

PYEXOCROSS: a Python program for generating spectra and cross-sections from molecular line lists

Jingxin Zhang ^{*}, Jonathan Tennyson ^{*} and Sergei N. Yurchenko 

Department of Physics and Astronomy, University College London, Gower Street, WC1E 6BT London, UK

Accepted 2024 April 22. Received 2024 April 2; in original form 2023 October 10

ABSTRACT

PYEXOCROSS is a Python adaptation of the EXOCROSS Fortran application, PYEXOCROSS is designed for post-processing the huge molecular line lists generated by the ExoMol project and other similar initiatives such as the HITRAN and HITEMP data bases. PYEXOCROSS generates absorption and emission stick spectra, cross-sections, and other properties (partition functions, specific heats, cooling functions, lifetimes, and oscillator strengths) based on molecular line lists. PYEXOCROSS calculates cross-sections with four line profiles: Doppler, Gaussian, Lorentzian, and Voigt profiles in both sampling and binned methods; a number of options are available for computing Voigt profiles which we test for speed and accuracy. PYEXOCROSS supports importing and exporting line lists in the ExoMol and HITRAN/HITEMP formats. PYEXOCROSS also provides conversion between the ExoMol and HITRAN data formats. In addition, PYEXOCROSS has extra code for users to automate the batch download of line list files from the ExoMol data base.

Key words: Software – Molecular Data – Exoplanets – Line Profiles – Cross-sections – ExoMol.

1. INTRODUCTION

A major motivation of modern astronomy is the desire to characterize and model the atmospheres of the many extrasolar planets discovered over the previous three decades (Tinetti, Encrenaz & Coustenis 2013; Madhusudhan 2019; Tennyson & Yurchenko 2022). Doing this requires significant quantities of spectroscopic data with hot objects requiring billions of lines to accurately reproduce spectroscopic features of their atmospheres (Yurchenko et al. 2014). The ExoMol project (Tennyson & Yurchenko 2012) is systematically constructing comprehensive molecule line lists to aid the study of the atmospheres of exoplanets and other hot objects which contain molecules such as brown dwarfs, cool stars, and photon-dominated regions of the interstellar medium. ExoMol systematically produces its data in a series of papers which are summarized in data release papers (Tennyson et al. 2016b, 2020). Similar endeavours include several well-known molecular spectroscopy data bases used by astronomers such as HITRAN (Gordon et al. 2022), HITEMP (Rothman et al. 2010; Hargreaves et al. 2019), GEISA (Jacquinet-Husson et al. 2011), CDMS (Müller et al. 2013; Endres et al. 2016), the JPL spectroscopic catalogue (Pickett et al. 1998), TheoReTS (Rey et al. 2016), and NASA Ames (Huang, Schwenke & Lee 2021). A particular characteristic of ExoMol, which is shared by HITEMP, TheoReTS, and NASA Ames, is that their line lists are designed for studies involving hot molecules. The line lists are often huge; the ExoMol data base contains in excess of 10^{12} transitions. This means for many purposes it is desirable to have a means to rapidly and flexibly post-process these transition data to provide them in forms, such as cross-sections, suitable for use in astronomical models and spectroscopic studies.

The ExoMol data base is very widely used by scientists studying exoplanets. It has been used to create opacities for a variety of exoplanet codes (see e.g. Gandhi et al. 2020; Chubb et al. 2021; Grimm et al. 2021). ExoMol data have been instrumental in many molecular detections in exoplanet atmospheres such as the recent detection of CO₂ and SO₂ in the atmosphere of WASP-39b by Rustamkulov et al. (2023) using *JWST*. ExoMol is increasingly being used for studies of cool stars and brown dwarfs, as well as in a variety of non-astronomical applications. Yurchenko, Al-Refaie & Tennyson (2018a) provided a post-processing code EXOCROSS capable of treating data in both ExoMol and HITRAN/HITEMP formats. EXOCROSS is written in Fortran which makes it efficient for processing large data sets but somewhat inflexible. For example, it is hard to take part of the current EXOCROSS code and include as a utility in some other application; PYEXOCROSS was developed in response to this issue.

The main practical challenge that confronts us is processing huge amounts of data because of the extremely large data sets for polyatomic molecular hot line lists. In particular, many molecules have more than one Transitions file in the ExoMol data base, such as the parent isotopologue ²³Na¹⁶O¹H of the molecule NaOH, its OYT5 data set due to Owens, Tennyson & Yurchenko (2021) containing an excess of 400 billion transitions stored in 90 Transitions files, giving a total file size of 463 Gb. To overcome this problem, automation and batch data processing

* E-mail: jingxin.zhang.19@ucl.ac.uk (JZ); j.tennyson@ucl.ac.uk (JT)

are important for improving efficiency and reducing errors caused by operational faults. In the PYEXOCROSS program, the approach to working with tens of billion lines and optimizing the program for high throughput involves the use of efficient parallelization and vectorization.

PYEXOCROSS is a general Python adaptation of the EXOCROSS Fortran application (Yurchenko et al. 2018a) which aims to provide the same functionalities. Like EXOCROSS, PYEXOCROSS is also designed to manage huge molecular line lists generated by the ExoMol project. PYEXOCROSS calculates spectra and spectral properties on the basis of the underlying spectroscopic data including temperature and pressure-dependent absorption and emission cross-sections, temperature-dependent cooling functions, state-dependent radiative lifetimes, oscillator strengths, and thermodynamic properties such as partition functions and specific heats. PYEXOCROSS is designed to post-process data from the ExoMol data base with the ExoMol, HITRAN, and HITEMP data bases all serving as input data bases for PYEXOCROSS. PYEXOCROSS exports the results in the ExoMol format which can also be read by effective Hamiltonian spectral simulation program PGOPHER (Western 2017). The ExoMol and HITRAN/HITEMP formats are the line lists formats that are initially offered; these data bases are characterized by very extensive wavelength coverage and well-developed data structures.

There are several similar programs available for post-processing line lists including the HITRAN interface HAPI (Kochanov et al. 2016, 2018), SPECTRAPLOT.COM (Goldenstein et al. 2017), and SPECTRA (Tennyson & Miller 1989; Tennyson, Miller & Le Sueur 1993). However these programs would struggle to process the huge line lists required for simulations of atmospheres at elevated temperatures. We note that RADIS (van den Bekerom & Pannier 2021), a fast line-by-line code for infrared emission and absorption spectra at equilibrium and non-local thermodynamic equilibrium (non-LTE) which works with HITRAN/HITEMP and ExoMol data, is designed for this purpose, as is the graphical processing unit (GPU) based opacity code of Grimm et al. (2021). PYEXOCROSS imports the input line lists in both ExoMol (Tennyson & Yurchenko 2012; Tennyson et al. 2016b) and HITRAN (Rothman et al. 2005) formats and returns data in the ExoMol format. Future development will support more data base formats in PYEXOCROSS.

PYEXOCROSS performs a variety of functions. It can produce molecular absorption and emission cross-sections on a grid at a given set of temperatures and pressures (see Section 2.8); PYEXOCROSS can apply a range of line profiles namely Gaussian, Lorentzian, and Voigt profiles (see Section 3). PYEXOCROSS also implements the computations of other useful functions including partition functions (Section 2.2), specific heats (Section 2.3), cooling functions (Section 2.4), radiative lifetimes (Section 2.5), oscillator strengths (Section 2.6), and line intensities in the form of spectra (Section 2.7). PYEXOCROSS supports converting data formats between the ExoMol and HITRAN data bases (Section 2.11). PYEXOCROSS provides some additional functionalities according to the ExoMol data base including automating the batch download of the ExoMol line list files, finding the uncertainty available data sets, and comparing updates of the ExoMol data base (Section 3.8). The paper is organized into six sections as follows. The main functionality of PYEXOCROSS is described in Section 2. Section 3 discusses the line profile for calculating the cross-sections implemented in PYEXOCROSS. Section 4 gives the details of the calculation steps. The units used in PYEXOCROSS are specified in Section 5.2. Finally, Section 6 presents some conclusions. The PYEXOCROSS manual (<https://pyexocross.readthedocs.io/>), which is given in the supplementary data and in the PYEXOCROSS GitHub repository (<https://github.com/ExoMol/PyExoCross.git>), provides comprehensive working instructions and use case examples.

2. MAIN FUNCTIONALITY

This section introduces the main functionalities of the PYEXOCROSS program, including partition functions, specific heats, cooling functions, radiative lifetimes, oscillator strengths, intensities, stick spectra, and cross-sections. Some extra functionalities include filters, data format conversion, the batch download of line list files, data manipulation based on their uncertainty, and checking updates of the ExoMol data base. The line profiles for calculating the cross-sections will be illustrated in Section 3.

2.1 ExoMol format

ExoMol format (Tennyson, Hill & Yurchenko 2013) assumes that a line list consists of a States file (.states) and Transitions files (.trans). A States file contains the state IDs, energy term values, total degeneracies, total angular momenta, uncertainties, lifetimes, Lande- g factors, and any other quantum state descriptions. As an example, an excerpt from the recently updated States file for $^{27}\text{Al}^1\text{H}$ (Yurchenko et al. 2018b) is shown in Table 1. Transitions files consist of the upper IDs, lower IDs, the Einstein A -coefficient, and (optionally) reference transition wavenumbers. An extract of a Transitions file for $^{27}\text{Al}^1\text{H}$ is shown in Table 2.

2.2 Partition functions

Although crucial for hot molecular models, partition functions are not always straightforward to determine (Neale & Tennyson 1995; Sousa-Silva et al. 2014). Irwin (1981), Sauval & Tatum (1984), Barklem & Collet (2016), and Gamache et al. (2017) present compilations of partition functions for key astrophysical species. The partition function $Q(T)$ can be defined as the sum over states:

$$Q(T) = \sum_n g_n^{\text{tot}} e^{-c_2 \tilde{E}_n / T}, \quad (1)$$

where $g_n^{\text{tot}} = g_n^{\text{ns}}(2J_n + 1)$ is the total degeneracy. g_n^{ns} is the nuclear-spin statistical weight factor; note that ExoMol and HITRAN follow the ‘physicist’s’ convention and use the full nuclear spin degeneracy for g_n^{ns} , see Pavlenko, Yurchenko & Tennyson (2020) for a discussion of this. J_n is the corresponding total angular momentum. $c_2 = hc/k_B$ is the second radiation constant (cm K) where h is the Planck constant (erg s), c is the speed of light (cm s $^{-1}$), and k_B is the Boltzmann constant (erg K $^{-1}$). $\tilde{E}_n = E_n/hc$ is the energy term value (cm $^{-1}$). T is the temperature

Table 1. Extract from the .states States file of the $^{27}\text{Al}^1\text{H}$ line list (Yurchenko et al. 2018b).

i	\tilde{E} (cm $^{-1}$)	g	J	Unc (cm $^{-1}$)	τ (s)	+/−	e/f	State	v	$ \Lambda $	$ \Sigma $	$ \Omega $	Ecalc (cm $^{-1}$)	Source label
1	0.000000	12	0	0.000001	Inf	+	e	X(1Sigma+)	0	0	0	0	0.000000	Ma
2	1625.061501	12	0	0.000597	4.9288E−03	+	e	X(1Sigma+)	1	0	0	0	1625.069321	Ma
3	3194.213550	12	0	0.000365	2.6227E−03	+	e	X(1Sigma+)	2	0	0	0	3194.213685	Ma
4	4708.808866	12	0	0.000196	1.8633E−03	+	e	X(1Sigma+)	3	0	0	0	4708.817022	Ma
5	6170.189966	12	0	0.029900	1.4910E−03	+	e	X(1Sigma+)	4	0	0	0	6170.193041	Ma
6	7579.564189	12	0	0.250000	1.2740E−03	+	e	X(1Sigma+)	5	0	0	0	7579.564189	Ca
7	8938.046805	12	0	0.300000	1.1350E−03	+	e	X(1Sigma+)	6	0	0	0	8938.046805	Ca
8	10246.644027	12	0	0.350000	1.0410E−03	+	e	X(1Sigma+)	7	0	0	0	10246.644027	Ca
9	11506.245734	12	0	0.400000	9.7520E−04	+	e	X(1Sigma+)	8	0	0	0	11506.245734	Ca
10	12717.633797	12	0	0.450000	9.2834E−04	+	e	X(1Sigma+)	9	0	0	0	12717.633797	Ca

Notes. i : State counting number;
 \tilde{E} : Term value (in cm $^{-1}$);
 g_{tot} : Total state degeneracy;
 J : Total angular momentum quantum number;
 Unc: Estimated uncertainty of energy level (in cm $^{-1}$);
 τ : Lifetime (in s $^{-1}$);
 +/−: Total parity;
 e/f : Rotationless parity;
 State: Electronic term value;
 v : Vibrational quantum number;
 $|\Lambda|$: Absolute value of the projection of electronic angular momentum;
 $|\Sigma|$: Absolute value of the projection of the electronic spin;
 $|\Omega|$: Absolute value of the projection of the total angular momentum;
 Ecalc: Calculated state energy (in cm $^{-1}$);
 Source label: Method used to generate term value.

Table 2. Extract from the .trans Transitions file of the $^{27}\text{Al}^1\text{H}$ line lists (Yurchenko et al. 2018b).

f	i	A_{fi} (s $^{-1}$)	$\tilde{\nu}_{fi}$ (cm $^{-1}$)
594	598	9.1644e−14	3.731850
60	28	1.7183e−05	4.209495
59	27	2.1667e−05	4.325081
58	26	3.2749e−05	4.605786
57	25	4.9130e−05	4.917133
56	24	7.0449e−05	5.228273
54	23	9.6527e−05	5.537220
53	22	1.2674e−04	5.844347
51	21	1.5996e−04	6.150075

Notes. f : Upper state counting number;
 i : Lower state counting number;
 A_{fi} : Einstein A-coefficient in s $^{-1}$;
 $\tilde{\nu}_{fi}$: Transition wavenumber in cm $^{-1}$. (This column is optional).

Table 3. Specification of the .pf partition function file format.

Field	Fortran format	C format	Description
T	F8.1	%8.1f	Temperature in K
$Q(T)$	F15.4	%15.4f	Partition function (dimensionless)

Note. Fortran format: (F8.1, 1x, F15.4).

(K). PyEXOCROSS provides the functionality to compute a molecular partition function with equation (1) using the energy term values stored in the corresponding States file of a line list in question on a grid of temperatures.

In spectroscopic applications, where a partition function $Q(T)$ is required, it can be read directly from an ExoMol partition function (.pf) file; this is recommended as in some cases these partition functions have improved accuracy compared with the simple sum of over the states in the States file. If the .pf file is not available, $Q(T)$ for a given temperature can be computed using equation (1) as part of the spectral simulations.

Table 3 specifies the partition function file format and Table 4 gives an example of the partition function file. Fig. 1 displays examples of partition functions computed using the ExoMol line lists with temperatures from 0 to 5000 K. Here and elsewhere examples are based on use of the following line lists: H₂O POKAZATEL (Polyansky et al. 2018), $^{12}\text{C}^{16}\text{O}_2$ UCL-4000 (Yurchenko et al. 2020), $^{32}\text{S}^{16}\text{O}_2$ ExoAmes

Table 4. Extract from the .pf partition function file generated from the $^{27}\text{Al}^1\text{H}$ line lists (Yurchenko et al. 2018b).

T	$Q(T)$
1.0	12.0000
2.0	12.0042
3.0	12.0858
4.0	12.3885
5.0	12.9621
6.0	13.7647
7.0	14.7312
8.0	15.8065

Notes. T : Temperature in K;
 $Q(T)$: Partition function (dimensionless).

Table 5. Specification of the .cp specific heat file format.

Field	Fortran format	C format	Description
T	F8.1	%8.1f	Temperature in K
$C_p(T)$	F15.4	%15.4f	Specific heat (dimensionless)

Note. Fortran format: (F8.1, 1x, F15.4).

Table 6. Extract from the .cp specific heat file generated from the $^{27}\text{Al}^1\text{H}$ line lists (Yurchenko et al. 2018b).

T	$C_p(T)$
200.0	29.1538
201.0	29.1545
202.0	29.1553
203.0	29.1560
204.0	29.1567
205.0	29.1575
206.0	29.1583
207.0	29.1591
208.0	29.1955
209.0	29.1973

Notes. T : Temperature in K;
 $C_p(T)$: Specific heat ($\text{J K}^{-1} \text{mol}^{-1}$).

Table 7. Specification of the .cf cooling function file format.

Field	Fortran format	C format	Description
T	F8.1	%8.1f	Temperature in K
$W(T)$	ES12.4	%12.4E	Cooling function in $\text{erg (s molecule sr)}^{-1}$

Note. Fortran format: (F8.1, 1x, ES12.4).

Table 8. Extract from the .cf cooling function file generated from the $^{27}\text{Al}^1\text{H}$ line lists (Yurchenko et al. 2018b).

T	$W(T)$
1.0	1.04 219 598E-28
2.0	8.95 038 739E-25
3.0	1.82 111 455E-23
4.0	8.06 932 723E-23
5.0	1.94 455 237E-22
6.0	3.52 589 825E-22
7.0	5.56 853 398E-22
8.0	8.18 964 607E-22
9.0	1.15 275 411E-21
10.0	1.56 910 825E-21

Notes. T : Temperature in K;
 $W(T)$: Cooling function in $\text{erg (s molecule sr)}^{-1}$.

Table 9. Specification of the .trans Transitions file format.

Field	Fortran format	C format	Description
f	I12	%12d	Upper state ID
i	I12	%12d	Lower state ID
A	ES10.4	%10.4E	Einstein A-coefficient in s ⁻¹
$\tilde{\nu}_{fi}$	F15.6	%15.6f	Transition wavenumber in cm ⁻¹ (optional)

Note. Fortran format: (I12, 1x, I12, 1x, ES10.4, 1x, F15.6)

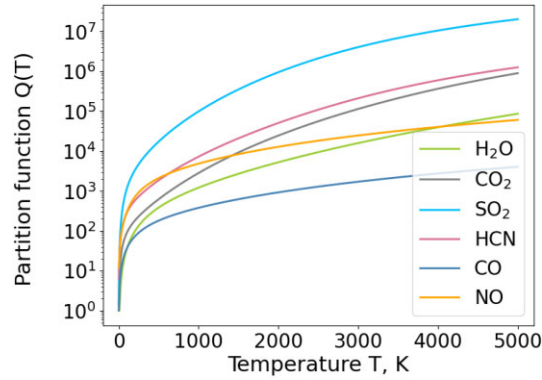


Figure 1. Partition functions of ¹H₂¹⁶O, ¹²C¹⁶O₂, ³²S¹⁶O₂, ¹H¹²C¹⁴N, ¹²C¹⁶O, and ¹⁴N¹⁶O with temperatures from 0 to 5000 K from the ExoMol data base.

(Underwood et al. 2016, Tóbiás et al. 2018), ¹H¹²C¹⁴N (Barber et al. 2014), ¹²C¹⁶O Li2015 (Li et al. 2015; Somogyi, Yurchenko & Yachmenev 2021), and NO XABC (Qu, Yurchenko & Tennyson 2021).

PYEXOCROSS provides various ways to obtain the partition function: reading the local .pf file or the ExoMol website directly, computing the partition function from the ExoMol .states, or by reading it from the HITRAN website. We note that recent analysis by Wang et al. (2023) suggests that the maximum temperature quoted by HITRAN for a converged partition function is, in many cases, too high.

