes et al. 2015; Gibson et al. 2020).

★ E-mail: florian.debras@irap.omp.eu (FD); baptiste.klein@physics.ox.ac.uk (BK)

**Table 1.** Physical parameters for Gl 15 A, HD 189733 b, and for the simulated hot Jupiter used in the study. When taken from the literature, the reference of each parameter is indicated in the right-hand column†.

| Stellar parameters | Gl 15 A | |
|---|---|---|
| | Value | Reference |
| Mass ($M_\odot$) | $0.400 \pm 0.008$ | Ro21 |
| Radius ($R_\odot$) | $0.375 \pm 0.007$ | Ro21 |
| Effective temperature (K) | $3742 \pm 30$ | Ro21 |
| *H* magnitude | $4.476 \pm 0.2$ | Cu03 |
| Systemic velocity [km.s$^{-1}$] | $11.73 \pm 0.0001$ | Fo18 |
| Limb Darkening (Quadratic) | 0.0156, 0.313 | Cl11 |
| **Planet parameters** | | |
| | HD 189 733 b | Synthetic planet | Reference |
| Transit depth (per cent) | $2.2 \pm 0.1$ | 2.2 | Ad19 |
| Radius ($R_J$) | $1.142 \pm 0.04$ | 0.55 | – |
| Mass ($M_J$) | $1.13 \pm 0.05$ | 0.572 | – |
| g (m.s$^{-2}$) | $22.45 \pm 1.5$ | 49.18 | – |
| Orbital period (d) | $2.218577 \pm 0.000001$ | 2.218577 | – |
| Mid transit time (BJD TBD) | $2458334.990899 \pm 0.0007$ | 2459130.8962180 | Ad19 |
| Inclination (°) | $85.7 \pm 0.1$ | 90.0 | – |
| Eccentricity | 0.0 | 0.0 | – |
| Equilibrium temperature (K) | $1209 \pm 11$ | 1209 | - |
| Orbital semi-amplitude (km.s$^{-1}$) | $151.2 \pm 4.5$ | 120.0 | Ad19 |
| Transit duration (h) | $1.84 \pm 0.04$ | 1.84 | – |

†To gain some space in the table, we use aliases for the references. Ro21, Cu03, Fo18, Cl11, and Ad19 stand, respectively for Rosenthal et al. (2021), Cutri et al. (2003), Fouqué et al. (2018), Claret & Bloemen (2011), and Addison et al. (2019).

More complex models would therefore be needed to be representative of actual observations but the more complex the model, the more degenerate the problem. It is therefore an important task to understand the sources of degeneracies in atmospheric retrievals in order to provide the most reliable parameter estimates. Such degeneracies have already been studied in low-resolution spectroscopy (LRS) for more than 20 yr (see e.g. Brown 2001 and the references in the introduction of Welbanks & Madhusudhan 2019), but are less extensively studied in HRS, particularly in the infrared. Fisher & Heng (2019) have studied the information that can be obtained through the sodium doublet in the visible, concluding that HRS alone is not enough to determine appropriately the pressure that are probed by the sodium lines. The combination with LRS in Pino et al. (2018) might allow to resolve some of these questions. We obtained similar conclusions when including clouds in our companion paper, where the loss of the continuum by HRS exacerbated a degeneracy between cloud top pressure and water content. Clouds, in general, are a major source of work to understand plausible degeneracies in the spectra, both at high and low resolution (see e.g. Kitzmann & Heng (2018); Barstow (2020)). Inclusion of multidimensional effects further complicates this picture (Line & Parmentier 2016; Pluriel et al. 2020; Welbanks & Madhusudhan 2022).

In this paper, we therefore focus on a few degeneracies and potential biases that are inherent to HRS with application on synthetic SPIRou transit data. We test three cases: uncertainties in the planet's mass and radius, non-isothermal vertical structures, and models with multiple molecular species with comparable mixing ratios. We first recall the process of data generation and reduction in Section 2. In Section 3, we then present our test cases and the results of template matching and Bayesian retrieval on the atmospheric parameters. This leads us to discuss how to optimize the detection and the ways forward in Section 4, before concluding in Section 5.

## 2 DATA GENERATION AND ANALYSIS

The generation of the synthetic spectra and their reduction are extensively described in paper I. They are very shortly reminded here.

### 2.1 Creation and reduction of synthetic data

We simulate the observations of a planetary transit with a near-infrared (nIR) high-resolution spectrograph. This is done by injecting a synthetic planet atmosphere spectrum into a sequence of 192 spectra of the bright M dwarf Gl 15A collected during 5 h with SPIRou in 2020 October and divided in two sets (see Table 1 and Fig. 1). Gl 15 A is chosen both because we have many spectra of it with SPIRou and because it is a well-studied star in radial velocity. If its system contains a short-period Earth-like planet, it is shown that no Jovian planet orbits this star for periods of less than 10 yr (see for example fig. 2 of Pinamonti et al. (2022).The observations sample the [0.9,2.5] $\mu$m wavelength range in 49 diffraction orders with a typical pixel size of 2.28 km. s$^{-1}$ and a spectral resolution of $\sim$70 000. Data are reduced through the APERO pipeline (Cook et al. 2022) that calibrates the data in wavelength and applies state-of-the art telluric correction.

The synthetic planet is based on the classical hot Jupiter HD 189733 b (Bouchy et al. 2005) injected on a circular orbit and we decided to conserve four planetary and transit parameters to obtain data with consistent expected level of detection: (i) the transit depth, (ii) the transit duration, (iii) the ratio between the stellar radius and the atmospheric scale height, and (iv) the atmospheric temperature. The injected planet spectra are all shifted by 30 km.s$^{-1}$ so that stellar molecular features and planet atmosphere absorption lines are not mixed.

Once the planet is injected, we remove the stellar spectra and remaining telluric contaminations by dividing each observed spectrum within each order by a median spectrum. This step is performed successively in the Earth rest and stellar rest frames, and an additional

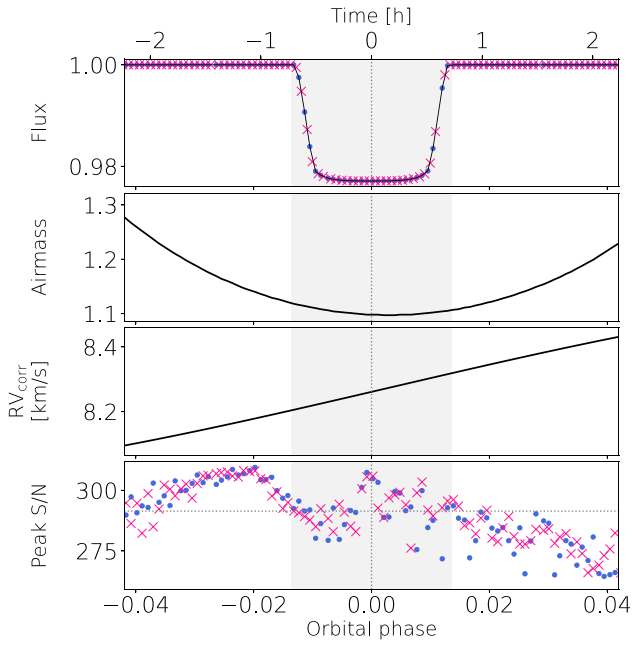


**Figure 1.** Variations of photometric flux (top panel), airmass (panel 2), Geocentric-to-stellar rest frame RV correction (panel 3), and peak SNR per velocity bin during the two simulated transits of the HD 189733 b analog. On panels 1 and 4, the two different transits are respectively shown as blue dots and pink crosses. The vertical grey band indicates the primary transit of the simulated planet. The horizontal grey dashed line on the bottom panel indicates the average value of the peak S/N for the observed spectra.

high-pass filter is applied to the residual spectra, in order to correct for low-frequency variations in the continuum. Outliers are flagged and masked using a sigma-clipping procedure, and the residual time-varying telluric flux is corrected with an airmass detrending in the log-flux space. We then apply a principal component analysis (PCA) to get rid of the remaining correlated noise. An auto-encoder can be applied instead, although it is not yet mature for parameter retrieval and is limited to detection of molecular species as we cannot reproduce its effect efficiently on the models. Diffraction orders 57 to 54 (i.e. ∼1300 to ∼1500 nm) and 42 to 40 (i.e. ∼1800 to ∼2000 nm), located within nIR water absorption bands, are discarded.

### 2.2 Uncovering the planetary signature

Once the reduced data have gone through the PCA or auto-encoder step, the planetary signal is still largely buried under the noise. We either perform a template matching method between theoretical models and the reduced data, or a statistical exploration of the parameter space through nested sampling using the PYTHON module PYMULTINEST (Feroz & Hobson 2008; Feroz, Hobson & Bridges 2009; Buchner et al. 2014; Feroz et al. 2019). The models are created with PETITRADTRANS (Mollière et al. 2019) which provides the planerary radius as a function of wavelength. They are next trasnformed into an absorption by calculating the ratio of planetary to stellar radius squared, the so-called transit depth. The correlation function (CCF) calculated for different planet velocimetric semi-amplitude ($K_p$) and systemic Doppler shift ($V_{sys}$) writes:

$$\mathrm{CCF} = \sum_i \frac{d_i m_i}{\sigma_i^2}, \tag{1}$$

where $m_i$, $d_i$, and $\sigma_i$ are respectively the flux in the model spectrum, the observed flux and the flux uncertainty at pixel $i$ (corresponding to time $t$ and wavelength $\lambda$: $d_i = d(t, \lambda)$.). This function is calculated and summed for every SPIRou order. More precisely:

$$\sigma_i^2 = \sigma^2(t, \lambda) = \frac{\sum_t \left(d(t, \lambda) - \overline{d(\lambda)}\right)^2}{N_{\mathrm{spectra}}} \frac{\overline{\mathrm{SNR}}}{\mathrm{SNR}(t)}, \tag{2}$$

where the bar denotes a time average and $N_{\mathrm{spectra}}$ is the number of spectra. The barred SNR values are calculated for each order. In order to convert correlation value to significance of detection, we perform as is frequently done in literature, i.e. divide by the standard deviation of the correlation map away from the planetary signal.

