

Improved assessment of H_2^{16}O transitions in the region $6600\text{--}7500\text{ cm}^{-1}$ Tibor Furtenbacher^a, Scott C. Egbert^b, Keeyoon Sung^c, Sean C. Coburn^b, Brian J. Drouin^c, Gregory B. Rieker^b, Attila G. Császár^{a,c,*}^a Institute of Chemistry, ELTE Eötvös Loránd University, H-1117 Budapest, Pázmány Péter sétány 1/A, Hungary^b Precision Laser Diagnostics Laboratory, University of Colorado Boulder, 1111 Engineering Drive, Boulder, CO 80309, USA^c Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

ARTICLE INFO

Keywords:

Rovibrational transitions

Dual-comb spectroscopy

 H_2^{16}O

Line positions

CW2024 line list

Empirical energy levels

ABSTRACT

Due to the very large number of experimental, empirical, and computational studies addressing the rovibrational transitions and the underlying energy levels of the main isotopologue, H_2^{16}O , of the water molecule, the room-temperature coverage offered by the canonical line-by-line spectroscopic database HITRAN2020 is almost complete, 99.87%, up to 6600 cm^{-1} . This study provides a detailed comparison of two recent datasets, one with measured line positions and intensities, obtained using a dual-comb spectrometer (DCS), from 6600 to 7500 cm^{-1} at temperatures up to 1300 K (neglecting the experimentally covered region of $7500\text{--}7650\text{ cm}^{-1}$ with noisy spectral features), and the other a composite empirical line list, called CW2024, covering the spectrum from the microwave to the ultraviolet. The aim has been the mutual validation of the two datasets, correction of possible errors in both of them, and providing additional validated experimental data for inclusion in line-by-line databases. At the end, 5419 transitions from the DCS measurement database were investigated with respect to those in the CW2024 line list, with the majority of the lines falling within the respective individual uncertainties of the two datasets. Using the composite empirical CW2024 database as a reference, 423 assignments were made for a subset of the 616 lines measured at high temperatures which are not part of HITRAN2020. Thus, in the region $6600\text{--}7500\text{ cm}^{-1}$ all but seven transitions, with intensities above a room-temperature cutoff of $1 \times 10^{-28}\text{ cm molecule}^{-1}$, have become experimentally known. In the Supplementary Material to this paper, an updated DCS database is provided, showcasing the improvements described in the article.

1. Introduction

Detailed knowledge of the many characteristics of the rovibrational spectra of water vapour, from the microwave to the ultraviolet, in widely different environments and conditions, including variations in temperature, pressure, and composition, is required in a number of scientific and engineering fields; see, for example, the introductory sections of Refs. [1–3]. Therefore, details about rovibrational lines measured in spectra of water isotopologues, including positions, intensities, lineshapes, self and foreign broadenings and shifts, mostly as a function of temperature and pressure, have been collected in a number of line-by-line (LBL) databases, including HITRAN [4], GEISA [5], ReSpecTh [6], and JPL [7]. The present version of the most widely used LBL spectroscopic database, with more than 32 000 users, is called HITRAN2020 [2]. When judging the coverage offered by HITRAN2020 for water isotopologues, one must take into account that the original purpose of the HITRAN database was to cover and characterize transitions governing the radiative balance of (relatively) cold atmospheres.

However, for the overly important case of water isotopologues [1,3], for the sake of completeness and in response to the large number of related experimental studies, HITRAN2020 already contains quite a few transitions which have minuscule one-photon absorption intensities at room temperature, down to $10^{-30}\text{ cm molecule}^{-1}$, and even below at the longest wavelengths. There exists a high-temperature extension of HITRAN, called HITEMP [8], that was last updated in 2010. There are LBL datasets focusing on elevated temperatures from the ExoMol community [9,10], as well. Improving further the huge number of entries in the HITRAN2020 water dataset, and eventually those of a future iteration of HITEMP, has been an ongoing project in a number of research laboratories around the world. Such advancements usually require joint experimental, empirical, theoretical, and computational efforts [11], especially for a molecule so thoroughly studied by all feasible approaches as water.

Recently, the groups involved in this project, UCB, ELTE, and JPL (see author affiliations for abbreviations), reported complementary

* Corresponding author at: Institute of Chemistry, ELTE Eötvös Loránd University, H-1117 Budapest, Pázmány Péter sétány 1/A, Hungary.

E-mail address: attila.csaszar@ttk.elte.hu (A.G. Császár).

results related to the spectroscopy of the parent water isotopologue, H_2^{16}O . The UCB team, using a dual-comb spectrometer (DCS), observed a large number of rovibrational spectra, spanning the extended wavenumber range of $6600\text{--}7650\text{ cm}^{-1}$ [12]. They extracted from the spectra, with the help of elaborate, mostly multicomponent fittings, a large number of spectral parameters, including line centre positions and line intensities. During this study, the emphasis was not on the accuracy of the line positions deduced, but rather on line shapes and line broadenings as a function of temperature (the highest temperature at which spectra were recorded was 1300 K). To assign the spectral lines, the positions were taken mainly from the HITRAN2020 database [2]. At approximately the same time, the ELTE team completed an analysis of all the measured rovibrational spectra of H_2^{16}O [13], based on the MARVEL (Measured Active Rotational–Vibrational Energy Levels) [14–17] procedure. This analysis produced an extensive set of validated rovibrational transitions and empirical energy levels, both are referred to as W2024 [13]. W2024 updated the previous set of water spectroscopic data called W2020 [18,19], which in fact (a) was the successor of the IUPAC compilation of water (in particular, H_2^{16}O) data [3,20–23], and (b) formed one of the main sources upon which the line positions of HITRAN2020 [2] were based. In the same paper, the ELTE team published an extensive room-temperature H_2^{16}O line list called CW2024. Most of the CW2024 lines are predicted from accurate empirical (W2024) rovibrational energy levels; thus, they have the usual experimental accuracy of better than about 0.005 cm^{-1} , certainly comparable to or even better than the accuracy of the UCB measurements. In Ref. [13], the W2024 rovibrational energies and the corresponding CW2024 transitions were compared with entries of HITRAN2020 [2], clearly supporting each other, save a few cases, where not enough experimental data are available, hindering a MARVEL-type global empirical analysis.

The coverage of the rovibrational transitions of H_2^{16}O provided by HITRAN2020 [2], when compared to a complete first-principles computed line list, depends heavily on the chosen one-photon absorption intensity cutoff [24]. Most importantly, with a cut-off value of $1 \times 10^{-28}\text{ cm molecule}^{-1}$ at room temperature, the coverage for H_2^{16}O in the interval $6600\text{--}7650\text{ cm}^{-1}$ is almost 100%; there are just a small number of missing transitions (of course, with a cut-off value lowered to $1 \times 10^{-30}\text{ cm molecule}^{-1}$, many more transitions are missing from HITRAN2020, but they are less relevant, as these lines are basically invisible at room temperature, though certainly not at the higher temperatures of the DCS measurements, *vide infra*). The extensive set of higher-temperature DCS measurements [12], supplemented by the emission measurements [25] at 1943 K performed between 6600 and 7050 cm^{-1} , offers the chance to increase the coverage provided by HITRAN2020 in this region. Note that assuming room temperature and an intensity cutoff of $1 \times 10^{-28}\text{ cm molecule}^{-1}$, HITRAN2020 has a coverage of 99.87% for atmospherically relevant transitions with wavenumbers lower than 6600 cm^{-1} .

The principal aim of this study has been a detailed comparison of the spectroscopic data of Refs. [12] and [13], of experimental and empirical origin, respectively, the hope being the identification and rectification of incorrect line-position entries in the two datasets. Most importantly, a detailed comparison of the experimental rovibrational transitions reported in Ref. [12], between 6600 and 7650 cm^{-1} , with the corresponding entries of the CW2024 list [13] confirmed the correctness of the great majority of the observed and fitted transitions, both in position and in intensity. In fact, for about 75% of the transitions a perfect match, within the uncertainties, was found between the wavenumbers of HITRAN2020, CW2024, and 24EgSyCoDr (this is how we will refer to Ref. [12] from now on, following the convention introduced in Ref. [20] about tagging published papers for MARVEL-type global empirical analyses). Just as important from the perspective of this article, a total of 616 newly measured lines were published in the Supplementary Material to Ref. [12], which were claimed not to be part of, not “catalogued” in the HITRAN2020 database [2]. A

particularly important aim of the joint project of the three teams was thus the confirmation of as many uncatalogued lines as feasible, so that they can be recommended, with full confidence and complete assignment, for inclusion in spectroscopic databases of water as validated, measured transitions. For many of these lines, that only appear in the measured data at one (high) temperature, 24EgSyCoDr [12] reported a placeholder lower energy, E'' , of 5000 cm^{-1} (since both E'' and the intensity cannot be determined from a measurement at a single temperature). Here we were able to add meaningful E'' values based on the CW2024 line list and enable the calculation of informed intensities based on those E'' values. All higher intensity lines missing from HITRAN2020 were also expected to be identified as measured data, increasing the coverage even closer to 100% for the atmospherically relevant transitions up to 7500 cm^{-1} . As a side product, the detailed comparison of the 24EgSyCoDr, CW2024, and HITRAN2020 data resulted in several useful and general observations, which we decided to present and advocate here. These findings and the related recommendations go beyond the actual spectroscopic data of H_2^{16}O , as they should help future comparisons of experimental, empirical, and computed spectroscopic data.