2.3 Specific heats

The molar specific heat (J K⁻¹ mol⁻¹) at constant pressure C_p for a given temperature using the molecule partition function as given by Martin, François & Gijbels (1992) is defined calculated on a grid of

$$C_p(T) = R \left[\frac{Q''}{Q} - \left(\frac{Q'}{Q} \right)^2 \right] + \frac{5R}{2}, \quad (2)$$

where R is the molar gas constant ($= 8.31446261815324 \text{ J K}^{-1} \text{ mol}^{-1}$). The partition function $Q(T)$ is defined in equation (1). The first two moments of the partition function can be evaluated directly from the energy term values :

$$Q'(T) = T \frac{dQ}{dT} = \sum_n g_n^{\text{tot}} \left(\frac{c_2 \tilde{E}_n}{T} \right) \exp \left(- \frac{c_2 \tilde{E}_n}{T} \right),$$

$$Q''(T) = T^2 \frac{d^2 Q}{dT^2} + 2Q' = \sum_n g_n^{\text{tot}} \left(\frac{c_2 \tilde{E}_n}{T} \right)^2 \exp \left(- \frac{c_2 \tilde{E}_n}{T} \right). \quad (3)$$

PYEXOCROSS uses these equation in connection with the molecular States file to compute specific heats on a grid of temperatures. Table 5 lists the specific heat file format and Table 6 gives an example of the specific heat file of AlH. Fig. 2 shows the specific heats of NO, MgH, and AlH at $T \in [0, 5000]$ K.

We note that following the work of Wang et al. (2023) a near complete set of specific heats are now available on the ExoMol website.

2.4 Cooling functions

Cooling functions play an important role in a whole range of astronomical environments ranging from the early Universe to the interstellar medium, star formation, and issues with planetary stability (Coppola, Lodi & Tennyson 2011; Miller et al. 2013; Wang et al. 2014; Nguyen et al. 2022; Gérard et al. 2023). The temperature-dependent cooling function $W(T)$ [units erg (s molecule sr)⁻¹ = 10⁻⁷ Watts (molecule sr)⁻¹] is the total energy per unit solid angle (in steradians) emitted by a molecule at temperature T (Melnikov et al. 2016). Tables 7 and 8 give the file format and an example of the cooling function file. Fig. 3 shows the temperature-dependent cooling function of NO, MgH, and AlH from the

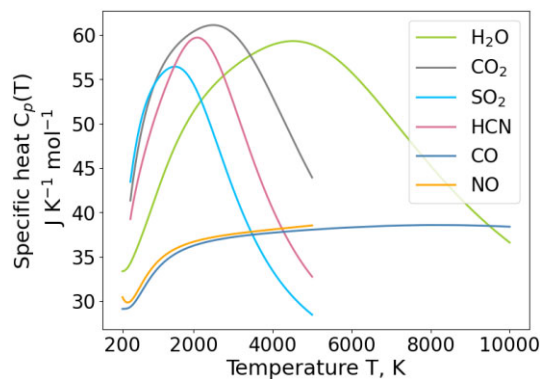


Figure 2. Specific heats of $^1\text{H}_2^{16}\text{O}$, $^{12}\text{C}^{16}\text{O}_2$, $^{32}\text{S}^{16}\text{O}_2$, $^1\text{H}^{12}\text{C}^{14}\text{N}$, $^{12}\text{C}^{16}\text{O}$, and $^{14}\text{N}^{16}\text{O}$ with temperatures from 200 to 10 000 K or T_{\max} for the given line list, using the ExoMol line lists.

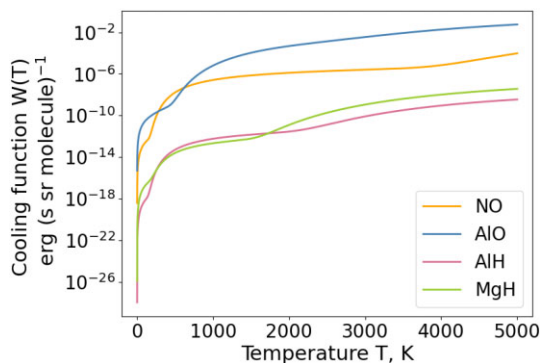


Figure 3. The temperature-dependent cooling functions of $^{14}\text{N}^{16}\text{O}$, $^{27}\text{Al}^{16}\text{O}$, $^{27}\text{Al}^1\text{H}$, and $^{24}\text{Mg}^{16}\text{O}$ with temperatures from 0 to 5000 K using the ExoMol line lists.

ExoMol data base. The cooling functions is produced by the emissivity (Neale, Miller & Tennyson 1996):

$$W(T) = \frac{1}{4\pi Q(T)} \sum_{f,i} A_{fi} h c \tilde{\nu}_{fi} g'_f e^{-c_2 \tilde{E}'_f / T}. \quad (4)$$

In equation (4), $Q(T)$ is the partition function at temperature T ; g'_f is the upper state degeneracy and \tilde{E}'_f is the upper state energy which can be read from the `.states` States file; and A_{fi} and $\tilde{\nu}_{fi}$ are the Einstein A -coefficient (s^{-1}) and transition wavenumber (cm^{-1}) in the `.trans` Transitions file (see Tables 9 and 2). PYEXOCROSS requires the States files, Transitions files, and partition function files to compute $W(T)$ on a grid of temperatures. Although some smaller Transitions files provide the wavenumbers as column 4 (see next), it is important that they evaluated using the upper and lower state term values from the States files as given by

$$\tilde{\nu}_{fi} = \tilde{E}'_f - \tilde{E}''_i. \quad (5)$$

2.5 Radiative lifetimes

Radiative life times are important for determining the critical density and as they represent a competing process to collisional thermalization in environments including plasmas such as the interstellar medium. The radiative lifetime τ_i (s) of state i can be calculated by inverting the sum of Einstein A -coefficients A_{fi} (s^{-1}) with $\tilde{E}_j < \tilde{E}_i$ from the (`.trans`) Transitions file(s), see Table 2 (Tennyson et al. 2016a):

$$\tau_i = \frac{1}{\sum_f A_{fi}}. \quad (6)$$

To calculate the state-dependent lifetime, the process has been divided into three steps. Initially, transitions are reorganized with different upper state categories by `pandas.DataFrame.groupby` (McKinney & et al. 2010). Each upper state f corresponds to one or more than one Einstein A -coefficient A_{fi} . Secondly, the sum over the Einstein A -coefficients per upper state is computed. Finally, the inverse of each sum is calculated. In principle, τ_i can be calculated for each state i in the States file (see Table 10). In practice, however, the `.trans` files are not always complete. In this case the lifetime is considered to be undefined with `NaN`. Some lowest states, particularly for high symmetry molecules such methane or H_3^+ , have infinite lifetimes, for which ‘inf’ is used. The inclusion of lifetimes is being made a standard part of the (recommended) States files in ExoMol as part of the generalization of the ExoMol data format to allow for predissociation lifetimes (Tennyson et al. 2023).

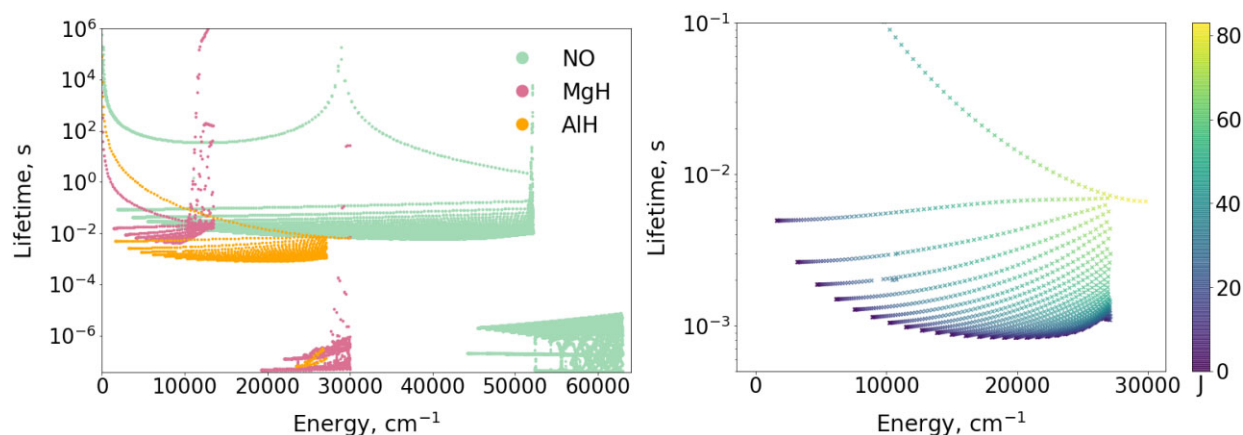


Figure 4. Left-hand panel: Lifetime of $^{14}\text{N}^{16}\text{O}$, $^{24}\text{Mg}^1\text{H}$, and $^{27}\text{Al}^1\text{H}$ using the ExoMol line lists. Right-hand panel: State-dependent lifetime in range of 10^{-4} – 10^{-1} s of $^{27}\text{Al}^1\text{H}$ from the ExoMol data base. The left had scale give the J values for which life times are plotted.

Table 10. Specification of the mandatory part of the `.states` States file with extra data options `unc`, `τ` , and `g` .

Field	Fortran format	C format	Description
i	I12	%12d	State ID
E	F12.6	%12.6f	State energy in cm^{-1}
g_{tot}	I6	%6d	State degeneracy
J	I7/F7.1	%7d/%7.1f	J -quantum number (integer/half integer)
(Unc)	F12.6	%12.6f	Uncertainty in state energy in cm^{-1} (optional)
(τ)	ES12.4	%12.4E	Lifetime in s (optional)
(g)	F10.6	%10.6f	Landé g -factor (optional)
(Extra)			Extra quantum numbers, any format (optional)

Notes. ID: State identifier – a non-negative integer index, starting at 1;

J : Total angular momentum quantum, excluding nuclear spin;

Fortran format, J integer: (I12, 1x, F12.6, 1x, I6, I7, 1x, ES12.4, 1x, F10.6)

or J half integer: (I12, 1x, F12.6, 1x, I6, F7.1, 1x, ES12.4, 1x, F10.6).

Table 11. Specification of the `.os` oscillator strength file format.

Field	Fortran format	C format	Description
f	I12	%12d	Upper state degeneracy
i	I12	%12d	Lower state degeneracy
gf	ES10.4	%10.4E	Oscillator strength
$\tilde{\nu}_i$	F15.6	%15.6f	Central bin wavenumber, cm^{-1} (optional)

Note. Fortran format: (I12, 1x, I12, 1x, ES10.4, 1x, F15.6).

PYEXOCROSS adds the lifetimes to the States file in the ExoMol data format. PYEXOCROSS also provides a user option to save either an uncompressed `.states` file or a compressed `.states.bz2` file. Fig. 4 (left-hand panel) shows states-dependent lifetimes for different molecules while the right-hand panel shows the lifetimes computed from AIH line list. These lifetimes are fully state specific; recently Owens et al. (2023) used ExoMol data to construct a lifetime data base (LiDB) for plasma modelling which stores lifetimes and decay routes for vibronic states by lumping them over rotational and fine structure states.

2.6 Oscillator strengths

The oscillator strength is a dimensionless quantity used in spectroscopy to indicate the probability of electromagnetic radiation being emitted or absorbed in transitions between energy levels of an atom or molecule (Demtröder 1981). The fractional amount of effective electrons participating in the transition is represented by the oscillator strength f (Orloski & Trigueiros 2001). The weighted oscillator strength gf is defined as

$$gf = \frac{g'_{\text{tot}} A_{fi}}{(c\tilde{\nu}_{fi})^2}, \quad (7)$$

where $g'_{\text{tot}} = g_{\text{ns}}(2J' + 1)$ is the degeneracy of the upper states with upper angular momentum J' ; A_{fi} and $\tilde{\nu}_{fi}$ are the Einstein A -coefficient (s^{-1}) and transition wavenumber (cm^{-1}), respectively; and c is the velocity of light (cm s^{-1}).

Table 12. Extract from the .os oscillator strength file generated from the $^{14}\text{N}^{16}\text{O}$ line lists (Qu et al. 2021).

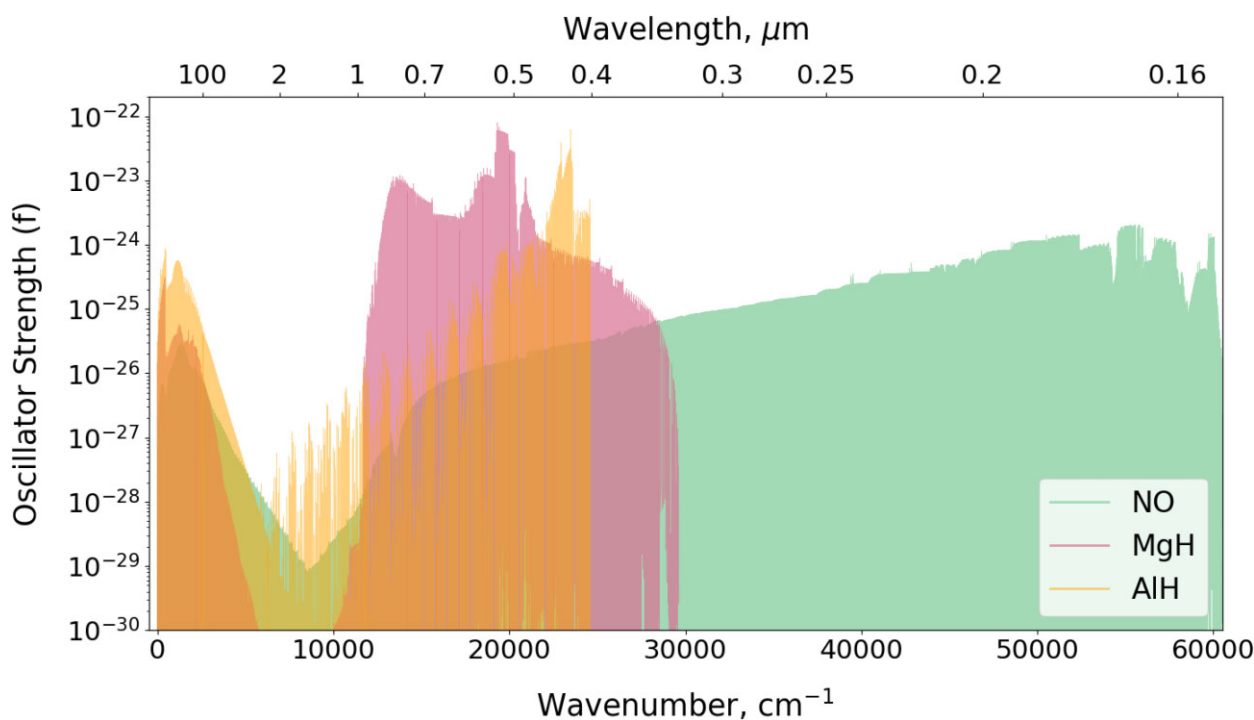
f	i	gf	$\tilde{\nu}_i$
2237	2604	2.3938E-29	32 406.776 335
16 945	16 990	2.6529E-30	32 406.778 880
11 519	11 563	9.3397E-30	32 406.843 732
15 241	14 758	4.6102E-26	32 406.846 700
15 986	15 778	1.0011E-25	32 406.866 875
13 921	13 689	9.0563E-31	32 406.888 211
7437	7184	1.1891E-30	32 406.897 702
11 504	11 847	1.0124E-28	32 406.925 712
16 602	16 639	4.0684E-30	32 406.926 676
16 353	16 394	5.5179E-31	32 406.942 107

Notes. f : Upper state ID;

i : Lower state ID;

gf : Oscillator strength;

$\tilde{\nu}_i$: Central bin wavenumber, cm^{-1} (optional).

**Figure 5.** Oscillator strength of $^{14}\text{N}^{16}\text{O}$, $^{24}\text{Mg}^{16}\text{O}$, and $^{27}\text{Al}^1\text{H}$ from the ExoMol line lists.

The actual oscillator strength is defined by $f = gf/g''_{\text{tot}}$, where g''_{tot} is the degeneracy of the lower state. Tables 11 and 12 give the file format and an example of the oscillator strength file. Fig. 5 shows the oscillator strength of NO, MgH, and AlH from the ExoMol data base.

2.7 Line intensities and stick spectra

2.7.1 Absorption coefficient

The absorption coefficient, also known as the absorption line intensity I_{fi} (cm molecule^{-1}), can be calculated at a general temperature T using (Jørgensen et al. 1996; Yurchenko et al. 2018a)

$$I_{f \leftarrow i} = \frac{g'_f A_{fi}}{8\pi c \tilde{\nu}_{fi}^2} \frac{e^{-c_2 \tilde{E}_i''/T} (1 - e^{-c_2 \tilde{\nu}_{fi}/T})}{Q(T)}, \quad (8)$$

where g'_f is the upper state degeneracy; \tilde{E}_i'' is the lower state energy; A_{fi} is the Einstein A-coefficient (s^{-1}); $\tilde{\nu}_{fi}$ is the transition wavenumber (cm^{-1}); $c_2 = hc/k_B$ is the second radiation constant (cm K), where h is the Planck constant (erg s), c is the speed of light (cm s^{-1}), and k_B is the Boltzmann constant (erg K^{-1}); and T is the temperature in K. The temperature-dependent partition function $Q(T)$ can be accessed from the

Table 13. Specification of the mandatory part of the `.stick` stick spectra file with the extra columns for the optional quantum numbers.

Field	Fortran format	C format	Description
$\bar{\nu}$	F12.6	%12.6f	Frequency (wavenumber) in cm^{-1}
I	ES14.8	%14.8E	Intensity in cm molecule^{-1}
J'	I7/F7.1	%7d/%7.1f	J -quantum number (integer/half integer) for upper state
E'	F12.4	%12.4f	Upper state energy in cm^{-1}
J''	I7/F7.1	%7d/%7.1f	J -quantum number (integer/half integer) for lower state
E''	F12.4	%12.4f	lower state energy in cm^{-1}
(QN')			Extra quantum numbers for upper state, any format (optional)
(QN'')			Extra quantum numbers for lower state, any format (optional)

Table 14. Extract from the `.stick` stick spectra file generated from the absorption $^{24}\text{Mg}^1\text{H}$ line list (Owens et al. 2022).

