

Line positions, intensities, and Einstein *A* coefficients for 3-0 band of $^{12}C^{16}O$: A spectroscopy learning method

Zhi-Xiang Fan(范志祥), Zhi-Zhang Ni(倪志樟), Jie-Jie He(贺洁洁), Yi-Fan Wang(王一凡), Qun-Chao Fan(樊群超), Jia Fu(付佳), Yong-Gen Xu(徐勇根), Hui-Dong Li(李会东), Jie Ma(马杰), and Feng Xie(谢锋)

Citation: Chin. Phys. B, 2021, 30 (**12**): 123301. DOI: 10.1088/1674-1056/ac3069
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

R-branch high-lying transition emission spectra of SbNa molecule

Chun-Run Luo(罗春润), Qun-Chao Fan(樊群超), Zhi-Xiang Fan(范志祥), Jia Fu(付佳), Jie Ma(马杰), Hui-Dong Li(李会东), and Yong-Gen Xu(徐勇根)

Chin. Phys. B, 2021, 30 (**1**): 013301. DOI: 10.1088/1674-1056/abc3b2

Study on the $A^2\Pi_{3/2u}$, $B^2\Delta_{3/2u}$, and $X^2\Pi_{3/2g}$ states of Cl_2^+ including its isotopologues

Wu Ling, You Su-Ping, Shao Xu-Ping, Chen Gang-Jin, Ding Ning, Wang You-Mei, Yang Xiao-Hua

Chin. Phys. B, 2015, 24 (**8**): 083301. DOI: 10.1088/1674-1056/24/8/083301

Assignment of terahertz vibrational modes of L-glutamine using density functional theory within generalized-gradient approximation

Zhang Han, Zhang Zhao-Hui, Zhao Xiao-Yan, Zhang Tian-Yao, Yan Fang, Shen Jiang

Chin. Phys. B, 2015, 24 (**7**): 073301. DOI: 10.1088/1674-1056/24/7/073301

High-resolution photoassociation spectroscopy of ultracold Cs_2 long-range 0_g^- state： The external well potential depth

Liu Wen-Liang, Wu Ji-Zhou, Ma Jie, Xiao Lian-Tuan, Jia Suo-Tang

Chin. Phys. B, 2014, 23 (**1**): 013301. DOI: 10.1088/1674-1056/23/1/013301

High resolution photoassociation spectra of an ultracold Cs_2 long-range 0_u^+ ($6S_{1/2}$+$6P_{1/2}$) state

Chen Peng, Li Yu-Qing, Zhang Yi-Chi, Wu Ji-Zhou, Ma Jie, Xiao Lian-Tuan, Jia Suo-Tang

Chin. Phys. B, 2013, 22 (**9**): 093301. DOI: 10.1088/1674-1056/22/9/093301

Line positions, intensities, and Einstein A coefficients for 3–0 band of $^{12}C^{16}O$: A spectroscopy learning method*

Zhi-Xiang Fan(范志祥)[1], Zhi-Zhang Ni(倪志樟)[1], Jie-Jie He(贺洁洁)[1], Yi-Fan Wang(王一凡)[1], Qun-Chao Fan(樊群超)[1,†], Jia Fu(付佳)[1,‡], Yong-Gen Xu(徐勇根)[1], Hui-Dong Li(李会东)[1], Jie Ma(马杰)[2], and Feng Xie(谢锋)[3]

[1] *School of Science, Key Laboratory of High Performance Scientific Computation, Xihua University, Chengdu* 610039, *China*

[2] *State Key Laboratory of Quantum Optics and Quantum Optics Devices, Laser Spectroscopy Laboratory, College of Physics and Electronics Engineering, Shanxi University, Taiyuan* 030006, *China*

[3] *Institute of Nuclear and New Energy Technology, Collaborative Innovation Center of Advanced Nuclear Energy Technology, Key Laboratory of Advanced Reactor Engineering and Safety of Ministry of Education, Tsinghua University, Beijing* 100084, *China*

(Received 31 August 2021; revised manuscript received 2 October 2021; accepted manuscript online 18 October 2021)

Based on the model- and data-driven strategy, a spectroscopy learning method that can extract the novel and hidden information from the line list databases has been applied to the R branch emission spectra of 3–0 band of the ground electronic state of $^{12}C^{16}O$. The labeled line lists such as line intensities and Einstein A coefficients quoted in HITRAN2020 are collected to enhance the dataset. The quantified spectroscopy-learned spectroscopic constants is beneficial for improving the extrapolative accuracy beyond the measurements. Explicit comparisons are made for line positions, line intensities, Einstein A coefficients, which demonstrate that the model- and data-driven spectroscopy learning approach is a promising and an easy-to-implement strategy.

Keywords: carbon monoxide, line lists, a model- and data-driven strategy, spectroscopy learning

PACS: 33.20.Vq, 33.15.Mt, 33.20.Sn, 33.20.–t **DOI:** 10.1088/1674-1056/ac3069

1. Introduction

Carbon monoxide (CO) is the most abundant polar molecule in understanding of interstellar medium, planetary atmospheres and interstellar clouds, in which line positions and line intensities are important in a variety of applications.[1,2] The vibrational–rotational transition focusing on the 3–0 band in the ground state of $^{12}C^{16}O$ has been the subject of numerous experimental and theoretical investigations due to its relatively weak intensity.[3–16]

For example, the ATMOS Fourier transform spectrometer has been used to record accurate transition lines of vibrational–rotational bands, from which the Dunham coefficients have been generated for $^{12}C^{16}O$.[3] Later, the strongest transitions of the 3–0 band were measured by using Fourier transform spectrometer (FTS) with lines from P17 to R17,[4,5] external cavity diode laser (ECL) with lines from R0 to R20,[6] FTS and a multi-line-fitting technique from P25 to R25.[7] In 2014, Mondelain *et al.* reported experimental line positions with sub-MHz accuracy using comb-assisted cavity ring-down spectroscopy.[8] Later, molecular transition frequencies including the line intensities for 3–0 band of $^{12}C^{16}O$ have been measured by Cygan *et al.* using cavity-enhanced spectroscopy techniques for R23,[9] R24, and R28[10] with relative uncertainties at the level of 10^{-10}. Recently, the spectral sensitivity of the comb-locked cavity ring-down spectrometer established by Wang *et al.* allowed the detection of lines with an accuracy of tens of kHz.[11]

It should be noted that comprehensive theoretical investigations have been performed for various line lists, such as Goorvitch's line list (G94),[12] Velichko *et al.* (2012) line list (V12),[13] Coxon and Hajigeorgiou's line list (CH04) derived from an empirical potential function,[14] and Li *et al.* line list.[15] These scientific tasks and high-quality spectral data are helpful for understanding the observational astronomy better, description of model-building, constructing dipole moment function,[16,17] and proposing data analysis techniques.

