PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2019, **21**, 3473

Accurate empirical rovibrational energies and transitions of $H_2{}^{16}O$

Roland Tóbiás, [a] Tibor Furtenbacher, [a] Jonathan Tennyson [b] and Attila G. Császár *[c]

Several significant improvements are proposed to the computational molecular spectroscopy protocol MARVEL (Measured Active Rotational–Vibrational Energy Levels) facilitating the inversion of a large set of measured rovibrational transitions to energy levels. The most important algorithmic changes include the use of groups of transitions, blocked by their estimated experimental (source segment) uncertainties, an inversion and weighted least-squares refinement procedure based on sequential addition of blocks of decreasing accuracy, the introduction of spectroscopic cycles into the refinement process, automated recalibration, synchronization of the combination difference relations to reduce residual uncertainties in the resulting dataset of empirical (MARVEL) energy levels, and improved classification of the lines and energy levels based on their accuracy and dependability. The resulting protocol, through handling a large number of measurements of similar accuracy, retains, or even improves upon, the best reported uncertainties of the spectroscopic transitions employed. To show its advantages, the extended MARVEL protocol is applied for the analysis of the complete set of highly accurate $H_2{}^{16}O$ transition measurements. As a result, almost 300 highly accurate energy levels of $H_2{}^{16}O$ are reported in the energy range of 0–6000 cm^{-1}. Out of the 15 vibrational bands involved in accurately measured rovibrational transitions, the following three have definitely highly accurate empirical rovibrational energies of 8–10 digits of accuracy: (v_1 v_2 v_3) = (0 0 0), (0 1 0), and (0 2 0), where v_1, v_2, and v_3 stand for the symmetric stretch, bend, and antisymmetric stretch vibrational quantum numbers. The dataset of experimental rovibrational transitions and empirical rovibrational energy levels assembled during this study, both with improved uncertainties, is considerably larger and more accurate than the best previous datasets.

Received 14th August 2018,
Accepted 21st November 2018

DOI: 10.1039/c8cp05169k

rsc.li/pccp

1 Introduction

In the era of optical frequency combs and frequency-comb spectroscopy[1,2] it has become increasingly realistic to determine experimental line positions with an accuracy on the order of 10 kHz or even better. Previously, such experimental accuracy was reserved to pure rotational transitions measured in the microwave (MW) region of the electromagnetic spectrum but nowadays there is growing evidence that this accuracy can be extended all the way from the MW to the near infrared (NIR) region[3–5] and beyond.[6,7]

Complementing the experimental advances in high-resolution molecular spectroscopy in the theoretical front is the introduction of spectroscopic networks,[8–10] and the Measured Active Rotational–Vibrational Energy Levels (MARVEL) scheme.[8,9,11–13] The MARVEL algorithm and code, originally presented in ref. 11 and 12, allows the efficient determination of accurate empirical energy levels with well-defined uncertainties from a set of measured and assigned rovibronic transitions of known uncertainty, the global analysis of measured spectra,[14] and the validation of related spectroscopic datasets and databases.[13] MARVEL has been employed for a number of molecular systems, including $^{12}C_2$,[15] nine isotopologues of water,[16–20] three isotopologues of SO_2,[21] three isotopologues of H_3^+,[22,23] $^{14}NH_3$,[24] and parent ketene.[25] A web-based version of the MARVEL code has allowed the involvement of high-school students[26] in spectroscopic research projects, leading to published studies on $^{48}Ti^{16}O$,[27] $^{90}Zr^{16}O$,[28] $H_2{}^{32}S$,[29] and $^{12}C_2H_2$,[30] molecules of considerable astronomical interest.

Energy levels derived using the MARVEL procedure are increasingly being used to improve not only transition wavenumbers[31] but also partition functions[32] present in standard information systems. Just considering water, MARVEL energy levels have been used extensively by the ExoMol project[33] to improve predicted line positions in line lists for $H_2{}^nO$ (n = 16, 17, 18).[34,35] Similarly, the most recent edition of the HITRAN

[a] *Laboratory of Molecular Structure and Dynamics, Institute of Chemistry, ELTE Eötvös Loránd University and MTA-ELTE Complex Chemical Systems Research Group, P.O. Box 32, H-1518 Budapest 112, Hungary*
[b] *Department of Physics and Astronomy, University College London, London WC1E 6BT, UK*
[c] *MTA-ELTE Complex Chemical Systems Research Group, P.O. Box 32, H-1518 Budapest 112, Hungary. E-mail: csaszarag@caesar.elte.hu*

database, HITRAN2016,[31] makes use of MARVEL energy levels for $H_2{}^nO$ (n = 16, 17, 18),[36,37] as well as for the deuterated isotopologues.[38] MARVEL energy levels formed a key part of a procedure used to greatly improve predicted energy levels for water isotopologues.[34] Empirical (MARVEL) rovibrational energies have also been used to generate the most accurate partition functions available for $H_2{}^{16}O$,[39] heavy water,[40] and its three constituent isotopologues.[40] As they are present in a large number of spectra, water lines can also be used for calibration in different spectral regions.[41,42]

The second-generation MARVEL code[12] runs extremely fast, treats transition data and the related matrices (*vide infra*) on the order of 100 000 in less than a minute on a single CPU, and it is platform independent. This efficiency of the MARVEL code facilitates enhancements, such as those discussed below, of the MARVEL algorithm.

In all the MARVEL studies cited above accurate empirical rovibronic energy levels have been determined from the simultaneous treatment of all measured transitions. Nevertheless, it has repeatedly been observed that when transitions based on the empirical (MARVEL) energy levels and their uncertainties were compared to the best measurements, slight distortions and occasionally unnecessarily large, but sometimes too small, uncertainties characterize the MARVEL energy levels. These inconsistencies, at least in part, are the result of the inclusion of many transitions of orders of magnitude lower accuracy than the best measurements in the MARVEL inversion and refinement procedure. In this study we attempt to devise a MARVEL-based protocol which retains the accuracy of the best measurements while still working with as complete spectroscopic networks as feasible during the inversion and refinement process.

Next, let us review the state-of-the-art of high-resolution spectroscopy for the $H_2{}^{16}O$ isotopologue of water, the subject molecule of our feasibility study in which the extended MARVEL (extMARVEL) protocol is employed to gain highly accurate rovibrational energy levels meeting or even exceeding the accuracy of the best measurements.

The highest-quality database of accurate rovibrational energy levels and transitions of water vapor is the spectroscopic information system maintained at the Jet Propulsion Laboratory (JPL).[43,44] Nowadays the $H_2{}^{16}O$ dataset of JPL energy levels is based on a study of Lanquetin *et al.*[45] In particular, the experimental JPL energy levels of $H_2{}^{16}O$ are exactly those reported in ref. 45. The JPL transitions are the results of an effective Hamiltonian fit of a considerable number of transitions, as detailed on the JPL website.[44] The most comprehensive evaluation of measured water transitions was performed by an IUPAC Task Group.[16–20] Ref. 20 gives a summary of this work. Part III of this series[18] contains a validated and recommended set of measured $H_2{}^{16}O$ transitions (about 200 000) and empirical energy levels (about 20 000), based on experimental data available prior to 2013. These sets were used to update the water data in HITRAN2016.[31] Since 2013, there have been many new studies of water spectra and, in particular, the use of optical-frequency-comb-based measurements to determine very accurate wavenumbers for selected transitions in the infrared, see, *e.g.*, ref. 5, 46 and 47. The advent and availability of new, precise experimental techniques act as an impetus to further improve the MARVEL treatment of measured transitions allowing the determination of highly accurate empirical energy levels.

As of today no *ortho–para* rovibrational transitions have been observed in water vapor.[67] This has the consequence that water spectra define two principal components[10] within the measured spectroscopic network of $H_2{}^{16}O$, corresponding to *ortho*- and *para*-$H_2{}^{16}O$. Note that the number $q = v_3 + K_a + K_c$ is even for *para* and odd for *ortho* rovibrational states, where v_3 is the vibrational quantum number corresponding to the antisymmetric stretch motion, while K_a and K_c are the usual rigid-rotor quantum numbers of an asymmetric-top molecule.

As emphasized in ref. 42, despite the fact that *ortho* lines may have three times higher intensity than *para* lines (and thus about two times more of them have been determined experimentally[18]), the best frequency standards correspond to *para*-$H_2{}^{16}O$, as there is no hyperfine splitting of the *para*-$H_2{}^{16}O$ lines and the minimum-energy level of the *para* PC can properly be set to zero with zero uncertainty.

High-resolution radial-velocity-shift transition measurements used to detect molecular species in the atmospheres of exoplanets[68,69] has greatly increased the need for accurate laboratory transition frequencies. While this technique has been used successfully to identify water in exoplanets[70,71] with high confidence, standard water line lists used for exoplanet modelling appear to be not sufficiently accurate for this task.[72] It is also true that many of the most important line positions have been measured sufficiently accurately. In the extended list[73] of the astrophysically most important spectral lines, the International Astronomical Union listed 13 $H_2{}^{16}O$ lines, with assignments and rest frequencies recalled in Table 1. For all these frequencies several independent, highly accurate experimental determinations are available (see Table 1). Nevertheless, not all of these transitions are part of accurate cycles satisfying the law of energy conservation[13] to the accuracy of the best measurements, which calls for optical-frequency-comb-based remeasurement of certain rovibrational transitions on the ground vibrational state of water.

The present study has been executed with two principal goals in mind. First, we wanted to improve the MARVEL protocol,[11,12] with particular emphasis on a better reproduction of the most accurately measured rovibrational transitions. Second, to test the utility of the improved protocol, we chose $H_2{}^{16}O$, the admittedly most important polyatomic molecule for high-resolution spectroscopy, for which the accurate knowledge of the rovibrational energy levels and transitions is important in a number of scientific and engineering applications. Therefore, first a detailed description of the extMARVEL scheme is given in Section 2. To aid the reader of the methodological section, some of the more technical details are moved to appendices. Those interested only in the spectroscopic results of this study, and not the way they were determined, can skip Section 2. Then, Section 3 discusses the $H_2{}^{16}O$ experimental input data used, while Section 4 presents results of our extMARVEL analysis of a large number of old and new sources.[46–66,74–103] This analysis

Table 1 The 13 astrophysically most important water ($H_2{}^{16}O$) lines recommended by the International Astronomical Union and their different experimental determinations. The experimental (Expt.) frequencies of multiple measurements follow the order of their increased uncertainties. In the assignment column, the standard spectroscopic notation $J'_{K_a',K_c'} \leftarrow J''_{K_a'',K_c''}$ is used, where $J'_{K_a',K_c'}$ and $J''_{K_a'',K_c''}$ are the rotational labels for the upper and lower energy levels, respectively. All transitions, with AF = approximate frequency, belong to the ground vibrational state

AF/GHz	Component	Assignment	extMARVEL/kHz	01LaCoCa[45]/kHz	Expt./kHz
22.235	*para*	$6_{16} \leftarrow 5_{23}$	22 235 079.85(6)	22 235 007(4240)	22 235 079.85(6)[48]
					22 235 080(29)[49]
					22 235 200(600)[50]
					22 235 220(154)[51]
					22 235 000(190)[52]
183.310	*ortho*	$3_{13} \leftarrow 2_{20}$	183 310 090.4(1)	183 309 897(2544)	183 310 090.6(1)[53]
					183 310 087(4)[54]
					183 310 117(29)[55]
					183 310 150(65)[49]
					183 310 200(190)[52]
					183 311 300(1330)[56]
325.153	*para*	$5_{15} \leftarrow 4_{22}$	325 152 899(2)	325 153 101(3392)	325 152 899(2)[54]
					325 152 888(13)[57]
					325 152 919(22)[55]
					325 153 700(882)[52]
380.197	*para*	$4_{14} \leftarrow 3_{21}$	380 197 359.8(6)	380 197 395(3610)	380 197 359.8(6)[58]
					380 197 356(4)[54]
					380 197 365(13)[57]
					380 197 372(16)[55]
					380 196 800(621)[52]
439.151	*para*	$6_{43} \leftarrow 5_{50}$	439 150 794.8(6)	439 151 582(5528)	439 150 794.8(6)[58]
					439 150 795(2)[54]
					439 150 812(19)[55]
448.001	*para*	$4_{23} \leftarrow 3_{30}$	448 001 077.5(6)	448 001 155(3610)	448 001 077.5(6)[58]
					448 001 075(16)[55]
					448 000 300(904)[52]
474.689	*ortho*	$5_{33} \leftarrow 4_{40}$	474 689 108(2)	474 689 879(3816)	474 689 108(2)[54]
					474 689 127(21)[55]
556.936	*para*	$1_{10} \leftarrow 1_{01}$	556 935 987.6(6)	556 935 841(1824)	556 935 987.7(6)[58]
					556 935 985(3)[59]
					556 935 995(8)[54]
					556 936 002(16)[55]
					556 935 819(185)[60]
					556 935 800(190)[52]
					556 935 800(206)[61]
620.700	*para*	$5_{32} \leftarrow 4_{41}$	620 700 954.9(6)	620 701 697(4664)	620 700 954.9(6)[58]
					620 700 950(35)[62]
					620 700 844.1(150)[63]
					620 700 807(163)[55]
752.033	*ortho*	$2_{11} \leftarrow 2_{02}$	752 033 113(15)	752 032 978(2544)	752 033 113(15)[64]
					752 033 104(19)[60]
					752 033 227(125)[55]
					752 033 300(190)[52]
916.172	*ortho*	$4_{22} \leftarrow 3_{31}$	916 171 581(21)	916 171 448(3392)	916 171 580(21)[65]
					916 171 582(21)[66]
					916 171 405(194)[60]
970.315	*ortho*	$5_{24} \leftarrow 4_{31}$	970 315 021(21)	970 315 165(3392)	970 315 022(21)[66]
					970 315 020(21)[65]
					970 314 968(58)[60]
987.927	*ortho*	$2_{02} \leftarrow 1_{11}$	987 926 762(21)	987 926 473(2341)	987 926 764(21)[66]
					987 926 760(21)[65]
					987 926 743(23)[60]

also includes comparisons with previously determined accurate water levels and lines.[45,60,95] The paper ends with some concluding remarks in Section 5.

2 Methodology

2.1 Spectroscopic networks

Sets of rovibrational transitions, whether they are measured or computed, can be treated as building blocks of spectroscopic networks (SN, see ref. 8–13), whereby (a) the vertices are energy levels, (b) the edges correspond to transitions, oriented from the lower energy level to the upper one (independently whether the transition was recorded in absorption, emission, or by means of an action spectroscopy), and (c) the (positive) edge weights are the wavenumbers of the transitions. Note that many other weighting schemes (*e.g.*, weighting by line intensities) can be utilized in practical applications of SNs.

If SNs are formed by experimental transitions, called experimental SNs, the edge weights should be associated with

measurement uncertainties. However, in many data sources, instead of providing line-by-line uncertainties, only an average "expected" accuracy is provided, usually corresponding to an intense unblended line, *i.e.*, to a best-case measurement scenario. Based on these experimental line positions and approximate uncertainties—together with assignments to the lower and upper energy levels of the transition—one is able to derive empirical (MARVEL) energy levels with uncertainties *via* the MARVEL (Measured Active Rotational–Vibrational Energy Levels) procedure,[11,12] utilizing a weighted least-squares technique (*e.g.*, robust reweighting[104]) and the Rydberg–Ritz combination principle.[105]

The detailed analysis of experimental SNs is very important as one would like to treat all the measured rovibrational lines of a molecule simultaneously and make the validated lines and their energy levels available to spectroscopists and spectroscopic database developers. For this purpose, it is necessary to explore (a) the components (sets of energy levels not connected by any measured transition), (b) the bridges (transitions whose deletion increases the number of components), and (c) the cycles (collections of connected edges within which every vertex has two neighboring vertices) characterizing the given experimental SN.

Since the energy levels of the different components are not connected, during the MARVEL analysis we are forced to set the lowest-lying energy level (core) of each component to zero. If the core of a component corresponds to the lowest-energy level of a particular nuclear-spin isomer of the molecule examined, this component is a principal component (PC), otherwise it is called a floating component (FC). Clearly, one is most interested in the detailed characterization of those energy levels and transitions which are part of PCs.

Due to the fact that bridges of experimental SNs may compromise the accuracy of the empirical energy levels determined during a MARVEL analysis, they require special attention. That is, if a bridge is determined incorrectly or inaccurately, the energy levels connected to the core of their component through this bridge will be shifted or scattered. Energy levels "behind" a bridge cannot be known more accurately than the bridge itself, a considerable hindrance in the derivation of highly accurate empirical energy levels.

Cycles are extremely useful when compatibility of the transitions and their associated uncertainties obtained in different groups under often widely different experimental conditions are examined.[106] It is important to check, for all the cycles of the SN, whether the law of energy conservation (LEC, see ref. 13) is satisfied within the experimental accuracy. Accordingly, in the cases where (a) a transition is assigned improperly, (b) its wavenumber is measured inaccurately, or (c) the uncertainty of this wavenumber is underestimated, the discrepancy (absolute signed sum of the line positions) becomes higher than the threshold (the sum of the uncertainties) in the given cycle, indicating a conflict among the related transitions.

All the cycles of an experimental SN can be expressed in a cycle basis (CB) with the symmetric differences of the basic cycles in the CB, which contains all non-bridge lines of this SN.[13] Accordingly, construction of such CBs and evaluation of their entries could be sufficient to test the compatibility of the measured rovibrational lines in light of the LEC. For details about the use of CBs to explore inconsistencies in SNs, see ref. 13. Note also that in contrast to ref. 13, cycles of length 2, which correspond to repeated measurements, are also permitted here, as this extension makes the present formalism simpler than the previous one.[13]

Subnetworks are of great importance during the treatment of SNs. These derived graph structures are (a) represented with a participation matrix, $\mathbf{P} = \mathrm{diag}(P_1, P_2, \ldots, P_{N_\mathrm{T}})$, where N_T is the number of lines in the SN, and $P_i = 1$ if the ith transition of the SN is inserted in the subnetwork, otherwise $P_i = 0$, and (b) filled with all the energy levels of the SN, among which there may also be isolated nodes (vertices without inserted lines).

Due to the presence of some outliers, which should be excluded from the database of transitions during our analyses (see also Section 2.3.2), it is necessary to introduce the leading subnetwork of the SN, denoted with $\mathcal{N}_\mathrm{LS}$, which contains all the non-excluded lines of the database. In the present description, **P** will always denote the participation matrix assigned to the $\mathcal{N}_\mathrm{LS}$ subnetwork.

2.2 The MARVEL procedure

To understand the improvements proposed in this study better, first the traditional MARVEL approach,[11,12] upon which our novel MARVEL algorithm is built, is recalled.

During a MARVEL analysis of measured rovibronic transitions the following objective function is minimized:

$$\Omega(\mathbf{E}) = (\boldsymbol{\sigma} - \mathbf{RE})^\mathrm{T}\mathbf{PW}(\boldsymbol{\sigma} - \mathbf{RE}), \quad (1)$$

where (a) $\mathbf{E} = \{E_1, E_2, \ldots, E_{N_\mathrm{L}}\}^\mathrm{T}$ is the vector of energies of the N_L energy levels, (b) $\boldsymbol{\sigma} = \{\sigma_1, \sigma_2, \ldots, \sigma_{N_\mathrm{T}}\}^\mathrm{T}$ is the vector of wavenumbers of the N_T transitions, (c) **R** is the Ritz matrix (see also eqn (3) of ref. 11), (d) $\mathbf{W} = \mathrm{diag}(w_1, w_2, \ldots, w_{N_\mathrm{T}})$ is a diagonal weight matrix with $w_i = \delta_i^\eta$, (e) δ_i is the uncertainty of σ_i, and (f) η is a nonpositive exponent, set to -2 in our calculations.

Since the minimum of $\Omega(\mathbf{E})$, denoted by $\bar{\mathbf{E}} = \{\bar{E}_1, \bar{E}_2, \ldots, \bar{E}_{N_\mathrm{L}}\}^\mathrm{T}$, is not unique,

$$\bar{E}_{\mathrm{core}(1)} = \bar{E}_{\mathrm{core}(2)} = \cdots = \bar{E}_{\mathrm{core}(N_\mathrm{c})} = 0 \quad (2)$$

is set for the cores indexed as core(1), core(2),..., core(N_c), where N_c is the number of components in $\mathcal{N}_\mathrm{LS}$. Furthermore, $\bar{\mathbf{E}}$ must also satisfy the relation

$$\mathbf{G}\bar{\mathbf{E}} = \mathbf{F}, \quad (3)$$

where $\mathbf{G} = \mathbf{R}^\mathrm{T}\mathbf{PWR}$ is the weighted Gram–Schmidt matrix of **R** and $\mathbf{F} = \mathbf{R}^\mathrm{T}\mathbf{PW}\boldsymbol{\sigma}$ is the vector of free terms. Note that eqn (2) and (3) can be solved with Cholesky decomposition, using, for example, the standard Eigen package,[107] and can be reduced by transforming **G** into a block diagonal form, which consists of N_c independent diagonal blocks. Note that during the testing of the code we experimented with several decomposition procedures, but no significant differences were found.

To eliminate the linear dependencies of the block-diagonal form of **G**, one has to leave out a (row, column) pair from each block. Although, in principle, these pairs could be chosen arbitrarily, in practice it is best to select those (row, column) combinations which are associated with the largest diagonal entries of the corresponding blocks; thus, improves the numerical stability of the solution.

In practice, the core indices of $\mathcal{N}_{\mathrm{LS}}$ are not necessarily known to the user. Then, from any solution, $\mathbf{E}'=\left\{E'_1, E'_2, \ldots, E'_{N_{\mathrm{L}}}\right\}^{\mathrm{T}}$, of eqn (3), the core($i$) index can be obtained for all $1 \leq i \leq N_{\mathrm{c}}$ as follows:

$$\mathrm{core}(i) = \underset{\substack{j=1 \\ \mathrm{comp}(j)=i}}{\overset{N_{\mathrm{L}}}{\mathrm{argmin}}} E'_j, \tag{4}$$

where comp(j) is the component index of the jth energy level, *i.e.*, the index of the component containing this energy level. (In eqn (4), the operation 'argmin' returns the j index of the smallest E_j' value for which comp(j) = i.) Based on the core indices, the $\bar{E}_j$ values can be expressed for all $1 \leq j \leq N_{\mathrm{L}}$ as

$$\bar{E}_j = E_j' - E'_{\mathrm{core}(j)}. \tag{5}$$

In what follows, we always assume that the core indices of $\mathcal{N}_{\mathrm{LS}}$ are available.