$\bar{\nu}_{fi}$	I_{fi}	J'	\bar{E}'	J''	\bar{E}''	par'	e/f'	State'	v'	$ \Lambda '$	$ \Sigma '$	$ \Omega '$	par''	e/f''	State''	v''	$ \Lambda ''$	$ \Sigma ''$	$ \Omega ''$
1.242966	7.42574603E-49	20.5	10749.7776	20.5	10748.5346	-	f	X(2Sigma+)	8	0	-0.5	-0.5	+	e	X(2Sigma+)	9	0	0.5	0.5
1.252056	7.53243618E-46	20.5	10749.7776	19.5	10748.5256	-	f	X(2Sigma+)	8	0	-0.5	-0.5	+	f	X(2Sigma+)	9	0	0.5	0.5
1.337756	9.08537842E-46	21.5	10749.8724	20.5	10748.5346	-	e	X(2Sigma+)	8	0	-0.5	-0.5	+	e	X(2Sigma+)	9	0	0.5	0.5
1.569618	2.65779920E-52	36.5	12150.5019	36.5	12148.9323	-	f	X(2Sigma+)	7	0	-0.5	-0.5	+	e	X(2Sigma+)	8	0	0.5	0.5
1.597018	1.38115517E-48	36.5	12150.5019	35.5	12148.9049	-	f	X(2Sigma+)	7	0	-0.5	-0.5	+	f	X(2Sigma+)	8	0	0.5	0.5
1.816673	1.86393524E-48	37.5	12150.7490	36.5	12148.9323	-	e	X(2Sigma+)	7	0	-0.5	-0.5	+	e	X(2Sigma+)	8	0	0.5	0.5
2.859016	2.89098064E-45	0.5	10353.9808	0.5	10352.2738	-	f	X(2Sigma+)	11	0	-0.5	-0.5	+	e	X(2Sigma+)	11	0	0.5	0.5
2.862028	5.84278951E-45	1.5	10354.0094	0.5	10352.2738	-	e	X(2Sigma+)	11	0	-0.5	-0.5	+	e	X(2Sigma+)	11	0	0.5	0.5
3.589218	8.68228140E-47	33.5	11929.6911	34.5	11926.1019	-	e	X(2Sigma+)	9	0	-0.5	-0.5	+	e	X(2Sigma+)	8	0	0.5	0.5
3.606156	8.58388055E-47	32.5	11929.6121	33.5	11926.0060	-	f	X(2Sigma+)	9	0	-0.5	-0.5	+	f	X(2Sigma+)	8	0	0.5	0.5

Notes. $\bar{\nu}_{fi}$: Absorption frequency (wavenumber) (in cm^{-1});
 I_{fi} : Absorption intensity (in cm molecule^{-1});
 J' and J'' : Total angular momentum quantum number for upper and lower state;
 \bar{E}' and \bar{E}'' : Term value (in cm^{-1}) for upper and lower state;
 par' and par'': Total parity for upper and lower state;
 e/f' and e/f'': Rotationless parity for upper and lower state;
 State' and State'': Electronic term value for upper and lower state;
 v' and v'' : Vibrational quantum number for upper and lower state;
 $|\Lambda|'$ and $|\Lambda|''$: Absolute value of the projection of electronic angular momentum for upper and lower state;
 $|\Sigma|'$ and $|\Sigma|''$: Absolute value of the projection of the electronic spin for upper and lower state;
 $|\Omega|'$ and $|\Omega|''$: Absolute value of the projection of the total angular momentum for upper and lower state.

partition function file (`.p.f`). When computing $I_{f \leftarrow i}$ with PyEXOCROSS, the energy term values are read from the ExoMol States (`.states`) file, which are also used to evaluate $\bar{\nu}_{fi}^2$, the Einstein A-coefficient are read from the ExoMol Transitions (`.trans`) file, and the partition function is either taken from the partition function file or calculated on the fly using the energy term values and equation (1), as in the case for the cooling functions.

We note that the definitions of line intensities given in equations (8) and (9) are for 100 per cent abundance and, unlike HITRAN and HITEMP, are not scaled to allow for isotopic fractional abundances. The results should therefore be multiplied by the appropriate fractional abundance for many applications.

2.7.2 Emission coefficient

The emission coefficient is also called the emission line intensity and emissivity ϵ [units $\text{erg (s molecule sr)}^{-1} = 10^{-7} \text{ Watts (molecule sr)}^{-1}$] (Yurchenko et al. 2018a).

$$\epsilon(i \leftarrow f) = \frac{g'_f h c A_{fi} \bar{\nu}_{fi} e^{-c\bar{E}'_f/T}}{4\pi Q(T)}, \tag{9}$$

where g'_f is the upper state degeneracy and \bar{E}'_f is the upper state energy term value. As in the case of the absorption coefficients, the energy term values \bar{E}_f are read from the ExoMol States file, which are also used for $\bar{\nu}_{fi}$, A_{fi} are taken from the Transitions files, and $Q(T)$ is from the partition function file or computed using equation (1). When calculating the emission coefficient for the HITRAN data base, the HITRAN format line lists file (`.par`) only has the lower state energy $E_{\text{lower}} (E''_i)$. Then the upper state energy must be evaluated using equation (5) and the partition function is taken from the partition function file.

2.7.3 Stick spectra

Stick spectra are lists of frequencies and line intensities, absorption or emission, calculated using equations (8) and (9), respectively, along with their descriptions (required quantum numbers) for the upper and lower states (Yurchenko et al. 2018a). PyEXOCROSS can apply filters of uncertainties, thresholds, and quantum numbers when producing stick spectra. Table 13 gives the mandatory part of the stick (`.stick`) spectra file. Table 14 shows an example of a stick spectrum output. Lines are plotted as ‘sticks’, with their height given by their intensity (see Fig. 6).

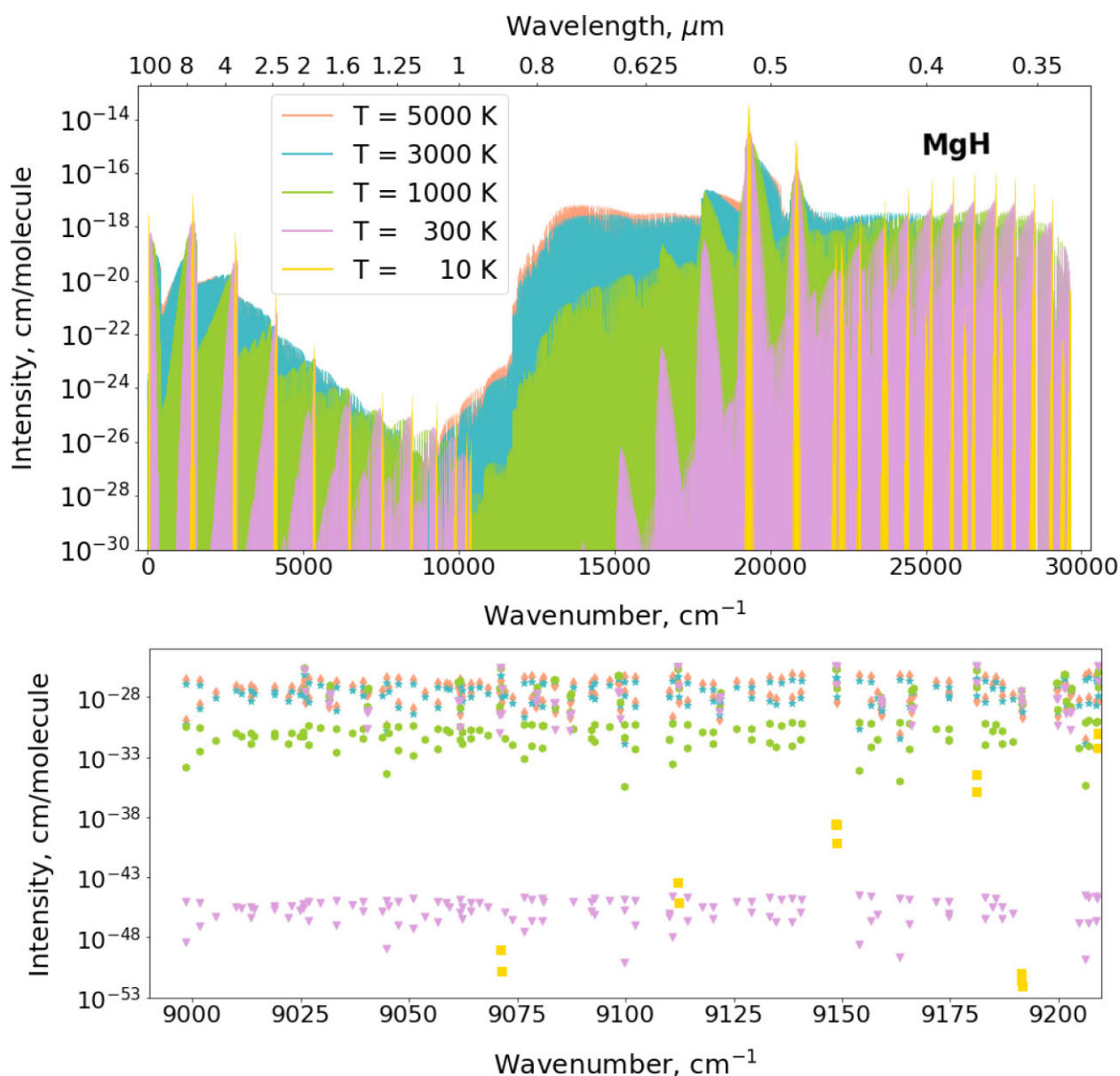


Figure 6. Stick spectra: Absorption intensities of $^{24}\text{Mg}^1\text{H}$ at $T = 10, 300, 1000, 3000,$ and 5000 K simulated using the ExoMol line list of data set XAB (Owens et al. 2022). Upper panel: Intensities in wavenumber range $[0, 30000]$ cm^{-1} . Lower panel: Intensities zoom at $[9000, 9200]$ cm^{-1} .

2.7.4 Intensity thresholds

PYEXOCROSS provides an intensity threshold filter to skip weak lines. Using an intensity threshold can speed up the cross-section calculations or reduce the output of stick spectra calculations. In PYEXOCROSS, the default value of the intensity threshold is 10^{-30} cm molecule^{-1} but this value can be changed in the input file.

2.8 Cross-sections

PYEXOCROSS computes the cross-sections (σ_{ab} or σ_{em}) for the given wavenumber range, temperature, pressure, line intensity (absorption coefficient I_{fi} in equation 8 or emission coefficient ϵ_{if} in equation 9) and line profile ($f_{\tilde{\nu}}(\tilde{\nu})$, more details in Section 3). PYEXOCROSS provides uncertainty, threshold, and quantum number filters for calculating cross-sections (see Section 2.9). Figs 7 and 8 show examples of cross-sections of MgH and H₂O at $P = 1$ bar and $T = 10, 300, 1000, 3000,$ and 5000 K simulated using the ExoMol line list XAB (Owens et al. 2022) and POKAZATEL (Polyansky et al. 2018). Fig. 9 compares cross-sections of H₂O (Polyansky et al. 2018), CO₂ (Yurchenko et al. 2020), and SO₂ (Underwood et al. 2016) at $T = 300$ and 3000 K. Table 15 shows the cross-section file format and Table 16 specifies a part of the cross-section file.

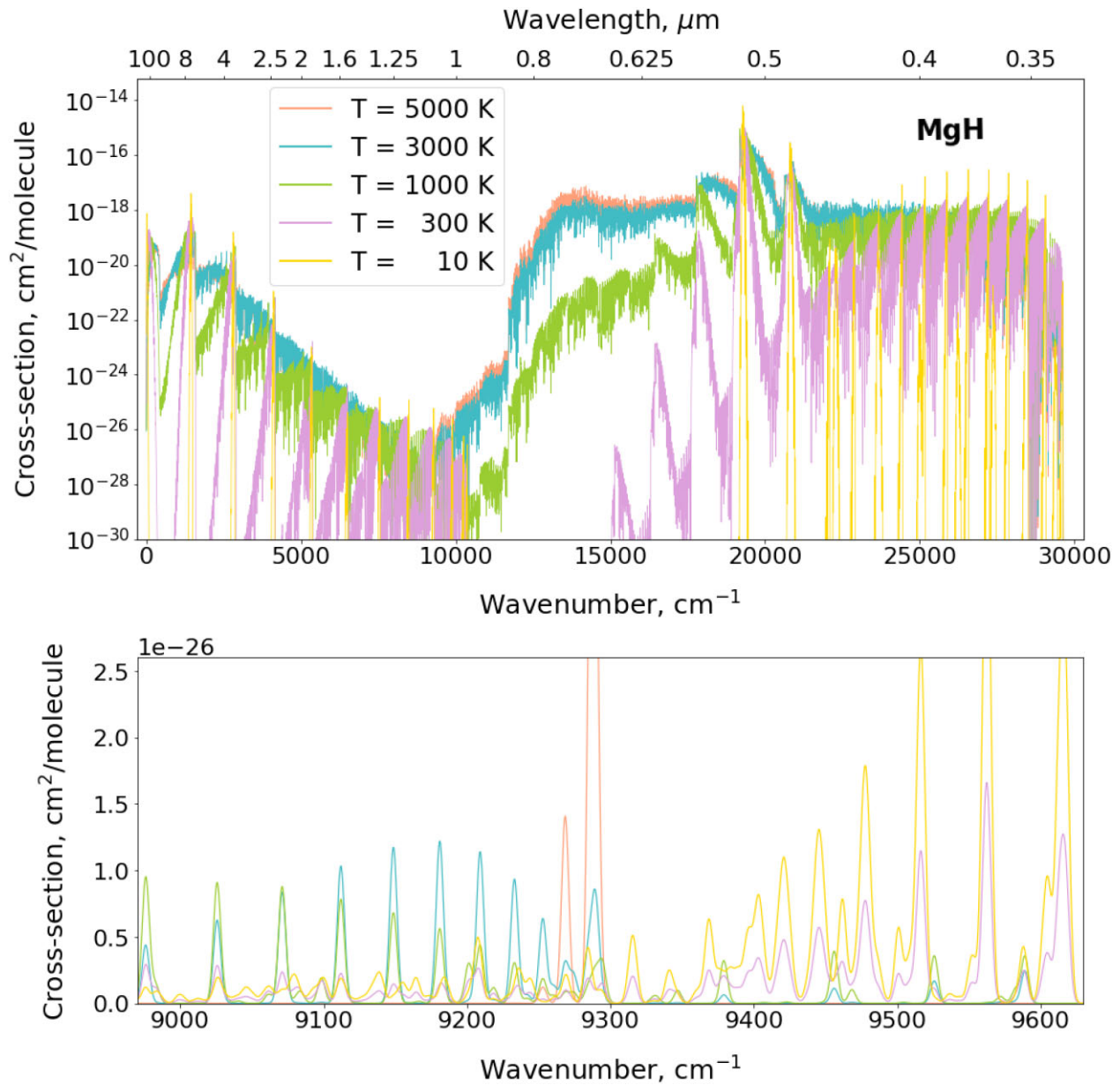


Figure 7. Cross-section: Absorption spectrum of $^{24}\text{Mg}^1\text{H}$ at $T = 10, 300, 1000, 3000,$ and 5000 K simulated using the ExoMol line list of the data set XAB (Owens et al. 2022) broadened using a Gaussian line profile with its HWHM $\alpha_G = 3 \text{ cm}^{-1}$. Upper panel: Cross-sections in wavenumber range $[0, 30000] \text{ cm}^{-1}$. Lower panel: Cross-sections zoom for $[9000, 9600] \text{ cm}^{-1}$.

2.8.1 Wavenumber grids

PYEXOCROSS uses uniform wavenumber grids when calculating cross-sections. The grids can be defined using two methods for a given wavenumber range $[\tilde{\nu}_{\min}, \tilde{\nu}_{\max}]$: (1) using the number of grid points N_{points} with $N_{\text{points}} - 1$ as the number of grid intervals or (2) the size of the grid bin $\Delta\tilde{\nu}$. The wavenumber step size is defined as

$$\Delta\tilde{\nu} = \frac{\tilde{\nu}_{\max} - \tilde{\nu}_{\min}}{N_{\text{points}} - 1}. \quad (10)$$

In order to let $\Delta\tilde{\nu}$ be a ‘round’ value, the number of grid points is usually set to an odd number (Yurchenko et al. 2018a).

2.8.2 Absorption and emission cross-sections

The absorption cross-section σ_{ab} at wavenumber $\tilde{\nu}$ for a transition line $f \leftarrow i$ with intensity I_{fi} (equation 8) and line profile ($f_{\tilde{\nu}fi}$) is given by

$$I_{fi} = \int_{-\infty}^{\infty} \sigma_{fi}(\tilde{\nu}) d\tilde{\nu} \Rightarrow \sigma_{\text{ab}}(\tilde{\nu}) = I_{fi} f_{\tilde{\nu}fi}(\tilde{\nu}). \quad (11)$$

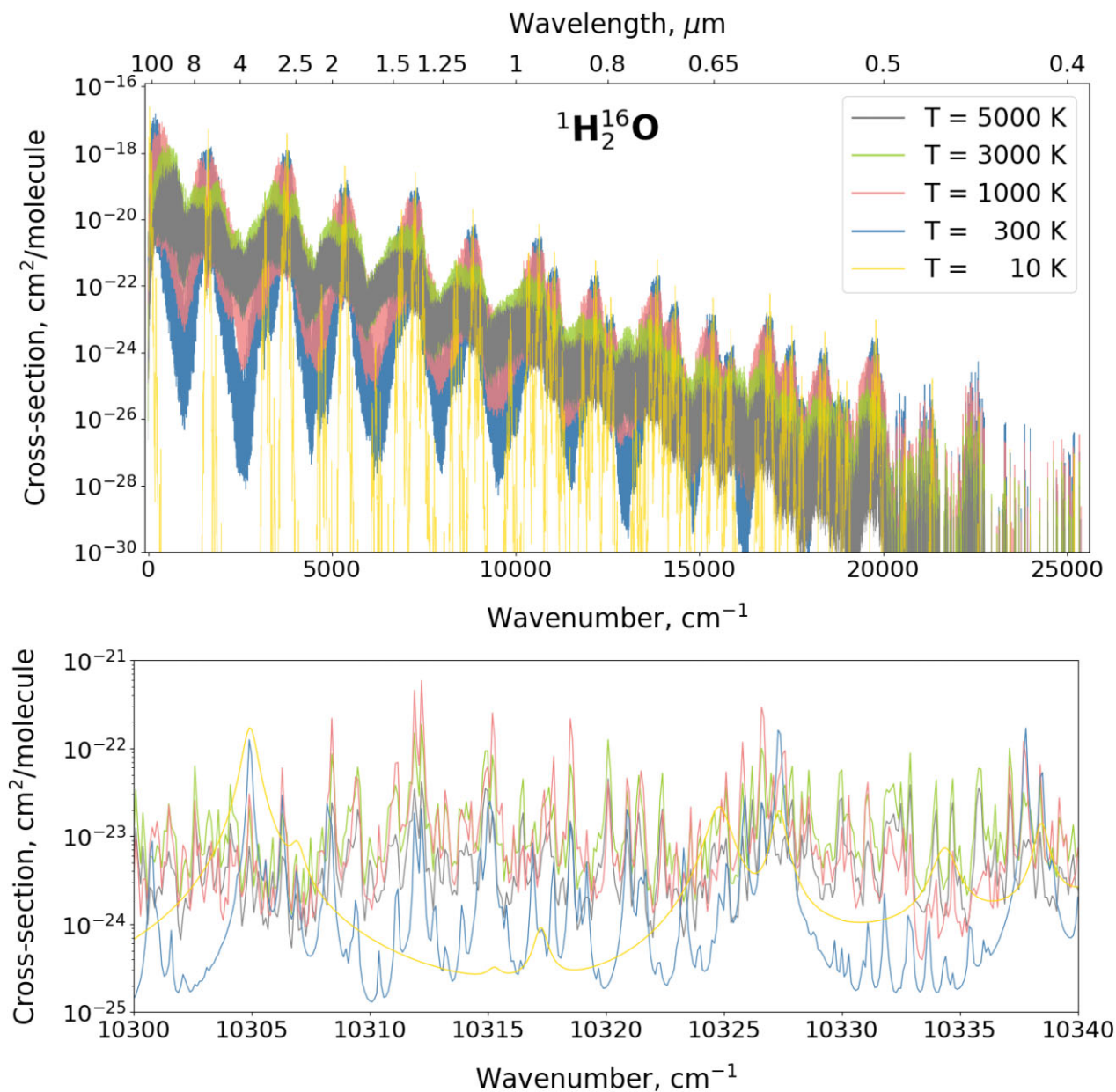


Figure 8. Cross-section: Absorption spectrum of $^1\text{H}_2^{16}\text{O}$ at $T = 10, 300, 1000, 3000,$ and 5000 K simulated using the ExoMol line list of the data set POKAZATEL (Polyansky et al. 2018) broadened using a SciPy Voigt line profile with the bin size 0.1 cm^{-1} at $P = 1$ bar, HWHM mixed with H_2 and He broadeners and also applied uncertainty and threshold filters (0.01 cm^{-1} and $10^{-30} \text{ cm molecule}^{-1}$). Upper panel: Cross-sections in wavenumber range $[0, 26000] \text{ cm}^{-1}$. Lower panel: Cross-sections zoom for $[10300, 10340] \text{ cm}^{-1}$.

