



**HAL**  
open science

# The RISOTTO radiative transfer and retrieval pipeline for the analysis of occultation spectra

Ashwin Braude, Stéphane Ferron, Franck Montmessin

► **To cite this version:**

Ashwin Braude, Stéphane Ferron, Franck Montmessin. The RISOTTO radiative transfer and retrieval pipeline for the analysis of occultation spectra. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 2021, 274 (November), pp.107848. 10.1016/j.jqsrt.2021.107848 . insu-03310333

**HAL Id: insu-03310333**

**<https://insu.hal.science/insu-03310333>**

Submitted on 30 Jul 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## Journal Pre-proof

The RISOTTO radiative transfer and retrieval pipeline for the analysis of occultation spectra

A.S. Braude, S. Ferron, F. Montmessin

PII: S0022-4073(21)00341-1  
DOI: <https://doi.org/10.1016/j.jqsrt.2021.107848>  
Reference: JQSRT 107848



To appear in: *Journal of Quantitative Spectroscopy & Radiative Transfer*

Received date: 16 March 2021  
Revised date: 23 July 2021  
Accepted date: 25 July 2021

Please cite this article as: A.S. Braude, S. Ferron, F. Montmessin, The RISOTTO radiative transfer and retrieval pipeline for the analysis of occultation spectra, *Journal of Quantitative Spectroscopy & Radiative Transfer* (2021), doi: <https://doi.org/10.1016/j.jqsrt.2021.107848>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier Ltd.

## Highlights

- New radiative transfer and retrieval pipeline designed for occultation data.
- Able to correct for uncertainties in multiple instrumental parameters.
- Results validated to be within 2% of existing radiative transfer codes.

Journal Pre-proof

# The RISOTTO radiative transfer and retrieval pipeline for the analysis of occultation spectra

A. S. Braude<sup>a,\*</sup>, S. Ferron<sup>b</sup>, F. Montmessin<sup>a</sup>

<sup>a</sup>*Laboratoire Atmosphères, Milieux, Observations Spatiales (LATMOS), UVSQ Université  
Paris-Saclay, Sorbonne Université, CNRS, Paris, France*

<sup>b</sup>*ACRI-ST, 11 Boulevard d'Alembert, 78280 Guyancourt, France*

---

## Abstract

We present a new radiative transfer and retrieval pipeline, referred to as RISOTTO, designed for observations made in occultation geometry. RISOTTO can retrieve accurate vertical profiles of the constituents of a planetary atmosphere using a Bayesian approach while simultaneously correcting for instrumental artefacts and uncertainties. The algorithm makes use of a rapid Abel integration scheme involving matrix multiplication, using pre-calculated gaseous cross-section data to calculate the integration along the line of sight, while also separating out any effects of transmission baseline variation. We assess its accuracy using synthetic solar occultation spectra of Mars and comparing it with the older more established NEMESIS radiative transfer and retrieval pipeline (Irwin et al. 2008, *J. Quant. Spec. Radiat. Transf.* 109, 1136-1150), and show that it can match its accuracy while also being able to more easily retrieve gases with substantial continuum absorption.

**Keywords:** Radiative transfer, Retrievals, Atmospheres, composition

---

\*Corresponding author

*Email address:* ashwin.braude@latmos.ipsl.fr (A. S. Braude)

## 1 **1. Introduction**

2 Solar and stellar occultation observations of a planetary atmosphere using  
3 high-resolution spectroscopy provide a valuable means of measuring vertically  
4 resolved atmospheric abundance profiles of different gaseous species[1], with the  
5 geometry of the spacecraft designed to detect radiation that passes through the  
6 atmosphere tangentially to the surface. This geometry maximises the optical path  
7 of radiation that passes through the atmosphere, and thus the total absorption of  
8 a given gaseous species, allowing for the detection of trace gases at quantities  
9 far lower than ground-based or nadir observations are able to achieve. The oc-  
10 cultation process involves viewing a bright source with a very well-characterised  
11 emission spectrum, such as the sun or a distant star, providing a simple baseline  
12 reference from which the transmission of radiation through the atmosphere can be  
13 calculated. To extract scientific information out of these spectral data, a radiative  
14 transfer and retrieval pipeline is required. The aim of the radiative transfer model  
15 is to simulate gaseous absorption along the line of sight, while the retrieval stage  
16 then uses this model to convert input values of transmission into realistic vertical  
17 gas profiles and, where applicable, to extract information on vertical temperature  
18 and pressure variations.

19 In this work, we present the code for Retrievals in the Infrared of Solar Occul-  
20 TaTiOns (RISOTTO), a new retrieval algorithm based on Bayesian inference with  
21 an implemented forward model that can retrieve atmospheric profiles using a rapid  
22 and optimised matrix inversion scheme. This code is also capable of simultane-  
23 ously correcting for minor instrumental artefacts in the data, together with any  
24 distortions of the spectral baseline (for instance due to aerosol absorption), and  
25 can provide a first-order quantification of the signal-to-noise ratio (SNR) of each  
26 spectrum in the occultation without needing any prior information on sources of  
27 noise. Unlike many other similar radiative transfer and retrieval tools, it is gen-

28 eralisable and does not rely on large amounts of prior information on technical  
29 instrumental parameters. This therefore makes it run more efficiently, have a sim-  
30 pler user interface, and is easier to tailor and customise to the instrument or the  
31 spectral domain at hand.

32 Occultation observations usually have two main aims: firstly, to find evidence  
33 of trace gases in a planetary atmosphere and provide accurate detection limits in  
34 the case of non-detection; and secondly to constrain the climatology of the planet  
35 through vertically-resolved observations of known gases in the atmosphere. These  
36 two objectives require very different retrieval techniques. In the former case, the  
37 retrieval algorithm has to be robust to a very large uncertainty on *a priori* vertical  
38 variations of gaseous abundances (if the molecule has never been detected on the  
39 planet before, there is by definition no prior information on the vertical profile of  
40 the gas, and so the *a priori* uncertainty is close to infinite) while also having to  
41 clearly distinguish a detection from a non-detection. In the latter case, there is less  
42 uncertainty on the prior profile but the algorithm has to produce vertical profiles  
43 of gaseous compounds to very high accuracy and precision.

44 In Section 2 we describe how the code estimates the signal-to-noise ratio of  
45 the spectrum, together with reference molecular absorption data and instrumental  
46 parameters, and thereby calculates a forward model of the spectrum through path  
47 integration along the line of sight. In Section 3 we then describe the inversion  
48 algorithm that results in atmospheric profiles that provide an optimal fit to the  
49 observed spectra given the aforementioned forward model. We show that the re-  
50 sults obtained using this method are accurate by comparing them with the results  
51 obtained from similar data using a different, more established, radiative transfer  
52 and retrieval code called NEMESIS [2] in Section 4. Finally, we provide some  
53 concluding remarks in Section 5.

## 54 2. Forward model

### 55 2.1. Definition

56 The aim of a forward model is to be able to compute a simulated set of trans-  
 57 mission values given a vector of known variables that we refer to as the *state vec-*  
 58 *tor*. We distinguish two sets of variables that make up the state vector: *scientific*  
 59 variables which define the properties of the atmosphere that we wish to investigate  
 60 in the first place, and *instrumental* variables which define properties of the spectra  
 61 that are related to calibration defects or uncertainties. We focus on three of these  
 62 corrections in particular: those due to uncertainties in the transmission baseline  
 63 (continuum modelling), those due to uncertainties in the instrumental line shape  
 64 (ILSF modelling), and those due to uncertainties in the spectral registration (spec-  
 65 tral law derivation). We assume the following model to derive the corrected trans-  
 66 mission values  $T_{cor}$  from a set of transmission values  $T_{mod}$  modelled according to  
 67 the wavenumber grid  $\lambda_{mod}$  with respect to which the molecular cross-sections are  
 68 stored:

$$T_{cor}(\lambda_{cor})_i = C(\lambda_{cor})_i \sum_j K(\lambda_{cor}, \lambda_{mod})_{ij} T_{mod}(\lambda_{mod})_j \quad (1)$$

69 where  $C$  is the continuum model and  $K$  is the instrument line shape model.  
 70 This parameter  $K$  takes into account both the modelled wavenumber grid  $\lambda_{mod}$  and  
 71 the estimated wavenumber grid of the observed data  $\lambda_{cor}$  (both in units of  $\text{cm}^{-1}$ ).  
 72 We discuss these three instrumental variables individually in section 2.2, and then  
 73 describe how scientific variables are used to calculate  $T_{mod}$  in sections 2.3 - 2.5.

### 74 2.2. Instrumental variables

#### 75 2.2.1. Spectral Law Derivation

76 A first guess of spectral registration can be made using reference solar or stel-  
 77 lar lines that are present in the spectrum (eg. [3]), but is not usually sufficient

78 for accurate fitting of high-resolution spectra, especially in wavenumber regions  
 79 where these reference lines are lacking. This means that a second spectral reg-  
 80 istration process is needed to further refine the first guess. The forward model  
 81 assumes a polynomial relationship between  $\lambda_{mod}$  and  $\lambda_{cor}$ :

$$\lambda_{cor} - \lambda_{ref} = \begin{pmatrix} a_0 & a_1 & a_2 & \dots & a_n \end{pmatrix} \begin{pmatrix} 1 \\ \lambda_{mod} - \lambda_{ref} \\ (\lambda_{mod} - \lambda_{ref})^2 \\ \vdots \\ (\lambda_{mod} - \lambda_{ref})^n \end{pmatrix} \quad (2)$$

82 where the polynomial coefficients  $\begin{pmatrix} a_0 & a_1 & a_2 & \dots & a_n \end{pmatrix}$  are variables of the  
 83 model to be retrieved. For relatively small wavenumber ranges, a linear relation-  
 84 ship ( $n = 1$ ) is usually sufficient, while retrievals of higher-order terms may be  
 85 required for larger wavenumber ranges. The wavenumber  $\lambda_{ref}$  is a single arbitrary  
 86 wavenumber value (usually fixed to the average of  $\lambda_{mod}$ ) that acts as a centralising  
 87 factor for when the wavenumber values in  $\lambda_{mod}$  are of several orders of magnitude  
 88 larger than zero (e.g. in the near-infrared where  $\lambda_{mod} \sim 10^5 \text{ cm}^{-1}$ ), and to thereby  
 89 make the retrieval of higher-order polynomial coefficients more efficient.

90 This spectral shift correction process only works when  $\lambda_{mod}$  and  $\lambda_{cor}$  are al-  
 91 ways close enough that a modelled gas absorption peak cannot be fully resolved  
 92 spectrally from an observed gas absorption peak, that is to say that  $\lambda_{mod} - \lambda_{cor}$  can-  
 93 not be greater than the full-width half maximum (FWHM) of a given absorption  
 94 line. In cases where this is not satisfied, an approximate correction may need to  
 95 be made manually after a single forward model, which the retrieval algorithm can  
 96 then refine further as necessary.

