

Data structures for ExoMol: Molecular line lists for exoplanet and other atmospheres

Jonathan Tennyson, Christian Hill, and Sergei N. Yurchenko

Citation: *AIP Conf. Proc.* **1545**, 186 (2013); doi: 10.1063/1.4815853

View online: <http://dx.doi.org/10.1063/1.4815853>

View Table of Contents: <http://proceedings.aip.org/dbt/dbt.jsp?KEY=APCPCS&Volume=1545&Issue=1>

Published by the AIP Publishing LLC.

Additional information on AIP Conf. Proc.

Journal Homepage: <http://proceedings.aip.org/>

Journal Information: http://proceedings.aip.org/about/about_the_proceedings

Top downloads: http://proceedings.aip.org/dbt/most_downloaded.jsp?KEY=APCPCS

Information for Authors: http://proceedings.aip.org/authors/information_for_authors

ADVERTISEMENT



AIPAdvances

Submit Now

**Explore AIP's new
open-access journal**

- **Article-level metrics
now available**
- **Join the conversation!
Rate & comment on articles**

Data structures for ExoMol: Molecular line lists for exoplanet and other atmospheres

Jonathan Tennyson, Christian Hill and Sergei N. Yurchenko

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

Abstract. At elevated temperatures the spectra of polyatomic molecules become extremely complicated with millions, or even billions, of transitions potentially playing an important role. The atmospheres of cool stars and "hot Jupiter" extrasolar planets are rich with molecules in the temperature range 1000 to 3000 K and their properties are strongly influenced by the infrared and visible spectra of these molecules. Access to extensive lists of transitions is essential for interpreting even the rather simple spectra that can be obtained from exoplanets. So far there are extensive, reliable lists of spectral lines for a number species including some stable diatomics, water and ammonia. Data are almost completely lacking for many key species such as methane. The ExoMol project aims to construct line lists of molecular transitions suitable for spectroscopic and atmospheric modelling of cool stars and exoplanets. At high temperatures it is necessary to consider huge numbers of lines even for a single species. Examples of line lists are given; data protocols defined and data handling issues which arise from trying to distribute these huge datasets discussed. In particular, a uniform but flexible format is given for the representation of line lists and cross sections resulting from the ExoMol project.

Keywords: Molecular spectroscopy

PACS: 33.20.-t, 31.15.ag,31.15.V-, 31.50.-x, 33.15.Mt

INTRODUCTION

At elevated temperatures the spectra of polyatomic molecules become extremely complicated with millions, or even billions, of transitions potentially playing an important role. Such spectra are important in many environments including astrophysical ones which are both hot and rich in molecules such as the atmospheres of cool stars and brown dwarfs, and the atmospheres extrasolar planets, particularly "hot Jupiters". Similar data are needed to study radiative transport and the spectral signatures of flames, engines and exhaust gases to name just a few examples. The data demands for modelling the radiative transport and spectra by bodies which are rich with molecules in the temperature range 1000 to 3000 K are thus huge [1]. Access to extensive lists of transitions is essential for interpreting even the rather simple spectra that can be obtained from exoplanets [2, 3]. So far our group has generated extensive, reliable lists of spectral lines for a number species including some stable diatomics [4, 5, 6, 7], water [8], H_3^+ [9], HCN/HNC [10], and recently ammonia [11]. However there are a number of key molecules for which data are almost completely lacking, of which methane is a particularly pressing example [3].

The ExoMol project [6] aims to construct line lists of molecular transitions suitable for spectroscopic and atmospheric modelling of cool stars and exoplanets. As at high temperatures it is necessary to consider huge numbers of lines even for a single species,