Similarly, the emission cross-section σ_{em} at wavenumber $\tilde{\nu}$ for a transition line $i \rightarrow f$ is calculated with the emission coefficient ϵ_{if} (equation 9) and line profile ($f_{\tilde{\nu}fi}$):

$$\epsilon_{if} = \int_{-\infty}^{\infty} \epsilon_{if}(\tilde{\nu}) d\tilde{\nu} \Rightarrow \sigma_{\text{em}}(\tilde{\nu}) = \epsilon_{if} f_{\tilde{\nu}fi}(\tilde{\nu}). \quad (12)$$

2.9 Filters

PYEXOCROSS can apply filters including uncertainty, threshold, quantum number labels, formats, and values based on the information given in the `.states` file if users require the filters.

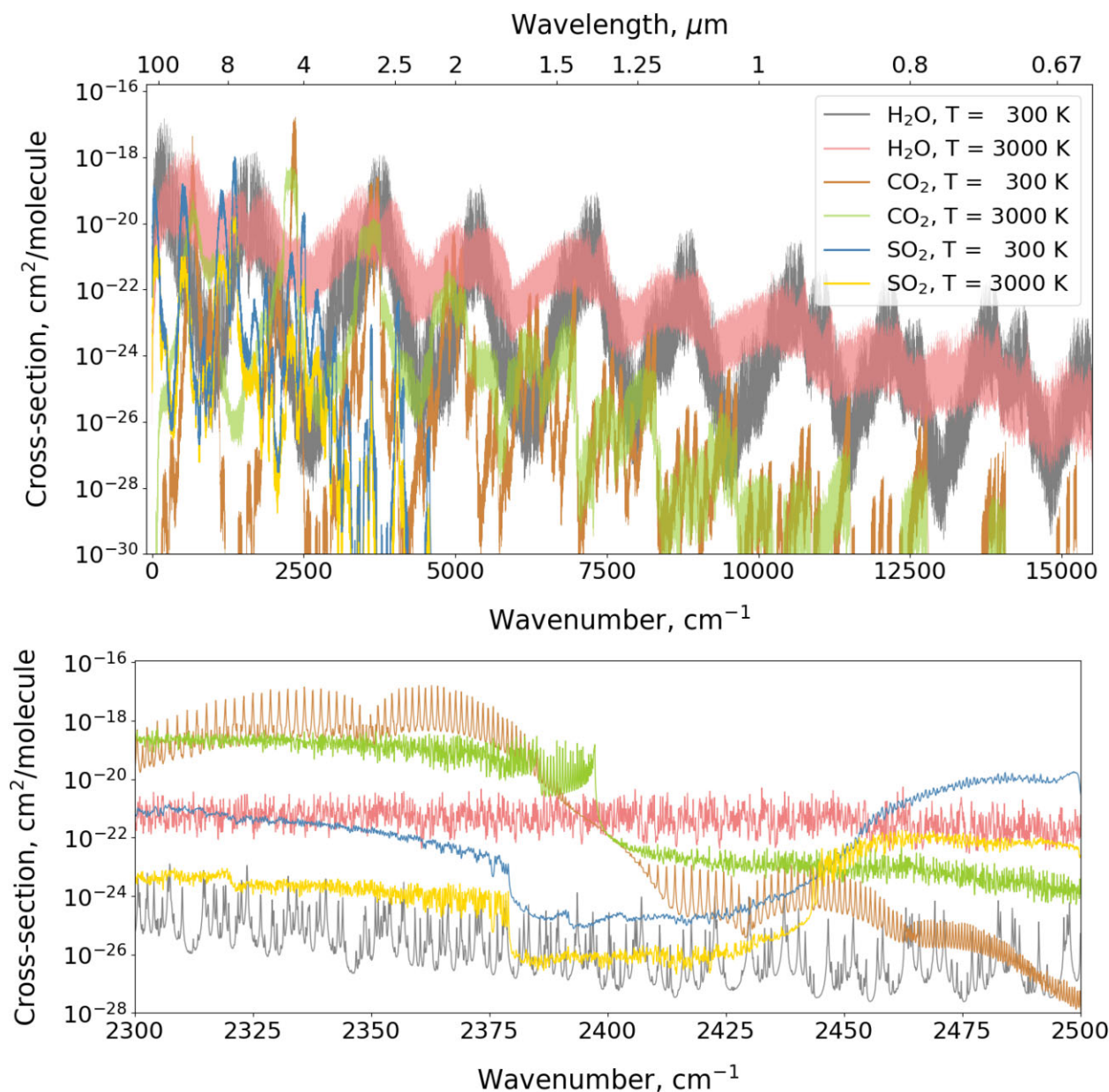


Figure 9. Cross-section: Absorption spectrum of $^1\text{H}_2^16\text{O}$, $^{12}\text{C}^{16}\text{O}_2$, and $^{32}\text{S}^{16}\text{O}_2$, at $T = 300$ and 3000 K simulated using the ExoMol line list of the data set POKAZATEL (Polyansky et al. 2018), UCL-4000 (Yurchenko et al. 2020), and ExoAmes (Underwood et al. 2016) broadened using a SciPy Voigt line profile with bin size 0.1 cm^{-1} at $P = 1$ bar, HWHM mixed with H_2 and He broadeners and also applied uncertainty and threshold filters (0.01 cm^{-1} and $10^{-30}\text{ cm}^2\text{ molecule}^{-1}$). Upper panel: Cross-sections in wavenumber range $[0, 15500]\text{ cm}^{-1}$. Lower panel: Cross-sections zoom for $[2300, 2500]\text{ cm}^{-1}$.

Table 15. Specification of the `.cross` cross-section file format.

Field	Fortran format	C format	Description
$\bar{\nu}_i$	F12.6	%12.6f	Central bin wavenumber, cm^{-1}
σ_i	ES14.8	%14.8E	Absorption cross-section, $\text{cm}^2\text{ molecule}^{-1}$

Note. Fortran format: (F12.6,1x,ES14.8).

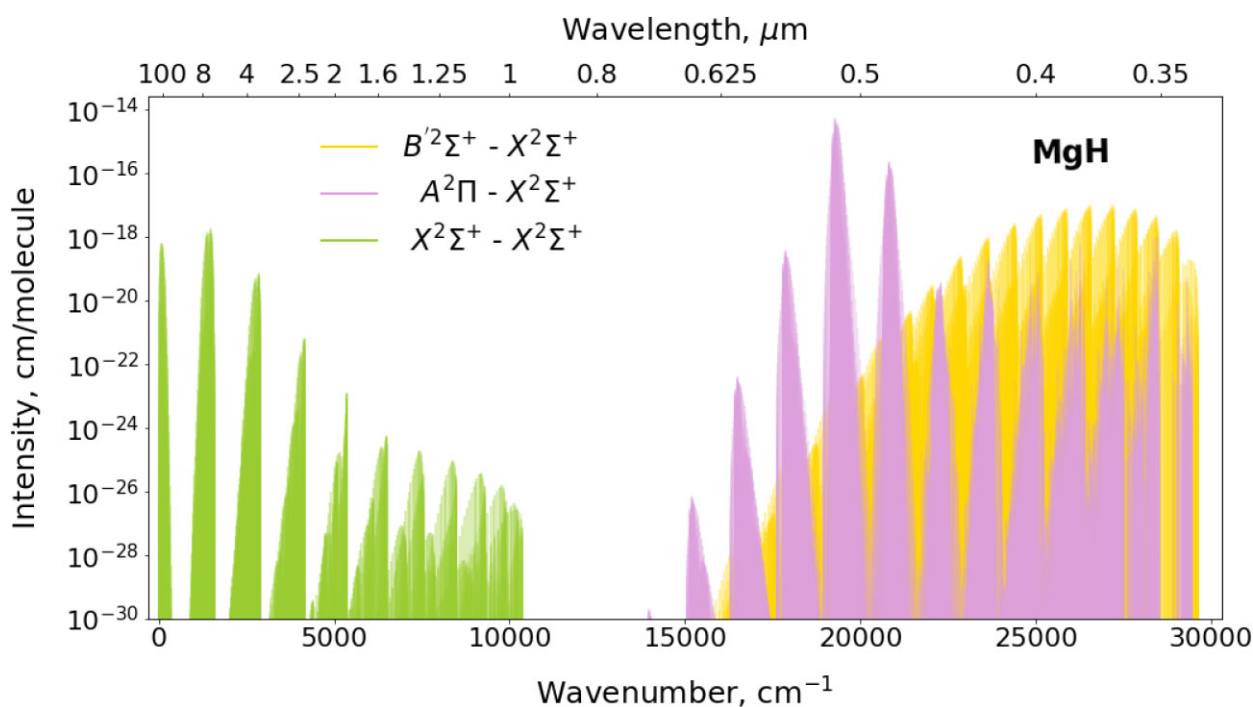
2.9.1 Uncertainty filter

An uncertainty filter can be applied when computing stick spectra, cross-sections, or performing format conversion. The uncertainty refers to the accuracy of each individual transition wavenumbers. PYEXOCROSS applies an uncertainty filter with a given value which is usually taken to be the upper bound of the selected uncertainties. Laboratory spectroscopy uses uncertainties in cm^{-1} and the uncertainties less than 0.01 cm^{-1} are usually assumed to be accurate, but even uncertainties of 0.05 cm^{-1} are probably good. Astronomers use the resolving power

Table 16. Extract from the `.cross` cross-section file generated from the $^1\text{H}_2^{16}\text{O}$ POKAZATEL ExoMol line lists (Polyansky et al. 2018).

$\bar{\nu}$ (cm $^{-1}$)	σ_i (cm 2 molecule $^{-1}$)
5000.000 000	8.03 246 668E-25
5000.100 000	1.84 420 150E-24
5000.200 000	2.00 318 098E-23
5000.300 000	3.84 642 766E-24
5000.400 000	8.77 939 177E-25
5000.500 000	4.33 350 935E-25
5000.600 000	2.91 607 795E-25
5000.700 000	2.30 487 176E-25
5000.800 000	2.01 726 401E-25
5000.900 000	2.32 892 097E-25

Notes. $\bar{\nu}$: Central bin wavenumber, cm $^{-1}$;
 σ_i : Absorption cross-section, cm 2 molecule $^{-1}$.

**Figure 10.** Absolute absorption line intensities of different electronic states X-X, A-X, and B-X bands of the ExoMol $^{24}\text{Mg}^1\text{H}$ simulated at $T = 300$ K.

$R = \frac{\lambda}{\Delta\lambda}$ in wavelength as a measure of the spectral resolution of an instrument and this is approximately the same as $R = \frac{\bar{\nu}}{\Delta\bar{\nu}}$ in wavenumbers. $R > 100\,000$ indicates high resolution, but for some astronomy fields, $R > 50\,000$ could be seen as high enough resolution as well.

2.9.2 Intensity threshold filter

An intensity threshold filter can be applied when computing stick spectra, cross-sections, or performing format conversion. A large intensity filter means only stronger intensities are processed in the task in question. PYEXOCROSS extracts and process lines, stronger than intensity threshold requested. The default intensity threshold value is $I > 10^{-30}$ cm molecule $^{-1}$. Users are recommended to use the threshold filter when processing the molecules with large data sets to reduce run times; however, we should note that including weak lines with $I < 10^{-30}$ cm molecule $^{-1}$ can significantly alter the results (Yurchenko et al. 2014).

2.9.3 Quantum number filter

A quantum number filter can be applied in calculations of stick spectra and cross-sections. This filter can be used to select transitions for a given set of the upper/lower quantum numbers. The user can also select the quantum numbers to be printed out as well as the printing format. Fig. 10 shows the stick spectra of MgH using the electronic states as the quantum number filter.

Table 17. HITRAN line-by-line format in 2004 edition (Rothman et al. 2005) (160-character record).

Parameter	M	I	ν	S	A	γ_{air}	γ_{self}	E''	n_{air}	δ_{air}	V'	V''	Q'	Q''	I_{err}	I_{ref}	*	g'	g''
Field length	2	1	12	10	10	5	5	10	4	8	15	15	15	15	6	12	1	7	7
Fortran format	I2	I1	F12.6	E10.3	E10.3	F5.4	F5.4	F10.4	F4.2	F8.6	A15	A15	A15	A15	6I1	6I2	A1	F7.1	F7.1
C format	%2d	%1d	%12.6f	%10.3e	%10.3e	%5.4f	%5.4f	%10.4f	%4.2f	%8.6f	%15s	%15s	%15s	%15s	%6d	%12d	%1s	%7.1f	%7.1f

Table 18. Description of the parameters presented in the 160-character records in the line-by-line section of the HITRAN data base (Rothman et al. 2005).

Parameter	Meaning	Field length	Type
M	Molecule number	2	Integer
I	Isotopologue number	1	Integer
ν	Vacuum wavenumber	12	Real
S	Intensity	10	Real
A	Einstein A -coefficient	10	Real
γ_{air}	Air-broadened half width	5	Real
γ_{self}	Self-broadened half width	5	Real
E''	Lower state energy	10	Real
n_{air}	Temperature-dependent exponent for γ_{air}	4	Real
δ_{air}	Air pressure-induced line shift	8	Real
V'	Upper state global 'quanta'	15	Hollerith
V''	Lower state global 'quanta'	15	Hollerith
Q'	Upper state local 'quanta'	15	Hollerith
Q''	Upper state local 'quanta'	15	Hollerith
I_{err}	Uncertainty indices	6	Integer
I_{ref}	Reference indices	12	Integer
*	Flag	1	Character
g'	The statistical weight of the upper state	7	Real
g''	The statistical weight of the lower state	7	Real

2.10 Plotting options

PYEXOCROSS provides the functionality to optionally plot figures for stick spectra and cross-sections. Users also can set the bottom y-axis view limits of data coordinates on plotting stick spectra and cross-sections. The default value of the bottom y-axis limit is 10^{-30} cm molecule $^{-1}$.

2.11 Format conversion

2.11.1 Converting data format from ExoMol to HITRAN

PYEXOCROSS can convert data from the ExoMol format to the HITRAN format (see Tables 17 and 18). This involves the following steps: (1) reading the States file (.states) from the ExoMol data base, (2) reading the ExoMol Transitions files (.trans) in chunks, (3) reading the ExoMol partition function file, (4) for each line, calculating the line intensity and applying the uncertainty and intensity thresholds, and (5) finally, printing out the transition descriptors in the HITRAN format, with the corresponding error codes for the given uncertainties, etc.

For the format conversion, the user has to provide the HITRAN molecule and isotopologue IDs. If the molecule is included in the HITRAN data base, PYEXOCROSS finds the necessary isotopic abundance from HITRANOnline, otherwise the program uses 1 as the abundance to calculate the intensities. PYEXOCROSS can extract Einstein A -coefficients from the ExoMol Transitions files, lower state energy; and upper and lower total degeneracy from the ExoMol States file. PYEXOCROSS also supports extracting γ_{self} and n_{air} from .self and .air broadening files, separately, if they exist. Otherwise, the default values $\gamma_{\text{self}} = 0.07$ cm $^{-1}$ per atm and $n_{\text{air}} = 0.5$ cm $^{-1}$ per atm are used. PYEXOCROSS only considers the first two HITRAN error codes which are error codes of the line position and intensity. The remaining four integers are set to zero. The line position uncertainty error codes are described in Table 19. For the intensity error codes, the record '0' is used for all states since they are undefined in the ExoMol data base. Table 20 gives some information on how the quantum number formats are converted.

PYEXOCROSS provides more flexibility to help users choose which quantum number labels they prefer to save in the HITRAN format as well as their formats. Users can provide global and local quantum number labels and their corresponding formats in the input file so that they can arbitrarily change the quantum number formats to satisfy the 60-character limit of the HITRAN data base format. If the total number of characters provided is less than 60, PYEXOCROSS will truncate them to be exactly 60 characters.

PYEXOCROSS currently cannot provide air pressure-induced line shifts (δ_{air}), reference indices (I_{ref}), and an associated uncertainty flag.

Table 19. The uncertainty codes used by the HITRAN data base (Gordon et al. 2022).

Code	Absolute uncertainty (unc) range
0	$1 \leq \text{unc}$ or Unreported
1	$0.1 \leq \text{unc} < 1$
2	$0.01 \leq \text{unc} < 0.1$
3	$0.001 \leq \text{unc} < 0.01$
4	$0.0001 \leq \text{unc} < 0.001$
5	$0.00001 \leq \text{unc} < 0.0001$
6	$0.000001 \leq \text{unc} < 0.00001$
7	$0.0000001 \leq \text{unc} < 0.000001$
8	$0.00000001 \leq \text{unc} < 0.0000001$
9	$0.000000001 \leq \text{unc} < 0.00000001$

Table 20. Conversion of the quantum number values between the ExoMol and HITRAN data base.

Quantum number value in ExoMol Description Label	Quantum value in HITRAN			Rotational parity (0 or 1) τ_i
	Total symmetry G_{tot}	Vibrational symmetry G_{vib}	Rotational symmetry G_{rot}	
0				s
1	A1'	A1'	A1'	a
2	A2'	A2'	A2'	
3	E'	E'	E'	
4	A1''	A1''	A1''	
5	A2''	A2''	A2''	
6	E''	E''	E''	

Table 21. Specification of the mandatory part of the `.broad` broadening file format.

Field	Fortran format	C format	Description
code	A2	%2s	Code identifying quantum number set following J
γ_{ref}	F6.4	%6.4f	Lorentzian half width at reference temperature and pressure in $\text{cm}^{-1}/\text{bar}$
n_L	F6.3	%6.3f	Temperature exponent
J''	I7/F7.1	%7d/%7.1f	Lower J -quantum number integer/half integer

Note. Fortran format: J integer: (A2,1x,F6.4,1x,F6.3,1x,I7) or J half integer: (A2,1x,F6.4,1x,F6.3,1x,F7.1).

2.11.2 Converting data format from HITRAN to ExoMol

In principle PYEXOCROSS can convert HITRAN `.par` files into the ExoMol format (`.states`, `.trans`, or `.pf`), but this is not straightforward since the HITRAN data model does not ensure that the tabulated lower state energies and transitions are self-consistent. The main challenge is therefore to construct the ExoMol States file and define the associated state IDs. To this end, the following steps are used.