97 *2.2.2. Instrument Line Shape Function model*

98 The finite spectral resolution of an instrument usually results in an instrumen-  
 99 tal signature, or Instrumental Line Shape Function (ILSF) that distorts an observed  
 100 spectrum through convolution in a manner that is often, but not always, well-  
 101 characterised. The algorithm is currently equipped to model an ILSF according to  
 102 a simple Gaussian, although additional types of ILSF can easily be added to the  
 103 algorithm according to the instrument at hand. The term  $K(\lambda_{cor}, \lambda_{mod})_{ij}$  in Equa-  
 104 tion 1, which solely depends on a single retrievable parameter  $p_{gauss}$  defined as the  
 105 spectral resolution in units of  $R = \frac{\lambda}{\Delta\lambda}$ , would therefore be represented as follows:

$$K(\lambda_{cor}, \lambda_{mod})_{ij} = \frac{1}{\sigma N_i} \exp\left(-\frac{1}{2} \left(\frac{\lambda_{cor(j)} - \lambda_{mod(i)}}{\sigma}\right)^2\right) \quad (3)$$

106 where  $\sigma = \frac{\lambda_{mod}}{2\sqrt{2\ln(2)}p_{gauss}}$  describes the standard deviation of the Gaussian in  
 107 wavenumber units, and  $N_i = \sum_j K(\lambda_{cor}, \lambda_{mod})_{ij}$  is a normalisation factor.

108 *2.2.3. Continuum model*

109 Often the baseline, or continuum level, of a given transmission spectrum can  
 110 exhibit unwanted distortions due to a number of factors such as instrumental noise,  
 111 poor calibration or the presence of aerosol. This provides an additional source of  
 112 uncertainty particularly when measuring the abundances of certain molecules that  
 113 exhibit substantial continuum absorption at the given spectral resolution.

114 A prior continuum model is therefore found by linearly interpolating over local  
 115 maxima in a spectrum and then smoothing the resulting line [4], which we usually  
 116 find provides a better first guess to the baseline level than the more established  
 117 convex and concave hull methods [5]. Although the shape of the baseline level is  
 118 refined further in the retrieval process, a constraint must be placed on it to prevent  
 119 it from overfitting the spectra. This is done by multiplying a diagonal matrix of  
 120 the continuum uncertainties  $\sigma_i^2$  with a kernel function to give a prior continuum

121 covariance matrix  $\mathbf{C}_{\text{cont}}$  that induces a correlation between neighbouring points on  
 122 the continuum. Empirically we find that a Matérn kernel [6] provides the best fit:

$$C_{contij} = \sigma_i \left( \frac{2^{1-\nu}}{\Gamma(\nu)} \left( 2 \frac{\sqrt{(i-j)^2 + \sqrt{2\nu}}}{\rho} \right)^\nu K_\nu \left( 2 \frac{\sqrt{(i-j)^2 + \sqrt{2\nu}}}{\rho} \right) \right)_{ij} \sigma_j \quad (4)$$

123 where  $\Gamma(\nu)$  is the Gamma function,  $K_\nu$  is the modified Bessel function of the  
 124 second kind and  $\nu$  is an integer that controls how the correlation changes with  
 125 distance from a given point. The parameter  $\rho$  is the key constraint on the fitting  
 126 of the continuum - the lower the value the higher the frequency of the baseline  
 127 variation in wavenumber space. Hence a value of  $\rho$  that is too large may underfit  
 128 the spectrum by neglecting medium-frequency variations in the baseline, while a  
 129 value of  $\rho$  that is too small may overfit the spectrum and mistake any instrumental  
 130 artefacts or genuine gas absorption features for variations in the baseline.

131 The advantage of this method is that it allows greater flexibility in the con-  
 132 tinuum level for a given retrieval while minimising the possibility of overfitting.  
 133 This enables a much more efficient fitting procedure of species with substantial  
 134 continuum absorption where the baseline level is uncertain - upper limits on these  
 135 species can be derived directly through a single retrieval without the need for  
 136 a complex continuum subtraction procedure, as explained further in section 4.2.  
 137 However, if total flexibility of the continuum over all wavenumbers and tangent  
 138 heights is required, the speed of the retrieval may be much slower since the num-  
 139 ber of state vector variables required will be equal to the total number of spectral  
 140 points.

### 141 2.3. Scientific variables

142 Given the information available in the mid-infrared, we distinguish three sets  
 143 of scientific variables in the model: vertical gaseous abundance profiles, vertical  
 144 temperature profiles and a reference pressure value  $P_0$  known *a priori* at a fixed

145 altitude, usually the surface ( $z_0 = 0$  km). In order to calculate the vertical pressure  
 146 profile  $P(z)$ , we make two assumptions. Firstly, we assume that the atmosphere is  
 147 an ideal gas. Secondly, we assume that the relationship between temperature and  
 148 pressure at a given altitude grid  $z_j$  can be described according to the hydrostatic  
 149 approximation:

$$P(z_j) = P_0 \exp\left(-\sum_{i=0}^{j-1} \frac{z_{i+1} - z_i}{2} \left(\frac{Mg(z_{i+1})}{RT_{i+1}} + \frac{Mg(z_i)}{RT_i}\right)\right) \quad (5)$$

150 where  $N$  is the total number of altitude gridpoints,  $M$  is the molar mass of the  
 151 planetary atmosphere (assumed constant with altitude and location),  $g(z_i)$  is the  
 152 gravitational acceleration,  $R$  is the universal gas constant and  $T_i$  is the temperature  
 153 at altitude  $z_i$ .

154 If there is some uncertainty in the ILSF, gaseous abundance retrievals may  
 155 be degenerate with temperature and pressure. In this case, either a fixed pressure-  
 156 temperature profile must be assumed (using a prior obtained either from a separate  
 157 instrument or from a theoretical model) or the profile of another gas (such as  $\text{CO}_2$   
 158 in the case of Mars or  $\text{CH}_4$  in the case of Jupiter and Saturn) must be assumed to  
 159 be fixed and relatively well-mixed in the atmosphere at the required altitudes. An  
 160 example of the latter case will be discussed further in section 4.1.

#### 161 2.4. Molecular cross-sections

162 For a given molecule, a Humlíček algorithm [7] is used to calculate molecular  
 163 absorption cross-sections  $\sigma(p_i, t_j, \lambda_k)$  as a function of wavenumber  $\lambda_k$  from a set  
 164 of source line data (e.g. the HITRAN 2016 database [8]), assuming a Voigt line-  
 165 shape. These are calculated independently for a set of  $n_i$  pressure grid values ( $p_i$ ,  
 166 in units of Pascals) and  $n_j$  temperature grid values ( $t_j$ , in units of Kelvin), chosen  
 167 to efficiently sample the variation in molecular absorption for a temperature and  
 168 pressure parameter space suitable for the planetary atmosphere in question. The

169 cross-sections could then be linearly interpolated to intermediate temperature and  
 170 pressure values during the forward modelling process, but this is both slow and  
 171 increases the probability of inducing errors in the optical path integration due to  
 172 aliasing or undersampling.

173 If we therefore assume that, for a given molecule, the molecular absorption  
 174 cross-section at a given wavenumber  $\lambda_k$  can be represented as a quartic surface  
 175 with respect to temperature  $T$  and the logarithm of pressure  $\ln(P)$ :

$$\sigma_k(P, T) \approx C_{00(k)} + C_{10(k)} \ln(P) + C_{01(k)} T + C_{11(k)} \ln(P) T \\ + C_{20(k)} \ln(P)^2 + C_{02(k)} T^2 \dots + C_{44(k)} \ln(P)^4 T^4 \quad (6)$$

176 or equivalently:

$$\sigma_k(P, T) \approx \text{trace}(\mathbf{C}_k \mathbf{M}) \quad (7)$$

177 with  $\mathbf{M}$  defined as follows:

$$\mathbf{M} = \begin{pmatrix} \ln(P)^0 \\ \ln(P)^1 \\ \vdots \\ \ln(P)^4 \end{pmatrix} \begin{pmatrix} T^0 & T^1 & \dots & T^4 \end{pmatrix} \quad (8)$$

178 then one can solve a set of  $n_i \times n_j$  simultaneous equations for the coefficients  
 179 in  $\mathbf{C}_k$  substituting the cross-section values pre-calculated at the aforementioned  
 180 temperature and pressure grid into equation 6:

$$\left\{ \begin{array}{l}
\sigma_k(p_1, t_1) = C_{00(k)} + C_{10(k)} \ln(p_1) + C_{01(k)} t_1 + C_{11(k)} \ln(p_1) t_1 \\
\quad + C_{20(k)} \ln(p_1)^2 + C_{02(k)} t_1^2 \dots + C_{44(k)} \ln(p_1)^4 t_1^4 \\
\sigma_k(p_2, t_1) = C_{00(k)} + C_{10(k)} \ln(p_2) + C_{01(k)} t_1 + C_{11(k)} \ln(p_2) t_1 \\
\quad + C_{20(k)} \ln(p_2)^2 + C_{02(k)} t_1^2 \dots + C_{44(k)} \ln(p_2)^4 t_1^4 \\
\vdots \\
\sigma_k(p_{n_i}, t_{n_j}) = C_{00(k)} + C_{10(k)} \ln(p_{n_i}) + C_{01(k)} t_{n_j} + C_{11(k)} \ln(p_{n_i}) t_{n_j} \\
\quad + C_{20(k)} \ln(p_{n_i})^2 + C_{02(k)} t_{n_j}^2 \dots + C_{44(k)} \ln(p_{n_i})^4 t_{n_j}^4
\end{array} \right. \quad (9)$$

181 These coefficients  $C_k$  can be pre-calculated once and then stored in look-up ta-  
182 bles as a function of wavenumber. The look-up tables can then be directly loaded  
183 into the forward model when needed, and then the cross-section values calcu-  
184 lated at the required temperature and pressure grid using the matrix multiplication  
185 scheme in equation 7. We find that this method is usually around 3-4 times faster  
186 than linear interpolation.

### 187 2.5. Path integration through an Abel approximation scheme

188 Our forward model relies on two major assumptions. Firstly, we assume that  
189 all radiation from the sun passes directly through the atmosphere into the detector  
190 without any diffuse component from scattering in the atmosphere or from surface  
191 reflection. This is often an assumption made for solar occultation retrievals con-  
192 ducted in the mid-infrared (e.g. [9, 10]) and a further analysis on the applicability  
193 of this assumption to planetary solar occultation spectra can be found in [11]. Any  
194 extinction due to the presence of aerosol in the atmosphere is assumed to be broad  
195 relative to the observed gas absorption peaks given the spectral resolution, and is  
196 therefore removed during the baseline correction process. Secondly, we assume  
197 that the atmosphere is spherically symmetric and that gas abundances only vary

198 with altitude. Since we are observing the atmosphere at the limb, we can assume  
 199 an optical path that is exactly perpendicular to a tangent point at a distance  $z_T$  di-  
 200 rectly above the planetary surface, and therefore simplify the calculation of slant  
 201 opacity using the Beer-Lambert Law to an Abel integral equation (e.g. [12]):

$$\tau_{T\lambda} = 2 \int_{R_T}^{\infty} \sum_{m=1}^{N_m} \frac{\rho(z)x_m}{M} \sigma_{m\lambda}(z) \frac{R_T + z}{\sqrt{(R_T + z)^2 - (R_T + z_T)^2}} dz \quad (10)$$

202 where  $\sigma$  is the extinction cross-section of a molecule (of volume mixing ratio  
 203  $x_m$ ) at a given wavenumber  $\lambda$  and as a function of altitude  $z$  from the surface, and  
 204  $R_T$  is the planetary radius below the observed tangent height.  $M$  is the molar mass  
 205 of the planetary atmosphere which is assumed to be constant with altitude. The  
 206 factor of 2 in the integral is to account for both the incoming and outgoing stream  
 207 of radiation through the atmosphere, which is illustrated further in figure 1.