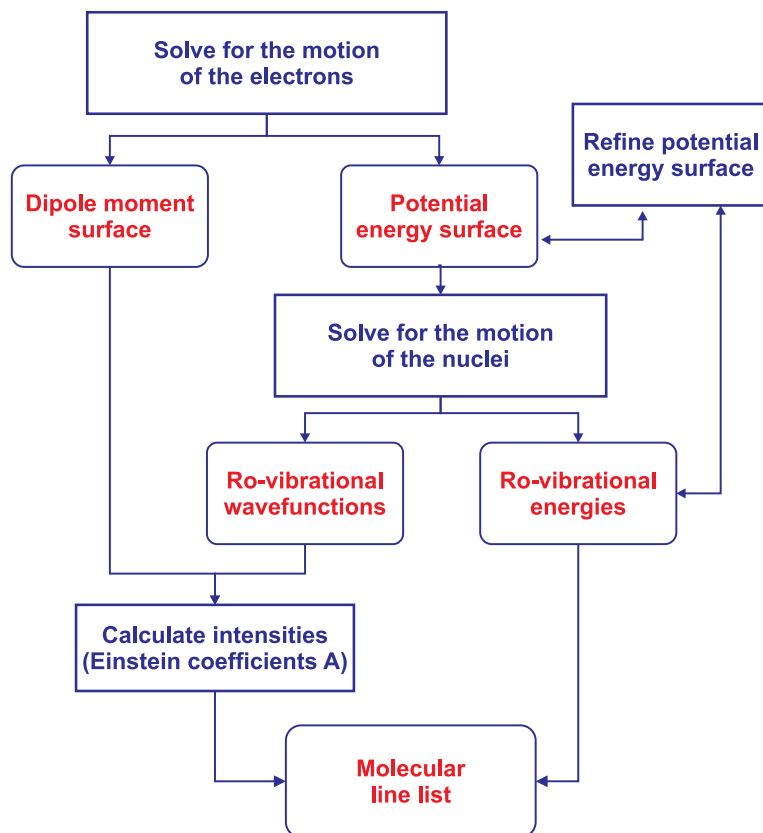


FIGURE 1. Schematic representation of the general procedure used by the ExoMol project to produce molecular line lists.

direct determination of these line lists by laboratory measurements is problematic. Line lists are therefore best computed on the basis of robust theoretical models tested against available laboratory data rather than constructed experimentally.

METHODOLOGY

Our general strategy for generating line lists is illustrated in Fig. 1. The strategy is based on the use of high quality *ab initio* electronic structure calculations with standard quantum chemistry packages [12, 13] and direct solution of the resulting nuclear motion problem. For diatomics this is done by direct numerical integration of single [14] or coupled [15] potential energy curves; for triatomics [16] and some tetratomics [16] variational solutions based on the use of exact kinetic operators [17] are used; for larger systems numerically defined kinetic energy operators as calculated by the program TROVE [18] are employed.

With a few exceptions [4, 5, 9], the use of pure first-principles quantum mechanics does not yield energy levels and hence transition frequencies of the desired accuracy. Energies levels can be improved using one of three strategies. Firstly, the *ab initio* potential energy surface can be improved by fitting to experimental data using appropriately

defined procedures [19, 20, 21]. Secondly, it is possible, at least in TROVE, to substitute observed vibrational band origins into the calculation immediately before the step which generates the rotational energy levels [11]. Finally, experimental energy levels determined, for example, using the MARVEL procedure [22], can be substituted in place of the calculated ones, see [4, 23] for example.

While *ab initio* procedures for determining energy levels are improved by the use of experimental data, this is not so for transition intensities. Systematic tests [24] have shown that it is hard to improve on high quality, *ab initio* dipole moment surfaces by fitting to experimental data. Indeed there is growing evidence [25, 26, 27, 28] that carefully determined dipole moment surfaces can be calculated which give transition intensities with an accuracy competitive with, or possibly better than, experiment.

A particular feature of the line lists required for modelling molecular spectra in hot atmospheres is their sheer size. For example, the most complete available line list for the diatomic TiO contains 58 million transitions [29], while both the BT2 line list for water [8] and the recent CDS-4000 for CO₂ [30] each contain over 500 million lines. The problem gets worse for large molecules: the recent BYTe hot ammonia line list has over a billion transitions [11]. Although attempts have been made to reduce these numbers by importance sampling [31], these do not change the orders of magnitudes involved and, indeed, run the risk of deleting important transitions: see [32] for an example. Data handling for data sets of this size therefore becomes a serious problem. For this reason the remainder of this article is concerned with the issues of data formats and data compaction.