Recently, the performance analyses of a joint data- and model-driven machine learning approaches were presented for the prediction of full diatomic vibrational spectra including dissociation energy, which were extracted from a wide range of existing heat capacity data.[18] In this article, a model- and data-driven spectroscopy learning method is opted to de-

*Project supported by the Open Research Fund of Computational Physics Key Laboratory of Sichuan Province, Yibin University (Grant No. YBXYJSWL-ZD-2020-006), the Funds for Sichuan Distinguished Scientists of China (Grant Nos. 2019JDJQ0050 and 2019JDJQ0051), the National Natural Science Foundation of China (Grant Nos. 61722507 and 11904295), the National Undergraduate Innovation and Entrepreneurship Training Program of China (Grant No. S202110650046), the State Key Laboratory Open Fund of Quantum Optics and Quantum Optics Devices, Laser Spectroscopy Laboratory (Grant No. KF201811), and the Open Research Fund Program of the Collaborative Innovation Center of Extreme Optics (Grant No. KF2020003).

†Corresponding author. E-mail: fanqunchao@mail.xhu.edu.cn

‡Corresponding author. E-mail: fujiayouxiang@126.com

© 2021 Chinese Physical Society and IOP Publishing Ltd

termine the line positions, intensities, and Einstein A coefficients based on information from the measurements and the HITRAN2020 line list.[19] This paper is organized as follows. Section 2 introduces the model- and data-driven spectroscopy learning strategy. Section 3 gives the application for 3–0 band of $^{12}C^{16}O$. Finally, summary is described in Section 4.

2. Theory and method of calculations

2.1. Modeling of transition lines

The one-dimensional Schrödinger equation can be presented as

$$H(r)\Psi_{\upsilon,J}(r) = E_{\upsilon,J}\Psi_{\upsilon,J}(r), \tag{1}$$

with the Hamiltonian operator

$$H(r) = -\frac{\hbar^2}{2\mu}\frac{\mathrm{d}^2}{\mathrm{d}r^2} + V_J(r), \tag{2}$$

where μ denotes the reduced mass of the given diatomic system, υ and J are the vibrational and rotational quantum numbers, respectively. $V_J(r)$ is the effective potential energy that is a sum of the rotationless potential $V(r)$ plus the centrifugal term from the kinetic operator. The eigenfunctions $\Psi_{\upsilon,J}(r)$ of the potential $V_J(r)$ along with the transition dipole moment curves (TDMCs) were then applied for calculating the TDM matrix elements, yielding transition dipole moment $R_{\upsilon'J',\upsilon''J''}$ to compute the line intensities, Einstein A coefficients for the transitions of a given molecule (see Subsection 2.2). The rovibrational energy levels $E_{\upsilon,J}$ may be calculated by the Dunham formula[20] or by the Herzberg expression,[21] and the latter one is defined as

$$\begin{aligned} E(\upsilon,J) &= G_\upsilon + F_J = G_\upsilon + B_\upsilon\left[J(J+1)-\Omega^2\right] \\ &\quad - D_\upsilon\left[J(J+1)-\Omega^2\right]^2 \\ &\quad + H_\upsilon\left[J(J+1)-\Omega^2\right]^3 \\ &\quad + L_\upsilon[J^4(J+1)-\Omega^2]^4 + \cdots, \end{aligned} \tag{3}$$

where G_υ and F_J correspond to the vibrational and rotational level, respectively. B_υ, D_υ, and $H_\upsilon, \ldots$ are the effective rotational and centrifugal constants. Ω is the projection of the electronic angular momentum onto the internuclear axis. The transition lines from a higher energy state to a lower energy state of a given molecule may then be written as

$$v = v_0 + F'(J') - F''(J''), \tag{4}$$

where $v_0\,(=\{(T_e' - T_e'') + (G' - G'')\})$ is the band origin, single primes and double primes denote the upper and lower states respectively. According to the selection rules, $\Delta J = 0, 1, -1$ stand for the Q, R, and P branches, respectively. Considering the R branch transitions of $^{12}C^{16}O$ molecule in the ground state with Ω being zero, whereby equation (4) becomes

$$\begin{aligned} v &= v_0 + [B_{\upsilon'}\pi_{1,J+1} - D_{\upsilon'}\pi_{2,J+1} \\ &\quad + H_{\upsilon'}\pi_{3,J+1} + L_{\upsilon'}\pi_{4,J+1} + \cdots] \\ &\quad - [B_{\upsilon''}\pi_{1,J} - D_{\upsilon''}\pi_{2,J} \\ &\quad + H_{\upsilon''}\pi_{3,J} + L_{\upsilon''}\pi_{4,J} + \cdots] \end{aligned} \tag{5}$$

defining expansion coefficients $\pi_{i,J}$

$$\pi_{i,J} = [J(J+1)]^i \tag{6}$$

and for two lines from rotational states J_1 and J_2 the transition line difference[22] is given by

$$\begin{aligned} v_{J_2} - v_{J_1} &= [B_{\upsilon'}\pi_{1,J_2+1} - D_{\upsilon'}\pi_{2,J_2+1} \\ &\quad + H_{\upsilon'}\pi_{3,J_2+1} + L_{\upsilon'}\pi_{4,J_2+1} + \cdots] \\ &\quad - [B_{\upsilon''}\pi_{1,J_2} - D_{\upsilon''}\pi_{2,J_2} \\ &\quad + H_{\upsilon''}\pi_{3,J_2} + L_{\upsilon''}\pi_{4,J_2} + \cdots] \\ &\quad - [B_{\upsilon'}\pi_{1,J_1+1} - D_{\upsilon'}\pi_{2,J_1+1} \\ &\quad + H_{\upsilon'}\pi_{3,J_1+1} + L_{\upsilon'}\pi_{4,J_1+1} + \cdots] \\ &\quad + [B_{\upsilon''}\pi_{1,J_1} - D_{\upsilon''}\pi_{2,J_1} \\ &\quad + H_{\upsilon''}\pi_{3,J_1} + L_{\upsilon'}\pi_{4,J_1} + \cdots]. \end{aligned} \tag{7}$$