The nested sampling relies on the calculation of a likelihood $L$, defined following the frameworks of Brogi & Line (2019) and Gibson et al. (2020):

$$\mathcal{L} = \prod_{i=0}^{N} \frac{1}{\sqrt{2\pi\sigma_i}} \exp\left\{-\frac{[m_i - a d_i]^2}{b\sigma_i^2}\right\}, \tag{3}$$

where $a$ and $b$ are scaling factors to account for incorrect modelling, a is set to 1 in this paper, and b is optimized globally as in Gibson et al. (2020).

In order to account for the fact that the observed planet atmosphere spectrum is affected by the data reduction procedure, we degrade the model before comparing it to the data, following the procedure detailed in Gibson et al. (2022). First, we create the projector P on the vector space defined by our subset of PCA eigenvectors obtained in the data analysis. At each iteration of the nested sampling process, we then compute a sequence of model spectra, called M, matching the wavelength and time grids of the observations, and Doppler shifted are the values of $K_p$ and $V_{sys}$. We finally subtract the projection by P of M. Our final, degraded sequence of theoretical model M arcmin is:

$$M' = \exp\left(\log M - P \log M\right). \tag{4}$$

As we show in our companion paper, this step is crucial not to bias the retrieved planet parameters.

Finally, as explained in our companion paper, we have the possibility to include a proxy for planetary rotation and winds. We simply convolve our 1D atmospheric models by a rotation kernel that considers the latitudinal speed variation due to rotation. This kernel can be modulated to take into account any latitudinal wind shape, such as superrotation. We expressed this kernel as two convolution products so that it is very efficient numerically speaking and allows one to retrieve the planet rotation rate in a parameter space exploration algorithm.

## 3 APPLICATION TO SIMULATED DATA

### 3.1 Uncertainties in mass and radius

Our first test was to keep a simple, isothermal model containing only water as in paper I but to change the radius $R$ and gravitational acceleration $g$ (proxy for mass $M$ here as $g \propto M/R^2$) of the planet in the retrieval compared to the injected planet. We tested three different cases: in Model A, gravity, and radius are treated as free parameters in the retrieval. Model B imposes a wrong gravity and a good radius and we look at the effect on the temperature and water composition. Model C imposes a wrong gravity but allows only the radius to change. The results and their comparison with the isothermal model of Paper 1 are provided in Table 2. Through the section, $R$ and $g$ are expressed in planetary units.

**Table 2.** Summary of the retrieved parameters when varying mass and radius. The radius is in true planetary radius ($R_p$), gravity in planetary gravity ($g_p$), and water mass mixing ratio (MMR) in log. For each parameter, the first column is the model true value. The second column represent input values when only one number is provided or uniform prior range when the parameter is included in the retrieval. The third column is the retrieved values with $1\sigma$ uncertainty. The description of the models is given in the text. P I is paper one, and M x means Model x.

| | R ($R_p$) | | | g ($g_p$) | | | $T_{eq}$ (K) | | | Water MMR | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | True | Input | Retrieved | True | Input | Retrieved | True | Input | Retrieved | True | Input | Retrieved |
| P I | 1.0 | 1.0 | – | 1.0 | 1.0 | – | 900 | [200,2000] | $1013 \pm 117$ | −2.11 | [−8,−1] | $-2.49 \pm 0.41$ |
| M A | 1.0 | [0.6,1.4] | $0.94 \pm 0.17$ | 1.0 | [0.6,1.4] | $1.04 \pm 0.18$ | 900 | [200,2000] | $1026 \pm 124$ | −2.11 | [−8,−1] | $-2.57 \pm 0.44$ |
| M B | 1.0 | 1.0 | – | 1.0 | 0.75 | – | 900 | [200,2000] | $967 \pm 119$ | −2.11 | [−8,−1] | $-3.05 \pm 0.37$ |
| M C | 1.0 | [0.5,1.5] | $0.725 \pm 0.13$ | 1.0 | 0.75 | – | 900 | [200,2000] | $1003 \pm 130$ | −2.11 | [−8,−1] | $-2.61 \pm 0.44$ |

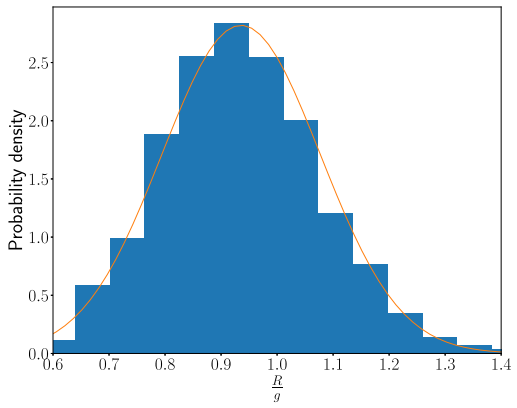


**Figure 2.** Posterior distribution of $\frac{R}{g}$ from Model A in planetary units. The injected ratio is 1. The orange line is a Gaussian with mean 0.936 and standard deviation 0.14.

Regarding Model A, gravity, and radius are recovered within a few per cents of error. There is a large degeneracy between both parameters which translates into a smaller error bar on the normalized $R/g$ ratio: the posterior distribution of $R/g$ shown in Fig. 2 shows that the distribution is well matched with a Gaussian of standard deviation 0.14. The uncertainties on other physical parameters are comparable with paper I and the retrieval is globally comparable with paper I.

When we simply imposed a lower gravity in the nested sampling algorithm compared to the injected model in Model B, the retrieval of composition, and temperature gave lower values for MMR and temperature. The water MMR is more affected, with the retrieved value $3\sigma$ away from the injected value. However, contrary to the results of paper I, the input parameters are outside of the posterior distribution in the temperature-composition joint posterior (figure not shown): there is an actual bias that was not present in our analysis with correct mass and radius.

Finally, Model C shows that varying only gravity or radius allows to obtain comparable results with Model 1 for water and temperature, with retrieved $R/g$ close to 1.

The results of this section show that, within our high-SNR framework, we are sensitive to more than the sole amplitude of molecular lines. Indeed, observationnaly speaking we are sensitive to the variations of the transit depth with wavelength hence the important quantity is:

$$O \sim \frac{H R_p}{R_s^2}, \tag{5}$$

where $H$ is the typical scale height of the atmosphere, $R_p$ the planetary radius, and $R_s$ the stellar radius. For the simple case of an isothermal atmosphere:

$$H = \frac{\mathcal{R}T}{Mg}, \tag{6}$$

where $\mathcal{R}$ is the ideal gas constant, $T$ the temperature, $M$ the molecular mass, and $g$ the gravitational acceleration. If the amplitude was the only concern, we would find that the mass and radius of the planet would be correlated with temperature as the code tries to match the $HR_p$ value to that of the injected planet. The fact that temperature and composition uncertainties remains globally insensitive to mass and radius as long as $R/g \approx 1$ shows that we are not only matching amplitudes, but the shape of the lines as well which are affected by temperature and composition only.

This analysis points towards the fact that the uncertainty in mass and radius should be included in the retrieval of atmospheric parameters rather than chosen as constants. It allows to avoid biases and, at least in our simple case, does not degrade the obtained atmospheric properties. For optimization purposes, only one of these quantities can be included in the retrieval, remembering that we are only sensitive to the ratio of radius and gravity. This will be particularly stringent for low-mass, distant exoplanets or planets around very active/young stars where the complicated radial velocity signature might lead to large uncertainties in the mass.

### 3.2 Non-vertically isothermal models

In paper I, we only considered vertically constant models in temperature and composition. Here, we test whether we are able to retrieve parameters that vary vertically, focusing on non-isothermal profiles. We implement a vertical temperature profile taken from 3D GCM simulations of HD 189 733 b (Drummond et al. 2018). The temperature structure is averaged at both limbs and used as an input in the 1D `PETITRADTRANS` modelling (see Fig. 3). We created two models with constant water volume mixing ratio (VMR) of $10^{-3}$ and $10^{-5}$, respectively.

For our retrieval, we have tested four temperature prescription: in case 'Isotherm' we assumed an isothermal profile. Case 'Linear' used a four points temperature profile, where we retrieve the temperature at four different pressures (1, 100, and $10^4$ Pa and 1 bar) and linearly interpolate in log pressure between these points. The temperature at lower pressures than 1 Pa and higher than 1 bar is set as constant. Case 'Lagrange' also used the same four points prescription, but the interpolation was made through a Lagrange polynomial, ensuring a smooth temperature profile. Finally in case 'Guillot', we have used the widely applied two temperature model of Guillot (2010) and retrieved four parameters: the internal temperature, the equilibrium temperature, the infrared opacity and the infrared to visible opacity ratio. As we wanted to focus on the temperature structure, we present results where we have fixed the water composition in the

