Journal of Quantitative Spectroscopy & Radiative Transfer 315 (2024) 108878

Contents lists available at ScienceDirect

Journal of Quantitative Spectroscopy and Radiative Transfer

journal homepage: www.elsevier.com/locate/jqsrt

Spectroscopic heat maps reveal how to design experiments to improve the uncertainties of transitions and energy levels present in line-by-line databases

Péter Árendás [a,b,*], Tibor Furtenbacher [b,c], Attila G. Császár [b,c]

[a] Budapest Business University, Budapest, Hungary
[b] Institute of Chemistry, ELTE Eötvös Loránd University, Budapest, Hungary
[c] HUN-REN–ELTE Complex Chemical Systems Research Group, Budapest, Hungary

ARTICLE INFO

Keywords:
HITRAN
Spectroscopic networks
Spectroscopic heat map
Transition uncertainties

ABSTRACT

Spectroscopic information systems contain enormous sets of line-by-line (LBL) data on characteristics of rovibronic transitions. Developers of these datasets are continuously improving transitions data, for example, *via* lowering the uncertainties of the transitions. Improved transition uncertainties may also result in a decrease in the uncertainties of the energies of some of the quantum states. This paper introduces a method that assesses the existing line lists and suggests wavenumber intervals for possible (re)measurements, based on their usefulness to improve the uncertainties of the transitions and/or the quantum-state energies. This is called the spectroscopic heat map method, which characterizes with two values all wavenumber intervals of a fixed length. The spectroscopic heat map method is tested on the example of seven water isotopologues and the most abundant isotopologue of acetylene, based on data present in the HITRAN information system.

1. Introduction

The line-by-line datasets in the spectroscopic databank HITRAN [1,2] are widely recognized as the *de facto* international standards in high-resolution molecular spectroscopy. The importance of the HITRAN data can be appreciated when considering the large number of ground-based, airborne, and space applications which heavily depend on them. HITRAN updates are published on a quadrennial basis, with smaller updates being made in intervening years. Important aims of these updates include the extension of the data and the improvement of the accuracy of the transitions already included in the dataset [2].

In this paper, we restrict our discussion to the question what extent newly acquired spectroscopic data would help to improve the information about existing line positions. We are considering the following two improvements that might happen after replacing existing lines: (1) improving, that is, lowering uncertainties with which a transition is known, and (2) improving the uncertainty of the energies of quantum states, calculated from the complete set of transitions in the line list, like done in the MARVEL algorithm [3,4].

New data about transition wavenumbers intended to replace old entries usually come from new measurements (for example, from carefully designed precision-spectroscopy measurements [5], which result in line positions of kHz accuracy). They seldom come from theoretical calculations, and especially not from first-principles ones. Measurements are always associated with restricted wavenumber intervals allowed by the experimental setup. Thus, to help experimentalists, it is important to find wavenumber intervals selected for remeasurement of transitions that are most useful to improve the existing line list.

There are various physical attributes that describe a spectroscopic measurement. The mathematical model introduced in this paper uses a subset of these attributes and provides a useful starting point for the design of experiments to improve line lists.

Let us discuss the consequences of obtaining new information about transitions existing in a database, whose (a) intensity is above a certain minimum value, (b) wavenumber is in a certain 100 cm^{-1}-wide interval, and (c) accuracy is known within certain limits. For example, there are 2282 rovibrational transitions of $H_2{}^{16}O$ between 6970 cm^{-1} and 7070 cm^{-1} in the HITRAN database, and from these 2282 transitions 460 transitions have an intensity value above 10^{-25} cm molecule^{-1}. The first question we are aiming to answer concerns what happens if we obtain line-position data with an uncertainty significantly better than the line list has at present? The second question addresses possible improvements about the uncertainty of energy values. For this, a brief discussion about the concept of spectroscopic networks (SNs) [6,7] is necessary. Briefly, spectroscopic networks are graph representations of a line list, whereby the vertices of the graph correspond to quantum states present in the line list, and the edges of the graph represent the transitions measured between the (assigned) quantum states. Various weight functions can be attached both to the vertices and the

* Corresponding author at: Budapest Business University, Budapest, Hungary.
E-mail address: Arendas.Peter@uni-bge.hu (P. Árendás).

https://doi.org/10.1016/j.jqsrt.2023.108878
Received 1 December 2023; Received in revised form 14 December 2023; Accepted 19 December 2023
Available online 30 December 2023

edges, based on what is needed in a particular situation. Spectroscopic networks have proved to be particularly useful to aid high-resolution molecular spectroscopy [8–14]. The theoretical background about SNs, necessary to appreciate the current study, is provided in Section 2.

The structure of the rest of this paper is as follows. Section 2 explains the methodology used, including a brief description of the concept of spectroscopic networks. Section 3 introduces a small, artificial line list, upon which the concepts of this paper are demonstrated. Section 4 discusses the spectroscopic heat maps of seven water isotopologues present in HITRAN, $H_2{}^{16}O$, $H_2{}^{17}O$, $H_2{}^{18}O$, $HD^{16}O$, $HD^{17}O$, $HD^{18}O$, and $D_2{}^{16}O$, and also one on the most abundant isotopologue of acetylene, $^{12}C_2H_2$. Section 5 describes an additional analysis for $H_2{}^{16}O$, using the first-principles POKAZATEL [15] line list. Section 6 discusses the conclusions of this work and contains further development ideas.

2. Methodology

2.1. General considerations

Let L be a line list of a molecule. Let us denote a transition of L by t, and the wavenumber, the uncertainty, and the intensity of t by $w(t)$, $u(t)$, and $i(t)$, respectively. Moreover, let $q(t) = 1$ if both the upper and lower quantum states of the transition are assigned, else, let $q(t) = 0$. Note that in the case of HITRAN the intensity value $i(t)$ of the transition t is obtained after dividing the 'S' parameter of the database by the isotopologue abundance.

Let u_a and i_a, where 'a' stands for 'assumed', denote an arbitrarily chosen uncertainty and intensity value, in units of cm^{-1} and $cm\,molecule^{-1}$, respectively (these are the usual units used in high-resolution spectroscopy and in the rest of the paper the units will usually not be given explicitly). These two values represent the assumption that a transition can be found in the line list L with an uncertainty of u_a and with an intensity of at least i_a.

Next, let l denote a wavenumber interval, for example, $l = 100$ cm^{-1}. For a line list L, this wavenumber interval is slided, in 1 cm^{-1} increments, from 0 cm^{-1} until it reaches the largest wavenumber value in L. Let s denote the left endpoint of the interval. In each step, the consequences of improving transition uncertainties to u_a, for transitions t for which $w(t) \in [s, s+l]$ and $i(t) \geq i_a$, is investigated. To avoid an inconvenient overloading of the notation, in what follows the parameters L, w, u_a, and i_a are considered to be fixed.

2.2. Quantifying transition uncertainty improvements

For a fixed L, w, u_a, and i_a, let us begin by defining the transition set

$$T_1(s) = \{t \in L : w(t) \in [s, s+l]\}, \tag{1}$$

i.e., $T_1(s)$ contains all transitions that have a wavenumber value in the $[s, s+l]$ interval.