DATA FORMAT FOR EXOMOL

Up until now there has been no uniform format for providing line lists. Instead the preferred format has evolved with time and hence changed, albeit often only slightly, from line list to line list. Below we propose a uniform but flexible format which we plan to use for all future large line lists. For small problems, such as diatomics with less than a million transitions, use of single line list file will continue to be sufficient. For other published lists we will provide updated versions in the new format via our website, <http://www.exomol.com>.

The division of the line list data into a states file and a transitions file was originally employed by Viti *et al.* in their follow-up to the VTP1 water line list [33] as a means to facilitate the efficient sorting of large number of transitions, with associated tags, into wavelength order. However it was soon realised that this format gave advantages which go beyond simple compactness of data representation. In particular, the use of a separate states file allows the *post hoc* manipulation of the energy data which has been used, for example, to both improve the energies in the HCN/HNC line list [34] and to transform an H¹²C¹⁴N line list into one appropriate for H¹³C¹⁴N [23].

All data files are provided in text format, using only the 95 printable ASCII characters (decimal codes 32-126). Each data *record* consists of fixed-width data columns which are separated by *single spaces* (not tabs) into *fields*. Missing data for any field are indicated by a single asterisk, *.

Transitions are divided into two types of file. One, with the extension `.states`,

TABLE 1. Specification of the mandatory part of the states file.

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†

* The state ID is a non-negative integer index, starting at 1

† J is quantum number for the total angular momentum excluding nuclear spin and is a good quantum number (in the absence of hyperfine coupling).

contains the molecular states, each of which is identified by an integer ID. The other, with the extension `.trans`, references the upper and lower states of a transition using their IDs, and provides the Einstein A -coefficient for the transition.

ExoMol also provides absorption cross section data in a file with extension `.sigma`, described below.

.states file format

The first four fields of the states file are mandatory for all records; these are illustrated in Table 1. Subsequent fields are molecule-dependent and documented in a `README` file supplied with the data. The good quantum numbers and labels should precede the approximate ones; for example, where *parity* is a helpful label it should occur in a field preceding any vibrational quantum numbers. Note that we have chosen to follow the convention, as adopted in the HITRAN database [35], that the state degeneracy, g_{tot} , is the *total* statistical degeneracy including any nuclear spin degeneracy resulting from unresolved hyperfine coupling and the $(2J + 1)$ factor from the rotational motion of the molecule.

The J quantum number may be a non-negative integer or half-odd integer depending on the state, and so either integer or floating-point format may be used for its field.

Finally, we recognize that in some cases transition frequencies and hence energy levels are required to better than $1 \times 10^{-1} \text{ cm}^{-1}$. In these cases higher precision energy levels will be placed in an extra field beyond the mandatory 4 fields. This will be noted in the accompanying `readme` file.

.trans file format

Table 2 gives the specification of the transitions file. There are three mandatory fields: the first two label the upper and lower state IDs and the third holds the transition's Einstein A -coefficient. Each record represents a single quantum transition, and the records are sorted in ascending order of wavenumber. Very large datasets may be split into more than one transitions file; the filename (and `README` file) should indicate the wavenumber range of the data in each file.

TABLE 2. Specification of the transitions file.

Field	Fortran Format	C Format	Description
i'	I12	%12d	Upper state ID
i''	I12	%12d	Lower state ID
A	ES10.4	%10.4e	Einstein A -coefficient in s^{-1}

TABLE 3. Specification of the absorption cross sections file.

Field	Fortran Format	C Format	Description
$\tilde{\nu}_i$	F12.6	%12.6f	Central wavenumber of bin in cm^{-1}
σ	ES14.8	%14.8e	Absorption cross section in $cm^2 molec^{-1}$