2. Methodological details

2.1. Measurement conditions and data

Measurements presented in Ref. [12] used a dual-comb spectrometer capable of simultaneously covering the wide region of $6600\text{--}7650\text{ cm}^{-1}$ with low frequency-axis uncertainty. Historically, broadband measurements have required a thermal light source to illuminate the sample, complicating efforts to measure high-temperature gases, where the surrounding optics and equipment also emit strong thermal radiation. Laser spectrometers provide an intense, collimated, and often modulated light source to measure gas absorption at high temperatures. DCS is an emerging technique that is well suited for high-temperature, database-oriented spectroscopic studies, combining the high intensity of laser sources with a bandwidth of 1000 cm^{-1} .

Summarizing Ref. [12], 29 transmission measurements of pure water were collected at conditions ranging from 300 to 1300 K and 0.5 to 16 Torr (near saturation) over a 91.4 cm pathlength, double-passed through a 45.7 cm quartz spectroscopy cell. The spectroscopy cell was placed in the middle of a 100 cm three-zone furnace with refractory baffles to reduce the temperature variation, resulting in a maximum temperature uncertainty of 1.5% at 1300 K . Likewise, pressure was carefully monitored during each nominally 60 min measurement, resulting in a maximum pressure uncertainty of 0.37% .

The 200 MHz repetition rate of the frequency combs in the DCS exactly sets the optical frequency spacing between the ‘teeth’ of the spectrometer, and thus provides a spectral point spacing in the recorded spectrum of 0.0068 cm^{-1} . The repetition rate of both frequency combs was actively referenced to the GPS time base, resulting in a frequency uncertainty of $1.7 \times 10^{-4}\text{ cm}^{-1}$. Each of the 29 measurements presented in Ref. [12] is the average of nominally one hour of data collection, producing an average absorbance noise of 7.4×10^{-4} across the spectrum for all the conditions measured.

The 1300 K measurement at a pressure of 16 Torr is shown in the top panel of Fig. 1. The lower panel of Fig. 1 shows a 3 cm^{-1} portion of the six 16 Torr measurements, representing 0.3% of the total spectral bandwidth measured. The upper and the lower plots contain the same data, but with different vertical axis scaling. The scaling of the upper plot emphasizes the strong transitions in this region, while the zoomed vertical axis of the lower plot shows the dozens of weak transitions in the region, many of which are only visible at higher temperatures. The bulk of the discussion in this work focuses on these weak high-temperature transitions, many of which are experimentally reported for the first time. For additional information about the spectroscopic measurements, see Ref. [12].

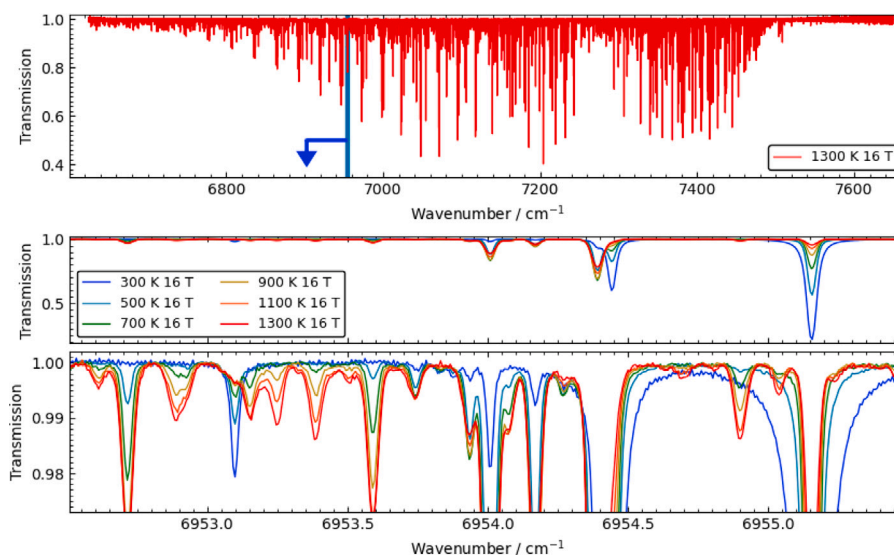


Fig. 1. Top panel: A sample of the measured spectra from Ref. [12], measured at a temperature of 1300 K and a pressure of 16 Torr. Bottom panel: Highlights of the large number of transitions measured, many of which are only visible at the highest temperatures.

In total, 4803 newly measured line centre positions were provided in the supplemental data of 24EgSyCoDr [12], with line assignments taken from the HITRAN2020 database, with reported standard uncertainties corresponding to a 1σ bound. Additionally, 616 measured lines that did not have a clear assignment in the HITRAN2020 database, the reference used by 24EgSyCoDr, were added to the reported database. Most of these lines correspond to transitions with high lower-state energies that are only visible at elevated temperatures, including some observed only at a single temperature, such as those shown in Fig. 1. In addition to measuring line centre positions and line strengths, relevant for the present study, self-widths, shifts, and line-shape parameters were also reported for thousands of spectral lines in Ref. [12], along with the temperature dependence of many self-widths and shifts.

Labelling of the water lines, here as well as in HITRAN2020, is based on the standard rigid-rotor-harmonic-oscillator picture [26]. Thus, the quantum states of H_2^{16}O are denoted as $(v_1 v_2 v_3)[J K_a K_c]$, where v_1 and v_3 are vibrational normal-mode quantum numbers characterizing the symmetric and antisymmetric stretches, v_2 corresponds to the bending motion of the molecule, J is the exact rotational quantum number, and K_a and K_c are the standard [27] approximate asymmetric-top rotational quantum numbers.

2.2. Multispectrum fitting

In Ref. [12], Egbert et al. employed a multispectrum fitting routine (MSFR), developed by Benner et al. [28,29] and extended with a graphical user interface by Drouin et al. [30], to fit the thousands of transitions measured across the 29 temperature and pressure combinations. This technique uses information from all measurements simultaneously to extract the best fit spectral parameters across all conditions measured. In 24EgSyCoDr [12], we first fit parameters for strong lines with large residual errors between the measurement and the model, aiming to avoid interference with smaller, neighbouring transitions (e.g., strong absorptions impacted by the lack of temperature dependence values for collisional widths in this region of the HITRAN2020 database for H_2O absorption). Transitions not present in the data were then added to the line list as needed.

After these initial steps, an incremental fitting approach was used to determine further spectroscopic parameters. Beginning with intensity, line centres, widths, and finally shifts, the MSFR was used to determine spectral parameters for all lines above the noise floor in a given spectral region. If parameters for a line could be measured to within a specified

statistical fit uncertainty (see Ref. [12]), these values were retained and an attempt was made to fit additional parameters. If the uncertainty of the most-recently fit parameter exceeded the specified uncertainty threshold, that parameter was not retained in the updated line list or reported in Ref. [12].

The standard uncertainty values for the database of Ref. [12] were calculated by quadrature addition of the statistical fit uncertainty reported by MSFR (unique for each transition) and measurement-wide uncertainties, including temperature (instrument, spatial, and temporal), pressure (instrument and temporal), and spectrometer wavenumber uncertainties. For weak lines in particular, the total uncertainty is dominated by the reported MSFR statistical fit value. As observed during comparisons with the CW2024 and MARVEL results, MSFR uncertainties obtained by the Benner et al. algorithm [28] for weak lines, especially those in the wings of larger transitions, appear to be too small.

During MSFR, the fourfold variation in the measured temperature presented unique challenges and provided clear opportunities. These are highlighted in the treatment of the 616 uncatalogued transition lines published in the Supplementary Material of Ref. [12], but affect all weaker measured lines presented in Ref. [12].

For the newly measured transitions observed only at a single temperature, it is not possible to infer both the lower state energy and the intensity from the measurement alone. The lower state energy informs how the population of the energy level, measured through the intensity of the line, will change as a function of temperature. When a line is only observed at one temperature, the measured intensity is inextricably coupled to the assumed lower-state energy. In Ref. [12], the authors used a placeholder lower-state energy, $E'' = 5000 \text{ cm}^{-1}$, to provide an anchor for quantifying an intensity value from the measurement. This assumption provided good agreement with the values measured for new transitions that were visible at multiple temperatures, and for which both lower-state energy and intensity could be measured. For those lines where E'' was fixed at 5000 cm^{-1} , the measured intensity (an assumed strength at 296 K) is only meaningful in the context of the measurement temperature and the assumed E'' . MARVEL is used in this work to improve upon this assumption, adding meaningful E'' values and enabling the calculation of informed intensities based on those E'' values. Comparable to the coupled nature of intensity and E'' for high-temperature lines, width and shift parameters for high-temperature transitions are also coupled to their corresponding temperature dependence values, as discussed in greater detail in Ref. [12].