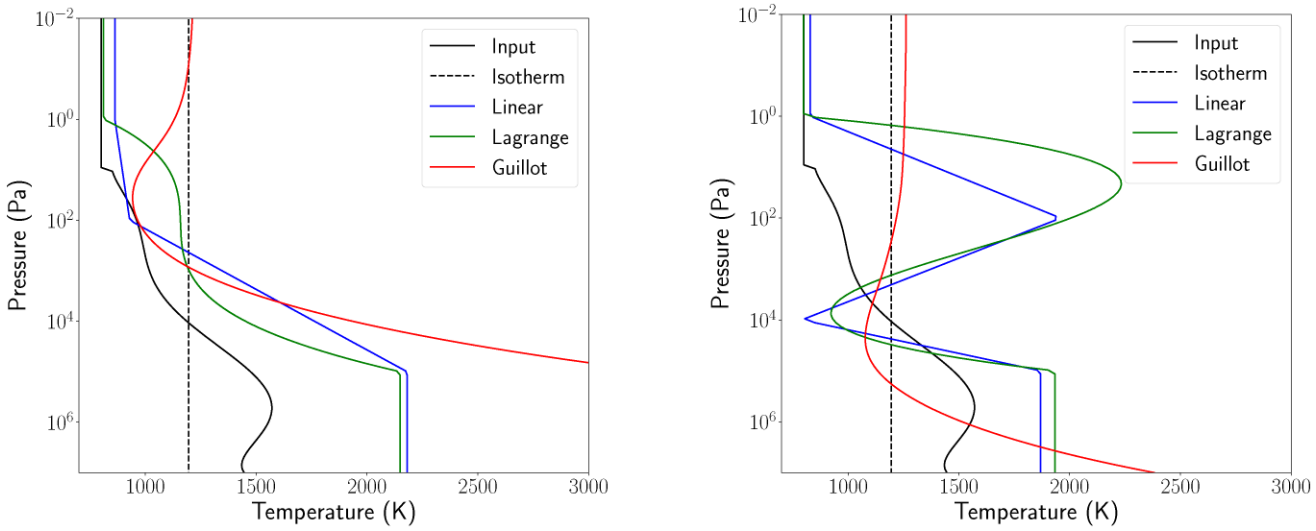


**Figure 3.** Temperature as a function of pressure for the input profile (black) and the mean of the retrieved profiles with the three possible models discuted in the text (colours). Left: water VMR of $10^{-3}$. Right: water VMR of $10^{-5}$.

retrieval to the model composition. When we let this value as a free parameter, we always retrieved the appropriate water composition but the temperature retrieval is worsen due to expected (and already presented in paper 1) degeneracies. The retrieved temperatures profiles are shown in Figs 3 and 4.

We first notice that the temperature profile is poorly constrained: the standard deviation with pressure can easily reach hundreds of Kelvin, which is much more than the ≈115K we obtained in the isothermal cases of paper I. The temperature is on average higher than the injected profile, but this mainly arises from the choice of priors which are not centred around the injected profile. In both cases, as seen in Fig. 4, the deep (≥1 bar) and shallow (≤1 Pa) atmosphere temperatures are just given by the priors: the distribution is close to the uniform prior distribution we chose. However, what we clearly see on Figs 3 and 4 is that the retrievals are sensitive to different regions for the two models: the high (low) water concentration model is more sensitive to the higher (deeper) atmosphere. This was expected as the retrieved radius depends mostly on the regions which contribute the most to the water absorption lines, which correspond to pressures where the water column becomes optically thick (optical depth becoming greater than 1). This roughly corresponds to 100 Pa in the dense water model, and $10^4$ in the other.

Interestingly, although the mean profile is closer to the injected profile around $10^4$ Pa in the low-VMR case, the standard deviation of recovered temperature with pressure is always lower in the high-VMR case. This arises from the lower SNR in the low VMR case: even at the pressures which contains most of the water information, the amplitude of the planetary signal is too low to permit a precise fit to the data. This is consistent with the isothermal retrieval, whose posterior temperature distribution is well matched by a Gaussian with mean and variance 1041 ± 63 K in the high-VMR case, and 1195 ± 143 K in the low-VMR case.

This test therefore shows that, on average, we are poorly sensitive to the temperature structure and that we are primarily probing pressures where the optical depth of molecular lines reaches 1. This means that we could potentially obtain a better understanding of the temperature profile by combining the information from different molecules: depending on their density and opacity, they will probe different pressure levels. A retrieval with a unique temperature profile for different molecules might therefore be less informative than trying to retrieve the temperature for different species and estimating where they provide most of their signal.

### 3.3 Recovering a multispecies model

We now consider the case of multiple species, namely $H_2O$, CO, $CH_4$, and $NH_3$. Through this section, unless specified otherwise our synthetic atmosphere always contains these four molecules (in addition to $H_2$ and He which are largely dominant) with isothermal profile and we only vary the VMRs of each individual species.

We created three synthetic transit sequences with three injected models labelled 1, 2, and 3. Model 1 used the MMRs reported in the table 4 of Giacobbe et al. (2021), Model 2 kept the same MMR for water but the other molecules were a factor of 10 lower and Model 3 a factor of 100 lower. The characteristics of the three models are summarized in Table 3 and plotted on Fig. 5, where we show the whole models and two zooms: one where water has low-amplitude absorption lines (around 1640 nm) and one where water absorption is dominant (around 1860 nm).

In Appendix A, we show the cross-correlation maps for the three synthetic models. When we correlated the synthetic data with the injected models in Fig. A1, we recovered ≈5$\sigma$ detection in all cases, with Model 3 having the highest SNR and ratio between maximum positive and minimum negative value of correlation. This is not surprising as wee see on Fig. 5 that the amplitude of the lines is higher for this model, where the lower level of other species impacts less the global shape of the spectrum.

We then correlated the synthetic data with models containing only one of the species. This is usually done in the literature for atmospheric characterization to validate a detection of an individual species even if the atmosphere contains other constituents. Our results of individual detection for each models are gathered in Table 4 When using all of the orders, $CH_4$ was detected (significance larger than 3$\sigma$) for all three models as shown in Fig. A2 for Model 3. $NH_3$ on the other hand is detected in Model 1, only marginally detected (detection around 2$\sigma$) in Model 2 as shown in Fig. A3 and not detected in Model 3. Water is not detected in Model 1, detected in

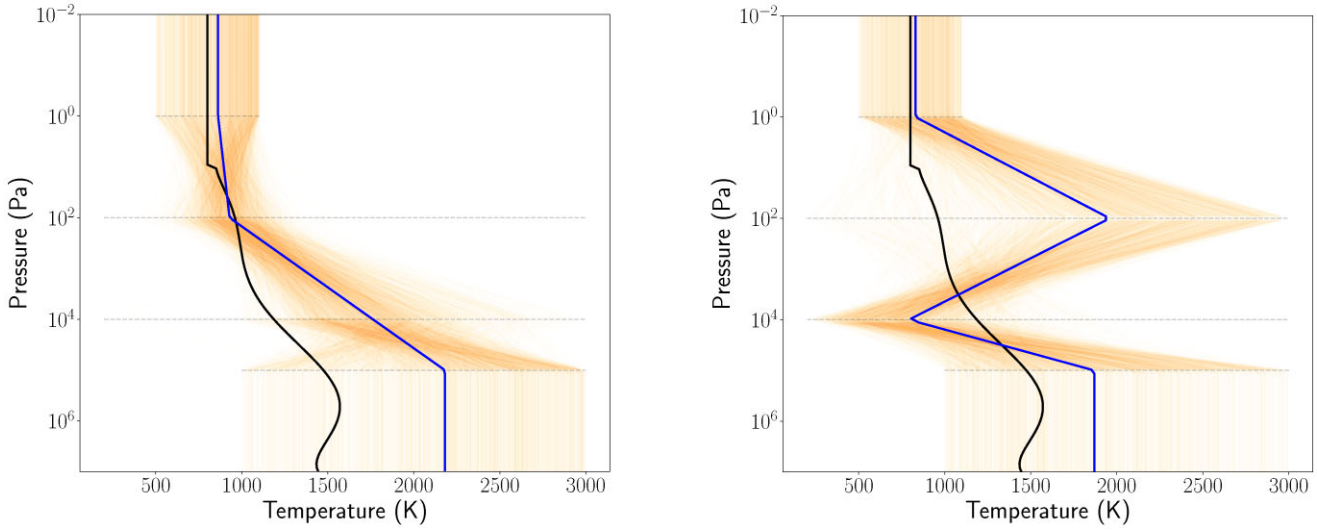


**Figure 4.** Temperature as a function of pressure for the input profile (black) and all of the retrieved profiles from the linearly interpolated four points retrieval, with the mean profile in blue. The grey dashed lines represent the uniform prior ranges at the four pressures. Left: water VMR of $10^{-3}$. Right: water VMR of $10^{-5}$.

**Table 3.** Physical parameters for the multispecies models included used in the nested sampling retrieval. The temperature is in Kelvin and the abundances in MMRs, to be easily comparable with the posterior figures.

| | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| Temperature | 900 | 900 | 900 |
| $\log_{10}(H_2O)$ | −3.05 | −3.05 | −3.05 |
| $\log_{10}(CO)$ | −1.8 | −2.8 | −3.8 |
| $\log_{10}(NH_3)$ | −3.0 | −4.0 | −5.0 |
| $\log_{10}(CH_4)$ | −1.5 | −2.5 | −3.5 |

**Table 4.** Detection of individual species when correlating single-component models with the synthetic data detailed in Table 3. A detection means a SNR superior to 3, whereas a marginal detection is between 2 and 3.

| | $H_2O$ | $NH_3$ | $CH_4$ |
|---|---|---|---|
| Model 1 | No | Yes | Yes |
| Model 2 | Yes | Marginal | Yes |
| Model 3 | Yes | No | Yes |

Model 2 as shown in Fig. A4 and detected over $4\sigma$ in Model 3. Finally, we never detected CO as we explain in the next paragraph. These rather poor results led us to consider selecting orders, as we detail in the next section.

Importantly, $NH_3$ and $H_2O$ in Model 1 and 2 were not robustly detected although, when we performed injection-recovery test with only these molecules at the same VMR they were easily detected (larger than $4\sigma$). This shows that the non-detection of a given individual species does not systematically mean that it is absent from the atmosphere but can simply reflect a mismatch between a complex observed atmosphere spectrum and a too simplistic single-species model.

For CO, we realized that the issue came from the stellar CO which prevents the detection of planetary CO. When we divide by the mean stellar spectra in the data reduction process we affect the planetary lines and hamper the detection. However, we also tested that the presence of CO in the synthetic data only marginally affected the retrieval of other species, due to its limited wavelength range of absorption and well separated absorption lines. We will therefore not consider CO in the rest of this section although it is in the models.