(i) Construct the ExoMol States file: (i) Read HITRAN line list `.par` file; (ii) for each line, calculate the upper state energy \tilde{E}'_i with the wavenumber $\tilde{\nu}_j$ and lower state energy \tilde{E}''_j by equation (5); (iii) convert the uncertainty error code into the upper bound of the absolute uncertainty range; (iv) process quantum numbers and extract angular momentum F' and F'' from the HITRAN quantum numbers; (v) split HITRAN data into the upper and lower data sets, called DataFrames in Python, which include the Einstein A -coefficients (A), the statistical weights (g), uncertainties, state energies \tilde{E} , global quantum number and local quantum numbers, and then concatenate these two DataFrames with the same columns; (vi) create a new DataFrame which merges levels with the same statistical weights g_{tot} , state energies \tilde{E} , uncertainties, and quantum numbers, in this process, we use the lines with the lowest uncertainty when they have the same quantum numbers and statistical weights; (vii) sort with increasing state energies and use the continuum-increasing integers as the state IDs; and (viii) create the ExoMol States file in the ExoMol data format with the state ID i , state energy \tilde{E} , total degeneracy g_{tot} , angular momentum J or F , uncertainties unc , and quantum numbers (see Table 10).

(i) Construct the ExoMol Transitions file: (i) For each line in the HITRAN DataFrame, reconstruct the upper and lower state IDs; (ii) calculate new transitions frequencies with upper and lower state energies; and (iii) write an ExoMol Transitions file with the upper ID, lower state ID, Einstein A -coefficient A , and frequency wavenumber $\tilde{\nu}$ (see Table 9).

(i) Construct an ExoMol air-broadening file (`.air`): (i) Concatenate the initial HITRAN DataFrame with the γ_{air} , n_{air} , and J'' columns, and then drop duplicated rows and (ii) fill 'a0' in the first column called 'code' and save code, γ_{air} , n_{air} , and J'' to the ExoMol air broadening file with the ExoMol data format (see Table 21). Note this procedure assumes the relatively simple a0 J -diet for line broadening (Barton et al. 2017).

(i) Construct ExoMol self-broadening files (.self): (i) Concatenate the initial HITRAN DataFrame with the γ_{self} , n_{air} , and J'' columns, and then drop duplicated rows and (ii) fill 'a0' in the first column called 'code' and save code, γ_{air} , n_{air} , and J'' to the ExoMol self-broadening file with the ExoMol data format (see Table 21, see next for more details on the line broadening codes).

PYEXOCROSS can only convert the standard 160-character HITRAN format data to the ExoMol format and for reasons given above the results need to be reviewed carefully. As the HITRAN data base has added extra H₂, He, CO₂, and H₂O broadening, users can replace air and self-broadening columns with these extra broadening data. The quantum numbers are formatted following tables S1 and S2 in the HITRAN 2020 supplementary material (Karlovets et al. 2021). The quantum numbers are written using the same format as used in the HITRAN.par file, therefore there is no need for the format specification of the quantum numbers.

3. LINE BROADENING AND LINE PROFILES

An individual spectral line at high resolution is characterized by (1) its transition frequency, (2) transition integrated intensity, and (3) line profile. The line profile is governed by line broadening effects which are functions of temperature and pressures, and whose inclusion is essential for most practical applications. The temperature-dependent Doppler broadening is represented by a Doppler line profile in the form of a Gaussian function. The pressure broadening can be modelled by a Lorentzian line profile, while the Voigt line profiles provide a convolution of the Doppler and Lorentzian profiles (Schreier 2017). It may also be necessary to include lifetime effects but these also produce a Lorentzian so do not alter the discussion ahead (Tennyson et al. 2023).

Line profiles $f_{\tilde{\nu}}(\tilde{\nu})$ are integrable functions whose areas are normalized to unity:

$$\int_{-\infty}^{\infty} f_{\tilde{\nu}}(\tilde{\nu})d\nu = 1. \tag{13}$$

PYEXOCROSS offers a choice of four profiles namely Doppler, generalized Gaussian, Lorentzian, and Voigt. Of course the Doppler profile is a Gaussian but with the line width defined based on the mass of the species concerned, temperature, and the line position, while Gaussian allows the user to specify the half width at half-maximum (HWHM; α_G). Section 3.5.1 explores seven different methods of calculating Voigt profiles.

3.1 Wing cutoff

As line profiles have in principle infinite spread, frequency (or wavelength) cutoffs are often used in practical calculations to limit the calculation region to the line centre only and to allow for the fact that Voigt/Lorentzian functions do not generally give a good representation of the far wings (Yurchenko et al. 2018a). Aside from influencing computation time and the accuracy of the cross-sections, a wing cutoff is also necessary if functions representing, e.g. the water continuum are to be used, as was recently advocated for exoplanet studies by Anisman et al. (2022). Gharib-Nezhad et al. (2024) discuss this issue in detail. PYEXOCROSS uses a default cutoff of 25 cm⁻¹ in line with HITRAN, HITEMP, and most representations of the water continuum. This choice is a matter of convention and can be changed by the user; however, we note that the default value is in line with the recent recommendations of Gharib-Nezhad et al. (2024).

3.2 Broadening parameters

Pressure broadening in the ExoMol data base is currently based on the ExoMol diet (Barton et al. 2017); the data base stores line broadening parameters in .broad broadening files; broadeners available for a given species are specified in the .def definition file. Users can provide as many broadening files with appropriate file suffixes, e.g. H2O.broad and CO2.broad. PYEXOCROSS can then use suffixes to read the corresponding broadening files and, if necessary, process mixtures of broadeners (see Section 3.4.2). PYEXOCROSS can process user-supplied broadening files provided they use the ExoMol .broad file format. The format of the broadening file is expressed in Table 21 and the specification of the broadening file is given in Table 22. Currently, the ExoMol data base contains .broad files for air, Ar, CH₄, CO, CO₂, H₂, H₂O, He, N₂, NH₃, NO, O₂, and the rapidly increasing number of broadeners being included in the ExoMol data base. The HITRAN data base has air and self-broadening for all species as well as parameters for the broadening by He, H₂, H₂O, and CO₂ (Ngo et al. 2013; Wilzewski et al. 2016; Tan et al. 2019, 2022; Wcislo et al. 2021) for a small number of key species.

3.3 Doppler and Gaussian profile

The Gaussian line profile is defined as

$$f_{\tilde{\nu}_{fi},\alpha_G}(\tilde{\nu}) = \sqrt{\frac{\ln 2}{\pi}} \frac{1}{\alpha_G} \exp\left(-\frac{(\tilde{\nu} - \tilde{\nu}_{fi})^2 \ln 2}{\alpha_G^2}\right), \tag{14}$$

where α_G is the HWHM and $\tilde{\nu}_{fi}$ is the line centre. α_G is related to the Gaussian standard deviation σ_G as (Squires 2001)

$$\sigma_G = \frac{\alpha_G}{\sqrt{2 \ln 2}} \tag{15}$$

and can be specified in the PYEXOCROSS input.

Table 22. Extract from the `.broad` broadening file of the H₂ broadening file `14N-1H3_H2.broad` (Barton et al. 2017).

Code	γ_{ref}	n_L	J''
a0	0.0908	0.583	0
a0	0.0935	0.559	1
a0	0.0926	0.527	2
a0	0.0850	0.508	3
a0	0.0800	0.529	4
a0	0.0768	0.506	5
a0	0.0728	0.501	6
a0	0.0698	0.506	7
a0	0.0671	0.514	8
a0	0.0655	0.500	9

Notes. Code: Code identifying the ExoMol diet type;
 γ_{ref} : Lorentzian half width at reference temperature and pressure in $\text{cm}^{-1}/\text{bar}$;
 n_L : Temperature exponent;
 J'' : Lower J -quantum number integer/half integer.

The Doppler profile is based on the Gaussian shape specified in equation (14) with HWHM α_D given by (Squires 2001)

$$\alpha_D = \sqrt{\frac{2N_A k_B T \ln 2}{M}} \frac{\tilde{\nu}_{fi}}{c}, \quad (16)$$

which depends on the temperature, mass of the molecule, and the line position. Here, k_B (erg K) is the Boltzmann constant, c (cm s^{-1}) is the speed of light, T (K) is the temperature, M (Da) (g mol^{-1}) is the molar mass of the isotopologue in grams (Ngo et al. 2013) which can be taken from the HITRANOnline website or from the `.def` definition file of the ExoMol data base, and N_A (mol^{-1}) is Avogadro's number. N_A/M can be replaced by $1/1000m$ where isotopologue mass m (kg) is available from the ExoMol definition file. For Doppler as well as for the generalized Gaussian, the HWHMs do not depend on the (upper/lower) molecular states.

3.4 Lorentzian profiles

The Lorentzian profile is the Cauchy distribution whose probability density function is (Feller 1971; Johnson, Kotz & Balakrishnan 1994)

$$f_{\tilde{\nu}_{fi}, \gamma_L}^L(\tilde{\nu}) = \frac{1}{\pi} \frac{\gamma_L}{(\tilde{\nu} - \tilde{\nu}_{fi})^2 + \gamma_L^2}, \quad (17)$$

where $\tilde{\nu}_{fi}$ is the centroid in wavenumbers identifying the position of the peak of the distribution and γ_L is Lorentzian HWHM, which in PYEXOCROSS conventionally assumes the following temperature and pressure-dependent form:

$$\gamma_L = \gamma_{\text{ref}} \left(\frac{T_{\text{ref}}}{T} \right)^{n_L} \frac{P}{P_{\text{ref}}}, \quad (18)$$

where $\tilde{\nu}_{fi}$ is the line centre. As in HITRAN, $T_{\text{ref}} = 296$ K is the reference temperature and $P_{\text{ref}} = 1$ bar is the reference pressure (Schreier & Hochstaffl 2021), and γ_{ref} and n_L are the reference HWHM (see Sections 3.2 and 3.4.1) and the temperature exponent, which are the broadening parameters for the given broadeners in the ExoMol data base. In the HITRAN data base, γ_{air} and γ_{self} are used to calculate γ_{ref} with mixing ratios provided by users, see Section 3.4.2. n_L is as n_{air} in the HITRAN line lists. In general, γ_{ref} and n_L are (upper/lower) state-dependent. The specification of the broadeners and the broadener mixture ratio must be set in the PYEXOCROSS `.inp` input file.

Users can choose to input the HWHM (γ_L) of the Lorentzian profile directly as a constant for calculations instead of calculating with the parameters from the line-broadening files of the ExoMol data base.

3.4.1 Extracting broadening parameters from `.broad`

Calculation of the Lorentzian/Voigt broadening requires information from the States (`.states`) file and the parameters from the broadening file (`.broad`). If the value of J'' in the States file is larger than J''_{max} in the broadening file, then the broadening parameters γ_L and n_L corresponding to J''_{max} are assumed. If J'' is half integer, the parameters corresponding to the index value $J'' - \frac{1}{2}$ in the broadening file are used.

This process of extracting values based on their indices is especially efficient in Python. A part of the Python code for extracting broadening parameters from the .broad file is shown in Listing 1.

```
def extract_broad(broad_df, states_trans_df):
    max_broad_J = max(broad_df['Jpp'])
    Jpp = states_trans_df['J'''].values
    Jpp[Jpp > max_broad_J] = max_broad_J
    id_broad = (states_trans_df['J''']-0.1).round(0).astype('int32').values
    gamma_L = broad_df['gamma_L'][id_broad]
    n_air = broad_df['n_air'][id_broad]
    return(gamma_L, n_air)
```

Listing 1: A sample of the Python code for extracting broadeners from the .broad broadening file. PYEXOCROSS reads the .broad broadening file and saves the data as a DataFrame broad_df. states_trans_df is a DataFrame storing the line lists of the .states States file merged by upper and lower states IDs from the .trans Transitions file. Jpp and J'' means lower J and id_broad indicates the index of the broad DataFrame. gamma_L and n_air are γ_{ref} and n_{air} denoting the reference HWHM and the temperature exponents.

3.4.2 Mixtures of broadeners

PYEXOCROSS allows for inclusion of different broadeners or their mixtures. For mixtures, in the input file, users can specify the fractional abundance (X_i) of the i th broadener. Then, the overall value of γ_L is the weighted sum of the values of γ_i^L from each broadener (Yurchenko et al. 2018a):

$$\gamma_L = \sum_i \gamma_i^L X_i \quad \text{and} \quad \sum_i X_i = 1. \tag{19}$$

3.4.3 Lifetime broadening

In principle, the natural lifetime contributes to the line profile alongside the effects of the temperature-dependent Doppler broadening and the pressure-dependent collision broadening. In practice, the natural lifetime generally contributes very little to the line profile and is consequently ignored. However, for predissociating states, the effect of the lifetime broadening can be important, see Pavlenko et al. (2022) and Yurchenko et al. (2024) for example; these can easily be included in the Voigt profile (Tennyson et al. 2023). Like the pressure broadening, lifetime broadening γ_τ produces a Lorentzian line shape with the HWHM given by

$$\gamma_\tau = \frac{\hbar}{2\tau'hc}, \tag{20}$$

where h is the Planck constant (erg s) and $\hbar = h/(2\pi)$ is the reduced Planck constant. c is the speed of light (cm s^{-1}). τ' is the upper state lifetime (s). The sum of the pressure broadening and lifetime broadening $\gamma_{tot} = \gamma_p + \gamma_\tau$ can be then used as the total Lorentzian component of the line profile. Since a Voigt profile is already being utilized, this has no computational impact on a calculation, implying that routine use of γ_{tot} for the half width on a routine basis would eliminate the need to worry about whether or not predissociation has to be considered.

3.5 Voigt profiles

The Voigt profile results from the convolution of the Gaussian and Lorentzian profiles and is given by (Armstrong 1967; Herbert 1974)

$$f_{\tilde{\nu}_{fi}, \sigma_G, \gamma_L}^V(\tilde{\nu}) = f_{\tilde{\nu}_{fi}, \sigma_G}^G(\tilde{\nu}) \otimes f_{\tilde{\nu}_{fi}, \gamma_L}^L(\tilde{\nu}) = \int_{-\infty}^{+\infty} d\tilde{\nu}' f^G(\tilde{\nu} - \tilde{\nu}', \sigma_G) \times f^L(\tilde{\nu}' - \tilde{\nu}_{fi}, \gamma_L) = \frac{y}{\pi^{3/2}\sqrt{2}\sigma_G} \int_{-\infty}^{+\infty} \frac{e^{-t^2}}{(x-t)^2 + y^2} dt, \tag{21}$$

where

$$x = \frac{\tilde{\nu} - \tilde{\nu}_{fi}}{\sqrt{2}\sigma_G} \quad \text{and} \quad y = \frac{\gamma_L}{\sqrt{2}\sigma_G}. \tag{22}$$

Here x gives the wavenumber scale in units of the Gaussian standard deviation, y is the ratio of Lorentzian widths to the Gaussian standard deviation, σ_G is the Gaussian standard deviation, and γ_L is HWHM of the Lorentzian profile. The central wavenumber or frequency, $\tilde{\nu}_{fi}$ is the energy difference of the transition (Schreier 2018).

PYEXOCROSS offers nine methods for calculating the Voigt profile with five of them as pseudo-Voigt profiles as described in the following.

3.5.1 *SciPy Voigt profile*

Use of the Python open-source library `SCIPY` (Virtanen et al. 2020) gives the Voigt profile directly, see Listing 2.

```
from scipy.special import voigt_profile
def scipy_Voigt_profile(dv, sigma, gamma):
    return voigt_profile(dv, sigma, gamma)
```

Listing 2: Sample of the Python code used to calculate Voigt Profile by `SCIPY` package (Virtanen et al. 2020). $dv = \bar{v} - \bar{v}_{fi}$ and \bar{v}_{fi} is the line centre. σ is the standard deviation σ_G of the Gaussian profile. γ is the half width for pressure broadening γ_L .

`scipy.special.voigt_profile` uses `scipy.signal.convolve` to compute the Voigt profile with the convolution of a one-dimensional (1D) Normal distribution (Gaussian profile) with the standard deviation σ_G and a 1D Cauchy distribution (Lorentzian profile) with HWHM γ_L (Virtanen et al. 2020), see equation (21). The method of `scipy.signal.convolve` is given in Listing 3.

```
from scipy.signal import convolve
convolved=dx*convolve(Lorentzian_profile, Gaussian_profile, mode="same")
```

Listing 3: Method for computing Voigt profile by the convolution of the Gaussian and Lorentzian profile using the `SCIPY` Python package as used by the `scipy.special.voigt_function`.

3.5.2 *SciPy wofz Voigt profile*

The Python open-source library `SCIPY` can be used to calculate the Voigt profile, see Listing 4. The `SCIPY` function `scipy.special.wofz` calculates the Voigt profile for σ_G and γ_L as given by

$$f_{\bar{v}_{fi}, \sigma_G, \gamma_L}^V(\bar{v}) = \frac{1}{\sqrt{2\pi}\sigma_G} K(x, y), \quad (23)$$

where x and y are given in equation (22), $K(x, y)$ is the Voigt function

$$K(x, y) = \frac{y}{\pi} \int_{-\infty}^{+\infty} \frac{e^{-t^2}}{(x-t)^2 + y^2} dt \quad (24)$$

satisfying $\int K dx = \sqrt{\pi}$. $K(x, y)$ comprises the real part of the complex error function (Faddeeva function) $\omega(z)$ (Faddeeva et al. 1961):

$$\omega(z) \equiv K(x, y) + iL(x, y) = \frac{i}{\pi} \int_{-\infty}^{+\infty} \frac{e^{-t^2}}{z-t} dt. \quad (25)$$

where

$$z = x + iy = \frac{(v - \bar{v}_{fi} + i\gamma_L)}{\sqrt{2}\sigma_G} \quad (26)$$

and

$$f_{\bar{v}_{fi}, \sigma_G, \gamma_L}^V(\bar{v}) = \frac{\text{Re}[\omega(z)]}{\sigma_G \sqrt{2\pi}}. \quad (27)$$

In equation (26), in the limiting cases of $\sigma_G = 0$ and $\gamma_L = 0$, equation (27) is simplified to the Lorentzian profile and Gaussian profile, respectively.

```
import numpy as np
from scipy.special import wofz
def scipy_wofz_Voigt_profile(dv, sigma, gamma):
    z = (dv + 1j * gamma) / sigma / np.sqrt(2)
    voigt = (np.real(wofz(z)) / sigma / np.sqrt(2 * np.pi))
    return voigt
```

Listing 4: A sample of the Python code used to calculate the Voigt profile (Hill 2020; Virtanen et al. 2020) with the real part of the Faddeeva function. `wofz` in `scipy.special.wofz` can be computed with error function $\text{erf}(-iz)$ given by Pierluissi, Vanderwood & Gomez (1977): $\omega(z) = e^{-z^2}(1 - \text{erf}(-iz))$.