208 We assume that the total number density of molecules  $\rho$  in the atmosphere at  
 209 each altitude gridpoint  $z_j$  can be described according to the Ideal Gas Law:

$$\rho(z_j) = \frac{P_j}{k_B T_j} \quad (11)$$

210 with  $k_B$  referring to the Boltzmann constant.

211 We also assume that the molecular absorption cross-sections  $\sigma_{m\lambda}(z)$  vary con-  
 212 tinuously with altitude  $z$ :

$$\sigma_{m\lambda}(z) = \sum_{p=0}^{n_p} a_{mpj\lambda} z^p \quad (12)$$

213 The values  $a_{mpj\lambda}$  are found by calculating sample reference values  $\sigma_{m\lambda u}$  at dis-  
 214 crete altitude gridpoints  $z_u$  (with associated pressure values  $P_u$  and temperature  
 215 values  $T_u$ ), and then performing an interpolation of degree  $n_p$  to an arbitrary inter-  
 216 mediate altitude  $z_j$ :

$$a_{mpj\lambda} \equiv \sum_{u=1}^N C_{ju}^{(p)} \sigma_{m\lambda u} \quad (13)$$

217 The set of matrices  $C_{ju}^{(n_p)}$  therefore represents a function for linear interpolation  
 218 (if  $n_p = 1$ ) or cubic spline interpolation (if  $n_p = 3$ ) of  $\sigma_{m\lambda}$  from one set of altitude  
 219 gridpoints  $z_u$  to another set of altitude gridpoints  $z_j$ . This has two purposes: firstly,  
 220 to remove an integration singularity when  $z = z_T$ , and secondly to account for  
 221 the fact that the cross-sections are dependent on pressure and temperature which  
 222 themselves vary continuously with altitude. We find empirically that cubic spline  
 223 interpolation provides the best trade-off between accuracy and speed.

224 We can therefore define a matrix  $A_{Tu}$ , referred to as the *Abel kernel*, which is  
 225 derived from the substitution of equations 12-13 into equation 10:

$$A_{Tu} = 2 \sum_{j=1}^{N-1} \sum_{p=0}^3 C_{ju}^{(p)} \int_{\delta_{jt}z_T + (1-\delta_{jt})z_j}^{z_{j+1}} z^p \frac{R_T + z}{\sqrt{(R_T + z)^2 - (R_T + z_T)^2}} dz \quad (14)$$

226 Hence, equation 10 can be approximated through quadrature as follows:

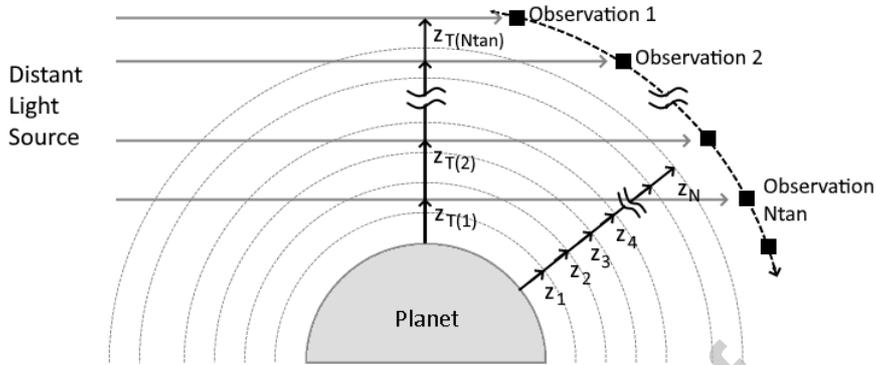
$$\tau_{\lambda T} \approx \sum_{u=1}^N \sum_{m=1}^{N_m} \frac{\rho_u}{M} A_{Tu} \sigma_{m\lambda u} x_{mu} \quad (15)$$

227 and so:

$$T_{mod}(\lambda_{mod})_j = \exp(-\tau_{\lambda T}) \quad (16)$$

228 where  $t$  in this equation is equal to the index of the altitude gridpoint that is  
 229 located just below  $z_T$ , and  $\delta_{jt}$  is the Kronecker delta function. The integral in  
 230 equation 14 is solved through global adaptive quadrature.

231 Equation 15 can therefore allow us to calculate modelled transmission spectra  
 232  $\exp(\tau_{\lambda T})$  from a series of vertical profiles of gas abundance  $x_{mu}$  through a simple  
 233 matrix multiplication, where the Abel kernel  $A_{Tu}$  only needs to be calculated once



**Figure 1:** A diagram of a given occultation consisting of  $N_{tan}$  individual observations of the atmosphere of a planet, adapted from Figure 1 of Takagi et al. [13]. A spacecraft in orbit around the planet performs multiple observations of a distant light source directly through the atmosphere, with the optical path of the radiation for each observation, marked in grey, passing through the atmosphere parallel to the surface at a given tangent point. The altitude of the path above the tangent point is referred to as a ‘tangent height’  $z_T$ , with each observation associated with a unique tangent height. This is not to be confused with the designation  $z$  which refers to the height above the surface of a given atmospheric layer (marked in the semicircular dotted lines) at which a gas abundance or temperature value is evaluated.

234 at the beginning of the retrieval. We find an altitude grid sampling  $z_{i+1} - z_i$  com-  
 235 parable to the difference in altitude between adjacent observations in the occulta-  
 236 tion is sufficient to provide adequate vertical resolution without compromising on  
 237 speed.

### 238 2.6. Estimation of the signal-to-noise ratio

239 In order to accurately constrain detection limits for minor species, a reasonable  
 240 estimation of the signal-to-noise ratio (SNR) of the spectrum at each diffraction  
 241 order and tangent height is required. We intend to approximately estimate the  
 242 total SNR of a spectrum without any prior information either on sources of instru-  
 243 mental noise (eg. thermal dark current) or on the forward model. To this end, we  
 244 compared three different noise quantification models:

245 **Method 1** Calculate the standard deviation of the difference between the original  
 246 spectrum and the spectrum smoothed by a Gaussian filter of a set number

247 of pixel widths, removing anomalous values of the difference greater than 3  
248 sigma. The estimate of the SNR is then calculated by dividing the median  
249 of the unsmoothed spectrum by the difference value extrapolated to a filter  
250 width of 0.

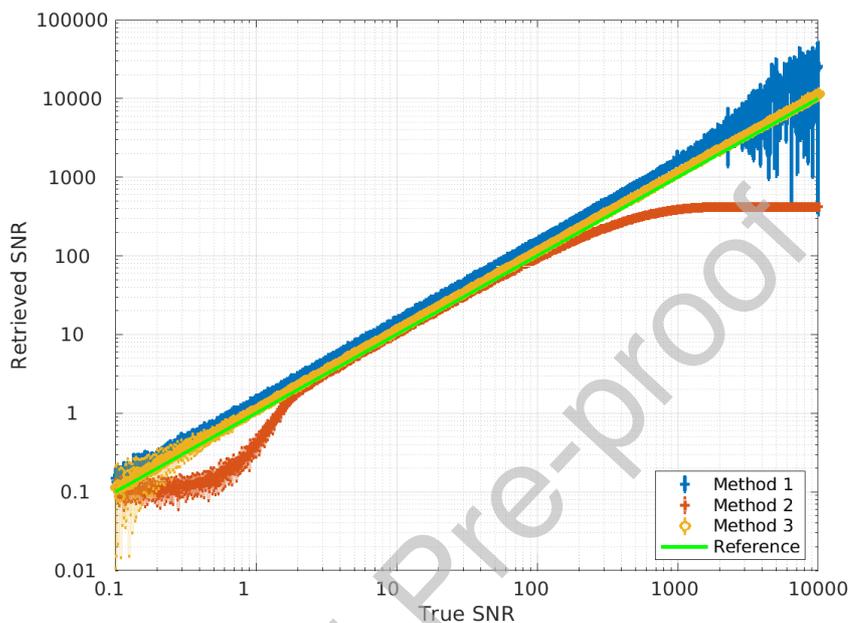
251 **Method 2** Calculate the difference between the original spectrum and the spec-  
252 trum smoothed by a Gaussian filter, and then derive the SNR through the  
253 inverse of a robust estimate of the difference value ( $1.4826 * \text{mean absolute}$   
254  $\text{deviation}(\text{unsmoothed spectrum} - \text{smoothed spectrum})$ )

255 **Method 3** Find all the local maxima present in the spectrum smoothed by a Gaus-  
256 sian filter, and then estimate the SNR from the median of the spectrum di-  
257 vided by the average variance in the spectrum around each local maximum.

258 In Figure 2 we compare the accuracy of these three methods for a number of  
259 synthetic occultations fixed to a given SNR. We find that Method 3 generally  
260 provides the most accurate estimations of SNR over the widest range of noise  
261 values, especially for solar occultation spectra of Mars like those observed by  
262 the Trace Gas Orbiter/Atmospheric Chemistry Suite (ACS) instrument with SNR  
263 values in the thousands [14]. However, Method 2 may be more accurate for values  
264 of SNR  $\sim 10$ -100 as can be found more typically in mid-ultraviolet observations  
265 of stellar occultations of Mars by the SPICAM instrument [15] on board Mars  
266 Express, as well as of similar observations by the SPICAV instrument on board  
267 Venus Express [16].

268 The main advantage in our quantification of SNR compared with other re-  
269 trieval models is that it does not either require any prior knowledge of the instru-  
270 mentation nor does it require multiple retrievals to accurately refine. However,  
271 it cannot easily take into account wavenumber dependency in SNR. In addition,  
272 the noise profile cannot account for either large individual artefacts or systematic

273 offsets in the spectrum, perhaps resulting from uncertainties in the forward model  
 or errors in background correction.



**Figure 2:** SNR estimation for a number of synthetic transmission spectra, with random Gaussian noise of a given SNR added to the spectrum as shown on the x-axis, and then re-evaluated using the three different methods explained in the text as shown on the y-axis. The green line shows the optimal case where the SNR is calculated to perfect accuracy and precision. We find that method 2 tends to provide the most accurate estimates for SNR values of  $\sim 10$ -100, but method 3 can provide reasonable estimates over a much wider range of SNR values.

274

### 275 **3. Retrieval model**

#### 276 *3.1. Gauss-Newton iteration scheme*

277 The aim of a retrieval algorithm is to find a set of atmospheric parameters  
 278 for which the forward model results in a set of simulated transmission values  
 279 that provides the best fit to the observed transmission data. In practice, this is  
 280 complicated by the presence of noise and systematics in the spectrum and by the

281 fact that the problem does not always have a unique solution. For instance, a  
 282 gas profile that oscillates wildly over a characteristic length that is smaller than  
 283 the vertical resolution will result in the same observed spectrum as a gas profile  
 284 that varies smoothly with altitude. Indeed, if too little constraint is placed on the  
 285 retrieval this is exactly what will occur, in a process known as *ill-conditioning*  
 286 [2]. We therefore implement a Bayesian optimal estimation approach that relies  
 287 on some prior constraints on the atmospheric parameters we wish to retrieve. As  
 288 we essentially follow a similar procedure to the one described by equations 2.16 -  
 289 2.31 in Rodgers et al. [17], we will only summarise the main steps here and leave  
 290 the reader to consult said work for a more detailed derivation of each equation.