Replacing J_1 and J_2 with J_2 and J_3, and substituting them into Eq. (7) gives the 2nd transition line difference

$$\begin{aligned} &v_{J_3} - v_{J_2} - (v_{J_2} - v_{J_1}) \\ &= [B_{\upsilon'}\pi_{1,J_3+1} - D_{\upsilon'}\pi_{2,J_3+1} \\ &\quad + H_{\upsilon'}\pi_{3,J_3+1} + L_{\upsilon'}\pi_{4,J_3+1} + \cdots] \\ &\quad - [B_{\upsilon''}\pi_{1,J_3} - D_{\upsilon''}\pi_{2,J_3} \\ &\quad + H_{\upsilon''}\pi_{3,J_3} + L_{\upsilon''}\pi_{4,J_3} + \cdots] \\ &\quad - 2[B_{\upsilon'}\pi_{1,J_2+1} - D_{\upsilon'}\pi_{2,J_2+1} \\ &\quad + H_{\upsilon'}\pi_{3,J_2+1} + L_{\upsilon'}\pi_{4,J_2+1} + \cdots] \\ &\quad + 2[B_{\upsilon''}\pi_{1,J_2} - D_{\upsilon''}\pi_{2,J_2} \\ &\quad + H_{\upsilon''}\pi_{3,J_2} + L_{\upsilon''}\pi_{4,J_2} + \cdots] \\ &\quad + [B_{\upsilon'}\pi_{1,J_1+1} - D_{\upsilon'}\pi_{2,J_1+1} \\ &\quad + H_{\upsilon'}\pi_{3,J_1+1} + L_{\upsilon'}\pi_{4,J_1+1} + \cdots] \\ &\quad - [B_{\upsilon''}\pi_{1,J_1} - D_{\upsilon''}\pi_{2,J_1} \\ &\quad + H_{\upsilon''}\pi_{3,J_1} + L_{\upsilon''}\pi_{4,J_1} + \cdots], \end{aligned} \tag{8}$$

in which the inertial rotational constant $B_{\upsilon'}$ can be constructed as

$$\begin{aligned} B_{\upsilon'} &= (v_{J_3} - 2v_{J_2} + v_{J_1})\pi_{11}^{J+1} + B_{\upsilon''}\pi_{11}^{J} + D_{\upsilon'}\pi_{12}^{J+1} - D_{\upsilon''}\pi_{12}^{J} \\ &\quad - H_{\upsilon'}\pi_{13}^{J+1} + H_{\upsilon''}\pi_{13}^{J} - L_{\upsilon'}\pi_{14}^{J+1} + L_{\upsilon''}\pi_{14}^{J} + \cdots. \end{aligned} \tag{9}$$

Analogously to Eq. (6), the inward expansion coefficients of Eq. (9) are defined as

$$\pi_{11}^{J+1} = \frac{1}{\pi_{1,J_3+1} - 2\pi_{1,J_2+1} + \pi_{1,J_1+1}},$$

$$
\begin{aligned}
\pi_{11}^{J} &= \frac{\pi_{1,J_3} - 2\pi_{1,J_2} + \pi_{1,J_1}}{\pi_{1,J_3+1} - 2\pi_{1,J_2+1} + \pi_{1,J_1+1}},\\
\pi_{12}^{J+1} &= \frac{\pi_{2,J_3+1} - 2\pi_{2,J_2+1} + \pi_{2,J_1+1}}{\pi_{1,J_3+1} - 2\pi_{1,J_2+1} + \pi_{1,J_1+1}},\\
\pi_{12}^{J} &= \frac{\pi_{2,J_3} - 2\pi_{2,J_2} + \pi_{2,J_1}}{\pi_{1,J_3+1} - 2\pi_{1,J_2+1} + \pi_{1,J_1+1}},\\
\pi_{13}^{J+1} &= \frac{\pi_{3,J_3+1} - 2\pi_{3,J_2+1} + \pi_{3,J_1+1}}{\pi_{1,J_3+1} - 2\pi_{1,J_2+1} + \pi_{1,J_1+1}},\\
\pi_{13}^{J} &= \frac{\pi_{3,J_3} - 2\pi_{3,J_2} + \pi_{3,J_1}}{\pi_{1,J_3+1} - 2\pi_{1,J_2+1} + \pi_{1,J_1+1}},\\
\pi_{1_{i_{\max}}^{J+1}} &= \frac{\pi_{i_{\max},J_3+1} - 2\pi_{i_{\max},J_2+1} + \pi_{i_{\max},J_1+1}}{\pi_{1,J_3+1} - 2\pi_{1,J_2+1} + \pi_{1,J_1+1}},\\
\pi_{1_{i_{\max}}^{J}} &= \frac{\pi_{i_{\max},J_3} - 2\pi_{i_{\max},J_2} + \pi_{i_{\max},J_1}}{\pi_{1,J_3+1} - 2\pi_{1,J_2+1} + \pi_{1,J_1+1}}.
\end{aligned}
\tag{10}
$$

Combing Eqs. (5) and (9) gives

$$
\begin{aligned}
&\upsilon_0 + B_{\upsilon''}\delta(J)_2 + D_{\upsilon'}\delta(J)_3 + (-D_{\upsilon''})\,\delta(J)_4\\
&+ (-H_{\upsilon'})\,\delta(J)_5 + H_{\upsilon''}\delta(J)_6\\
&+ (-L_{\upsilon'})\,\delta(J)_7 + L_{\upsilon''}\delta(J)_8 + \cdots\\
&= \upsilon - \left(\upsilon_{J_3} - 2\upsilon_{J_2} + \upsilon_{J_1}\right)\pi_{11}^{J+1}\pi_{1,J+1}.
\end{aligned}
\tag{11}
$$

Therefore, one can simplify Eq. (11) into a matrix form

$$
\boldsymbol{\delta\chi} = \boldsymbol{k}, \tag{12}
$$

where $\boldsymbol{\delta}$ indicates the elements matrix relying on the rotational quantum number J, $\boldsymbol{\chi}$ is the matrix of molecular constants, $\boldsymbol{k}$ is the column vector of transition lines

$$
\boldsymbol{\chi} = \begin{pmatrix}
\upsilon_0\\
B_{\upsilon''}\\
D_{\upsilon'}\\
-D_{\upsilon''}\\
-H_{\upsilon'}\\
H_{\upsilon''}\\
\cdots\\
\chi_{i_{\max}}
\end{pmatrix},
$$

$$
\boldsymbol{k} = \begin{pmatrix}
\upsilon_{J_1'} - \left(\upsilon_{J_3} - 2\upsilon_{J_2} + \upsilon_{J_1}\right)\pi_{11}^{J+1}\pi_{1,J_1'+1}\\
\upsilon_{J_2'} - \left(\upsilon_{J_3} - 2\upsilon_{J_2} + \upsilon_{J_1}\right)\pi_{11}^{J+1}\pi_{1,J_2'+1}\\
\upsilon_{J_3'} - \left(\upsilon_{J_3} - 2\upsilon_{J_2} + \upsilon_{J_1}\right)\pi_{11}^{J+1}\pi_{1,J_3'+1}\\
\upsilon_{J_4'} - \left(\upsilon_{J_3} - 2\upsilon_{J_2} + \upsilon_{J_1}\right)\pi_{11}^{J+1}\pi_{1,J_5'+1}\\
\upsilon_{J_5'} - \left(\upsilon_{J_3} - 2\upsilon_{J_2} + \upsilon_{J_1}\right)\pi_{11}^{J+1}\pi_{1,J_5'+1}\\
\upsilon_{J_6'} - \left(\upsilon_{J_3} - 2\upsilon_{J_2} + \upsilon_{J_1}\right)\pi_{11}^{J+1}\pi_{1,J_6'+1}\\
\cdots\\
\upsilon_{J_{i_{\max}}'} - \left(\upsilon_{J_3} - 2\upsilon_{J_2} + \upsilon_{J_1}\right)\pi_{11}^{J+1}\pi_{1,J_{i_{\max}}'+1}
\end{pmatrix},
$$

$$
\delta(J_l)_i = \begin{cases}
1, & i = 1,\\
\pi_{1_{\frac{i+1}{2}}^{J+1}}\pi_{1,J+1} - \pi_{\frac{i+1}{2},J+1}, & i = 3,5,7,\ldots,\text{odd},\\
\pi_{1_{\frac{i}{2}}^{J}}\pi_{1,J+1} - \pi_{\frac{i}{2},J}, & i = 2,4,6,\ldots,\text{even}.
\end{cases}
\tag{13}
$$