3.5.3 *Humlíček Voigt profile*

Humlíček's algorithm (Humlíček 1979) for approximating the Voigt profile function is used extensively. Humlíček's Voigt profile gives a method for computing the complex probability function $\omega(z)$ where $z = x + iy$ (see equations 21, 25, and 26). $x = \frac{(\bar{v} - \bar{v}_{fi})\sqrt{\ln 2}}{\alpha_D}$ is the wavenumber scale in units of the Doppler HWHM α_D and $y = \frac{\gamma_L \sqrt{\ln 2}}{\alpha_D}$ is the ratio of the Lorentzian HWHM γ_L to the Doppler HWHM α_D . A rational

approximation is used in Humlíček’s w4 algorithm, in which the imaginary part on the real axis is minimized in order to minimize its relative error (Humlíček 1982). The following procedure can evaluate both the real and imaginary parts of $\omega(z)$ with a high degree of relative accuracy (Humlíček 1979, 1982; Kuntz 1997).

$$\omega(z) = \begin{cases} \frac{t}{(0.5+t^2)\sqrt{\pi}}, & \text{if } s \geq 15, \\ \frac{t(1.4104739589 + \frac{u}{\sqrt{\pi}})}{0.75+u(3+u)}, & \text{if } 5.5 \leq s < 15, \\ \frac{16.4955+t(20.20933+t(11.96482+t(3.778987+0.5642236t)))}{16.4955+t(38.82363+t(39.27121+t(21.69274+t(6.699398+t)))}), & \text{if } s < 5.5 \text{ and } y \geq 0.195|x| - 0.176, \\ e^u - \frac{t(36183.31-u(3321.99-u(1540.787-u(219.031-u(35.7668-u(1.320522-u(0.56419))))))}{32066.6-u(24322.8-u(9022.23-u(2186.18-u(364.219-u(61.5704-u(1.84144-u))))))}, & \text{if } s < 5.5 \text{ and } y < 0.195|x| - 0.176, \end{cases} \tag{28}$$

where complex function $t = y - xi$, $s = |x| + y$, and $u = t^2$. Zaghoul (2017) and Hui, Armstrong & Wray (1978) defined a simpler, more efficient, and relatively accurate approximation for the evaluation of the Faddeyeva function with six regions of the proposed partitioning.

The Humlíček Voigt line profile is then given by (Humlíček 1979, 1982)

$$f_{\tilde{\nu}_{fi}, \alpha_D, \gamma_L}^V(\tilde{\nu}) = \frac{\text{Re}[\omega(z)]\sqrt{\ln 2}}{\pi\alpha_D} \tag{29}$$

3.5.4 Pseudo-Voigt profiles

PyEXOCROSS provides five pseudo-Voigt line profiles. They are not the recommended algorithms and we do not consider pseudo-Voigt profiles in comparison Section 3.7 because they have lower accuracy and speed of run time than other Voigt profiles.

Thompson pseudo-Voigt profile

The pseudo-Voigt function $V(d\nu, \gamma_V)$ is frequently used for calculating experimental spectral line shapes (Ida, Ando & Toraya 2000). As an approximation to the Voigt profile $V(x)$, the pseudo-Voigt profile combines a Gaussian curve $G(x)$ and a Lorentzian curve $L(x)$ linearly, rather than by convolution (Ida et al. 2000). The following equation provides the mathematical definition of the normalized pseudo-Voigt profile $V(d\nu, \gamma_V)$ (Thompson, Cox & Hastings 1987; Sánchez-Bajo & Cumbreña 1997):

$$V(d\nu, \gamma_V) = \eta L(d\nu, \alpha_D) + (1 - \eta)G(d\nu, \gamma_L) \quad \text{with } 0 < \eta < 1, \tag{30}$$

where $d\nu = \tilde{\nu} - \tilde{\nu}_{fi}$; α_D and γ_L are the HWHM of the Gaussian and Lorentzian profiles.

The accuracy of the pseudo-Voigt profile depends on several possible choices for calculating the parameter η and HWHM of the Voigt profile γ_V . For the Thompson pseudo-Voigt profile η is given in equation (31) (Thompson et al. 1987):

$$\eta \approx 1.36603(\gamma_L/\gamma_V) - 0.47719(\gamma_L/\gamma_V)^2 + 0.11116(\gamma_L/\gamma_V)^3. \tag{31}$$

From the associated Gaussian and Lorentzian HWHMs, the value of γ_V of the Thompson pseudo-Voigt profile can be determined using (Thompson et al. 1987)

$$\gamma_V = [\alpha_D^5 + 2.69269\alpha_D^4\gamma_L + 2.42843\alpha_D^3\gamma_L^2 + 4.47163\alpha_D^2\gamma_L^3 + 0.07842\alpha_D\gamma_L^4 + \gamma_L^5]^{1/5}. \tag{32}$$

Kielkopf pseudo-Voigt profile

Kielkopf (1973) proposed a weighted-sum method to approximate the value of γ_V of the pseudo-Voigt profile via equations (30 and 31) as given by (see also Schreier (2011)

$$\gamma_V \approx \frac{1}{4} \left[(1 + 0.099 \ln 2)\gamma_L + \sqrt{(1 - 0.099 \ln 2)^2\gamma_L^2 + 4 \ln 2\alpha_D^2} \right] \approx 0.5346\gamma_L + \sqrt{0.2166f\gamma_L^2 + \alpha_D^2}. \tag{33}$$

Olivero pseudo-Voigt profile

Olivero & Longbothum (1977) provided another alternative to equations (30 and 31) with the value of HWHM γ_V approximated by

$$\gamma_V \approx [1 - 0.18121(1 - d^2) - (0.023665e^{0.6d} + 0.00418e^{-1.9d}) \sin(\pi d)](\alpha_D + \gamma_L), \tag{34}$$

where d is a dimensionless parameter related to α_D and γ_L (Olivero & Longbothum 1977) as

$$d = \frac{\gamma_L - \alpha_D}{\gamma_L + \alpha_D}. \tag{35}$$

Liu-Lin pseudo-Voigt profile

Liu et al. (2001) used d related to the Gaussian and Lorentzian half widths (α_D and γ_L) defined in equation (35) to provide a new conversion of η from Kielkopf (1973), see equation (31). The half width of the Voigt profile is the same as equation (34). The Liu-Lin pseudo-Voigt profile is calculated from equations (30, 34, 35, and 36) using

$$\eta \approx 0.68188 + 0.61293d - 0.18384d^2 - 0.11568d^3. \tag{36}$$

Rocco pseudo-Voigt profile

Di Rocco & Cruzado (2012) provided a theoretically justified method for calculating γ_V and η in the pseudo-Voigt methodology and defined them as follows:

$$\begin{aligned}\gamma_V &= \omega_G b_{1/2}(y), \\ \eta &\approx \frac{\sqrt{\pi} V_y(0) \gamma_V - \sqrt{\ln 2}}{\sqrt{\pi} V_y(0) \gamma_V (1 - \sqrt{\pi \ln 2})},\end{aligned}\quad (37)$$

where

$$\omega_G = \frac{\alpha_D}{\sqrt{\ln 2}},$$

$V_y(0)$ is the Voigt profile in equation (21) at $x = 0$

$$V_y(0) = \frac{e^{y^2} (1 - \operatorname{erf}(y))}{\sqrt{\pi} \omega_G}.$$

Here,

$$y = \frac{\gamma_L}{\omega_G}$$

and

$$b_{1/2}(y) = y + \sqrt{\ln 2} e^{-0.6055y + 0.0718y^2 - 0.0049y^3 + 0.000136y^4}.$$

3.6 Binned line profiles

All the above methods use a sampling algorithm to evaluate the Voigt profile on a set of frequency wavenumbers $\tilde{\nu}_i$. The disadvantage of these methods is that they can lead to undersampling of the opacity (absorption or emission) at low sampling resolution (see e.g. Hill, Yurchenko & Tennyson 2013). In order to avoid this problem and guarantee that the full opacity is preserved, the following binned approach can be used by defining the analytically integrated (averaged) cross-sections within a bin $[\tilde{\nu}_k - \Delta\tilde{\nu}/2, \tilde{\nu}_k + \Delta\tilde{\nu}/2]$ as a sum over the contributions from each line (Hill et al. 2013) as given by

$$\begin{aligned}\bar{\sigma}_k^{\text{ab},fi} &= \frac{I_{fi}}{\Delta\tilde{\nu}} \int_{\tilde{\nu}_k - \Delta\tilde{\nu}/2}^{\tilde{\nu}_k + \Delta\tilde{\nu}/2} f_{\tilde{\nu}_{fi}}(\tilde{\nu}) d\tilde{\nu}, \\ \bar{\sigma}_k^{\text{em},if} &= \frac{\epsilon_{if}}{\Delta\tilde{\nu}} \int_{\tilde{\nu}_k - \Delta\tilde{\nu}/2}^{\tilde{\nu}_k + \Delta\tilde{\nu}/2} f_{\tilde{\nu}_{if}}(\tilde{\nu}) d\tilde{\nu},\end{aligned}\quad (38)$$

where $\tilde{\nu}_{fi}$ is the line centre, I_{fi} and ϵ_{if} are the line intensities (absorption and emission) from equations (8 and 9).

The related functions for the binned cross-sections for the Doppler, Gaussian, Lorentzian, and Voigt profiles are specified below.

3.6.1 Binned Doppler and Gaussian profile

Both the Doppler and the Gaussian profiles use the Gaussian function. Taking advantage of the fact that the integral of the Gaussian function has an analytic solution, the error function erf can be used:

$$\int e^{-x^2} dx = \frac{\sqrt{\pi}}{2} \operatorname{erf}(x).\quad (39)$$

Therefore, we apply the analytical integral to the Gaussian profile (Hill et al. 2013):

$$\bar{\sigma}_k^{fi} = \frac{I_{fi}}{\Delta\tilde{\nu}} \int_{\tilde{\nu}_k - \Delta\tilde{\nu}/2}^{\tilde{\nu}_k + \Delta\tilde{\nu}/2} f_{\tilde{\nu}_{fi}}^G(\tilde{\nu}) d\tilde{\nu} = \frac{I_{fi}}{2\Delta\tilde{\nu}} [\operatorname{erf}(x_{k,fi}^+) - \operatorname{erf}(x_{k,fi}^-)],\quad (40)$$

where $x_{k,fi}^\pm$ are the scaled limits of the frequency bin k centred on $\tilde{\nu}_k$ related to the line centre $\tilde{\nu}_{fi}$ given by Hill et al. (2013):

$$x_{k,fi}^\pm = \frac{\sqrt{\ln 2}}{\alpha_D} \left[\tilde{\nu}_k - \tilde{\nu}_{fi} \pm \frac{\Delta\tilde{\nu}}{2} \right].\quad (41)$$

3.6.2 Binned Lorentzian profile

The integral of the Lorentzian function has an analytical solution in the format

$$\int \frac{dx}{x^2 + \gamma^2} = \frac{1}{\gamma} \arctan\left(\frac{x}{\gamma}\right) + C.\quad (42)$$

If we calculate profiles in the range of the wings which have large distances from the line centre, then the normalization correction factor is unity. To solve the far wings in calculating profiles and save an excessive amount of computing time, we truncate the profiles at some distance $\Delta\tilde{\nu}/2$ from the line centre (Sharp & Burrows 2007). Therefore, PyEXOCROSS provides two methods for calculating binned Lorentzian and Voigt profiles with or without a cutoff filter for users.

The same approach of an analytical binned integral can be applied to the Lorentzian profile $f_{\tilde{\nu}_{fi}}^L(\tilde{\nu})$ as follows (Yurchenko et al. 2018a):

$$\bar{\sigma}_k^{fi} = \frac{I_{fi}}{\Delta\tilde{\nu}} \int_{\tilde{\nu}_k - \Delta\tilde{\nu}/2}^{\tilde{\nu}_k + \Delta\tilde{\nu}/2} f_{\tilde{\nu}_{fi}}^L(\tilde{\nu}) d\tilde{\nu} = \frac{I_{fi}}{\pi \Delta\tilde{\nu}} [\arctan(y_{k,fi}^+) - \arctan(y_{k,fi}^-)], \tag{43}$$

where $\bar{\sigma}_k^{fi}$ is an averaged cross-section for the bin k centred at $\tilde{\nu}_k$ over the interval $\Delta\tilde{\nu}$, $\tilde{\nu}_{fi}$ is the centre of the line, and $y_{k,fi}^\pm$ are given by

$$y_{k,fi}^\pm = \frac{\tilde{\nu}_k - \tilde{\nu}_{fi} \pm \Delta\tilde{\nu}/2}{\gamma_L}. \tag{44}$$

Because the line profiles are usually cut off at a distance $\Delta\tilde{\nu}/2$ from the line centre, it is common to preserve the total strength of the line by re-normalizing its area to unity (Sharp & Burrows 2007). Here we multiply equation (43) by the correction factor $b_{\text{corr}} \geq 1$ given by

$$b_{\text{corr}} = \frac{\pi}{\arctan(y_{\text{end},fi}) - \arctan(y_{\text{start},fi})}, \tag{45}$$

where $y_{\text{start},fi}$ and $y_{\text{end},fi}$ are defined as

$$y_{\text{start},fi} = \frac{\tilde{\nu}_{\text{start}} - \tilde{\nu}_{fi}}{\gamma_L} \quad \text{and} \quad y_{\text{end},fi} = \frac{\tilde{\nu}_{\text{end}} - \tilde{\nu}_{fi}}{\gamma_L}. \tag{46}$$

3.6.3 Binned Voigt profile

Gauss–Hermite quadrature is a form of the Gaussian quadrature used in numerical analysis to approximate the value of integrals (Abramowitz & Stegun 1972):

$$\int_{-\infty}^{+\infty} e^{-t^2} f(t) dt \approx \sum_{m=1}^n \omega_m f(t_m), \tag{47}$$

where n is the number of the used sample points and t_m are the roots of the Hermite polynomial $H_n(t)$ ($m = 1, 2, \dots, n$) (physicists' version) and the Gauss–Hermite quadrature weights ω_m are given by (Abramowitz & Stegun 1972):

$$\omega_m = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 [H_{n-1}(t_m)]^2}. \tag{48}$$

Combined with the Gaussian and Lorentzian profiles, we use Humlíček's algorithm (Humlíček 1979) as the Gauss–Hermite quadratures to create a formulation of a similar integral method for the Voigt profile (Yurchenko et al. 2018a) with HWHM of Doppler (and Gaussian) profile α_D and the HWHM of Lorentzian profile γ_L (see equation 21):

$$f_{\tilde{\nu}_{fi}, \alpha_D, \gamma_L}^V(\tilde{\nu}) = \frac{y \ln 2}{\pi^{3/2} \alpha_D^2} \int_{-\infty}^{+\infty} \frac{e^{-t^2}}{(x-t)^2 + y^2} dt = \frac{y \ln 2}{\pi^{3/2} \alpha_D^2} \sum_{m=1}^{N_{G-H}} \frac{\omega_m^{G-H}}{(x-t_m)^2 + y^2}, \tag{49}$$

where t_m are the Gauss–Hermite quadrature points (roots) and ω_m^{G-H} are the Gauss–Hermite quadrature weights. Using equation (43) to form the area-conserved integrals, for the binned Voigt profile we obtain (Yurchenko et al. 2018a)

$$\bar{\sigma}_k^{fi} = \frac{I_{fi}}{\pi^{3/2} \Delta\tilde{\nu}} \sum_{m=1}^{N_{G-H}} \omega_m^{G-H} [\arctan(z_{k,fi,m}^+) - \arctan(z_{k,fi,m}^-)], \tag{50}$$

where the scaled limits $z_{k,fi,m}^\pm$ of the bin k centred at the grid point $\tilde{\nu}_k$ is given by

$$z_{k,fi,m}^\pm = \frac{\tilde{\nu}_k - \tilde{\nu}_{fi} - t_m \sigma_G \pm \Delta\tilde{\nu}/2}{\gamma_L}.$$

We commonly take $N_{G-H} = 20$ as the number of Gauss–Hermite points used.

This expression can be further corrected for the missing area due to the wing's cutoff similarly to the Lorentzian example above. Here, we multiply $\bar{\sigma}_k^{fi}$ in equation (50) with the normalization correction factor b_{corr} given by

$$b_{\text{corr}} = \frac{\pi}{\arctan(z_{\text{end},fi,m}) - \arctan(z_{\text{start},fi,m})}, \tag{51}$$

where

$$z_{\text{start},fi,m} = \frac{\tilde{\nu}_{\text{start}} - \tilde{\nu}_{fi} - t_m \sigma_G}{\gamma_L} \quad \text{and} \quad z_{\text{end},fi,m} = \frac{\tilde{\nu}_{\text{end}} - \tilde{\nu}_{fi} - t_m \sigma_G}{\gamma_L}. \tag{52}$$

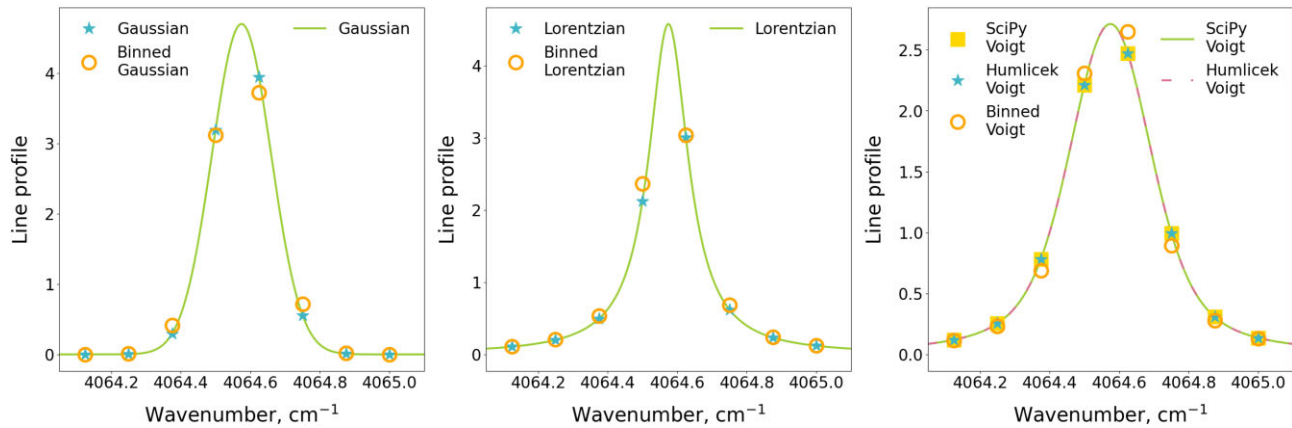


Figure 11. Comparison of the sampling and binned methods with different line profiles for a single normalized transition with a cutoff = 100 cm⁻¹ and the bin size of 0.125 cm⁻¹ (8 points per 1 cm⁻¹) at the same temperature $T = 300$ K and pressure $P = 1$ bar. Left-hand panel: Doppler profile; middle panel: Lorentzian profile; and right-hand panel: Voigt profile.

3.7 Line profile comparisons

3.7.1 Comparison of the sampling and binned line profile methods

When the cross-sections resolution (i.e. the number of points per wavenumber interval) is low, the result of cross-sections calculations becomes very sensitive to the way how profiles are evaluated, using sampling or binning. While the sampling line profile methods are constructed to accurately reproduce the cross-sections at the given sampling points, the binned methods guarantee the integrated cross-section within a given bin, regardless of the number of grid points or integration intervals. Accordingly, the binned methods are especially advantageous when the conservation of the total flux, absorption, or emission coefficients, is required. The sampling method are preferred when the accuracy of the cross-sections at the given frequencies is critical.

Fig. 11 gives examples of the Doppler, Lorentzian, and Voigt line profiles calculated using the sampling and binned methods for a single normalized transition in the wavenumber range [4064, 4065] cm⁻¹ at the temperature $T = 300$ K and pressure $P = 1$ bar using the bin size of 0.125 cm⁻¹ (8 points per cm⁻¹) with a cutoff of 100 cm⁻¹. The lines show the correct line profiles modelled, while points represent the corresponding calculated values. For the sampling method, the profile lines go exactly through the calculated cross-section points, while for the binned method, the points are shifted away (up or down) from the curve in order to maintain the integrated cross-section (unity in this case) as follows:

$$\int_{-\infty}^{\infty} f_{\tilde{\nu}_{fi}}(\tilde{\nu})d\tilde{\nu} \equiv \sum_i f_{\tilde{\nu}_{fi}}(\tilde{\nu}_i)\Delta\tilde{\nu}, \quad (53)$$

where the rectangular integration rule is used.