291 Let  $x$  be a state vector and  $y$  be a vector of observed transmission values (the  
 292 *measurement vector*), so that  $P(y|x)$  describes the likelihood function of the trans-  
 293 mission values calculated using the forward model. To provide some constraint  
 294 on the retrieval, we provide an initial state vector (also known as the *prior state*  
 295 *vector*) consisting of educated guesses of each of the scientific and instrumental  
 296 variables to be retrieved. In addition, we provide a corresponding set of *prior co-*  
 297 *variance* values that estimate the degree of certainty on each of the initial guessed  
 298 values and the degree to which they correlate with each other. These prior state  
 299 vector and covariance values therefore describe an *a priori* probability function  
 300  $P(x)$ . We can therefore easily derive an *a posteriori* probability function  $P(x|y)$ ,  
 301 that is to say the probability of obtaining a set of instrumental and scientific vari-  
 302 ables given a set of observed transmission values, according to Bayes' theorem:

$$P(x|y)P(y) = P(y|x)P(x) \quad (17)$$

303 where  $P(y)$  is a normalisation factor.

304 Let  $\mathbf{S}_y$  be a covariance matrix where the diagonal values give the estimated  
 305 spectral uncertainty as calculated according to the signal-to-noise estimation method

306 previously described in section 2.6 (we generally do not assume spectral correla-  
 307 tion and so set non-diagonal values to zero). If the probability functions are all  
 308 assumed to follow a Gaussian distribution, it can be shown that the most likely  
 309 value of  $x$ , that is to say the maximum value of  $P(x|y)$ , involves minimising a cost  
 310 function  $\phi$ :

$$\phi = (\mathbf{y} - \mathbf{K}_n \mathbf{x}_n)^T \mathbf{S}_y^{-1} (\mathbf{y} - \mathbf{K}_n \mathbf{x}_n) + (\mathbf{x}_n - \mathbf{x}_0)^T \mathbf{S}_0^{-1} (\mathbf{x}_n - \mathbf{x}_0) \quad (18)$$

311 The left-hand side ensures a good fit to the spectrum by minimising the differ-  
 312 ence between the measurement vector  $\mathbf{y}$  and the forward model  $\mathbf{K}_n \mathbf{x}_n$  computed  
 313 following the  $n$ th iteration of the retrieval, while the right-hand side prevents ill-  
 314 conditioning by using the prior state vector  $\mathbf{x}_0$  (of covariance  $\mathbf{S}_0$ ) as a constraint  
 315 on the optimal retrieved state vector solution  $\mathbf{x}_n$ .

316 Nonetheless, if the optimal state vector is too far from the prior, the problem  
 317 becomes non-linear. Assuming low non-linearity, the optimal solution can be  
 318 found through a Gauss-Newton iteration scheme:

$$\mathbf{x}_{i+1} = \mathbf{x}_0 + \mathbf{S}_0 \mathbf{K}_i^T (\mathbf{K}_i \mathbf{S}_0 \mathbf{K}_i^T + \mathbf{S}_y)^{-1} (\mathbf{y} - \mathbf{K}_i \mathbf{x}_i + \mathbf{K}_i (\mathbf{x}_i - \mathbf{x}_0)) \quad (19)$$

319 and the *a posteriori* covariance on the final state vector can then be found as  
 320 follows:

$$\mathbf{S}_n = (\mathbf{S}_0^{-1} + \mathbf{K}_n^T \mathbf{S}_y^{-1} \mathbf{K}_n)^{-1} \quad (20)$$

321 where the derivation of the Jacobian matrix  $\mathbf{K}_n = \frac{d\mathbf{y}}{d\mathbf{x}_n}$  is described in the Ap-  
 322 pendix for each variable type.

323 The quality of each forward model is quantified by a reduced  $\chi^2$  parameter,  
 324 which is equal to the left-hand side of the cost function described in equation 18:

$$\frac{\chi^2}{n_y} = \frac{1}{n_y} (\mathbf{y} - \mathbf{K}_i \mathbf{x}_i)^T \mathbf{S}_y^{-1} (\mathbf{y} - \mathbf{K}_i \mathbf{x}_i) \quad (21)$$

325 where  $n_y$  is the total number of data points. Ideally, the value of the reduced  
 326  $\chi^2$  parameter should decrease with each iteration of equation 19, however this  
 327 may not be the case if the problem is highly non-linear. Hence, if the reduced  $\chi^2$   
 328 increases after a single iteration, a step damping procedure is conducted where a  
 329 fraction of the nominal step  $\mathbf{x}_{i+1} - \mathbf{x}_i$  is chosen that minimises the reduced  $\chi^2$ .

330 The iteration scheme ends either when a local minimum in the reduced  $\chi^2$  is  
 331 found, when the maximum number of iterations has been exceeded, or when the  
 332 change between two subsequent forward models is below a threshold value  $\sqrt{F_{tol}}$ :

$$\frac{1}{n_y} (\mathbf{K}_{i+1} \mathbf{x}_{i+1} - \mathbf{K}_i \mathbf{x}_i)^T \mathbf{S}_i^{-1} (\mathbf{K}_{i+1} \mathbf{x}_{i+1} - \mathbf{K}_i \mathbf{x}_i) < F_{tol} \quad (22)$$

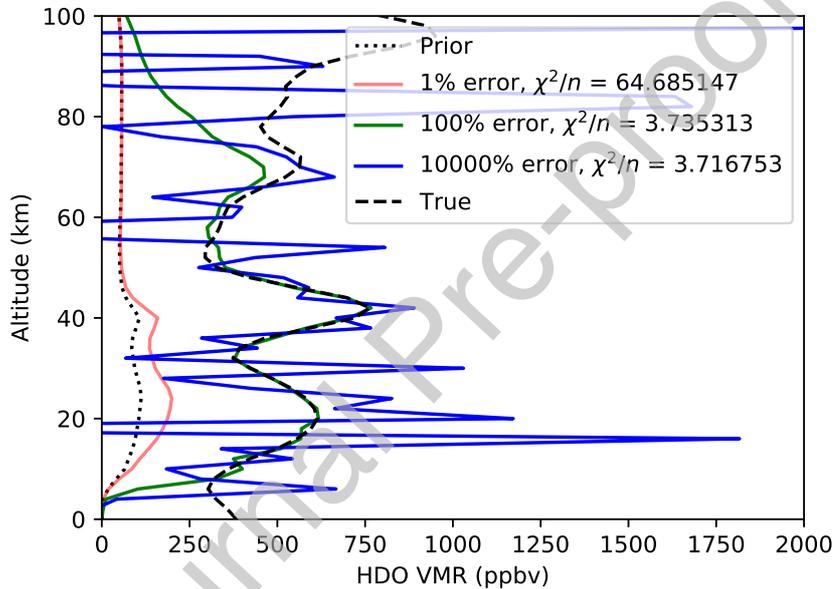
### 333 3.2. Constraints on the prior likelihood function

334 In this model, we are dealing with two separate aims, as previously described  
 335 in Section 1, that require a different treatment of the prior likelihood function  
 336 described by  $\mathbf{x}_0$  and  $\mathbf{S}_0$ . The first aim is to constrain vertical profiles of gases that  
 337 are known to exist in the atmosphere. In this case, gas abundance profiles extracted  
 338 from theoretical models usually provide an adequate prior. Empirically we find  
 339 that the standard deviation on the prior abundance of each gas  $m$  can usually be  
 340 set to equal around 100% as shown in Figure 3. Non-diagonal covariances are  
 341 derived by convolving the diagonal terms of the covariance matrix ( $\sigma(z)$ ) with a  
 342 Gaussian kernel:

$$S_{0m(ij)} = \sigma_i \exp\left(\frac{-4 \ln(2)(z_i - z_j)^2}{\rho^2}\right) \sigma_j \quad (23)$$

343 where the term  $\rho$  is referred to as a ‘correlation length’ (in km), a higher value  
 344 of which results in a smoother profile. The value of this term should be set accord-

345 ing to the vertical sampling of the occultation data. The second aim is to derive  
 346 stringent upper limits in the case of a non-detection of a particular gas species.  
 347 Since these species are undetected, we have zero prior information on their abun-  
 348 dances by definition. In this case, we wish to make the prior abundances equal  
 349 to 0 at all altitudes, with diagonal variances  $S_{0m(ii)}$  made as large as possible and  
 350 non-diagonal covariances  $S_{0m(i\neq j)}$  derived through a similar kernel convolution.



**Figure 3:** The effect of changing the prior uncertainty on a retrieved profile of HDO to best fit a set of synthetic spectra produced by a forward model of a synthetic ‘true’ profile of HDO (black, dashed). In the case of the retrieved profile marked in red, the prior uncertainty is only 1%, resulting in insufficient flexibility in the HDO profile to enable it to fit the observed spectra and hence a very poor  $\chi^2/n$  value. In the case of the retrieved profile in blue, you have the opposite extreme, where the prior error is 10,000% and for which there is insufficient constraint on the retrieval to converge to a physical solution, resulting in an ill-conditioned, highly oscillating profile. A 100% prior error results in the most realistic profile that provides the best fit to the occultation.

351 In the case of a non-detection of a gaseous species, the probability distribution  
 352 of retrieved abundances should approximate a Gaussian around a mean volume  
 353 mixing ratio of zero. The standard deviation would roughly equal the instrumen-

tal detection limit of the species, which itself would be dependent on both the presence of noise in the spectrum and the sensitivity of the spectrum to that particular species. A consequence of this Gaussian probability assumption is that the retrieved abundance of a gas that is not present in the atmosphere would be equally likely to be negative as it would to be positive. Negative gas abundances are of course unphysical, but are nonetheless useful when assessing the sensitivity of a spectrum to a particular gas. A retrieved VMR that is negative is usually an indication either of ill-conditioning or of a genuine lack of sensitivity at a given altitude. For this reason, we do not choose to restrict our gas volume mixing ratios to be exclusively positive by performing retrievals in logarithmic space. This also removes a source of bias and allows smoother convergence to a solution.