Outside of the challenge of extrapolating parameters away from the reference temperature (296 K) to the temperature where that transition was observed (up to 1300 K), the measurements presented in 24EgSyCoDr [12] have the unique advantage of being able to separate the contributions of different lines by simultaneously leveraging measured absorption at various temperatures. In the spectral region and at the conditions measured, many H_2^{16}O lines overlap, as observable in Fig. 1. In these instances, lines corresponding to transitions with different lower-state energies that might have similar strengths at one temperature are likely to diverge in strength as the measured temperature changes, as shown at 6953.1 and 6953.5 cm^{-1} in Fig. 1. Even though the pairs are nearly indistinguishable at 500–700 K, the transitions diverge dramatically at other measured temperatures. Access to measurements at six temperatures thus gives additional insight into the relative contributions of transitions with overlapping absorption contributions.

2.3. MARVEL

The MARVEL algorithm [14,15,17], used for the comparison, global analysis, and subsequent validation of the experimentally measured transitions, is based upon the theory of spectroscopic networks (SN) [16,31]. In an SN, each energy level in a measured and assigned transition serves as a vertex, while the vertices are interconnected by the observed transitions, serving as edges.

The large number of experimentally observed transitions of H_2^{16}O allow the creation of a well-connected SN, linking all transitions to the two roots of the principal components, corresponding to the two nuclear spin isomers (*ortho*- and *para*- H_2^{16}O), of the SN [32,33]. The W2024 dataset contains 19 027 empirical rovibrational energy values, with individual uncertainties and confidence intervals, whose determination utilized 309 286 (non-unique) transition wavenumbers, reported in 189 (almost exclusively experimental) data sources.

The MARVEL protocol was originally set up to detect inconsistencies among experimentally measured transitions: the simultaneous, global analysis of the SN allows the user to detect lines that are in conflict with other data collected from the literature. This feature proves especially valuable during the identification of issues with experimental data coming from a large number of diverse sources, providing transitions with uncertainties varying many orders of magnitude. In this study, MARVEL has been used to validate both the catalogued and the uncatalogued entries of 24EgSyCoDr [12] (*vide infra*). At the same time, the lines identified by Ref. [12] can be used to check the correctness of the transitions collected during the W2024 study [13], updating the results of the W2020 analysis [18,19].

2.4. The C(H)W2024 line list

Almost all of the energy levels defining the W2024 list of Ref. [13] come from the experimental SN forming part of the W2024 database. The coverage in the rovibrational energy levels is almost complete up to 9500 cm^{-1} . This statement is based on finding one-to-one correspondences with results of sophisticated first-principles computations, which by design provide full coverage [34], though the computed energy levels and transitions are intrinsically significantly less accurate than their measured counterparts.

Based on (a) the empirical W2024 energies [13], supplemented with a few computed ones, and (b) first-principles one-photon absorption intensities, an extensive composite line list, called CW2024, was published in the same study. The CW2024 dataset was compared in Ref. [13] to lines in the HITRAN2020 spectroscopic database [2] and provided important hints about possible improvements of the HITRAN2020 dataset.

The CW2024 line list helps to verify measured transitions in the 6600–7500 cm^{-1} region investigated in this study; this list is believed

to be complete down to room-temperature one-photon absorption intensities of 10^{-31} cm molecule^{-1} . Since Ref. [12] contains transitions measured at temperatures beyond 1000 K, for the purposes of the present study we had to extend the CW2024 line list with transitions that become measurable at these high temperatures. Therefore, all transitions with an intensity larger than 10^{-26} cm molecule^{-1} at $T = 1100$ K were added to the CW2024 line list. The resulting line list is called here CHW2024, where H indicates that this database contains hot lines.

3. Results and discussion

The Supplementary Material of Ref. [12] contains 44 083 transitions corresponding to spectral lines of different water isotopologues, of which 23 120 belong to H_2^{16}O . From the set of H_2^{16}O transitions, 5419 correspond to lines found in the DCS spectra which underwent MSFR fitting. These transitions can be divided into two categories. The first set contains 4803 so-called *catalogued lines*, these are lines for which the assignment was simply taken from the HITRAN2020 database. The second set, of so-called *uncatalogued lines*, contains the remaining 616 transitions. Of these transitions, 576 were strong enough to be fit by MSFR. The remaining 40 uncatalogued lines were clearly visible in the measurements, but were too weak to generate a stable line centre using MSFR. For this subset of lines, a line centre (to the nearest 0.01 cm^{-1}) was reported without a corresponding uncertainty. The uncatalogued lines are perhaps the more important ones for the spectroscopic community. In what follows, first the results obtained for the catalogued and uncatalogued lines during the present study are considered, where the discussion is governed by the CHW2024 line list. This is followed by a presentation of problems with the CHW2024 line list itself.

The Supplementary Material of this paper contains a file called 25FuEgSuCo.csv. It is basically the same as the Supplementary Material of Ref. [12], but a second column, called ‘action’, has been added. If no action was taken on the entry of the Supplementary Material of Ref. [12], the value given in this column is ‘NONE’. Further action categories, such as ‘newly assigned’, ‘reassigned’, ‘refit’, ‘W2024 incompatible’, ‘OK’, and ‘assigned uncat’ are defined in the following two sections. In addition to the annotated 25FuEgSuCo.csv, we have included the CHW2024 line list in the Supplementary Material as CHW2024_6600_7500.txt.

3.1. Measured catalogued line centres in 24EgSyCoDr [12]

Of the 4803 catalogued transitions of Ref. [12], 4770 have full assignments, while 33 lines have only a partial assignment (typically one or more quantum numbers with a value of -2 in HITRAN2020). Of the latter lines, we were able to assign 17 with the help of the CHW2024 line list (note that for these lines the 24EgSyCoDr [12] entries usually have higher accuracy than those in the CHW2024 line list). These lines, with their expanded (two-sigma) uncertainties, can be found in Table 1 and in the file 25FuEgSuCo.csv of the Supplementary Material, tagged as ‘newly assigned’. Consideration of these changes in the assignments is recommended during the next edition of the HITRAN database.

For the remaining 16 lines, no assignment could be found; thus, we were able to check 4787 catalogued lines against those in the CHW2024 database. To achieve this, these 4787 lines were added to the MARVEL input file corresponding to the W2024 database [13]. After executing several MARVEL runs for this extended dataset, we found conflicts for several catalogued lines, which differed from the W2024 lines by significantly more than their stated (expanded) uncertainties. Most of the problems found could be fixed by following two strategies.

The first technique used was simple reassignment. We found two particular issues necessitating these reassignments. The first one is related to *ortho-para* transition pairs, as in some cases the *ortho* and

Table 1

The 17 newly assigned, measured lines among the “catalogued” ones, which had incomplete assignments in the reference HITRAN2020 database [2].

24EgSyCoDr [12] position/cm ⁻¹	CHW2024 position/cm ⁻¹	Assignment
6663.915(4)	6663.915(1)	(0 2 1) [20 3 17] ← (0 0 0) [20 5 16]
6723.531(2)	6723.532(5)	(1 2 1) [16 1 15] ← (0 2 0) [17 1 16]
6745.566(2)	6745.565(3)	(2 1 0) [15 1 14] ← (0 1 0) [16 2 15]
6745.840(4)	6745.841(1)	(1 2 1) [14 2 12] ← (0 2 0) [15 2 13]
6777.7041(8)	6777.7044(6)	(2 1 0) [15 0 15] ← (0 1 0) [16 1 16]
6786.28(1)	6786.2867(4)	(1 3 0) [15 4 11] ← (0 1 0) [16 3 14]
6879.313(6)	6879.317(6)	(2 0 0) [15 9 7] ← (0 0 0) [16 8 8]
6965.284(4)	6965.286(6)	(0 3 1) [15 7 9] ← (0 1 0) [15 7 8]
6968.868(6)	6968.868(1)	(0 2 2) [9 8 2] ← (0 0 1) [9 8 1]
6969.264(2)	6969.265(1)	(2 0 0) [12 9 4] ← (0 0 0) [13 8 5]
6987.956(4)	6987.959(5)	(0 4 0) [14 10 5] ← (0 0 0) [14 11 4]
6999.723(6)	6999.72(1)	(1 0 1) [18 9 9] ← (0 0 0) [18 9 10]
7035.935(4)	7035.93(1)	(1 0 1) [16 10 7] ← (0 0 0) [16 10 6]
7078.522(4)	7078.52(2)	(0 0 2) [14 8 6] ← (0 0 0) [14 9 5]
7148.51(1)	7148.520(4)	(1 3 0) [12 7 6] ← (0 1 0) [12 6 7]
7252.351(6)	7252.33(2)	(2 0 0) [16 9 7] ← (0 0 0) [16 8 8]
7359.543(6)	7359.5402(6)	(2 1 0) [15 0 15] ← (0 1 0) [14 1 14]