The profile normalization error given as the difference of the integrated area from unity of a (normalized) line profile is given by definition

$$\epsilon_{\text{area}} = 1 - \sum f_{\tilde{\nu}_{fi}}(\tilde{\nu})\Delta\tilde{\nu}. \quad (54)$$

For the binned methods, ϵ_{area} is zero by construction, while for the sampling methods, it depends on the number of sampling points. Choosing an adequate number of sampling points is critical in radiative transfer applications (Chubb et al. 2021). Undersampling can lead to underestimation of opacity, while too many points can make the calculations too slow. The profile sampling error can be defined as the root-mean-squares error deviation of the calculated from the exact profiles at the sampling points as given by

$$\epsilon_{\text{sampling}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (f(\tilde{\nu}_i) - f_{\text{calc}}(\tilde{\nu}_i))^2}, \quad (55)$$

where N is the number of sampling points. The sampling error $\epsilon_{\text{sampling}}$ should be zero for the sampling methods by construction (within the numerical accuracy).

Fig. 12 gives a detailed analysis of the line profile errors ϵ_{area} and $\epsilon_{\text{sampling}}$ for a single normalized transition at $\tilde{\nu}_{if} = 4064.574\,702$ cm⁻¹ for different temperatures T , pressures P , number of points N_{points} , and bin sizes $\Delta\tilde{\nu}$; and a large wavenumber range of [0, 8000] cm⁻¹ is used. All sampling methods have negligible $\epsilon_{\text{sampling}}$, while all binned methods should lead to perfect $\epsilon_{\text{sampling}}$, at least within the numerical error.

For the cross-comparisons, the Gaussian line profile is always the most accurate sampling line profile to preserve the profile area even with smaller number of points due to its compact shape. The SciPy Voigt and Humlíček Voigt profiles provide similar accuracy. For constant pressure and grid number, the line profiles uncertainty decreases with increasing temperature, see Fig. 12 (first panel). In contrast, line profiles show a loss of accuracy as the pressure or grid number increase, see the second and third panels. The binned Lorentzian and Voigt profiles appear to degrade at higher pressures due to their non-negligible wings cutoff beyond the interval [0, 8000] cm⁻¹.

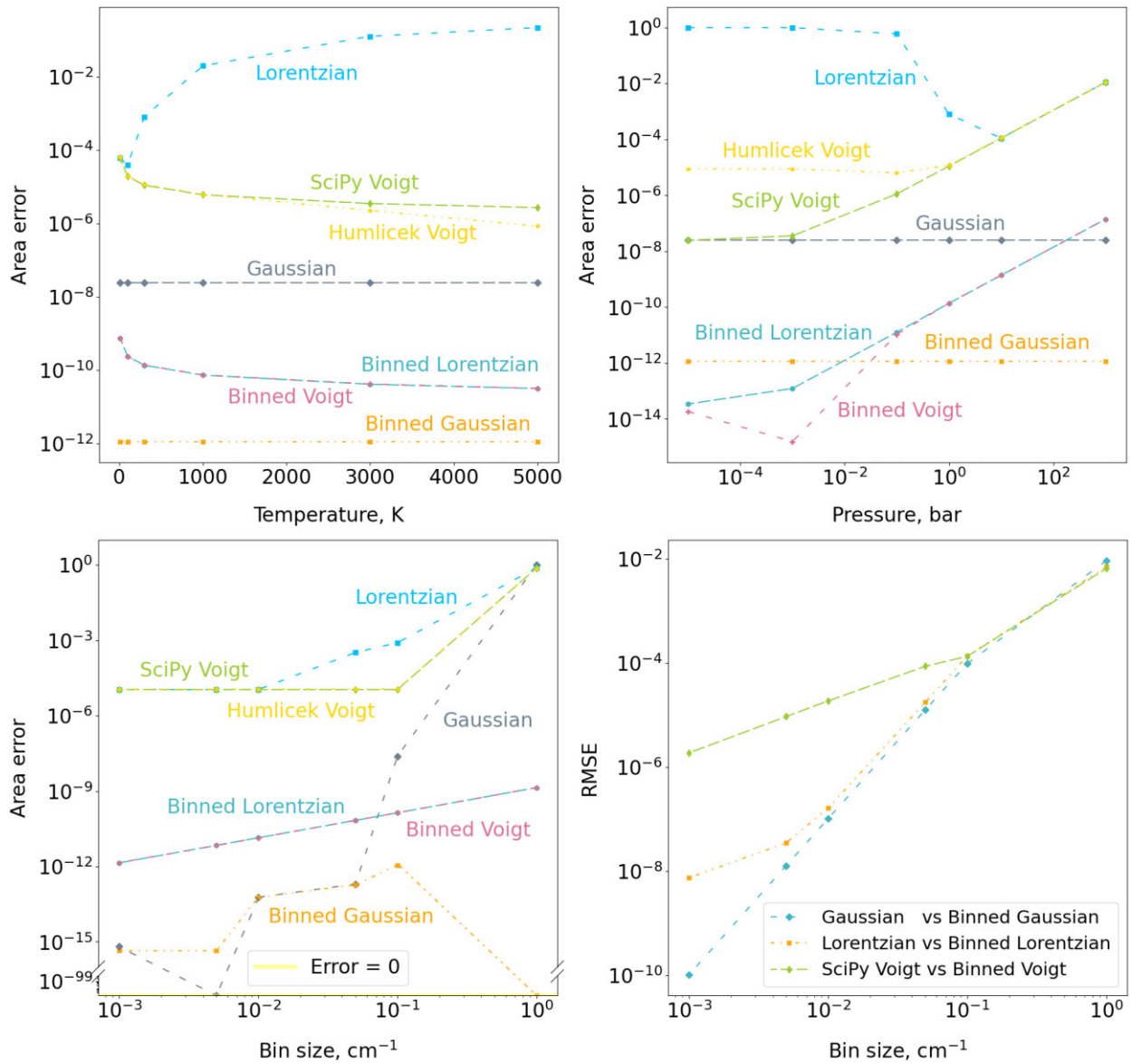


Figure 12. Line profile errors ϵ_{area} (equation 54) and $\epsilon_{\text{sampling}}$ (equation 55) for different line profiles representing the same single normalized transition ($\tilde{\nu}_{if} = 4064.574\,702\text{ cm}^{-1}$), including Gaussian ($\alpha_G = 0.1$), Lorentzian (γ_L is estimated by $\gamma_{\text{ref}} = 0.07$ and $n_L = 0.5$), and Voigt, for the large wavenumber range of $[0, 8000]\text{ cm}^{-1}$ at different temperatures T , pressures P , and bin sizes $\Delta\tilde{\nu}$. Top-left panel: ϵ_{area} for different temperatures T at $P = 1$ bar and number of grid points $N_{\text{points}} = 80\,001$ (bin size $\Delta\tilde{\nu} = 0.1\text{ cm}^{-1}$) without a cutoff. Top-right panel: ϵ_{area} for different pressures P at $T = 300\text{ K}$ and $\Delta\tilde{\nu} = 0.1\text{ cm}^{-1}$ without a cutoff. Bottom-left panel: ϵ_{area} for different number of grid points N_{points} (bin sizes $\Delta\tilde{\nu}$) at $T = 300\text{ K}$, $P = 1$ bar without a cutoff. Bottom-right panel: $\epsilon_{\text{sampling}}$ for different number of grid points N_{points} (bin sizes $\Delta\tilde{\nu}$) at $T = 300\text{ K}$, $P = 1$ bar, and cutoff = 100 cm^{-1} .

The sampling errors of different line profiles $\epsilon_{\text{sampling}}$ are shown in the fourth panel of Fig. 12. While the sampling profiles are almost perfect and perform better with more grid points, the binned profile can introduce a substantial error especially when the resolution is low. Generally, in radiative transfer applications direct sampling line profiles are preferred (Yurchenko et al. 2018a) especially for coarser grids.

We ignore the pseudo-Voigt profiles in this analysis as having too low accuracy.

3.7.2 Comparison of program run times with different line profiles

To compare the program running times for computing cross-sections with different line profiles, in Table 23 we give examples of cross-sections calculations for different numbers of transitions, bin sizes (numbers of grid points), and filters used.

From Table 23, we can find that the Lorentzian profile is the fastest followed by the Doppler, Gaussian, and SCIPY Voigt profiles. Pseudo-Voigt profiles process data with NumExpr (Cooke et al. 2009) which is a fast numerical expression evaluator for NUMPY. As an approximation method commonly in use, the Humlíček Voigt profile runs rather slowly. Because of the increased number of steps in binned profile calculations, the program costs more time than the sampling methods. The SCIPY wofz Voigt, Humlíček Voigt, and Rocco pseudo-Voigt

Table 23. Example of the program run times on 32 cores, in s, for computing cross-sections using different line profiles at temperatures $T = 300$ K, pressures $P = 1$ bar, and cutoff $= 25 \text{ cm}^{-1}$ for the ExoMol H_2O POKAZATEL line list (Polyansky et al. 2018), for the wavenumber range $[20\,000, 22\,000] \text{ cm}^{-1}$. The left-hand panel compares the run times for different numbers of transitions 10^4 , 10^6 , and 10^8 with bin size of 0.1 cm^{-1} (20 001 grid points) without any filters. The middle panel compares the run times for different numbers of transitions 10^4 , 10^6 , and 10^8 with bin size of 0.1 cm^{-1} (20 001 grid points) with uncertainty and threshold filters. The right four columns compare the run times for different bin sizes (0.1, 0.01, 0.005, and 0.001 cm^{-1}) and the numbers of grid points (20 001, 200 001, 400 001, and 2000 001) using 10^8 transitions with uncertainty and threshold filters.

Line profiles	Uncertainty filter: none Threshold filter: none Bin size $\Delta\bar{\nu} = 0.1 \text{ cm}^{-1}$ $N_{\text{points}} = 20\,001$ $N_{\text{transitions}} =$			Uncertainty filter: 0.01 cm^{-1} Threshold filter: $10^{-30} \text{ cm molecule}^{-1}$ Bin size $\Delta\bar{\nu} = 0.1 \text{ cm}^{-1}$ $N_{\text{points}} = 20\,001$ $N_{\text{transitions}} =$			Uncertainty filter: 0.01 cm^{-1} Threshold filter: $10^{-30} \text{ cm molecule}^{-1}$ $N_{\text{transitions}} = 10^8$ $N_{\text{points}} = 2000/\Delta\bar{\nu} + 1$ Bin size (cm^{-1}) $\Delta\bar{\nu} =$		
	10^4	10^6	10^8	10^4	10^6	10^8	0.01	0.005	0.001
Doppler	2.70	6.79	54.63	3.01	2.64	8.00	7.66	9.42	13.43
Binned Doppler	4.37	10.52	114.23	3.98	4.00	10.12	9.45	12.65	19.66
Gaussian	2.84	7.43	52.53	3.74	3.85	9.37	8.72	11.09	15.00
Binned Gaussian	4.42	10.85	119.77	3.85	3.96	10.08	9.39	12.28	20.03
Lorentzian	2.83	7.39	51.69	3.87	3.97	9.47	8.77	10.46	14.15
Binned Lorentzian	4.14	9.16	61.24	3.95	3.95	9.67	9.24	11.31	14.73
SciPy Voigt	2.99	9.77	83.19	3.87	3.91	9.34	8.85	10.83	13.21
SciPy Wofz Voigt	3.10	10.26	121.01	3.80	3.97	9.71	9.09	11.73	17.27
Humlíček Voigt	3.24	10.69	161.25	3.90	4.05	10.15	12.62	20.03	46.62
Thompson pseudo-Voigt	2.99	9.89	85.86	3.80	3.90	10.14	9.27	11.63	19.85
Kielkopf pseudo-Voigt	4.01	10.15	85.98	3.82	3.94	10.08	9.57	13.09	19.72
Olivero pseudo-Voigt	3.93	9.93	88.55	3.81	3.89	10.01	9.58	12.48	20.44
Liu-Lin pseudo-Voigt	3.98	10.03	85.20	3.88	3.99	10.18	9.40	14.02	19.63
Rocco pseudo-Voigt	3.97	10.09	85.47	3.97	3.98	9.83	9.56	13.08	20.16
Binned Voigt	6.19	13.80	657.68	3.95	4.35	10.32	22.38	38.78	107.06

profiles are all have similar accuracy but significantly slower than the SciPy Voigt profile. Binned Voigt profiles are more accurate than standard sampling Voigt profiles, but take much more computer time.

To summarize, considering its higher accuracy and relatively short run times, we recommend the SciPy Voigt profile for calculating cross-sections.

3.7.3 Comparison of program run times with PyExoCross, ExoCross, and RADIS

We tested run times of PyExoCross, ExoCross, and RADIS programs on 32 cores. PyExoCross makes good use of the cores and costs 371 s for reading files, processing data, calculating absorption cross-sections, and saving results for H_2O at temperature $T = 3000$ K, pressure $P = 1$ bar, cutoff $= 25 \text{ cm}^{-1}$, uncertainty $\leq 0.01 \text{ cm}^{-1}$, and threshold $\geq 10^{-40} \text{ cm molecule}^{-1}$; wavenumbers range from 0 to $41\,200 \text{ cm}^{-1}$ with SciPy Voigt line profile using the ExoMol $^1\text{H}_2^{16}\text{O}$ line list of data set POKAZATEL (Polyansky et al. 2018). Fortran program ExoCross costs 3600 s for whole processes with the same conditions; however, only about 300 s of this is actually processing data with the wall clock time dominated by reading and unpacking the files.

RADIS downloads ExoMol compressed bz2 files then converts them into HDF5 files and saves both original files and converted files to local for next time calculating. To compare with RADIS, PyExoCross costs 27 s (including reading compressed bz2 files, processing data, calculating cross-sections, and saving results) and RADIS costs 25 s (not including times for downloading, decompressing, and converting files; includes reading HDF5 files, processing data, and calculating cross-sections) on calculating absorption cross-sections of ExoMol $^1\text{H}_2^{16}\text{O}$ line list of data set POKAZATEL in the wavenumber range $[20\,000, 20\,500] \text{ cm}^{-1}$ at temperature $T = 300$ K, pressure $P = 1$ bar, cutoff $= 10^{-27} \text{ cm}^{-1}$, and threshold $\geq 10^{-40} \text{ cm molecule}^{-1}$ with Voigt line profile.

3.8 Additional functionality

PyExoCross provides the following additional functionality for users: (1) locating line lists in the ExoMol data base with uncertainties provided as part of the States file; (2) checking for updated line lists compared with the last or new version of the master file, or the version the users saved last time; and (3) downloading the States and Transitions files via the ExoMol API instead of downloading files one by one.

4. CALCULATION PROTOCOL

The PyExoCross program processes line lists with the following steps (see Fig. 13):

- (1) Populate the input file with the necessary data and settings to be read by PyExoCross;
- (2) Obtain (download) data from the ExoMol and/or HITRAN websites (if required);

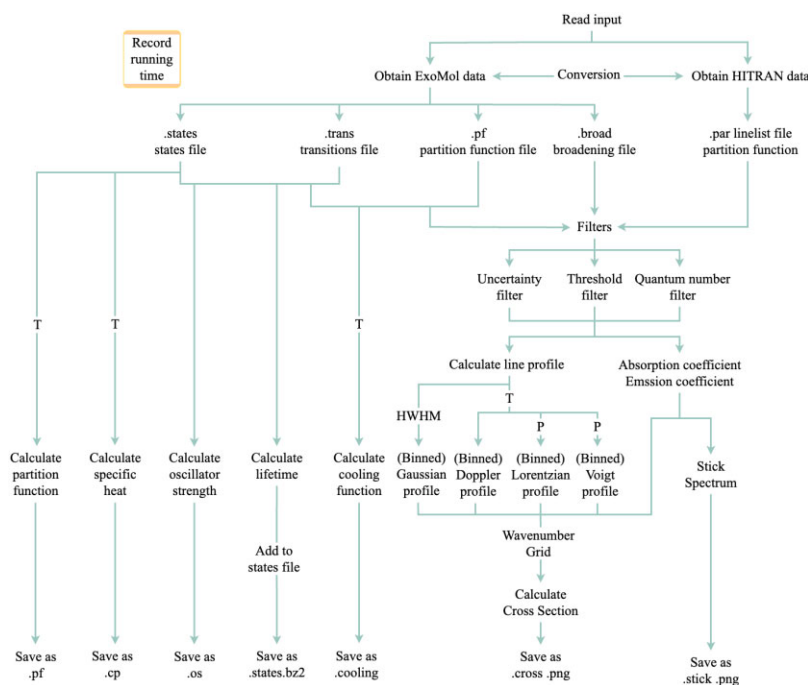


Figure 13. Calculation protocol of the PYEXOCROSS program. *T* and *P* refer to temperature and pressure, separately.

- (3) Read the ExoMol line lists files: (.states) States file, (.trans) Transitions file(s), (.pf) partition function file, and (.broad) broadening file or the HITRAN line lists (.par) files and partition function;
- (4) Compute partition function (if required);
- (5) Compute specific heat (if required);
- (6) Compute cooling function (if required);
- (7) Compute radiative lifetime (if required);
- (8) Compute oscillator strength (if required);
- (9) For intensities, apply the uncertainty or quantum numbers filters (if required);
- (10) Compute a line intensity (absorption or emission coefficient, if required);
- (11) Compute line profile (if required);
- (12) Compute cross-sections on a grid of wavenumbers (if required);
- (13) Save the results of partition functions, specific heats, cooling functions, radiative lifetimes, and cross-sections into the separate files;
- (14) Plot cross-sections (if required);
- (15) Report program running time.

5. DATA FORMAT

PYEXOCROSS can use the input data in either ExoMol or HITRAN format. It can also convert between different formats including HITRAN or ExoMol. Table 24 gives all file types in the ExoMol data base. PYEXOCROSS available file types are described in the aforementioned sections.

5.1 HITRAN format

Compared with the ExoMol format line list files (.states and .trans), the HITRAN format line list file (.par) includes the parameters related to both upper state and lower state. Table 17 gives the format of the HITRAN line-by-line records. The description of the HITRAN parameters is specified in Table 18. PYEXOCROSS processes the HITRAN format data to obtain the parameters for calculating cooling functions, oscillator strength, stick spectra, and cross-sections. The partition function, molar mass, and fractional abundance are available from <https://hitran.org/docs/iso-meta/>. PYEXOCROSS extracts these parameters from the webpage directly and automatically instead of entering these parameters by hand. The partition functions at different temperatures are recorded in the .txt file online which can be mined by PYEXOCROSS.

5.2 Units

The default labels and units used in the PYEXOCROSS program are listed in Table 25. The physical and astronomical constants in calculations are shown in Listing 5. astropy.constants is a sub-package of ASTROPY where ASTROPY is a community Python package for astronomy (Robitaille et al. 2013). Some of the parameters used by PYEXOCROSS are not in SI units but in cgs units (e.g. erg).

Table 24. Specification of the ExoMol file types (contents in brackets are optional) (Tennyson et al. 2023).