Next, let us filter out those transitions which cannot be detected because of their low intensity values:

$$T_2(s) = \{t \in T_1(s) : i(t) \geq i_a\} = \\ = \{t : t \in [s, s+l], i(t) \geq i_a\}. \tag{2}$$

Thus, $T_2(s)$ contains only those transitions from the given interval which also have the required intensity. Now, we are ready to select the transitions whose uncertainty can be improved. These are the elements of the set

$$T_3(s) = \{t \in T_2(s) : u(t) > u_a\} = \\ = \{t : t \in [s, s+l], i(t) \geq i_a, u(t) > u_a\}. \tag{3}$$

As a convenient endpoint in this chain, let us quantify the transition improvements in the $[s, s+l]$ wavenumber interval by the function

$$f(s) = |T_3(s)|. \tag{4}$$

Thus, $f(s) \in \mathbb{N}$ is equal to the number of transitions of L, in the interval $[s, s+l]$, which have (a) an intensity value of at least the assumed intensity i_a, and (b) an uncertainty value larger than the assumed uncertainty u_a.

2.3. Spectroscopic networks

To quantify energy uncertainty improvements, we rely on the spectroscopic-network representation of the line list. The *spectroscopic network* of a molecule is a graph $G(V, E)$, in which the vertex set V represents the rovibronic quantum states of the molecule, and the edge set E corresponds to allowed transitions between the quantum states. Physical quantities can be added as weight functions: most notably, quantum-state energy as vertex weights, and transition intensity and wavenumber as edge weights.

Therefore, we represent the line list L by the graph $G(V, E)$, where $E = \{t \in L : q(t) = 1\}$. The quantum state of the molecule that is defined to have zero energy value is the *root* of the spectroscopic network. For a graph $G(V, E)$ a *path* $P \subseteq E$ of length $k-1$ is an edge set $\{e_1, e_2, \ldots, e_{k-1}\} \subseteq E$, for which there exists a vertex set $\{v_1, \ldots, v_k\} \subseteq V$ such that for $1 \leq i \leq k$ the endpoints of e_i are v_i and v_{i+1}.

Lastly, let us call a quantum state X *ε-reachable* if there exists a path from the root to X that contains transitions whose uncertainty is at most ε. For example, if a quantum state is 10^{-5}-reachable, then there exists a chain of transitions, a path, from the root to the quantum state such that the uncertainty of each transition in the chain is at most 10^{-5}.

2.4. Quantifying energy uncertainty improvements

Let us define the transition set

$$T' = \{t \in L : q(t) = 1, u(t) \leq u_a\}. \tag{5}$$

The set T' contains the assigned transitions which have an uncertainty value already smaller than u_a. Let the set V' denote the u_a-reachable quantum states of the spectroscopic network representing T'.

We are interested in vertices which are not in V', but become u_a-reachable after improving the uncertainty of the assigned transitions of a $T_3(s)$ edge set to u_a, for some s. Let

$$T_4(s) = \{t \in T_3(s) : q(t) = 1\}. \tag{6}$$

Let V_s denote the u_a-reachable quantum states of the spectroscopic network representing $T' \cup T_4(s)$. Note that $|V_s| \geq |V'|$.

Then, as a sister-value of $f(s)$ (see Eq. (4)), let us define

$$g(s) = |V_s| - |V'|. \tag{7}$$

The $g(s)$ value expresses the number of quantum states whose reachability could be improved to u_a after obtaining u_a-accurate data of the assigned transitions in the wavenumber interval $[s, s+l]$.

Using the definitions given above, we define the *spectroscopic heat map* of a molecule as the pair of its $f(s)$ and $g(s)$ values, which can be calculated under various chosen parameter values u_a and i_a.

2.5. Remarks

The value of $f(s)$ shows how many transition uncertainties could be improved to u_a in the $[s, s+l]$ interval by new measurements. The value of $g(s)$ expresses that after the improvement of the transitions in the above sense, in the resulting line list how many energy levels become reachable from the root using transitions of at most u_a uncertainty.

Note that there are additional obstacles in obtaining accurate information about rovibronic transitions; for example, the density of lines in the chosen wavenumber range and disturbances caused by impurities in the sample, including isotopologues. The spectroscopic-heat-map method presented in this paper should be supplemented with a secondary investigation, which can also be based on the exact

Table 1
An artificial line list, used during the discussion of the methological concepts of this paper.

$w(t)$	$u(t)$	$i(t)$	Upper	Lower	Label
0.5	10^{-7}	10^{-28}	A	Root	#1
1.5	10^{-3}	10^{-26}	B	Root	#2
2.5	10^{-6}	10^{-22}	C	A	#3
3.5	10^{-4}	10^{-24}	?	?	#4
4.5	10^{-3}	10^{-23}	C	B	#5
5.5	10^{-7}	10^{-26}	E	D	#6
6.5	10^{-3}	10^{-22}	D	C	#7

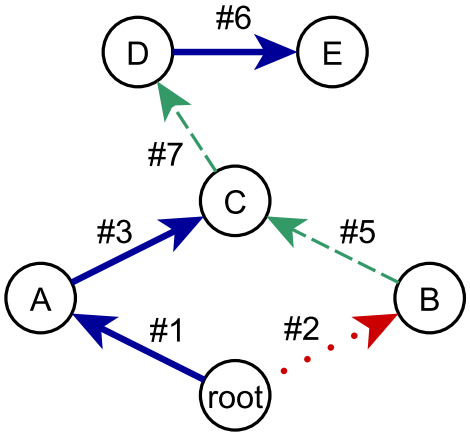

Fig. 1. The spectroscopic network representation of the line list of Table 1; line #4 is not shown because its quantum states are not assigned.

specifications of the experimental apparatus to be used to remeasure the line(s). Refining the spectroscopic-heat-map method to improve its capability to suggest wavenumber intervals for measurement seems like a promising future research idea.

3. A small artificial example

Let us demonstrate the methodology suggested above through the small, artificial line list of Table 1. Let us set the assumed uncertainty value to $u_a = 10^{-5}$, and the assumed intensity value to $i_a = 10^{-25}$. The spectroscopic network representation of this artificial line list is shown in Fig. 1.

First, let us discuss the transition uncertainty improvements. We can improve lines where $u(t)$ is larger than $u_a = 10^{-5}$ and $i(t)$ is at least $i_a = 10^{-25}$. The uncertainty condition only removes lines #1, #3, and #6 of Fig. 1. These transitions, represented by continuous blue edges in Fig. 1, have already better uncertainties than the assumed uncertainty. From the remaining lines, line #2 is removed, since we cannot improve this line due to its small intensity value. This line is represented by a dotted red edge in Fig. 1.

Therefore, the transitions that are subject to uncertainty improvements are #4, #5, and #7. One of these lines, line #4, has no assignment; thus, it cannot be represented in the spectroscopic network. The other two lines are represented by dashed green edges in Fig. 1.

Next, let us concentrate on the uncertainty improvements of the energies of the quantum states. Fig. 1 suggests that there are two quantum states currently reachable from the root *via* edges with at least u_a uncertainty: nodes A and C. In other words, the set of the 10^{-5}-reachable vertices is $\{A, C\}$.