Together equations (11) and (12) allow us to determine the $\{i_{\max}\}$ ($i_{\max}$ = even) molecular constants with $\{i_{\max}+3\}$ accurate experimental transition lines. Thus, it becomes clear that the biggest challenge in calculating the arguments $\boldsymbol{\chi}$ is to obtain the final form of algebraic equation (12), which offers alternative choices for the maximal degrees of the multi-expansions. Accordingly, the model described above that consists of a difference and algebraic approach (DAA) is constructed to calculate R branch transition lines.

2.2. Spectroscopy learning strategy

It appeared that the DAA model with fewer and explicit parameters which is desirable in the prediction of wild spectral range is better than the artificial neural networks. Therefore, understanding and characterizing the proper parameters $\boldsymbol{\chi}$ can improve the overall performance of the proposed model that is generally associated with problems of avoiding under- and over-fitting. According to the DAA model, it is possible to characterize the parameters $\boldsymbol{\chi}$ from the one-to-one mapping

$$
\boldsymbol{\chi} = \boldsymbol{k}\boldsymbol{\delta}^{-1}, \tag{14}
$$

in which the restricted value $\boldsymbol{\chi}$ was manipulated according to the following rule:

$$
\Delta\boldsymbol{\chi} = \Delta\boldsymbol{k}\boldsymbol{\delta}^{-1}. \tag{15}
$$

It is understood that the development of the laser spectroscopy and frequency-comb allows for precisely determined line positions v with high resolution, which play essential roles in determining the emission Einstein A coefficients for the transition from a lower state (υ'', J'') to an upper state (υ', J')[23]

$$
A_{\upsilon'J',\upsilon''J''} = \frac{64\pi^4}{3h}\frac{|m|}{2J'+1}v^3 R^2_{\upsilon'J',\upsilon''J''}, \tag{16}
$$

where the SI unit of Einstein coefficient A is s^{-1}, h is Planck's constant, and $m = [J'(J'+1) - J''(J''+1)]/2$. $R_{\upsilon'J',\upsilon''J''}$ is the transition dipole moment between the lower and upper states, and can be extracted from the freely available computer code LEVEL by Le Roy.[24]

Then the absorption line intensity S (in unit $\frac{\mathrm{cm}^{-1}}{\mathrm{molecule}\cdot\mathrm{cm}^{-2}}$ at temperature T_0 in unit K) of a transition can be expressed in terms of the Einstein A coefficient as[23]

$$
S_{\upsilon'J',\upsilon''J''} = \frac{I_{\mathrm{a}}(2J'+1)\,g_{\mathrm{s}}g_{\mathrm{i}}\exp\left(-hcE_{\upsilon'',J''}/kT_0\right)\left[1-\exp\left(-hcv/kT_0\right)\right]}{8\pi cv^2 Q} \times A_{\upsilon'J',\upsilon''J''}, \tag{17}
$$

where g_{s}, g_{i} are the state-dependent weight and state-independent weight, respectively. c is the speed of light, k is Boltzmann's constant, $E_{v'',J''}$ is the lower energy state, Q is the partition function at T_0 (= 296 K), and I_{a} (= 98.6544%) is the isotopic abundance of $^{12}\mathrm{C}^{16}\mathrm{O}$ molecule.

An alternative approach which corresponds to the calculation of line intensity of a transition at temperature T relies on that of reference temperature T_0[23]

$$S(T)=S(T_0)\frac{\exp\left(-hcE_{v'',J''}/kT\right)\left[1-\exp\left(-hcv/kT\right)\right]Q(T_0)}{\exp\left(-hcE_{v'',J''}/kT_0\right)\left[1-\exp\left(-hcv/kT_0\right)\right]Q(T)}. \tag{18}$$

Finally, it is thus devoted to explore spectroscopy learning that helps to extract the novel and hidden information from the data sets such as $\{v^{\text{expt}}\}$, $\{A^{\text{Ref}}\}$, $\{S^{\text{Ref}}\}$, where the learning and optimization algorithms can be seen in the next section.

2.3. Learning and optimization algorithms

According to Eq. (11), the DAA model embedded in spectroscopy learning algorithm is practically heuristic, yielding many possibilities for $\{\chi\}$, which may suffer from different optimization problems that require some effective implementation algorithms to evaluate the final determination.

Thus, to get the optimal solution, the main idea of spectroscopy learning is to find accessible and explicit rules to minimize the objective functions

$$\chi^*=\arg\min_{\chi}||v^{\text{expt}}-v^{\text{cal}}||, \tag{19}$$

$$\chi^*=\arg\min_{\chi}||A^{\text{expt}}-A^{\text{cal}}||, \tag{20}$$

$$\chi^*=\arg\min_{\chi}||S^{\text{expt}}-S^{\text{cal}}||, \tag{21}$$

where the parameters χ are unknown to be computed through Eq. (12). The Einstein coefficient $A^{\text{cal}}(\chi)$, line intensity $S^{\text{cal}}(\chi)$ both associate strongly with χ and their own definitions. In order to estimate the precision of the reconstruction, we used the residual of the root mean squares (RMS) of the deviations as the distance for transition lines

$$\overline{\Delta\nu}=\sqrt{\frac{1}{N}\sum_{J=1}^{N}\left|\nu_{J,\text{expt}}-\nu_{J,\text{cal}}\right|^2}, \tag{22}$$

where N is the number of data points. According to other spectroscopic quantities, the distance can be arranged into the similar forms, written as

$$\overline{\Delta A}=|A^{\text{Ref}}-A^{\text{cal}}|, \tag{23}$$

$$\overline{\Delta S}=|S^{\text{Ref}}-S^{\text{cal}}|. \tag{24}$$

Equations (22)–(24) listed above are straightforward optimal goals to evaluate χ^*. From another point of view, there is no doubt that the Einstein A coefficient and line intensity are the critical basis for understanding the spectral information, and more importantly, the ideal physical convergence of extrapolative prediction. As mentioned earlier, spectroscopy that is an essential investigative technique allows for satisfactory transition lines $\{v_n^{\text{expt}}\}$ with sub-MHz or even kHz accuracy, which can be used to analyze molecular constants at a local fitting by the least square strategy, however, resulting in poor ability in extrapolating line positions beyond the measurement.