We then tried to retrieve the parameters with our nested sampling algorithm. We show the resulting posterior distributions in Appendix B. Several things can be noted:

(i) The $H_2O$ abundance is poorly constrained in the two first models. This is not surprising as we exclude orders where tellurics are dominant, hence where water has its major impact on the spectrum. It was not an issue in the single species model but becomes problematic when other species are considered with the same VMR as water and the absorption lines of water are reduced in amplitude. In Model 3, we recover results comparable to those of paper I.

(ii) The temperature is always overestimated and leads to large degeneracies with composition. When we fix the temperature, as shown in Fig. B3 compared to Fig. B2, the retrieval of other parameters is largely improved.

(iii) There is some degeneracy between $H_2O$, $CH_4$, and $NH_3$, although we expect HRS to distinguish between molecular lines of different species. This rather counter-intuitive correlation is easily explained: for a given quantity of, say, $H_2O$, increasing the quantity of $CH_4$ or $NH_3$ decreases the line depth by increasing the mean radius of the planet as seen in Fig. 5. The algorithm thus does not differentiate properly between low quantity of $H_2O$ and $CH_4$/$NH_3$ or high quantities of all of them.

(iv) $NH_3$ is not recovered in Model 3 and only an upper limit for its content is obtained. However, we serendipitously removed a few orders where water has the highest signal (around the bands of water at 1.4 and 1.8 microns) and retrieved ammonia in Fig. B5. This retrieval shows two peaks: one at low temperature (close to the injected 900 K) where $NH_3$ is poorly recovered but water and methane are, and one at much larger temperatures (few thousands kelvins) where the fit of the $NH_3$ composition is tightly centred around its injected composition, but at the cost of losing the detection of methane and having a degeneracy between water and temperature to ensure a constant line depth for ammonia. We understand this as the fact that the first peak has a maximum of likelihood from the

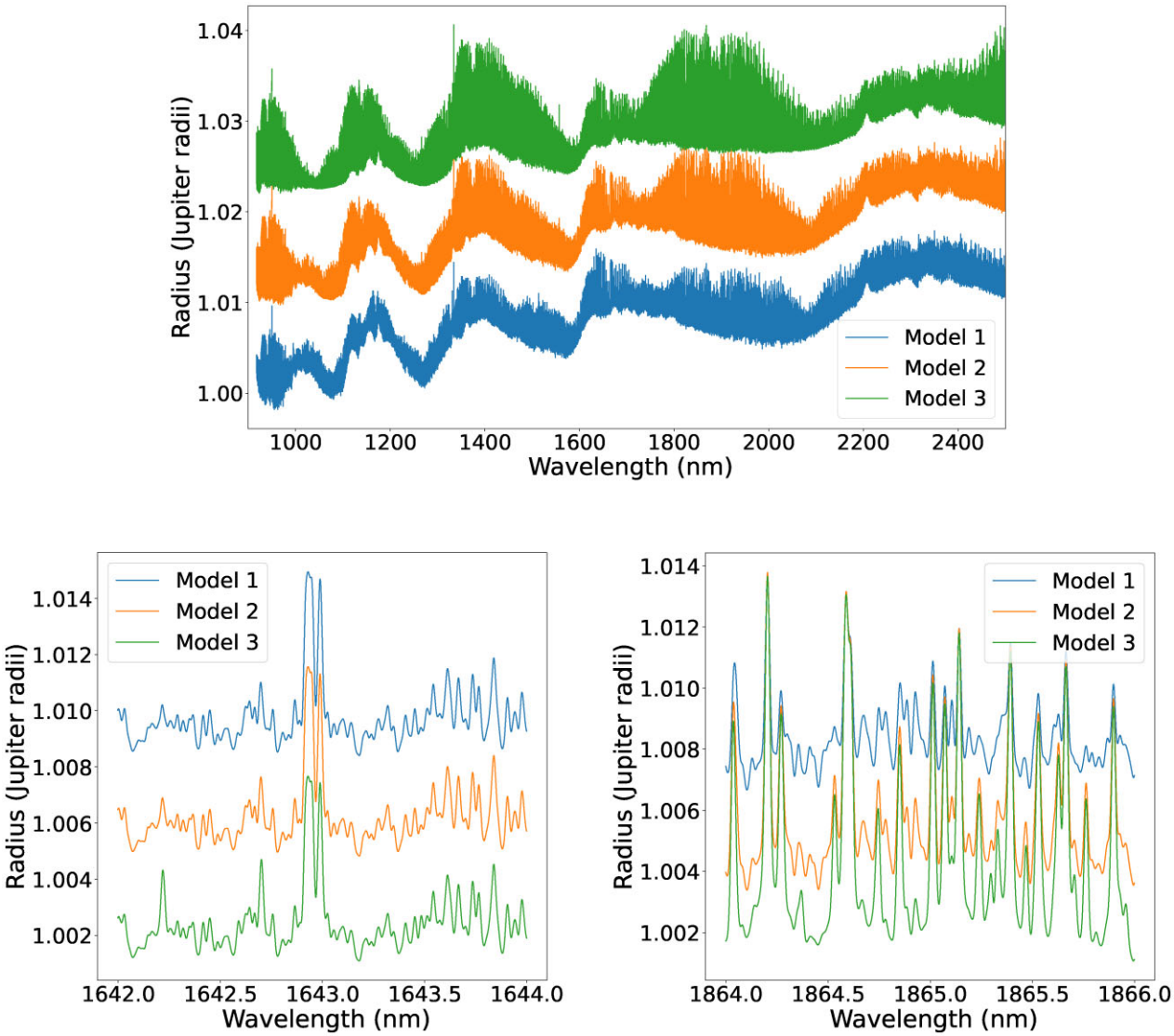


**Figure 5.** Top: transit radius as a function of wavelength over the SPIRou domain for the three models of Table 3. The models have been shifted for visual comparison. Bottom: zoom on two different wavelength ranges. The models have not been shifted in the zooms.

fit of the fewer remaining water lines, whereas the second peak fits perfectly ammonia and provides a secondary maximum. We have verified that, when removing these orders in a pure water model, we do not obtain this second peak and simply recover water at the injected VMR. It is not clear how this result would translate to real planetary observation and whether we could potentially detect ammonia in secondary maxima, but it further confirms the degeneracies between composition and temperature in the amplitude of the lines and that care must be taken from the use of HRS only with simple priors.

(v) In all cases, the typical error bar of log-VMR is 2 dex. It shows that an atmospheric retrieval with transmission spectroscopy is powerful to identify species but does not give a precise value of the composition (and temperature), except for much higher SNRs as in Line et al. (2021) or by coupling with LRS.

If we now use the nested sampling process to retrieve individual species from the four-species synthetic data, we confirm the results obtained with the cross-correlation method. If one species is dominating, we are able to recover it with the nested sampling algorithm but if the three have comparable VMRs, they are not always recovered individually. Additionally, even if one species is dominant, we often recover too low an abundance compared to the injected value. This is to be expected: as we see on Fig. 5, the depth of the absorption lines is reduced by the presence of other species. This translates into a lower recovered value of the VMR compared to the injected one, which is not an error of the algorithm but rather comes from too simple an assumption (that the multicomponent model is equivalent to a combination of single-component models). Hence, low VMRs of given species in planetary atmospheres can simply arise from an erroneous chemical composition.

Finally, we also performed a test in which we aimed to retrieve species that were not included in the model. We used the retrieval with $H_2O$, $CH_4$, and $NH_3$ with an atmosphere model containing only $H_2O$. The MULTINEST algorithm converged towards a low ($\leq 10^{-5}$) but non-zero composition in $CH_4$ and $NH_3$, because of the degeneracies we already mentioned. It therefore shows that the best-fitting might not lead to a real individual detection and care must be taken when analysing only posterior data.

### 3.4 Order selection

As we could not always detect molecules individually when considering all orders, we tried to define a merit function that would select or weight the wavelength range for each molecule. Two methods were tested: (i) we created a model with VMR $= 10^{-4}$ for $H_2O$, $NH_3$, and $CH_4$ and calculated the pearson correlation coefficient with the single species model. For each molecule, we then only selected the orders where this correlation was larger than 0.5 and (ii) we calculated the autocorrelation of the spectra of each individual species order by order and used it as weights in the CCF.

For $CH_4$, the second method slightly improved the detection but only marginally. For water and $NH_3$, there was no reliable improvement by using either of the two methods. The second method works best for water in Model 2 but the first method is best in Model 1, whereas it is the opposite for $NH_3$, and in all cases the improvement is only marginal. We therefore could not rely on these methods to improve our detection limits in the general cases.

However, as we mentioned in the previous section, removing a few orders dominated by water helped detecting and constraining the $NH_3$ composition with the nested sampling algorithm in Model 3. This shows that, although for individual molecules we did not find reliable ways to improve the significance of correlation, order selection can improve the retrieval for multiple species models. This is to be kept in mind when trying to rule out the presence of certain molecules from Bayesian exploration of the parameter space.

Finally, it is interesting to note that our preferred orders for water detection (following method (i) or taking the 15 best orders of methods (ii)) are very different from those of Giacobbe et al. (2021). We actually don't recover water in Model 1 or 2, and only marginally in Model 3 with their wavelength domain. This points out toward either a difference in our two analyses, a much larger signal as they combine five transits of high atmospheric signature or that HD 209 458 b has a much higher water VMR than $CH_4$ and $NH_3$.

## 4 DISCUSSION

### 4.1 Combining transits

Since addressing all sources of uncertainties and degeneracies is out of reach, we have focused on a few cases but have not mentioned the impact of stacking transits on the retrieval. Obviously, adding up many transits helps in identifying the atmospheric absorption by increasing the SNR as long as there is no (or low) variability in the planetary signature. The combination of several transits will be discussed in detail in forthcoming papers of the ATMOSPHERIX consortium with real data (Masson et al., in preparation, Hood et al., in preparation).