Quantum-state energy uncertainty improvements refer to the extension of this original set by new vertices. Observe in Fig. 1 that the improvement of the uncertainty of line #5 to 10^{-5} implies that B becomes 10^{-5}-reachable, and a similar improvement of line #7 leads to vertices D and E becoming 10^{-5}-reachable.

The results related to the line list of Table 1 are collected in Table 2. Each row of Table 2 corresponds to a wavenumber interval, and the metrics defined in Section 2 are calculated for each interval. Among these metrics are the $f(s)$ and $g(s)$ values, together forming a spectroscopic heat map for this small, artificial example. The results show, for example, that the $[4, 7]$ wavenumber interval yields the highest $g(s)$ value, 3. The maximum $g(s)$ value highlights this wavenumber interval as a suitable target for a future experiment if the goal is to improve quantum-state energy uncertainties derived from the line list.

4. Spectroscopic heat maps of water isotopologues and $^{12}C_2H_2$

This section contains results obtained when the spectroscopic-heat-map method is used on line lists of seven water isotopologues and the main isotopologue of acetylene. All the experimental data are taken from the HITRAN 2020 [1] databank.

Before discussing the isotopologues one after the other, it is worth pointing out that the complete set of results on the seven water isotopologues are found in the Supplementary Material to this paper. Readers are encouraged to utilize these data sets, based on their own interest and measurement capabilities. The data are organized as follows. Each of the seven water isotopologues have five corresponding text files, one for each of the following u_a values: 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, and 10^{-7}. The first line of each text file explicitly identifies both the isotopologue and the u_a parameter of the data. The column structure of the individual text files is shown in Table 3. An additional file containing a detailed example is also included in the Supplementary Material.

In this section, along a short description of the obtained results, each isotopologue is characterized by a corresponding figure. Figs. 2–8 visualize the opportunities which arise to improve transition or quantum-state uncertainties by improving the current HITRAN line uncertainties of water isotopologues in the corresponding line lists. The layout of all these four-panel figures is as follows. The two panels on the left are about transition uncertainty improvements, *i.e.*, they show $f(s)$ values. The two panels on the right highlight opportunities for quantum-state energy uncertainty improvements, *i.e.*, $g(s)$ values. The panels in the first row correspond to the assumed intensity of $i_a = 10^{-24}$, while the second-row panels to $i_a = 10^{-25}$. In each panel, the three assumed uncertainty values are represented by different colors: blue, red, and orange curves correspond to $u_a = 10^{-3}$, 10^{-4}, and 10^{-5}, respectively.

4.1. $H_2{}^{16}O$

Fig. 2 illustrates selected results derived for the main water isotopologue, $H_2{}^{16}O$. The first observation one can make is that the blue lines in each panel, corresponding to $u_a = 10^{-3}$, show $f(s) = 0$ values. This means that all the assigned transitions in HITRAN whose intensity is larger than 10^{-24} cm molecule^{-1} are known with at least

Table 2
Showing the calculations on the line list of Table 1, with the interval size set to $l = 3$ cm^{-1}.

s	$[s, s+l]$	$\lvert T_1\rvert$	T_1	$\lvert T_2\rvert$	T_2	$f(s)$	T_3	$g(s)$	$V_s \setminus V'$
0	[0, 3]	3	#1, #2, #3	1	#3	0	∅	0	∅
1	[1, 4]	3	#2, #3, #4	1	#3, #4	1	#4	0	∅
2	[2, 5]	3	#3, #4, #5	2	#3, #4, #5	2	#4, #5	1	B
3	[3, 6]	3	#4, #5, #6	1	#4, #5	2	#4, #5	1	B
4	[4, 7]	3	#5, #6, #7	2	#5, #7	2	#5, #7	3	B, D, E

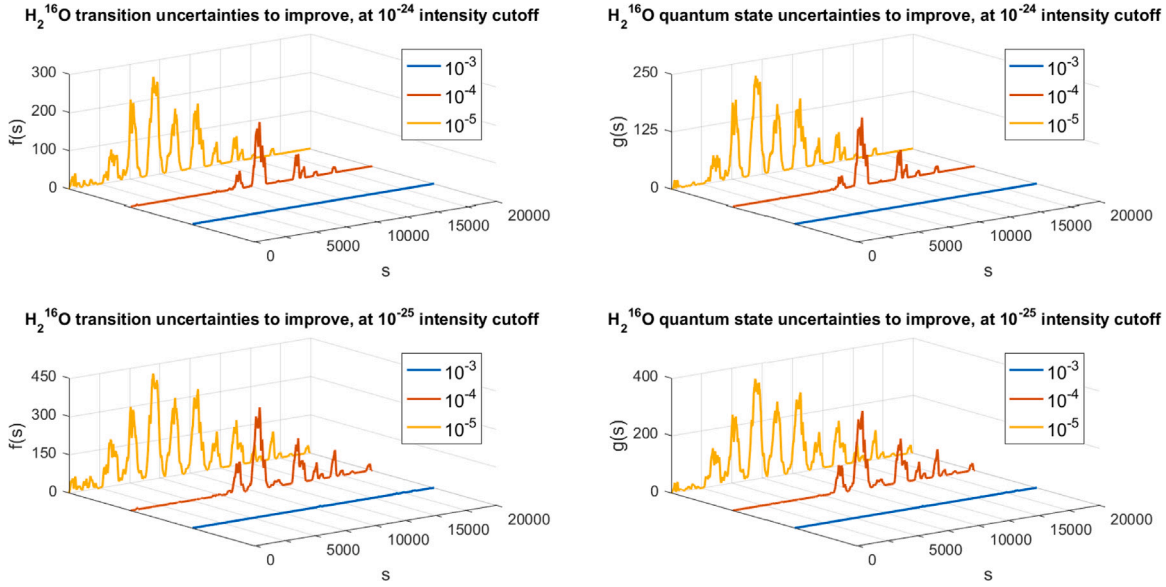

Fig. 2. The $f(s)$ and $g(s)$ values for the $H_2{}^{16}O$ molecule, at assumed intensities of 10^{-24} (top panels) and 10^{-25} cm^{-1} (bottom panels), and assumed uncertainties of 10^{-3}, 10^{-4}, and 10^{-5} (blue, red, and orange curves, respectively), for $s \in [0, 19\,900]$.

Table 3
The format of the individual text files in the Supplementary Material of this paper.

Column no.	Content
1	s
2	$T_1(s)$
3	$T_2(s)$ if $i_a = 10^{-24}$
4	$f(s)$ if $i_a = 10^{-24}$
5	$g(s)$ if $i_a = 10^{-24}$
6	$T_2(s)$ if $i_a = 10^{-25}$
7	$f(s)$ if $i_a = 10^{-25}$
8	$g(s)$ if $i_a = 10^{-25}$
9	$T_2(s)$ if $i_a = 10^{-26}$
10	$f(s)$ if $i_a = 10^{-26}$
11	$g(s)$ if $i_a = 10^{-26}$
12	$T_2(s)$ if $i_a = 10^{-27}$
13	$f(s)$ if $i_a = 10^{-27}$
14	$g(s)$ if $i_a = 10^{-27}$
15	$T_2(s)$ if $i_a = 10^{-28}$
16	$f(s)$ if $i_a = 10^{-28}$
17	$g(s)$ if $i_a = 10^{-28}$

10^{-3} cm^{-1} accuracy. This is an excellent state of affairs for the main water isotopologue.