One must note in passing that measurement uncertainty is a combination of precision and absolute accuracy. Precision of measured line positions varies considerably according to a number of factors: (a) type and quality of the spectrometer, (b) the spectral resolution relative to the observed line width, (c) the signal to noise ratio, (d) the choice and especially the control of experimental conditions, (e) the complexity of the spectrum, and (f) the line retrieval methods employed. Since all of these factors must be considered as carefully as possible, it is not at all surprising that experimentalists measuring a large number of lines often refrain from reporting line-by-line uncertainties.

As soon as $\bar{\mathbf{E}}$ is determined, it is necessary to examine the $\Delta_i = \sigma_i - \bar{\sigma}_i$ (fitting) residuals and the $d_i = |\Delta_i| - \delta_i$ (fitting) defects, where $\bar{\sigma}_i = \bar{E}_{\mathrm{up}(i)} + \bar{E}_{\mathrm{low}(i)}$ is the ith wavenumber estimate, while up(i) and low(i) are the indices of the upper and lower energy levels of the ith transition, respectively. $\mathcal{N}_{\mathrm{LS}}$ is called consistent if the largest defect of $\mathcal{N}_{\mathrm{LS}}$, $d_{\max}$, is not positive. In the case of $d_{\max} > 0$, we have to either exclude certain lines or increase their uncertainties in order to ensure the consistency of $\mathcal{N}_{\mathrm{LS}}$.

2.3 The extended MARVEL (extMARVEL) approach

The approach outlined here is an extension of the MARVEL procedure,[11,12] which proved to be a powerful tool to obtain correct and reliable empirical rovibrational energy levels and associated uncertainties in a number of applications, as detailed in the Introduction. Our improved protocol is depicted in Fig. 1, the following subsections explain each of the schemes shown in Fig. 1.

2.3.1 Isolation of source segments (ISS). Having collated the transitions into a SN, good estimates for the uncertainties of the line positions are required for the extMARVEL analysis.

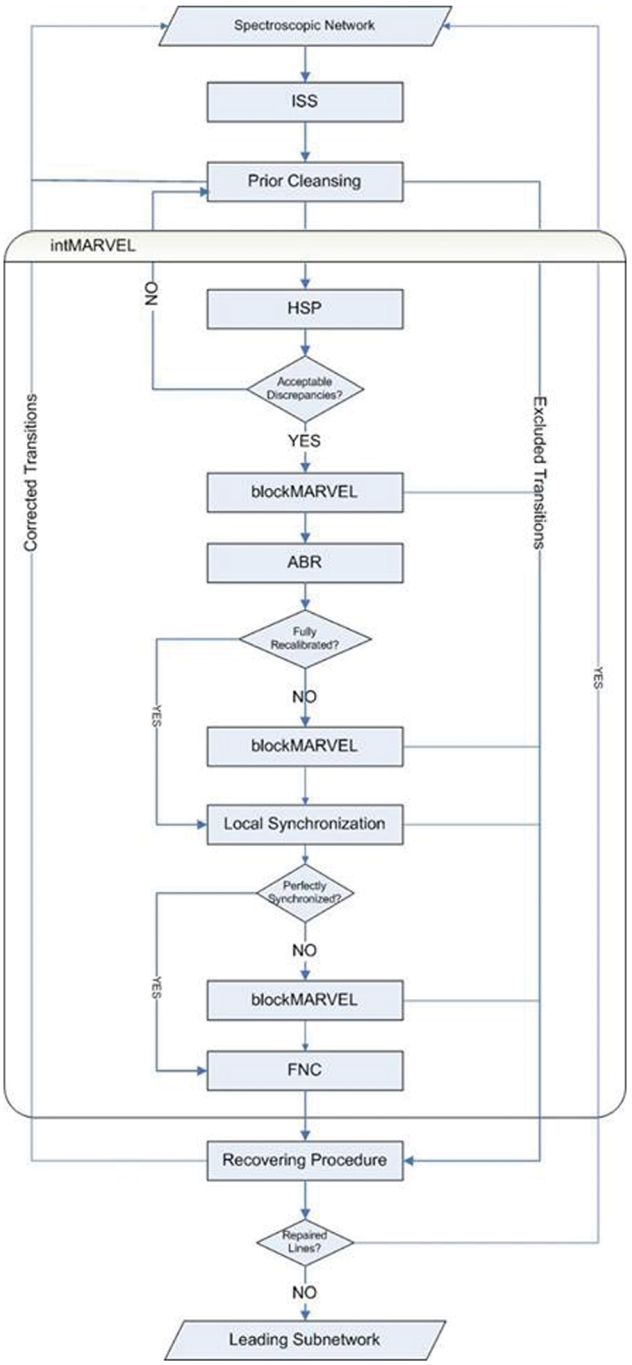


Fig. 1 The extended MARVEL (extMARVEL) approach, see text for an explanation of each step.

Fortunately, in practice it is often sufficient to find a suitable uncertainty for each subset of transitions of the same (data) source which are chosen to have identical uncertainties, at least as a first approximation.

Then, one needs to divide the transitions of the sources into segments, in which the uncertainties of the wavenumbers are considered to be (approximately) within a factor of 10. These approximate uncertainties are referenced as estimated segment uncertainties (ESU). A particular segment is denoted in this

study with a character sequence obtained by the concatenation of the source tag with one of the strings '_S1', '_S2', 'S3', *etc.* By convention, (a) segments are indexed in increasing order of their ESUs, and (b) '_S1' is never written out explicitly in the first segment.

It is important to establish the ESU values with caution to avoid the problem of under- and overutilization of the data. If explicit information on the accuracy is not available from the sources, one can use the simple approximate relations 'uncertainty = resolution/10' and 'uncertainty = resolution' for the unblended and blended lines, respectively. (It would help spectroscopic database developers if experimentalists published their lines partitioned at least into blended and unblended transitions and reported average uncertainty estimates for at least these two categories.)

At the end of this process, two input files need to be created for further analysis. The first one is the transition database in MARVEL format[11,12] with the distinction that the ''uncertainty'' column is not evaluated. The second one is the list of the source segments with their ESUs and a ternary flag for each segment. The value of this flag is (a) 0, if the given segment must not be subjected to recalibration using a single recalibration factor, (b) −1, if this segment should be recalibrated with the same factor as the segment in the previous record of the segment file, or (c) 1, otherwise.

2.3.2 Prior cleansing. Before the empirical energy levels are determined, it is mandatory to perform an extensive cleansing of the collated dataset to weed out clear outliers. As to inconsistencies concerning the dataset, both errors (misprints or transcription errors) and inaccuracies may occur (a) in the wavenumbers, (b) in the line assignments, and (c) in the ESU values of the segments. All these problems need to be identified. For this purpose, several checks must be carried out, including (a) a test whether the selection rules hold properly for the input transitions, (b) use of the ECART (Energy Conservation Analysis of Rovibronic Transitions, see ref. 13) algorithm to list and delete the incorrect cycles, or (c) an analysis of the fitting residuals derived from the MARVEL procedure.

2.3.3 Integrated MARVEL (intMARVEL) analysis. The integrated MARVEL (intMARVEL) algorithm (see Fig. 1) is a fully automated procedure able to (a) generate refined segment uncertainties (RSU) from the ESU values, (b) determine the energy levels of the SN based on their most accurate lines, (c) exclude outlier transitions from the SN, (d) calculate the individual wavenumber uncertainties of the lines originating from the RSU values, (f) recalibrate ill-calibrated segments, (e) synchronize the combination difference (CD) relations, and (g) assess the lines, segments, and energy levels on the basis of their accuracy and dependability.

This procedure was programmed into a code, written in the C++ language, called intMARVEL, which requires the two input files mentioned in Section 2.3.1. The output of the intMARVEL code contains the transitions, the energy levels, and the segments characterized, all in separate text files. A detailed description of the theoretical background of the intMARVEL algorithm is given in Section 2.4 and the corresponding appendices.

2.3.4 Recovering procedure. Once the intMARVEL analysis has been completed, the transitions excluded during certain stages of the extMARVEL algorithm need to be re-assessed and, if possible, corrected. In order to achieve data correction, the wavenumbers and assignments of the excluded lines are to be checked carefully based on the original sources.

After this revision, reassignments can be made for the uncorrected, but still excluded lines, using (a) the wavenumber-sorted experimental transition dataset, and (b) a MARVEL or an effective Hamiltonian (EH) linelist ignoring weak and forbidden transitions. At the end of the recovering step, the previous stages of the extMARVEL procedure should be repeated with the corrected lines until no transitions can be ''repaired''.

2.4 Background of the intMARVEL analysis

2.4.1 Hiearachical segment perturbation (HSP). To obtain the RSU values, the hierarchical segment perturbation (HSP) scheme is introduced. During HSP, for all the s segments, (a) a $\mathcal{N}_{\mathrm{per}}$ perturbation subnetwork is constructed by including all the non-excluded lines from s and those segments which are much more accurate than s (hierarchical perturbation), and (b) the RSU value is provided for s from the discrepancies and thresholds of those basic cycles in a given CB of $\mathcal{N}_{\mathrm{per}}$ which contain transitions from s (representative cycles). If there are too few representative cycles for the refinement of a particular ESU, then the corresponding RSU will be identical to this ESU value. The consecutive stages of the HSP procedure are described in Appendix A.

2.4.2 BlockMARVEL refinement. Since inaccurately measured wavenumber entries may deteriorate the accuracy of the estimated energy levels, a novel procedure, called the constrained MARVEL scheme, was designed with which these harmful effects can be avoided or at least substantially reduced. This technique ensures that the wavenumber estimates derived from a given $\mathcal{N}_{\mathrm{p}}$ (previous) subnetwork of $\mathcal{N}_{\mathrm{LS}}$, *via* the MARVEL algorithm, should remain unaffected when the MARVEL analysis is repeated for a $\mathcal{N}_{\mathrm{a}}$ (actual) subnetwork of $\mathcal{N}_{\mathrm{LS}}$, which also contains $\mathcal{N}_{\mathrm{p}}$. By definition, the cores of $\mathcal{N}_{\mathrm{a}}$ are the cores of $\mathcal{N}_{\mathrm{p}}$, as well (see also Section 2.1).

Designate the core indices in $\mathcal{N}_{\mathrm{p}}$ and $\mathcal{N}_{\mathrm{a}}$ with $\mathrm{core_p}(1)$, $\mathrm{core_p}(2)$,..., $\mathrm{core_p}(N_{\mathrm{p,c}})$ and $\mathrm{core_a}(1)$, $\mathrm{core_a}(2)$,..., $\mathrm{core_a}(N_{\mathrm{a,c}})$, respectively, where $N_{\mathrm{p,c}}$ and $N_{\mathrm{a,c}}$ are the number of components in $\mathcal{N}_{\mathrm{p}}$ and $\mathcal{N}_{\mathrm{a}}$, respectively. Let $\mathbf{P}^{\langle\mathcal{N}_{\mathrm{a}}\rangle}$ be the participation matrix of $\mathcal{N}_{\mathrm{a}}$. With the notation introduced, eqn (2) and (3) can be reformulated for $\mathcal{N}_{\mathrm{a}}$ as follows:

$$\bar{E}_{\mathrm{core_a}(1)} = \bar{E}_{\mathrm{core_a}(2)} = \cdots = \bar{E}_{\mathrm{core_a}(N_{\mathrm{a,c}})} = 0 \tag{6}$$

and

$$\mathbf{G}^{\langle\mathcal{N}_{\mathrm{a}}\rangle}\bar{\mathbf{E}} = \mathbf{F}^{\langle\mathcal{N}_{\mathrm{a}}\rangle}, \tag{7}$$

where $\mathbf{G}^{\langle\mathcal{N}_{\mathrm{a}}\rangle} = \mathbf{R}^{\mathrm{T}}\mathbf{P}^{\langle\mathcal{N}_{\mathrm{a}}\rangle}\mathbf{W}\mathbf{R}$ and $\mathbf{F}^{\langle\mathcal{N}_{\mathrm{a}}\rangle} = \mathbf{R}^{\mathrm{T}}\mathbf{P}^{\langle\mathcal{N}_{\mathrm{a}}\rangle}\mathbf{W}\boldsymbol{\sigma}$. Since the wavenumber estimates produced from $\mathcal{N}_{\mathrm{p}}$ with the MARVEL technique should remain unchanged, the following constraints must be enforced for all $1 \le i \le N_{\mathrm{L}}$:

$$\bar{E}_i - \bar{E}_{\mathrm{core_p}(\mathrm{comp_p}(i))} = \beta_i, \tag{8}$$

where $\mathrm{comp}_\mathrm{p}(i)$ is the component index of the ith energy level in $\mathcal{N}_\mathrm{p}$ and β_i is the bound for $\bar{E}_i$. (Bounds correspond to empirical energy levels derived from $\mathcal{N}_\mathrm{p}$ using the MARVEL analysis.) In other words, only the cores of $\mathcal{N}_\mathrm{p}$ are linearly independent variables, all the other energy levels can be expressed from eqn (8). Accordingly, these constraints can be written in a matrix-vector form as

$$\bar{\mathbf{E}} = \mathbf{C}\bar{\mathbf{H}} + \boldsymbol{\beta}, \tag{9}$$

where $\bar{\mathbf{H}} = \{\bar{H}_1, \bar{H}_2, \ldots, \bar{H}_{N_{\mathrm{p,c}}}\}^\mathrm{T}$ and $\mathbf{C} = \{c_{ij}\}$ are specified as $\bar{H}_j = \bar{E}_{\mathrm{core_p}(j)}$ and

$$c_{ij} = \begin{cases} 1, & \text{if } j = \mathrm{comp}_\mathrm{p}(i), \\ 0, & \text{otherwise}, \end{cases} \tag{10}$$

respectively. Applying eqn (7)–(9) and after some algebraic manipulations, the following formulae (constrained MARVEL equations) can be deduced:

$$\bar{H}_1 = \bar{H}_2 = \cdots = \bar{H}_{N_{\mathrm{a,c}}} = 0, \tag{11}$$

$$\mathbf{G}_\mathbf{C}^{\langle\mathcal{N}_\mathrm{a}\rangle}\bar{\mathbf{H}} = \mathbf{F}_\mathbf{C}^{\langle\mathcal{N}_\mathrm{a}\rangle}, \tag{12}$$

where $\mathbf{G}_\mathbf{C}^{\langle\mathcal{N}_\mathrm{a}\rangle} = \mathbf{C}^\mathrm{T}\mathbf{G}^{\langle\mathcal{N}_\mathrm{a}\rangle}\mathbf{C}$ and $\mathbf{F}_\mathbf{C}^{\langle\mathcal{N}_\mathrm{a}\rangle} = \mathbf{C}^\mathrm{T}\left(\mathbf{F}^{\langle\mathcal{N}_\mathrm{a}\rangle} - \mathbf{G}^{\langle\mathcal{N}_\mathrm{a}\rangle}\boldsymbol{\beta}\right)$.

Note that $\mathbf{G}_\mathbf{C}^{\langle\mathcal{N}_\mathrm{a}\rangle}$ is a much smaller matrix than $\mathbf{G}$; thus, solving eqn (11) and (12) is much less expensive than solving eqn (2) and (3). In fact, for a database of approximately 260 000 $H_2{}^{16}O$ transitions, the traditional MARVEL and the constrained MARVEL codes were executed with fixed uncertainties. For the latter procedure, transitions were divided into eight blocks, indexed with −9, −8, −7, −6, −5, −4, −3, and −2, which contained 1, 9, 45, 666, 8304, 71 990, 174 965, and 3626 lines, respectively. The computational time concerning the older and the newer approaches was 17.8 and 4.5 s on a single processor, respectively, which corresponds to a speed-up of about four. Nevertheless, in contrast to the $\mathbf{G}$ matrix, $\mathbf{G}_\mathbf{C}^{\langle\mathcal{N}_\mathrm{a}\rangle}$ is not diagonally dominant; thus, a pivoting strategy must be employed to yield the solution of eqn (11) and (12). Once $\bar{\mathbf{H}}$ is determined, $\bar{\mathbf{E}}$ can easily be calculated by means of eqn (9).

Utilizing eqn (11) and (12), a divide-and-conquer-style algorithm (blockMARVEL procedure) was designed, whose successive steps are listed in Appendix B. During this refinement, (a) lines are divided into blocks according to the orders of magnitude of the associated RSUs, (b) $\mathcal{N}_\mathrm{a}$ is extended block by block through linking the transitions from the upcoming block to this subnetwork, and (c) the uncertainties of the added lines are adjusted to achieve consistency within $\mathcal{N}_\mathrm{a}$. At the end of this process, accurate constrained empirical energy levels and reliable wavenumber uncertainties are obtained, provided that (a) all the segments are well calibrated, and (b) the CD relations are synchronized (see the next two subsections).

2.4.3 Accuracy-based recalibration (ABR). Having the constrained energy levels derived, each ill-calibrated source segment, for which the average of the residuals related to its transitions significantly differs from zero, has to be recalibrated.

To recalibrate the segments, they have to be placed into the $R(1)$, $R(2)$, …, $R(N_\mathrm{R})$ recalibration classes, where N_R is the number of recalibration classes. This is achieved the following way: (a) those segments which must be recalibrated with the same factor are grouped into the same class, and (b) all the segments for which simple recalibration (recalibration with a single factor) is not permitted, are distributed into separate classes. For the $R(i)$ class, a regional subnetwork, $\mathcal{N}_\mathrm{reg}$, which contains the non-bridge lines of the segments in $R(i)$, is introduced. To assign a recalibration factor to $R(i)$, the following objective function should be optimized:

$$\Omega_\mathrm{reg}(f) = \sum_{j=1}^{N_\mathrm{reg}} w_{\mathrm{reg},j}\left(f\sigma_{p_j} - \bar{\sigma}_{p_j}\right)^2, \tag{13}$$

where (a) p_j is the index of the jth line in $\mathcal{N}_\mathrm{reg}$, (b) $w_{\mathrm{reg},j} = \delta_{\mathrm{reg},j}{}^{-2}$, (c) $\delta_{\mathrm{reg},j}$ is an approximate uncertainty of the p_jth transition during recalibration, and (d) N_reg is the number of lines in $\mathcal{N}_\mathrm{reg}$. The best estimate of the recalibration factor, f_best, can be expressed as

$$f_\mathrm{best} = \frac{\sum_{j=1}^{N_\mathrm{reg}} w_{\mathrm{reg},j}\sigma_{p_j}\bar{\sigma}_{p_j}}{\sum_{j=1}^{N_\mathrm{reg}} w_{\mathrm{reg},j}\bar{\sigma}_{p_j}{}^2}. \tag{14}$$

Starting from proper initial values, the $\delta_{\mathrm{reg},j}$ uncertainties are iteratively adjusted to provide an f_best value satisfying $\delta_{\mathrm{reg},j} \leq |\Delta_{\mathrm{reg},j}|$, where $\Delta_{\mathrm{reg},j} = f_\mathrm{best}\sigma_{p_j} - \bar{\sigma}_{p_j}$ for all $1 \leq j \leq N_\mathrm{reg}$. If (a) the recalibration of $R(i)$ is permitted, (b) N_reg is sufficiently large, and (c) the f_best parameter significantly reduces the distance of the residuals from zero in $\mathcal{N}_\mathrm{reg}$, then f_best is used as the recalibration factor of the $R(i)$ recalibration class, otherwise $R(i)$ is not recalibrated.

Eqn (14) leads to an improved recalibration technique, called accuracy-based recalibration (ABR). Its sequential steps are presented in Appendix C. It is planned that the intMARVEL code will handle input recalibration factors to speed up its running.

If there are no segments that can be recalibrated, then our SN is fully recalibrated. In the case that the investigated SN is not recalibrated fully, the blockMARVEL analysis is re-executed (only one more time) at the termination of the ABR method.

2.4.4 Local synchronization procedure. It may occur that those constrained empirical energy levels which were determined only by very few transitions during the blockMARVEL refinement must be substituted with more dependable energy values using the CD relations of the energy levels. For the ith energy level, a local subnetwork, $\mathcal{N}_\mathrm{loc}$ is defined, where all the non-bridge transitions incident to this energy level are included. To calculate the best energy value, e_best, available from the lines of $\mathcal{N}_\mathrm{loc}$, the following objective function is minimized:

$$\Omega_\mathrm{loc}(e) = \sum_{j=1}^{N_\mathrm{loc}} w_{\mathrm{loc},j}\left(e_j - e\right)^2, \tag{15}$$

where (a) e_j is the trial energy of the jth line in $\mathcal{N}_\mathrm{loc}$, (b) $w_{\mathrm{loc},j} = \delta_{\mathrm{loc},j}^{-2}$, (c) $\delta_{\mathrm{loc},j}$ is an approximate uncertainty of the jth

transition inserted into $\mathcal{N}_{\rm loc}$, and (d) $\mathcal{N}_{\rm loc}$ is the number of lines in $\mathcal{N}_{\rm loc}$. The e_j energy is specified as

$$e_j = \begin{cases} \sigma_{u_j} + \bar{E}_{\rm low(u_j)}, & \text{if up}(u_j) = i, \\ \bar{E}_{\rm up(u_j)} - \sigma_{u_j}, & \text{otherwise}, \end{cases} \quad (16)$$

where u_j is the index of the jth line in $\mathcal{N}_{\rm loc}$. Similarly to $f_{\rm best}$ in eqn (14), the $e_{\rm best}$ value is expanded in a closed form:

$$e_{\rm best} = \frac{\sum_{j=1}^{N_{\rm loc}} w_{{\rm loc},j} e_j}{\sum_{j=1}^{N_{\rm loc}} w_{{\rm loc},j}}. \quad (17)$$

As in Section 2.4.3, an iterative refinement can be performed for the $\delta_{{\rm loc},j}$ uncertainties for which reasonable guesses are employed. Obviously, the final $e_{\rm best}$ value should fulfill the relation $\delta_{{\rm loc},j} \leq |\Delta_{{\rm loc},j}|$, where $\Delta_{{\rm loc},j} = e_j - e_{\rm best}$, for all $1 \leq j \leq N_{\rm loc}$.