## 4. Validation

### 4.1. Comparison with the NEMESIS radiative transfer and retrieval tool

We validated the retrieval model by generating a set of synthetic spectra for a Mars-like atmosphere using the forward model of a well-established radiative transfer and retrieval tool known as NEMESIS [2, 18], and then attempting to retrieve them using RISOTTO starting from a relatively distant prior. In this section we present two sets of retrievals that demonstrate the accuracy of RISOTTO: one of vertical abundance profiles given a fixed temperature-pressure profile, and one of temperature given a fixed CO<sub>2</sub> volume mixing ratio and base pressure value. An arbitrary ILSF and linear offset value of the wavenumber grid is specified in the NEMESIS forward model itself, and then we add random Gaussian noise corresponding to a deliberately high SNR (~100,000). However, NEMESIS is not currently designed for automatic transmission baseline correction, and can only output and retrieve spectra with a flat baseline. For both sets of retrievals, we therefore compare two scenarios: one in which the transmission baseline is cali-

380 brated perfectly to unity at all wavenumbers, and one in which the transmission  
 381 baseline is distorted in order to test the continuum subtraction model. In the latter  
 382 case, we perform our own low-frequency baseline distortion using a set of ran-  
 383 dom values before then attempting the retrieval with RISOTTO. This distortion  
 384 model takes three separate effects into account: a) a polynomial component  $P(\lambda')$   
 385 representing simple baseline distortions due to calibration errors, b) a sinusoidal  
 386 component  $S(\lambda')$  representing distortions in the continuum due to the presence of  
 387 broad aerosol absorption features, and c) an exponential component  $A(z_T)$  repre-  
 388 senting attenuation along the line of sight at a given tangent height  $z_T$ , due to an  
 389 aerosol layer of fractional scale height  $z_{scale}$  present below an altitude  $z_{knee}$ :

$$T'(\lambda, z_T) = T(\lambda, z_T)A(z_T)[P(\lambda') + S(\lambda')] \quad (24)$$

390 Definitions are as follows:

$$P(\lambda') = 1 + a_0 + a_1\lambda' + a_2\lambda'^2 \quad (25)$$

$$S(\lambda') = A_1A_2 \sin(k_1\lambda' + \phi_1) \sin(k_2\lambda' + \phi_2) \quad (26)$$

$$A(z_T) = \begin{cases} \exp\left(\frac{z_T - z_{knee}}{z_{scale}}\right) & z_T < z_{knee} \\ 1 & z_T > z_{knee} \end{cases} \quad (27)$$

$$\lambda' = \lambda - \bar{\lambda} + \lambda_{rand} \quad (28)$$

391 where  $\bar{\lambda}$  is the central wavenumber (in  $\text{cm}^{-1}$ ) and the coefficients  $\{a_0, a_1, a_2\}$ ,  
 392  $\{A_1, A_2\}$ ,  $\{k_1, k_2\}$ ,  $\{\phi_1, \phi_2\}$  and  $\lambda_{rand}$  are all random numbers that vary from tangent  
 393 height to tangent height. For both sets of retrievals we have chosen arbitrary values  
 394 of  $z_{knee} = 32$  km and  $z_{scale} = 15$  km.

395 In the first set of retrievals, we chose a wavenumber range of 2932 - 2938  
396  $\text{cm}^{-1}$ , and injected a volume mixing ratio of 1% water vapour and 500 ppbv of  
397 formaldehyde ( $\text{H}_2\text{CO}$ ) into the NEMESIS forward model, both constant with al-  
398 titude. At this wavelength range, we are sensitive to three prominent absorption  
399 lines of  $\text{H}_2^{16}\text{O}$  centred on  $2933.7 \text{ cm}^{-1}$ ,  $2935.1 \text{ cm}^{-1}$  and  $2936.9 \text{ cm}^{-1}$  respectively,  
400 with a smaller broad absorption feature centred on  $2933.2 \text{ cm}^{-1}$ . Several smaller  
401  $\text{H}_2^{12}\text{C}^{16}\text{O}$  absorption lines are found scattered throughout this wavelength range  
402 as well, including a line that almost directly overlaps with the  $2933.7 \text{ cm}^{-1}$   $\text{H}_2^{16}\text{O}$   
403 absorption feature. The cross-section values for these gases were calculated ac-  
404 cording to the HITRAN 2016 database [8]. We assumed a fixed isothermal profile  
405 at a temperature of 180 K and a reference pressure at the planetary surface, esti-  
406 mated by the Mars Climate Database [19] given an arbitrary set of spatio-temporal  
407 variables, to be 5.46 mbar. The pressure values were then extrapolated to higher  
408 altitudes according to equation 5. The atmosphere was assumed to have a con-  
409 stant molar mass of  $43.34 \text{ g mol}^{-1}$ . Since the gaseous abundances are degenerate  
410 with respect to both temperature and pressure, we chose to completely fix the  
411 temperature-pressure profile in the RISOTTO model to the aforementioned val-  
412 ues, and only allow the instrumental variables and the molecular abundance pro-  
413 files to vary in the model, starting from a prior of 1%  $\text{H}_2\text{O}$  and 100 ppbv of  $\text{H}_2\text{CO}$   
414 at all altitudes. We chose to generate spectra with NEMESIS at tangent height  
415 values regularly spaced 1 km apart between 0 km and 20 km, and 2 km apart  
416 between 20 km up to the top of the atmosphere at 122 km. This tangent height  
417 grid spacing was chosen in order to reconcile speed and memory constraints with  
418 regards to NEMESIS on one hand, and a broad and regular vertical sampling of  
419 the atmosphere on the other hand, with the narrower spacing below 20 km chosen  
420 in order to better sample the vertical pressure variation which is taken into ac-  
421 count differently in NEMESIS. For comparison, we also ran an additional forward

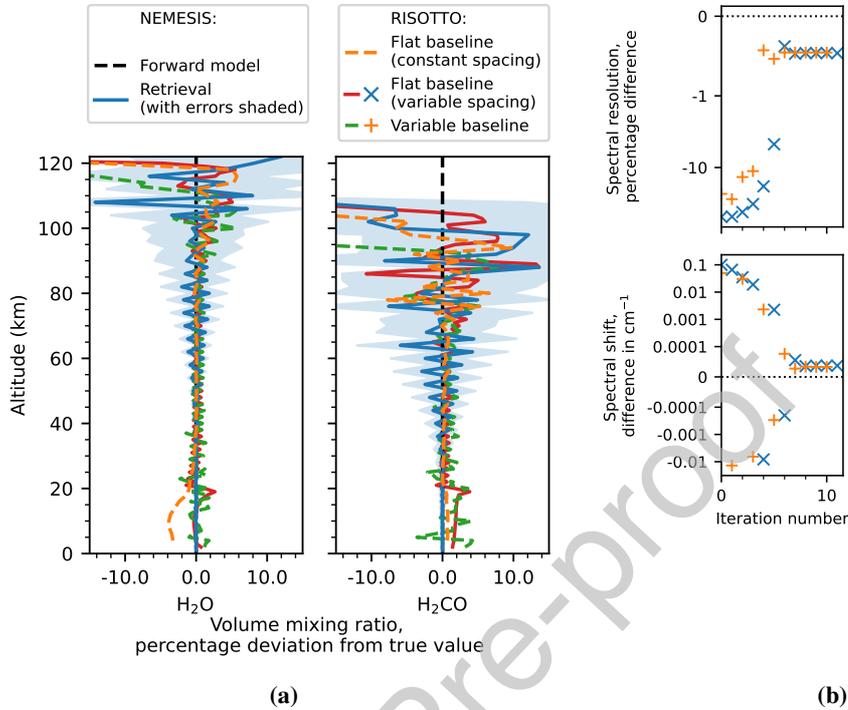
422 model using NEMESIS with tangent heights spaced apart at 2 km intervals over  
423 the whole vertical range, for which we attempted to retrieve the molecular abun-  
424 dance profiles using RISOTTO as in the previous cases. By contrast, the vertical  
425 altitude spacing for the abundance profiles in the RISOTTO retrieval was chosen  
426 to be 1km over the whole atmosphere in all cases. In addition, we attempted to re-  
427 retrieve the profiles using NEMESIS itself starting from the same prior values, in  
428 order to provide a visual indication of the uncertainty on the retrieved values due  
429 to the finite SNR, particularly at higher altitudes where sensitivity to molecular  
430 absorption is weak.

431 Figure 4a shows a comparison between the vertical water vapour and formalde-  
432 hyde profiles retrieved in all the aforementioned cases, with examples of the cor-  
433 responding spectral fits in figure 5. For these retrievals the correlation lengths of  
434 the molecular profiles, in both the NEMESIS and RISOTTO cases, were set delib-  
435 erately lower than they would in a real retrieval, in order to reduce the smoothness  
436 of the profiles so that the horizontal oscillations in abundance roughly reflect the  
437 *a posteriori* uncertainties on the retrieved values. In addition, there is a differ-  
438 ence in the lower boundary condition between NEMESIS and RISOTTO due to  
439 differences in layer splitting, which should only affect the retrieved abundance at  
440 the lowest altitude gridpoint, but the difference can affect higher altitudes if the  
441 profile is set to be too smooth. For altitudes above around 20 km we find that  
442 the correspondence between the original forward model from NEMESIS and the  
443 retrieved abundance profiles from RISOTTO is consistently accurate to within 2%  
444 at most. The retrieved water vapour and formaldehyde values then start becoming  
445 less precise above an altitude of around 90 km and 70 km respectively, with the  
446 difference relating to the fact that the H<sub>2</sub>CO lines are weaker and so the spectral  
447 sensitivity decreases more with increasing altitude. Above these altitudes, the er-  
448 rors in the retrieved values are dominated by noise, a lack of spectral sensitivity

449 and, for the very highest altitudes, differences in where the top of the atmosphere  
450 is defined. This is also reflected in the attempt at re-retrieval of the profiles using  
451 NEMESIS itself, with the degree of oscillation around the forward modelled value  
452 roughly equalling that of RISOTTO.

453 Allowing the continuum to be retrieved induces little error on the vertical pro-  
454 files above  $z_{knee}$ , apart from at very high altitudes where the noise level and the  
455 decreasing sensitivity to gaseous absorption increases the uncertainty on the base-  
456 line level. Below  $z_{knee}$ , some increase in uncertainty is observed, especially with  
457  $H_2CO$  where the attenuation of the signal results in a decrease in spectral sensi-  
458 tivity to the absorption lines, approximately doubling the total uncertainty at the  
459 lowest altitudes.

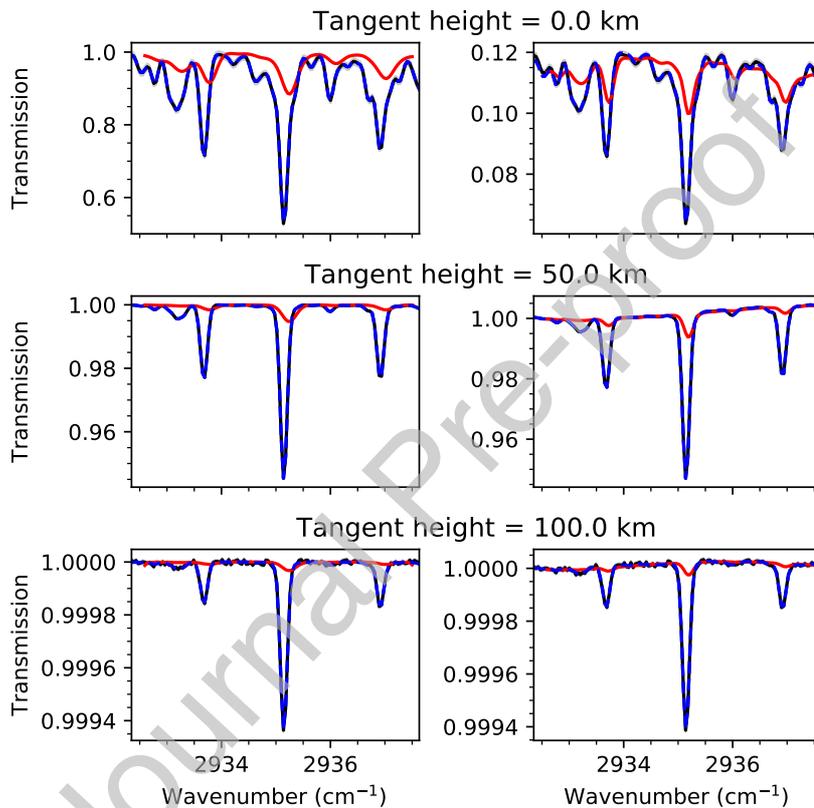
460 Below 20 km, however, there are some small systematic differences in the  
461 retrieved values between NEMESIS and RISOTTO. This is mostly due to differ-  
462 ences in how the gas absorption cross-sections are calculated at a given pressure.  
463 In NEMESIS, these are calculated through a simple linear interpolation with re-  
464 spect to cross-sections at a known pressure and temperature grid. In addition,  
465 NEMESIS calculates the line of sight integral through a layer-by-layer summa-  
466 tion, where each layer is completely homogeneous and the pressure and tem-  
467 perature values of each layer are calculated according to the Curtis-Godson ap-  
468 proximation. By contrast, in RISOTTO the calculated absorption cross-sections  
469 vary continuously with respect to altitude according to equation 12. This is more  
470 clearly illustrated in Figure 4a by the fact that decreasing the spacing of spectral  
471 observations below 20 km dramatically reduces the discrepancy between the for-  
472 ward model and the water vapour profile retrieved by NEMESIS. Nonetheless,  
473 even in the worst case the error this pressure degeneracy induces is only of the  
474 order of around 5% at most, which in practice for a real retrieval would be sub-  
475 stantially lower than the error due to noise and other systematics in the spectra.