File extension	PYEXOCROSS available	N_{files}	File DSname	Contents
.all	Input	1	Master	Single file defining contents of the ExoMol data base.
.def	Input	N_{tot}	Definition	Defines contents of other files for each isotopologue.
.states	Input/Output	N_{tot}	States	Energy levels, quantum numbers, lifetimes, (Landé g -factors, uncertainties).
.trans	Input/Output	a	Transitions	Einstein A coefficients, (wavenumber).
.broad	Input/Output	N_{mol}	Broadening	Parameters for pressure-dependent line profiles.
.cross	Output	b	Cross-sections	Temperature- or temperature and pressure-dependent cross-sections.
.kcoef	No	c	k -coefficients	Temperature and pressure-dependent k -coefficients.
.pf	Input/Output	N_{tot}	Partition function	Temperature-dependent partition function, (cooling function).
.cf	Output	N_{tot}	Cooling function	Temperature-dependent cooling function.
.cp	Output	N_{tot}	Specific heat	Temperature-dependent specific heat.
.dipoles	No	N_{tot}	Dipoles	Transition dipoles including phases.
.super	No	d	Super-lines	Temperature-dependent super-lines (histograms) on a wavenumber grid.
.nm	No	e	VUV cross-sections	Temperature and pressure-dependent VUV cross-sections (wavelength, nm).
.fits, .h5, .kta	No	f	Opacities	Temperature and pressure-dependent opacities for radiative-transfer applications.
.overview	No	N_{mol}	Overview	Overview of data sets available.
.readme	No	N_{iso}	Readme	Specifies data formats.
.model	No	N_{iso}	Model	Model specification.
.cont	No	$N_{\text{iso}}/0$	Continuum	Continuum contribution to the photoabsorption.
.pdef	No	N_{tot}	Definition	Defines contents of .photo files for each isotopologue.
.photo	No	N_{iso}	Photodissociation	Photodissociation cross-sections for each lower state.

Notes. N_{files} : Total number of possible files.

N_{mol} : Number of molecules in the data base.

N_{tot} : It is the sum of N_{iso} for the N_{mol} molecules in the data base.

N_{iso} : Number of isotopologues considered for the given molecule.

^aThere are N_{tot} sets of .trans files but for molecules with large numbers of transitions the .trans files are subdivided into wavenumber regions.

^bThere are N_{cross} sets of .cross files for isotopologue.

^cThere are N_{kcoef} sets of .kcoef files for each isotopologue.

^dThere are N_T sets of T -dependent super-lines.

^eThere are N_{VUV} sets of VUV cross-sections.

^fSet of opacity files in the format native to specific radiative-transfer programs.

Table 25. Default units used in the PYEXOCROSS program.

Quantity	Label	Units
Partition function	$Q(T)$	None
Specific heat	$C_p(T)$	$\text{J K}^{-1} \text{mol}^{-1}$
Cooling function	$W(T)$	$\text{erg (s molecule sr)}^{-1}$
Radiative lifetime	τ_i	s
Oscillator strength	f	None
Molar mass	M	Dalton
Wavenumber		cm^{-1}
Wavelength		μm
Temperature	T	K
Pressure	P	bar
Absorption coefficient	I_{fi}	cm molecule^{-1}
Emission coefficient	ϵ_{if}	$\text{erg (s molecule sr)}^{-1}$
Absorption cross-sections	σ_{ab}	$\text{cm}^2 \text{molecule}^{-1}$
Emission cross-sections	σ_{em}	$\text{erg cm (s molecule sr)}^{-1}$

```

import astropy.constants as ac
Tref = 296.0 # Reference temperature is 296 K
Pref = 1.0 # Reference pressure is 1 bar
N_A = ac.N_A.value # Avogadro number (1/mol)
h = ac.h.to('erg s').value # Planck's constant (erg s)
c = ac.c.to('cm/s').value # Velocity of light (cm/s)
kB = ac.k_B.to('erg/K').value # Boltzmann's const (erg/K)
R = ac.R.to('J / (K mol)').value # Molar gas constant (J/(K mol))
c2 = h * c / kB # Second radiation constant (cm K)

```

Listing 5: Python code for calculating the physical and astronomical constants.

6. CONCLUSIONS

We have developed PYEXOCROSS, a python new program for calculating the spectroscopic properties of molecules with spectral line lists. PYEXOCROSS is the Python version of the Fortran program EXOCROSS (Yurchenko et al. 2018a). Some CPU parallelization is available, at present there is no GPU support. The spectral line lists of the ExoMol data base are used to generate partition functions, specific heats, cooling functions, radiative lifetimes, oscillator strengths, stick spectra, and cross-sections by PYEXOCROSS. At present, PYEXOCROSS can ingest data from the ExoMol, HITRAN, and HITEMP data bases to provide parts of the functionalities supported by the EXOCROSS program. Other features will be developed later, such as other data formats while this project will also implement the processing of other data bases in the future.

Future enhancements to PYEXOCROSS include non-LTE calculations, generation of k -tables, and opacity production. A library of fractional isotopic abundances will be built for the PYEXOCROSS program to calculate the intensities for different distributions of isotopologues. PYEXOCROSS will also support more molecular spectral data bases in future versions, along with support for data type and data format conversions among different data bases.

ACKNOWLEDGEMENTS

We thank Ahmed Al-Refaie for helpful discussions and suggestions on the methods of this work. This work was supported by the European Research Council under Advanced Investigator Project 883830.

DATA AVAILABILITY

The PYEXOCROSS program is publicly accessible on GitHub (<https://github.com/ExoMol/PyExoCross.git>) and its program manual webpage is <https://pyexocross.readthedocs.io/>.

The EXOCROSS Fortran program is freely available from the GitHub (<https://github.com/ExoMol/ExoCross.git>). The EXOCROSS manual webpage is <https://exocross.readthedocs.io/>.

The molecular line lists discussed in this paper and many more are available from the ExoMol website (<https://www.exomol.com/>).

REFERENCES

- Abramowitz M., Stegun I., 1972, Handbook of Mathematical Functions. National Bureau of Standards Applied Mathematics Series - 55. Natl. Bur. Stand. Washington DC
- Anisman L., Chubb K. L., Elsej J., Al-Refaie A., Changeat Q., Yurchenko S. N., Tennyson J., Tinetti G., 2022, *J. Quant. Spectrosc. Radiat. Transfer*, 278, 108013
- Armstrong B., 1967, *J. Quant. Spectrosc. Radiat. Transfer*, 7, 61
- Barber R. J., Strange J. K., Hill C., Polyansky O. L., Mellau G. C., Yurchenko S. N., Tennyson J., 2014, *MNRAS*, 437, 1828
- Barklem P. S., Collet R., 2016, *A&A*, 588, A96
- Barton E. J., Hill C., Czurylo M., Li H.-Y., Hyslop A., Yurchenko S. N., Tennyson J., 2017, *J. Quant. Spectrosc. Radiat. Transfer*, 203, 490
- Chubb K. L. et al., 2021, *A&A*, 646, A21
- Cooke D., Hochberg T., Alted F., Vilata I., Wiebe M., de Menten G., Valentino A., McLeod R., 2009, NumExpr: Fast Numerical Expression Evaluator for NumPy (2.6.8). Zenodo. <https://doi.org/10.5281/zenodo.1492916>
- Coppola C. M., Lodi L., Tennyson J., 2011, *MNRAS*, 415, 487
- Demtröder W., 1981, NASA STI/Recon Tech. Rep. A, 82, 12273
- Di Rocco H. O., Cruzado A., 2012, *Acta Phys. Pol. A*, 122, 666
- Endres C. P., Schlemmer S., Schilke P., Stutzki J., Müller H. S. P., 2016, *J. Mol. Spectrosc.*, 327, 95
- Faddeeva V. N., Terent'ev N. M., Fok V. A., Fry D. G., Hons B. A., 1961, Tables of Values of the Function $\omega(z) = \exp(-z^2)(1 + 2i/\sqrt{\pi} \int_0^z \exp(t^2) dt)$ for Complex Argument. Pergamon Press, Oxford
- Feller W., 1971, An Introduction to Probability Theory and its Applications, Vol. 2, 2nd edn. John Wiley & Sons Inc., New York, p. 704
- Gamache R. R. et al., 2017, *J. Quant. Spectrosc. Radiat. Transfer*, 203, 70
- Gandhi S. et al., 2020, *MNRAS*, 495, 224
- Gérard J.-C. et al., 2023, *Icarus*, 389, 115261
- Gharib-Nezhad E. S. et al., 2024, *RASTI*, 3, 44
- Goldenstein C. S., Miller V. A., Spearrin R. M., Strand C. L., 2017, *J. Quant. Spectrosc. Radiat. Transfer*, 200, 249
- Gordon I. E. et al., 2022, *J. Quant. Spectrosc. Radiat. Transfer*, 277, 107949
- Grimm S. L. et al., 2021, *ApJS*, 253, 30
- Hargreaves R. J. et al., 2019, *J. Quant. Spectrosc. Radiat. Transfer*, 232, 35
- Herbert F., 1974, *J. Quant. Spectrosc. Radiat. Transfer*, 14, 943
- Hill C., 2020, Learning Scientific Programming with Python. Cambridge Univ. Press, Cambridge
- Hill C., Yurchenko S. N., Tennyson J., 2013, *Icarus*, 226, 1673
- Huang X., Schwenke D. W., Lee T. J., 2021, *Acc. Chem. Res.*, 54, 1311
- Hui A. K., Armstrong B. H., Wray A. A., 1978, *J. Quant. Spectrosc. Radiat. Transfer*, 19, 509
- Humlíček J., 1979, *J. Quant. Spectrosc. Radiat. Transfer*, 21, 309
- Humlíček J., 1982, *J. Quant. Spectrosc. Radiat. Transfer*, 27, 437
- Ida T., Ando M., Toraya H., 2000, *J. Appl. Crystallogr.*, 33, 1311
- Irwin A. W., 1981, *ApJS*, 45, 621

- Jacquinet-Husson N. et al., 2011, *J. Quant. Spectrosc. Radiat. Transfer*, 112, 2395
- Johnson N. L., Kotz S., Balakrishnan N., 1994, *Continuous Univariate Distributions*, Vol. 1. John Wiley & Sons Inc., New York
- Jørgensen U. G., Larsson M., Iwamae A., Yu B., 1996, *A&A*, 315, 204
- Karlovets E. V. et al., 2021, *J. Quant. Spectrosc. Radiat. Transfer*, 276, 107896
- Kielkopf J. F., 1973, *J. Opt. Soc. Am.*, 63, 987
- Kochanov R. V., Gordon I., Rothman L., Wcislo P., Hill C., Wilzewski J., 2016, *J. Quant. Spectrosc. Radiat. Transfer*, 177, 15
- Kochanov R. V., Gordon I. E., Rothman L. S., Tan Y., Karns J., Matt W., Hill C., 2018, *BJ McCall Proc. 73rd Int. Symp. Mol. Spectrosc.* Univ. Illinois at Urbana-Champaign, Urbana, IL, p. MI07
- Kuntz M., 1997, *J. Quant. Spectrosc. Radiat. Transfer*, 57, 819
- Li G., Gordon I. E., Rothman L. S., Tan Y., Hu S.-M., Kassi S., Campargue A., Medvedev E. S., 2015, *ApJS*, 216, 15
- Liu Y., Lin J., Huang G., Guo Y., Duan C., 2001, *J. Opt. Soc. Am. B*, 18, 666
- McKinney W. et al., 2010, *Proc. 9th Python Sci. Conf. Vol. 445, Data Structures for Statistical Computing in Python*. TX, Austin, p. 51
- Madhusudhan N., 2019, *ARA&A*, 57, 617
- Martin J. M. L., François J.-P., Gijbels R., 1992, *J. Chem. Phys.*, 96, 7633
- Melnikov V. V., Yurchenko S. N., Tennyson J., Jensen P., 2016, *Phys. Chem. Chem. Phys.*, 18, 26268
- Miller S., Stallard T., Tennyson J., Melin H., 2013, *J. Phys. Chem. A*, 117, 9770
- Müller H. S., Endres C., Stutzki J., Schlemmer S., 2013, *AIP Conf. Proc. Vol. 1545, Eighth International Conference on Atomic and Molecular Data and Their Applications: ICAMDATA-2012*. Am. Inst. Phys., New York, p. 96
- Neale L., Tennyson J., 1995, *ApJ*, 454, L169
- Neale L., Miller S., Tennyson J., 1996, *ApJ*, 464, 516
- Ngo N., Lisak D., Tran H., Hartmann J.-M., 2013, *J. Quant. Spectrosc. Radiat. Transfer*, 129, 89
- Nguyen T. G., Cowan N. B., Pierrehumbert R. T., Lupu R. E., Moores J. E., 2022, *MNRAS*, 513, 6125
- Olivero J. J., Longbothum R., 1977, *J. Quant. Spectrosc. Radiat. Transfer*, 17, 233
- Orloski R., Trigueiros A., 2001, *J. Quant. Spectrosc. Radiat. Transfer*, 69, 13
- Owens A., Tennyson J., Yurchenko S. N., 2021, *MNRAS*, 502, 1128
- Owens A., Dooley S., McLaughlin L., Tan B., Zhang G., Yurchenko S. N., Tennyson J., 2022, *MNRAS*, 511, 5448
- Owens A., He T., Hanicinec M., Hill C., Mohr S., Tennyson J., 2023, *Plasma Sources Sci. Technol.*, 32, 085015
- Pavlenko Y. V., Yurchenko S. N., Tennyson J., 2020, *A&A*, 633, A52
- Pavlenko Y., Tennyson J., Yurchenko S. N., Schmidt M. R., Jones H. R. A., Lyubchik Y., Suárez Mascareño A., 2022, *MNRAS*, 516, 5655
- Pickett H. M., Poynter R. L., Cohen E. A., Delitsky M. L., Pearson J. C., Müller H. S. P., 1998, *J. Quant. Spectrosc. Radiat. Transfer*, 60, 883
- Pierluissi J. H., Vanderwood P. C., Gomez R. B., 1977, *J. Quant. Spectrosc. Radiat. Transfer*, 18, 555
- Polyansky O. L., Kyuberis A. A., Zobov N. F., Tennyson J., Yurchenko S. N., Lodi L., 2018, *MNRAS*, 480, 2597
- Qu Q., Yurchenko S. N., Tennyson J., 2021, *MNRAS*, 504, 5768
- Rey M., Nikitin A. V., Babikov Y. L., Tyuterev V. G., 2016, *J. Mol. Spectrosc.*, 327, 138
- Robitaille T. P. et al., 2013, *A&A*, 558, A33
- Rothman L. et al., 2010, *J. Quant. Spectrosc. Radiat. Transfer*, 111, 2139
- Rothman L. S. et al., 2005, *J. Quant. Spectrosc. Radiat. Transfer*, 96, 139
- Rustankulov Z. et al., 2023, *Nature*, 614, 659+
- Sánchez-Bajo F., Cumbreira F., 1997, *J. Appl. Crystallogr.*, 30, 427
- Sauval A. J., Tatum J. B., 1984, *ApJS*, 56, 193
- Schreier F., 2011, *J. Quant. Spectrosc. Radiat. Transfer*, 112, 1010
- Schreier F., 2017, *J. Quant. Spectrosc. Radiat. Transfer*, 187, 44
- Schreier F., 2018, *MNRAS*, 479, 3068
- Schreier F., Hochstaff P., 2021, *J. Quant. Spectrosc. Radiat. Transfer*, 258, 107385
- Sharp C. M., Burrows A., 2007, *ApJS*, 168, 140
- Somogyi W., Yurchenko S. N., Yachmenev A., 2021, *J. Chem. Phys.*, 155, 214303
- Sousa-Silva C., Hesketh N., Yurchenko S. N., Hill C., Tennyson J., 2014, *J. Quant. Spectrosc. Radiat. Transfer*, 142, 66
- Squires G. L., 2001, *Practical Physics*. Cambridge Univ. Press, Cambridge
- Tan Y., Kochanov V. R., Rothman L. S., Gordon I. E., 2019, *J. Quant. Spectrosc. Radiat. Transfer*, 124, 11580
- Tan Y., Skinner F. M., Samuels S., Hargreaves R. J., Hashemi R., Gordon I. E., 2022, *ApJS*, 262, 40
- Tennyson J., Miller S., 1989, *Comput. Phys. Commun.*, 55, 149
- Tennyson J., Yurchenko S. N., 2012, *MNRAS*, 425, 21
- Tennyson J., Yurchenko S. N., 2022, *Frontiers Astron. Space Sci.*, 8, 795040
- Tennyson J., Miller S., Le Sueur C. R., 1993, *Comput. Phys. Commun.*, 75, 339
- Tennyson J., Hill C., Yurchenko S. N., 2013, *AIP Conf. Proc. Vol. 1545, Eighth International Conference on Atomic and Molecular Data and their Applications: ICAMDATA-2012*. Am. Inst. Phys., New York, p. 186
- Tennyson J., Hulme K., Naim O. K., Yurchenko S. N., 2016a, *J. Phys. B: At. Mol. Opt. Phys.*, 49, 044002
- Tennyson J. et al., 2016b, *J. Mol. Spectrosc.*, 327, 73
- Tennyson J. et al., 2020, *J. Quant. Spectrosc. Radiat. Transfer*, 255, 107228
- Tennyson J., Pezzella M., Zhang J., Yurchenko S. N., 2023, *RASTI*, 2, 231
- Thompson P., Cox D., Hastings J., 1987, *J. Appl. Crystallogr.*, 20, 79
- Tinetti G., Encrenaz T., Coustenis A., 2013, *A&AR*, 21, 1
- Tóbiás R., Furtenbacher T., Császár A. G., Naumenko O. V., Tennyson J., Flaud J.-M., Kumard P., Poirier B., 2018, *J. Quant. Spectrosc. Radiat. Transfer*, 208, 152
- Underwood D. S., Tennyson J., Yurchenko S. N., Huang X., Schwenke D. W., Lee T. J., Clausen S., Fateev A., 2016, *MNRAS*, 459, 3890
- van den Bekerom D., Pannier E., 2021, *J. Quant. Spectrosc. Radiat. Transfer*, 261, 107476
- Virtanen P. et al., 2020, *Nature Methods*, 17, 261
- Wang Y., Ferland G. J., Lykins M., Porter R., van Hoof P. A., Williams R., 2014, *MNRAS*, 440, 3100

- Wang R., Balcunaite U., Chen J., Yuan C., Owens A., Tennyson J., 2023, *J. Quant. Spectrosc. Radiat. Transfer*, 306, 108617
- Weislo P. et al., 2021, *J. Quant. Spectrosc. Radiat. Transfer*, 260, 107477
- Western C. M., 2017, *J. Quant. Spectrosc. Radiat. Transfer*, 186, 221
- Wilzewski J. S., Gordon I. E., Kochanov R. V., Hill C., Rothman L. S., 2016, *J. Quant. Spectrosc. Radiat. Transfer*, 168, 193
- Yurchenko S. N., Tennyson J., Bailey J., Hollis M. D. J., Tinetti G., 2014, *Proc. Natl. Acad. Sci.*, 111, 9379
- Yurchenko S. N., Al-Refai A. F., Tennyson J., 2018a, *A&A*, 614, A131
- Yurchenko S. N., Williams H., Leyland P. C., Lodi L., Tennyson J., 2018b, *MNRAS*, 479, 1401
- Yurchenko S. N., Mellor T. M., Freedman R. S., Tennyson J., 2020, *MNRAS*, 496, 5282
- Yurchenko S. N., Szajna W., Hakalla R., Semenov M., Sokolov A., Tennyson J., Pavlenko Y., Schmidt M. R., 2024, *MNRAS*, 527, 9736
- Zaghloul M. R., 2017, *ACM Trans. Math. Softw.*, 44, 22

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.