Assuming $u_a = 10^{-4}$, one starts finding intervals in which both transition uncertainties and quantum state uncertainties could be improved. The best opportunity is to conduct measurements around 10 650 cm^{-1}, which is the approximate wavenumber position of the largest peak of the red curves in all four panels of Fig. 2.

The most obvious, but not too surprising feature of the left panels of Fig. 2 is that most transitions can be improved with new measurements with an accuracy of 10^{-5} or better. Unfortunately, these measurements are not trivial, though with modern experimental techniques involving saturation spectroscopy and frequency combs such accuracy is certainly achievable for a significant number of transitions [5,16–19].

Another feature of the $H_2{}^{16}O$ data is that the shape of the corresponding $f(s)$ and $g(s)$ curves is very similar. Thus, there is a strong correlation between improving transition uncertainties and improving quantum-state energy uncertainties. Therefore, conducting a measurement in a useful wavenumber interval helps to achieve both goals.

In addition to Fig. 2, results about $H_2{}^{16}O$ are also shown in Table 4, providing a different viewpoint on the same data. Table 4 contains the maxima of the $f(s)$ values for $H_2{}^{16}O$ at various chosen intensity and uncertainty values. For example, the top suggestion regarding transition improvements, under the assumed intensity of 10^{-25} and the assumed uncertainty of 10^{-5}, is the interval from 6970 cm^{-1} to 7070 cm^{-1}: a total of 414 transition uncertainties could be improved within this interval. There are readily available lasers which would facilitate such measurements [5].

4.2. $H_2{}^{17}O$

Fig. 3 displays selected spectroscopic-heat-map results obtained for the isotopologue $H_2{}^{17}O$. One can observe immediately that in this case there are opportunities to improve transition uncertainties already at the $u_a = 10^{-3}$ level. This situation is unlike that of $H_2{}^{16}O$, with no such opportunities. The $f(s)$ values are roughly the same for the case of $u_a = 10^{-4}$. Assuming 10^{-4} measurement uncertainty, a huge $f(s)$ peak appears around 6000 – 7000 cm^{-1}.

Another important observation is that in the interval considered the energies of the quantum states of $H_2{}^{17}O$ are known quite accurately, offering practically no opportunities for measurements with 10^{-3} and 10^{-4} accuracy. However, if one considered measurements with $u_a = 10^{-5}$, the 6000 – 7000 cm^{-1} wavenumber interval, mentioned in the previous paragraph, has huge $g(s)$ values.

4.3. $H_2{}^{18}O$

Fig. 4 shows selected spectroscopic-heat-map results on the second most abundant water isotopologue, $H_2{}^{18}O$. Regarding the transition improvement opportunities of $H_2{}^{18}O$, a vital factor is the assumed intensity value. Comparing the two panels of the left column of Fig. 4, there are large differences at all assumed uncertainty values between detecting transitions above intensity magnitudes 10^{-24} and 10^{-25}. By measuring wavenumber intervals around 10 000 cm^{-1} or 15 000 cm^{-1}, and detecting transitions above 10^{-25} intensity, we can already make transition uncertainty improvements with 10^{-3} accuracy.

The energies of the quantum states of $H_2{}^{18}O$ are well known with an uncertainty of 10^{-3} and above 10^{-24} in intensity. Opportunities to improve quantum-state energy uncertainties arise after changing the assumed intensity to 10^{-25} and the assumed uncertainty to 10^{-4}.

4.4. $HD^{16}O$

Fig. 5 displays selected spectroscopic-heat-map results for $HD^{16}O$. Regarding both the improvement of transition uncertainties and

Table 4
Top suggestions for the $H_2{}^{16}O$ isotopologue. Column i_a contains the assumed intensities, u_a is the assumed uncertainty, s is the starting wavenumber of the interval, and $f(s)$ is the number of transitions which can be improved.

i_a	$u_a = 10^{-3}$		$u_a = 10^{-4}$		$u_a = 10^{-5}$		$u_a = 10^{-6}$		$u_a = 10^{-7}$	
	s	$f(s)$	s	$f(s)$	s	$f(s)$	s	$f(s)$	s	$f(s)$
10^{-24}	N/A		10 665	164	6 998	256	3 674	340	3 674	340
10^{-25}	10 561	3	10 677	320	6 970	414	3 674	482	3 674	482
10^{-26}	22 556	30	10 677	551	6 958	612	6 958	613	6 958	613
10^{-27}	16 906	130	10 599	835	10 599	942	10 599	942	10 599	942
10^{-28}	16 904	325	10 603	1141	10 603	1303	10 603	1310	10 603	1310

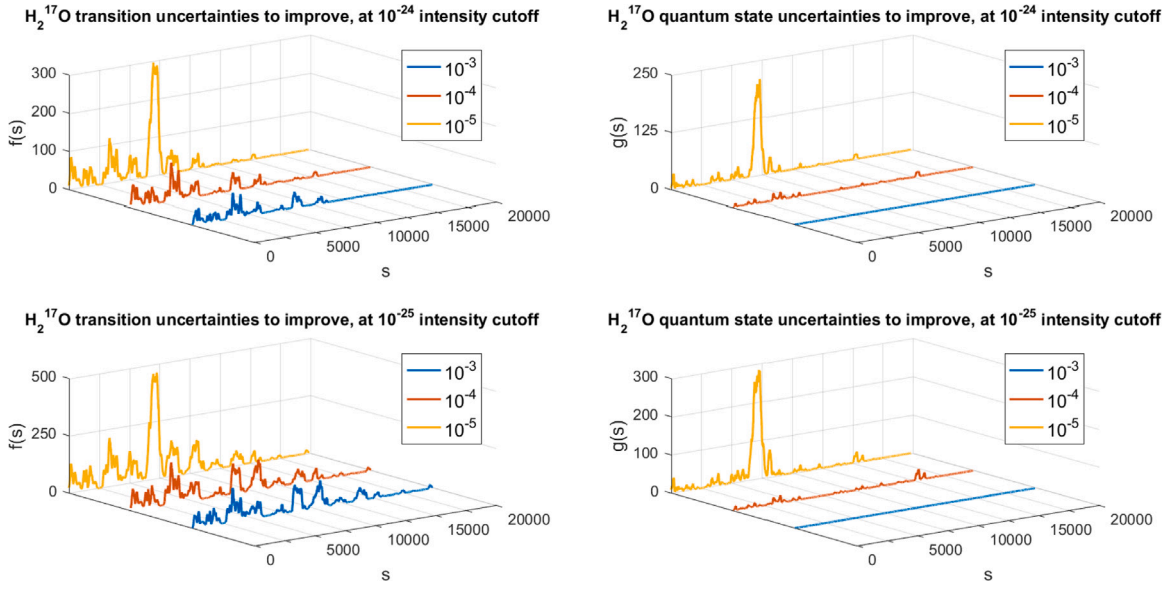

Fig. 3. The $f(s)$ and $g(s)$ values for the $H_2{}^{17}O$ molecule, at chosen intensities of 10^{-24} and 10^{-25}, and assumed uncertainties of 10^{-3}, 10^{-4}, and 10^{-5}, for the wavenumber range of $s \in [0, 19\,900]$.

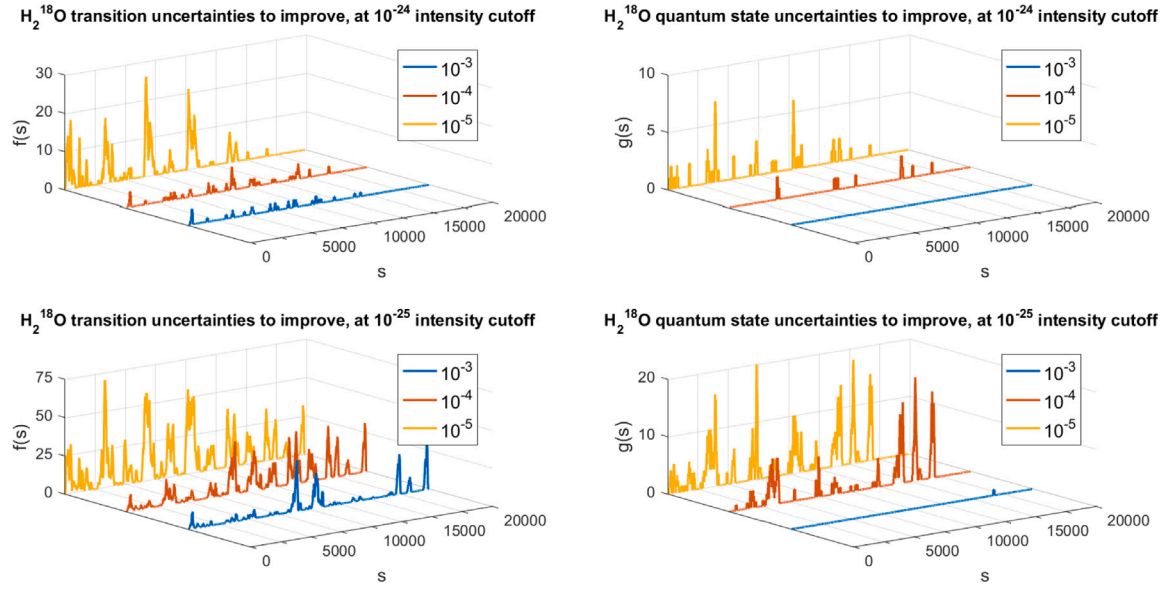

Fig. 4. The $f(s)$ and $g(s)$ values for the $H_2{}^{18}O$ molecule, at assumed intensities of 10^{-24} and 10^{-25}, and assumed uncertainties of 10^{-3}, 10^{-4}, and 10^{-5}, for $s \in [0, 19\,900]$.

quantum-state energy uncertainties, $HD^{16}O$ only offers real opportunities for remeasurements with $u_a = 10^{-5}$ accuracy. At this accuracy, however, there is a wide selection of useful intervals between 0 and 10 000 cm^{-1} for remeasurements, as indicated by the large $f(s)$ and $g(s)$ values.

Above 10 000 cm^{-1}, however, both the $f(s)$ and the $g(s)$ values are very small if not zero. Therefore, conducting measurements above 10 000 cm^{-1} do not seem to be useful to improve the already measured transitions or the energies of the quantum states of $HD^{16}O$.

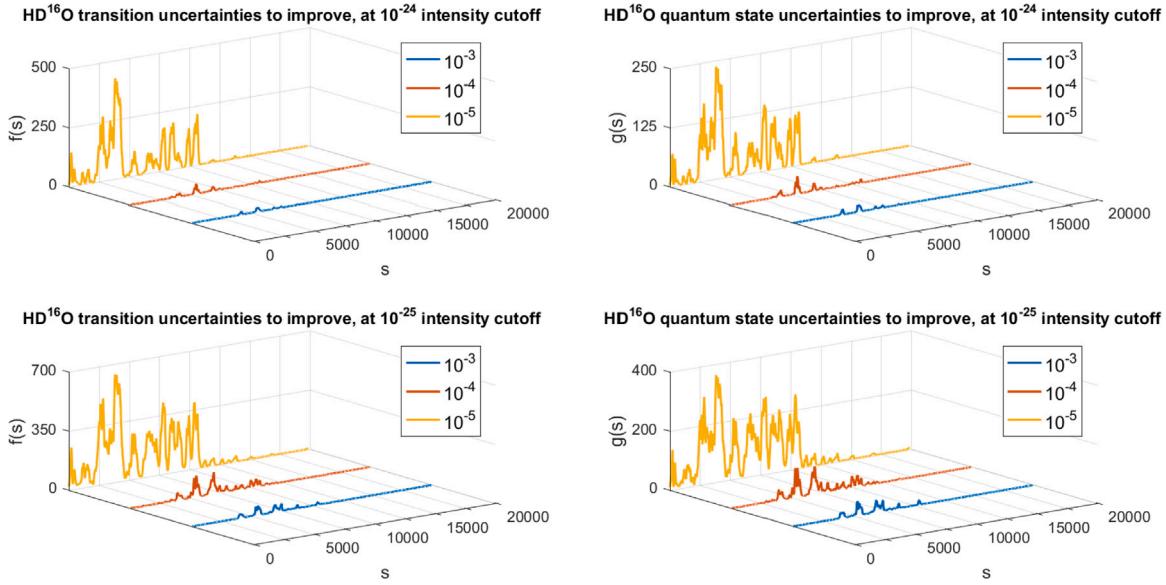

Fig. 5. The $f(s)$ and $g(s)$ values for the $HD^{16}O$ molecule, at assumed intensities of 10^{-24} and 10^{-25}, and assumed uncertainties of 10^{-3}, 10^{-4}, and 10^{-5}, for $s \in [0, 19\,900]$.

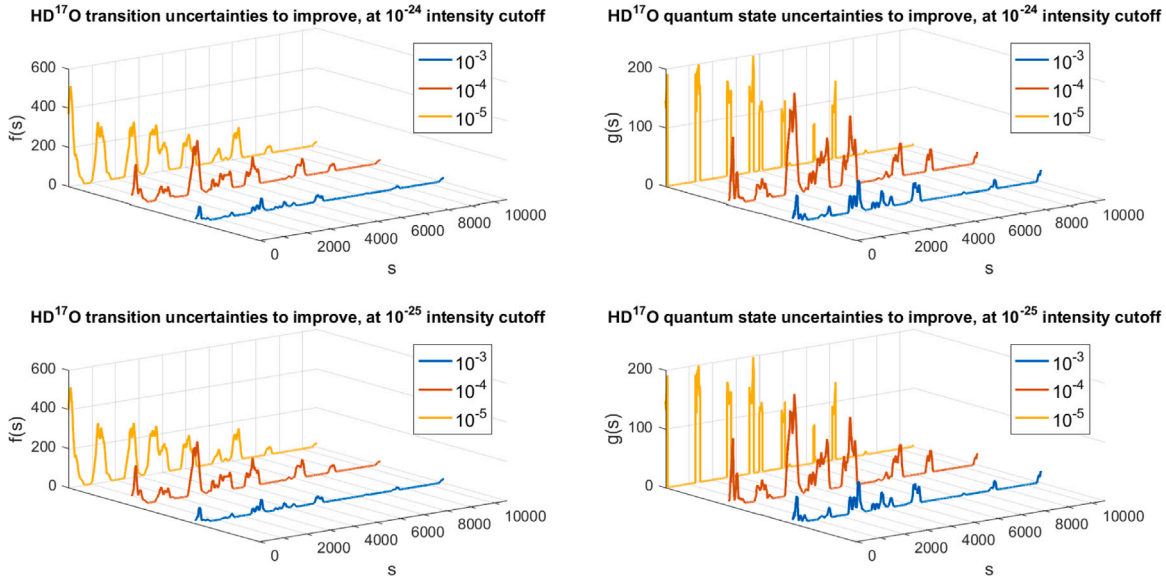

Fig. 6. The $f(s)$ and $g(s)$ values for the $HD^{17}O$ molecule, at assumed intensities of 10^{-24} and 10^{-25}, and assumed uncertainties of 10^{-3}, 10^{-4}, and 10^{-5}.

4.5. $HD^{17}O$

Fig. 6 contains selected spectroscopic-heat-map results on the rare water isotopologue $HD^{17}O$. The most spectacular result of the analysis is that, in all cases, both the $f(s)$ and the $g(s)$ functions show "oscillating" behavior between large and very small values. This is especially true in the case of $u_a = 10^{-5}$. Consequently, one should be very precise when selecting wavenumber intervals to measure, as a small difference could mean improving zero quantum state uncertainties instead of improving around 200 of them.

This oscillatory behavior of both functions stop after around 8000 cm^{-1}. For wavenumber intervals above 8000 cm^{-1}, there are only a few and globally not so useful measurement opportunities, as shown in Fig. 6.

4.6. $HD^{18}O$

Selected spectroscopic-heat-map results obtained for $HD^{18}O$ are shown in Fig. 7. Apart from the wavenumber interval between 0 and 1000 cm^{-1}, measurements with 10^{-3} accuracy offer practically no improvement opportunities. Under 10^{-4} accuracy, the $f(s)$ and $g(s)$ values in a few additional wavenumber intervals arise also moderately, but in general, the possible contribution of a 10^{-4}-uncertainty measurement is still low.

However, measuring with 10^{-5} accuracy offer great opportunities to improve both transition uncertainties and quantum state energy uncertainties in the case of $HD^{18}O$. However, one should be very precise in selecting the wavenumber interval, as the $f(s)$ and $g(s)$ values show oscillatory behavior, alternating between large and very small (or zero) values.

4.7. $D_2{}^{16}O$

Fig. 8 illustrates selected spectroscopic-heat-map results derived for $D_2{}^{16}O$. The transitions and quantum state energies of $D_2{}^{16}O$ are well known at the 10^{-3} uncertainty level, offering no real improvement opportunities for measurements with this accuracy. For 10^{-4} accuracy,

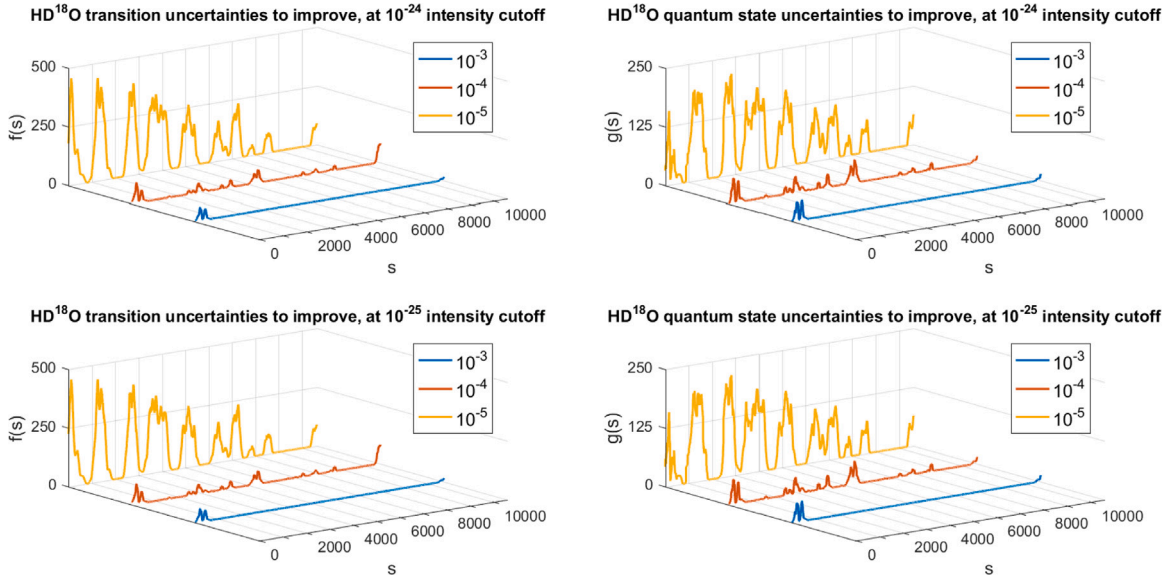

Fig. 7. The $f(s)$ and $g(s)$ values for the $HD^{18}O$ molecule, at assumed intensities of 10^{-24} and 10^{-25}, and assumed uncertainties of 10^{-3}, 10^{-4}, and 10^{-5}.

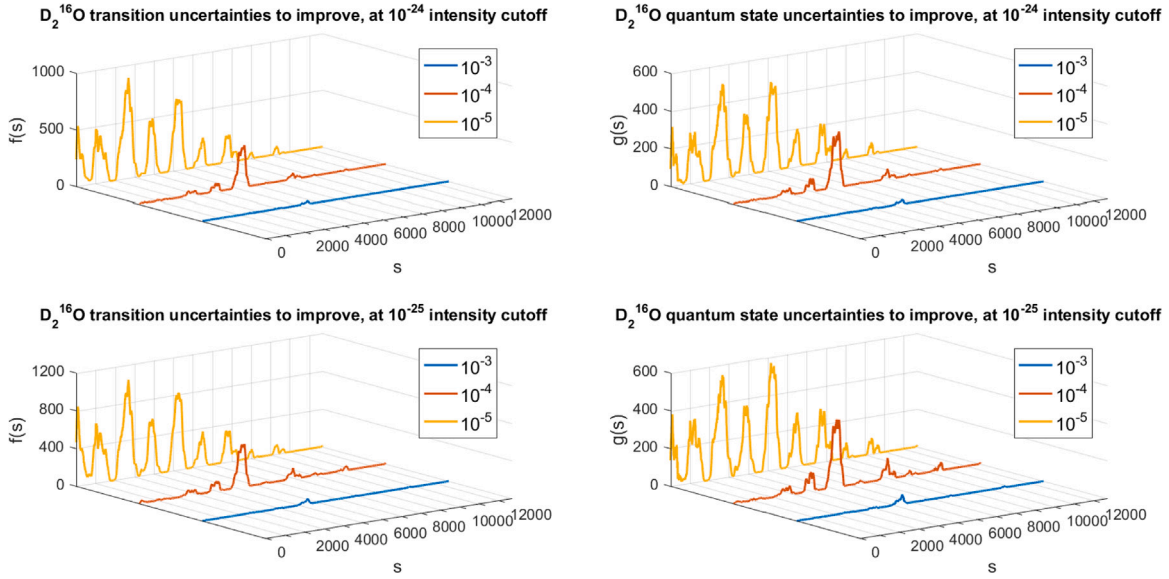

Fig. 8. The $f(s)$ and $g(s)$ values for the $D_2{}^{16}O$ molecule, at assumed intensities of 10^{-24} and 10^{-25}, and assumed uncertainties of 10^{-3}, 10^{-4}, and 10^{-5}.

there is one relatively large peak in both the $f(s)$ and $g(s)$ curves on all panels of Fig. 8, centered around 5200 cm^{-1}.

New measurements between 0 and 9000 cm^{-1} start to be useful after reaching the 10^{-5} accuracy level. Notice the oscillatory behavior of both functions in the corresponding (yellow) curves. A consequence of this oscillation is that one should be precise when selecting the wavenumber interval for measurement, as a small change could mean the difference between a very useful measurement from the viewpoint of this analysis, and a not very useful one.

4.8. $^{12}C_2H_2$

At last, let us discuss another molecular example, acetylene, to underline the usefulness of the spectroscopic-heat-map approach.

Fig. 9 shows the $g(s)$ values obtained for the $^{12}C_2H_2$ HITRAN line list, at $u_a = 10^{-7}$ and $i_a = 10^{-25}$. In 2023, new, kHz-accuracy measurements were performed in the spectral range of 7125–7230 cm^{-1} for the $^{12}C_2H_2$ molecule [19], following the ideas of the so-called SNAPS

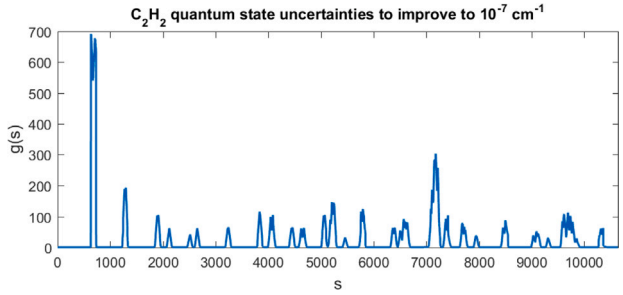

Fig. 9. The $g(s)$ values for the $^{12}C_2H_2$ molecule, at the assumed intensity of 10^{-25} and the assumed uncertainty of 10^{-7}.

(spectroscopic network assisted precision spectroscopy) approach [5]. In this work the authors could improve the uncertainties of 57 energy

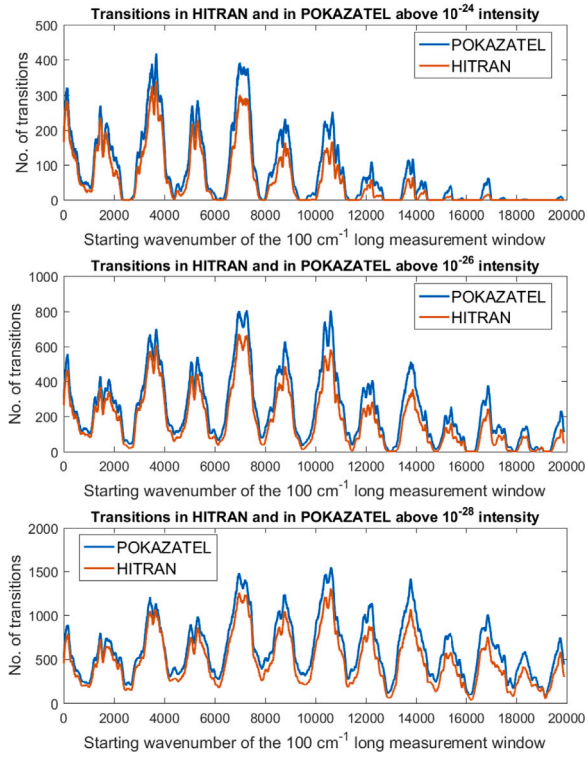

Fig. 10. Comparison of the number of $H_2{}^{16}O$ transitions in the HITRAN and POKAZATEL line lists between 0 and 20 000 cm^{-1}.

levels using 20 newly measured transitions (*i.e.*, in our terminology, $g(s) \gg f(s)$). If we check the shape of the $g(s)$ curve for the $^{12}C_2H_2$ HITRAN line list, see Fig. 9, we can clearly see that there is a high peak in the 7000–7500 cm^{-1} region. This means that this region is a very good choice to improve the uncertainties of the energy levels. Furthermore, it can be seen that further 200 energy levels can be improved by measurements in this region.

5. Completeness analysis of $H_2{}^{16}O$

To provide additional insight about the accuracy of the transitions and energy levels of $H_2{}^{16}O$, this study also contains an analysis that compares the incomplete HITRAN line list with the complete first-principles POKAZATEL [15] line list of $H_2{}^{16}O$. Using the definitions from Section 2.2, this comparison was done by calculating the $|T_2(s)|$ values between 0 and 19 900 cm^{-1} for both line lists, with assumed intensity values of 10^{-24}, 10^{-26}, and 10^{-28}. The results obtained are shown in Fig. 10.

Besides offering a viewpoint on the completeness of the HITRAN line list, this comparison also sheds light on several other issues. A welcome observation is that, overall, the transitions are well mapped in HITRAN until 20 000 cm^{-1}: the two curves are close to each other in all three panels.

The number of possible transitions grow in a slow pace as the assumed intensity value increases, with a growth factor of only 2 on the vertical axis from 10^{-24} to 10^{-26} and from 10^{-26} to 10^{-28}.

The atmospheric windows (zero transitions in an interval) fade as the intensity value decreases, completely disappearing at 10^{-28}. The presence of the less transition-dense intervals seen at 10^{-28} might also offer a good opportunity of determining the assignment for these transitions, as lower line density could make the identification easier. Such a comparison is also possible for the other water isotopologues.

6. Summary and conclusions

This paper presents a novel method for the analysis of large line-by-line datasets, for example those in the canonical HITRAN [1] spectroscopic databank. This analysis, which we call the 'spectroscopic-heat-map approach', can help the design of new spectroscopic measurements by highlighting wavenumber intervals that seem particularly promising for (re)measurement if the goal is to improve the accuracy of transitions and energy levels of the line list of the given molecule.

There are two parameters which govern the design of new measurements, namely the minimum intensity (i_a), above which the lines are assumed to be detectable, and the assumed wavenumber uncertainty (u_a) which the measured lines will have. The main idea is to observe what happens after improving the uncertainties of certain transitions, that meet a few criteria, in a 100 cm^{-1} wavenumber interval, and the interval window is slided from 0 cm^{-1} to the maximum wavenumber in the line list. The method outputs text files, with multiple numbers describing each 100 cm^{-1} interval; these data sets are called spectroscopic heat maps, well suited for visualization.

The algorithm was applied to line lists of seven water isotopologues and to acetylene of the HITRAN spectroscopic databank. As part of the Supplementary Material, the readers are provided with comprehensive spectroscopic heat maps for the seven water isotopologues at various assumed uncertainty and intensity parameters. The hope is that they could help in the design of actual future spectroscopic measurements.