**Figure 4:** (a) A comparison of the percentage error on the abundances of water vapour and formaldehyde retrieved from a set of spectra generated using a NEMESIS forward model, showing the percentage deviation from the ‘true’ profiles used as an input in the NEMESIS forward model. We note that, for altitudes where there is reasonable sensitivity to the retrievals, the RISOTTO profiles are true to the forward model within around 1-2%, with some greater uncertainties possible at lower altitudes where the change in absorption cross-section with pressure starts to become substantial (the spike in the red and green profiles around 20 km is due to the discontinuity in the tangent height grid). (b) Convergence of the RISOTTO flat and variable baseline retrieval cases to the NEMESIS forward modelled values of ILSF and spectral offset (black dotted lines), as a function of iteration number, where iteration number 0 is equal to the prior value. We should note that both graphs in (b) are plotted on a symmetric log-scale on the y-axis, that is to say that the values vary logarithmically except within the lowest major tick marks where they vary linearly with  $y$ . We therefore show that the ILSF converges to within around 0.5% of the ‘true’ ILSF value and within around  $5 \times 10^{-5} \text{ cm}^{-1}$  of the ‘true’ wavenumber offset value.

476 In Figure 4b we plot the convergence of the ILSF and spectral law respec-  
 477 tively as a function of iteration number for the RISOTTO flat and variable con-  
 478 tinuum cases respectively. Both cases converge within around 0.5% of the ‘true’  
 479 ILSF value from the NEMESIS forward model (equivalent to a Gaussian with a  
 480 FWHM of  $0.15 \text{ cm}^{-1}$  as specified by NEMESIS, or a spectral resolution power of

481 approximately 19550 in RISOTTO terms), and to within an order of magnitude of  
 482  $10^{-5} \text{ cm}^{-1}$  of the true wavenumber offset value. In these cases the most important  
 483 source of error would be the spectral sampling (equivalent to around  $0.05 \text{ cm}^{-1}$  in  
 484 Figure 5).

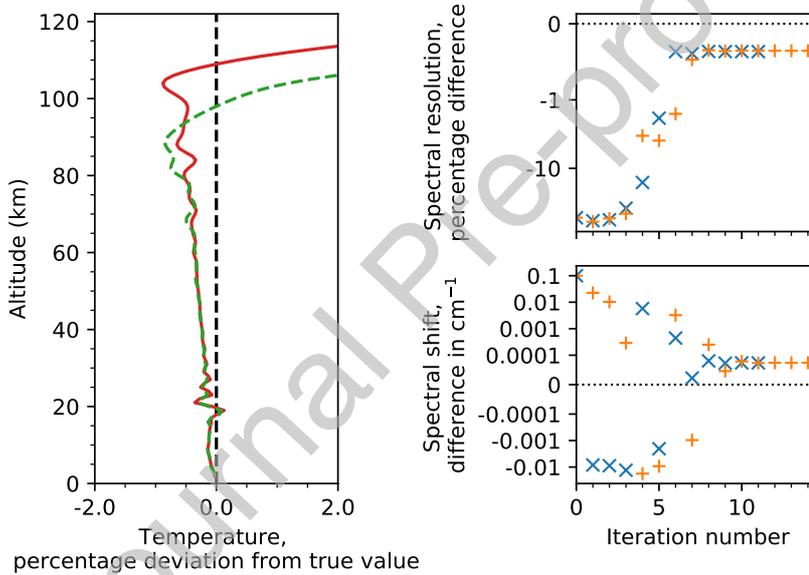


**Figure 5:** Diagram of convergence of the RISOTTO code to a synthetic spectrum (in black) generated by NEMESIS at a set of sample tangent heights above a surface value, for the set of retrievals shown in Figure 4a. For the spectra in the left column, the baseline was left unaltered and fixed in the RISOTTO retrieval (corresponding to the retrieval in red from Figure 4a), for the spectra in the right column, the continuum was distorted and retrieved (corresponding to the equivalent retrieval in green from Figure 4a). The prior RISOTTO forward model, in red, starts from  $\text{H}_2\text{O}$  and  $\text{H}_2\text{CO}$  volume mixing ratios a factor of 10 and 5 respectively below the ‘true’ NEMESIS forward model values, as well as a spectral resolution of  $R \sim 10,000$  (vs approximately 19,550 from the NEMESIS forward model) and a constant spectral shift of  $0.1 \text{ cm}^{-1}$  to the right of the NEMESIS wavenumber grid. We can see a good convergence, in the blue dashed line, to the true spectrum in both cases, with both the true and the retrieved spectral fits barely distinguishable.

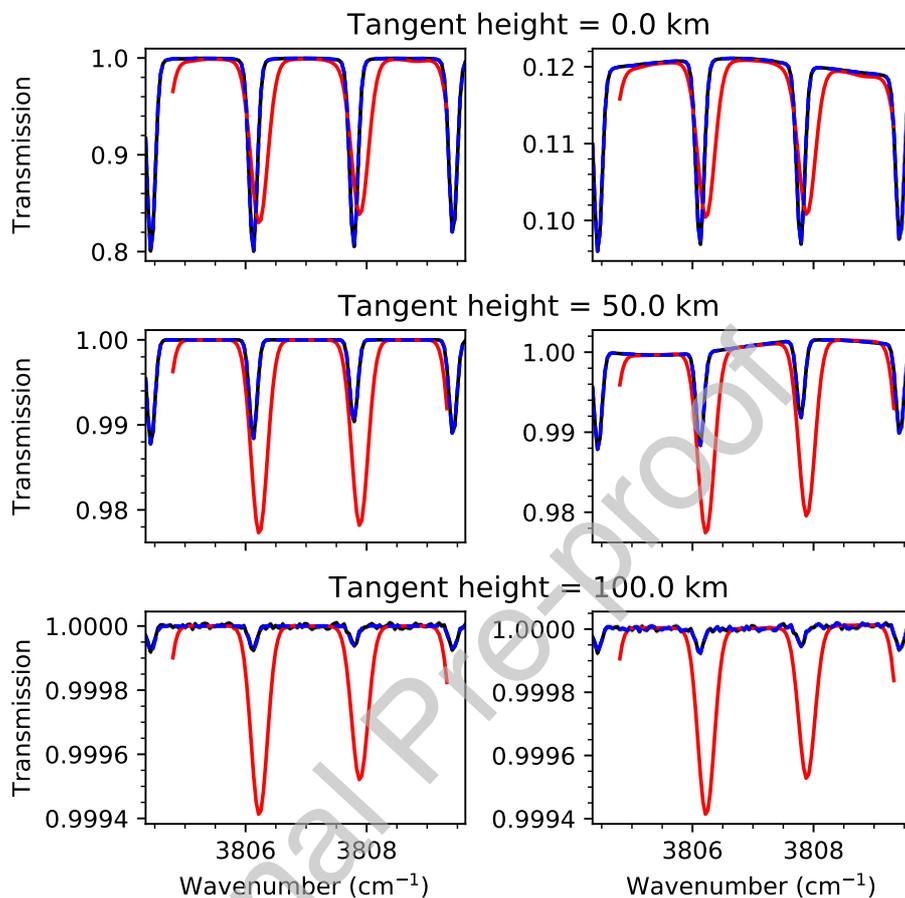
485 In order to validate the temperature retrieval, we chose to perform a second for-  
486 ward model of a set of CO<sub>2</sub> lines using NEMESIS between a wavenumber range of  
487 3804 - 3810 cm<sup>-1</sup> and then retrieve the temperature profile again using RISOTTO.  
488 We used CO<sub>2</sub> as a proxy for a gas with a known abundance that is well-mixed in  
489 the atmosphere, and so for which a temperature profile can be retrieved without  
490 worrying about degeneracies between abundance and temperature. We assumed  
491 an atmosphere with a constant volume mixing ratio of 95% CO<sub>2</sub> with altitude,  
492 and a fixed reference pressure value of 5.46 mbar at the planetary surface as with  
493 the previous set of retrievals. These were all made invariable in the RISOTTO  
494 retrieval model as well. Given these constraints, we modelled an isothermal at-  
495 mosphere of 180 K using NEMESIS, and then offset the wavenumber grid, added  
496 noise and distorted the transmission baseline as before. The only scientific vari-  
497 able that was therefore allowed to vary in the RISOTTO retrievals was the vertical  
498 temperature profile (with instrumental variables retrieved as before), for which we  
499 started with an isothermal prior profile of 210 K.

500 Figure 6 shows the resulting retrieved profiles from RISOTTO, in both flat and  
501 variable continuum cases. Although there is a consistent increase in error with alti-  
502 tude due to small differences in how the density scale height is calculated between  
503 RISOTTO and NEMESIS, we still find that below around 100 km the profiles re-  
504 trieved with RISOTTO are consistently within about 1% of the NEMESIS forward  
505 modelled values, with the error tending towards the prior only at higher altitudes  
506 where sensitivity to the spectral lines begins to be lost. The systematic difference  
507 in the upper atmosphere can of course be mitigated by increasing the altitude at  
508 which the reference pressure value is fixed, although this will of course come at the  
509 price of lesser accuracy at lower altitudes. The disadvantage of this method is that  
510 it does usually require a single pressure value to be known *a priori*. Nonetheless  
511 this is a common assumption made in atmospheric retrievals where there is not

512 enough information to completely disambiguate temperature and pressure (e.g.  
 513 [20]). The ILSF and spectral law were observed to converge to within a similar  
 514 margin of error as in the previous set of retrievals of water vapour and formalde-  
 515 hyde abundance. Differences in the retrieved temperature profile between the flat  
 516 and variable continuum cases were seen to be negligible below around 90 km,  
 517 above which point the noise level and the decreasing sensitivity to the CO<sub>2</sub> lines  
 518 induces greater error on the location of the baseline. In any case we find a good  
 519 fit to the spectra in both cases below said altitudes as shown in figure 7.