.sigma file format

Absorption cross section data (σ) are provided by the ExoMol project as a function of wavenumber ($\tilde{\nu}$, in cm^{-1}) in units of $cm^2 molec^{-1}$ [36]. The ExoMol website allows cross sections to be produced at a user-defined resolution (binning interval) and wavenumber range, and output is in a file with the `.sigma` extension. This file consists of a series of either two-field records ($\tilde{\nu}_i$ and σ_i , separated by a single space) or single field records (one column of values, σ_i). In the latter case, the wavenumber corresponding to the centre of each bin may be calculated as

$$\tilde{\nu}_i = \tilde{\nu}_{\min} + (i - 1)\Delta\tilde{\nu}; \quad i = 1, 2, \dots, n$$

for the n wavenumber bins, each of width $\Delta\tilde{\nu}$, with the first at $\tilde{\nu}_{\min} \equiv \tilde{\nu}_1$. The values of $\tilde{\nu}_{\min}$, $\tilde{\nu}_{\max} \equiv \tilde{\nu}_n$ and $\Delta\tilde{\nu}$ are given in the `.sigma` file name and in an auxiliary file supplied with the `.sigma` file. The field formats are given in Table 3.

EXAMPLES

The BT2 water line list

The BT2 high-temperature line list [8] for $H_2^{16}O$ is presented to users as a `.states` file, and 16 `.trans` files. The transitions file is split into wavelength regions simply to aid downloads. Table 4 gives a portion of the BT2 states file and defines the fields used in this case. The fields give both rigorous quantum numbers as specified in Table 1 and the approximate quantum numbers for water in standard notation. Note that for most states the approximate quantum numbers are not actually known, as illustrated by line 980 in the example.

Table 5 gives a portion of the transitions file in the new format, which in this case corresponds to the format used when the line list was originally published [8].

Example: the H_3^+ absorption cross section

The absorption cross section of H_3^+ in the region of its fundamental vibrational modes, calculated at three different temperatures is shown in Fig. 2. A portion of the `.sigma` file for the 300 K cross section, in two-column format is shown below.

```
3100.000000 1.72265132e-20
3110.000000 1.35347996e-20
3120.000000 2.40026647e-20
3130.000000 1.27560496e-20
3140.000000 1.01366742e-22
3150.000000 2.83699709e-25
```

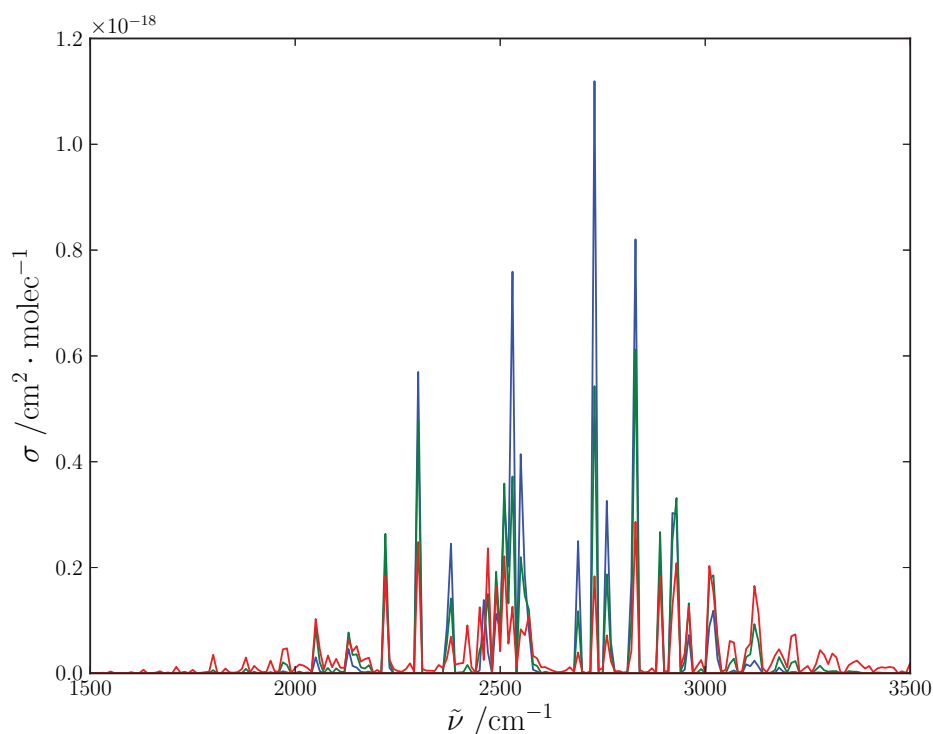


FIGURE 2. Comparison of the absorption cross section of H_3^+ at 300 K (blue), 500 K (green) and 1000 K (red), binned to a 10 cm^{-1} wavenumber grid.