If (a) $N_{\rm loc}$ is sufficiently large, and (b) the $|\Delta_{u_j}|$ absolute residuals are systematically larger than the corresponding $|\Delta_{{\rm loc},j}|$ values, then (a) the ith energy level is synchronized with $\mathcal{N}_{\rm loc}$ (*i.e.*, $\bar{E}_i$ is substituted with $e_{\rm best}$), (b) all the transitions of $\mathcal{N}_{\rm loc}$ are excluded from $\mathcal{N}_{\rm LS}$ whose recalculated fitting defects are positive, and (c) those lines of $\mathcal{N}_{\rm loc}$ which were incorrectly excluded from $\mathcal{N}_{\rm LS}$ are reincluded in $\mathcal{N}_{\rm LS}$. Based on eqn (17), a so-called local synchronization procedure is constructed. The stages of this algorithm are detailed in Appendix D.

If there is no energy level left which should be synchronized with its $\mathcal{N}_{\rm loc}$, then the SN is said to be perfectly synchronized. If our SN is not perfectly synchronized, the blockMARVEL procedure has to be repeated upon completion of the local synchronization process.

2.4.5 Full network characterization (FNC). As the closing step of the intMARVEL procedure, the energy levels, the transitions, and the segments are given quality labels *via* the so-called full network characterization (FNC) technique. During this step, described in detail in Appendix E, (a) the energy levels are equipped with uncertainties, dependability grades, and useful connectivity parameters, (b) the transitions are rated using the grades of their upper and lower energy levels, and (c) segments are associated with statistical data concerning their wavenumber ranges and their accuracy. The FNC completes the intMARVEL process.

3 Experimental data sources

An almost complete set of data sources, available up to 2013, about recorded and assigned high-resolution spectra of $H_2{}^{16}O$ is given in the IUPAC compilation, ref. 18. Since then, partly helped by the data of ref. 18, results of a number of new spectroscopic measurements and a number of new analyses have been published on $H_2{}^{16}O$, some highly relevant for this study.[46,47,63,99–103] Note that a complete update of the IUPAC Part III paper[18] is in progress.[108]

In Table 2 segments of experimental sources, old and new, are collected reporting highly accurate measured transitions, defined as having ESU $\leq 10^{-4}$ cm^{-1}. These are the segments utilized during this study. For the experimental data sources we follow the naming convention of a IUPAC study.[16] The representation of the segment accuracies can be found in Table 2. It is comfortable to note that good agreement is seen among the ESU, RSU, ASU, and MSU values.

There is one source which reports hyperfine transitions characterizing the spectra of *ortho*-$H_2{}^{16}O$, 09CaPuHaGa.[58] Not the hyperfine split transitions but their weighted averages, also reported in 09CaPuHaGa,[58] have been utilized in this study.

Whenever possible, the available experimental information was used to determine the ESU values of the source segments. In some cases, when the accuracy of the source segment was not explicitly given, an educated guess had to be employed for the ESU value. When line-by-line uncertainties were reported, *e.g.*, in the cases of 95MaOdIwTs,[60] 12YuPeDrMa,[98] 11DrYuPeGu,[97] 13YuPeDr,[63] and 06MaToNaMo,[95] we formed an average from these data. The extended MARVEL protocol requires that less accurate segments of sources are ignored from the analysis of more accurate data. In the case of the sources 93Toth,[109] 93Tothb,[83] and 91Toth,[81] only the transitions given with five digits after the decimal point were included in our initial dataset.

Our extMARVEL refinements indicated that certain sources, namely 82KaJoHo,[110] 83Guelachv,[111] 14ReOuMiWa,[100] 17MoMiKaBe,[46] and 18MiMoKaKa,[103] are not sufficiently accurate for the purposes of the present study. Therefore, the transitions these sources contain were omitted from our analysis.

4 Results and discussion

Let us start by providing some statistical information about the present analysis of the $H_2{}^{16}O$ measured transitions based on the highly accurate data source segments reported in Table 2. The number of transition entries in the leading subnetwork is 3099 and 3988 for *ortho*- and *para*-$H_2{}^{16}O$, respectively. As to transitions of A^+/A^- grade (the detailed definition of the grades of the empirical energy levels determined in this study is given in Appendix E), their number is 462/650 for *ortho*/*para*-$H_2{}^{16}O$. The number of lines which had to be excluded from our extMARVEL analysis from the source segments selected is only 8, three of which are in the THz region (*vide infra*). For *ortho*-$H_2{}^{16}O$, the number of energy levels determined is 725, out of which 97 have A^+/A^- grades. For *para*-$H_2{}^{16}O$, the extMARVEL analysis resulted in 857 energy levels, of which 117 received A^+/A^- grades. The largest J value, $J_{\rm max}$, is 17/18 for *para*/*ortho*-$H_2{}^{16}O$. The range of highly accurate $H_2{}^{16}O$ energy levels is quite extended, it is 0–3000 cm^{-1} for *para*-$H_2{}^{16}O$ and somewhat narrower, 0–1800 cm^{-1}, for *ortho*-$H_2{}^{16}O$.

One of the important practical results of this study, a set of highly accurate empirical energy levels of $H_2{}^{16}O$, is reported in Table 3. Only empirical energy levels of grade A^+ and A^- quality are included in Table 3. The highly accurate rovibrational energy levels belong to the vibrational states (v_1 v_2 v_3) = (0 0 0), (0 1 0), and (0 2 0), where v_1, v_2, and v_3 stand for the symmetric stretch, bend, and antisymmetric stretch vibrational quantum numbers.

Table 2 $H_2{}^{16}O$ data source segments employed during this work and some of their most important characteristics[a]

Segment tag	Range/cm^{-1}	*A*/*V*/*E*	ESU/cm^{-1}	RSU/cm^{-1}	ASU/cm^{-1}	MSU/cm^{-1}
69Kukolich[48]	0.74168–0.74168	1/1/1	2.000×10^{-9}	$2.000 \times 10^{-}9$	2.000×10^{-9}	2.000×10^{-9}
09CaPuHaGa[58]	10.715–20.704	7/7/7	2.000×10^{-8}	2.000×10^{-8}	2.000×10^{-8}	2.000×10^{-8}
71Huiszoon[53]	6.1146–6.1146	1/1/1	4.000×10^{-8}	4.000×10^{-8}	4.000×10^{-8}	4.000×10^{-8}
95MaKr[59]	18.577–18.577	1/1/1	7.000×10^{-8}	7.000×10^{-8}	9.475×10^{-8}	9.475×10^{-8}
06GoMaGuKn[54]	6.1146–18.577	12/12/11	2.000×10^{-7}	1.145×10^{-7}	8.827×10^{-8}	2.724×10^{-7}
18KaStCaCa[102]	7164.9–7185.6	8/8/8	4.000×10^{-7}	4.000×10^{-7}	4.000×10^{-7}	4.000×10^{-7}
09CaPuBuTa[96]	36.604–53.444	9/9/9	5.000×10^{-7}	5.000×10^{-7}	5.000×10^{-7}	5.000×10^{-7}
11Koshelev[64]	25.085–25.085	1/1/1	5.000×10^{-7}	5.000×10^{-7}	5.000×10^{-7}	5.000×10^{-7}
07KoTrGoPa[57]	10.715–12.682	3/3/3	6.000×10^{-7}	8.507×10^{-7}	8.584×10^{-7}	1.725×10^{-6}
83HeMeLu[65]	13.013–32.954	7/7/7	7.000×10^{-7}	7.000×10^{-7}	7.000×10^{-7}	7.000×10^{-7}
83MeLuHe[66]	16.797–32.954	5/5/5	7.000×10^{-7}	7.000×10^{-7}	7.000×10^{-7}	7.000×10^{-7}
71StBe[49]	0.74168–6.1146	3/3/3	9.000×10^{-7}	1.941×10^{-6}	5.186×10^{-6}	1.241×10^{-5}
87BaAlAlPo[77]	14.199–19.077	6/6/5	1.000×10^{-6}	1.000×10^{-6}	4.244×10^{-6}	2.046×10^{-5}
95MaOdIwTs[60]	18.577–162.44	139/138/135	1.000×10^{-6}	1.281×10^{-6}	2.520×10^{-6}	3.638×10^{-5}
18ChHuTaSu[47]	12.622–12.665	6/6/6	1.000×10^{-6}	1.000×10^{-6}	1.000×10^{-6}	1.000×10^{-6}
54PoSt[51]	0.74169–0.74169	1/1/1	2.000×10^{-6}	4.673×10^{-6}	5.142×10^{-6}	5.142×10^{-6}
80Kuze[75]	0.40057–4.0026	5/5/4	2.000×10^{-6}	2.000×10^{-6}	9.021×10^{-6}	3.440×10^{-5}
91AmSc[79]	8.2537–11.835	5/5/5	2.000×10^{-6}	2.000×10^{-6}	2.000×10^{-6}	2.000×10^{-6}
97NaLoInNo[88]	118.32–119.07	5/5/4	2.000×10^{-6}	2.000×10^{-6}	6.032×10^{-6}	2.216×10^{-5}
91PeAnHeLu[80]	4.6570–19.804	30/30/30	3.000×10^{-6}	3.000×10^{-6}	3.907×10^{-6}	1.652×10^{-5}
83BuFeKaPo[62]	16.797–21.545	5/5/5	4.000×10^{-6}	2.316×10^{-6}	2.120×10^{-6}	5.070×10^{-6}
95Pearson[85]	4.3300–17.220	9/8/8	4.000×10^{-6}	4.000×10^{-6}	4.000×10^{-6}	4.000×10^{-6}
06MaToNaMo[95]	28.685–165.31	104/104/102	4.000×10^{-6}	1.745×10^{-6}	2.637×10^{-6}	3.066×10^{-5}
12YuPeDrMa[98]	9.8572–90.767	103/102/101	4.000×10^{-6}	4.000×10^{-6}	6.643×10^{-6}	2.337×10^{-4}
72LuHeCoGo[55]	6.1146–25.085	14/14/13	5.000×10^{-6}	1.062×10^{-6}	1.451×10^{-6}	5.427×10^{-6}
13YuPeDr[63]	17.690–67.209	182/182/181	5.000×10^{-6}	5.000×10^{-6}	5.992×10^{-6}	6.152×10^{-5}
11DrYuPeGu[97]	82.862–90.843	26/25/25	6.000×10^{-6}	6.000×10^{-6}	6.734×10^{-6}	1.156×10^{-5}
00ChPePiMa[90]	28.054–52.511	17/17/17	8.000×10^{-6}	1.600×10^{-6}	3.184×10^{-6}	1.422×10^{-5}
79HeJoMc[74]	0.072049–0.072049	1/1/1	1.000×10^{-5}	1.000×10^{-5}	1.000×10^{-5}	1.000×10^{-5}
51Jen[50]	0.74169–0.74169	1/1/0	2.000×10^{-5}	2.000×10^{-5}	2.000×10^{-5}	2.000×10^{-5}
72FlCaVa[52]	0.74168–25.085	7/7/0	2.000×10^{-5}	1.266×10^{-5}	1.509×10^{-5}	3.015×10^{-5}
06JoPaZeCo[94]	1485.1–1486.2	2/2/2	2.000×10^{-5}	2.000×10^{-5}	3.116×10^{-5}	4.232×10^{-5}
70StSt[61]	18.577–18.577	1/1/0	3.000×10^{-5}	6.241×10^{-6}	6.883×10^{-6}	6.883×10^{-6}
87BeKoPoTr[78]	7.7616–19.850	5/5/5	3.000×10^{-5}	4.103×10^{-5}	2.858×10^{-5}	4.663×10^{-5}
96Belov[86]	28.054–30.792	5/5/5	3.000×10^{-5}	7.376×10^{-6}	7.849×10^{-6}	1.765×10^{-5}
05HoAnAlPi[92]	212.56–594.95	166/164/60	3.000×10^{-5}	1.640×10^{-5}	2.668×10^{-5}	3.643×10^{-4}
13LuLiWaLi[99]	12.573–12.752	73/73/65	3.000×10^{-5}	9.660×10^{-5}	6.039×10^{-5}	5.243×10^{-4}
54KiGo[56]	6.1146–6.1146	1/1/0	5.000×10^{-5}	3.710×10^{-5}	4.437×10^{-5}	4.437×10^{-5}
96BrMa[87]	5206.3–5396.5	28/28/28	5.000×10^{-5}	1.327×10^{-5}	2.534×10^{-5}	9.346×10^{-5}
91Toth[81]	1072.6–2265.3	740/738/722	6.000×10^{-5}	2.117×10^{-5}	5.183×10^{-5}	1.460×10^{-3}
97MiTyKeWi[89]	2507.2–4402.8	935/935/190	6.000×10^{-5}	1.639×10^{-4}	4.774×10^{-4}	1.050×10^{-2}
95PaHo[84]	177.86–519.59	246/246/5	7.000×10^{-5}	1.072×10^{-4}	2.200×10^{-4}	9.827×10^{-3}
93Totha[82]	1316.1–4260.4	587/587/575	8.000×10^{-5}	3.317×10^{-5}	8.755×10^{-5}	9.449×10^{-4}
93Tothb[83]	1881.1–4306.7	1076/1076/1061	8.000×10^{-5}	3.428×10^{-5}	8.151×10^{-5}	1.485×10^{-3}
05Toth[93]	2926.5–7641.9	1895/1895/1832	8.000×10^{-5}	6.031×10^{-5}	1.396×10^{-4}	4.447×10^{-3}
85BrTo[76]	1323.3–1992.7	71/71/69	1.000×10^{-4}	2.042×10^{-5}	8.836×10^{-5}	1.877×10^{-3}
03ZoVa[91]	3010.2–4044.9	469/469/456	1.000×10^{-4}	9.533×10^{-5}	2.385×10^{-4}	2.465×10^{-3}
15SiHo[101]	7714.8–7919.9	71/71/62	1.000×10^{-4}	1.000×10^{-4}	2.266×10^{-4}	4.984×10^{-3}

[a] Tags denote experimental data-source segments used in this study. The column 'Range' indicates the range corresponding to validated wavenumber entries within the experimental linelist. '*A*/*V*/*E*' is an ordered triplet with *A* = the number of assigned transitions in the source segments, *V* = the number of validated transitions, and *E* = the number of exploited transitions (see also Appendix E). In the heading of this table, ESU, RSU, ASU, and MSU designate the estimated, the refined, the average, and the maximum segment uncertainties, respectively. Rows of this table are arranged in the order of the ESUs.

All energy levels reported have at least an 8-digit accuracy, often considerably better (up to 10 digits of accuracy). J_{max} is 13/14 for the most dependable *ortho*/*para* rovibrational energy levels presented in Table 3.

4.1 Highly accurate sources

Fig. 2 shows the unsigned residuals (differences between the observed wavenumbers and their extMARVEL estimates) for those segments having RSU $< 10^{-5}$ cm^{-1}. Clearly, the extMARVEL refinement retained the high accuracy of the original measurements.

Some of the extMARVEL rovibrational energy levels may have an accuracy considerably lower than that of the transitions determining it. The simplest example that highlights the difficulties of providing uncertainties to energy levels is as follows. Let us have two separate 4-cycles, both formed by highly accurate transitions, connected by a bridge of lower measured accuracy. If one of the 4-cycles contains a core of the leading subnetwork, then all its energy levels have the same high accuracy as the measured transitions. However, this is not true for the other 4-cycle, where the transitions are known with high accuracy, but the overall accuracy of the energy levels is determined

Table 3 Accurate empirical (extMARVEL) energy levels, and their assignments, for *para*-$H_2{}^{16}O$ (first four columns) and *ortho*-$H_2{}^{16}O$ (last four columns), of grade A^+ and A^- quality. All of the energy levels are of grade A^+ quality, except those with an asterisk, which are of grade A^- quality