### 4.2 Improving the detection

Throughout the two first articles of the ATMOSPHERIX consortium, we have focused on optimising the data reduction process. Further improvement of the data analysis framework will be required to characterise the atmospheres of the most challenging targets of the ATMOSPHERIX sample, either because of their low-amplitude atmosphere signals or due to the host star being too faint and/or active. The community is devoting substantial efforts to enhance the significance of molecular detection and get as much information as possible from the data. The use of an auto-encoder, introduced in paper I, is one of such example. Among the possible improvements, we want to mention the works of Meech et al. (2022) and Rasmussen et al. (2022). Both teams use Gaussian processes to perform a spectrum retrieval and improve the data reduction process.

Other technics have been presented in the literature although they have not yet been applied as systematically as template matching. We notably think about tomography (Watson et al. 2019) which is an interesting prospect to characterize exoplanets. If we were able to retrieve a mean line-profile, Doppler imaging techniques inspired from stellar studies (e.g. Vogt, Penrod & Hatzes 1987) could also be used to study the multidimensional structure of planets. This prospect is particularly interesting in the visible, where the lines have higher SNR, in emission spectroscopy and in the fortchoming era of 30+ meter telescopes.

Globally, the use of HRS to characterize exoplanet atmospheres is less than 15 years old, and there are still lots of possibilities to improve the techniques. Such improvements might mitigate the conclusions of this paper as the detection level will be increased. We still expect degeneracies to be present and important in the process as we exposed them here and we advocate that there is a lack of studies focusing on the inherent degeneracies and limitations of the method thus far.

## 5 CONCLUSION

In this paper, we have extended on the work of paper I, presenting our data analysis pipeline, by stuying different sources of uncertainty and degeneracy inherent to our analysis. We have shown that we are able to retrieve the correct model but that numerous degeneracies can drastically increase the error bars. We have focused on three issues: inaccuracies in the mass and radius, non-vertically isothermal profiles and the retrieval of multiple species. The conclusions of our tests are as follows:

(i) The mass and radius of the planet should be included in the retrieval if they are uncertain, as this leads to a more reliable atmospheric retrieval.

(ii) The vertical temperature distribution of the planet's atmosphere is not easily retrieved as we are mostly sensitive to pressures where the optical depth approaches to 1. However, this also means that different molecules will be sensitive to different pressure levels which might allow one to probe the atmosphere at different depths, and to reconstruct a global temperature profile by combining information of different molecules at different pressures.

(iii) Models with multiple species introduce several degeneracies which can lead to erroneous conclusions: one can identify molecules that are not present or estimate inaccurately their mixing ratios.

(iv) When imposing the temperature, the retrieval is significantly improved.

(v) Although transmission spectroscopy is good at detecting molecules, the $1\sigma$ uncertainty on the VMR can reach up to 2 orders of magnitudes for our typical SNR of 200. Stacking many observations or using independent diagnostics such as LRS is necessary to reduce these uncertainties.

(vi) We did not find a reliable way to weight or select the SPIRou orders in order to improve the molecular detection for single species. We found however than selecting order can improve the retrieval of one species at the cost of a worse retrieval of another one and temperature in multispecies models.

The combination of this paper and paper I gives an overview of the capacity of our pipeline to analyse SPIRou data of exoplanet atmospheres through transmission spectroscopy. They will serve as

a basis for forthcoming papers of the ATMOSPHERIX consortium on real targets, whose studies are ongoing.

## ACKNOWLEDGEMENTS

Based on observations obtained at the Canada-France-Hawaii Telescope (CFHT) which is operated from the summit of Maunakea by the National Research Council of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique of France, and the University of Hawaii. The observations at the Canada–France–Hawaii Telescope were performed with care and respect from the summit of Maunakea which is a significant cultural and historic site. FD thanks the CNRS/INSU Programme National de Planétologie (PNP) and Programme National de Physique Stellaire (PNPS) for funding support. This work is supported by the Action Spécifique Numérique of CNRS/INSU. This work was granted access to the HPC resources of CALMIP supercomputing center under the allocation 2021-P21021. BK acknowledges funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 865624, GPRV). JFD, CM, IB, XB, AC, XD, GH, and FS acknowledge funding from Agence Nationale pour la Recherche (ANR, project ANR-18-CE31-0019 SPlaSH). AM, BC, BB, and SV acknowledge funding from Programme National de Planétologie (PNP) and the Scientific Council of the Paris Observatory. XD and AC acknoweldge funding by the French National Research Agency in the framework of the Investissements d'Avenir program (ANR-15-IDEX-02), through the funding of the 'Origin of Life' project of the Université de Grenoble Alpes. JL acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 679030/WHIPLASH), and from the french state: CNES, Programme National de Planétologie (PNP), and the ANR (ANR-20-CE49-0009: SOUND). JFD and BK acknowledge funding from the European Research Council (ERC) under the H2020 research & innovation programme (grant agreement no. #740651 NewWorlds). O.V. acknowledges funding from the Agence National de la Recherche through the ANR project 'EXACT' (ANR-21-CE49-0008-01), from the Centre National d'Études Spatiales (CNES), and from the CNRS/INSU Programme National de Planétologie (PNP). Part of the work has also been carried out in the frame of the National Centre for Competence in Research Planets supported by the Swiss National Science Foundation (SNSF). WP acknowledge financial support from the SNSF for project 200021_200726. PT acknowledges supports by the European Research Council under grant Agreement no. ATMO 757858. EM acknowledges funding from FAPEMIG under project number APQ-02493-22 and research productivity grant no. 309829/2022-4 awarded by the CNPq, Brazil. We wish to thank J. Seidel for her constructive comments and questions.

## DATA AVAILABILITY

The data are available upon request to the author and the code to analyse them is publicly available.

## REFERENCES

Addison B. et al., 2019, Publ. Astron. Soc. Pac., 131, 115003
Barstow J. K., 2020, MNRAS, 497, 4183
Bouchy F. et al., 2005, A&A, 444, L15
Brogi M., Line M. R., 2019, AJ, 157, 114
Brown T. M., 2001, ApJ, 553, 1006
Buchner J. et al., 2014, A&A, 564, A125
Claret A., Bloemen S., 2011, A&A, 529, A75
Cook N. J. et al., 2022, Publ. Astron. Soc. Pac., 134, 114509
Cutri R. M. et al., 2003, VizieR Online Data Catalog, II/246
Drummond B., Mayne N. J., Manners J., Baraffe I., Goyal J., Tremblin P., Sing D. K., Kohary K., 2018, ApJ, 869, 28
Feroz F., Hobson M. P., 2008, MNRAS, 384, 449
Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601
Feroz F., Hobson M. P., Cameron E., Pettitt A. N., 2019, The Open J. Astrophys., 2, 10
Fisher C., Heng K., 2019, ApJ, 881, 25
Fouqué P. et al., 2018, MNRAS, 475, 1960
Giacobbe P. et al., 2021, Nature, 592, 205
Gibson N. P. et al., 2020, MNRAS, 493, 2215
Gibson N. P., Nugroho S. K., Lothringer J., Maguire C., Sing D. K., 2022, MNRAS, 512, 4618
Guillot T., 2010, A&A, 520, A27
Guillot T., Fletcher L. N., Helled R., Ikoma M., Line M. R., Parmentier V., 2022, preprint (arXiv:2205.04100)
Haynes K., Mandell A. M., Madhusudhan N., Deming D., Knutson H., 2015, ApJ, 806, 146
Kitzmann D., Heng K., 2018, MNRAS, 475, 94
Line M. R., Parmentier V., 2016, ApJ, 820, 78
Line M. R. et al., 2021, Nature, 598, 580
Meech A., Aigrain S., Brogi M., Birkby J. L., 2022, MNRAS, 512, 2604
Mollière P., Wardenier J. P., van Boekel R., Henning T., Molaverdikhani K., Snellen I. A. G., 2019, A&A, 627, A67
Pinamonti M. et al., 2022, A&A, 664, A65
Pino L. et al., 2018, A&A, 612, A53
Pluriel W., Zingales T., Leconte J., Parmentier V., 2020, A&A, 636, A66
Rasmussen K. C., Brogi M., Rahman F., Beltz H., Currie M., Rauscher E., Ji A. P., 2022, AJ, 164, 35
Rosenthal L. J. et al., 2021, ApJS, 255, 8
Vogt S. S., Penrod G. D., Hatzes A. P., 1987, ApJ, 321, 496
Watson C. A., de Mooij E. J. W., Steeghs D., Marsh T. R., Brogi M., Gibson N. P., Matthews S., 2019, MNRAS, 490, 1991
Welbanks L., Madhusudhan N., 2019, AJ, 157, 206
Welbanks L., Madhusudhan N., 2022, ApJ, 933, 79