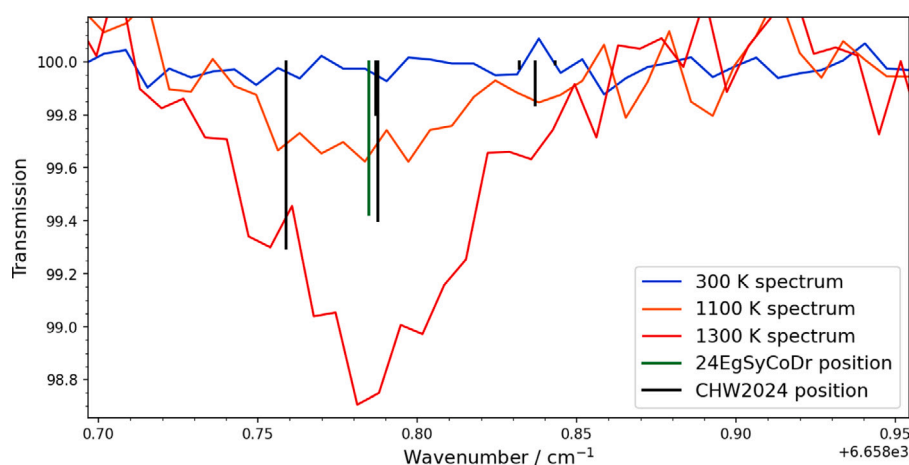


Fig. 2. An example for an incorrectly fitted high-temperature 24EgSyCoDr [12] catalogued line, near 6658.785 cm⁻¹: instead of the single green line, the two black CWH2024 lines should have been fitted.

para labels were swapped in 24EgSyCoDr [12]. Having access to high-quality first-principles results, readily available for H₂¹⁶O [35], one can be confident about the proper ordering of these *ortho-para* transition pairs [36–40]. Another reason behind the incorrect labels is that 24EgSyCoDr [12] used labels of room-temperature HITRAN2020 transitions even in those cases where, due to the high temperature ($T > 1000$ K) at which their lines were recorded, these lines are simply missing from the HITRAN2020 database. After a careful analysis, we could reassign 36 of the 24EgSyCoDr catalogued lines. They are tagged as ‘reassigned’ in the file 25FuEgSuCo.csv of the Supplementary Material.

The second, more complicated method involved the refitting of the problematic transitions. Fig. 2 illustrates a case where refitting became necessary. In this particular case, in 24EgSyCoDr [12] only one transition was fitted to the peak at 6658.785 cm⁻¹, but under the observed peak there should be, as suggested by the CHW2024 line list, at least two transitions with significant and comparable intensity. During this particular refitting, it was assumed that there are two transitions with substantial intensity below the observed peak. As a result of these refittings, we obtained wavenumbers in good agreement with the CHW2024 entries. In total, 26 24EgSyCoDr lines were refitted during this work; they are denoted by the tag ‘refit’ in the file 25FuEgSuCo.csv of the Supplementary Material. Adding these refitted lines to the catalogued lines, we obtain a total of 4801 lines.

At the end, 88 conflicts among the catalogued lines had to be declared to be unsolvable, based on information presently available. They are labelled as ‘W2024 incompatible’. It is likely that in the case

of these deleted transitions, the problem is not with the measured line centre positions, but with the labels assigned to them.

The result of our MARVEL analysis of the W2024 transitions augmented with a set of slightly modified 24EgSyCoDr catalogued transitions can be summarized as follows. In the case of 3530 24EgSyCoDr catalogued lines, that is for more than 73% of the lines reported in 24EgSyCoDr, the stated experimental uncertainties appear to be perfect (labelled as ‘OK’ in the file 25FuEgSuCo.csv). For 923 catalogued lines, the uncertainty had to be increased by a factor of up to 5, and there were only 217 24EgSyCoDr lines whose uncertainty had to be increased by a factor between 5 and 10. These lines are generally small and/or congested by other overlapping absorption, likely resulting in under-reported statistical uncertainties by the MSFR technique.

3.2. Uncatalogued lines

From the original 616 uncatalogued lines published in the Supplementary Material of Ref. [12], 423 could be assigned straightforwardly; both the measured line centres and the intensities were found to be in very good agreement with the CWH2024 results, labelled as ‘assigned uncat’ in the file 25FuEgSuCo.csv. As to the remaining 193 lines, we could not come up with feasible labels based on the information available to us, and have labelled them as ‘unknown in W2024’ in the file 25FuEgSuCo.csv. In general, these transitions have very small room-temperature absorption intensities, with a median of 7×10^{-31} cm molecule⁻¹, and are only visible at elevated temperatures. Correct

Table 2

The 10 lines incorrectly assumed to be uncatalogued in 24EgSyCoDr [12]. The descriptors of the transitions are taken from CW2024 dataset. All line positions (pos.) are in cm^{-1} .

Transition	24EgSyCoDr [12] pos.	CW2024 [13] pos.	HITRAN2020 [2] pos.
(1 0 1) [16 7 10] – (0 0 0) [17 7 11]	6693.491(4)	6693.494(1)	6693.357 500
(0 1 2) [9 1 9] – (0 1 0) [10 2 8]	6829.34(1)	6829.3509(4)	6829.354 800
(1 4 0) [5 5 1] – (0 2 0) [5 4 2]	6983.862(4)	6983.858(1)	6983.866 884
(1 3 0) [14 5 10] – (0 1 0) [14 4 11]	7005.93(2)	7005.929(2)	7005.909 711
(0 4 0) [14 7 7] – (0 0 0) [14 4 10]	7175.653(8)	7175.659(2)	7175.664 798
(0 1 2) [14 2 13] – (0 1 0) [13 3 10]	7212.596(4)	7212.597(1)	7212.590 016
(2 1 0) [12 9 4] – (0 1 0) [12 8 5]	7273.75(3)	7273.75(1)	7273.746 842
(2 1 0) [4 4 0] – (0 1 0) [5 1 5]	7345.43(1)	7345.4360(1)	7345.436 303
(0 0 2) [14 8 6] – (0 0 0) [13 9 5]	7416.161(6)	7416.16(2)	7416.207 612 ^a
(1 0 1) [16 7 10] – (0 0 0) [15 7 9]	7478.655(3)	7478.657(1)	7478.519 890

^a The descriptors of this transition in HITRAN2020 are (–2 –2 –2) [14 –1 –1] ← (0 0 0) [13 9 5].

Table 3

Lines missing from HITRAN2020 in the 6600–7500 cm^{-1} region, corresponding to a room-temperature intensity cutoff of 10^{-28} cm molecule^{-1} .

Wavenumber/ cm^{-1}		Intensity/ cm molecule^{-1}		Transition
24EgSyCoDr	CW2024	24EgSyCoDr	PoKaZaTeL	
6924.606(1)	6924.607(1)	2.36×10^{-28}	2.80×10^{-28}	(1 2 0) [14 8 7] – (0 0 0) [15 7 8]
– ^b	6949.670 8(1)	–	7.48×10^{-28}	(0 4 0) [10 9 1] – (0 0 0) [11 8 4]
– ^b	7115.571 7(1)	–	3.11×10^{-27}	(0 4 0) [9 9 0] – (0 0 0) [10 8 3]
– ^b	7124.300(1)	–	1.29×10^{-28}	(2 0 0) [12 9 4] – (0 0 0) [11 10 1]
7217.648(1)	7217.647 7(1)	3.79×10^{-26}	9.11×10^{-27}	(0 4 0) [10 9 1] – (0 0 0) [10 8 2]
7217.648(1)	7217.648 3(1)	3.79×10^{-26}	2.73×10^{-26}	(0 4 0) [10 9 2] – (0 0 0) [10 8 3]
– ^a	7283.616(1)	–	1.84×10^{-26}	(2 0 0) [12 9 4] – (0 0 0) [12 8 5]
– ^a	7353.3844(4)	–	4.55×10^{-28}	(2 1 0) [12 1 11] – (0 1 0) [11 2 10]
7360.048(2)	7360.050 3(1)	7.92×10^{-26}	5.27×10^{-26}	(0 4 0) [9 9 0] – (0 0 0) [9 8 1]
7360.048(2)	7360.050 4(1)	7.92×10^{-26}	1.76×10^{-26}	(0 4 0) [9 9 1] – (0 0 0) [9 8 2]
– ^b	7436.566(1)	–	3.44×10^{-28}	(0 0 2) [14 7 7] – (0 0 0) [13 8 6]
– ^b	7462.127 1(1)	–	6.32×10^{-28}	(0 4 0) [10 9 1] – (0 0 0) [9 8 2]

^a This line is buried under a transition with a similar lower-state energy.

^b This line is not visible.

assignment of these uncatalogued lines would be assisted by additional high-resolution measurements, especially above 1300 K, a challenging requirement [12,25].

It is important to note that 10 of the uncatalogued lines reported in 24EgSyCoDr [12] can actually be found in the HITRAN2020 database. At the time of the writing of the 24EgSyCoDr article, these lines could not be assigned (the lower energy levels were unknown), resulting in inadequate intensities, and thus the measured lines could not be matched with lines in the HITRAN2020 database. Table 2 contains the 24EgSyCoDr, and the associated HITRAN2020 and CW2024, lines which were incorrectly assumed to be uncatalogued. Table 2 shows that the HITRAN2020 line positions can differ significantly from both the 24EgSyCoDr and CW2024 results, which may be part of the reason why these transitions are listed as uncatalogued lines in Ref. [12].

Table 3 contains the lines missing from HITRAN2020 in the 6600–7500 cm^{-1} region, corresponding to a room-temperature intensity cutoff of 1×10^{-28} cm molecule^{-1} . They are listed together with the uncatalogued lines assigned in this work. As seen in Table 3, after adding the uncatalogued lines of Ref. [12] to the HITRAN2020 database, there are only seven transitions remaining which have not been measured experimentally. Note that during the measurements detailed in Ref. [12], the two *ortho-para* doublet pairs shown in Table 3 were not uniquely distinguishable. The line centre and intensity of the measured absorption is shown, nominally matching the sum of the component transitions taken from a first-principles calculation (PoKaZaTeL [10]). The positions of these experimentally unmeasured lines are known with an accuracy better than 10^{-3} cm^{-1} from the CW2024 line list [13]. Of the seven lines that were not measured, five were not visible at any of the measurement conditions reported in Ref. [12] and another two are either buried under a transition with a similar E'' or are simply not visible. Spectra free from Doppler broadening might help to observe these transitions.

3.3. Incorrect (C)W2024 entries

As detailed in Section 3.1, the catalogued and uncatalogued lines reported in 24EgSyCoDr [12] were checked using the MARVEL procedure. As a result, we found several cases where the 24EgSyCoDr lines appear to be in conflict with previous measurements. In a number of these cases, it is not the interpretation of the 24EgSyCoDr lines, but those of previous measurements that appear to be incorrect. While checking the measured spectra of Ref. [12], we found 72 (out of 5419) measured lines present in the W2024 input file, or lines predicted from measured lines corresponding to other spectral regions, for which there is either definitively no measured peak in the 24EgSyCoDr spectra or where the (C)W2024 lines were too far from the measured peak to be considered a match. As highlighted in this section, many of these prior measurements being removed from CHW2024 are high-temperature emission measurements, which can be less reliable than absorption measurements for a number of reasons (see, e.g., Ref. [25]). Figs. 3–6 illustrate cases of previous incorrect measurements and lines predicted from incorrect measurements.

Fig. 3, for example, shows an *ortho-para* line pair at 6985.126 cm^{-1} , measured and assigned by 20MiKaMoCa [41]. The label of the *ortho* line is (1 0 1) [14 13 1] – (0 0 0) [14 13 2], while that of the *para* pair is (1 0 1) [14 13 2] – (0 0 0) [14 13 1]. As seen in Fig. 3, the line centre position reported in Ref. [12] fits perfectly the observed peak, but the 20MiKaMoCa [41] line is off by about 0.03 cm^{-1} , while its claimed uncertainty is 0.001 cm^{-1} . Therefore, we deleted both the *ortho* and the *para* lines of 20MiKaMoCa from the W2024 database.

The incorrectly predicted line of Fig. 4 is the (1 0 1) [15 11 5] – (0 0 0) [15 11 4] transition of *ortho*-H₂¹⁶O. The upper energy level of this transition is determined by eight measured transitions, reported by 02TeBeZoSh [45], 06ZoShPoBa [46], 08ZoShOvPo [44], and 18Ru-FoScJo [42]. This transition has a *para* pair, whose upper state is also determined by the same sources. At the end, we decided to delete these

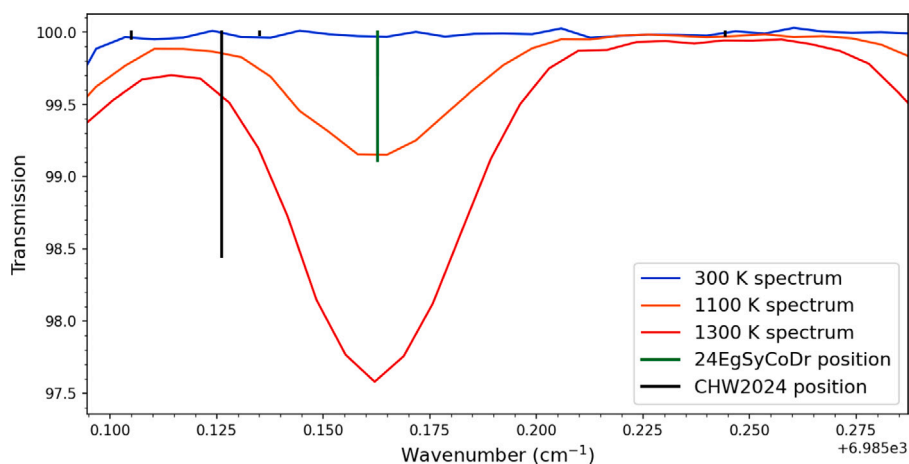


Fig. 3. An incorrect measured line (black) at 6985.126 cm^{-1} reported by 20MiKaMoCa [41].

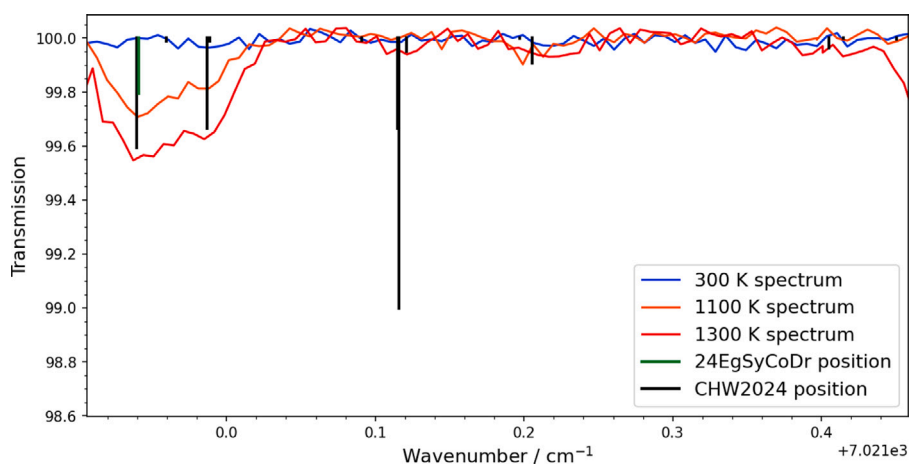


Fig. 4. Incorrect line prediction from 18RuFoScJo [42] at 7021.115 cm^{-1} .

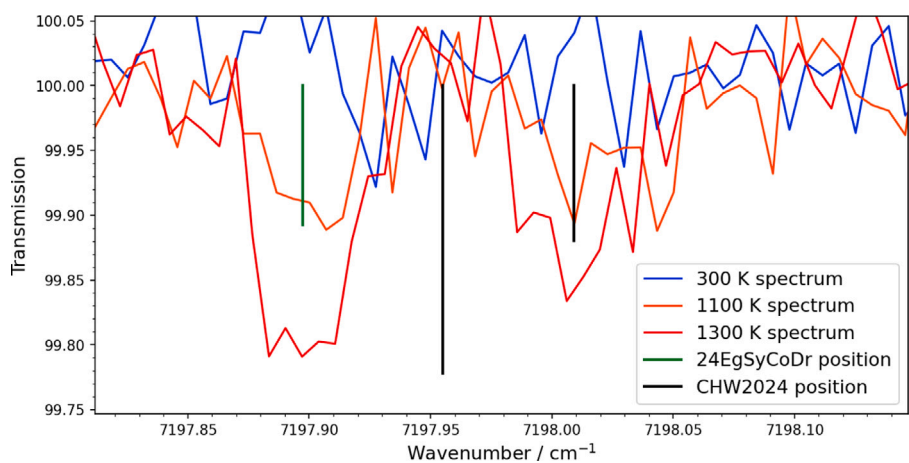


Fig. 5. Incorrect line prediction from 15CaMiLoKa [43] at 7197.955 cm^{-1} .

eight plus eight transitions from W2024 and accept the 24EgSyCoDr measurement at 7020.842 cm^{-1} .

The incorrectly predicted (blue) line of Fig. 5 is based on a measured line reported at $7938.1173 \text{ cm}^{-1}$ in 15CaMiLoKa [43]. As seen in Fig. 5, the line which is predicted to be at 7197.955 cm^{-1} should be at 7197.897 cm^{-1} according to the measurements of Ref. [12]. Therefore, we deleted this 15CaMiLoKa line from the W2024 database.

Fig. 6 shows a very interesting case. An *ortho-para* pair was measured at 7248.274 cm^{-1} by 08ZoShOvPo [44]. The label of the *ortho* line is $(200)[14114] - (000)[14105]$, while its *para* pair is $(200)[14113] - (000)[14104]$. The upper states of these transitions are confirmed by nine plus nine other transitions, measured by 00ZoPoTeSh [47], 06ZoShPoBa [46], 08ZoShOvPo [44], and 18RuFoScJo [42]. There was only one measurement at 7248.094 by 08ZoShOvPo, but it was deleted during the W2024 study. According to Fig. 6, the

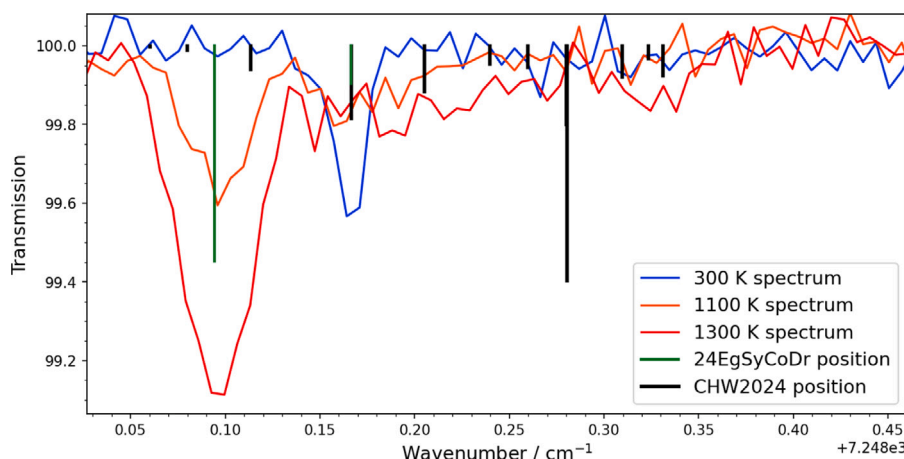


Fig. 6. An incorrect line reported by 08ZoShOvPo [44] at 7248.274 cm^{-1} and its correct line position at 7248.094 cm^{-1} , reported in 24EgSyCoDr [12] as an uncatalogued line.

Table 4

Incorrect W2024 [13] empirical rovibrational energy levels and their newly determined values. The levels are labelled according to $(v_1 v_2 v_3) [J K_a K_c]$. The numbers in brackets show the expanded (two-sigma) uncertainty of the energy levels. Difference is defined as ‘This work’ – ‘W2024’.

Label	Energy/ cm^{-1}		Difference/ cm^{-1}
	W2024	This work	
(0 0 2) [14 9 5]	10932.457(6)	10932.483(2)	0.0262
(0 0 2) [14 9 6]	10932.436(6)	10932.4622(6)	0.0266
(0 0 2) [15 10 5]	11495.272(5)	11495.248(7)	-0.0239
(0 0 2) [15 10 6]	11495.265(5)	11495.241(7)	-0.0240
(0 0 2) [15 9 6]	11282.942(5)	11282.918(2)	-0.0239
(0 0 2) [15 9 7]	11282.876(8)	11282.852(7)	-0.0239
(0 0 2) [16 10 7]	11868.798(1)	Deleted	–
(0 4 1) [13 7 7]	13263.81(1)	Deleted	–
(0 4 1) [15 2 14]	12557.102(1)	12557.109(6)	0.0061
(0 6 0) [14 0 14]	11035.8(1)	11036.203(4)	0.3677
(1 0 1) [14 13 1]	11682.788(1)	11682.8238(4)	0.0363
(1 0 1) [14 13 2]	11682.788(1)	11682.8238(4)	0.0363
(1 0 1) [15 10 6]	11328.495(7)	Deleted	–
(1 0 1) [15 11 4]	11556.079(1)	11555.806(4)	-0.2728
(1 0 1) [15 11 5]	11556.079(1)	11555.807(4)	-0.2728
(1 0 1) [16 5 11]	10905.1090(1)	10905.1161(2)	0.0071
(1 0 1) [16 9 8]	11449.536(6)	11449.562(2)	0.0251
(1 0 1) [17 8 10]	11643.406(1)	11643.4522(9)	0.0461
(1 0 1) [18 10 8]	12461.31(1)	Deleted	–
(1 0 1) [18 3 15]	11301.817(1)	11301.8226(4)	0.0051
(1 0 1) [20 4 16]	12356.419(1)	12356.410(6)	-0.0091
(1 1 1) [13 11 2]	12615.630(6)	12615.646(2)	0.0159
(1 1 1) [13 11 3]	12615.630(6)	12615.646(2)	0.0159
(1 1 1) [18 2 17]	12339.372(2)	12339.3821(8)	0.0104
(1 1 1) [21 1 21]	13055.118(1)	13055.125(6)	0.0071
(1 2 0) [15 9 6]	11022.953(1)	11022.895(4)	-0.0576
(1 2 0) [15 9 7]	11022.920(3)	11022.862(5)	-0.0576
(1 2 0) [16 9 8]	11396.54(1)	11396.512(4)	-0.0320
(1 2 1) [15 0 15]	12575.181(1)	12575.248(7)	0.0672
(1 2 1) [15 1 15]	12575.122(5)	12575.183(1)	0.0611
(1 2 1) [9 9 0]	12717.408(3)	12717.422(6)	0.0132
(1 2 1) [9 9 1]	12717.408(3)	12717.422(6)	0.0132
(2 0 0) [14 11 3]	11170.611(1)	11170.43(2)	-0.1853
(2 0 0) [14 11 4]	11170.611(1)	11170.43(2)	-0.1853
(2 0 0) [16 10 6]	11651.19(1)	11651.206(2)	0.0204
(2 0 0) [16 10 7]	11651.17(1)	11651.192(1)	0.0206
(2 0 0) [16 11 5]	11893.971(6)	11894.007(3)	0.0360
(2 0 0) [16 11 6]	11893.971(6)	11894.007(3)	0.0360
(2 0 0) [16 8 9]	11217.235(6)	11217.362(6)	0.1273
(2 0 0) [17 7 11]	11465.50(9)	Deleted	–
(2 0 0) [19 1 19]	10738.585(2)	10738.5730(5)	-0.0121
(3 0 0) [14 2 13]	12799.294(6)	12799.248(1)	-0.0461

deleted line is the correct one; therefore, we deleted the 10 + 10 measured lines from the *ortho-para* pairs in W2024.

In summary, due to an extensive comparison of W2024 [13] and 24EgSyCoDr [12] lines, a total of 72 W2024 lines had to be deleted and the uncertainty of 58 previously measured lines had to be increased by more than four times. As to the related energy levels, Table 4 contains new energy values that differ significantly, by more than 0.005 cm^{-1} , from the W2024 energy levels, after adding the correct 24EgSyCoDr lines and deleting the incorrect W2024 lines. It can be seen that (a) all problematic energy levels are above 10700 cm^{-1} , (b) the lowest J value is 9, but usually it is significantly higher, (c) there are a large number of *ortho-para* pairs with small but significant splittings, and (d) at the end we had to delete five energy levels from the W2024 energy level set.

4. Summary and conclusions

This paper describes a comprehensive, critical, line-by-line analysis of rovibrational transitions of H_2^{16}O in the wavenumber region 6600–7500 cm^{-1} . The study is based on 29 temperature-dependent spectra published in a paper referred to here as 24EgSyCoDr [12], using a dual-comb spectrometer (DCS), from which the line centre positions and the line intensities are taken. The Supplementary Material of Ref. [12] contains 23 120 spectral lines belonging to H_2^{16}O , from which 5419 lines underwent multispectrum fitting (MSFR), updating parameters such as line centre and intensity. Of those, 4803 catalogued lines have assignments simply taken from the HITRAN2020 [2] database. The so-called uncatalogued lines of Ref. [12] are the remaining 616 transitions. Of the uncatalogued transitions, 576 were strong enough to be fit by MSFR, 40 were visible but too weak to generate a stable line centre using MSFR. These uncatalogued lines are perhaps the most important ones for the spectroscopic community.

In comparing the catalogued lines of Ref. [12] with the empirical lines of the composite CW2024 line list [13] directly, we observed agreement within the respective uncertainties for 73% of the 4803 catalogued lines. Assignments were proposed for 17 catalogued lines contained in 24EgSyCoDr [12], where the HITRAN2020 assignment was incomplete.

From the 616 uncatalogued lines reported in 24EgSyCoDr [12], it was possible to assign 423 using the CW2024 line list. If the uncatalogued lines measured and reported in Ref. [12] were added to HITRAN2020, all but seven lines have become known experimentally in the 6600–7500 cm^{-1} interval, assuming a room temperature intensity cutoff of 1×10^{-28} cm molecule^{-1} . A few problems were identified with the empirical energy levels of the W2024 list, yielding problematic transitions in the CW2024 list. In fact, a total of 72 W2024 lines, and five empirical energy levels above 10700 cm^{-1} with the lowest J value of 9, had to be deleted and the uncertainty of 58 previously measured lines had to be increased by more than four times.

Detailed comparison of the 24EgSyCoDr, CW2024, and HITRAN2020 lines resulted in a couple of important observations and recommendations. (1) Analysis of experimental spectra, especially those recorded at higher temperatures and using multispectrum fitting routines, is a highly complex task. Thus, joint consideration of all available experimental, empirical, and computed data is highly recommended. In the present case, this allowed refitting of certain spectral features, for example, involving a different number of lines in the fitting. (2) MARVEL-type empirical analysis of even the full set of measured transitions suffers from the unavailability of the actually measured spectra. In general, it is recommended to publish complete observed spectra, with appropriate metadata, which would also facilitate their later (re)analysis. (3) Reporting individual uncertainties for the measured lines is highly desirable. This could be achieved here after redoing the MSFR, taking advantage of an extended version of the CW2024 empirical line list, including computed intensities. (4) The large number of additional transitions visible at high temperatures helps to move towards the complete characterization of the region covered by the experiments.

An updated database superseding that of 24EgSyCoDr [12] is provided in the Supplementary Material of this work. It is basically the same as the Supplementary Material of Ref. [12], but a second column, called ‘action’, has been added.

CRedit authorship contribution statement

Tibor Furtenbacher: Writing – review & editing, Visualization, Validation, Methodology, Investigation, Formal analysis. **Scott C. Egbert:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Keeyoon Sung:** Writing – review & editing, Validation, Software, Methodology. **Sean C. Coburn:** Writing – review & editing, Validation, Investigation. **Brian J. Drouin:** Writing – review & editing, Validation, Software, Methodology, Investigation. **Gregory B. Rieker:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Attila G. Császár:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Acknowledgements

The work of AGC received support from the National Research, Development, and Innovation Office of Hungary (NKFIH, grant no. K138233) and from the Fulbright Foundation, funding his stay at the Jet Propulsion Laboratory, California Institute of Technology in 2024. The work of SE and GBR received support through NASA's Exoplanetary Research Program (80NSSC18K0351), the Air Force Research Laboratory (AFRL) (FA8650-20-2-2418), and Air Force Office of Scientific Research (AFOSR) (FA9550-20-1-0328). Part of this research was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration and California Institute of Technology.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jqsrt.2025.109513>.

Data availability

Data will be made available on request.

References

- [1] Bernath PF. The spectroscopy of water vapour: experiment, theory and applications. *Phys Chem Chem Phys* 2002;4:1501–9. <http://dx.doi.org/10.1039/B200372D>.
- [2] Gordon IE, Rothman LS, Hargreaves RJ, Hashemi R, Karlovets EV, Skinner FM, Conway EK, Hill C, Kochanov RV, Tan Y, Wcisło P, Finenko AA, Nelson K, Bernath PF, Birk M, Boudon V, Campargue A, Chance KV, Coustenis A, Drouin BJ, Flaud J-M, Gamache RR, Hodges JT, Jacquemart D, Mlawer EJ, Nikitin AV, Perevalov VI, Rotger M, Tennyson J, Toon GC, Tran H, Tyuterev VG, Adkins EM, Baker A, Barbe A, Cané E, Császár AG, Dudaryonok A, Egorov O, Fleisher AJ, Fleurbaey H, Foltynowicz A, Furtenbacher T, Harrison JJ, Hartmann J-M, Horneman V-M, Huang X, Karman T, Karns J, Kass S, Kleiner I, Kofman V, Kwabia-Tchana F, Lavrentieva NN, Lee TJ, Long DA, Lukashvskaya AA, Lyulin OM, Makhnev VY, Matt W, Massie ST, Melosso M, Mikhailenko SN, Mondelain D, Müller HSP, Naumenko O, Perrin A, Polyansky OL, Raddaoui E, Raston PL, Reed ZD, Rey M, Richard C, Tóbiás R, Sadiek I, Schwenke DW, Starikova E, Sung K, Tamassia F, Tashkun SA, Vander Auwera J, Vasilenko IA, Vigasin AA, Villanueva GL, Vispoel B, Wagner G, Yachmenev A, Yurchenko SN. The HITRAN2020 molecular spectroscopic database. *J Quant Spectr Radiat Transf* 2022;277:107949. <http://dx.doi.org/10.1016/j.jqsrt.2021.107949>.
- [3] Tennyson J, Bernath PF, Brown LR, Campargue A, Császár AG, Daumont L, Gamache RR, Hodges JT, Naumenko OV, Polyansky OL, Rothman LS, Vandaele AC, Zobov NF. A database of water transitions from experiment and theory (IUPAC technical report). *Pure Appl Chem* 2014;86:71–83. <http://dx.doi.org/10.1515/pac-2014-5012>.
- [4] Rothman LS. History of the HITRAN database. *Nat Rev Phys* 2021;3:302–4. <http://dx.doi.org/10.1038/s42254-021-00309-2>.
- [5] Delahaye T, Armante R, Scott N, Jacquinet-Husson N, Chédin A, Crépeau L, Crevoisier C, Douet V, Perrin A, Barbe A, Boudon V, Campargue A, Coudert L, Ebert V, Flaud J-M, Gamache R, Jacquemart D, Jolly A, Kwabia Tchana F, Kyuberis A, Li G, Lyulin O, Manceron L, Mikhailenko S, Moazzen-Ahmadi N, Müller H, Naumenko O, Nikitin A, Perevalov V, Richard C, Starikova E, Tashkun S, Tyuterev V, Vander Auwera J, Vispoel B, Yachmenev A, Yurchenko S. The 2020 edition of the GEISA spectroscopic database. *J Mol Spectrosc* 2021;380:111510. <http://dx.doi.org/10.1016/j.jms.2021.111510>.
- [6] Turányi T, Zsély IG, Papp M, Nagy T, Furtenbacher T, Tóbiás R, Árendás P, Császár AG. ReSpecTh: Reaction kinetics, spectroscopy, and thermochemical datasets. *Sci Data* 2025. in press.
- [7] JPL catalog. 2025. (last accessed 2025)<http://spec.jpl.nasa.gov/ftp/pub/catalog/catdir.html>.
- [8] Rothman LS, Gordon IE, Barber RJ, Dothe H, Gamache RR, Goldman A, Perevalov V, Tashkun SA, Tennyson J. HITEMP, the high-temperature molecular spectroscopic database. *J Quant Spectrosc Radiat Transfer* 2010;111:2139–50.
- [9] Tennyson J, Yurchenko SN, Al-Refaie AF, Barton EJ, Chubb KL, Coles PA, Diamantopoulou S, Gorman MN, Hill C, Lam AZ, Lodi L, McKemmish LK, Na Y, Owens A, Polyansky OL, Sousa-Silva C, Underwood DS, Yachmenev A, Zak E. The ExoMol database: molecular line lists for exoplanet and other hot atmospheres. *J Mol Spectrosc* 2016;327:73–94. <http://dx.doi.org/10.1016/j.jms.2016.05.002>.
- [10] Polyansky OL, Kyuberis AA, Zobov NF, Tennyson J, Yurchenko SN, Lodi L. ExoMol molecular line lists XXX: A complete high-accuracy line list for water. *Mon Not R Astron Soc* 2018;480:2597–608. <http://dx.doi.org/10.1093/mnras/sty1877>.
- [11] Furtenbacher T, Tóbiás R, Árendás P, Császár AG, Past, present. And future of joint utilization of experimental, empirical, and computed spectroscopic data: challenges, advantages, and recommendations. *Adv Quant Chem* 2025;93. <http://dx.doi.org/10.1016/bs.aiq.2025.01.002>, in press.
- [12] Egbert SC, Sung K, Coburn SC, Drouin BJ, Rieker GB. Water-vapor absorption database using dual comb spectroscopy from 300 to 1300 K part I: Pure H₂O, 6600 to 7650 cm⁻¹. *J Quant Spectrosc Radiat Transfer* 2024;318:108940. <http://dx.doi.org/10.1016/j.jqsrt.2024.108940>.
- [13] Furtenbacher T, Tóbiás R, Tennyson J, Gamache RR, Császár AG. The W2024 database of the water isotopologue H₂¹⁶O. *Sci Data* 2024;11:1058. <http://dx.doi.org/10.1038/s41597-024-03847-3>.
- [14] Császár AG, Czákó G, Furtenbacher T, Mátyus E. An active database approach to complete spectra of small molecules. *Annu Rep Comput Chem* 2007;3:155–76. [http://dx.doi.org/10.1016/S1574-1400\(07\)03009-5](http://dx.doi.org/10.1016/S1574-1400(07)03009-5).
- [15] Furtenbacher T, Császár AG, Tennyson J. MARVEL: measured active rotational-vibrational energy levels. *J Mol Spectrosc* 2007;245:115–25. <http://dx.doi.org/10.1016/j.jms.2007.07.005>.
- [16] Császár AG, Furtenbacher T. Spectroscopic networks. *J Mol Spectrosc* 2011;266:99–103. <http://dx.doi.org/10.1016/j.jms.2011.03.031>.