Energy/cm^{-1}	$(v_1\ v_2\ v_3)J_{K_a,K_c}$	Energy/cm^{-1}	$(v_1\ v_2\ v_3)J_{K_a,K_c}$	Energy/cm^{-1}	$(v_1\ v_2\ v_3)J_{K_a,K_c}$	Energy/cm^{-1}	$(v_1\ v_2\ v_3)J_{K_a,K_c}$
0	$(0\ 0\ 0)0_{0,0}$	1806.671529(30)	$(0\ 0\ 0)13_{1,13}$	0	$(0\ 0\ 0)1_{0,1}$	1765.248467(64)	$(0\ 0\ 0)8_{8,1}$
37.13712384(68)	$(0\ 0\ 0)1_{1,1}$	1810.583280(79)	$(0\ 0\ 0)9_{7,3}$	18.57738488(39)	$(0\ 0\ 0)1_{1,0}$	1782.875649(18)	$(0\ 0\ 0)13_{0,13}$
70.09081349(65)	$(0\ 0\ 0)2_{0,2}$	1813.787601(17)	$(0\ 1\ 0)3_{2,2}$	55.7020277(99)	$(0\ 0\ 0)2_{1,2}$	1786.7935598(52)	$(0\ 0\ 0)9_{7,2}$
95.1759380(31)	$(0\ 0\ 0)2_{1,1}$	1817.451194(21)	$(0\ 1\ 0)4_{0,4}$	111.1072838(60)	$(0\ 0\ 0)2_{2,1}$	1789.429033(13)	$(0\ 0\ 0)11_{3,8}$
136.1639195(12)	$(0\ 0\ 0)2_{2,0}$	1843.029604(33)	$(0\ 0\ 0)11_{4,8}$	112.967305(10)	$(0\ 0\ 0)3_{0,3}$	1795.5407549(85)	$(0\ 1\ 0)3_{2,1}$
142.2784859(10)	$(0\ 0\ 0)3_{1,3}$	1875.461821(19)	$(0\ 0\ 0)10_{6,4}$	149.5714542(92)	$(0\ 0\ 0)3_{1,2}$	1797.802452(11)	$(0\ 1\ 0)4_{1,4}$
206.301428(21)	$(0\ 0\ 0)3_{2,2}$	1875.469719(18)	$(0\ 1\ 0)4_{1,3}$	188.36200986(47)	$(0\ 0\ 0)3_{2,1}$	1851.178614(11)	$(0\ 0\ 0)10_{6,5}$
222.052757(13)	$(0\ 0\ 0)4_{0,4}$	1907.451421(19)	$(0\ 1\ 0)3_{3,1}$	201.04402869(42)	$(0\ 0\ 0)4_{1,4}$	1875.213801(11)	$(0\ 0\ 0)11_{4,7}$
275.497042(18)	$(0\ 0\ 0)4_{1,3}$	1922.829071(12)	$(0\ 1\ 0)5_{1,5}$	261.6242179(88)	$(0\ 0\ 0)3_{3,0}$	1883.821404(23)	$(0\ 1\ 0)3_{3,0}$
285.219339(10)	$(0\ 0\ 0)3_{3,1}$	1922.901125(11)	$(0\ 1\ 0)4_{2,2}$	276.56792527(55)	$(0\ 0\ 0)4_{2,3}$	1884.221976(10)	$(0\ 1\ 0)4_{2,3}$
315.77953341(82)	$(0\ 0\ 0)4_{2,2}$	1960.207413(33)	$(0\ 0\ 0)12_{2,10}$	301.5535473(94)	$(0\ 0\ 0)5_{0,5}$	1896.972180(16)	$(0\ 1\ 0)5_{0,5}$
326.62546666(58)	$(0\ 0\ 0)5_{1,5}$	1985.784894(12)	$(0\ 0\ 0)11_{5,7}$	358.722528(11)	$(0\ 0\ 0)4_{3,2}$	1938.712538(10)	$(0\ 0\ 0)12_{3,10}$
383.842515(13)	$(0\ 0\ 0)4_{3,1}$	2005.917050(16)	$(0\ 1\ 0)4_{3,1}$	375.663155(11)	$(0\ 0\ 0)5_{1,4}$	1975.200945(19)	$(0\ 0\ 0)11_{5,6}$
416.2087402(12)	$(0\ 0\ 0)5_{2,4}$	2024.152654(39)	$(0\ 1\ 0)5_{2,4}$	422.71630882(88)	$(0\ 0\ 0)5_{2,3}$	1977.068674(18)	$(0\ 1\ 0)5_{1,4}$
446.696589(13)	$(0\ 0\ 0)6_{0,6}$	2041.780551(28)	$(0\ 1\ 0)6_{0,6}$	423.4579912(56)	$(0\ 0\ 0)6_{1,6}$	1981.0213341(95)	$(0\ 1\ 0)4_{3,2}$
488.134178(15)	$(0\ 0\ 0)4_{4,0}$	2042.374098(77)	$(0\ 0\ 0)13_{2,12}$	464.313340(12)	$(0\ 0\ 0)4_{4,1}$	1986.010725(72)	$(0\ 0\ 0)9_{8,1}$
503.9681027(59)	$(0\ 0\ 0)5_{3,3}$	2054.368667(66)	$(0\ 0\ 0)10_{7,3}$	485.01769568(72)	$(0\ 0\ 0)5_{3,2}$	2018.516215(38)	$(0\ 0\ 0)13_{1,12}$
542.905778(13)	$(0\ 0\ 0)6_{1,5}$	2105.867908(44)	$(0\ 0\ 0)12_{3,9}$	529.117030(21)	$(0\ 0\ 0)6_{2,5}$	2018.958981(16)	$(0\ 1\ 0)6_{1,6}$
586.4791835(77)	$(0\ 0\ 0)7_{1,7}$	2126.407724(37)	$(0\ 1\ 0)5_{3,3}$	562.449182(22)	$(0\ 0\ 0)7_{0,7}$	2030.174344(14)	$(0\ 1\ 0)5_{2,3}$
602.7734936(28)	$(0\ 0\ 0)6_{2,4}$	2129.618682(26)	$(0\ 1\ 0)4_{4,0}$	586.546806(40)	$(0\ 0\ 0)5_{4,1}$	2030.550848(17)	$(0\ 0\ 0)10_{7,4}$
610.114430(14)	$(0\ 0\ 0)5_{4,2}$	2142.597661(55)	$(0\ 0\ 0)11_{6,6}$	625.184327(15)	$(0\ 0\ 0)6_{3,4}$	2101.157030(17)	$(0\ 0\ 0)12_{4,9}$
661.5489133(67)	$(0\ 0\ 0)6_{3,3}$	2146.263726(36)	$(0\ 1\ 0)6_{1,5}$	680.4196881(12)	$(0\ 0\ 0)7_{1,6}$	2105.804845(43)	$(0\ 1\ 0)4_{4,1}$
709.608213(12)	$(0\ 0\ 0)7_{2,6}$	2205.652716(69)	$(0\ 0\ 0)12_{4,8}$	718.281923(24)	$(0\ 0\ 0)5_{5,0}$	2106.699939(14)	$(0\ 1\ 0)5_{3,2}$
742.073025(29)	$(0\ 0\ 0)5_{5,1}$	2211.1906371(89)	$(0\ 1\ 0)6_{2,4}$	720.368303(20)	$(0\ 0\ 0)8_{1,8}$	2120.251898(17)	$(0\ 0\ 0)11_{6,5}$
744.063661(23)	$(0\ 0\ 0)8_{0,8}$	2248.064567(84)	$(0\ 0\ 0)13_{3,11}$	732.93041648(96)	$(0\ 0\ 0)6_{4,3}$	2137.4916811(42)	$(0\ 1\ 0)6_{2,5}$
757.78018789(88)	$(0\ 0\ 0)6_{4,2}$	2300.685002(61)	$(0\ 0\ 0)12_{5,7}$	758.6154628(87)	$(0\ 0\ 0)7_{2,5}$	2223.090456(41)	$(0\ 0\ 0)13_{2,11}$
816.694236(52)	$(0\ 0\ 0)7_{3,5}$	2321.813015(61)	$(0\ 0\ 0)11_{7,5}$	818.562227(10)	$(0\ 0\ 0)7_{3,4}$	2228.068156(19)	$(0\ 1\ 0)5_{4,1}$
882.890327(21)	$(0\ 0\ 0)8_{1,7}$	2327.883775(49)	$(0\ 0\ 0)14_{1,13}$	861.805844(21)	$(0\ 0\ 0)8_{2,7}$	2230.489516(24)	$(0\ 0\ 0)10_{8,3}$
888.632650(14)	$(0\ 0\ 0)6_{5,1}$	2399.165477(17)	$(0\ 1\ 0)6_{4,2}$	864.804377(26)	$(0\ 0\ 0)6_{5,2}$	2247.917879(18)	$(0\ 1\ 0)6_{3,4}$
920.210001(25)	$(0\ 0\ 0)9_{1,9}$	2426.19618(21)	$(0\ 0\ 0)13_{4,10}$	896.373983(24)	$(0\ 0\ 0)9_{0,9}$	2251.578491(52)	$(0\ 0\ 0)12_{5,8}$
927.743902(11)	$(0\ 0\ 0)7_{4,4}$	2522.261331(64)	$(0\ 0\ 0)11_{8,4}$*	907.442739(12)	$(0\ 0\ 0)7_{4,3}$	2285.935851(16)	$(0\ 1\ 0)7_{1,6}$
982.911714(13)	$(0\ 0\ 0)8_{2,6}$	2613.104573(35)	$(0\ 0\ 0)12_{7,5}$*	982.321572(21)	$(0\ 0\ 0)8_{3,6}$	2298.111383(77)	$(0\ 0\ 0)11_{7,4}$
1045.0583403(41)	$(0\ 0\ 0)6_{6,0}$	2670.789689(15)	$(0\ 1\ 0)8_{3,5}$	1021.2635806(10)	$(0\ 0\ 0)6_{6,1}$	2304.119654(21)	$(0\ 0\ 0)14_{2,13}$
1050.157663(19)	$(0\ 0\ 0)8_{3,5}$	2748.099560(76)	$(0\ 0\ 0)13_{6,8}$*	1036.0410822(48)	$(0\ 0\ 0)7_{5,2}$	2313.872430(23)	$(0\ 1\ 0)8_{1,8}$
1059.646655(15)	$(0\ 0\ 0)7_{5,3}$	2920.132087(14)	$(0\ 1\ 0)8_{5,3}$	1055.285214(19)	$(0\ 0\ 0)9_{1,8}$	2368.7982056(85)	$(0\ 1\ 0)7_{2,5}$
1080.385444(19)	$(0\ 0\ 0)9_{2,8}$			1090.755547(36)	$(0\ 0\ 0)10_{1,10}$	2374.587138(32)	$(0\ 1\ 0)6_{4,3}$
1114.532248(50)	$(0\ 0\ 0)10_{0,10}$			1098.914168(13)	$(0\ 0\ 0)8_{4,5}$	2390.929063(14)	$(0\ 0\ 0)13_{3,10}$
1131.775573(22)	$(0\ 0\ 0)8_{4,4}$			1178.127134(11)	$(0\ 0\ 0)9_{2,7}$	2410.006047(55)	$(0\ 0\ 0)12_{6,7}$
1216.189769(53)	$(0\ 0\ 0)7_{6,2}$			1192.4001374(68)	$(0\ 0\ 0)7_{6,1}$	2439.0809341(94)	$(0\ 1\ 0)7_{3,4}$
1216.231260(21)	$(0\ 0\ 0)9_{3,7}$			1231.3723839(61)	$(0\ 0\ 0)8_{5,4}$	2471.371510(47)	$(0\ 1\ 0)8_{2,7}$
1255.911545(14)	$(0\ 0\ 0)8_{5,3}$			1259.12474186(98)	$(0\ 0\ 0)9_{3,6}$	2498.470791(42)	$(0\ 0\ 0)11_{8,3}$
1293.018138(13)	$(0\ 0\ 0)10_{1,9}$			1269.8396771(11)	$(0\ 0\ 0)10_{2,9}$	2509.998800(16)	$(0\ 0\ 0)13_{4,9}$
1327.117604(24)	$(0\ 0\ 0)11_{1,11}$			1303.315623(28)	$(0\ 0\ 0)11_{0,11}$	2527.689134(62)	$(0\ 0\ 0)14_{3,12}$
1340.884880(14)	$(0\ 0\ 0)9_{4,6}$			1336.440968(13)	$(0\ 0\ 0)9_{4,5}$	2548.3448242(97)	$(0\ 1\ 0)7_{4,3}$
1394.814159(54)	$(0\ 0\ 0)7_{7,1}$			1371.0198349(43)	$(0\ 0\ 0)7_{7,0}$	2589.005435(18)	$(0\ 0\ 0)12_{7,6}$
1411.641890(12)	$(0\ 0\ 0)8_{6,2}$			1387.8170727(57)	$(0\ 0\ 0)8_{6,3}$	2605.540121(60)	$(0\ 0\ 0)13_{5,8}$
1437.968586(14)	$(0\ 0\ 0)10_{2,8}$			1422.333875(12)	$(0\ 0\ 0)10_{3,8}$	2606.398257(16)	$(0\ 1\ 0)8_{3,6}$
1474.980787(12)	$(0\ 0\ 0)9_{5,5}$			1453.502995(12)	$(0\ 0\ 0)9_{5,4}$	2607.474510(66)	$(0\ 0\ 0)15_{1,14}$
1525.135991(43)	$(0\ 0\ 0)11_{2,10}$			1501.053523(22)	$(0\ 0\ 0)11_{1,10}$	2664.285572(38)	$(0\ 1\ 0)9_{1,8}$*
1538.149477(27)	$(0\ 0\ 0)10_{3,7}$			1534.053335(22)	$(0\ 0\ 0)12_{1,12}$	2700.372781(12)	$(0\ 1\ 0)7_{5,2}$
1557.844418(26)	$(0\ 0\ 0)12_{0,12}$			1557.541614(18)	$(0\ 0\ 0)10_{4,7}$	2794.603744(30)	$(0\ 1\ 0)9_{2,7}$
1590.69071(14)	$(0\ 0\ 0)8_{7,1}$			1566.89570(11)	$(0\ 0\ 0)8_{7,2}$	2894.450561(85)	$(0\ 0\ 0)14_{5,10}$*
1616.453054(22)	$(0\ 0\ 0)10_{4,6}$			1594.762769(66)	$(0\ 1\ 0)1_{0,1}$	2895.8386511(81)	$(0\ 1\ 0)8_{5,4}$
1631.245487(18)	$(0\ 0\ 0)9_{6,4}$			1607.588639(18)	$(0\ 0\ 0)9_{6,3}$	3368.954987(18)	$(0\ 2\ 0)3_{2,1}$
1634.967095(11)	$(0\ 1\ 0)1_{1,1}$			1616.711503(44)	$(0\ 1\ 0)1_{1,0}$	4383.251954(52)	$(0\ 2\ 0)6_{6,1}$
1664.964587(11)	$(0\ 1\ 0)2_{0,2}$			1653.267093(20)	$(0\ 1\ 0)2_{1,2}$		

by the accuracy of the bridge. This also means that if the transitions are reconstructed from MARVEL energy levels, they may have considerably higher uncertainties than the directly measured transitions. This characteristics of the MARVEL protocol cannot be easily circumvented without new, accurate measurements. Thus, in MARVEL determinations of empirical rovibronic energy levels it can happen that the reconstructed lines should have a better uncertainty than indicated by the energy-level-based uncertainties. For the present study this means that uncertainties of transitions determined from uncertainties of rovibrational energy levels may not be fully realistic, they may provide inflated uncertainties, which must be considered when these transitions are used in an application.

For *para*-$H_2{}^{16}O$ one of the most important lines is the $5_{1,5} \leftarrow 4_{2,2}$ pure rotational transition at about 325.153 GHz (this is how this transition is usually reported) within the (0 0 0) ground vibrational state. Our final 8-digit-accuracy determination gives 325 152 899(2) kHz, based basically upon the experimental

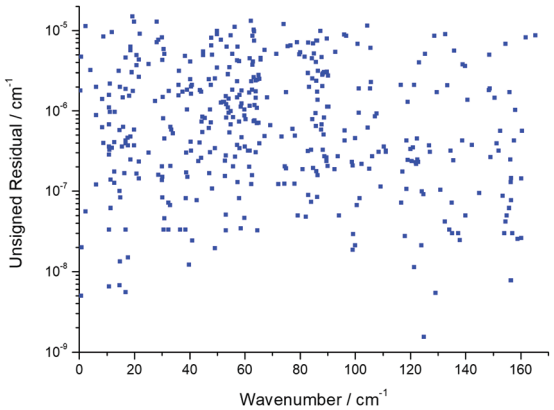


Fig. 2 Unsigned residuals (differences between the observed and the calculated wavenumbers) for the source segments with RSU $\approx 10^{-7}$ cm^{-1} and RSU $\approx 10^{-6}$ cm^{-1}, where RSU stands for "refined segment uncertainty". Residuals below 10^{-9} cm^{-1} are not plotted as these are artificial results.

result of 06GoMaGuKn,[54] but including the effects of all other relevant measured transitions.

Next, let us turn our attention to the THz region. In the THz region the following transitions had to be excluded from the extMARVEL analysis (all wavenumbers in cm^{-1}): 39.003 0,[98] 82.638 9,[60] and 86.467 2.[97] 95MaOdIwTs[60] and 06MaToNaMo[95] are the two most dominant sources which report accurately measured purely rotational transitions in the THz region. The unsigned residuals, *i.e.*, the unsigned differences of the observed and the extMARVEL-predicted wavenumbers, are plotted in Fig. 3. The estimated ESU of the transitions in 95MaOdIwTs[60] is 1×10^{-6} cm^{-1}, *i.e.*, 30 kHz, while the RSU, according to our study, is slightly higher, 38 kHz. Fig. 3 shows that extMARVEL is basically able to confirm the claimed accuracy of the measurements when they are refined together with all the other relevant source segments. Only very few transitions have absolute residuals higher than 100 kHz. This means that the extMARVEL procedure is able to retain the accuracy of the original measurements.

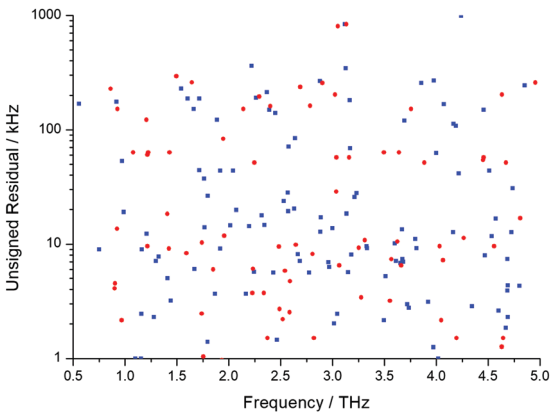


Fig. 3 Unsigned residuals for 95MaOdIwTs[60] and 06MaToNaMo,[95] denoted with blue and red dots, respectively. Residuals below 1 kHz are not plotted as these are artificial results.

It should prove beneficial for future studies to know more and more highly accurate transitions in a wider and wider spectral range. For example, for *ortho*-$H_2{}^{16}O$, two often quoted pure rotational transitions are 321.23 GHz for $10_{2,9} \leftarrow 9_{3,6}$ on (0 0 0) and 336.23 GHz for $5_{2,3} \leftarrow 6_{1,6}$ on (0 1 0). The extMARVEL counterpart of the first transition determined in this study is 321 225 677(44) kHz, while the frequency of the second transition is determined to be 336 228 131(140) kHz, both of which can be considered dependable due to the fact that they are based on energy levels of A$^+$ grade.

4.2 IAU lines

For all 13 $H_2{}^{16}O$ lines recommended by the International Astronomical Union (IAU) empirical estimates have been derived in this study. These estimates are listed in Table 1. Note that all the IAU-selected frequencies are below 1 THz. This is due partly to the fact that when these lines were selected THz spectroscopy measurements were rarely available.

The data presented in Table 1 allows an assessment of the extMARVEL frequencies in relation to those deemed to be most important by IAU. As evident from the last column of Table 1, all these lines have been determined experimentally *via* multiple, independent, highly accurate measurements, contributing to an improved confidence in the astronomically important frequencies.

4.3 Comparison with the energy levels of Lanquetin *et al.*

Fig. 4 shows the unsigned differences of the extMARVEL and the IUPAC energy levels with respect to the experimental JPL data, which are the same as those given by Lanquetin *et al.*[45]

The first observation one can make about Fig. 4 is that the extMARVEL protocol is a significant improvement over the standard MARVEL protocol employed to obtain the IUPAC Part III energy levels.[18] Furthermore, it is expected that the relation

$$\mathrm{UD}_i^{(\mathrm{M})} = \left|\bar{E}_{01\mathrm{LaCoCa},i} - \bar{E}_{\mathrm{M},i}\right| \leq \sqrt{\varepsilon_{01\mathrm{LaCoCa},i}{}^2 + \varepsilon_{\mathrm{M},i}{}^2} \quad (18)$$

should be (approximately) satisfied for the unsigned deviations ($\mathrm{UD}_i^{(\mathrm{M})}$), where (a) $\bar{E}_{\mathrm{M},i}$ is the energy of the ith energy level

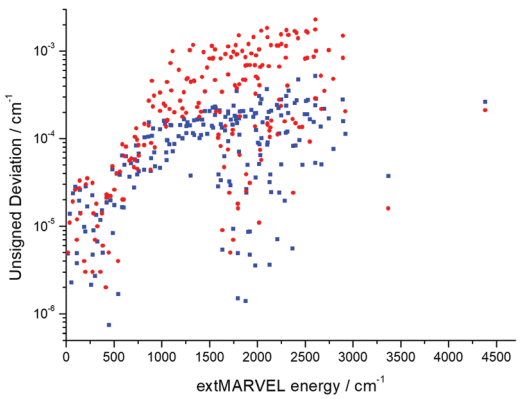


Fig. 4 Unsigned deviations of the extMARVEL (this work) and IUPAC Part III energy levels [see eqn (18)],[18] denoted with blue squares and red dots, respectively, referenced to those determined by Lanquetin *et al.*[45]

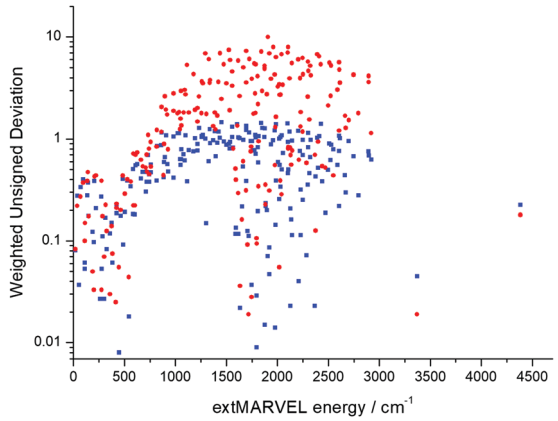


Fig. 5 Weighted unsigned deviations [see eqn (19)] of the extMARVEL (this work) and the IUPAC Part III[18] energy levels, denoted with blue squares and red dots, respectively, from those obtained in 01LaCoCa.[45]

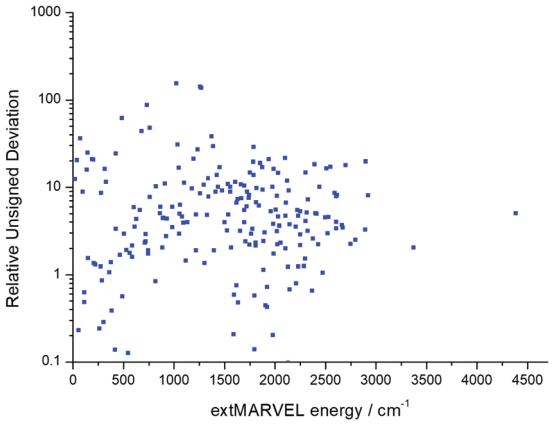


Fig. 6 Relative unsigned deviations [see eqn (21)] of the energy levels found in 01LaCoCa[45] from the extMARVEL energy levels.

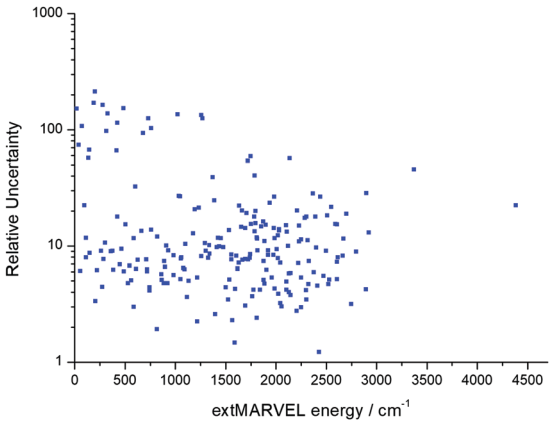


Fig. 7 Uncertainties of the energy levels in 01LaCoCa[45] relative to their extMARVEL counterparts.

obtained in this paper (M = extMARVEL) and in the IUPAC Part III paper[18] (M = IUPAC) with $\varepsilon_{M,i}$ uncertainty, and (b) $\bar{E}_{01LaCoCa,i}$ is the counterpart of $\bar{E}_{M,i}$ estimated in the source 01LaCoCa[45] with an $\varepsilon_{01LaCoCa,i}$ uncertainty. As a result, the corresponding weighted unsigned deviation (WUD), defined as

$$\mathrm{WUD}_i^{(M)} = \frac{\left|\bar{E}_{01LaCoCa,i} - \bar{E}_{M,i}\right|}{\sqrt{\varepsilon_{01LaCoCa,i}^2 + \varepsilon_{M,i}^2}}, \qquad (19)$$

must not be larger than 1. These weighted unsigned deviations are plotted in Fig. 5.

The substantial improvement achieved during the present study as compared to the IUPAC Part III study is obvious from Fig. 5. The relatively large WUDs related to the IUPAC Part III data (in fact, the uncertainties are smaller, by an order of magnitude, than the corresponding unsigned deviations), are attributed to the inaccuracy and the too optimistic energy uncertainties of some of the IUPAC Part III energy levels.

It is also worth examining whether the conditions

$$\left|\bar{E}_{01LaCoCa,i} - \bar{E}_{extMARVEL,i}\right| \leq \varepsilon_{extMARVEL,i} \qquad (20)$$

are (approximately) satisfied. Accordingly, the relative unsigned deviations (RUD),

$$\mathrm{RUD}_i = \frac{\left|\bar{E}_{01LaCoCa,i} - \bar{E}_{extMARVEL,i}\right|}{\varepsilon_{extMARVEL,i}} \leq 1, \qquad (21)$$

are formed and illustrated in Fig. 6. When these deviations are larger than 1, the 01LaCoCa[45] energies are not within the $\bar{E}_{extMARVEL,i} \pm \varepsilon_{extMARVEL,i}$ intervals. The protocol used in this study for the calculation of the uncertainties of the rovibrational energy levels is reliable; thus, relative deviations greater than 1 indicate that our energy levels are more accurate than those derived by Lanquetin *et al.*,[45] as also supported by the data of Table 1.