## APPENDIX A: MULTISPECIES MODEL-ALL ORDERS

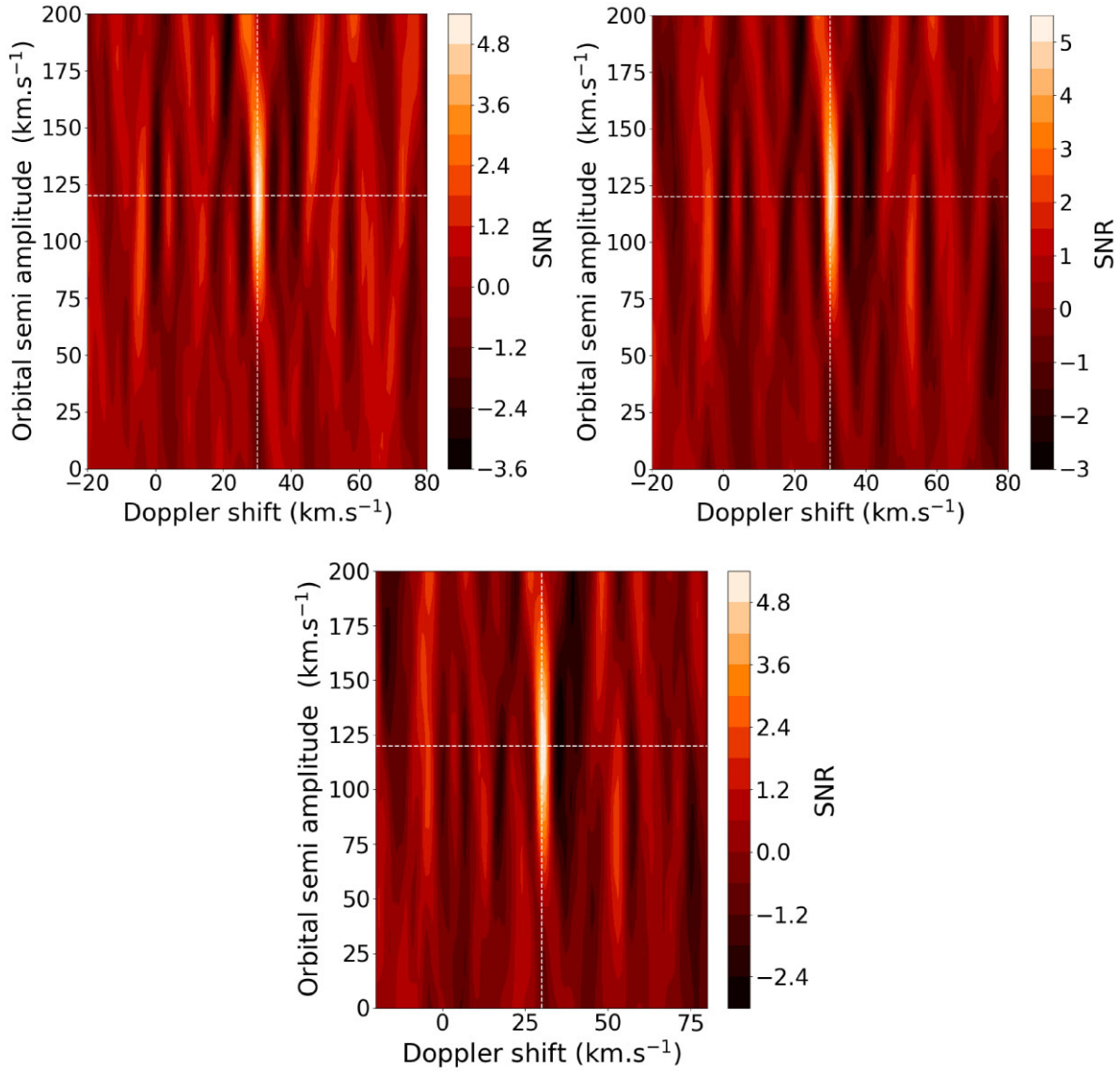


**Figure A1.** Cross correlation significance between synthetic data using Model 1 (top left), 2 (top right), and 3 (bottom) (see Section 2), and the same model with all molecules as a function of Doppler velocity and semi-amplitude.

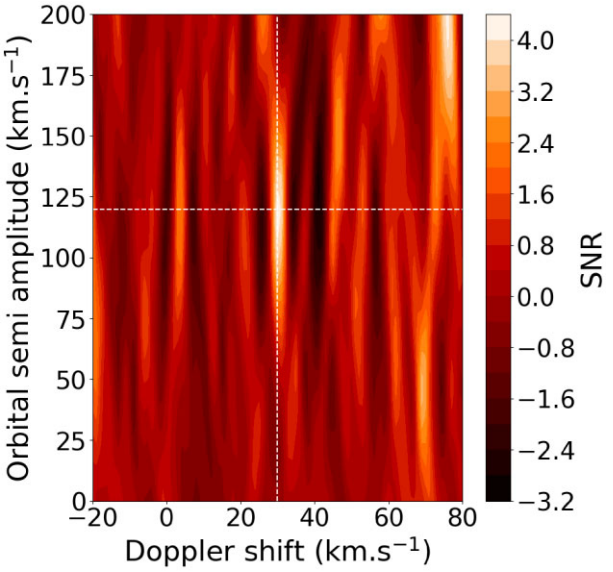


**Figure A2.** Cross correlation significance between synthetic data using Model 3 (see Section 2) and a model containing only $CH_4$ as a function of Doppler velocity and semi-amplitude.

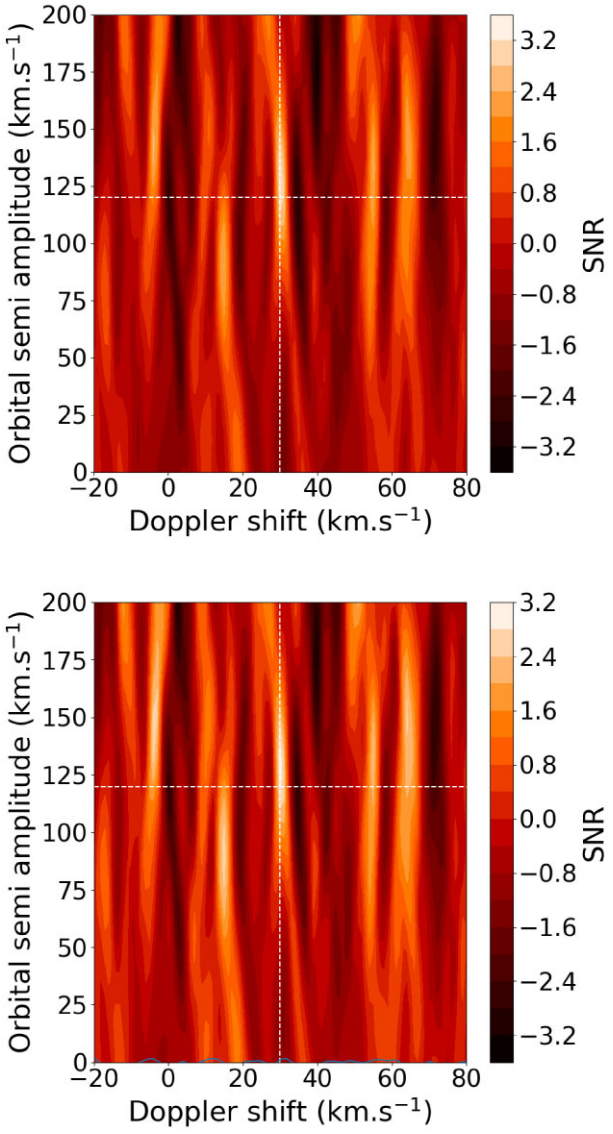


**Figure A3.** Cross correlation significance between synthetic data using Model 1 (top, see Section 2) or Model 2 (bottom) and a model containing only $NH_3$ as a function of Doppler velocity and semi-amplitude.

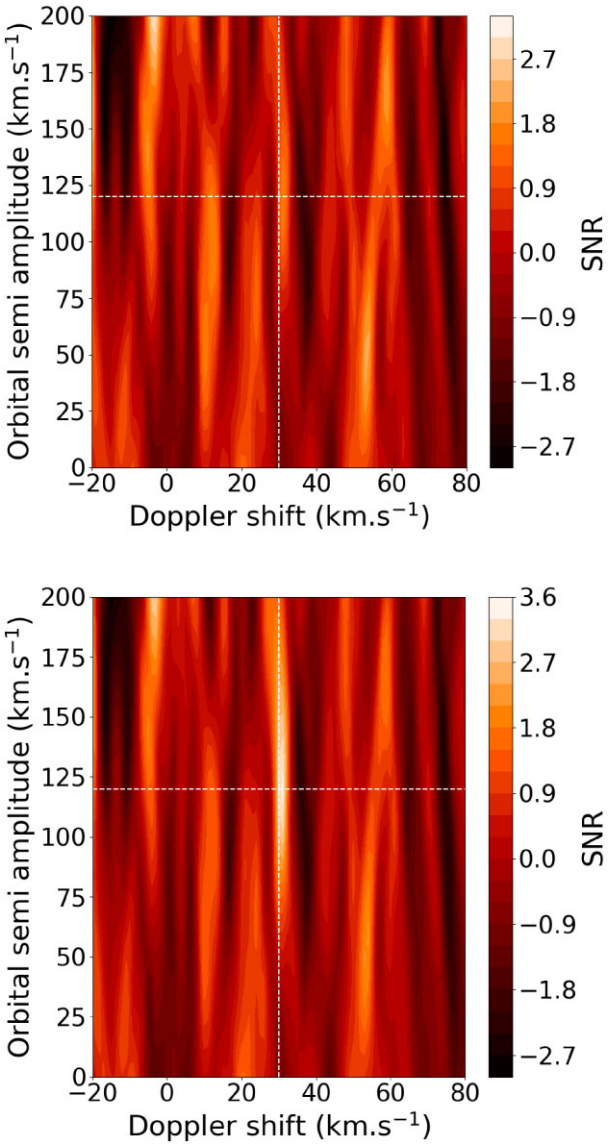


**Figure A4.** Cross correlation significance between synthetic data using Model 1 (top, see Section 2) or Model 2 (bottom) and a model containing only $H_2O$ as a function of Doppler velocity and semi-amplitude.