**Figure 6:** (left) Percentage error in the retrieved temperature profile from RISOTTO compared with the original NEMESIS forward model (black, dashed), assuming a surface pressure value of 5.46 mbar and a ‘true’ isothermal temperature profile of 180K at all altitudes. The RISOTTO retrievals in this case started from a prior isothermal temperature profile of 210 K at all altitudes. The red profile relates to the RISOTTO retrieval for a flat continuum while the green profile relates to the equivalent retrieval for a distorted baseline. We observe that the RISOTTO retrievals are within 1% error of the NEMESIS forward modelled values below 100 km, with a slight increase in error with altitude from the surface due to differences in the density scale height. On the right is shown the convergence of the RISOTTO flat (blue crosses) and variable (orange crosses) continuum retrievals to the NEMESIS forward modelled values of ILSF and spectral shift (black dotted lines, with both modelled values equivalent to those shown in figure 4b) as a function of iteration number. We find a similarly accurate retrieval of both quantities compared with the case in figure 4b.

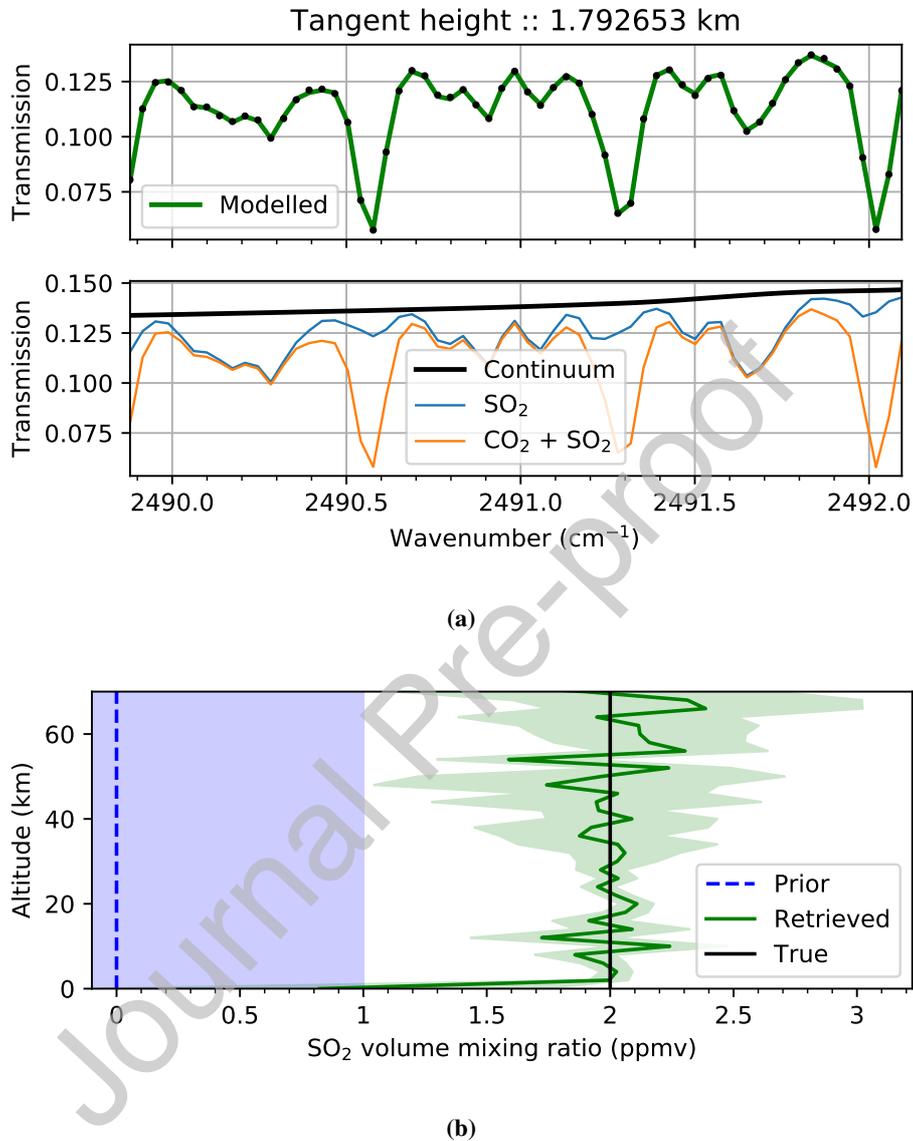


**Figure 7:** Fit to the NEMESIS forward modelled spectra (black) achieved by RISOTTO for the temperature retrievals shown in figure 6, (left) with a flat unity baseline and (right) with a distorted baseline, at a set of sample tangent heights. All absorption lines shown in this plot are due to the presence of  $^{12}\text{C}^{16}\text{O}^{16}\text{O}$ . The prior fit to the spectra is shown in red in both cases, for a prior isothermal temperature profile of 210K, a prior spectral resolution of  $R \sim 10,000$  (vs approximately 19,550 from the NEMESIS forward model) and a constant spectral shift of  $0.1 \text{ cm}^{-1}$  to the right of the NEMESIS wavenumber grid. As with Figure 5 we find a barely distinguishable fit of the RISOTTO model (in blue, dashed) to the NEMESIS forward model.

## 520 4.2. Tests on molecules with continuum absorption

521 One major advantage in using this radiative transfer and retrieval pipeline over  
522 many existing codes, including the aforementioned NEMESIS, is that the trans-  
523 mission baseline can be retrieved iteratively as a state vector parameter in the  
524 model, without having to pre-process the spectra to estimate and fix the baseline  
525 level before retrievals of scientific information from the data can be carried out.  
526 This is important when performing retrievals in spectral regions that have multiple  
527 overlapping lines that are either very saturated, or are so dense that they cannot  
528 be resolved from each other at the given spectral resolution, and so where there is  
529 an absence of regions of the spectrum that are transparent to gaseous absorption  
530 which would allow the baseline level to be known *a priori*. This can lead to a  
531 degeneracy between the abundance of the gaseous compounds that we are aiming  
532 to retrieve and the baseline level, and hence inaccurate gas retrievals.

533 Here, we use SO<sub>2</sub> as an example of a gaseous compound that has substantial  
534 continuum absorption in the infrared wavenumber range, and also exhibits ab-  
535 sorption features that overlap strongly with those of CO<sub>2</sub>. While the presence of  
536 SO<sub>2</sub> in the atmosphere of Venus is well-established, accurate measurements in this  
537 wavenumber region previously required a very complex procedure to fully decou-  
538 ple CO<sub>2</sub>, SO<sub>2</sub> and aerosol extinction [21, 22]. In Figure 8 we show the results of a  
539 retrieval of a synthetic spectrum with a distorted continuum and with a ‘true’ SO<sub>2</sub>  
540 volume mixing ratio of 2 ppmv at all altitudes, starting from a prior SO<sub>2</sub> value of  
541 0 ppmv. We can see that we can faithfully decouple the SO<sub>2</sub> values from varia-  
542 tions in the baseline, with only some 10% variation around the true value. This  
543 variation is due to ill-conditioning resulting from a retrieval starting from a prior  
544 that assumes no SO<sub>2</sub> at all, as would be the case in an atmosphere such as Mars  
545 where SO<sub>2</sub> remains undetected as of writing.



**Figure 8:** An example retrieval of a vertical profile of  $\text{SO}_2$  from a synthetic set of spectra. In figure (a) we show, on the top row, the combined fit (in green) of all gaseous species to a measured spectrum (in black) at a given tangent height. On the bottom row we show the breakdown of the spectral contribution of each gaseous species to the spectrum, together with the retrieved continuum level, noting the considerable offset between both spectra and the continuum level due to the presence of continuum absorption of  $\text{SO}_2$ . Nonetheless, as can be seen in figure (b), the retrieved  $\text{SO}_2$  abundance profile is within error of the 'true' profile even though the prior is far from the true value.

## 546 **5. Conclusion**

547 In this paper we describe a new retrieval algorithm with an implemented for-  
548 ward model, that is generalisable to spectra obtained in solar occultation geometry.  
549 Using synthetic Martian spectra, we show how it could be used to retrieve sensitive  
550 vertically-resolved profiles of multiple different gaseous compounds, while also  
551 simultaneously correcting for uncertainties in instrumental calibration artefacts  
552 and the presence of broad aerosol absorption. This is particularly advantageous  
553 in retrievals of gaseous species with near-saturated lines or substantial continuum  
554 absorption at the instrumental resolution . This model is shown to be accurate to  
555 within 1-2% of an existing, well-established radiative transfer and retrieval tool,  
556 while also being faster and simpler to use.

557 To summarise, we note the following key features of the model:

- 558 • The model is able to retrieve several vertical gaseous abundance profiles  
559 together with temperature and pressure where applicable, including when  
560 multiple gases overlap, while correcting for uncertainties in instrumental  
561 parameters such as the transmission baseline, the line spread function and  
562 small errors in the spectral registration, all simultaneously.
- 563 • The model is suitable both for vertical gas profile retrievals where high ac-  
564 curacy is required, and for detection of as yet undetected trace species where  
565 no prior information is available.
- 566 • The model can automatically provide a good estimate of the signal-to-noise  
567 ratio of a given spectrum without any prior knowledge of the noise profile.

## 568 **Acknowledgements**

569 The original code was first developed and conceptualised by Stéphane Fer-  
570 ron, with Ashwin Braude responsible for major changes to the model in order to

571 best tailor it for the available data, as well as for benchmarking the model against  
572 NEMESIS. All authors acknowledge funding from the Centre Nationale d'Études  
573 Spatiales (CNES). This paper relies on data from the TGO ACS/MIR instrument  
574 which was developed jointly by the Space Research Institute (IKI) in Moscow,  
575 Russia, and the Laboratoire Atmosphères, Milieux, Observations Spatiales (LAT-  
576 MOS/CNRS) in Paris, France, and received funding from Roscosmos, CNES and  
577 the Russian Ministry of Science and Education. We thank Juan Alday Parejo for  
578 his cross-section data used to validate the model and with his aid in the usage of  
579 NemesisSO, as well as Kevin Olsen for his advice on occultation retrievals in the  
580 infrared.