Fig. 7 shows the 01LaCoCa[45] uncertainties relative to their extMARVEL counterparts. Clearly, the present study represents a significant improvement over the uncertainties of the energies given in the source 01LaCoCa.[45]

Finally, let us support the conclusions based on the figures with some statistical data related to the UD, WUD, and RUD values. The average and the maximum UD decreased from 4.4×10^{-4} and 2.3×10^{-3} to 1.1×10^{-4} and 5.2×10^{-4} cm^{-1}, respectively. The average and the maximum WUD decreased from 2.21 to 0.59 and from 9.99 to 1.46, respectively. These substantial decreases in the deviations clearly prove that the extMARVEL treatment allows the full utilization of the most accurate spectroscopic measurements, as now we exceed the internal accuracy of the data presented by Lanquetin *et al.*[45]

5 Conclusions

During this study the standard Measured Active Rotational–Vibrational Energy Levels (MARVEL) algorithm[11,12] has been extended, in order to improve its performance toward allowing the automatic determination of highly accurate empirical energy levels, matching or exceeding the accuracy of the best underlying spectroscopic measurements. This is an especially important undertaking in the era of optical-frequency-comb

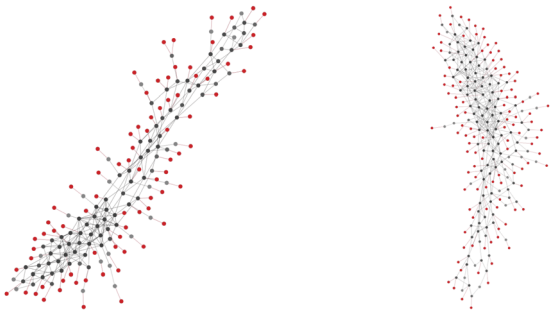
spectroscopies, as they can consistently yield orders of magnitude more accurate transitions than traditional high-resolution spectroscopic techniques.

There are several important algorithmic changes introduced in this study, resulting in what we call the extended MARVEL (extMARVEL) protocol. It is worth reiterating these improvements one by one as they all contributed to increase the utility of the MARVEL analysis of experimental transitions.

First, unlike in the standard version, in the new, extended protocol MARVEL-type analyses are performed based on the use of groups of transitions blocked by their estimated experimental uncertainties. This requires that the user segments the input sources based on assumed uncertainties of the different groups of transitions (a line-by-line analysis yielding individual initial line uncertainties would be ideal but this appears to be unrealistic). Second, the inversion and weighted least-squares refinement procedure is now based on sequential addition of blocks of decreasing accuracy. Wavenumber estimates determined in a given block are not allowed to be changed by the inclusion of less accurate measurements. Third, spectroscopic cycles are introduced during the refinement process. This is a particularly important advancement as due to the law of energy conservation[13] one can detect straightforwardly the best as well as the worst transitions in the collated set of experimentally measured and assigned transitions. Fourth, automated recalibration of the segments requiring this adjustment is performed. As shown before,[18] MARVEL is able to perform this job quite reliably. Fifth, synchronization of the combination difference relations is performed to reduce residual uncertainties in the resulting dataset of empirical (MARVEL) energy levels. Sixth, an improved classification scheme, providing seven grades decreasing in quality from A^+ to D, of the empirical energy levels is introduced, the grading is based on the assumed accuracy and dependability of the energy levels. This grading of the energy levels directly results in a grading of the measured transitions, as the lower grade of the lower and the upper energy level is attached to the line they define.

We used $H_2{}^{16}O$ as the molecule of choice for our feasibility study testing the extMARVEL protocol. Since the International Astronomical Union selected 13 water lines as highly important for astrophysical applications, it is important to note that all these transitions are reproduced now perfectly well by the extMARVEL protocol. From an application point of view it is also important to note that all these transitions are measured extremely accurately, though they are not part of cycles of similar high accuracy. This calls for new measurements, most likely involving optical-frequency-comb spectroscopic techniques. For *ortho*- and *para*-$H_2{}^{16}O$, we determined 97 and 117 energy levels with grades of at least A^- quality, meaning an accuracy better than 10^{-4} cm^{-1}. The range covered by these highly accurate rovibrational energies is quite substantial, 0–3000 cm^{-1} for *para*-$H_2{}^{16}O$ and 0–1800 cm^{-1} for *ortho*-$H_2{}^{16}O$.

The present dataset of highly accurate energy levels is larger than the experimental energy-level dataset maintained at JPL,



(a) *para*-$H_2{}^{16}O$ **(b)** *ortho*-$H_2{}^{16}O$

Fig. 8 The *para*- and *ortho*-$H_2{}^{16}O$ principal components of the experimental spectroscopic network of highly accurate measured transitions, including only those energy levels and transitions which have a grade of at least A^- quality. Note that the energy levels were determined by a much larger number of transitions, all those which have an uncertainty of better than 10^{-4} cm^{-1}.

which is the same as the one published by Lanquetin *et al.*[45] Furthermore, on average our data have an accuracy about an order of magnitude better than the data of Lanquetin *et al.*[45] Note also that the highly accurate energy levels are part of a large number of cycles, as clear from Fig. 8.

Conflicts of interest

There are no conflicts to declare.

Appendix A: stages of the HSP method (see Section 2.4.1)

S.1 for all $1 \le i \le N_s$: set $RSU^{[s(i)]} = ESU^{[s(i)]}$, where
- (a) $ESU^{[s(i)]}$ is the estimated segment uncertainty (ESU) of the $s(i)$ segment,
- (b) $RSU^{[s(i)]}$ is the (actual) refined segment uncertainty (RSU) of $s(i)$, and
- (c) N_s is the number of segments in the SN;

S.2 for all $1 \le i \le n_{mh}$, where n_{mh} (= 10) is the maximum number of HSP iterations:
- **S.2.1** save the indices of the segments into $k_1, k_2, \ldots, k_{N_s}$ so that $RSU^{[s(k_l)]} \le RSU^{[s(k_m)]}$ is enforced for all $1 \le l \le m \le N_s$;
- **S.2.2** for all $1 \le j \le N_s$:
  - **S.2.2.1** set $s = s(k_j)$;
  - **S.2.2.2** initialize $RSU^{[s]}_{prev}$ as $RSU^{[s]}$;
  - **S.2.2.3** if $n^{[s]} = 0$, then continue **S.2.2**, where $n^{[s]}$ is the number of rovibrational lines in $\mathcal{N}_{LS}$ and s;
  - **S.2.2.4** create a $\mathcal{N}_{per}$ perturbation subnetwork[I] with its $P^{\langle \mathcal{N}_{per} \rangle}$ participation matrix from $\mathcal{N}_{LS}$ as follows: $P_m^{\langle \mathcal{N}_{per} \rangle} = 1$ if $P_m = 1$ and either $s(i_m) = s$ or $\phi_{per} RSU^{[s]} \ge RSU^{[s(i_m)]}$, else $P_m^{\langle \mathcal{N}_{per} \rangle} = 0$ for all $1 \le m \le N_T$, where

(a) i_m is the mth segment index for which $\mathrm{s}(\mathrm{i}_m)$ includes the mth transition, and
(b) $\phi_{\mathrm{per}} \in (0,1]$ is the perturbation factor ($\phi_{\mathrm{per}} = 0.1$ is the default value);

S.2.2.5 construct a cycle basis (CB) for $\mathcal{N}_{\mathrm{per}}$, designated with $\mathcal{C}_{\mathrm{B}}$, using the breadth-first search (BFS) method;[112]

S.2.2.6 calculate the reduced discrepancy[II] of every $C \in \mathcal{C}_{\mathrm{B}}$ as

$$\mathrm{RD}_C^{[s]} = \begin{cases} 0, & \text{if } n_C^{[s]} = 0, \\ \left(\mathcal{D}_C - \sum_{\substack{m=1 \\ \mathrm{s}(m) \neq s}}^{N_{\mathrm{s}}} n_C^{[\mathrm{s}(m)]} \mathrm{RSU}^{[\mathrm{s}(m)]} \right) \Big/ n_C^{[s]}, & \\ \text{otherwise}, & \end{cases} \tag{22}$$

where
(a) $n_C^{[s]}$ is the number of transitions in C from s, and
(b) $\mathcal{D}_{\mathrm{C}}$ is the discrepancy of C;

S.2.2.7 collect the $C \in \mathcal{C}_{\mathrm{B}}$ representative cycles, for which $\mathrm{RD}_C^{[s]} > \phi_{\mathrm{hin}} \mathrm{RSU}^{[s]}$, into the set ρ, where ϕ_{hin} (= 0.01)[III] is the reduction hindrance factor;

S.2.2.8 if $|\rho| = 0$, then continue **S.2.2**, otherwise determine the average and maximum reduced discrepancy of the representative cycles for s:

$$\mathrm{ARD}^{[s]} = |\rho|^{-1} \sum_{C \in \rho} \mathrm{RD}_C^{[s]}, \tag{23}$$

$$\mathrm{MRD}^{[s]} = \max_{C \in \rho} \mathrm{RD}_C^{[s]}, \tag{24}$$

where $|\rho|$ is the cardinality of the set ρ;

S.2.2.9 if $\mathrm{MRD}^{[s]} > \phi_{\mathrm{ter}} \mathrm{RSU}^{[s]}$, where ϕ_{ter} (= 200) is the termination factor, then exit from intMARVEL;[IV]

S.2.2.10 if $n_{\mathrm{nb}}^{[s]} < \lfloor \phi_{\mathrm{rep}} n^{[s]} \rfloor$,[V] then set $\mathrm{RSU}^{[s]} = \mathrm{RSU}_{\mathrm{prev}}^{[s]}$, otherwise use $\mathrm{RSU}^{[s]} = \max(\mathrm{ARD}^{[s]}, \phi_{\mathrm{sep}} \mathrm{ESU}^{[s]})$, where
(a) $\lfloor\ \rfloor$ is the floor operation,
(b) $n_{\mathrm{nb}}^{[s]}$ is the number of non-bridge transitions from s in $\mathcal{N}_{\mathrm{per}}$, and
(c) ϕ_{rep} (= 0.1) and ϕ_{sep} (= 0.2) are the representativity and the separation factors, respectively;

S.2.3 if $r_{\max} < \phi_{\mathrm{conv}}$, then break S.2, where
(a) ϕ_{conv} (= 3) is the convergence factor,
(b) $r_{\max} = \max_{j=1}^{N_{\mathrm{s}}} r_j$, and
(c) $r_j = \max\left(\mathrm{RSU}^{[\mathrm{s}(j)]} / \mathrm{RSU}_{\mathrm{prev}}^{[\mathrm{s}(j)]}, \mathrm{RSU}_{\mathrm{prev}}^{[\mathrm{s}(j)]} / \mathrm{RSU}^{[\mathrm{s}(j)]} \right)$;

S.3 for all $1 \leq j \leq N_{\mathrm{T}}$: in the case that $\mathrm{RSU}^{[\mathrm{s}(\mathrm{i}_j)]} = \mathrm{ESU}^{[\mathrm{s}(\mathrm{i}_j)]}$, apply $\delta_{0,j} = \mathrm{RSU}^{[\mathrm{s}(\mathrm{i}_j)]}$, else set $\delta_{0,j} = \phi_{\mathrm{us}} \mathrm{RSU}^{[\mathrm{s}(\mathrm{i}_j)]}$, where
(a) $\delta_{0,j}$ is the initial uncertainty of the jth transition, and
(b) ϕ_{us} (= 0.5) is the uncertainty scaling factor;

S.4 for all $1 \leq j \leq N_{\mathrm{T}}$: if $P_j = 1$, then set $B_j = \lfloor \log_{10}\left(\mathrm{RSU}^{[\mathrm{s}(\mathrm{i}_j)]} \right) \rfloor$, otherwise use $B_j = 100$,[VI] where B_j is the block index of the jth line;

S.5 set $B_{\min} = \min_{j=1;P_j=1}^{N_{\mathrm{T}}} B_j$ and $B_{\max} = \max_{j=1;P_j=1}^{N_{\mathrm{T}}} B_j$;

S.6 identify the components of $\mathcal{N}_{\mathrm{LS}}$;

S.7 save the core(j)[VII] and comp(k) indices for all $1 \leq j \leq N_{\mathrm{c}}$ and $1 \leq k \leq N_{\mathrm{L}}$, respectively, by means of eqn (4);

S.8 end procedure.

Notes:

[I]The $\mathcal{N}_{\mathrm{per}}$ subnetwork allows the study of the lines arising from s and included in cycles where there is no transition from those s^* segments with $\mathrm{ESU}^{[s^*]} > \mathrm{ESU}^{[s]}$.

[II]The reduced discrepancy of a cycle C related to s represents the inaccuracy of the lines from s in C. Note that nonpositive reduced discrepancies cannot be used for refinement purposes.

[III]In what follows, the default values of the so-called control parameters, needed for the extMARVEL procedure, are mostly given in parentheses (see, *e.g.*, 'ϕ_{hin} (= 0.01)' in stage **S.2.2.7**).

[IV]In this case, the prior cleansing should be continued to decrease the reduced discrepancies.

[V]If there are at least $\lfloor \phi_{\mathrm{rep}} n^{[s]} \rfloor$ lines from s in the cycles of $\mathcal{N}_{\mathrm{per}}$, then its RSU is modified by the corresponding ARD value, else this RSU will remain unchanged in the given iteration.

[VI]If the block index of a transition is 100, this line will not participate in the blockMARVEL refinement (see Appendix B).

[VII]The core indices are determined *via* a traditional MARVEL analysis, using eqn (4). To accelerate this identification process, a logical variable, u_{ST}, can be introduced: if $u_{\mathrm{ST}} = 1$, then only lines related to a BFS-based ST of the LS are employed in eqn (4); otherwise, all the transitions of the LS are utilized in eqn (4). Obviously, $u_{\mathrm{ST}} = 0$ is a more stable choice than $u_{\mathrm{ST}} = 1$.

Appendix B: stages of the blockMARVEL refinement (see Section 2.4.2)

S.1 for all $1 \leq i \leq N_{\mathrm{L}}$: set $b_i = 100$, where b_i is the block index[I] of the ith energy level;

S.2 set $\mathcal{N}_{\mathrm{p}}$ to an empty subnetwork, where $P_j^{\langle \mathcal{N}_{\mathrm{p}} \rangle} = 0$ for all $1 \leq j \leq N_{\mathrm{T}}$;

S.3 assign $N_{\mathrm{p,c}} = N_{\mathrm{L}}$;

S.4 for all $1 \leq i \leq N_{\mathrm{L}}$: set $\mathrm{core_p}(i) = \mathrm{comp_p}(i) = i$ and $\beta_i = 0$;

S.5 for all $1 \leq j \leq N_{\mathrm{T}}$: initialize δ_j as $\delta_{0,j}$;

S.6 for all $B_{\min} \leq \mathscr{B} \leq B_{\max}$:

S.6.1 construct a $\mathcal{N}_{\mathrm{a}}$ subnetwork of $\mathcal{N}_{\mathrm{LS}}$ in the following fashion: if $B_j \leq \mathscr{B}$, then $P_j^{\langle \mathcal{N}_{\mathrm{a}} \rangle} = 1$, otherwise $P_j^{\langle \mathcal{N}_{\mathrm{a}} \rangle} = 0$ for all $1 \leq j \leq N_{\mathrm{T}}$;

S.6.2 if $\mathrm{P}^{\langle \mathcal{N}_{\mathrm{a}} \rangle} = \mathrm{P}^{\langle \mathcal{N}_{\mathrm{p}} \rangle}$, then continue **S.6**;

S.6.3 find the components of $\mathcal{N}_{\mathrm{a}}$;

S.6.4 archive the $\mathrm{comp_a}(1)$, $\mathrm{comp_a}(2)$,...,$\mathrm{comp_a}(N_{\mathrm{L}})$ component indices of the energy levels in $\mathcal{N}_{\mathrm{a}}$;

S.6.5 'infinite' loop:

S.6.5.1 solve eqn (11) and (12)[II] and determine $\bar{\mathbf{E}}$ using eqn (9);

S.6.5.2 for all $1 \leq j \leq N_{\mathrm{T}}$ with $B_j = \mathscr{B}$: calculate Δ_j and d_j;

S.6.5.3 search for $d_{\max} = \max_{j=1;B_j=\mathscr{B}}^{N_{\mathrm{T}}} d_j$;[III]

S.6.5.4 if $d_{\max} \leq \phi_{\text{noi}}$, then break **S.6.5**, where ϕ_{noi} (= 10^{-10}) designates the numerical noise factor;

S.6.5.5 for all $1 \leq j \leq N_{\text{T}}$: if $P_j^{\langle\mathcal{N}_{\text{a}}\rangle} = 1$ and $d_j > \phi_{\text{disc}} d_{\max}$, then substitute δ_j with $\phi_{\text{inc}}|\Delta_j|$, where
(a) ϕ_{disc} (= 0.1) is the discrimination factor, and
(b) ϕ_{inc} (= 1.1) denotes the increase factor;

S.6.5.6 create a $\mathcal{N}_{\text{F}}$ filtered subnetwork with its $\text{P}^{\langle\mathcal{N}_{\text{F}}\rangle}$ participation matrix as follows: if $P_j^{\langle\mathcal{N}_{\text{a}}\rangle} = 1$ and $\delta_j > \min(\phi_{\text{fil}}\text{RSU}^{[\mathfrak{s}(\mathfrak{i}_j)]}, \phi_{\text{co}})$, then set $P_j^{\langle\mathcal{N}_{\text{F}}\rangle} = 1$, otherwise use $P_j^{\langle\mathcal{N}_{\text{F}}\rangle} = 0$ for all $1 \leq j \leq N_{\text{T}}$, where
(a) ϕ_{fil} (= 10) is the filtration factor, and
(b) ϕ_{co} (= 0.05 cm^{-1}) is the cut-off factor;

S.6.5.7 save the indices of the transitions in $\mathcal{N}_{\text{F}}$ into l_1, l_2,…, $l_{N_{\text{F}}}$ in the order[IV,V] that either $\log_{10}(d_{l_i}) > \log_{10}(d_{l_j})$ or $\log_{10}(d_{l_i}) = \log_{10}(d_{l_j})$ and $B_{l_i} \geq B_{l_j}$ holds for all $1 \leq i \leq j \leq N_{\text{F}}$, where N_{F} is the number of lines in $\mathcal{N}_{\text{F}}$ (exclusion sort);

S.6.5.8 for all $1 \leq j \leq N_{\text{T}}$: set $P_j^{\langle\mathcal{N}_{\text{a}}\rangle}$ to $P_j^{\langle\mathcal{N}_{\text{a}}\rangle}\left(1 - P_j^{\langle\mathcal{N}_{\text{F}}\rangle}\right)$;

S.6.5.9 for all $1 \leq i \leq \mathcal{N}_{\text{F}}$:

S.6.5.9.1 identify the components of $\mathcal{N}_{\text{a}}$;

S.6.5.9.2 if the l_ith line connects components in $\mathcal{N}_{\text{a}}$, then reset $P_{l_i}^{\langle\mathcal{N}_{\text{a}}\rangle} = 1$, otherwise assign $P_{l_i} = 0$ and $B_{l_i} = 100$;

S.6.5.10 for all $1 \leq j \leq N_{\text{T}}$: set $P_j^{\langle\mathcal{N}_{\text{F}}\rangle}$ to $\left(1 - P_j^{\langle\mathcal{N}_{\text{a}}\rangle}\right)P_j^{\langle\mathcal{N}_{\text{F}}\rangle}$;

S.6.5.11 if $\text{tr}\left(\mathbf{P}^{\langle\mathcal{N}_{\text{F}}\rangle}\right) > 0$, then set $\delta_j = \delta_{0,j}$ for every $1 \leq j \leq N_{\text{T}}$ with $B_j = \mathscr{B}$, whereby $\text{tr}\left(\mathbf{P}^{\langle\mathcal{N}_{\text{F}}\rangle}\right)$ is the trace of $\mathbf{P}^{\langle\mathcal{N}_{\text{F}}\rangle}$;

S.6.6 for all $1 \leq i \leq N_{\text{L}}$: whenever $b_i = 100$ and $\text{core}_{\text{a}}(\text{comp}_{\text{a}}(i)) = \text{core}(\text{comp}(i))$ together with $\mathcal{S}_{\text{a}}(\text{comp}_{\text{a}}(i)) > 1$, then set the b_i index to $\mathscr{B}$, where
(a) $\text{core}_{\text{a}}(j)$ is the jth core index of $\mathcal{N}_{\text{a}}$, and
(b) $\mathcal{S}_{\text{a}}(j)$ is the jth component size;[VI]

S.6.7 replace $N_{\text{p,c}}$ with $N_{\text{a,c}}$, where $N_{\text{a,c}}$ is the number of components in $\mathcal{N}_{\text{a}}$;

S.6.8 for all $1 \leq i \leq N_{\text{p,c}}$: overwrite $\text{core}_{\text{p}}(i)$ with $\text{core}_{\text{a}}(i)$;

S.6.9 for all $1 \leq i \leq N_{\text{L}}$: set $\text{comp}_{\text{p}}(i) = \text{comp}_{\text{a}}(i)$ and $\beta_i = \bar{E}_i$;

S.7 for all $1 \leq j \leq N_{\text{T}}$ with $P_j = 0$:

S.7.1 recalculate Δ_j and d_j;

S.7.2 set $\delta_j = \max(\delta_{0,j}, \phi_{\text{inc}}|\Delta_j|)$;

S.8 end procedure.

Notes:

[I]At the end of the blockMARVEL process, the b_i block index denotes the $\mathscr{B}$ value when (a) the ith energy level becomes a non-isolated node in $\mathcal{N}_{\text{a}}$, and (b) the component of this energy level contains at least two vertices, among which at least one core of $\mathcal{N}_{\text{LS}}$ can also be found. In the case that b_i remains 100, the ith energy level became an isolated node of $\mathcal{N}_{\text{LS}}$ during the course of prior cleansing.

[II]The core indices of $\mathcal{N}_{\text{a}}$ are set in analogy to eqn (4).

[III]Since the residuals of the lines in $\mathcal{N}_{\text{p}}$ cannot be changed, it is sufficient to restrict $d_{\max}$ to those lines satisfying $B_j = \mathscr{B}$.

[IV]At this point, the values of the defects are the same as in stage **S.6.5.2**.

[V]For convenience, steps **S.6.5.7**–**S.6.5.10** will be referred to as the exclusion procedure and denoted with $\texttt{ex\_proc}(\mathcal{N}_{\text{a}}, \mathcal{N}_{\text{F}})$ in the remaining of this Appendix. During the exclusion procedure, lines are excluded one by one from $\mathcal{N}_{\text{a}}$ so as to avoid increasing the number of components in $\mathcal{N}_{\text{a}}$.

[VI]It is the number of energy levels in the jth component of $\mathcal{N}_{\text{a}}$.

Appendix C: stages of the ABR method (see Section 2.4.3)

S.1 find the bridges of the (whole) SN *via* the BFS method;

S.2 for all $1 \leq i \leq N_{\text{T}}$: if the ith transition is a bridge of the SN, then set $\mathcal{B}_i = 1$, otherwise use $\mathcal{B}_i = 0$;

S.3 for all $1 \leq i \leq N_{\text{R}}$: determine the $\text{ESU}_{\min}^{[\![R(i)]\!]}$ values as

$$\text{ESU}_{\min}^{[\![R(i)]\!]} = \min_{\substack{j=1 \\ R(\Re_j)=R(i)}}^{N_{\mathfrak{s}}} \text{ESU}^{[\mathfrak{s}(j)]}, \tag{25}$$

where $\Re_j$ is the recalibration class index of the $\mathfrak{s}(j)$ segment, for which $R(\Re_j)$ contains $\mathfrak{s}(j)$;

S.4 rearrange the recalibration classes such that $\text{ESU}_{\min}^{[\![R(m_i)]\!]} \leq \text{ESU}_{\min}^{[\![R(m_j)]\!]}$ should be satisfied for all $1 \leq i \leq j \leq N_{\text{R}}$;[I]

S.5 set the $\mathscr{F}$ flag[II] to 1;

S.6 for $1 \leq i \leq N_{\text{R}}$:

S.6.1 assign $\mathcal{R}$ to $R(m_i)$;

S.6.2 set $f^{[\![\mathcal{R}]\!]} = 1$, where $f^{[\![\mathcal{R}]\!]}$ is the recalibration factor for $\mathcal{R}$;

S.6.3 in the case that simple recalibration is not permitted for $\mathcal{R}$, continue **S.6**;[III]

S.6.4 for all $1 \leq j \leq N_{\text{T}}$: if $R(\Re_{\text{i}_j}) = \mathcal{R}$ and $\mathcal{B}_j = 0$, then set $P_j^{\langle\mathcal{N}_{\text{reg}}\rangle} = 1$, otherwise use $P_j^{\langle\mathcal{N}_{\text{reg}}\rangle} = 0$;[IV]

S.6.5 save the indices of the lines in $\mathcal{N}_{\text{reg}}$ into $p_1, p_2, \ldots, p_{N_{\text{reg}}}$;

S.6.6 if $N_{\text{reg}} < \max(\mu_{\text{rec}}, \phi_{\text{cdd}} n^{[\![\mathcal{R}]\!]})$, then continue **S.6**;[III,V], where
(a) $n^{[\![\mathcal{R}]\!]}$ is the number of transitions in the segments of $\mathcal{R}$,
(b) $\mu_{\text{rec}} \geq 2$ is the recalibration margin with the default value of $\mu_{\text{rec}} = 5$, and
(c) ϕ_{cdd}(= 0.2) is the critical data-density factor;

S.6.7 for all $1 \leq j \leq N_{\text{reg}}$: initialize $\delta_{\text{reg},j}$ as $\text{ESU}^{[\mathfrak{s}(\mathfrak{i}_{p_j})]}$;

S.6.8 'infinite' loop:[VI]

S.6.8.1 determine f_{best} by means of eqn (14);

S.6.8.2 for all $1 \leq j \leq N_{\text{reg}}$: calculate the $\Delta_{\text{reg},j}$ residual and the $d_{\text{reg},j} = |\Delta_{\text{reg},j}| - \delta_{\text{reg},j}$ defect;

S.6.8.3 search for $d_{\text{reg,max}} = \max_{j=1}^{N_{\text{reg}}} d_{\text{reg},j}$;

S.6.8.4 if $d_{\text{reg,max}} \leq \phi_{\text{noi}}$, then break **S.6.8**;

S.6.8.5 for all $1 \leq j \leq N_{\text{reg}}$: if $d_{\text{reg},j} > \phi_{\text{rd}} d_{\text{reg,max}}$, where ϕ_{rd} (= 0.1) is the regional discrimination factor, then substitute $\delta_{\text{reg},j}$ with $\phi_{\text{inc}}\delta_{\text{reg},j}$;

S.6.9 specify the recalibrated and non-recalibrated absolute median residuals (AMR) as follows:

$$\mathrm{AMR}_{\mathrm{rec}}^{[\![\mathcal{R}]\!]} = \left|\underset{j=1}{\overset{N_{\mathrm{reg}}}{\mathrm{median}}}\, \Delta_{\mathrm{reg},j}\right|, \tag{26}$$

$$\mathrm{AMR}_{\mathrm{nr}}^{[\![\mathcal{R}]\!]} = \left|\underset{j=1}{\overset{N_{\mathrm{reg}}}{\mathrm{median}}}\, \Delta_{p_j}\right|; \tag{27}$$

S.6.10 whenever $\mathrm{AMR}_{\mathrm{nr}}^{[\![\mathcal{R}]\!]} \leq \max(\phi_{\mathrm{acc}}\mathrm{ESU}_{\mathrm{min}}^{[\![\mathcal{R}]\!]}, \phi_{\mathrm{ass}}\mathrm{AMR}_{\mathrm{rec}}^{[\![\mathcal{R}]\!]})$ holds for $\mathrm{AMR}_{\mathrm{nr}}^{[\![\mathcal{R}]\!]}$, then continue **S.6**,[III,VII] where
(a) ϕ_{acc} (= 0.8) is the acceptability factor, and
(b) ϕ_{ass} (= 3) is the assimilation factor;
S.6.11 reassign $\mathscr{F} = 0$;
S.6.12 reset $f^{[\![\mathcal{R}]\!]} = f_{\mathrm{best}}$;
S.6.13 for $1 \leq j \leq N_{\mathrm{T}}$ with $R(\Re_{\mathrm{i}_j}) = \mathcal{R}$: replace σ_j with $f^{[\![\mathcal{R}]\!]}\sigma_j$;[VIII]
S.6.14 for $1 \leq j \leq N_{\mathrm{reg}}$: exchange δ_{p_j} with $\delta_{\mathrm{reg},j}$;
S.6.15 build an empty $\mathcal{N}_{\mathrm{p}}$ subnetwork in the following form: $P_j^{\langle\mathcal{N}_{\mathrm{p}}\rangle} = 0$ for all $1 \leq j \leq N_{\mathrm{T}}$;[IX]
S.6.16 set $N_{\mathrm{p,c}} = N_{\mathrm{L}}$;
S.6.17 for every $1 \leq j \leq N_{\mathrm{L}}$: assign $b_j = 100$, $\mathrm{core_p}(j) = \mathrm{comp_p}(j) = j$, and $\beta_j = 0$;
S.6.18 for all $B_{\min} \leq \mathscr{B} \leq B_{\max}$:
S.6.18.1 construct a $\mathcal{N}_{\mathrm{a}}$ subnetwork of $\mathcal{N}_{\mathrm{LS}}$ according to the following scheme: if $B_j \leq \mathscr{B}$, then $P_j^{\langle\mathcal{N}_{\mathrm{a}}\rangle} = 1$, else $P_j^{\langle\mathcal{N}_{\mathrm{a}}\rangle} = 0$ for all $1 \leq j \leq N_{\mathrm{T}}$;
S.6.18.2 if $\mathrm{P}^{\langle\mathcal{N}_{\mathrm{a}}\rangle} = \mathrm{P}^{\langle\mathcal{N}_{\mathrm{p}}\rangle}$, then continue **S.6.18**;
S.6.18.3 find the components of the $\mathcal{N}_{\mathrm{a}}$ subnetwork;
S.6.18.4 save the $\mathrm{comp_a}(1)$, $\mathrm{comp_a}(2)$,...,$\mathrm{comp_a}(N_{\mathrm{L}})$ component indices in $\mathcal{N}_{\mathrm{a}}$;
S.6.18.5 yield the solution for eqn (11) and (12);
S.6.18.6 calculate $\bar{\mathbf{E}}$ with the help of eqn (9);
S.6.18.7 for each $1 \leq j \leq N_{\mathrm{L}}$: if $b_j = 100$, $\mathcal{S}_{\mathrm{a}}(\mathrm{comp_a}(j)) > 1$, and $\mathrm{core_a}(\mathrm{comp_a}(j)) = \mathrm{core}(\mathrm{comp}(j))$, then set $b_j = \mathscr{B}$;
S.6.18.8 overwrite $N_{\mathrm{p,c}}$ with $N_{\mathrm{a,c}}$;
S.6.18.9 for all $1 \leq j \leq N_{\mathrm{p,c}}$: exchange $\mathrm{core_p}(j)$ with $\mathrm{core_a}(j)$;
S.6.18.10 for all $1 \leq j \leq N_{\mathrm{L}}$: set $\mathrm{comp_p}(j) = \mathrm{comp_a}(j)$ and $\beta_j = \bar{E}_j$;
S.6.18.11 for all $1 \leq j \leq N_{\mathrm{T}}$: update Δ_j;
S.6.19 for all $1 \leq j \leq N_{\mathrm{T}}$: replace δ_j with $\max(\delta_j, \phi_{\mathrm{inc}}|\Delta_j|)$;
S.6.20 for all $1 \leq j \leq N_{\mathfrak{s}}$ with $R(\Re_j) = \mathcal{R}$:
S.6.20.1 set $s = \mathfrak{s}(j)$;
S.6.20.2 place the indices of the non-bridge transitions in s into $q_1, q_2, \ldots, q_{N^{[s]}}$, where $N^{[s]}$ is the number of non-bridge lines in s;
S.6.20.3 define the recalibated median absolute residual (MAR) for the s segment as follows:

$$\mathrm{MAR}_{\mathrm{rec}}^{[s]} = \underset{k=1}{\overset{N^{[s]}}{\mathrm{median}}}\, |\Delta_{q_k}|; \tag{28}$$

S.6.20.4 if $\phi_{\mathrm{dec}}\mathrm{MAR}_{\mathrm{rec}}^{[s]} < \mathrm{RSU}^{[s]}$, where ϕ_{dec} (= 2) is the declination factor, then[X]
S.6.20.4.1 substitute $\mathrm{RSU}^{[s]}$ with $\max(\phi_{\mathrm{dec}}\mathrm{MAR}_{\mathrm{rec}}^{[s]}, \phi_{\mathrm{sep}}\mathrm{ESU}^{[s]})$;
S.6.20.4.2 for all $1 \leq k \leq N_{\mathrm{T}}$ with $\mathfrak{s}(\mathrm{i}_k) = s$: set $\delta_{0,k} = \phi_{\mathrm{us}}\mathrm{RSU}^{[s]}$ and $B_k = \lfloor \log_{10}(\mathrm{RSU}^{[s]}) \rfloor$;
S.7 end procedure.

Notes:
[I]The recalibration classes are recalibrated in decreasing order of their minimum ESU values to reduce the distortion effects caused by highly uncertain spectral lines on the recalibration factors.
[II]At stage **S.7**, $\mathscr{F}$ denotes whether $\mathcal{N}_{\mathrm{LS}}$ is fully recalibrated, *i.e.*, none of the recalibration classes are recalibrated during the ABR procedure.
[III]These empirical conditions, namely steps **S.6.3**, **S.6.6**, and **S.6.10**, have to be violated so that the $\mathcal{R}$ class can be recalibrated with the ABR algorithm.
[IV]The bridges of the SN are not included in $\mathcal{N}_{\mathrm{reg}}$ because they are reproduced with zero residuals.
[V]If this condition is true, then there are too few lines in $\mathcal{N}_{\mathrm{reg}}$ for the safe recalibration of $\mathcal{R}$.
[VI]The f_{best} value is obtained from an iterative refinement scheme similar to that presented in Appendix B.
[VII]If the $\mathcal{R}$ recalibration class is well calibrated, the condition of step **S.6.10** must be satisfied.
[VIII]If $\mathcal{R}$ is recalibrated, then the wavenumbers of its lines should be multiplied with $f^{[\![\mathcal{R}]\!]}$.
[IX]The constrained energy levels are recalculated in an analogous way as described in Appendix B, the only difference is that the wavenumber uncertainties are not refined here. In what follows, stages **S.6.15–S.6.18** are designated with $\texttt{fix\_bm}(\phi_{\mathrm{rd}})$ and are referred to as the fixed blockMARVEL procedure.
[X]If necessary, the RSU values, the initial line uncertainties, and the block indices associated to the recalibrated $\mathcal{R}$ class should be modified to properly describe the accuracy of the improved wavenumbers.

Appendix D: stages of the local synchronization method (see Section 2.4.4)

S.1 set flag $\mathscr{P}$ to 1;[I]
S.2 store the indices of the energy levels in $t_1, t_2, \ldots, t_{N_{\mathrm{L}}}$ in the order that either $\mathrm{comp}(t_i) < \mathrm{comp}(t_j)$ or $\mathrm{comp}(t_i) = \mathrm{comp}(t_j)$ and $\bar{E}_{t_i} \leq \bar{E}_{t_j}$ is valid for all $1 \leq i \leq j \leq N_{\mathrm{L}}$;
S.3 for all $1 \leq i \leq N_{\mathrm{L}}$:
S.3.1 set $T = t_i$;
S.3.2 if $b_T = 100$ or $\mathrm{core}(\mathrm{comp}(T)) = T$, then continue **S.3**;[II]
S.3.3 for all $1 \leq j \leq N_{\mathrm{T}}$: if $T \in \chi_j$ and $\mathcal{B}_j = 0$, then set $P_j^{\langle\mathcal{N}_{\mathrm{loc}}\rangle} = 1$, otherwise use $P_j^{\langle\mathcal{N}_{\mathrm{loc}}\rangle} = 0$,[III] where $\chi_j = \{\mathrm{up}(j), \mathrm{low}(j)\}$;
S.3.4 place the indices of the lines in $\mathcal{N}_{\mathrm{loc}}$ into $u_1, u_2, \ldots, u_{N_{\mathrm{loc}}}$;

S.3.5 if $N_{\mathrm{loc}} < \mu_{\mathrm{sync}}$, where μ_{sync} (= 4) is the synchronization margin, then continue **S.3**;[II,IV]

S.3.6 for all $1 \le j \le N_{\mathrm{loc}}$: calculate e_j by means of eqn (16);

S.3.7 search for the minimum ESU value in $\mathcal{N}_{\mathrm{loc}}$:

$$\mathrm{ESU}_{\mathrm{loc,min}} = \min_{j=1}^{N_{\mathrm{loc}}} \mathrm{ESU}\left[\mathrm{s}\left(\mathrm{i}_{u_j}\right)\right]; \tag{29}$$

S.3.8 for all $1 \le j \le N_{\mathrm{loc}}$: initialize $\delta_{\mathrm{loc},j}$ as $\mathrm{ESU}_{\mathrm{loc,min}}$;

S.3.9 'infinite' loop:[V]

S.3.9.1 calculate e_{best} with the help of eqn (17);

S.3.9.2 for all $1 \le j \le N_{\mathrm{loc}}$: determine the $\Delta_{\mathrm{loc},j}$ residual and the $d_{\mathrm{loc},j} = |\Delta_{\mathrm{loc},j}| - \delta_{\mathrm{loc},j}$ defect;

S.3.9.3 seek for $d_{\mathrm{loc,max}} = \max_{j=1}^{N_{\mathrm{loc}}} d_{\mathrm{loc},j}$;

S.3.9.4 if $d_{\mathrm{loc,max}} \le \phi_{\mathrm{noi}}$, then break **S.3.9**;

S.3.9.5 for every $1 \le j \le N_{\mathrm{loc}}$: if $d_{\mathrm{loc},j} > \phi_{\mathrm{disc}} d_{\mathrm{loc,max}}$, then replace $\delta_{\mathrm{loc},j}$ with $\phi_{\mathrm{inc}}\delta_{\mathrm{loc},j}$;

S.3.10 define the adjusted and non-adjusted MAR values of the Tth energy level in the following form:

$$\mathrm{MAR}_{\mathrm{adj}}^{(T)} = \operatorname*{median}_{j=1}^{N_{\mathrm{loc}}} |\Delta_{\mathrm{loc},j}|, \tag{30}$$

$$\mathrm{MAR}_{\mathrm{na}}^{(T)} = \operatorname*{median}_{j=1}^{N_{\mathrm{loc}}} |\Delta_{u_j}|; \tag{31}$$

S.3.11 if $\mathrm{MAR}_{\mathrm{na}}^{(T)} \le \phi_{\mathrm{sync}}\mathrm{MAR}_{\mathrm{adj}}^{(T)}$, where ϕ_{sync} (= 4) is the synchronization factor, then continue **S.3**;[II,VI]

S.3.12 reset $\mathscr{P} = 0$;

S.3.13 exchange $\bar{E}_T$ with e_{best};

S.4 if $\mathscr{P} = 0$, then:

S.4.1 initialize $\mathcal{N}_{\mathrm{F}}$ as $P_j^{\langle\mathcal{N}_{\mathrm{F}}\rangle} = 0$ for all $1 \le j \le N_{\mathrm{T}}$;

S.4.2 for all $1 \le j \le N_{\mathrm{T}}$:

S.4.2.1 update Δ_j and d_j;

S.4.2.2 if $P_j = 0$, then:

S.4.2.2.1 reset $\delta_j = \max(\delta_{0,j},\phi_{\mathrm{inc}}|\Delta_j|)$;

S.4.2.2.2 if $\delta_j \le \min(\phi_{\mathrm{fil}}\mathrm{RSU}^{[\mathrm{s}(\mathrm{i}_j)]}, \phi_{\mathrm{co}})$, then[VII]

S.4.2.2.2.1 set $P_j = 1$;

S.4.2.2.2.2 assign $B_j = \left\lfloor \log_{10}\left(\mathrm{RSU}^{[\mathrm{s}(\mathrm{i}_j)]}\right)\right\rfloor$;

S.4.2.3 if $P_j = 1$ and $d_j > 0$, then reassign $P_j^{\langle\mathcal{N}_{\mathrm{F}}\rangle} = 1$;[VIII]

S.4.2.4 reset $\delta_j = \max(\delta_j,\phi_{\mathrm{inc}}|\Delta_j|)$;

S.4.3 `ex_proc`$(\mathcal{N}_{\mathrm{LS}},\mathcal{N}_{\mathrm{F}})$;

S.4.4 `fix_bm`(ϕ_{ld}) where ϕ_{ld} (= 0.9) is the local discrimination factor;

S.4.5 for all $1 \le j \le N_{\mathrm{T}}$: replace δ_j with $\max(\delta_j,\phi_{\mathrm{inc}}|\Delta_j|)$;

S.5 end procedure.

Notes:

[I] Upon completion of the local synchronization procedure, $\mathscr{P}$ will designate whether all the constrained empirical energy levels are supported by this algorithm ($\mathscr{P} = 1$) or there are energy values adjusted during the synchronization process ($\mathscr{P} = 0$).

[II] If any of the conditions **S.3.2**, **S.3.5**, and **S.3.11** are not met, the synchronization of the Tth energy level with its $\mathcal{N}_{\mathrm{loc}}$ subnetwork is ignored.

[III] The $\mathcal{N}_{\mathrm{loc}}$ subnetwork contains all the non-bridge transitions of the whole SN which are included in the CD relations of the Mth energy level. Bridges are not inserted into $\mathcal{N}_{\mathrm{loc}}$ for the same reason detailed in Note IV of Appendix C.

[IV] μ_{sync} provides a reasonable lower limit for N_{loc}.

[V] The e_{best} value is determined by the aid of an iterative refinement technique (see also Appendices B and C).

[VI] If the $\bar{E}_{\mathrm{T}}$ value conforms with their trial energies, then the inequality of stage **S.3.11** has to be fulfilled. In the opposite case, e_{best} is used instead of $\bar{E}_{\mathrm{T}}$ in the rest of the local synchronization process.

[VII] In the case that certain constrained empirical energy levels were altered during the local synchronization, there may be transitions which can be reincluded in $\mathcal{N}_{\mathrm{LS}}$.

[VIII] If there are energy levels adjusted by the local synchronization procedure, then there must be lines of positive (updated) defects, for which the `ex_proc` algorithm needs to be called.

Appendix E: stages of the FNC method (see Section 2.4.5)

S.1 for all $1 \le i \le N_{\mathrm{T}}$: if $P_i = 0$ and $\delta_i \le \min\left(\phi_{\mathrm{rei}}\mathrm{RSU}^{[\mathrm{s}(\mathrm{i}_i)]}, \phi_{\mathrm{co}}\right)$, then reset $P_i = 1$,[I] where ϕ_{rei}(= 100) is the reinclusion factor;

S.2 detect the bridges of $\mathcal{N}_{\mathrm{LS}}$ *via* the BFS method;

S.3 for all $1 \le i \le N_{\mathrm{T}}$: if $P_i = 1$ and the ith transition is a bridge of $\mathcal{N}_{\mathrm{LS}}$, then set $\mathcal{B}_{\mathrm{LS},i} = 1$, otherwise use $\mathcal{B}_{\mathrm{LS},i} = 0$;

S.4 build the maximum bridgeless subnetwork (MBS)[II] of $\mathcal{N}_{\mathrm{LS}}$, denoted with $\mathcal{N}_{\mathrm{B}}$, as follows: $P_i^{\langle\mathcal{N}_{\mathrm{B}}\rangle} = P_i\left(1 - \mathcal{B}_{\mathrm{LS},i}\right)$ for all $1 \le i \le N_{\mathrm{T}}$;

S.5 identify the components of $\mathcal{N}_{\mathrm{B}}$ (*i.e.*, the bridge components of $\mathcal{N}_{\mathrm{LS}}$);

S.6 archive the $\mathrm{comp}_{\mathrm{B}}(1)$, $\mathrm{comp}_{\mathrm{B}}(2)$,..., $\mathrm{comp}_{\mathrm{B}}(N_{\mathrm{L}})$ indices of $\mathcal{N}_{\mathrm{B}}$;

S.7 save the $\mathcal{S}_{\mathrm{B}}(1), \mathcal{S}_{\mathrm{B}}(2), \ldots, \mathcal{S}_{\mathrm{B}}(N_{\mathrm{B,c}})$ component sizes of N_{B}, where $N_{\mathrm{B,c}}$ is the number of components in $\mathcal{N}_{\mathrm{B}}$;

S.8 for all $1 \le i \le N_{\mathrm{L}}$:

S.8.1 specify the $\mathscr{R}_i$ resistance[III] of the ith energy level as

$$\mathscr{R}_i = \begin{cases} \mathrm{P}, & \text{if } \mathrm{comp}_{\mathrm{B}}(i) = \mathrm{comp}_{\mathrm{B}}(\mathrm{core}(\mathrm{comp}(i)), \\ \mathrm{U}, & \text{if } \mathcal{S}_{\mathrm{B}}(\mathrm{comp}_{\mathrm{B}}(i)) = 1, \\ \mathrm{S}, & \text{otherwise;} \end{cases} \tag{32}$$

S.8.2 define the following 'connectivity' parameters,[IV]

$$\begin{aligned} \mathrm{LTD}_i &= \left|\left\{j\colon 1 \le j \le N_{\mathrm{T}}, P_j = 1, \text{ and } i \in \chi_j\right\}\right|, \\ \mathrm{BTD}_i &= \left|\left\{j\colon 1 \le j \le N_{\mathrm{T}}, B_j \le b_i, \text{ and } i \in \chi_j\right\}\right|, \\ \mathrm{LSD}_i &= \left|\left\{\mathcal{S}(j)\colon 1 \le j \le N_{\mathrm{T}}, P_j = 1, \text{ and } i \in \chi_j\right\}\right|, \\ \mathrm{BSD}_i &= \left|\left\{\mathcal{S}(j)\colon 1 \le j \le N_{\mathrm{T}}, B_j \le b_i, \text{ and } i \in \chi_j\right\}\right|, \end{aligned} \tag{33}$$

where χ_j is defined in stage **S.3.3** of Appendix D, and the ith rovibrational energy level is associated with the LTD_i leading transition degree, the BTD_i block transition degree, the LSD_i leading source degree, and the BSD_i block source degree;

S.8.3 assign the dependability grade[V] to the ith energy level as

$$\gamma_i = \begin{cases} \mathrm{A}^+, \text{ if } b_i \neq 100, \mathscr{R}_i = \mathrm{P}, \mathrm{LSD}_i \geq \mathrm{LSD_c}, \text{and } \mathrm{LTD}_i \geq \mathrm{LTD_c}, \\ \mathrm{A}^-, \text{ if } b_i \neq 100, \mathscr{R}_i = \mathrm{P}, \mathrm{LSD}_i \geq \mathrm{LSD_c}, \text{and } \mathrm{LTD}_i < \mathrm{LTD_c}, \\ \mathrm{B}^+, \text{ if } b_i \neq 100, \mathscr{R}_i = \mathrm{P}, \mathrm{LSD}_i < \mathrm{LSD_c}, \text{and } \mathrm{LTD}_i \geq \mathrm{LTD_c}, \\ \mathrm{B}^-, \text{ if } b_i \neq 100, \mathscr{R}_i = \mathrm{P}, \mathrm{LSD}_i < \mathrm{LSD_c}, \text{and } \mathrm{LTD}_i < \mathrm{LTD_c}, \\ \mathrm{C}^+, \text{ if } b_i \neq 100, \mathscr{R}_i = \mathrm{S}, \\ \mathrm{C}^-, \text{ if } b_i \neq 100, \mathscr{R}_i = \mathrm{U}, \\ \mathrm{D}, \quad \text{if } b_i = 100, \end{cases} \tag{34}$$

where $\mathrm{LTD_c}(= 10)$ and $\mathrm{LSD_c}(= 6)$ are the critical LTD and LSD, respectively;

S.8.4 if core(comp(i)) = i, then assign $\varepsilon_i = 0$[VI] and continue **S.8**;

S.8.5 construct a $\mathcal{N}_{\mathrm{frac}}$ fractional subnetwork[VII] of $\mathcal{N}_{\mathrm{LS}}$ in the following form: if $P_j = 1$, $\mathcal{B}_{\mathrm{LS},j} = 0$, and $i \in \chi_j$, then set $P_j^{\langle \mathcal{N}_{\mathrm{frac}} \rangle} = 1$, otherwise use $P_j^{\langle \mathcal{N}_{\mathrm{frac}} \rangle} = 0$ for all $1 \leq j \leq N_{\mathrm{T}}$;

S.8.6 save the indices of the transitions in $\mathcal{N}_{\mathrm{frac}}$ into $\nu_1, \nu_2, \ldots, \nu_{N_{\mathrm{frac}}}$ such that $\delta_{\nu_j} \leq \delta_{\nu_k}$ holds for all $1 \leq j \leq k \leq N_{\mathrm{frac}}$, where N_{frac} is the number of lines in $\mathcal{N}_{\mathrm{frac}}$;

S.8.7 if $N_{\mathrm{frac}} < n_{\mathrm{cct}}$ or $|S^{(N_{\mathrm{frac}})}| < n_{\mathrm{ccs}}$, then set $\varepsilon_i = -1$[VIII] and continue **S.8**, where

(a) $S^{(N_{\mathrm{frac}})} = \{\mathcal{S}(\mathscr{I}_{\nu_j}) : 1 \leq j \leq N_{\mathrm{frac}}\}$ is the set of sources including transitions with the $\nu_1, \nu_2, \ldots, \nu_{N_{\mathrm{frac}}}$ indices,

(b) $\mathcal{S}(l)$ is the lth source of the SN,

(c) $\mathscr{I}_m$ is the mth source index for which $\mathcal{S}(\mathscr{I}_m)$ contains the mth transition,

(d) $n_{\mathrm{cct}}(= 5)$ is the critical number of confirmative lines, and

(e) $n_{\mathrm{ccs}}(= 3)$ is the critical number of confirmative sources;

S.8.8 specify the minimally required number of confirmative transitions, n_{mrct}, in this way:[IX]

$$n_{\mathrm{mrct}} = \min\{k: n_{\mathrm{cct}} \leq k \leq N_{\mathrm{frac}} \text{ and } |S^{(k)}| \geq n_{\mathrm{ccs}}\}; \tag{35}$$

S.8.9 introduce the number of confirmative transitions,[X]

$$n_{\mathrm{ct}} = \max\left\{k: n_{\mathrm{mrct}} \leq k \leq N_{\mathrm{frac}} \text{ and } \delta_{\nu_k} \leq \phi_{\mathrm{fe}} \delta_{\nu_{n_{\mathrm{mrct}}}}\right\}, \tag{36}$$

where $\phi_{\mathrm{fe}}(= 3)$ is the fractional extension factor;

S.8.10 estimate the ε_i uncertainty as[XI]

$$\varepsilon_i = \sqrt{\frac{1}{n_{\mathrm{ct}}} \sum_{k=1}^{n_{\mathrm{ct}}} \delta_{\nu_k}^2}; \tag{37}$$

S.9 for all $1 \leq i \leq N_{\mathfrak{s}}$:

S.9.1 set $s = \mathfrak{s}(i)$;

S.9.2 provide the $(A^{[s]}, V^{[s]}, E^{[s]})$ triplet[XII] for the s segment as

$$\begin{aligned} A^{[s]} &= \left|\left\{j: 1 \leq j \leq N_{\mathrm{T}} \text{ and } \mathfrak{s}(\mathrm{i}_j) = s\right\}\right|, \\ V^{[s]} &= \left|\left\{j: 1 \leq j \leq N_{\mathrm{T}}, P_j = 1, \text{ and } \mathfrak{s}(\mathrm{i}_j) = s\right\}\right|, \\ E^{[s]} &= \left|\left\{j: 1 \leq j \leq N_{\mathrm{T}}, P_j = 1, B_j \leq b_j^*, \text{ and } \mathfrak{s}(\mathrm{i}_j) = s\right\}\right|, \end{aligned} \tag{38}$$

where

(a) $b_j^* = \max(b_{\mathrm{up}(j)}, b_{\mathrm{low}(j)})$, and

(b) $A^{[s]}$, $V^{[s]}$, and $E^{[s]}$ symbolize the number of assigned, validated, and exploited transitions corresponding to the s segment, respectively;

S.9.3 define the following statistical quantities:[XIII]

$$\begin{aligned} \sigma_{\min}^{[s]} &= \min_{\substack{j=1 \\ \mathfrak{s}(\mathrm{i}_j)=s}}^{N_{\mathrm{T}}} \sigma_j, \\ \sigma_{\max}^{[s]} &= \max_{\substack{j=1 \\ \mathfrak{s}(\mathrm{i}_j)=s}}^{N_{\mathrm{T}}} \sigma_j, \\ \mathrm{ASU}^{[s]} &= \frac{1}{V^{[s]}} \sum_{\substack{j=1 \\ \mathfrak{s}(\mathrm{i}_j)=s}}^{N_{\mathrm{T}}} P_j \delta_j, \\ \mathrm{MSU}^{[s]} &= \max_{\substack{j=1 \\ \mathfrak{s}(\mathrm{i}_j)=s \\ P_j=1}}^{N_{\mathrm{T}}} \delta_j, \end{aligned} \tag{39}$$

where $\mathrm{ASU}^{[s]}$ and $\mathrm{MSU}^{[s]}$ are the average and maximum segment uncertainties of the s segment, respectively;

S.10 for all $1 \leq i \leq N_{\mathrm{T}}$: represent the dependability grade of the ith transition, Γ_i, with the lower[V] grade of $\gamma_{\mathrm{up}(i)}$ and $\gamma_{\mathrm{low}(i)}$;

S.11 end procedure.

Notes:

[I] Although the reincluded lines may have relatively large uncertainties, they could be useful for corroborating the dependability of the energy levels.

[II] $\mathcal{N}_{\mathrm{B}}$ is built by placing the non-bridge lines of $\mathcal{N}_{\mathrm{LS}}$ into this subnetwork.

[III] The resistance of the ith rovibrational energy level is (a) protected (P), if it lies in the same bridge component as the core of its component in $\mathcal{N}_{\mathrm{LS}}$, (b) unprotected (U), if it is alone in its bridge component, or (c) semiprotected (S), otherwise.

[IV] These quantities can be interpreted in the following way: (a) LTD_i is the total number of lines in $\mathcal{N}_{\mathrm{LS}}$ incident to the ith energy level, (b) BTD_i is the number of exploited lines from $\mathcal{N}_{\mathrm{LS}}$ including this rovibrational state, and (c) LSD_i and BSD_i correspond to the numbers of sources containing transitions used in the definitions of LTD_i and BTD_i, respectively. The jth transition of the SN is exploited if $B_j \leq \max(b_{\mathrm{up}(j)}, b_{\mathrm{low}(j)})$, that is, if this line is utilized during the determination of its upper or lower energy level.

[V] The particular grades are listed starting from the best grade down to the worst one. In fact, energy levels with grade D are undefined, as there is no line in $\mathcal{N}_{\mathrm{LS}}$ incident to these energy levels.

[VI] The uncertainties of the cores are set to zero.

[VII] $\mathcal{N}_{\mathrm{frac}}$ contains all the non-bridge lines from $\mathcal{N}_{\mathrm{LS}}$ which are incident to the ith energy level.

[VIII] $\varepsilon_i = -1$ indicates that a reliable uncertainty cannot be assigned to the ith energy level.

[IX] $n_{\mathrm{mrct}} \geq n_{\mathrm{cct}}$ is the smallest positive integer for which lines indexed with $\nu_1, \nu_2, \ldots, \nu_{n_{\mathrm{mrct}}}$ arise from at least n_{ccs} sources.

[X] As a result of this stage, a collection of $n_{\mathrm{ct}} \geq n_{\mathrm{mrct}}$ transitions, called confirmative linelist (CL), is obtained, whose lines are suitable for the estimation of the uncertainty of the ith energy level.

[xi]If $n_{\mathrm{mrct}} > 0$ and the ith energy level is not a core, then ε_i is calculated *via* eqn (37).

[xii]This triplet can be denoted with '*A*/*V*/*E*', as well.

[xiii]The $[\sigma_{\min}^{[s]}, \sigma_{\max}^{[s]}]$ interval represents the measurement range of the wavenumbers in the *s* segment, while the parameters $\mathrm{ASU}^{[s]}$ and $\mathrm{MSU}^{[s]}$ describe the estimated accuracy of *s* at the end of the intMARVEL procedure.

Acknowledgements

The authors thank Dr Csaba Fábri for fruitful discussions. The Budapest group gratefully acknowledges the financial support they received from NKFIH (grant number K119658). The Budapest group also received support from the grant VEKOP-2.3.2-16-2017-00014, supported by the European Union and the State of Hungary and co-financed by the European Regional Development Fund. The collaboration between ELTE and UCL was supported by the CM1405 COST action, MOLIM: Molecules in Motion.

References

1 J. L. Hall, Nobel Lecture: defining and measuring optical frequencies, *Rev. Mod. Phys.*, 2006, **78**, 1279–1295.
2 T. W. Hänsch, Nobel Lecture: passion for precision, *Rev. Mod. Phys.*, 2006, **78**, 1297–1309.
3 M. J. Thorpe, D. Balslev-Clausen, M. S. Kirchner and J. Ye, Cavity-enhanced optical frequency comb spectroscopy: application to human breath analysis, *Opt. Express*, 2008, **16**, 2387–2397.
4 A. Foltynowicz, P. Maslowski, A. J. Fleisher, B. J. Bork and J. Ye, Cavity-enhanced optical frequency comb spectroscopy in the mid-infrared application to trace detection of hydrogen peroxide, *Appl. Phys. B: Lasers Opt.*, 2013, **110**, 163–175.
5 C. A. Alrahman, A. Khodabakhsh, F. M. Schmidt, Z. Qu and A. Foltynowicz, Cavity-enhanced optical frequency comb spectroscopy of high-temperature H_2O in a flame, *Opt. Express*, 2014, **22**, 13889–13895.
6 D. Z. Kandula, C. Gohle, T. J. Pinkert, W. Ubachs and S. E. Eikema, Extreme ultraviolet frequency comb metrology, *Phys. Rev. Lett.*, 2010, **105**, 063001.
7 D. Z. Kandula, C. Gohle, T. J. Pinkert, W. Ubachs and S. E. Eikema, XUV frequency comb metrology on the ground state of helium, *Phys. Rev. A: At., Mol., Opt. Phys.*, 2011, **84**, 062512.
8 A. G. Császár, G. Czakó, T. Furtenbacher and E. Mátyus, An active database approach to complete spectra of small molecules, *Annu. Rep. Comput. Chem.*, 2007, **3**, 155–176.
9 A. G. Császár and T. Furtenbacher, Spectroscopic networks, *J. Mol. Spectrosc.*, 2011, **266**, 99–103.
10 A. G. Császár, T. Furtenbacher and P. Árendás, Small molecules-Big data, *J. Phys. Chem. A*, 2016, **120**, 8949–8969.
11 T. Furtenbacher, A. G. Császár and J. Tennyson, MARVEL: measured active rotational–vibrational energy levels, *J. Mol. Spectrosc.*, 2007, **245**, 115–125.
12 T. Furtenbacher and A. G. Császár, MARVEL: measured active rotational-vibrational energy levels. II. Algorithmic improvements, *J. Quant. Spectrosc. Radiat. Transfer*, 2012, **113**, 929–935.
13 R. Tóbiás, T. Furtenbacher and A. G. Császár, Cycle bases to the rescue, *J. Quant. Spectrosc. Radiat. Transfer*, 2017, **203**, 557–564.
14 E. Czinki, T. Furtenbacher, A. G. Császár, A. K. Eckhardt and G. C. Mellau, The 1943 K emission spectrum of $H_2{}^{16}O$ between 6600 and 7050 cm^{-1}, *J. Quant. Spectrosc. Radiat. Transfer*, 2018, **206**, 46–54.
15 T. Furtenbacher, I. Szabó, A. G. Császár, P. F. Bernath, S. N. Yurchenko and J. Tennyson, Experimental energy levels and partition function of the $^{12}C_2$ molecule, *Astrophys. J., Suppl. Ser.*, 2016, **224**, 44.
16 J. Tennyson, P. F. Bernath, L. R. Brown, A. Campargue, M. R. Carleer, A. G. Császár, R. R. Gamache, J. T. Hodges, A. Jenouvrier, O. V. Naumenko, O. L. Polyansky, L. S. Rothman, R. A. Toth, A. C. Vandaele, N. F. Zobov, L. Daumont, A. Z. Fazliev, T. Furtenbacher, I. F. Gordon, S. N. Mikhailenko and S. V. Shirin, Critical evaluation of the rotational–vibrational spectra of water vapor. Part I. Energy levels and transition wavenumbers for $H_2{}^{17}O$ and $H_2{}^{18}O$, *J. Quant. Spectrosc. Radiat. Transfer*, 2009, **110**, 573–596.
17 J. Tennyson, P. F. Bernath, L. R. Brown, A. Campargue, M. R. Carleer, A. G. Császár, R. R. Gamache, J. T. Hodges, A. Jenouvrier, O. V. Naumenko, O. L. Polyansky, L. S. Rothman, R. A. Toth, A. C. Vandaele, N. F. Zobov, A. Z. Fazliev, T. Furtenbacher, I. F. Gordon, S.-M. Hu, S. N. Mikhailenko and B. Voronin, Critical evaluation of the rotational–vibrational spectra of water vapor. Part II. Energy levels and transition wavenumbers for $HD^{16}O$, $HD^{17}O$, and $HD^{18}O$, *J. Quant. Spectrosc. Radiat. Transfer*, 2010, **110**, 2160–2184.
18 J. Tennyson, P. F. Bernath, L. R. Brown, A. Campargue, 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, IUPAC critical evaluation of the rotational–vibrational spectra of water vapor. Part III: energy levels and transition wavenumbers for $H_2{}^{16}O$, *J. Quant. Spectrosc. Radiat. Transfer*, 2013, **117**, 29–58.
19 J. Tennyson, P. F. Bernath, L. R. Brown, A. Campargue, A. G. Császár, L. Daumont, R. R. Gamache, J. T. Hodges, O. V. Naumenko, O. L. Polyansky, L. S. Rothmam, A. C. Vandaele, N. F. Zobov, N. Dénes, A. Z. Fazliev, T. Furtenbacher, I. E. Gordon, S.-M. Hu, T. Szidarovszky and I. A. Vasilenko, IUPAC critical evaluation of the rotational–vibrational spectra of water vapor. Part IV. Energy levels and transition wavenumbers for $D_2{}^{16}O$, $D_2{}^{17}O$ and $D_2{}^{18}O$, *J. Quant. Spectrosc. Radiat. Transfer*, 2014, **142**, 93–108.
20 J. Tennyson, P. F. Bernath, L. R. Brown, A. Campargue, A. G. Császár, L. Daumont, R. R. Gamache, J. T. Hodges, O. V. Naumenko, O. L. Polyansky, L. S. Rothman,

A. C. Vandaele and N. F. Zobov, A Database of Water Transitions from Experiment and Theory (IUPAC Technical Report), *Pure Appl. Chem.*, 2014, **86**, 71–83.

21 R. Tóbiás, T. Furtenbacher, A. G. Császár, O. V. Naumenko, J. Tennyson, J.-M. Flaud, P. Kumar and B. Poirier, Critical evaluation of measured rotational–vibrational transitions of four sulphur isotopologues of $S^{16}O_2$, *J. Quant. Spectrosc. Radiat. Transfer*, 2018, **208**, 152–163.

22 T. Furtenbacher, T. Szidarovszky, E. Mátyus, C. Fábri and A. G. Császár, Analysis of the rotational–vibrational states of the molecular ion H_3^+, *J. Chem. Theory Comput.*, 2013, **9**, 5471–5478.

23 T. Furtenbacher, T. Szidarovszky, C. Fábri and A. G. Császár, MARVEL Analysis of the rotational–vibrational states of the molecular ions H_2D^+ and D_2H^+, *Phys. Chem. Chem. Phys.*, 2013, **15**, 10181–10193.

24 A. R. Al Derzi, T. Furtenbacher, J. Tennyson, S. N. Yurchenko and A. G. Császár, MARVEL analysis of the measured high-resolution spectra of $^{14}NH_3$, *J. Quant. Spectrosc. Radiat. Transfer*, 2015, **161**, 117–130.

25 C. Fábri, E. Mátyus, T. Furtenbacher, L. Nemes, B. Mihály, T. Zoltáni and A. G. Császár, Variational quantum mechanical and active database approaches to the rotational–vibrational spectroscopy of ketene, *J. Chem. Phys.*, 2011, **135**, 094307.

26 C. Sousa-Silva, L. K. McKemmish, K. L. Chubb, J. Baker, E. J. Barton, M. N. Gorman, T. Rivlin and J. Tennyson, Original Research By Young Twinkle Students (ORBYTS): when can students start performing original research?, *Phys. Educ.*, 2018, **53**, 015020.

27 L. K. McKemmish, T. Masseron, S. Sheppard, E. Sandeman, Z. Schofield, T. Furtenbacher, A. G. Császár, J. Tennyson and C. Sousa-Silva, MARVEL analysis of the measured high-resolution spectra of $^{48}Ti^{16}O$, *Astrophys. J., Suppl. Ser.*, 2017, **228**, 15.

28 L. K. McKemmish, K. Goodhew, S. Sheppard, A. Bennet, A. Martin, A. Singh, C. Sturgeon, R. Godden, T. Furtenbacher, A. G. Császár and J. Tennyson, MARVEL analysis of the measured high-resolution spectra of $^{90}Zr^{16}O$, *Astrophys. J.*, 2018, **867**, 33.

29 K. L. Chubb, O. V. Naumenko, S. Keely, S. Bartolotto, S. MacDonald, M. Mukhtar, A. Grachov, J. White, E. Coleman, A. Liu, A. Z. Fazliev, E. R. Polovtseva, V. M. Horneman, A. Campargue, T. Furtenbacher, A. G. Császár, S. N. Yurchenko and J. Tennyson, MARVEL analysis of the measured high-resolution rovibrational spectra of H_2S, *J. Quant. Spectrosc. Radiat. Transfer*, 2018, **218**, 178–186.