This work offers various ideas for future research. The most straightforward idea is to investigate the possible improvements of transition and energy uncertainties for other species. One could also develop a more detailed method of quantifying the uncertainty improvements: for example, by including ab initio data, or by extending the method with other considerations, for example, line density. Another idea is to develop a method which explicitly suggests measurements, using the considerations of this paper, but adding an objective function to classify the usefulness of the possible measurements.

CRediT authorship contribution statement

Péter Árendás: Writing – review & editing, Writing – original draft, Visualization, Methodology, Data curation, Conceptualization. **Tibor Furtenbacher:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Data curation, Conceptualization. **Attila G. Császár:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Péter Árendás reports financial support was provided by National Research Development and Innovation Office. Attila G. Császár reports financial support was provided by National Research Development and Innovation Office. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The complete set of results on the seven water isotopologues are found in the Supplementary Material to this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jqsrt.2023.108878.

References

[1] Gordon IE, Rothman LS, Hargreaves RJ, Hashemi R, Karlovets EV, Skinner FM, Conway EK, Hill C, Kochanov RV, Tan Y, Wcisło P, Finenko AA, Nelson K, Bernath PF, Birk M, Boudon V, Campargue A, Chance KV, Coustenis A, Drouin BJ, Flaud J-M, Gamache RR, Hodges JT, Jacquemart D, Mlawer EJ, Nikitin AV, Perevalov VI, Rotger M, Tennyson J, Toon GC, Tran H, Tyuterev VG, Adkins EM, Baker A, Barbe A, Canè E, Császár AG, Dudaryonok A, Egorov O, Fleisher AJ, Fleurbaey H, Foltynowicz A, Furtenbacher T, Harrison JJ, Hartmann J-M, Horneman V-M, Huang X, Karman T, Karns J, Kassi S, Kleiner I, Kofman V, Kwabia-Tchana F, Lavrentieva NN, Lee TJ, Long DA, Lukashevskaya AA, Lyulin OM, Makhnev VY, Matt W, Massie ST, Melosso M, Mikhailenko SN, Mondelain D, Müller HSP, Naumenko O, Perrin A, Polyansky OL, Raddaoui E, Raston PL, Reed ZD, Rey M, Richard C, Tóbiás R, Sadiek I, Schwenke DW, Starikova E, Sung K, Tamassia F, Tashkun SA, Vander Auwera J, Vasilenko IA, Vigasin AA, Villanueva GL, Vispoel B, Wagner G, Yachmenev A, Yurchenko SN. The HITRAN2020 molecular spectroscopic database. J Quant Spectr Radiat Transf 2022;277:107949. http://dx.doi.org/10.1016/j.jqsrt.2021.107949.

[2] Rothman L. History of the HITRAN database. Nat Rev Phys 2021;3:302–4. http://dx.doi.org/10.1038/s42254-021-00309-2.

[3] Furtenbacher T, Császár AG, Tennyson J. MARVEL: measured active rotational-vibrational energy levels. J Mol Spectrosc 2007;245:115–25. http://dx.doi.org/10.1016/j.jms.2007.07.005.

[4] Furtenbacher T, Császár AG. MARVEL: Measured Active Rotational-Vibrational Energy Levels. II. Algorithmic improvements. J Quant Spectrosc Radiat Transf 2012;113:929–35. http://dx.doi.org/10.1016/j.jqsrt.2012.01.005.

[5] Tóbiás R, Furtenbacher T, Simkó I, Császár AG, Diouf ML, Cozijn FMJ, Staa JMA, Salumbides EJ, Ubachs W. Spectroscopic-network-assisted precision spectroscopy and its application to water. Nature Commun 2020;11:1708.

[6] Császár AG, Furtenbacher T. Spectroscopic networks. J Mol Spectrosc 2011;266:99–103. http://dx.doi.org/10.1016/j.jms.2011.03.031.

[7] Furtenbacher T, Császár AG. The role of intensities in determining characteristics of spectroscopic networks. J Mol Spectrosc 2012;1009:123–9. http://dx.doi.org/10.1016/j.molstruc.2011.10.057.

[8] Furtenbacher T, Árendás P, Mellau G, Császár AG. Simple molecules as complex systems. Sci Rep 2014;4:4654. http://dx.doi.org/10.1038/srep04654.

[9] Császár AG, Furtenbacher T, Árendás P. Small molecules – big data. J Phys Chem A 2016;120:8949–69. http://dx.doi.org/10.1021/acs.jpca.6b02293.

[10] Árendás P, Furtenbacher T, Császár AG. On spectra of spectra. J Math Chem 2016;54:806–22. http://dx.doi.org/10.1007/s10910-016-0591-1.

[11] Tóbiás R, Furtenbacher T, Császár AG. Cycle bases to the rescue. J Quant Spectrosc Radiat Transf 2017;203:557–64. http://dx.doi.org/10.1016/j.jqsrt.2017.03.031.

[12] Árendás P, Furtenbacher T, Császár AG. From Bridges to Cycles in Spectroscopic Networks. Sci Rep 2020;10:19489. http://dx.doi.org/10.1038/s41598-020-75087-5.

[13] Árendás P, Furtenbacher T, Császár AG. Selecting lines for spectroscopic (re)measurements to improve the accuracy of absolute energies of rovibronic quantum states. J Cheminform 2021;13:67. http://dx.doi.org/10.1186/s13321-021-00534-y.

[14] Tóbiás R, Bérczi K, Szabó C, Császár AG. autoECART: Automatic energy conservation analysis of rovibronic transitions. J Quant Spectrosc Radiat Transf 2021;272:107756. http://dx.doi.org/10.1016/j.jqsrt.2021.107756.

[15] Polyansky OL, Kyuberis AA, Zobov NF, Tennyson J, Yurchenko SN, Lodi L. ExoMol molecular line lists XXX: A complete high-accuracy line list for water. Mon Not R Astron Soc 2018;480:2597–608.

[16] Diouf ML, Tóbiás R, Simkó I, Cozijn FMJ, Salumbides EJ, Ubachs W, Császár AG. Network-based design of near-infrared lamb-dip experiments and the determination of pure rotational energies of $H_2{}^{18}O$ at kHz accuracy. J Phys Chem Ref Data 2021;50:023106. http://dx.doi.org/10.1063/5.0052744.

[17] Diouf ML, Tóbiás R, van der Schaaf TS, Cozijn FMJ, Salumbides EJ, Császár AG, Ubachs W. Ultraprecise relative energies in the (2 0 0) vibrational band of $H_2{}^{16}O$. Mol Phys 2022;120:e2050430. http://dx.doi.org/10.1080/00268976.2022.2050430.

[18] Diouf ML, Tóbiás R, Cozijn FMJ, Salumbides EJ, Fábri C, Puzzarini C, Császár AG, Ubachs W. Parity-pair-mixing effects in nonlinear spectroscopy of HDO. Opt Express 2022;30:46040. http://dx.doi.org/10.1364/oe.474525.

[19] Castrillo A, Fasci E, Furtenbacher T, D'Agostino V, Khan MA, Gravina S, Gianfrani L, Császár AG. On the $^{12}C_2H_2$ near-infrared spectrum: absolute transition frequencies and an improved spectroscopic network at the kHz accuracy level. Phys Chem Chem Phys 2023;25:23614–25. http://dx.doi.org/10.1039/D3CP01835K.