In this work, the most attention of spectroscopy learning has been devoted to the optimization of parameters $\{\chi\}$ with the experimental line positions lists to be divided into training and testing parts, where the optimization problems can be solved and designed in following two aspects.

The first is the algorithm layer where the greedy algorithm[18,25] is presented to identify the "true" lines $\{v_n^{\text{true}}\}$ for the optimal solution χ. In detail, as we know, direct fitting of existing experimental transition spectra may extract inevitable noise, which is an obstacle to the optimal solution χ. In this regard, there is the need to first locate the increment Δv_J from its initial value $\{v_J^{\text{expt}}\}$ for running the greedy algorithm before the optimization work can begin in earnest. Then, two small variations ($\delta v_J^{(1)}=\frac{1}{2}\Delta v_J$) are appended to get three tentative points $v_J^{(1)}\in\{v_J^{\text{expt}}-\frac{1}{2}\Delta v_J, v_J^{\text{expt}}, v_J^{\text{expt}}+\frac{1}{2}\Delta v_J\}$, from which the updated initial value $v_J^{\text{expt}(1)}$ will be with the least error determined for the next iteration. In the next iteration, three new tentative points $v_J^{(2)}\in\{v_J^{\text{expt}(1)}-(\frac{1}{2})^2\Delta v_J, v_J^{\text{expt}(1)}, v_J^{\text{expt}(1)}+(\frac{1}{2})^2\Delta v_J\}$ will be generated in terms of the greedy algorithm until the difference between the last and current iterations satisfies the convergence criteria (10^{-6} MHz). The schematic diagram of greedy algorithm can be seen in Fig. 1.

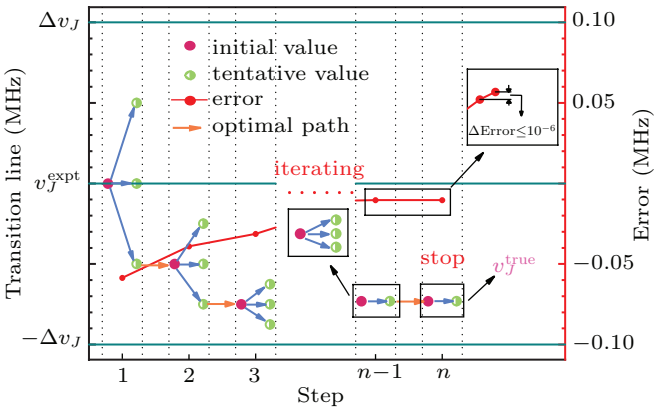


Fig. 1. The schematic diagram of greedy algorithm. The least error (right axis) for each iteration is shown.

The second is the design model that attempts to find a parameter set that performs well on a wide range. As such, the main objective of this part is to implement and investigate an appropriate size of the model to reach the optimal target values $\chi(i)$ According to Eq. (11), the scope of parameters $\chi(i)$ can vary dynamically, which provides a variety of possible alternatives and, however, sometimes may lead to a model collapse. For example, there will be C_n^{i+3} alternatives to determine $\chi(i)$ through Eq. (12) with $\{i+3\}$ training lines from $\{v_n^{\text{expt}}\}$ database. Thus, it is important to explore the size of

the model carefully within allowable tolerance range to avoid overfitting problem, which also inspires us to see whether variational degrees of the model can develop a comparably adaptive one and improve the capacity for extrapolative predictive accuracy of the model itself. Apart from the aforementioned two aspects, the Einstein A coefficient and line intensity have been implemented as supplementary physical quantities for the enhanced performance of the exploration beyond the measurement, especially reducing the risk of overfitting problems.

2.4. Calculation details

Step 1 Select an initial value (*e.g.*, $i = 4$), which is suggested for the corresponding $\chi(i)$.

Step 2 Record all possible alternatives C_N^{i+3}, utilizing greedy algorithm in each iteration.

Step 3 Using Eqs. (22)–(24) to quantify the parameters $\chi(i)$ that best meet the corresponding threshold requirements.

Step 4 If Step 3 is done, terminate the process and current i is the final degree of Eq. (12). Otherwise, a bigger adaptive value of degree (*e.g.*, $i = 6$) will be loaded to execute and repeat Steps 1–4.

Step 5 Here, the Einstein A coefficient and line intensity for each line position could be generated.

3. Application

3.1. Line positions

In the present article, the spectroscopy learning method has been applied to the R-branch of 3–0 band in the ground of $^{12}C^{16}O$, where excellent examples of line positions are collected by Mondelain *et al.*[8] and Wang *et al.*[11] Moreover, the former frequencies are quoted in HITRAN2016[26] and latter ones will be updated in HITRAN2020[19] soon. Here, it is possible to test the spectroscopy learning method directly on the Wang *et al.* line lists (W21)[11] and the labelled line intensities and the Einstein A coefficients from HITRAN2020.[19]

Table 1. The determined "true" transition lines $\nu_{J_k}^{\text{true}}$ of the 3–0 band in the ground electronic state of CO (in unit kHz).

J_k	ν_{J_k}	J_k	ν_{J_k}
19	192019842944.262	2	190708280671.213
21	192113077706.786	4	190910419589.804
23	192193333611.446	6	191099878013.489
25	192260575649.424	8	191276620759.635
27	192314768823.342	10	191440612666.217
0	190493496458.961		

In the spectroscopy learning method, the greedy algorithm motivates us to reach the "true" lines $\{\nu_n^{\text{true}}\}$, which are tabulated in Table 1, for optimal spectroscopic constants $\chi(i)$. And, the determined spectroscopic constants are summarized in Table 2, where the ones learned in the spectroscopy learning method are compared with experimental[8,11,27] and theoretical values[13,14] from literature. The comparison shows an acceptable agreement which can allow us to calculate the line positions with the present spectroscopic constants. Of course, one can check these line positions using molecular spectroscopy program (such as PGOPHER[28]) with spectroscopic constants listed in Table 2.

Table 2. Spectroscopic constants of the 3–0 band in the ground electronic state of CO (in unit MHz).

	Band origin	B_υ	D_υ	$H_\upsilon \times 10^7$	$L_\upsilon \times 10^{12}$	Ref.
Lower state $\upsilon = 0$	–	57635.9681468171	0.183506534	1.741247	–1.0560	this work
	–	57635.968019	0.18350489	1.7168	–	[27]
	–	57635.968146	0.18350653	1.7412	–1.05	[11]
	–	57635.968112	0.18350651	1.7429	–1.0952	[13]
	–	57635.968100	0.18350637	1.7417	–1.0942	[14]
Upper state $\upsilon = 3$	190381373.3725621	56061.9101554768	0.183477828	1.607910	–1.0504	this work
	190381373.343	56061.90989	0.1834756	1.578	–	[8]
	190381373.37256	56061.9101499	0.1834778	1.6079	–1.07	[11]
	190381372.985	56061.91080	0.1834778	1.6105	–1.1635	[13]
	190381373.182	56061.91095	0.1834777	1.6094	–1.1626	[14]

In this work, some line lists are introduced for comparison, such as Mondelain *et al.* line list (M15),[8] Farrenq and Guelachvili's line list (FG91),[3] W21 line list,[11] V12 line list,[13] CH04 line list,[14] Li's line list (Li15).[15] Indeed, we can also make effort to reproduce the transition frequencies of P branch. Figure 2 gives the differences of the lines *versus* m in the range $-31 \leq m \leq 30$. It can be seen that an excellent agreement can be obtained between the present transition frequencies with values from measurements of M15[8] within sub-MHz accuracy and W21[11] within kHz accuracy, and with values from calculations of V12[13] and CH04[14] line lists except for FG91 set.[3] However, transition frequencies from Li's line list[15] show poor agreement with those illustrated in Fig. 2. Details of all line positions obtained which incorporate data from literature are collected in supporting information Table A1. Recently, Cygan *et al.* reported their results of line positions for R23,[9] R24 and R28[10] in the 3–0 band. The determined value ν(R23) = 192 193 333 611 kHz

is close to, but slightly higher than the Cygan's value of 192 193 333 554 kHz.[9] We also compared the current line positions of ν(R24) = 192 228 583.55 MHz and ν(R28) = 192 336 961.15 MHz with the results obtained by Cygan *et al.* with ν(R24) = 192 228 583.47 MHz and ν(R28) = 192 336 961.03 kHz,[10] verifying the accuracy of our values.

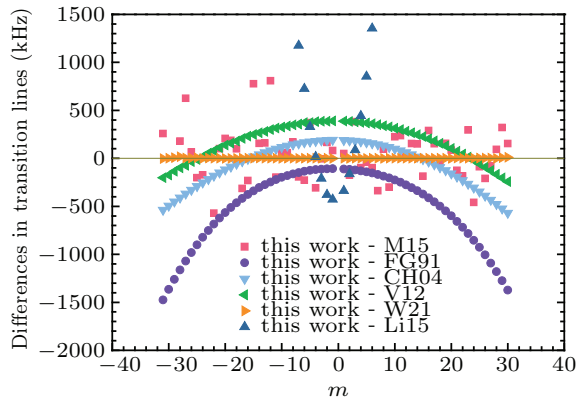


Fig. 2. The differences in the transition lines for the 3–0 band. The line differences display with $m = -J''$ in the P branch and $m = J'' + 1$ in the R branch.

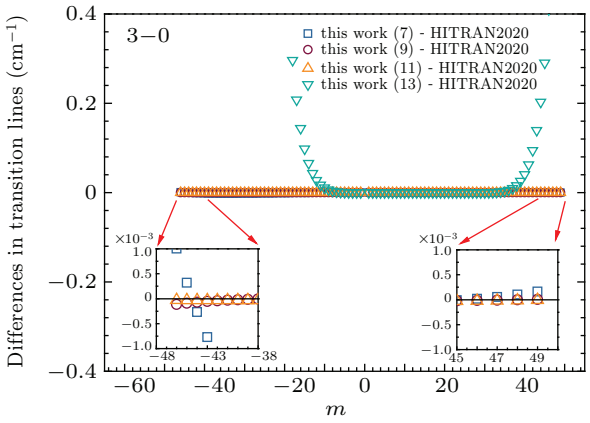


Fig. 3. The differences in the transition lines of "this work (n)" and the values from the HITRAN2020 for the 3–0 band. The line differences display with $m = -J''$ in the P branch and $m = J'' + 1$ in the R branch. It should be noticed that n in "this work (n)" means n experimental lines used in the training.

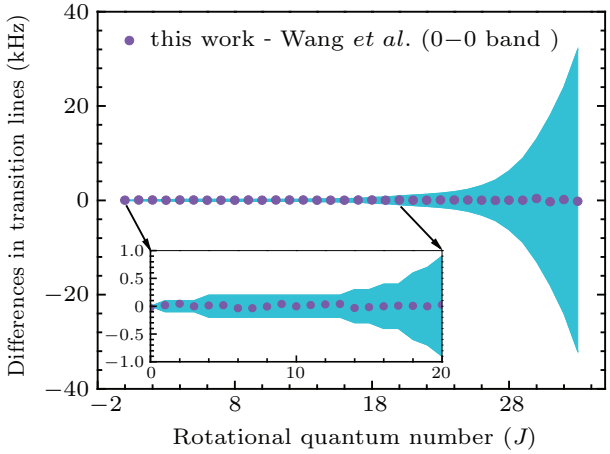


Fig. 4. The differences in the transition lines of this work and those from Wang *et al.*[11] for the 0–0 band. The line differences display with J''. The color shadow covers 1–σ uncertainty region.

As already pointed out, there is diversity in DAA model which can yield line positions with different choices of spectroscopic constants $\chi(i)$ $i = 4, 6, 8 \ldots$. And, comparisons of the differences in line positions are shown in Fig. 3, where lines from "This work (11)" show a best agreement with those of HITRAN2020 line list, giving the RMS value of 2.4×10^{-5} cm^{-1}.

Apparently, the molecular constants with respect to the vibrational states $\upsilon = 0, -3$ allow the calculations of rotational lines for 0–0 and 3–3 bands. The line differences between our results and those from Wang *et al.*[11] are plotted in Fig. 4 and Fig. 5, respectively, leading to an acceptable determination of the transition frequencies with a few kHz accuracy (see supporting information Table A2 for details).

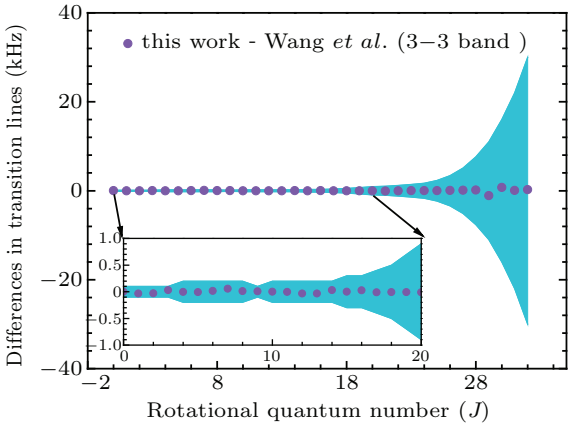


Fig. 5. Differences in the transition lines of this work and those from Wang *et al.*[11] for the 3–3 band. The line differences display with J''. The color shadow covers 1–σ uncertainty region.