CONCLUSIONS

The aim of the ExoMol project is the generation and distribution of comprehensive molecular line lists. Ideally one would want to do this using the protocols offered by the Virtual Atomic and Molecular Data Centre (VAMDC) project [37]. However the size of the datasets involved means that this is not at present practicable. As a compromise we

developed procedures, mentioned above, which transform the huge line list into much more manageable, temperature-dependent cross sections [36]. These offer the advantage of being small enough to quickly download and relatively straightforward to integrate into modelling codes. Our data in this format are integrated into the VAMDC by virtue of the option to download cross sections in the VAMDC's "XSAMS" format. However it has already been shown [38] that pressure-broadening effects are important for models of exoplanet atmospheres. At present the cross sections are generated using a zero pressure model and adding pressure effects, even approximately, represents a significant complication.

ACKNOWLEDGMENTS

We thank other members of the ExoMol team for the input into this work. This work was performed as part of ERC Advanced Investigator Project 267219 and the project VAMDC which is funded by the European Union INFRA-2008-1.2.2 Scientific Data Infrastructure program under Grant Agreement number 239108.

REFERENCES

1. C. M. Sharp, and A. Burrows, *Astrophys. J. Suppl. Ser.* **168**, 140–166 (2007).
2. G. Tinetti, A. Vidal-Madjar, M.-C. Liang, J.-P. Beaulieu, Y. Yung, S. Carey, R. J. Barber, J. Tennyson, I. Ribas, N. Allard, G. E. Ballester, D. K. Sing, and F. Selsis, *Nature* **448**, 169–171 (2007).
3. M. R. Swain, G. Vasisht, and G. Tinetti, *Nature* **452**, 329–331 (2008).
4. A. E. Engel, N. Doss, G. J. Harris, and J. Tennyson, *Mon. Not. R. Astr. Soc.* **357**, 471–477 (2005).
5. C. M. Coppola, L. Lodi, and J. Tennyson, *Mon. Not. R. Astr. Soc.* **415**, 487–493 (2011).
6. J. Tennyson, and S. N. Yurchenko, *Mon. Not. R. Astr. Soc.* **425**, 21–33 (2012).
7. B. Yadin, T. Vaness, P. Conti, C. Hill, S. N. Yurchenko, and J. Tennyson, *Mon. Not. R. Astr. Soc.* **425**, 34–43 (2012).
8. R. J. Barber, J. Tennyson, G. J. Harris, and R. N. Tolchenov, *Mon. Not. R. Astr. Soc.* **368**, 1087–1094 (2006).
9. L. Neale, S. Miller, and J. Tennyson, *Astrophys. J.* **464**, 516–520 (1996).
10. G. J. Harris, O. L. Polyansky, and J. Tennyson, *Astrophys. J.* **578**, 657–663 (2002).
11. S. N. Yurchenko, R. J. Barber, and J. Tennyson, *Mon. Not. R. Astr. Soc.* **413**, 1828–1834 (2011).
12. H. J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz, et al., MOLPRO, a package of ab initio programs (2010), see <http://www.molpro.net/>.
13. F. Aquilante, N. F. L. De Vico, G. Ghigo, P.-A. Malmqvist, P. Neogrady, T. B. Pedersen, M. Pitonak, M. Reiher, B. O. Roos, L. Serrano-Andrés, M. Urban, V. Veryazov, and R. Lindh, *J. Comput. Chem.* **31**, 224 (2010).
14. R. J. Le Roy, *LEVEL 8.0 A Computer Program for Solving the Radial Schrödinger Equation for Bound and Quasibound Levels*, University of Waterloo Chemical Physics Research Report CP-663, <http://leroy.uwaterloo.ca/programs/> (2007).
15. S. N. Yurchenko, L. Lodi, A. T. Patrascu, and A. V. Stoliarov, *DUO: a programme for calculating rovibronic spectra for diatomics* (2012).
16. I. N. Kozin, M. M. Law, J. Tennyson, and J. M. Hutson, *Comput. Phys. Commun.* **163**, 117–131 (2004).
17. J. Tennyson, *Computer Phys. Reports* **4**, 1–36 (1986).
18. S. N. Yurchenko, W. Thiel, and P. Jensen, *J. Mol. Spectrosc.* **245**, 126–140 (2007).
19. R. J. Le Roy, *DPotFit 1.1 A Computer Program for Fitting Diatomic Molecule Spectral Data to Potential Energy Functions*, University of Waterloo Chemical Physics Research Report CP-662R, <http://leroy.uwaterloo.ca/programs/> (2006).