## 581 References

- 582 [1] G. R. Smith, D. M. Hunten, Study of planetary atmospheres  
583 by absorptive occultations, *Rev. Geophys.* 28 (1990) 117–143.  
584 doi:10.1029/RG028i002p00117.
- 585 [2] P. G. J. Irwin, N. A. Teanby, R. De Kok, L. N. Fletcher, C. J. A. Howett,  
586 C. C. C. Tsang, C. F. Wilson, S. B. Calcutt, C. A. Nixon, P. D. Par-  
587 rish, The NEMESIS Planetary Atmosphere Radiative Transfer and Re-  
588 trieval Tool, *J. Quant. Spec. Radiat. Transf.* 109 (6) (2008) 1136–1150.  
589 doi:10.1016/j.jqsrt.2007.11.006.
- 590 [3] K. S. Olsen, F. Lefèvre, F. Montmessin, A. A. Fedorova, A. Trokhimovskiy,  
591 L. Baggio, O. Korablev, J. Alday, C. F. Wilson, F. Forget, D. A. Belyaev,  
592 A. Patrakeev, A. V. Grigoriev, A. Shakun, The vertical structure of CO in  
593 the Martian atmosphere from the ExoMars Trace Gas Orbiter, *Nat. Geosci.*  
594 14 (2021) 67–71. doi:10.1038/s41561-020-00678-w.
- 595 [4] F. Montmessin, O. Korablev, A. Trokhimovskiy, A. Fedorova, F. Lefèvre,

- 596 J.-L. Bertaux, L. Baggio, K. Olsen, A. Shakun, A stringent upper limit of  
597 20 pptv for methane on Mars and constraints on its dispersion outside Gale  
598 crater, *A&A* 650 (2021) A140. doi:10.1051/0004-6361/202140389.
- 599 [5] X. Xu, J. Cisewski-Kehe, A. B. Davis, D. A. Fischer, J. M. Brewer, Model-  
600 ing the Echelle Spectra Continuum with Alpha Shapes and Local Regression  
601 Fitting, *AJ* 157 (6) (2019) 243. doi:10.3847/1538-3881/ab1b47.
- 602 [6] M. G. Genton, Classes of kernels for machine learning: a statistics perspec-  
603 tive, *J. Mach. Learn. Res.* 2 (2001) 299–312.
- 604 [7] J. Humlíček, An efficient method for evaluation of the complex probabil-  
605 ity function: The Voigt function and its derivatives, *J. Quant. Spec. Ra-  
606 diat. Transf.* 21 (4) (1979) 309 – 313. doi:10.1016/0022-4073(79)90062-1.
- 607 [8] I. E. Gordon, L. S. Rothman, C. Hill, R. V. Kochanov, Y. Tan, P. F. Bernath,  
608 M. Birk, V. Boudon, A. Campargue, K. V. Chance, et al., The HITRAN2016  
609 molecular spectroscopic database, *J. Quant. Spec. Radiat. Transf.* 203 (2017)  
610 3 – 69. doi:https://doi.org/10.1016/j.jqsrt.2017.06.038.
- 611 [9] A. Bellucci, B. Sicardy, P. Drossart, P. Rannou, P. D. Nicholson, M. Hedman,  
612 K. H. Baines, B. Burrati, Titan solar occultation observed by Cassini/VIMS:  
613 Gas absorption and constraints on aerosol composition, *Icarus* 201 (2009)  
614 198–216. doi:10.1016/j.icarus.2008.12.024.
- 615 [10] G. Liuzzi, G. L. Villanueva, M. J. Mumma, M. D. Smith, F. Daerden, B. Ris-  
616 tic, I. Thomas, A. C. Vandaele, M. R. Patel, J.-J. Lopez-Moreno, et al.,  
617 Methane on Mars: New insights into the sensitivity of CH<sub>4</sub> with the NO-  
618 MAD/ExoMars spectrometer through its first in-flight calibration, *Icarus*  
619 321 (2019) 671–690. doi:10.1016/j.icarus.2018.09.021.

- 620 [11] R. J. de Kok, D. M. Stam, The influence of forward-scattered light in trans-  
621 mission measurements of (exo)planetary atmospheres, *Icarus* 221 (2012)  
622 517–524. doi:10.1016/j.icarus.2012.08.020.
- 623 [12] R. Gorenflo, S. Vessella, Abel Integral Equations, Vol. 1461 of Lecture Notes  
624 in Mathematics, Springer, 1991.
- 625 [13] S. Takagi, A. Mahieux, V. Wilquet, S. Robert, A.-C. Vandaele, N. Iwagami,  
626 An uppermost haze layer above 100 km found over Venus by the SOIR  
627 instrument onboard Venus Express, *Earth Planets Space* 71 (2019) 124.  
628 doi:10.1186/s40623-019-1103-x.
- 629 [14] O. Korablev, F. Montmessin, A. Trokhimovskiy, A. A. Fedorova, A. V.  
630 Shakun, A. V. Grigoriev, B. E. Moshkin, N. I. Ignatiev, F. Forget, F. Lefèvre,  
631 et al., The Atmospheric Chemistry Suite (ACS) of Three Spectrometers for  
632 the ExoMars 2016 Trace Gas Orbiter, *Space Sci. Rev.* 214 (1) (2018) 7.  
633 doi:10.1007/s11214-017-0437-6.
- 634 [15] B. R. Sandel, H. Gröller, R. V. Yelle, T. Koskinen, N. K. Lewis,  
635 J. L. Bertaux, F. Montmessin, E. Quémerais, Altitude profiles of O<sub>2</sub>  
636 on Mars from SPICAM stellar occultations, *Icarus* 252 (2015) 154–160.  
637 doi:10.1016/j.icarus.2015.01.004.
- 638 [16] F. Montmessin, J. L. Bertaux, F. Lefèvre, E. Marcq, D. Belyaev, J. C. Gérard,  
639 O. Korablev, A. Fedorova, V. Sarago, A. C. Vandaele, A layer of ozone  
640 detected in the nightside upper atmosphere of Venus, *Icarus* 216 (2011) 82–  
641 85. doi:10.1016/j.icarus.2011.08.010.
- 642 [17] C. D. Rodgers, Inverse Methods for Atmospheric Sounding: Theory and  
643 Practice, Vol. 2, World Scientific, 2000.

- 644 [18] J. Alday, C. F. Wilson, P. G. J. Irwin, K. S. Olsen, L. Baggio, F. Montmessin,  
645 A. Trokhimovskiy, O. Korablev, A. A. Fedorova, D. A. Belyaev, A. Grig-  
646 oriev, A. Patrakeev, A. Shakun, Oxygen isotopic ratios in Martian water  
647 vapour observed by ACS MIR on board the ExoMars Trace Gas Orbiter,  
648 *A&A* 630 (2019) A91. doi:10.1051/0004-6361/201936234.
- 649 [19] F. Forget, F. Hourdin, R. Fournier, C. Hourdin, O. Talagrand, M. Collins,  
650 S. R. Lewis, P. L. Read, J.-P. Huot, Improved general circulation models of  
651 the Martian atmosphere from the surface to above 80 km, *J. Geophys. Res.*  
652 104 (E10) (1999) 24155–24176. doi:10.1029/1999JE001025.
- 653 [20] E. Quémerais, J.-L. Bertaux, O. Korablev, E. Dimarellis, C. Cot,  
654 B. R. Sandel, D. Fussen, Stellar occultations observed by SPICAM  
655 on Mars Express, *J. Geophys. Res. Planets* 111 (2006) E09S04.  
656 doi:10.1029/2005JE002604.
- 657 [21] D. Belyaev, O. Korablev, A. Fedorova, J. L. Bertaux, A. C. Van-  
658 daele, F. Montmessin, A. Mahieux, V. Wilquet, R. Drummond, First  
659 observations of SO<sub>2</sub> above Venus' clouds by means of Solar Occulta-  
660 tion in the Infrared, *J. Geophys. Res. Planets* 113 (2) (2008) E00B25.  
661 doi:10.1029/2008JE003143.
- 662 [22] D. A. Belyaev, F. Montmessin, J.-L. Bertaux, A. Mahieux, A. A. Fedorova,  
663 O. I. Korablev, E. Marcq, Y. L. Yung, X. Zhang, Vertical profiling of SO  
664 2 and SO above Venus' clouds by SPICAV/SOIR solar occultations, *Icarus*  
665 217 (2) (2012) 740–751. doi:10.1016/j.icarus.2011.09.025.

666 **Appendix A. Derivatives of state vector variables**

667 *Appendix A.1. Continuum model*

668 As the corrected transmission spectrum is simply assumed to be equal to the  
 669 transmission spectrum of unity baseline multiplied by the continuum model, as  
 670 described in equation 1, the Jacobian for the continuum model is simply equal to  
 671 the transmission spectrum of unity baseline:

$$\frac{dT}{dC} = \frac{T(\lambda_{cor}, z_T)}{C(\lambda_{cor}, z_T)} \quad (\text{A.1})$$

672 *Appendix A.2. Spectral Law Derivation*

673 The Jacobians for the coefficients of the spectral law are calculated according  
 674 to equation 1:

$$\frac{dT_{cor}(\lambda_{cor})_i}{da_n} = C(\lambda_{cor})_i \sum_j \frac{dK(\lambda_{cor}, \lambda_{mod})_{ij}}{da_n} T_{mod}(\lambda_{mod})_j \quad (\text{A.2})$$

675 where the parameter  $\frac{dK(\lambda_{cor}, \lambda_{mod})_{ij}}{da_n}$  is found through finite differences.

676 *Appendix A.3. ILSF*

677 From equation 1 and assuming a Gaussian ILSF, the derivative with respect to  
 678 the spectral resolution parameter  $p_{gauss}$  can be calculated as follows:

$$\frac{dT_{cor}(\lambda_{cor})_i}{dp_x} = C(\lambda_{cor})_i \sum_j \frac{dK(\lambda_{cor}, \lambda_{mod})_{ij}}{dp_{gauss}} T_{mod}(\lambda_{mod})_j \quad (\text{A.3})$$

679 where:

$$\frac{dK_{(ij)}}{dp_{gauss}} = \frac{K_{(ij)} \left( 1 - \left( \frac{\lambda_{cor(j)} - \lambda_{mod(i)}}{\sigma} \right)^2 \right)}{N_i p_{gauss}} \quad (\text{A.4})$$

680 *Appendix A.4. Temperature*

681 From equations 11 and 15 we find that:

$$\frac{d\tau_{\lambda T}}{dT_u} = \sum_{m=1}^{N_m} \left[ \frac{P_u}{Mk_B} A_{T_u} x_{mu} \left( \frac{d\sigma_{m\lambda u}}{dT_u} - \frac{\sigma_{m\lambda u}}{T_u} \right) + \sum_{v=u}^N \frac{P_v}{Mk_B T_v} A_{T_v} x_{mv} k_{m\lambda v} \frac{d\ln(P_v)}{dT_u} \right] \quad (\text{A.5})$$

682 where the term  $\frac{d\sigma_{m\lambda u}}{dT_u}$  is solved through finite differences and  $\frac{d\ln(P_v)}{dT_u}$  is calculated  
 683 according to equation 5.

684 Hence:

$$\frac{dT_{cor}(\lambda_{cor})_i}{dT_u} = -T_{cor}(\lambda_{cor})_i \frac{d\tau_{\lambda T}}{dT_u} \quad (\text{A.6})$$

685 *Appendix A.5. Gas VMR*

686 Again, from equation 15 we find that:

$$\frac{d\tau_{\lambda T}}{dx_{mu}} \approx \frac{\rho_u}{M} A_{T_u} \sigma_{m\lambda u} \quad (\text{A.7})$$

687 and so:

$$\frac{dT_{cor}(\lambda_{cor})_i}{dx_{mu}} = -T_{cor}(\lambda_{cor})_i \frac{d\tau_{\lambda T}}{dx_{mu}} \quad (\text{A.8})$$

**Ashwin S. Braude:** Methodology, Software, Validation, Formal analysis, Writing – Original draft, Writing – Review & Editing, Visualisation. **Stéphane Ferron:** Conceptualisation, Methodology, Software, Formal analysis, Writing – Review & Editing, Visualisation. **Franck Montmessin:** Conceptualisation, Methodology, Resources, Writing – Review & Editing, Supervision, Project Administration, Funding acquisition.

Journal Pre-proof

**Declaration of interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof