Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **545,** 1–15 (2026) https://doi.org/10.1093/mnras/staf2135
Advance Access publication 2025 December 02

ExoMol line lists – LXIII. ExoMol line lists for 12 isotopologues of CO_2

Sergei N. Yurchenko ,★ Marco G. Barnfield, Charles A. Bowesman, Ryan P. Brady , Elizabeth R. Guest, Kyriaki Kefala, Qing-He Ni, Armando N. Perri , Oleksiy A. Smola, Andrei Solokov, Chenyi Tao and Jonathan Tennyson ★

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

Accepted 2025 November 26. Received 2025 November 24; in original form 2025 September 19

ABSTRACT

Extensive rovibrational line lists are constructed for 12 isotopologues of carbon dioxide: $^{12}C^{16}O_2$, $^{13}C^{16}O_2$, $^{12}C^{17}O_2$, $^{13}C^{17}O_2$, $^{12}C^{18}O_2$, $^{13}C^{18}O_2$, $^{16}O^{12}C^{17}O$, $^{16}O^{12}C^{18}O$, $^{16}O^{13}C^{17}O$, $^{16}O^{13}C^{18}O$, $^{17}O^{12}C^{18}O$, and $^{17}O^{13}C^{18}O$. The variational program TROVE was employed together with an exact kinetic energy operator, accurate empirical potential energy surface (Ames-2) and the *ab initio* dipole moment surface Ames-2021-40K. Empirical energy levels from the most recent MARVEL analyses, as well as from the HITRAN and CDSD data bases, are used to replace calculated values where available. The line lists are further supplemented by assigning AFGL quantum numbers using machine-learning based estimators. The resulting data were employed to generate opacities with four radiative transfer codes, TauREx, ARCiS, NEMESIS, and petitRADTRANS, both for individual isotopologues and for CO_2 at terrestrial isotopic natural abundance. All line lists and associated data are available at www.exomol.com.

Key words: molecular data – opacity – planets and satellites: atmospheres – stars: atmospheres – ISM: molecules.

1 INTRODUCTION

Carbon dioxide (CO_2) is a key atmospheric constituent across a wide range of planetary environments, including Earth (F. Oyafuso et al. 2017), Venus (M. Snels et al. 2014; A. Fedorova et al. 2015), and Mars (C. R. Webster et al. 2013; T. F. Refaat & U. N. Singh 2024), as well as exoplanets (M. R. Swain et al. 2009b, a; B. R. Oppenheimer et al. 2013; K. Heng & J. R. Lyons 2016) and brown dwarfs (R. S. Freedman et al. 2014; M. R. Line et al. 2017). In the Earth's atmosphere, CO_2 is the second most abundant greenhouse gas after water vapour (F. Oyafuso et al. 2017), and is therefore central to studies in environmental chemistry. Several dedicated missions (A. Butz et al. 2011; D. Crisp et al. 2004; J. B. Abshire et al. 2010) monitor Earth's atmospheric CO_2 content. Since CO_2 is a major component of the atmospheres of Earth, Mars, and Venus, remote sensing of this molecule is expected to be equally important for characterizing exoplanets with similar properties. For example, CO_2 dominates the opacity of the Venusian atmosphere, comprising approximately 95 per cent of its composition (J. B. Pollack et al. 1993). The critical role of CO_2 in atmospheric retrievals has been highlighted in multiple planetary science studies ranging from terrestrial remote sensing (e.g. S. Roche et al. 2021) to planetary atmospheres such as Mars (e.g. S. Jiménez-Monferrer et al. 2021).

In the context of exoplanetary science, CO_2 is anticipated to be a key molecular absorber across a broad range of planetary types, from temperate terrestrial planets to hot Jupiters. Its strong 4.3 μm band has been unambiguously detected with *James Webb Space Telescope* (*JWST*) in the atmospheres of WASP-39 b (E.-M. Ahrer et al. 2023), WASP-166 b (A. W. Mayo et al. 2025), and the sub-Neptune K2-18 b (N. Madhusudhan et al. 2023; M. Holmberg & N. Madhusudhan 2024). Consequently, comprehensive and accurate opacity data for CO_2 and its isotopologues, valid over wide spectral and temperature ranges, are essential for probing atmospheric composition and physical conditions such as temperature and pressure. Such data underpin high-accuracy atmospheric models and are crucial for interpreting spectroscopic observations from ground-based telescopes and space missions, including *JWST* and the future ESA space observatory Ariel (K. L. Chubb et al. 2024).

Detection of CO_2 with high-resolution spectroscopy has not yet been achieved, mainly due to challenges associated with telluric contamination (I. Carleo et al. 2022), although searches are ongoing (F. Biassoni et al. 2024). Isotopic ratios, in particular $^{13}C^{16}O_2/^{12}C^{16}O_2$, have been proposed as tracers of carbon isotope fractionation in exoplanetary atmospheres, with potential links to biological processes detectable by *JWST* (A. Glidden et al. 2023). Recently, four isotopologues of CO_2 – $^{12}C^{16}O_2$, $^{13}C^{16}O_2$, $^{16}O^{12}C^{18}O$, and $^{16}O^{12}C^{17}O$ – were detected in a terrestrial planet-forming region of an externally irradiated Herbig disc using VLT (J. Frediani et al. 2025).

Significant progress in developing high-resolution, variationally computed CO_2 line lists has been made in recent years by the NASA Ames group (X. Huang et al. 2012, 2013, 2014, 2017) and the ExoMol project (E. Zak et al. 2016; E. J. Zak et al. 2017a, b; S. N. Yurchenko et al. 2020). High-temperature line lists are particularly important for modelling exoplanetary atmospheres, and numerous works have targeted this goal (R. B. Wattson & L. S. Rothman 1986; S. Tashkun et al. 2003; L. S. Rothman et al. 2010; S. A. Tashkun & V. I. Perevalov 2011; X. Huang et al. 2013, 2014, 2017; S. N. Yurchenko et al. 2020). Recent high-temperature data sets include UCL-4000

★ E-mail: s.yurchenko@ucl.ac.uk (SNY); j.tennyson@ucl.ac.uk (JT)

(S. N. Yurchenko et al. 2020), Ames-2021 (X. Huang et al. 2022), the latest Ames line lists AI-3000K (X. Huang et al. 2023) and its HITEMP compilation (R. J. Hargreaves et al. 2024). An accurate room–temperature line list CDSD-2024-PI was recently reported by R. Kochanov & V. Perevalov (2025). While ExoMol and Ames line lists combine experimental results with variational calculations, improved with empirically derived energies, the CDSD-2024-PI (S. A. Tashkun et al. 2015) and HITRAN (E. V. Karlovets et al. 2021; I. E. Gordon et al. 2022) line lists rely on effective Hamiltonian models. All provide accurate and comprehensive rovibrational data for CO_2, but further improvements in accuracy and completeness remain necessary.

An additional aspect relevant to this work is the treatment of pressure broadening. A large body of literature exists on CO_2 broadening coefficients with different perturbing species; here we only highlight a few recent developments. R. R. Gamache & J. Lamouroux (2013) presented a detailed analysis of a large set of experimental broadening parameters for the main isotopologue colliding with N_2, O_2, and CO_2, with further analyses and reassessments reported by R. R. Gamache et al. (2014) and R. Hashemi et al. (2020).

In this work we present extensive rovibrational line lists for 12 isotopologues of CO_2: $^{12}C^{16}O_2$, $^{13}C^{16}O_2$, $^{12}C^{17}O_2$, $^{13}C^{17}O_2$, $^{12}C^{18}O_2$, $^{13}C^{18}O_2$, $^{16}O^{12}C^{17}O$, $^{16}O^{12}C^{18}O$, $^{16}O^{13}C^{17}O$, $^{16}O^{13}C^{18}O$, $^{17}O^{12}C^{18}O$, and $^{17}O^{13}C^{18}O$. The line lists were generated with TROVE (S. N. Yurchenko, W. Thiel & P. Jensen 2007) using an exact kinetic energy operator (KEO), the empirical potential energy surface (PES) Ames-2 (X. Huang et al. 2017), and the *ab initio* dipole moment surface (DMS) Ames-2021-40K (X. Huang et al. 2023). To improve line position accuracy, (i) we employed the empirical band-centre correction approach (S. N. Yurchenko et al. 2011; B. P. Mant et al. 2018), shifting the diagonal vibrational Hamiltonian terms to match experimental values, followed by (ii) substitution of calculated energy levels with empirical ones from MARVEL studies (M. T. I. Ibrahim et al. 2024; D. Alatoom et al. 2024; A. A. A. Azzam et al. 2024, 2025a, b, c; S. A. M. Obaidata et al. 2025; M. H. I. Mansour et al. 2025), HITRAN (I. E. Gordon et al. 2022; E. V. Karlovets et al. 2021), and CDSD-2024-PI (R. Kochanov & V. Perevalov 2025). As part of this work, empirical energies of the minor CO_2 isotopologues were extracted from HITRAN using the MARVEL procedure.

The resulting line lists, in ExoMol format (J. Tennyson, C. Hill & S. N. Yurchenko 2013; J. Tennyson et al. 2024b), contain energies, Einstein A coefficients, lifetimes, uncertainties, and quantum numbers for 12 isotopologues of CO_2. Partition functions and broadening parameters are also provided. Associated opacities were computed for the four radiative transfer codes ARCiS (M. Min et al. 2020), TauREx (A. F. Al-Refaie et al. 2021), NEMESIS (P. G. J. Irwin et al. 2008), and petitRADTRANS (P. Mollière et al. 2019) using the ExoMolOP procedure (K. L. Chubb et al. 2021), both for individual isotopologues and for a composite opacity assuming terrestrial isotopic abundances. All data are available from the ExoMol database at www.exomol.com.

2 ROVIBRATIONAL STATES OF CO_2 ISOTOPOLOGUES: QUANTUM NUMBERS AND SELECTION RULES

Accurate description of the rovibrational states is essential for the construction and assignment of spectroscopic line lists. Quantum numbers provide the link between the variationally computed eigenstates of CO_2 and the spectroscopic notation used in experimental and atmospheric studies. For CO_2 and its isotopologues, several labelling schemes exist, reflecting both the linear triatomic structure of the molecule and the strong Fermi-resonance interactions that complicate vibrational assignments. In addition, nuclear-spin statistics and molecular symmetry play a central role in determining the allowed rovibrational levels and transition selection rules. In the following subsections we outline the quantum number conventions, symmetry classifications, and selection rules adopted in this work, following both spectroscopic practice and the requirements of large-scale automatic assignments.

Vibrational states of symmetric (non-linear) XY_2 molecules can be described by three normal modes, conventionally denoted as ν_1 (symmetric stretch), ν_2 (bend), and ν_3 (asymmetric stretch). For linear symmetric triatomic molecules such as CO_2, Herzberg's notation is commonly used to label vibrational states. The three modes are associated with vibrational quantum numbers v_1, $v_2^{\ell_2}$, and v_3. The superscript ℓ_2 denotes the vibrational angular momentum associated with the doubly degenerate bending mode in the linear configuration. This quantum number represents the projection of the vibrational angular momentum along the molecular axis and takes values from $-v_2$ to $+v_2$ in steps of two. In spectroscopic notation, however, the absolute value $|\ell_2|$ is often used.

The most widely adopted scheme for labelling vibrational states of CO_2 isotopologues is that of the Air Force Geophysics Laboratory (AFGL) (G. Amat & M. Pimbert 1965; L. S. Rothman & L. D. G. Young 1981; R. A. Toth et al. 2008). The AFGL notation employs a quintuplet of quantum numbers (m_1, m_2, l, m_3, r) designed to simplify the description of strong Fermi resonances in CO_2 of states ν_1 and $2\nu_2$, including their overtones and combination states, i.e. between states of the type $(v_1, (v_2+2)^{\ell_2}, v_3)$ and $(v_1+1, v_2^{\ell_2}, v_3)$. The vibrational quantum numbers (m_1, m_2, m_3) are identical for all members of the same resonance polyad, while the fifth index r ranks the states within it. In AFGL notation, m_2 and l are always equal, meaning $l \geq l_2$. By convention, for a given Fermi multiplet, r starts from 1 and increases by removing one quantum of ν_1 and adding two quanta of ν_2. For example, the Fermi-resonant states $(2, 0^0, 1)$, $(1, 2^0, 1)$, and $(0, 4^0, 1)$ correspond to (2 0 0 1 1), (2 0 0 1 2), and (2 0 0 1 3), respectively (R. A. Toth et al. 2008). The polyad number $P = 2v_1 + v_2 + 3v_3$ is approximately conserved within a Fermi multiplet.

For asymmetric isotopologues such as $^{16}O^{12}C^{17}O$, the AFGL quantum numbers follow the same scheme, $(J, m_1, m_2, m_3, r, e/f)$, even though the concepts of symmetric and asymmetric stretching lose their strict meaning. While AFGL notation provides a useful framework for analysing resonance clusters, it is not well suited for the automatic assignment of very large datasets.

2.1 Rotational quantum number and parity

The total angular momentum of CO_2 is denoted by J, which includes contributions from both overall molecular rotation and vibrational angular momentum ($\ell_2 \neq 0$).

The Pauli principle imposes restrictions on symmetric isotopologues containing zero-spin oxygen nuclei, namely $C^{16}O_2$ and $C^{18}O_2$. For these species, vibrational states with even v_3 (symmetric under oxygen exchange) are only associated with even J, while states with odd v_3 (antisymmetric) only allow odd J. This reflects the requirement that the total rovibrational wavefunction be symmetric under exchange of the two oxygen nuclei. The rotationless parity of a state is labelled e or f and is determined by the sum $J + \ell_2 + v_3$: states with even values correspond to e parity, and those with odd values to f parity.

For $C^{17}O_2$ ($^{12}C^{17}O_2$ and $^{13}C^{18}O_2$), the ^{17}O nuclei have non-zero spin ($I = 5/2$). In this case, the Pauli principle does not restrict

odd or even levels, and both ortho and para nuclear-spin symmetries exist. All J values are allowed, and bands with $\ell_2 > 0$ exhibit ℓ-type doubling, giving rise to both e and f parity levels for each J.

2.2 Symmetry group classification

An alternative classification of rovibrational states uses molecular symmetry groups (P. R. Bunker & P. Jensen 1998), specifically C_{2v}(M) or C_s(M), and their irreducible representations (irreps). For symmetric CO_2, states transform according to the C_{2v}(M) group, which has four irreps: A_1, A_2, B_1, and B_2.

For isotopologues with zero-spin oxygen nuclei, the Pauli principle forbids rovibrational states of B_1 and B_2 symmetry, leaving only A_1 and A_2, which correlate with e/f parity according to:

$$e : (-1)^J = 1, \quad A_1, \tag{1}$$

$$e : (-1)^J = -1, \quad A_2, \tag{2}$$

$$f : (-1)^{J+1} = 1, \quad A_2, \tag{3}$$

$$f : (-1)^{J+1} = -1, \quad A_1. \tag{4}$$

For symmetric CO_2 with non-zero oxygen nuclear spin, all four irreps (A_1, A_2, B_1, B_2) are allowed. These can be correlated with ortho and para spin symmetries as summarized in Table 1. For the e/f labels, the convention $p = (-1)^J$ for e and $p = (-1)^{J+1}$ for f is applied, where p denotes molecular parity.

For asymmetric isotopologues, the rovibrational states transform as A' or A'' in the C_s(M) group. Their relation to e/f parity follows the same rules as in the symmetric, zero-spin case:

$$e : (-1)^J = 1, \quad A', \tag{5}$$

$$e : (-1)^J = -1, \quad A'', \tag{6}$$

$$f : (-1)^{J+1} = 1, \quad A'', \tag{7}$$

$$f : (-1)^{J+1} = -1, \quad A'. \tag{8}$$

2.3 Selection rules

Rotational selection rules for linear molecules are $\Delta J = 0, \pm 1$. For CO_2, Q-branch transitions ($\Delta J = 0$) occur only in parallel bands and involve $e \leftrightarrow f$ transitions, where the dipole moment change is parallel to the molecular axis. P- and R-branch transitions ($\Delta J = \pm 1$) occur in both parallel and perpendicular bands and involve $e \leftrightarrow e$ or $f \leftrightarrow f$.

For symmetric isotopologues with non-zero oxygen nuclear spin, additional nuclear-spin selection rules apply:

ortho $\leftrightarrow$ ortho, para $\leftrightarrow$ para.

In terms of molecular symmetry, these correspond to

$A_1 \leftrightarrow A_2$, $B_1 \leftrightarrow B_2$.

For asymmetric isotopologues, the e/f selection rule is general for all dipole allowed transitions and therefore still holds; in group-theoretical notation it is expressed as

$A' \leftrightarrow A''$.

The nuclear-spin degeneracy factors for the isotopologues considered in this work are summarized in Table 1.

3 LINE LIST PRODUCTION

The generation of high-quality line lists requires three key ingredients: (i) accurate PES and DMS to describe molecular structure and transition intensities; (ii) robust nuclear-motion calculations to solve the rovibrational Schrödinger equation; and (iii) empirical refinement procedures to bring calculated energy levels into agreement with experimental data. In this section, we outline the methodology adopted for CO_2 isotopologues, including the choice of potential and dipole moment surfaces, the details of the variational calculations, and the empirical corrections applied to improve the accuracy of the final line lists.

3.1 Potential energy surface

We initially considered using the more recent empirical PES Ames-X01d of CO_2 developed by X. Huang et al. (2023), which was optimized to treat all isotopologues on an equal footing. The earlier PES, Ames-2 (X. Huang et al. 2017), originally developed for the main isotopologue $^{12}C^{16}O_2$, was also available and had previously been used to construct the UCL-4000 line list (S. N. Yurchenko et al. 2020). However, out tests suggested that the optimization in Ames-X01d came at the cost of somewhat degraded line list quality. For $^{12}C^{16}O_2$ and $^{13}C^{18}O_2$, this is illustrated in Fig. 1, which compares TROVE-calculated term values with experimentally derived values (MARVEL or CDSD-2024-PI (R. Kochanov & V. Perevalov 2025)) using both PESs. Ames-2 provides a more compact and flatter J-dependence, indicating more accurate equilibrium values and greater suitability for subsequent band-centre corrections. Our tests confirmed this trend across all 12 isotopologues, with Ames-2 consistently giving better agreement with MARVEL and HITRAN energies, at least using TROVE. We therefore adopted Ames-2 for all line lists reported in this work.

3.2 Dipole moment surface

For transition intensities we used the *ab initio* DMS Ames-2021-40K of CO_2 by X. Huang et al. (2023), the latest in the Ames series of dipole moment surfaces. In our previous line list, UCL-4000, the DMS of O. Polyansky et al. (2015) was employed.

3.3 Variational nuclear-motion calculations

The rovibrational Schrödinger equation was solved using the program TROVE (S. N. Yurchenko et al. 2007). Our methodology follows S. N. Yurchenko & T. M. Mellor (2020) and S. N. Yurchenko, T. Mellor & J. Tennyson (2024), employing the exact KEO in the bisector frame together with basis functions constructed from associated Laguerre polynomials. For full details the reader is referred to these works; here we summarize only the main steps.

First, vibrational $J = 0$ eigenfunctions $\Psi_{\lambda,L}^{(J=0,\Gamma_{\rm vib})}$ were obtained variationally for $L = 0 \ldots L_{\rm max}$ by diagonalizing the $J = 0$ Hamiltonian $\hat{H}^{(J=0)}$ in all relevant irreducible representations (A_1, A_2, B_1, B_2 or A', A'' for symmetric or asymmetric species, respectively). The primitive basis was constructed as a product of 1D vibrational functions $\phi_{n_1}(r_1)$, $\phi_{n_2}(r_2)$, and $\phi_{n_3}^L(\rho)$, where r_1 and r_2 are stretching coordinates and $\rho = \pi - \alpha$ with α the bond angle (S. N. Yurchenko & T. M. Mellor 2020).

For $J > 0$, rovibrational basis functions were formed as contracted, symmetrized products of the $J = 0$ vibrational functions and rigid-rotor functions:

$$\Psi_{\lambda,K}^{(J,\Gamma)} = \{\Psi_{\lambda,K}^{(J=0,\Gamma_{\rm vib})} | J, K, \Gamma_{\rm rot}\rangle\}^{\Gamma}, \tag{9}$$

Table 1. Nuclear spin statistical weights g_{ns} for CO_2 isotopologues. Values are given for the relevant irreducible representations of the molecular symmetry groups. These factors are applied in partition function calculations and line intensity normalization.

			g_{ns}					
Γ	p/o	p	626	636	828	838	727	737
A_1	para	1	1	2	1	2	15	30
A_2	para	−1	1	2	1	2	15	30
B_1	ortho	−1	0	0	0	0	21	42
B_2	ortho	1	0	0	0	0	21	42
Γ			627	637	628	638	728	738
A'		1	6	12	1	2	6	12
A''		−1	6	12	1	2	6	12

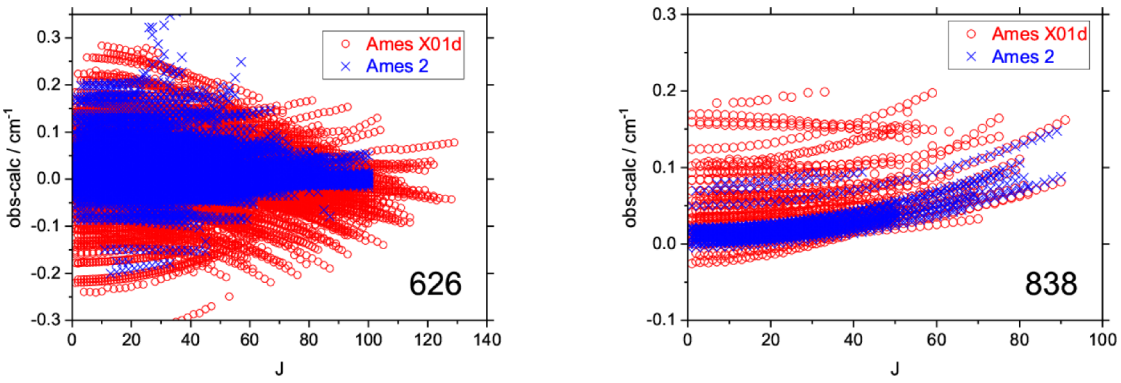

Figure 1. Obs.-Calc. residuals for $^{12}C^{16}O_2$ and $^{13}C^{18}O_2$ using two Ames PESs, Ames-2 (X. Huang et al. 2017), shown with red circles, and Ames-X01d (X. Huang et al. 2023), shown with blue crosses.

where $|J, K, \Gamma_{rot}\rangle$ are symmetrized rigid-rotor functions (S. N. Yurchenko, A. Yachmenev & R. I. Ovsyannikov 2017), K is constrained by L ($K = L$), and Γ, Γ_{vib}, Γ_{rot} are the total, vibrational, and rotational symmetries in C_{2v}(M) or C_s(M). Nuclear masses adopted were: $m_C = 11.99670909$ Da (^{12}C) and 13.000063355 Da (^{13}C); $m_O = 15.990525980$ Da (^{16}O), 16.99474312 Da (^{17}O), and 17.99477097 Da (^{18}O).

Rovibrational wavefunctions were computed up to J_{max} (see Table 2) and used to generate transition intensities for all dipole-allowed transitions. For $^{12}C^{16}O_2$, lower and upper state energies were limited to 16 000 and 36 000 cm^{-1}, respectively. For $^{13}C^{12}O_2$, the lower-state energy cut-off was set to 12 000 cm^{-1}. For all other isotopologues the lower-state cut-off was 10 000 cm^{-1}. Transition wavenumbers spanned 0–20 000 cm^{-1}. Nuclear spin statistical weights g_{ns} for all isotopologues are given in Table 1.

3.4 Empirical corrections of $J = 0$ energies

To improve calculated rovibrational energies, empirical band-centre corrections (EBCC; S. N. Yurchenko et al. 2009, 2011) were applied to the diagonal elements of the $J = 0$ Hamiltonian matrices. These corrections were obtained as average residuals between TROVE-calculated $J = 0$ band origins and empirical values (from MARVEL, CDSD-2024-PI, or HITRAN2020). For $^{12}C^{16}O_2$, the procedure is illustrated in Fig. 2.

Since $J = 0$ basis functions are eigenfunctions of the vibrational Hamiltonian, their diagonal elements can be straightforwardly replaced by empirical energies, thereby propagating improvements to higher-J rovibrational levels (S. N. Yurchenko et al. 2009; B. P. Mant et al. 2018). These also include states with $\ell_2 > 0$, which formally do not exist in Nature but are present in the basis set. This significantly reduced obs.-calc. residuals across most isotopologues. The exception was $^{16}O^{12}C^{18}O$, where some bands diverged at high J after EBCC (Fig. 2, green crosses). For this isotopologue, band-centre shifts were not applied.

3.5 MARVELization procedure

Further improvements were made using the hybrid MARVELization approach (L. K. McKemmish et al. 2024), in which calculated energies are replaced, where available, by more accurate values derived from MARVEL or effective Hamiltonian (EH) analyses. MARVEL energies were taken from the recent MARVEL studies of CO_2 isotopologues (M. T. I. Ibrahim et al. 2024; D. Alatoom et al. 2024; A. A. A. Azzam et al. 2024, 2025a, b, c; S. A. M. Obaidata et al. 2025; M. H. I. Mansour et al. 2025), while effective Hamiltonian energies were taken from CDSD-2024-PI (R. Kochanov & V. Perevalov 2025) ($^{12}C^{16}O_2$) and HITRAN2020 (I. E. Gordon et al. 2022) (all minor isotopologues). The EH energies are used to minimize gaps in available MARVEL energy levels.

To this end, the EH energies were generated using the MARVEL algorithm applied to HITRAN transitions (T. Furtenbacher, A. G. Császár & J. Tennyson 2007; R. Tobias et al. 2020; P. Arendas, T. Furtenbacher & A. G. Császár 2020, 2024). The HITRAN uncertainness were set based on the corresponding HITRAN error code (I. E. Gordon et al. 2022), for example, for code 4 (0.0001 cm^{-1} $\leq \epsilon <$ 0.001 cm^{-1}), the uncertainty of 0.0005 cm^{-1} was assumed. For the main isotopologue $^{12}C^{16}O_2$, the CDSD-2024-PI energies were readily provided by R. Kochanov & V. Perevalov (2025). The MARVEL procedure ensures self-consistency of the derived network, with uncertainties propagated through the bootstrap algorithm (J. Tennyson et al. 2024a). While MARVEL is usually applied to directly measured transitions, here it is used on effective Hamiltonian-derived frequencies to supplement the empirical data set.

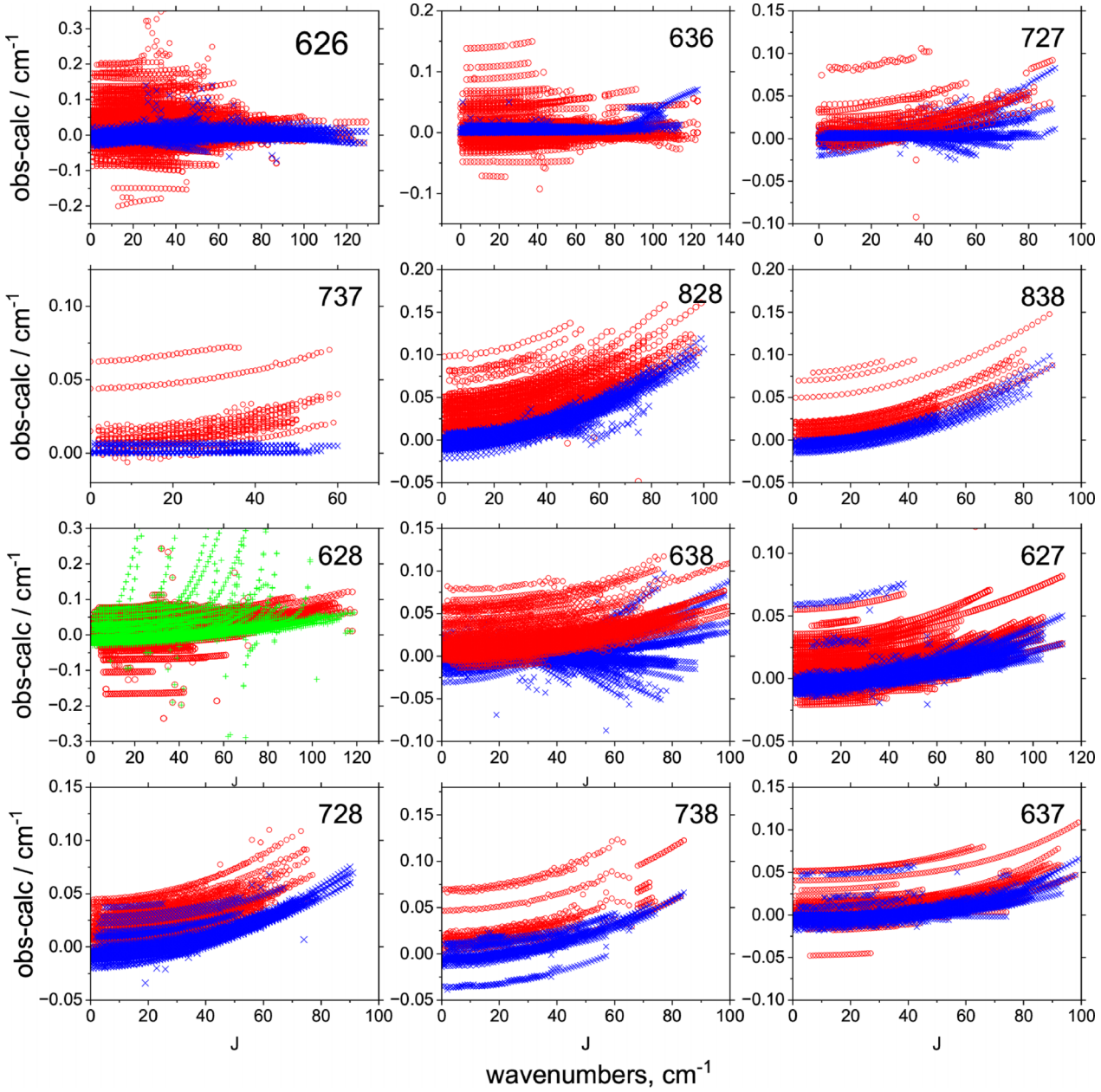

Figure 2. 'Obs.-calc'. residuals for different isotopologues using two Ames PESs, Ames-2 (X. Huang et al. 2017), before (red circles) and after (blue crosses) the band centre corrections (S. N. Yurchenko et al. 2009). The green crosses in the display 628 show the divergence of the band-centre-corrected energies with J increases (see the text).

In the ExoMol states files, purely MARVEL-derived levels are labelled 'Ma', HITRAN-derived levels as 'HI', and effective Hamiltonian values as 'EH' [see the description of the hybrid MARVELization approach of L. K. McKemmish et al. (2024)]. For the main isotopologue $^{12}C^{16}O_2$, MARVEL data from A. A. A. Azzam et al. (2025a) were supplemented by CDSD-2024-PI (R. Kochanov & V. Perevalov 2025). The HITRAN energies generated using the MARVEL approach are provided as part of the supplementary material.

Table 2 also compares the number of lines for each isotopologue with a transition intensity greater than 10^{-30} cm molecule^{-1} at 296 K. For some reason which remains unclear, given that our computed line lists are complete, HITRAN contains slightly more lines for each isotopologue.

4 ROVIBRATIONAL LINE LISTS FOR CO_2 ISOTOPOLOGUES

Using the TROVE rovibrational wavefunctions and the *ab initio* DMS Ames-2021-40K, we generated 12 line lists for CO_2 isotopologues, collectively referred to as Dozen, in the ExoMol format (J. Tennyson et al. 2024b). The coverage extends from 0 to 20 000 cm^{-1} (i.e. wavelengths ≥ 0.5 μm). For the parent and most important isotopologue, $^{12}C^{16}O_2$, we adopted a more extensive coverage of state excitations: a higher lower-state energy cut-off $E''_{\rm max}$ of 16 000 cm^{-1} (cf. 10 000–12 000 cm^{-1} for other isotopologues). A larger $E''_{\rm max}$ increases computational cost but improves population coverage of excited states and thus yields a larger, more complete line list (see

Table 2. Summary of CO_2 isotopologue line lists. L_{max} is the maximum vibrational angular momentum; J_{max} the maximum rotational quantum number; E''_{max} the maximum lower-state energy; N_{state} the number of states in the ExoMol `.states` file; N_{lines} the number of transitions in the `.trans` file; N_{HI} the number of HITRAN lines at $T = 296$ K; N_{Ca} the number of calculated ExoMol lines at $T = 296$ K; N_{Ma} the number of substituted lines with MARVEL values at $T = 296$ K; N_{Ma+HI} the total number of 'MARVELized' lines (Ma/HI or EH) at $T = 296$ K; and the terrestrial isotopic abundance from HITRAN (I. E. Gordon et al. 2022) is given in the final column. Only transitions stronger than $I = 10^{-30}$ cm molecule^{-1}) at $T = 296$ K are counted.

Iso	L_{max}	J_{max}	E''_{max} (cm^{-1})	N_{state}	N_{lines}	N_{HI} 296 K	N_{Ca} 296 K	N_{Ma} 296 K	N_{MA+HI} 296 K	Abundance (HITRAN)
626	30	250	16 000	3646 814	3607 555 029	175 891	174445	17 5109	17 5109	9.8420×10^{-1}
627	10	150	10 000	5065 351	1042 026 128	73 221	72228	36 582	73 082	7.3399×10^{-4}
628	20	150	10 000	7369 500	1112 362 652	116 222	11 8701	58 841	115 774	3.9471×10^{-3}
636	30	150	12 000	2670 885	753 056 397	70 884	69 870	38 525	70 779	1.1057×10^{-2}
637	10	150	10 000	7026 133	1169 286 790	22 830	22 577	4572	22 812	8.2462×10^{-6}
638	10	150	10 000	4072 374	1246 979 498	40 304	39 950	12 592	40 240	4.4345×10^{-5}
727	20	150	10 000	5168 133	583 350 654	6558	6493	5372	6551	1.3685×10^{-7}
728	16	150	10 000	6897 362	1141 805 035	5031	14 378	4186	5030	1.3685×10^{-7}
737	20	150	10 000	5133 983	92 365 743	1639	1501	903	1533	1.5375×10^{-9}
738	10	150	10 000	7432 457	1273 579 199	3580	3569	1288	3578	1.6535×10^{-8}
828	20	150	10 000	2527 339	321 570 612	10 541	1049	6267	10 520	3.9573×10^{-6}
838	20	150	10 000	2721 734	361 867 203	3053	2926	1146	2966	4.4460×10^{-8}

Table 2). Further control of the excitation coverage is provided by J_{max} and L_{max}, which were selected per isotopologue (Table 2). The choice of L_{max} affects the size of the basis set and thus its quality.

Each line list comprises a `.states` file, a set of `.trans` files, and a partition-function file `.pf`. As an example, an extract from the $^{12}C^{16}O_2$ `.states` file is given in Table 3. It contains state term values $\tilde{E}_i$, total degeneracies g_i, uncertainties (`unc`), lifetimes, and quantum numbers. Vibrational labels are provided in three conventions: Herzberg (v_1, $v_2^{\ell_2}$, v_3), AFGL (m_1, m_2, l_2, m_3, r), and TROVE local-mode indices (n_1, n_2, n_3, L). Both rotationless parity (e/f) and molecular symmetry (A_1, A_2, B_1, B_2, A', A'') are provided. States are indexed by an integer i (state ID), used in the transition files.

The `.states` files are MARVELized: where available, calculated term values are replaced by empirical values, MARVEL (`Ma`), HITRAN (`HI`), or effective-Hamiltonian (CDSD; `EH`), with the label supplied in the penultimate column; the original calculated value is retained in the final column. State-dependent lifetimes computed with Dozen are included (column 6; see Table 3).

For practicality, each isotopologue's transition set is divided into 20 `.trans` files, each covering a 1000 cm^{-1} bin over 0–20 000 cm^{-1}. An extract from a $^{12}C^{16}O_2$ `.trans` file is shown in Table 4, listing Einstein A coefficients and the upper/lower state IDs.

4.1 Using machine learning to reconstruct spectroscopic quantum numbers

In order to help with the usage of the Dozen line list, quantum numbers for all three assignment schemes are provided as part of the States file. While the TROVE quantum numbers are generated using the largest contribution to the associated eigenfunctions, to assign vibrational quantum numbers for AFGL and Herzberg they are not known, we developed a machine learning (ML) pipeline. To this end, a classifier was implemented in Pytorch (A. Paszke et al. 2019) using a multi-head neural network architecture containing a shared backbone and specialized output heads.

The ML model starts with a shared feature extraction network consisting of 6 fully connected layers (512 → 256 → 128 → 128 → 64 neurons) with GELU activations and dropout regularization to learn common representations from 32 molecular features (energy, quantum numbers, dipole moments, isotopic masses, symmetries). From the 64-dimensional shared representation, 8 separate classification heads branch out, each containing 3 additional layers (64 → 32 → output classes) to predict the specific quantum numbers (Herzberg: v_1, v_2, l_2, and v_3 and AFGL: m_1, m_2, m_3, and r). This design enables the backbone to capture shared molecular physics, while each head specializes in its respective quantum-number assignment task. Crucially, we enforce quantum-mechanical constraints within the loss function, limiting l_2 values according to the parity of v_2.

We assess model uncertainty using Monte Carlo (MC) Dropout (Y. Gal & Z. Ghahramani 2016), which keeps dropout layers active during inference by effectively sampling multiple network realizations. We perform 50 forward passes per sample, obtaining a distribution of class probabilities. From the mean probability vector $\bar{p}$, we compute predictive entropy: $H(\bar{p}) = -\sum \bar{p} \log \bar{p}$, normalized to [0,1] by dividing by $\log(n_{classes})$. A normalized entropy threshold of 0.4, determined from prior validation to optimize the trade-off between accepted correct predictions and rejected incorrect ones, was applied. Predictions below this threshold were accepted, enabling the assignment of over one million spectral lines across all targeted isotopologues.

For different isotopologues, the ML procedure resulted in 80 000–120 000 states additionally assignments with AFGL and Herzberg. While this represents only 1.4–3 per cent of the total number of states in each data set, this is a multifold increase comparing to the empirically assignment states, which is illustrated in Fig. 3.

4.2 Partition functions

Partition functions were computed for all 12 isotopologues on a 1 K grid from 0 to 5000 K using the direct sum

$$Q(T) = \sum_i g_{ns}^{(i)} \left(2J_i + 1\right) \exp\left(-\frac{c_2 \tilde{E}_i}{T}\right),$$

where $g_{ns}^{(i)}$ is the nuclear-spin statistical factor (Table 1), J_i is the rotational quantum number, c_2 is the second radiation constant (cm K), and $\tilde{E}_i$ is the energy term value (cm^{-1}) relative to the ground state ($J = 0$, $v_1 = v_2 = v_3 = 0$, $\ell_2 = 0$, e).

Fig. 4 compares the Dozen partition function for $^{12}C^{16}O_2$ with TIPS 2024 (total internal partition sum) of HITRAN2020 (R. R.

Table 3. Extract from the `.states` file of $^{12}C^{16}O_2$ Dozen line list.

i	$\tilde{E}$/cm^{-1}	g	J	unc/cm^{-1}	τ/s	$\Gamma_{\rm tot}$	e/f	v_1	v_2	l_2	v_3	$\lvert C_i^2\rvert$	m_1	m_2	l_2	m_3	r	n_1	n_2	n_3	Label	$\tilde{E}_{\rm calc.}$/cm^{-1}
1	0.000000	1	0	0.000001	inf	A1	e	0	0	0	0	1.00	0	0	0	0	1	0	0	0	Ma	0.000000
2	1285.408 294	1	0	0.000049	7.7985E-01	A1	e	0	2	0	0	0.54	1	0	0	0	2	0	0	1	Ma	1285.407 001
3	1388.184 102	1	0	0.000053	5.5304E-01	A1	e	1	0	0	0	0.54	1	0	0	0	1	1	0	0	Ma	1388.184 019
4	2548.363 900	1	0	0.001515	3.5601E-01	A1	e	1	2	0	0	0.47	2	0	0	0	3	1	0	1	Ma	2548.364 344
5	2671.143 088	1	0	0.001119	4.8854E-01	A1	e	2	0	0	0	0.66	2	0	0	0	2	1	1	0	Ma	2671.142 828
6	2797.136 399	1	0	0.001000	2.3234E-01	A1	e	1	2	0	0	0.52	2	0	0	0	2	1	0	1	EH	2797.135 151
7	3792.682 402	1	0	0.001000	2.3664E-01	A1	e	1	4	0	0	0.43	3	0	0	0	3	1	0	2	EH	3792.680 795
	–			–			–			–												
18	6016.690 108	1	0	0.001000	1.2107E-03	A1	e	1	0	0	2	0.60	1	0	0	2	0	0	3	0	EH	6016.701 190
19	6240.057 942	1	0	0.080000	1.5052E-01	A1	e	2	6	0	0	0.30	5	0	0	0	4	1	1	3	Ca	6240.057 942
20	6435.500 064	1	0	0.060000	1.7500E-01	A1	e	4	2	0	0	0.28	5	0	0	0	4	2	2	1	Ca	6435.500 064
21	6588.319 980	1	0	0.050000	2.2609E-01	A1	e	5	0	0	0	0.48	5	0	0	0	2	3	2	0	Ca	6588.319 980
	...			...			...			...												
30	7834.845 932	1	0	0.040000	1.7933E-01	A1	e	1	0	0	3	0.29	−1	−1	−1	−1	−1	2	2	0	Ca	7834.845 932
31	7974.351 788	1	0	0.040000	1.7354E-01	A1	e	1	0	0	3	0.51	−1	−1	−1	−1	−1	2	2	0	Ca	7974.351 788

Notes. i: State identifier (integer index used in `.trans` files).
$\tilde{E}$: State term value (cm^{-1}).
g: Total state degeneracy.
J: Total rotational quantum number.
unc: Energy uncertainty (cm^{-1}).
τ: State lifetime (s).
$\Gamma_{\rm tot}$: Total symmetry in C_{2v}(M) or C_s(M).
e/f: Rotationless parity label.
v_1, $v_2^{\ell_2}$, v_3: Normal-mode vibrational quantum numbers (Herzberg notation).
$|C_i|^2$: Largest expansion-coefficient weight used for assignment.
m_1, m_2, l_2, m_3, r: AFGL vibrational quantum numbers (−1 stands for non-available).
n_1, n_2, n_3, L: TROVE local-mode quantum numbers.
Label: MARVEL (`Ma`), Effective Hamiltonian (`EH`), HITRAN (`HI`), or calculated Dozen value (`Ca`).
$\tilde{E}_{\rm calc.}$: Original calculated Dozen term value (cm^{-1}).

Table 4. Extract from a `.trans` file of the $^{12}C^{16}O_2$ Dozen line list.

f	i	A_{fi}
872 090	849 449	3.7641e−16
1729 336	1740 374	1.9742e−09
443 590	423 498	2.7506e−15
89 228	95 628	1.9493e−12
1562 467	1577 182	1.1399e−08
1031 039	1081 028	3.1880e−14
1710 529	1691 051	4.1274e−14
2476 483	2488 556	1.5718e−09
1139 554	1151 915	5.3921e−11
1726 661	1675 412	4.8961e−11

Notes. f: Upper state counting number;
i: Lower state counting number;
A_{fi}: Einstein A coefficient (in s^{-1}).

Gamache et al. 2025) and with that obtained from the Ames AI-3000K line list (X. Huang et al. 2023). The Dozen and Ames curves agree closely and both lie systematically above the HITRAN TIPS values, reflecting improved completeness of the underlying states; the lack of convergence of the HITRAN TIPS partition function at high temperature has been noted previously (R. Wang et al. 2023).

4.3 Opacities

Temperature- and pressure-dependent opacities for all 12 isotopologues of CO_2, based on the Dozen line lists, were generated using the ExoMolOP procedure (K. L. Chubb et al. 2021). Opacity tables were produced for four widely used atmospheric retrieval codes: ARCiS (M. Min et al. 2020), TauREx (A. F. Al-Refaie et al. 2021), NEMESIS (P. G. J. Irwin et al. 2008), and petitRADTRANS (P. Mollière et al. 2019), and are provided alongside the line lists. In addition, we supply combined opacities constructed for a representative atmosphere containing multiple isotopologues in terrestrial abundance ratios (see Table 2).

4.4 Line broadening coefficients

Accurate treatment of collisional line broadening is essential for radiative-transfer modelling. We adopt a semi-empirical approach, combining theoretical predictions with carefully selected experimental data, to provide broadening and shifting coefficients for CO_2 transitions with common perturbers (N_2, O_2, CO_2, H_2, He, and H_2O) over the temperature range 200–2000 K.

Previous analyses have highlighted inconsistencies among experimental data sets. For example, R. R. Gamache & J. Lamouroux (2013) and R. R. Gamache et al. (2014) showed that different experiments on the same transition often disagree beyond the quoted uncertainties, complicating attempts to assess vibrational or isotopologue dependences of the half-width coefficients γ. More recent work by R. Hashemi et al. (2020) re-analysed the most reliable data, proposing a new set of semi-empirical parameters for CO_2- and air-broadening, suitable for Voigt and speed-dependent Voigt profiles, and found no significant vibrational dependence of γ_0.

Isotopologue effects are also small. For example, V. M. Devi et al. (1998) found no statistically significant difference between N_2-broadening of $^{16}O^{13}C^{18}O$ and that of the main isotopologue, while R. R. Gamache & J. Lamouroux (2013) estimated that isotopic mass effects reduce γ by less than 1 per cent. This was recently confirmed by cavity ringdown measurements of $^{13}C^{16}O_2$ by D. Mondelain et al. (2025), who found air-broadening coefficients only 0.4 per cent smaller than those of $^{12}C^{16}O_2$.

Given the very weak vibrational and isotopologue dependences, we provide only J-dependent broadening coefficients γ_0(296 K) and temperature exponents n for the single power-law parametrization. Speed dependence and line-mixing are not considered. The adopted γ_0 and n values are taken from HITRAN for broadening by H_2, He (Y. Tan et al. 2022), N_2, CO_2 (I. E. Gordon et al. 2022), and H_2O (Y. Tan et al. 2019), and are supplied in the `m0`-diet format for use with ExoCross (J. Tennyson et al. 2024b). In addition, we provide more recent theoretical water-broadening data (B. Vispoel & R. R. Gamache 2024) in the `m2`-diet format (A. Sokolov et al. 2025), which employs a double power-law to describe the temperature dependence of $\gamma(T)$ more accurately.

5 ILLUSTRATIONS AND ANALYSIS

In this section, we demonstrate the performance of the new line lists by comparing their predicted spectra with previous data sets and with available experimental information. A general overview is provided in Fig. 5, which shows spectra of all isotopologues at $T = 1000$ K on a logarithmic scale, without scaling to isotopic abundances.

Fig. 6 illustrates the coverage and quality of the Dozen $^{12}C^{16}O_2$ line list at $T = 2000$ K, compared with our previous UCL-4000 list and the Ames AI-3000K list (X. Huang et al. 2023). The Dozen results are in excellent agreement with AI-3000K. We also compared with HITEMP 2025 (R. J. Hargreaves et al. 2024), which was generated from AI-3000K using the super-lines and super-energies technique (R. J. Hargreaves et al. 2020). As expected, the HITEMP spectrum follows AI-3000K closely over 0-12 000 cm^{-1}, which is the range adopted by HITEMP for $^{12}C^{16}O_2$.

Compared to UCL-4000, the new line list exhibits more physical behaviour at higher wavenumbers. This was achieved by optimising the stretching basis set comparing to that used in S. N. Yurchenko et al. (2020) by extending the bond-length grid used in the numerical integrations from 2.2 to ∼2.6 Å. The wavenumber behaviour of Dozen is now consistent with the near-integrated dipole limit (NIDL) (E. S. Medvedev & V. G. Ushakov 2022; E. S. Medvedev et al. 2020), which predicts that overtone intensities decrease approximately exponentially, appearing as straight lines in a log-intensity plot (E. S. Medvedev et al. 2016). By contrast, the formation of plateaux at high overtones is a well-known indicator of numerical artefacts, as seen in UCL-4000 above 14 000 cm^{-1}.

A more subtle intensity problem in UCL-4000 was identified by A. A. Balashov et al. (2024). Transitions in the 700 nm region (∼14 300 cm^{-1}) were predicted with intensities above 10^{-30} cm molecule^{-1}, yet an extensive experimental search found no evidence for them. This band was therefore significantly overestimated in UCL-4000. The present line list, based on the improved basis set and a different DMS, brings these intensities below the 10^{-30} cm molecule^{-1} threshold value, as shown in Fig. 7.

Differences among isotopologues are illustrated in Fig. 8, which shows $T = 296$ K spectra scaled by natural terrestrial abundances. Apart from the dominant $^{12}C^{16}O_2$, only $^{13}C^{16}O_2$ and $^{16}O^{12}C^{18}O$ produce IR features of sufficient strength to be detectable under atmospheric conditions. Of course, this situation will change with the altered isotope abundances encountered elsewhere in space.

Finally, Figs 9 and 10 present stick spectra at $T = 296$ K for all 12 isotopologues of CO_2, separated into symmetric and asymmetric species. MARVELized transitions are highlighted, illustrating the extent of experimental coverage. Comparisons with HITRAN2020 (and CDSD-2024 for $^{12}C^{16}O_2$) further demonstrate the accuracy and

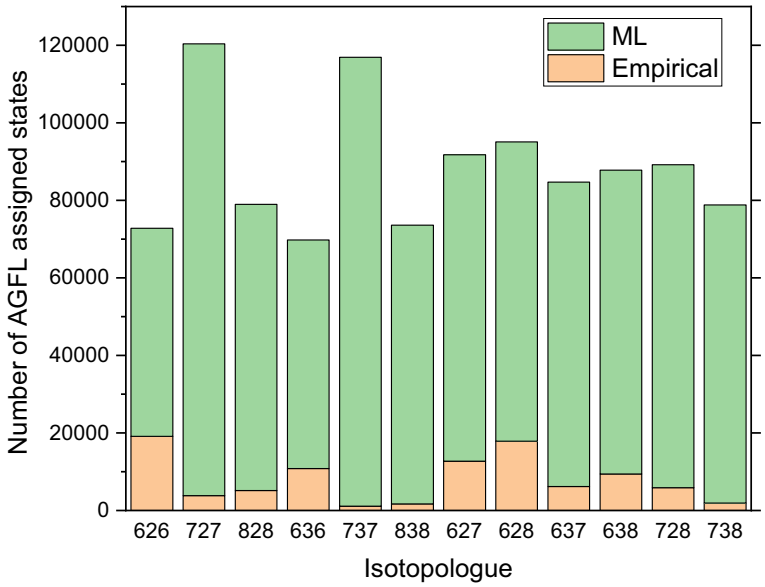

Figure 3. Coverage of the assigned states with the AFGL quantum numbers per isotopologue comparing to the corresponding empirical assignments: MARVEL, HITRAN or CDSD-2024-PI.

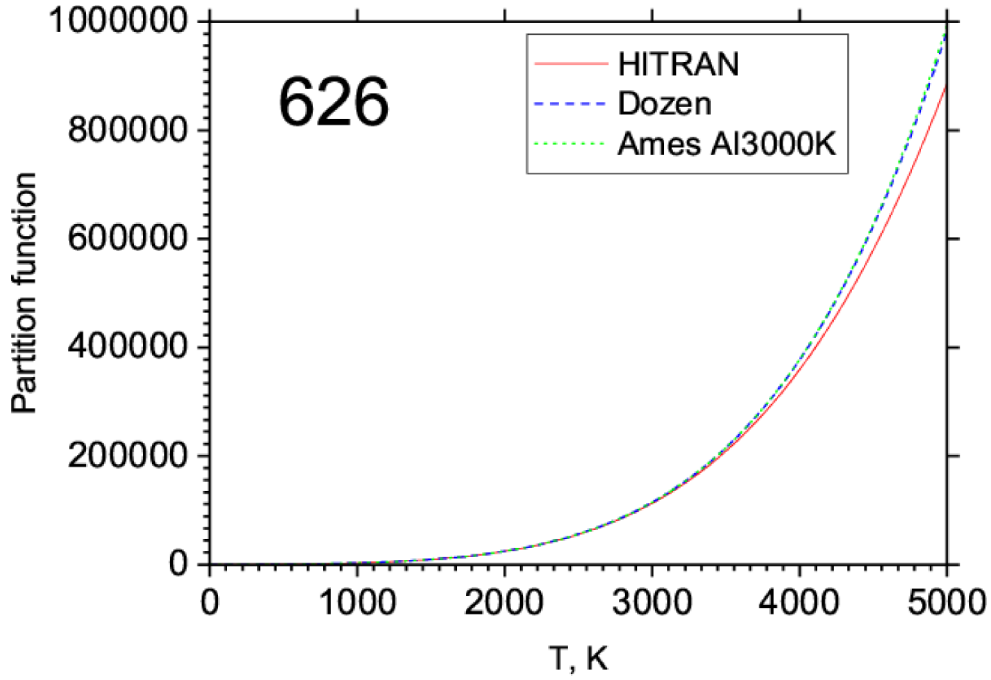

Figure 4. Partition functions for $^{12}C^{16}O_2$: comparison of Dozen with TIPS 2024 of HITRAN2024 (R. R. Gamache et al. 2025) and Ames AI-3000K.

completeness of the new line lists. A quantitative summary of line list statistics, including the number of empirically anchored transitions above the HITRAN intensity threshold of 10^{-30} cm molecule^{-1} at $T = 296$ K, is given in Table 2.

6 CONCLUSIONS

This work presents a new set of comprehensive rovibrational line lists for 12 isotopologues of CO_2 collectively know as the Dozen line list. These line lists were computed using accurate empirical PESs and the latest *ab initio* DMSs from the Ames group (X. Huang et al. 2017, 2023), covering the spectroscopic range up to 20 000 cm^{-1}. The line list for the main isotopologue $^{12}C^{16}O_2$ is applicable up to at least 3000 K, while those for the minor isotopologues are reliable up to about 2000 K. The accuracy of the datasets has been enhanced through empirical band-centre corrections and systematic MARVELization, incorporating experimental energies from MARVEL, HITRAN, and CDSD where available.

The new ExoMol CO_2 line lists represent a significant improvement over previous work. In particular, we now provide hot line lists for 12 isotopologues with an extended wavenumber coverage. Besides, spectra generated using Dozen display physically correct intensity behaviour for the high overtone bands thus avoiding the

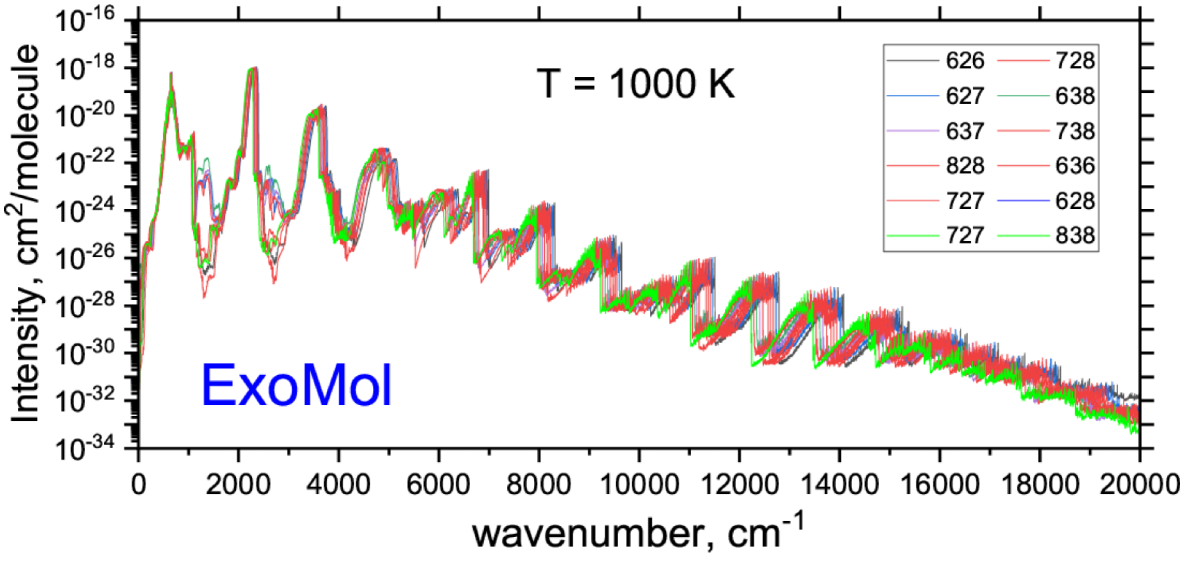

Figure 5. Spectra of CO_2 isotopologues at $T = 1000$ K (Gaussian profile, HWHM $= 1$ cm^{-1}). Intensities are not scaled to isotopic abundances.

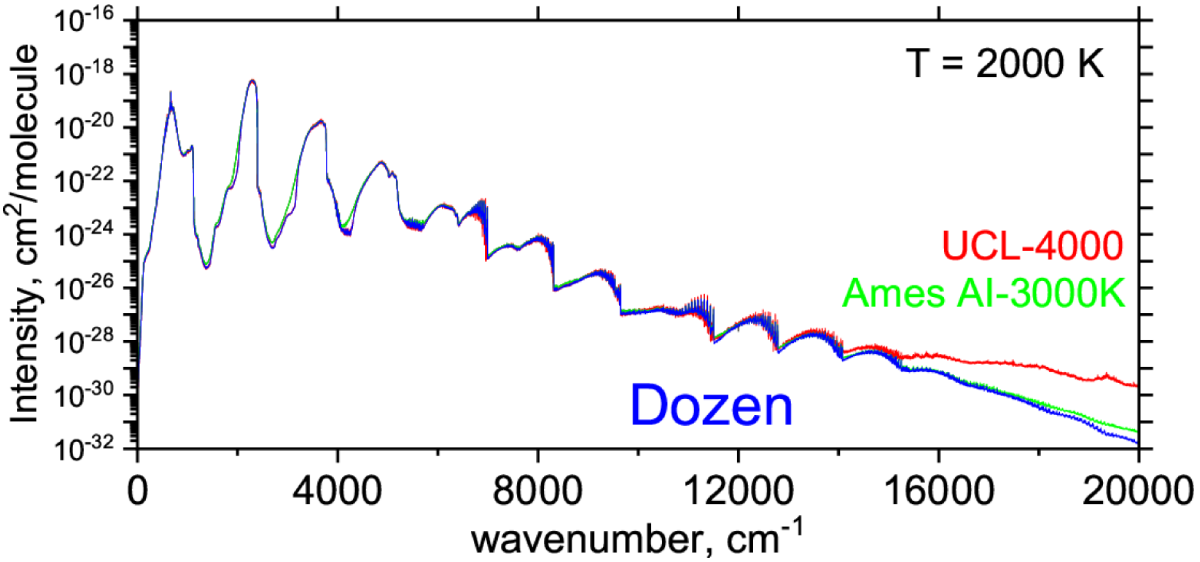

Figure 6. Cross-sections of $^{12}C^{16}O_2$ at $T = 2000$ K computed with UCL-4000, Dozen, AI-3000K, and HITEMP 2025 (Gaussian profile, HWHM $= 1$ cm^{-1}). The HITEMP 2025 and AI-3000K (X. Huang et al. 2023) spectra coincide over 0-12 000 cm^{-1}.