20. S. N. Yurchenko, B. A. Voronin, R. N. Tolchenov, N. Doss, O. V. Naumenko, W. Thiel, and J. Tennyson, *J. Chem. Phys.* **128**, 044312 (2008).
21. S. N. Yurchenko, R. J. Barber, J. Tennyson, W. Thiel, and P. Jensen, *J. Mol. Spectrosc.* **268**, 123–129 (2011).
22. T. Furtenbacher, A. G. Császár, and J. Tennyson, *J. Mol. Spectrosc.* **245**, 115–125 (2007).
23. G. J. Harris, F. C. Lerner, J. Tennyson, B. M. Kaminsky, Y. V. Pavlenko, and H. R. A. Jones, *Mon. Not. R. Astr. Soc.* **390**, 143–148 (2008).
24. A. E. Lynas-Gray, S. Miller, and J. Tennyson, *J. Mol. Spectrosc.* **169**, 458–467 (1995).
25. O. Asvany, E. Hugo, S. Schlemmer, F. Muller, F. Kuhnemann, S. Schiller, and J. Tennyson, *J. Chem. Phys.* **127**, 154317 (2007).
26. L. Lodi, J. Tennyson, and O. L. Polyansky, *J. Chem. Phys.* **135**, 034113 (2011).
27. L. Lodi, and J. Tennyson, *J. Quant. Spectrosc. Radiat. Transf.* **113**, 850–858 (2012).
28. M. Grechko, O. Aseev, T. R. Rizzo, N. F. Zobov, L. Lodi, J. Tennyson, O. L. Polyansky, and O. V. Boyarkina, *J. Chem. Phys.* **136**, 244308 (2012).
29. D. Schwenke, *Faraday Discuss.* **109**, 321–334 (1998).
30. S. A. Tashkun, and V. I. Perevalov, *J. Quant. Spectrosc. Radiat. Transf.* **112**, 1403–1410 (2011).
31. L. S. Rothman, I. E. Gordon, R. J. Barber, H. Dothe, R. R. Gamache, A. Goldman, V. I. Perevalov, S. A. Tashkun, and J. Tennyson, *J. Quant. Spectrosc. Radiat. Transf.* **111**, 2139–2150 (2010).
32. H. Kreckel, D. Schwalm, J. Tennyson, A. Wolf, and D. Zajfman, *New J. Phys.* **6**, 151 (2004).
33. S. Viti, J. Tennyson, and O. L. Polyansky, *Mon. Not. R. Astr. Soc.* **287**, 79–86 (1997).
34. G. J. Harris, J. Tennyson, B. M. Kaminsky, Y. V. Pavlenko, and H. R. A. Jones, *Mon. Not. R. Astr. Soc.* **367**, 400–406 (2006).
35. L. S. Rothman, I. E. Gordon, A. Barbe, D. C. Benner, P. F. Bernath, M. Birk, V. Boudon, L. R. Brown, A. Campargue, J. P. Champion, K. Chance, L. H. Coudert, V. Dana, V. M. Devi, S. Fally, J. M. Flaud, R. R. Gamache, A. Goldman, D. Jacquemart, I. Kleiner, N. Lacome, W. J. Lafferty, J. Y. Mandin, S. T. Massie, S. N. Mikhailenko, C. E. Miller, N. Moazzen-Ahmadi, O. V. Naumenko, A. V. Nikitin, J. Orphal, V. I. Perevalov, A. Perrin, A. Predoi-Cross, C. P. Rinsland, M. Rotger, M. Simeckova, M. A. H. Smith, K. Sung, S. A. Tashkun, J. Tennyson, R. A. Toth, A. C. Vandaele, and J. Vander Auwera, *J. Quant. Spectrosc. Radiat. Transf.* **110**, 533–572 (2009).
36. J. Tennyson, P. F. Bernath, L. R. Brown, A. Campargue, M. R. Carleer, A. G. Császár, L. Daumont, R. R. Gamache, J. T. Hodges, O. V. Naumenko, O. L. Polyansky, L. S. Rothman, A. C. Vandaele, N. F. Zobov, A. R. Al Derzi, C. Fábri, A. Z. Fazliev, T. Furtenbacher, I. E. Gordon, L. Lodi, and I. I. Mizus, *J. Quant. Spectrosc. Radiat. Transf.* **117**, 29–80 (2013).
37. M. L. Dubernet, V. Boudon, J. L. Culhane, M. S. Dimitrijevic, A. Z. Fazliev, C. Joblin, F. Kupka, G. Leto, P. Le Sidaner, P. A. Loboda, H. E. Mason, N. J. Mason, C. Mendoza, G. Mulas, T. J. Millar, L. A. Nunez, V. I. Perevalov, N. Piskunov, Y. Ralchenko, G. Rixon, L. S. Rothman, E. Roueff, T. A. Ryabchikova, A. Ryabtsev, S. Sahal-Brechot, B. Schmitt, S. Schlemmer, J. Tennyson, V. G. Tyuterev, N. A. Walton, V. Wakelam, and C. J. Zeippen, *J. Quant. Spectrosc. Radiat. Transf.* **111**, 2151–2159 (2010).
38. G. Tinetti, J. Tennyson, C. A. Griffiths, and I. Waldmann, *Phil. Trans. Royal Soc. London A* **370**, 2749–2764 (2012).