- [17] Furtenbacher T, Császár AG. MARVEL: measured active rotational-vibrational energy levels. II. Algorithmic improvements. *J Quant Spectrosc Radiat Transfer* 2012;113:929–35. <http://dx.doi.org/10.1016/j.jqsrt.2012.01.005>.
- [18] Furtenbacher T, Tóbiás R, Tennyson J, Polyansky OL, Császár AG. : A database of validated rovibrational experimental transitions and empirical energy levels of H_2^{16}O . *J Phys Chem Ref Data* 2020;49(2020):033101. <http://dx.doi.org/10.1063/5.0008253W>.
- [19] Furtenbacher T, Tóbiás R, Tennyson J, Polyansky OL, Kyuberis AA, Ovsyannikov RI, Zobov NF, Császár AG. The W2020 database of validated rovibrational experimental transitions and empirical energy levels of water isotopologues. Part II. H_2^{17}O and H_2^{18}O with an update to H_2^{16}O . *J Phys Chem Ref Data* 2020;49:043103. <http://dx.doi.org/10.1063/5.0030680>.
- [20] Tennyson J, Bernath PF, Brown LR, Campargue A, Carleer MR, Császár AG, Gamache RR, Hodges JT, Jenouvrier A, Naumenko OV, Polyansky OL, Rothman LS, Toth RA, Vandaele AC, Zobov NF, Fally S, Fazliev AZ, Furtenbacher T, Gordon IE, Mikhailenko SN, Shirin SV. IUPAC 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;110:573–96. <http://dx.doi.org/10.1016/j.jqsrt.2009.02.014>.
- [21] Tennyson J, Bernath PF, Brown LR, Campargue A, Császár AG, Daumont L, Gamache RR, Hodges JT, Naumenko OV, Polyansky OL, Rothman LS, Toth RA, Vandaele AC, Zobov NF, Al Derzi AR, Fábri C, Fazliev AZ, Furtenbacher T, Gordon IE, Hu S-M, Mikhailenko SN, Voronin BA. IUPAC 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;111:2160–84. <http://dx.doi.org/10.1016/j.jqsrt.2010.06.012>.
- [22] Tennyson J, Bernath PF, Brown LR, Campargue A, Császár AG, Daumont L, Gamache RR, Hodges JT, Naumenko OV, Polyansky OL, Rothman LS, Vandaele AC, Zobov NF, Al Derzi AR, Fábri C, Fazliev AZ, Furtenbacher T, Gordon IE, Lodi L, Mizus II. 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;117:29–80. <http://dx.doi.org/10.1016/j.jqsrt.2012.10.002>.
- [23] Tennyson J, Bernath PF, Brown LR, Campargue A, Császár AG, Daumont L, Gamache RR, Hodges JT, Naumenko OV, Polyansky OL, Rothman LS, Vandaele AC, Zobov NF, Dénes N, Fazliev AZ, Furtenbacher T, Gordon IE, S-MHu, Szidarovszky T, Vasilenko IA. 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. <http://dx.doi.org/10.1016/j.jqsrt.2014.03.019>.
- [24] Árendás P, Furtenbacher T, Császár AG. Spectroscopic heat maps reveal how to design experiments to improve the uncertainties of transitions and energy levels present in line-by-line databases. *J Quant Spectrosc Radiat Transfer* 2024;315:108878. <http://dx.doi.org/10.1016/j.jqsrt.2023.108878>.
- [25] Czinki E, Furtenbacher T, Császár AG, Eckhardt AK, Mellau GC. 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. <http://dx.doi.org/10.1016/j.jqsrt.2017.10.028>.
- [26] Wilson Jr EB, Decius JC, Cross PC. Molecular vibrations: the theory of infrared and Raman vibrational spectra. New York: McGraw Hill; 1955.
- [27] Krot HW. Molecular rotation spectra. New York: Dover; 1992.
- [28] Benner D, Rinsland CP, Devi V, Smith MAH, Atkins D. A multispectrum nonlinear least squares fitting technique. *J Quant Spectrosc Radiat Transfer* 1995;53:705–21. [http://dx.doi.org/10.1016/0022-4073\(95\)00015-d](http://dx.doi.org/10.1016/0022-4073(95)00015-d).
- [29] Letchworth KL, Benner DC. Rapid and accurate calculation of the voigt function. *J Quant Spectrosc Radiat Transfer* 2007;107:173–92. <http://dx.doi.org/10.1016/j.jqsrt.2007.01.052>.
- [30] Drouin BJ, Benner DC, Brown LR, Cich MJ, Crawford TJ, Devi VM, Guillaume A, Hodges JT, Mlawer EJ, Robichaud DJ, Oyafuso F, Payne VH, Sung K, Wishnow EH, Yu S. Multispectrum analysis of the oxygen A-band. *J Quant Spectrosc Radiat Transfer* 2017;186:118–38. <http://dx.doi.org/10.1016/j.jqsrt.2016.03.037>.
- [31] Furtenbacher T, Császár AG. The role of intensities in determining characteristics of spectroscopic networks. *J Mol Spectrosc* 2012;1009:123–9. <http://dx.doi.org/10.1016/j.molstruc.2011.10.057>.
- [32] Furtenbacher T, Árendás P, Mellau G, Császár AG. Simple molecules as complex systems. *Sci Rep* 2014;4:4654. <http://dx.doi.org/10.1038/srep04654>.
- [33] Császár AG, Furtenbacher T, Árendás P. Small molecules – big data. *J Phys Chem A* 2016;120:8949–69. <http://dx.doi.org/10.1021/acs.jpca.6b02293>.
- [34] Császár AG, Fábri C, Szidarovszky T, Mátyus E, Furtenbacher T, Czako G. The fourth age of quantum chemistry: molecules in motion. *Phys Chem Chem Phys* 2012;14:1085–106. <http://dx.doi.org/10.1039/C1CP21830A>.
- [35] Polyansky OL, Császár AG, Shirin SV, Zobov NF, Barletta P, Tennyson J, Schwenke DW, Knowles PJ. High-accuracy ab initio rotation-vibration transitions for water. *Science* 2003;299:539–42. <http://dx.doi.org/10.1126/science.1079558>.
- [36] Tóbiás R, Furtenbacher T, Simkó I, Császár AG, Diouf ML, Cozijn FMJ, Staa JMA, Salumbides EJ, Ubachs W. Spectroscopic-network-assisted precision spectroscopy and its application to water. *Nat Commun* 2020;11:1708. <http://dx.doi.org/10.1038/s41467-020-15430-6>.
- [37] Diouf ML, Tóbiás R, Simkó I, Cozijn FMJ, Salumbides EJ, Ubachs W, Császár AG. Network-based design of near-infrared lamb-dip experiments and the determination of pure rotational energies of H_2^{16}O at kHz accuracy. *J Phys Chem Ref Data* 2021;50:023106. <http://dx.doi.org/10.1063/5.0052744>.
- [38] Diouf ML, Tóbiás R, van der Schaaf TS, Cozijn FMJ, Salumbides EJ, Császár AG, Ubachs W. Ultraprecise relative energies in the (2 0 0) vibrational band of H_2^{16}O . *Mol Phys* 2022;120:e2050430. <http://dx.doi.org/10.1080/00268976.2022.2050430>.
- [39] Diouf ML, Tóbiás R, Cozijn FMJ, Salumbides EJ, Fábri C, Puzzarini C, Császár AG, Ubachs W. Parity-pair-mixing effects in nonlinear spectroscopy of HDO. *Opt Express* 2022;30:46040–59. <http://dx.doi.org/10.1364/OE.474525>.
- [40] Tóbiás R, Diouf ML, Cozijn FM, Ubachs W, Császár AG. All paths lead to hubs in the spectroscopic networks of water isotopologues H_2^{16}O and H_2^{18}O . *Commun Chem* 2024;7:34. <http://dx.doi.org/10.1038/s42004-024-01103-8>.
- [41] Mikhailenko SN, Kassi S, Mondelain D, Campargue A. Water vapor absorption between 5690 and 8340 cm^{-1} : accurate empirical line centers and validation tests of calculated line intensities. *J Quant Spectrosc Radiat Transfer* 2020;245:106840. <http://dx.doi.org/10.1016/j.jqsrt.2020.106840>.
- [42] Rutkowski L, Poltynowicz A, Schmidt FM, Johansson AC, Khodabakhsh A, Kyuberis AA, Zobov NF, Polyansky OL, Yurchenko SN, Tennyson J. An experimental water line list at 1950 K in the 6250–6670 cm^{-1} region. *J Quant Spectrosc Radiat Transfer* 2018;205:213–9. <http://dx.doi.org/10.1016/j.jqsrt.2017.10.016>.
- [43] Campargue A, Mikhailenko SN, Lohan BG, Karlovets EV, Mondelain D, Kassi S. The absorption spectrum of water vapor in the 1.25 μm atmospheric window (7911–8337 cm^{-1}). *J Quant Spectrosc Radiat. Transf* 2015;157:135–52. <http://dx.doi.org/10.1016/j.jqsrt.2015.02.011>.
- [44] Zobov NF, Shirin SV, Ovsyannikov RI, Polyansky OL, Barber RJ, Tennyson J, Coheur P-F, Bernath PF, Carleer M, Colin R. Spectrum of hot water in the 4750–13 000 cm^{-1} frequency range (0.769–2.1 μm). *Mon Not R Astron Soc* 2008;387:1093–8. <http://dx.doi.org/10.1111/j.1365-2966.2008.13234.x>.
- [45] Tereszchuk K, Bernath PF, Zobov NF, Shirin SV, Polyansky OL, Libeskind NI, Tennyson J, Wallace L. Laboratory spectroscopy of hot water near 2 μm and sunspot spectroscopy in the H-band region. *Astrophys J* 2002;577:496–500. <http://dx.doi.org/10.1086/342167>.
- [46] Zobov NF, Shirin SV, Polyansky OL, Barber RJ, Tennyson J, Coheur P-F, Bernath PF, Carleer M, Colin R. Spectrum of hot water in the 2000–4750 cm^{-1} frequency range. *J Mol Spectrosc* 237(2006):115–22. <http://dx.doi.org/10.1016/j.jms.2006.03.001>.
- [47] Zobov NF, Polyansky OL, Tennyson J, Shirin SV, Nassar R, Hirao T, Imajo T, Bernath PF, Wallace L. Using laboratory spectroscopy to identify lines in the K- and L-band spectrum of water in a sunspot. *Astrophys J* 2000;530:994–8. <http://dx.doi.org/10.1086/308419>.