30 K. L. Chubb, M. Joseph, J. Franklin, N. Choudhury, T. Furtenbacher, A. G. Császár, G. Gaspard, P. Oguoko, A. Kelly, S. N. Yurchenko, J. Tennyson and C. Sousa-Silva, MARVEL analysis of the measured high-resolution spectra of C_2H_2, *J. Quant. Spectrosc. Radiat. Transfer*, 2018, **204**, 42–55.

31 I. Gordon, L. Rothman, C. Hill, R. V. Kochanov, Y. Tan, P. Bernath, M. Birk, V. Boudon, A. Campargue, K. V. Chance, B. Drouin, J.-M. Flaud, D. Gamache, R. R. Jacquemart, V. I. Perevalov, A. Perrin, M. A. H. Smith, J. Tennyson, H. Tran, V. G. Tyuterev, G. C. Toon, J. T. Hodges, K. P. Shine, A. Barbe, A. G. Császár, M. V. Devi, T. Furtenbacher, J. J. Harrison, A. Jolly, T. Johnson, T. Karman, I. Kleiner, A. Kyuberis, J. Loos, O. M. Lyulin, S. N. Mikhailenko, N. Moazzen-Ahmadi, H. S. P. Müller, O. V. Naumenko, A. V. Nikitin, O. L. Polyansky, M. Rey, M. Rotger, S. Sharpe, E. Starikova, S. A. Tashkun, J. Vander Auwera, G. Wagner, J. Wilzewski, P. Wcislo, S. Yu and E. Zak, The HITRAN2016 molecular spectroscopic database, *J. Quant. Spectrosc. Radiat. Transfer*, 2017, **203**, 3–69.

32 R. R. Gamache, C. Roller, E. Lopes, I. E. Gordon, L. S. Rothman, O. L. Polyansky, N. F. Zobov, A. A. Kyuberis, J. Tennyson, S. N. Yurchenko, A. G. Császár, T. Furtenbacher, X. Huang, D. W. Schwenke, T. J. Lee, B. J. Drouin, S. A. Tashkun, V. I. Perevalov and R. V. Kochanov, Total Internal Partition Sums for 167 isotopologues of 53 molecules important in planetary atmospheres: application to HITRAN2016 and beyond, *J. Quant. Spectrosc. Radiat. Transfer*, 2017, **203**, 70–87.

33 J. Tennyson and S. N. Yurchenko, ExoMol: molecular line lists for exoplanet and other atmospheres, *Mon. Not. R. Astron. Soc.*, 2012, **425**, 21–33.

34 O. L. Polyansky, A. A. Kyuberis, L. Lodi, J. Tennyson, R. I. Ovsyannikov and N. Zobov, ExoMol molecular line lists XIX: high accuracy computed line lists for $H_2{}^{17}O$ and $H_2{}^{18}O$, *Mon. Not. R. Astron. Soc.*, 2017, **466**, 1363–1371.

35 O. L. Polyansky, A. A. Kyuberis, N. F. Zobov, J. Tennyson, S. N. Yurchenko and L. Lodi, ExoMol molecular line lists XXX: a complete high-accuracy line list for water, *Mon. Not. R. Astron. Soc.*, 2018, **480**, 2597–2608.

36 L. Lodi and J. Tennyson, Line lists for $H_2{}^{18}O$ and $H_2{}^{17}O$ based on empirically-adjusted line positions and *ab initio* intensities, *J. Quant. Spectrosc. Radiat. Transfer*, 2012, **113**, 850–858.

37 M. Birk, G. Wagner, J. Loos, L. Lodi, O. L. Polyansky, A. A. Kyuberis, N. F. Zobov and J. Tennyson, Accurate line intensities for water transitions in the infrared: comparison of theory and experiment, *J. Quant. Spectrosc. Radiat. Transfer*, 2017, **203**, 88–102.

38 A. A. Kyuberis, N. F. Zobov, O. V. Naumenko, B. A. Voronin, O. L. Polyansky, L. Lodi, A. Liu, S.-M. Hu and J. Tennyson, Room temperature linelists for deuterated water, *J. Quant. Spectrosc. Radiat. Transfer*, 2017, **203**, 175–185.

39 T. Furtenbacher, T. Szidarovszky, J. Hrubý, A. A. Kyuberis, N. F. Zobov, O. L. Polyansky, J. Tennyson and A. G. Császár, Definitive ideal-gas thermochemical functions of the $H_2{}^{16}O$ Molecule, *J. Phys. Chem. Ref. Data*, 2016, **45**, 043104.

40 I. Simkó, T. Furtenbacher, N. Dénes, T. Szidarovszky, J. Hrubý, N. F. Zobov, O. L. Polyansky, J. Tennyson and A. G. Császár, Recommended ideal-gas thermochemical functions for heavy water and its substituent isotopologues, *J. Phys. Chem. Ref. Data*, 2017, **46**, 023104.

41 C. Burgess and J. Hammond, Wavelength standards for the near-infrared spectral region, *Spectroscopy*, 2007, **22**, 40–48.

42 T. Furtenbacher and A. G. Császár, On employing $H_2{}^{16}O$, $H_2{}^{17}O$, $H_2{}^{18}O$, and $D_2{}^{16}O$ lines as frequency standards in the 15–170 cm^{-1} window, *J. Quant. Spectrosc. Radiat. Transfer*, 2008, **109**, 1234–1251.
43 H. M. Pickett, R. L. Poynter, E. A. Cohen, M. L. Delitsky, J. C. Pearson and H. S. P. Müller, Submillimeter, millimeter, and microwave spectral line catalog, *J. Quant. Spectrosc. Radiat. Transfer*, 1998, **60**, 883–890.
44 JPL catalog, http://spec.jpl.nasa.gov/ftp/pub/catalog/catdir.html, 2018.
45 R. Lanquetin, L. H. Coudert and C. Camy-Peyret, High-lying rotational levels of water: an analysis of the energy levels of the five first vibrational states, *J. Mol. Spectrosc.*, 2001, **206**, 54–67.
46 D. Mondelain, S. N. Mikhailenko, E. V. Karlovets, S. Béguier, S. Kassi and A. Campargue, Comb-assisted cavity ring down spectroscopy of 17o-enriched water between 7443 and 7921 cm^{-1}, *J. Quant. Spectrosc. Radiat. Transfer*, 2017, **203**, 206–212.
47 J. Chen, T.-P. Hua, L.-G. Tao, Y. Sun, A.-W. Liu and S.-M. Hu, Absolute frequencies of water lines near 790 nm with 10^{11} accuracy, *J. Quant. Spectrosc. Radiat. Transfer*, 2018, **205**, 91–95.
48 S. G. Kukolich, Measurement of the molecular *g* values in H_2O and D_2O and hyperfine structure in H_2O, *J. Chem. Phys.*, 1969, **50**, 3751–3755.
49 G. Steenbeckeliers and J. Bellet, Spectre micro-onde de molecules $H_2{}^{16}O$, $H_2{}^{17}O$ et $H_2{}^{18}O$, *C. R. Seances Acad. Sci., Vie Acad.*, 1971, **273**, 471–474.
50 C. K. Jen, Rotational magnetic moments in polyatomic molecules, *Phys. Rev.*, 1951, **81**, 197–203.
51 D. W. Posener and M. W. P. Strandberg, Centrifugal distortion effect in asymmetric top molecules III. H_2O, D_2O, and HDO, *Phys. Rev.*, 1954, **95**, 374–384.
52 J.-M. Flaud, C. Camy-Peyret and A. Valentin, Spectre infrarouge a haute résolution des bandes $\nu_1 + \nu_2$ et $\nu_2 + \nu_3$ de $H_2{}^{16}O$, *J. Phys.*, 1972, **33**, 741–747.
53 C. Huiszoon, A high resolution spectrometer for the shorter millimeter wavelength region, *Rev. Sci. Instrum.*, 1971, **42**, 477–481.
54 G. Y. Golubiatnikov, V. N. Markov, A. Guarnieri and R. Knöchel, Hyperfine structure of $H_2{}^{16}O$ and $H_2{}^{18}O$ measured by Lamb-dip technique in the 180–560 GHz frequency range, *J. Mol. Spectrosc.*, 2006, **240**, 251–254.
55 F. C. De Lucia, P. Helminger, R. L. Cook and W. Gordy, Submillimeter microwave spectrum of $H_2{}^{16}O$, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1972, **5**, 487–490.
56 W. C. King and W. Gordy, One-to-two millimeter wave spectroscopy. IV. Experimental methods and results for OCS, CH_3F, and H_2O, *Phys. Rev.*, 1954, **93**, 407–412.
57 M. A. Koshelev, M. Y. Tretyakov, G. Y. Golubiatnikov, V. V. Parshin, V. N. Markov and I. A. Koval, Broadening and shifting of the 321-, 325- and 380-GHz lines of water vapor by pressure of atmospheric gases, *J. Mol. Spectrosc.*, 2007, **241**, 101–108.
58 G. Cazzoli, C. Puzzarini, M. E. Harding and J. Gauss, The hyperfine structure in the rotational spectrum of water: lamb-dip technique and quantum-chemical calculations, *Chem. Phys. Lett.*, 2009, **473**, 21–25.
59 V. N. Markov and A. F. Krupnov, Measurements of the pressure shift of the (110)-(101) water line at 556 GHz produced by mixtures of gases, *J. Mol. Spectrosc.*, 1995, **172**, 211–214.
60 F. Matsushima, H. Odashima, T. Iwasaki, S. Tsunekawa and K. Takagi, Frequency measurement of pure rotational transitions of H_2O from 0.5 to 5 THz, *J. Mol. Spectrosc.*, 1995, **352/353**, 371–378.
61 D. A. Stephenson and R. G. Strauch, Water vapor spectrum near 600 GHz, *J. Mol. Spectrosc.*, 1970, **35**, 494–495.
62 A. V. Burenin, T. M. Fevralskikh, E. N. Karyakin, O. L. Polyansky and S. M. Shapin, Effective Pade Hamiltonian operator and its application for treatment of $H_2{}^{16}O$ rotational spectrum in the ground state, *J. Mol. Spectrosc.*, 1983, **100**, 182–192.
63 S. Yu, J. C. Pearson and B. J. Drouin, Terahertz spectroscopy of water in its second triad, *J. Mol. Spectrosc.*, 2013, **288**, 7–10.
64 M. A. Koshelev, Collisional broadening and shifting of the (211)–(202) transition of $H_2{}^{16}O$, $H_2{}^{17}O$, $H_2{}^{18}O$ by atmosphere gases, *J. Quant. Spectrosc. Radiat. Transfer*, 2011, **112**, 550–552.
65 P. Helminger, J. K. Messer and F. C. De Lucia, Continuously tunable coherent spectroscopy for the 0.1–1.0 THz region, *Appl. Phys. Lett.*, 1983, **42**, 309–310.
66 J. K. Messer, F. C. De Lucia and P. Helminger, The pure rotational spectrum of water vapor—A millimeter, submillimeter, and far infrared analysis, *Int. J. Infrared Millimeter Waves*, 1983, **4**, 505–539.
67 A. Miani and J. Tennyson, Can *ortho–para* transitions for water be observed?, *J. Chem. Phys.*, 2004, **120**, 2732–2739.
68 I. Snellen, High-dispersion spectroscopy of extrasolar planets: from CO in hot Jupiters to O_2 in exo-Earths, *Philos. Trans. R. Soc., A*, 2014, **372**, 20130075.
69 H. J. Hoeijmakers, R. J. de Kok, I. A. G. Snellen, M. Brogi, J. L. Birkby and H. Schwarz, A search for TiO in the optical high-resolution transmission spectrum of HD 209458b: hindrance due to inaccuracies in the line database, *Astron. Astrophys.*, 2015, **575**, A20.
70 J. L. Birkby, R. J. de Kok, M. Brogi, E. J. W. de Mooij, H. Schwarz, S. Albrecht and I. A. G. Snellen, Detection of water absorption in the day side atmosphere of HD 189733 b using ground-based high-resolution spectroscopy at 3.2 μm, *Mon. Not. R. Astron. Soc.*, 2013, **436**, L35–L39.
71 J. L. Birkby, R. J. de Kok, M. Brogi, H. Schwarz and I. A. G. Snellen, Discovery of water at high spectral resolution in the atmosphere of 51 Peg b, *Astrophys. J.*, 2017, **153**, 138.
72 J. Birkby, private communication, 2018.
73 https://www.craf.eu/iau-list-of-important-spectral-lines/#IAU%20list, 2018.
74 M. Herman, J. W. C. Johns and A. R. W. McKellar, High resolution laser Stark and infrared-radiofrequency double resonance spectroscopy of $H_2{}^{16}O$ at 6 μm, *Can. J. Phys.*, 1979, **57**, 397–401.

75 H. Kuze, Microwave spectrum of water in the ν_2 excited vibrational state, *Astrophys. J.*, 1980, **239**, 1131–1133.
76 L. R. Brown and R. A. Toth, Comparison of the frequencies of NH_3, CO_2, H_2O, N_2O, CO, and CH_4 as infrared calibration standards, *J. Opt. Soc. Am. B*, 1985, **2**, 842–856.
77 O. I. Baskakov, V. A. Alekseev, E. A. Alekseev and B. I. Polevoi, New submillimeter rotational lines of water and its isotopes, *Opt. Spektrosk.*, 1987, **63**, 1016–1018.
78 S. P. Belov, I. N. Kozin, O. L. Polyansky, M. Y. Tretyakov and N. F. Zobov, Rotational spectrum of the $H_2{}^{16}O$ molecule in the (010) excited vibrational state, *J. Mol. Spectrosc.*, 1987, **126**, 113–117.
79 T. Amano and F. Scappini, Millimeter-wave spectrum of rotationally excited H_2O, *Chem. Phys. Lett.*, 1991, **182**, 93–95.
80 J. C. Pearson, T. Anderson, E. Herbst, F. C. De Lucia and P. Helminger, Millimeter- and submillimeter-wave spectrum of highly excited states of water, *Astrophys. J.*, 1991, **379**, L41–L43.
81 R. A. Toth, ν_2 band of $H_2{}^{16}O$-line strengths and transition frequencies, *J. Opt. Soc. Am. B*, 1991, **8**, 2236–2255.
82 R. A. Toth, $2\nu_2$–ν_2 and $2\nu_2$ bands of $H_2{}^{16}O$, $H_2{}^{17}O$, and $H_2{}^{18}O$: line positions and strengths, *J. Opt. Soc. Am. B*, 1993, **10**, 1526–1544.
83 R. A. Toth, ν_1–ν_2, ν_3–ν_2, ν_1 and ν_3 bands of $H_2{}^{16}O$: line positions and strengths, *J. Opt. Soc. Am. B*, 1993, **10**, 2006–2029.
84 R. Paso and V.-M. Horneman, High-resolution rotational absorption spectra of $H_2{}^{16}O$, $HD^{16}O$, and $D_2{}^{16}O$ between 110 and 500 cm^{-1}, *J. Opt. Soc. Am. B*, 1995, **12**, 1813–1838.
85 J. C. Pearson, PhD thesis, Duke University, 1995, data from 00CHPePiMa.[90]
86 S. P. Belov, private communication, 2018, data from 00CHPePiMa.[90]
87 L. R. Brown and J. S. Margolis, Empirical line parameters of NH_3 from 4791 to 5294 cm^{-1}, *J. Quant. Spectrosc. Radiat. Transfer*, 1996, **56**, 283–294.
88 P. D. Natale, L. Lorini, M. Inguscio, I. G. Nolt, J. H. Park, G. D. Lonardo, L. Fusina, P. A. R. Ade and A. G. Murray, Accurate frequency measurement for H_2O and $^{16}O_3$ in the 119 cm^{-1} OH atmospheric window, *Appl. Opt.*, 1997, **36**, 8526–8532.
89 S. N. Mikhailenko, V. G. Tyuterev, K. A. Keppler, B. P. Winnewisser, M. Winnewisser, G. Mellau, S. Klee and K. N. Rao, The $2\nu_2$ band of water: analysis of new FTS measurements and high-K_a transitions and energy levels, *J. Mol. Spectrosc.*, 1997, **184**, 330–349.
90 P. Chen, J. C. Pearson, H. M. Pickett, S. Matsuura and G. A. Blake, Submillimeter-wave measurements and analysis of the ground and ν_2 = 1 states of water, *Astrophys. J., Suppl. Ser.*, 2000, **128**, 371–385.
91 Q. Zou and P. Varanasi, Laboratory measurement of the spectroscopic line parameters of water vapor in the 610–2100 and 3000–4050 cm^{-1} regions at lower-tropospheric temperatures, *J. Quant. Spectrosc. Radiat. Transfer*, 2003, **82**, 45–98.
92 V.-M. Horneman, R. Anttila, S. Alanko and J. Pietila, Transferring calibration from CO_2 laser lines to far infrared water lines with the aid of the ν_2 band of OCS and the ν_2, ν_1–ν_2, and ν_1 + ν_2 bands of $^{13}CS_2$: molecular constants of $^{13}CS_2$, *J. Mol. Spectrosc.*, 2005, **234**, 238–254.
93 R. A. Toth, Measurements of positions, strengths and self-broadened widths of H_2O from 2900 to 8000 cm^{-1}: line strength analysis of the 2nd triad bands, *J. Quant. Spectrosc. Radiat. Transfer*, 2005, **94**, 51–107.
94 L. Joly, B. Parvitte, V. Zéninari, D. Courtois and G. Durry, A spectroscopic study of water vapor isotopologues $H_2{}^{16}O$, $H_2{}^{18}O$ and HDO using a continuous wave DFB quantum cascade laser in the 6.7 μm region for atmospheric applications, *J. Quant. Spectrosc. Radiat. Transfer*, 2006, **102**, 129–138.
95 F. Matsushima, N. Tomatsu, T. Nagai, Y. Moriwaki and K. Takagi, Frequency measurement of pure rotational transitions in the ν_2 = 1 state of H_2O, *J. Mol. Spectrosc.*, 2006, **235**, 190–195.
96 G. Cazzoli, C. Puzzarini, G. Buffa and O. Tarrini, Pressure-broadening of water lines in the THz frequency region: improvements and confirmations for spectroscopic databases. Part II, *J. Quant. Spectrosc. Radiat. Transfer*, 2009, **110**, 609–618.
97 B. J. Drouin, S. Yu, J. C. Pearson and H. Gupta, Terahertz spectroscopy for space applications: 2.5–2.7 THz spectra of HD, H_2O and NH_3, *J. Mol. Struct.*, 2011, **1066**, 2–12.
98 S. Yu, J. C. Pearson, B. J. Drouin, M.-A. Martin-Drumel, O. Pirali, M. Vervloet, L. H. Coudert, H. S. P. Müller and S. Brünken, Measurement and analysis of new terahertz and far-infrared spectra of high temperature water, *J. Mol. Spectrosc.*, 2012, **279**, 16–25.
99 Y. Lu, X.-F. Li, J. Wang, A.-W. Liu and S.-M. Hu, H_2O line positions in the 784–795 nm region with 10^{-9} accuracy, *J. Quant. Spectrosc. Radiat. Transfer*, 2013, **118**, 96–101.
100 L. Regalia, C. Oudot, S. Mikhailenko, L. Wang, X. Thomas, A. Jenouvrier and P. Von der Heyden, Water vapor line parameters from 6450 to 9400 cm^{-1}, *J. Quant. Spectrosc. Radiat. Transfer*, 2014, **136**, 119–136.
101 V. T. Sironneau and J. T. Hodges, Line shapes, positions and intensities of water transitions near 1.28 μm, *J. Quant. Spectrosc. Radiat. Transfer*, 2015, **152**, 1–15.
102 S. Kassi, T. Stoltmann, M. Casado, M. Daeron and A. Campargue, Lamb dip CRDS of highly saturated transitions of water near 1.4 μm, *J. Chem. Phys.*, 2018, **148**, 054201.
103 S. N. Mikhailenko, D. Mondelain, E. V. Karlovets, S. Kassi and A. Campargue, Comb-assisted cavity ring down spectroscopy of ^{17}O enriched water between 6667 and 7443 cm^{-1}, *J. Quant. Spectrosc. Radiat. Transfer*, 2018, **206**, 163–171.
104 J. K. G. Watson, Robust weighting in least-squares fits, *J. Mol. Spectrosc.*, 2003, **219**, 326–328.
105 W. Ritz, On a new law of series spectra, *Astrophys. J.*, 1908, **28**, 237.
106 J. K. G. Watson, The use of term-value fits in testing spectroscopic assignments, *J. Mol. Spectrosc.*, 1994, **165**, 283–290.
107 G. Guennebaud, B. Jacob, *et al.*, *Eigen v3*, http://eigen.tuxfamily.org, 2010.

108 T. Furtenbacher, R. Tóbiás, J. Tennyson, O. V. Naumenko, O. L. Polyansky, N. F. Zobov and A. G. Császár, The 2018 update of the IUPAC database of water energy levels, *J. Quant. Spectrosc. Radiat. Transfer*, in preparation.

109 R. A. Toth, $D_2{}^{16}O$ and $D_2{}^{18}O$ transition frequencies and strengths in the ν_2 bands, *J. Mol. Spectrosc.*, 1993, **162**, 41–54.

110 J. Kauppinen, K. Jomena and V.-M. Horneman, New wavenumber calibration tables for H_2O, CO_2 and OCS lines between 400 cm^{-1} and 900 cm^{-1}, *Appl. Opt.*, 1982, **21**, 3332–3336.

111 G. Guelachvili, Experimental Doppler-limited spectra of the ν_2-bands of $H_2{}^{16}O$, $H_2{}^{17}O$, $H_2{}^{18}O$, and HDO by Fourier-transform spectroscopy: secondary wave-number standards between 1066 and 2296 cm^{-1}, *J. Opt. Soc. Am.*, 1983, **2**, 137–150.

112 C. Y. Lee, An algorithm for path connections and its applications, *IRE Trans. Electron. Comput.*, 1961, **10**, 346–365.