## APPENDIX B: MULTISPECIES MODEL-RETRIEVAL

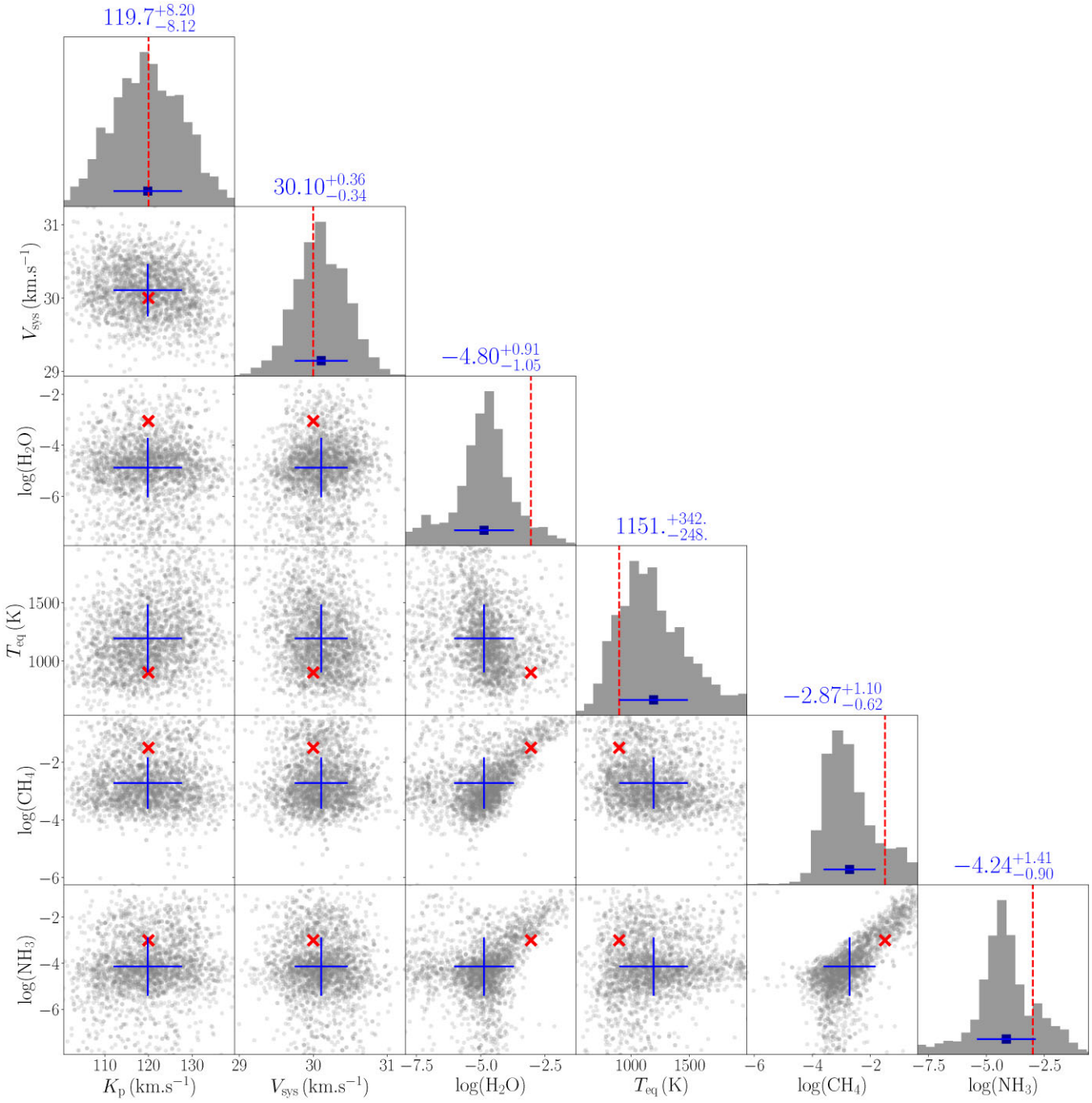


**Figure B1.** Corner plot of the result from a pymultinest retrieval with a model with multiple species, with VMR taken from the Model 1 of Table 3 (inspired by table 4 of the extended data of Giacobbe et al. (2021) as a fiducial example). The blue cross and bar shows the best-fitting value with the 1$\sigma$ error bar, the red line, and crosses show the injected values.