3.2. Einstein *A* coefficients and line intensities

The Einstein A coefficients and line intensities are used to evaluate the performance of the DAA model in predicting the line positions especially those beyond the measurements. Then, according to the definitions of the Einstein A coefficient and line intensity (see Eqs. (16) and (17)), the potential challenge is to determine the transition dipole moment $R_{\upsilon' J', \upsilon'' J''}$, which can be extracted from LEVEL[24] from the semi-empirical DMF suggested by Li *et al.*[15] and a PEF from Coxon & Hajigeorgiou[14] or calculated directly through Eq. (16) using Einstein A coefficient and transition line that both are provided by HITRAN2020 database.[19] In this work, we prefer the latter approach to confirm better predictive accuracy of the lines.

The accuracy of Einstein A coefficients and line intensities will be assessed by comparison between this work and those from HITRAN2020.[19] Moreover, in order to compare to literature values, several lists of Einstein A coefficients and line intensities are also collected, which were displayed in Fig. 6 and Fig. 7, respectively. And, detailed values of Einstein A coefficients and line intensities are presented in the supporting information Tables A3 and A4.

For the case of Einstein A coefficient comparisons, values from HITRAN2020,[19] G94,[12] Langhoff and Bauschlicher

(LB95),[29] Huré and Roueff (HR96)[30] are presented. The LB95 Einstein A coefficient values are those calculated by Eq. (16) using the DMF suggested by Langhoff and Bauschlicher.[29] And, the HR96 Einstein A coefficient values are those computed by Eq. (16) using the transition dipole moment $R_{v'J',v''J''}$ obtained by Huré and Roueff.[30] It should be noted that the lines of this work (11) were used in the calculation of Einstein A coefficient for both cases. The satisfactory agreement with HITRAN2020 values is viewed by the Einstein A coefficient comparisons plotted in Fig. 6 with m between −46 and 49. While G94 overestimated the Einstein A coefficients by about average 4.4%, and the Einstein A coefficients of HR96 and LB95 are approximately 1.8% and 1.5% lower than calculations of this work.

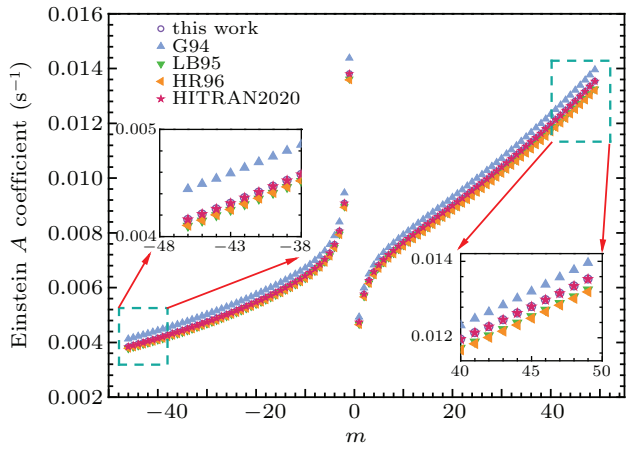


Fig. 6. Comparison of the Einstein A coefficients for the 3–0 band. The Einstein A coefficients display with $m = -J''$ in the P branch and $m = J'' + 1$ in the R branch.

For the case of line intensity comparison at T = 296 K, values from HITRAN2020,[19] G94,[12] LB95, HR96, SV04,[7] and Borkov *et al.* (B20)[17] are taken in account. The G94 line intensities performed at T = 296 K are computed by Eq. (18) using G94 line intensities at reference temperature T = 3000 K[12] with the Q (T = 3000 K) value of 1717.24391[15] in order to foster a fair comparison with available line intensities. The LB95 and HR96 line intensities are calculated by Eq. (17) with respective Einstein A coefficients derived before and with the Q (T = 296 K) value of 107.419824.[15] The B20 line intensities are the calculated values that can be found in supplementary material of Ref. [17].

The comparisons of line intensities *versus* m between −46 and 49 are shown in Fig. 7, in which the line intensity values of G94 and B20 are approximately 5.8% and 3.9% higher than present values, respectively, and the values of LB95 and HR96 are smaller than present values by about 1.5% and 1.8%, respectively. These slight discrepancies arose from that slightly different line position values and transition dipole moment are used for line intensity calculation. Whereas, line intensity values of this work are identical to the entries of the HITRAN2020 values[19] and deviate slightly from those measured by Sung and Varanasi.[7]

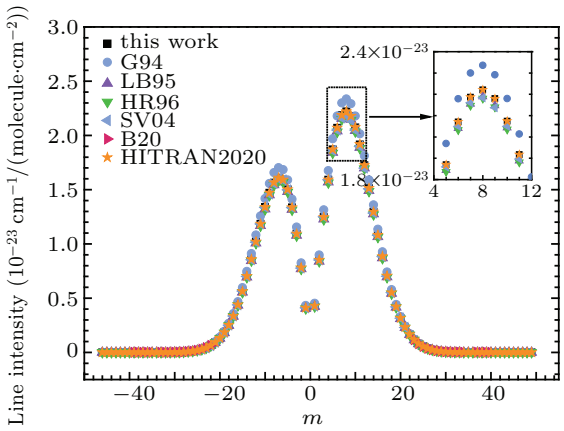


Fig. 7. Comparison of the line intensities for the 3–0 band. The line intensities display with $m = -J''$ in the P branch and $m = J'' + 1$ in the R branch.

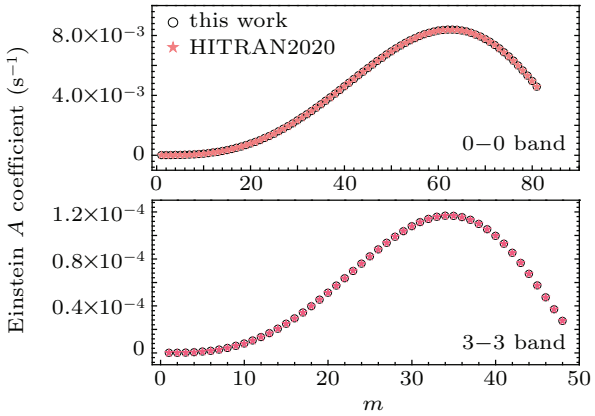


Fig. 8. Comparison of the Einstein A coefficients for the 0–0 and 3–3 bands. The Einstein A coefficients display with $m = J'' + 1$ in the R branch.