TABLE 4. BT2 [8] states file: portion of the file (upper part); field specification (lower part).

1	0.000000	1	0 + 1	1 p A1	0	0	0	0	0	0
2	1594.873096	1	0 + 1	2 p A1	0	1	0	0	0	0
3	3151.677108	1	0 + 1	3 p A1	0	2	0	0	0	0
4	3657.155752	1	0 + 1	4 p A1	1	0	0	0	0	0
...										
980	29437.246861	3	1 - 2	390 p A2	*	*	*	1	*	*
...										
Field				Fortran Format	C Format	Description				
<i>i</i>				I12	%12d	State ID				
<i>E</i>				F12.6	%12.6f	State energy in cm ⁻¹				
<i>g</i> _{tot}				I6	%6d	State degeneracy				
<i>J</i>				I7	%7d	<i>J</i> -quantum number				
parity				A1	%1s	Total parity*				
<i>i</i> _Γ				I2	%2d	Symmetry block ID				
<i>N</i>				I10	%10d	Index within symmetry block				
ns_isomer				A1	%1s	Nuclear spin isomer [†]				
Γ _{rve}				A2	%2s	Rovibronic symmetry label**				
<i>v</i> ₁				I2	%2d	Vibrational quantum number				
<i>v</i> ₂				I2	%2d	Vibrational quantum number				
<i>v</i> ₃				I2	%2d	Vibrational quantum number				
<i>K</i> _a				I2	%2d	Rotational quantum number				
<i>K</i> _c				I2	%2d	Rotational quantum number				

* total parity is the parity of the total molecular wavefunction (excluding nuclear spin) with respect to inversion through the molecular centre of mass of all particles' coordinates in the laboratory coordinate system, the E^* operation.

[†] 'o' for ortho-H₂O or 'p' for para-H₂O

** irreducible representation of the rovibronic wavefunction in the $C_{2v}(M)$ molecular symmetry group

TABLE 5. Portion of the first transitions file, covering the region $0 - 250 \text{ cm}^{-1}$, from the BT2 line list [8]. The columns contain the upper state ID, i , lower state ID, j and the Einstein A -coefficient linking these two states.

i	j	$A_{ij}(\text{s}^{-1})$
36534	35001	6.4700e-22
36356	34807	1.4010e-20
160476	165084	3.5010e-26
173882	169691	2.3110e-24
26507	25003	1.3020e-19