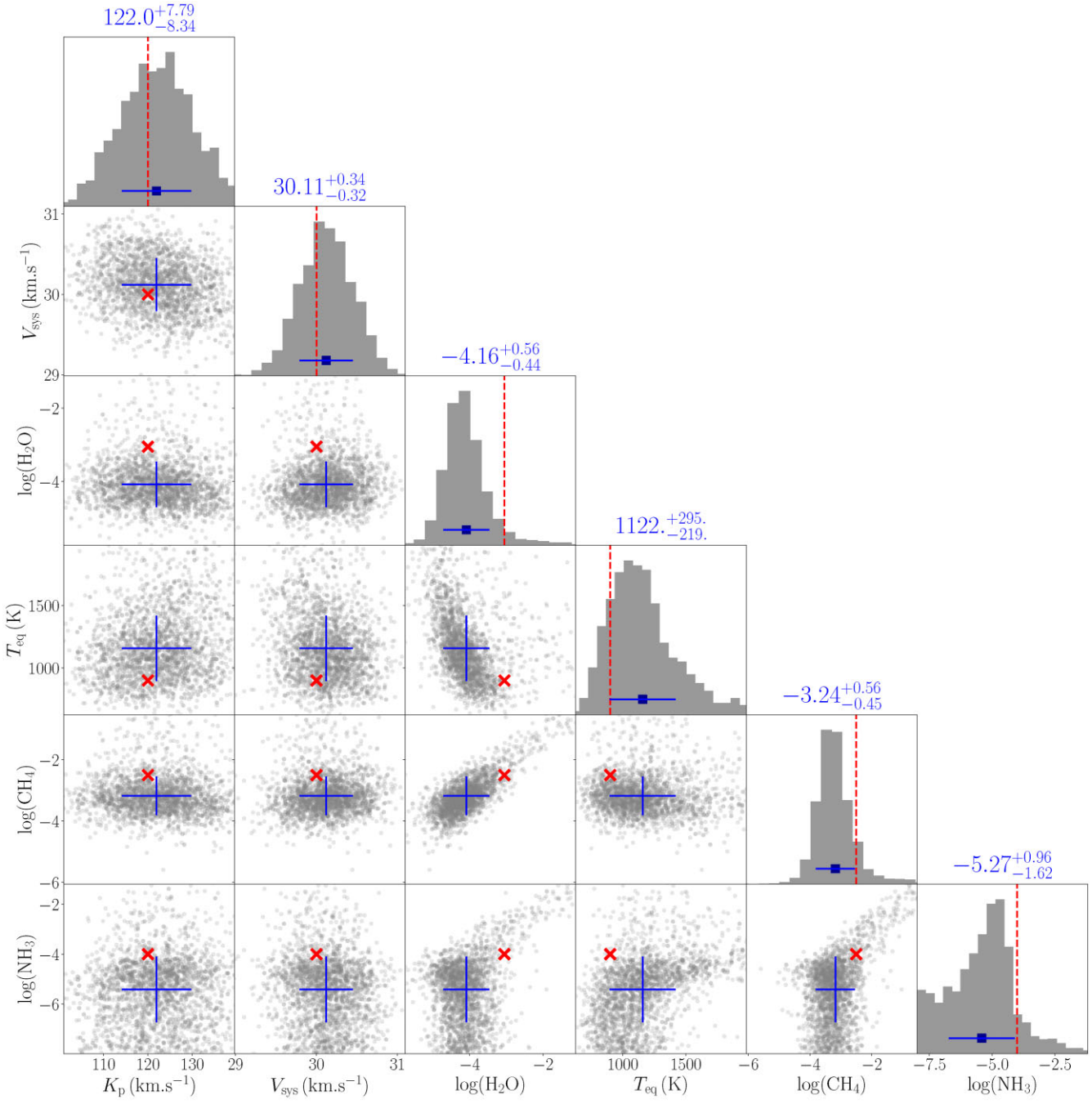


**Figure B2.** Same as Fig. B1 with Model 2 of Table 3.

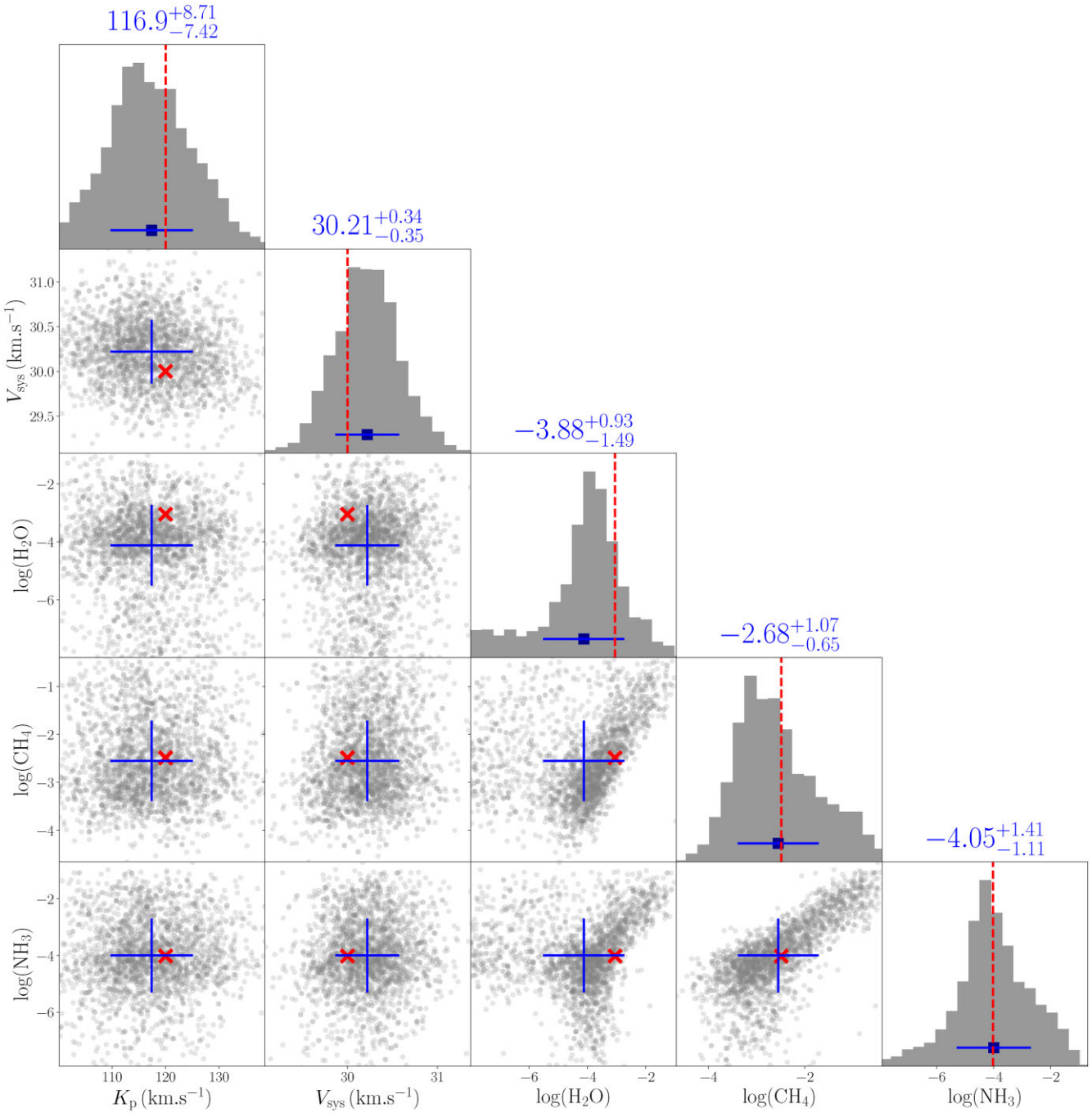


**Figure B3.** Same as Fig. B1 with Model 2 of Table 3 and no temperature retrieval.

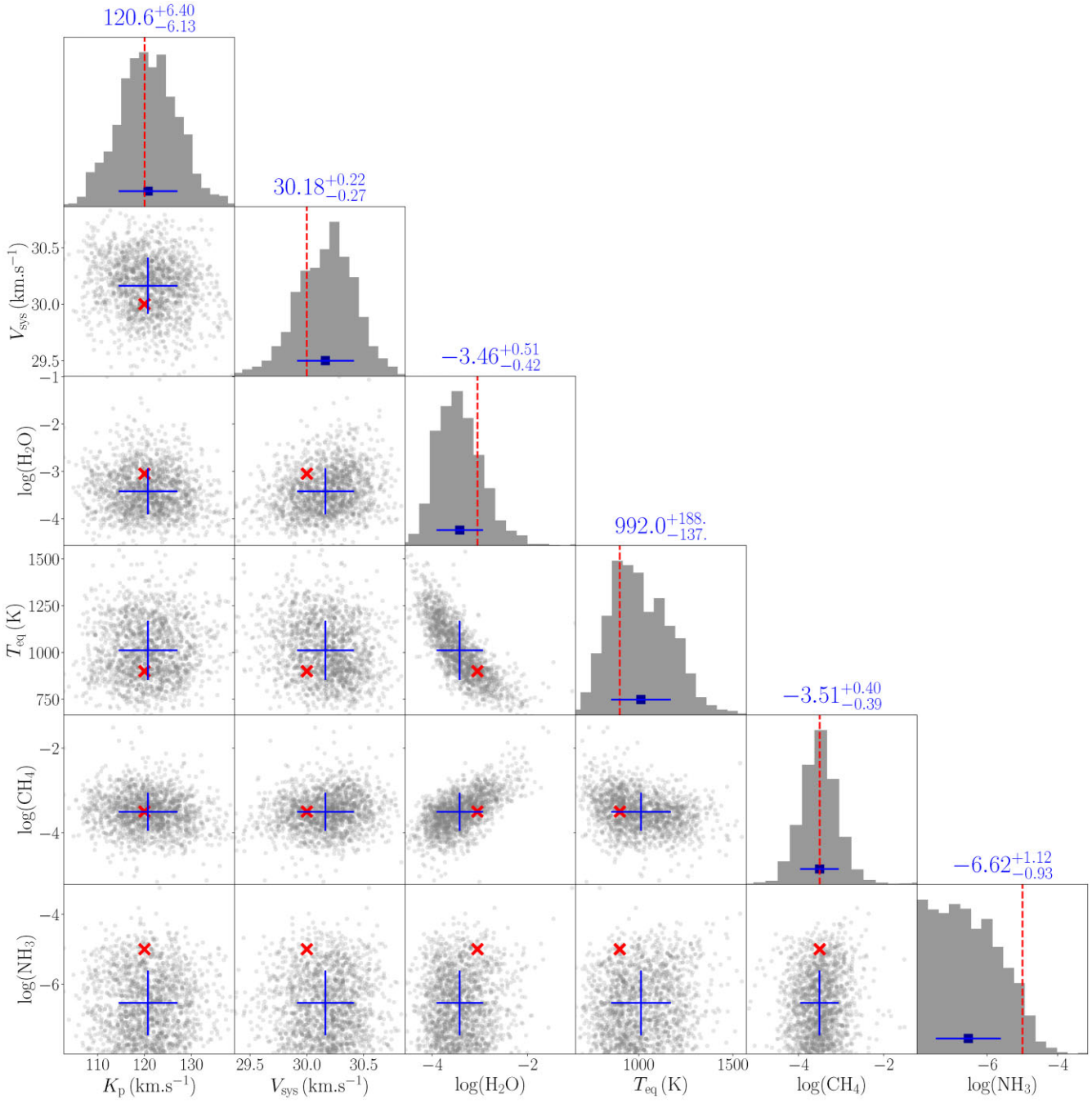


**Figure B4.** Same as Fig. B1 with Model 3 of Table 3. The shape of the $NH_3$ distribution confirms that this molecule is not detected in this model and we only obtain an upper limit for its content.

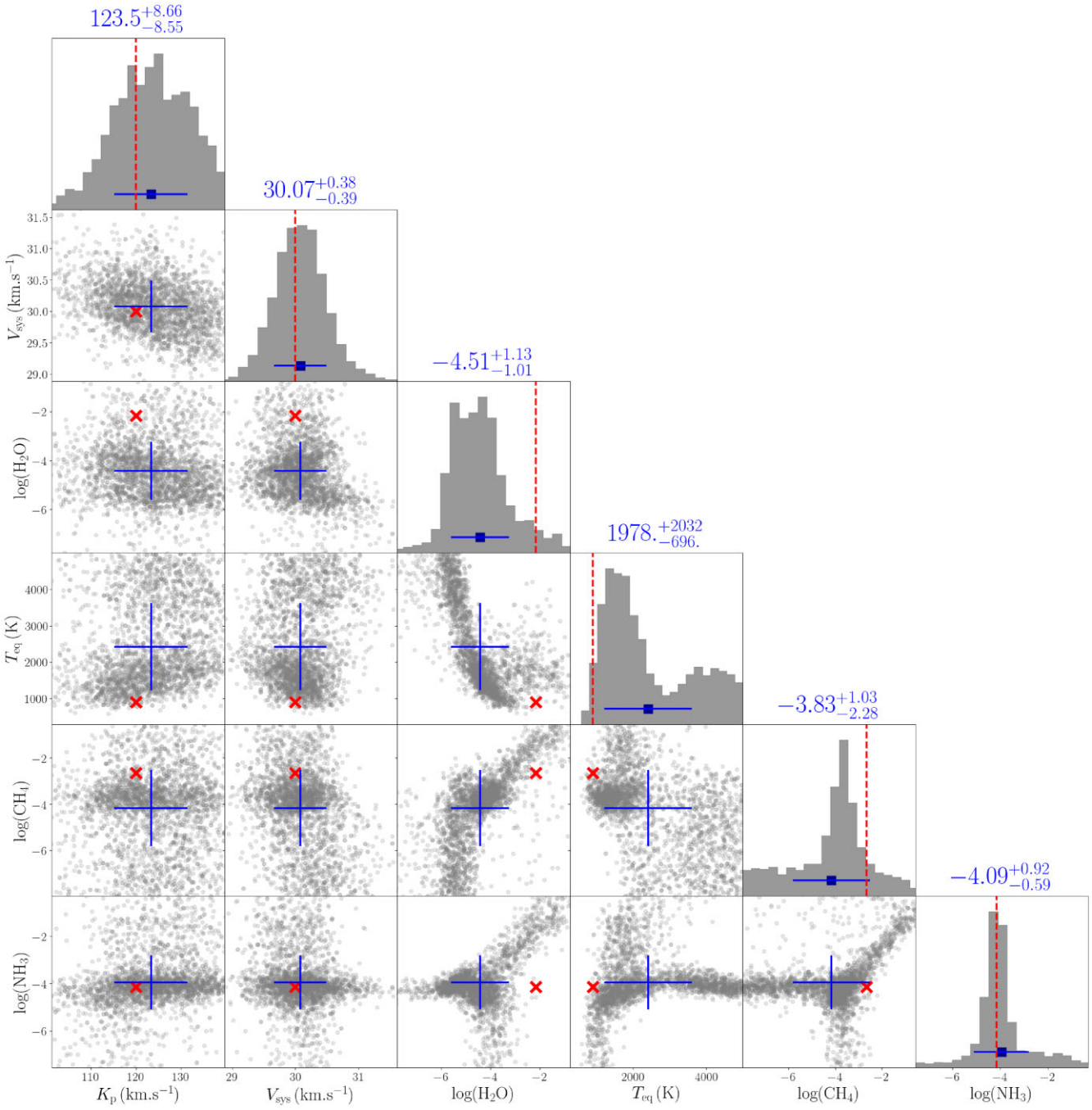


**Figure B5.** Same as Fig. B1 with Model 3 of Table 3 and removing orders 68, 52, and 39 (centred at 1.13, 1.47, and 1.96 microns). We recover a $NH_3$ detection, albeit with a very large temperature degeneracy as mentioned in the text.

This paper has been typeset from a TEX/LATEX file prepared by the author.