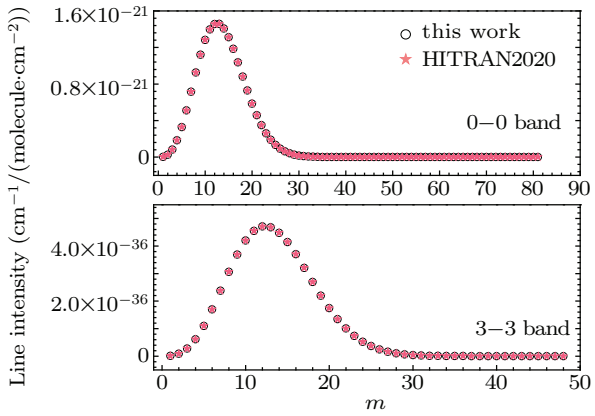


Fig. 9. Comparison of the line intensities for the 0–0 and 3–3 bands. The line intensities display with $m = J'' + 1$ in the R branch.

Let us finally consider the Einstein A coefficients, line intensities for the 0–0 and 3–3 bands. Now, spectroscopic constants reported in Table 2 are ready for these two tasks with the help of transition dipole moment $R_{v'J',v''J''}$ that were extracted from LEVEL[24] from the semi-empirical DMF suggested by Li *et al.*[15] and a PEF from Coxon & Hajigeorgiou.[14] Then, solving Eqs. (16) and (17) leads to the Einstein A coefficients and line intensities that are consistent with those from HITRAN2020.[19] These are clearly shown in Figs. 8 and 9 that compare these two value lists for the cases of Einstein A coefficients and line intensities, respectively, with m = 1 to

$m = 81$ for the 0–0 band, $m = 1$ to $m = 48$ for the 3–3 band. Those values of Einstein A coefficients and line intensities for the 0–0 and 3–3 bands are presented in the supporting information Tables A5 and A6.

4. Conclusions

In this article, a model- and data-driven strategy is proposed to learn line positions for diatomic molecules. The use of the strategy can help us unearth the hidden information behind the measurement and is applied for the 3–0 band in the ground electronic state of $^{12}C^{16}O$, enabling prediction with a few kHz accuracy. The present values of line positions, Einstein A coefficients and line intensities are compared to several other line lists, verifying the validity of the strategy. The results also suggest that the size in the learned-model can have different effects and diversities on the predictive accuracy of line positions. Moreover, in our forthcoming work, it could be interesting to see if the performance of the proposed technique is also reachable for CO isotopologues.

Acknowledgment

We appreciate Prof. V. I. Perevalov for his valuable suggestions for the line intensity calculations.

References

[1] Lyons J R, Gharib-Nezhad E and Ayres T R 2018 *Nat. Commun.* **9** 908
[2] Ayres T R, Lyons J R, Ludwig H G, Caffau E and Wedemeyer-Böhm S 2013 *Astrophys. J.* **765** 1
[3] Farrenq R, Guelachvili G, Sauval A J, Grevesse N and Farmer C B 1991 *J Mol. Spectrosc.* **149** 375
[4] Picqué N and Guelachvili G 1997 *J Mol. Spectrosc.* **185** 244
[5] Picqué N, Guelachvili G Dana V and Mandin J Y 2000 *J Mol. Spectrosc.* **517–518** 427
[6] Henningsen J, Simonsen H, Møgelberg T and Trudsø E 1999 *J Mol. Spectrosc.* **193** 354
[7] Sung K and Varanasi P 2004 *J. Quant. Spectrosc. Radiat. Transfer* **83** 445
[8] Mondelain D, Sala T, Kassi S, Romanini D, Marangoni M and Campargue A 2015 *J. Quant. Spectrosc. Radiat. Transfer* **154** 35
[9] Cygan A, Wcisło P, Wójtewicz S, Kowzan G, Zaborowski M, Charczun D, Bielska K, Trawiński R S, Ciuryło R, Masłowski P and Lisak D 2019 *Opt. Express* **27** 21810
[10] Cygan A, Wójtewicz S, Kowzan G, Zaborowski M, Wcisło P, Nawrocki J, Krehlik P, Śliwczyński Ł, Lipiński M, Masłowski P, Ciuryło R and Lisak D 2016 *J. Chem. Phys.* **144** 214202
[11] Wang J, Hu C L, Liu A W, Sun Y R, Tan Y and Hu S M 2021 *J. Quant. Spectrosc. Radiat. Transfer* **270** 107717
[12] Goorvitch D 1994 *Astrophys. J. Suppl. Ser.* **95** 535
[13] Velichko T I, Mikhailenko S N and Tashkun S A 2012 *J. Quant. Spectrosc. Radiat. Transfer* **113** 1643
[14] Coxon J A and Hajigeorgiou P G 2004 *J. Chem. Phys.* **121** 2992
[15] Li G, Gordon I E, Rothman L S, Tan Y, Hu S M, Kassi S, Campargue A and Medvedev E S 2015 *Astrophys. J. Supple. Ser.* **216** 15
[16] Perevalov V I and Karlovets E V 2019 *J. Mol. Spectrosc.* **364** 111184
[17] Borkov Yu G, Solodov A M, Solodov A A, Petrova T M, Karlovets E V and Perevalov V I 2020 *J. Quant. Spectrosc. Radiat. Transfer* **253** 107064
[18] Fu J, Long S S, Jian J, Fan Z X, Fan Q C, Xie F, Zhang Y and Ma J 2020 *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* **239** 118363
[19] Gordon I E, Rothman L S, Hargreaves R J, *et al.* 2021 *J. Quantum Spectrosc. Radiat. Transfer* **277** 107949
[20] Dunham J L 1932 *Phys. Rev.* **41** 721
[21] Herzberg G and Mrozowski S 1950 *Molecular Spectra and Molecular Structure. I. Spectra of Diatomic Molecules* (New York: Nostrand D Van Printing)
[22] Sun W G, Fan Q C, Li H D and Feng H 2011 *Spectrochim. Acta Part A* **79** 35
[23] Šimečková M, Jacquemart D, Rothman L S and Gamache R R 2006 *J. Quant. Spectrosc. Radiat. Transfer* **98** 130
[24] Le Roy R J 2007 *LEVEL 8.0: University of Waterloo Chemical physics research report CP-663*
[25] Fan Z X, He J J, Ni Z Z, Fan Q C, Fu J, Xu Y G, Li H D, Ma J and Xie F 2021 *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* **264** 120278
[26] Gordon I E, Rothman L S, Hill C, *et al.* 2017 *J. Quantum Spectrosc. Radiat. Transfer* **203** 3
[27] Winnewisser G, Belov S P, Klaus Th and Schieder C W Jr 1997 *J. Mol. Spectrosc.* **184** 468
[28] Western C M 2017 *J. Quant. Spectrosc. Radiat. Transfer* **186** 221
[29] Langhoff S R and Bauschlicher Jr. C W 1995 *J. Chem. Phys.* **102** 5220
[30] Huré J M and Roueff E 1996 *Astron. Astrophys. Suppl. Ser.* **117** 561