spurious plateau observed in UCL-4000 (S. N. Yurchenko et al. 2020). The intensity anomaly in the 700 nm region, is resolved by improving the stretching basis set used. Comparisons with HITRAN spectra at $T = 296$ K show excellent agreement, confirming both the reliability of line positions and the accuracy of intensities.

For the first time, we employed machine-learning methods to reconstruct AFGL and 'Herzberg' quantum numbers for CO_2 isotopologues, significantly extending the quantum assignment coverage in a consistent and physically constrained manner. We are currently using machine-learning techniques to improve the prediction transition wavenumbers for isotopologues (M. G. Barnfield et al. 2026); these results will be used to update Dozen states files for minor isotopologues with improved estimates for their energy levels in the near future. Partition functions have been computed on a fine temperature grid up to 5000 K, ensuring robust thermodynamic and opacity applications. Pressure-broadening coefficients with the most important perturbers (N_2, O_2, CO_2, H_2, He, H_2O, and air) have been collected from the literature and are provided in ExoMol's diet formats (`m0` and `m2`), enabling flexible use in radiative-transfer models.

Opacities were generated with ExoMolOP (K. L. Chubb et al. 2021) for four leading atmospheric retrieval codes (ARCiS, TauREx, NEMESIS, and petitRADTRANS). Combined opacity grids, scaled to terrestrial isotopic abundances, are also supplied for practical applications in planetary and stellar atmosphere modelling.

The line lists and associated data (partition functions, broadening parameters, opacities) are available in the ExoMol data base at www.exomol.com. Together, they provide the most complete and

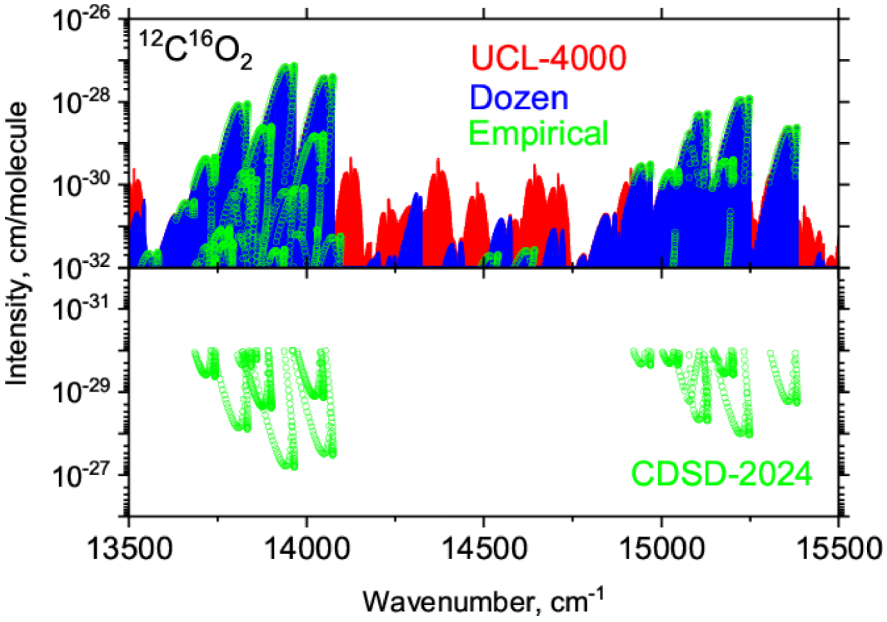

Figure 7. Intensity behaviour near 700 nm ($\sim$14 300 cm^{-1}). UCL-4000 overestimates this band, while the new Dozen list based on Ames-2021-40K yields significantly lower intensities, in better agreement with the experiment of A. A. Balashov et al. (2024).

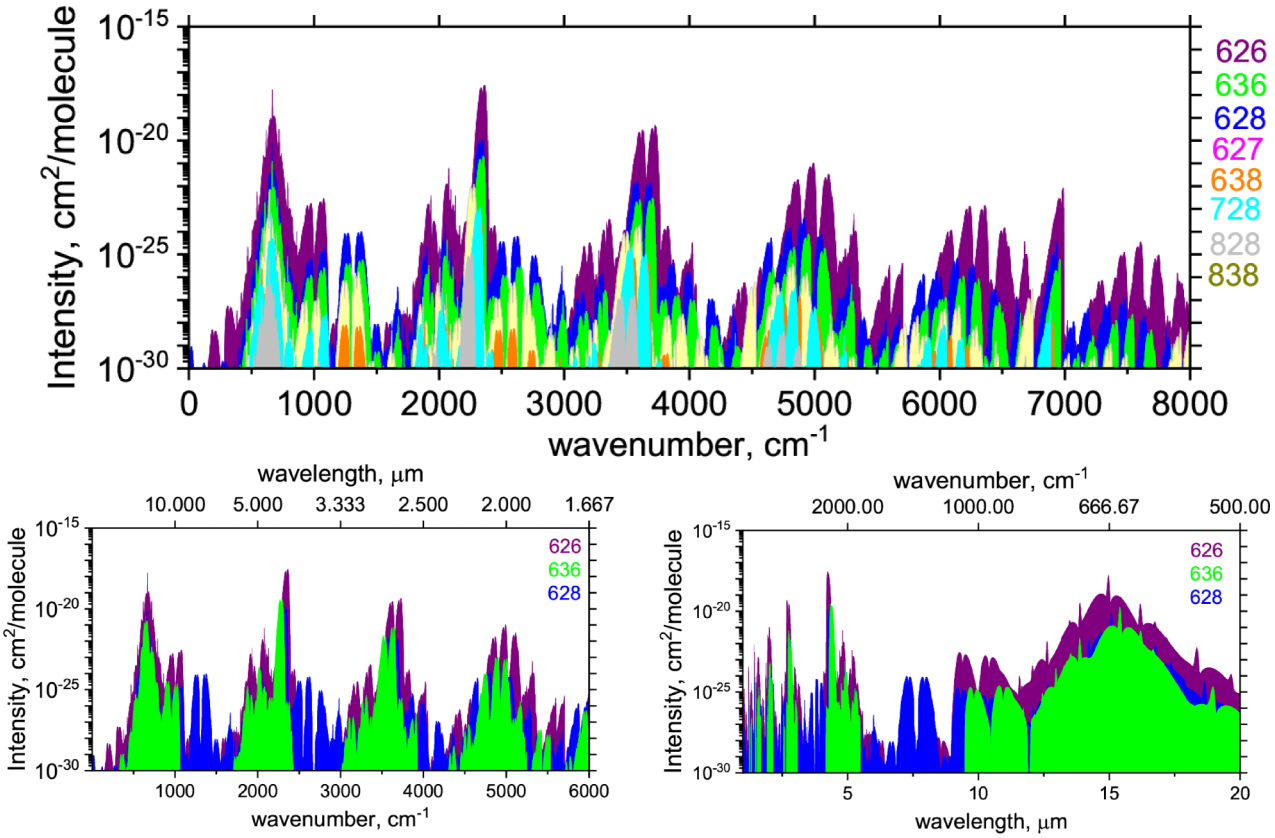

Figure 8. Spectra of selected CO_2 isotopologues at $T = 296$ K, scaled by terrestrial isotopic abundances. In addition to $^{12}C^{16}O_2$, only $^{13}C^{16}O_2$ and $^{16}O^{12}C^{18}O$ produce detectable IR features.

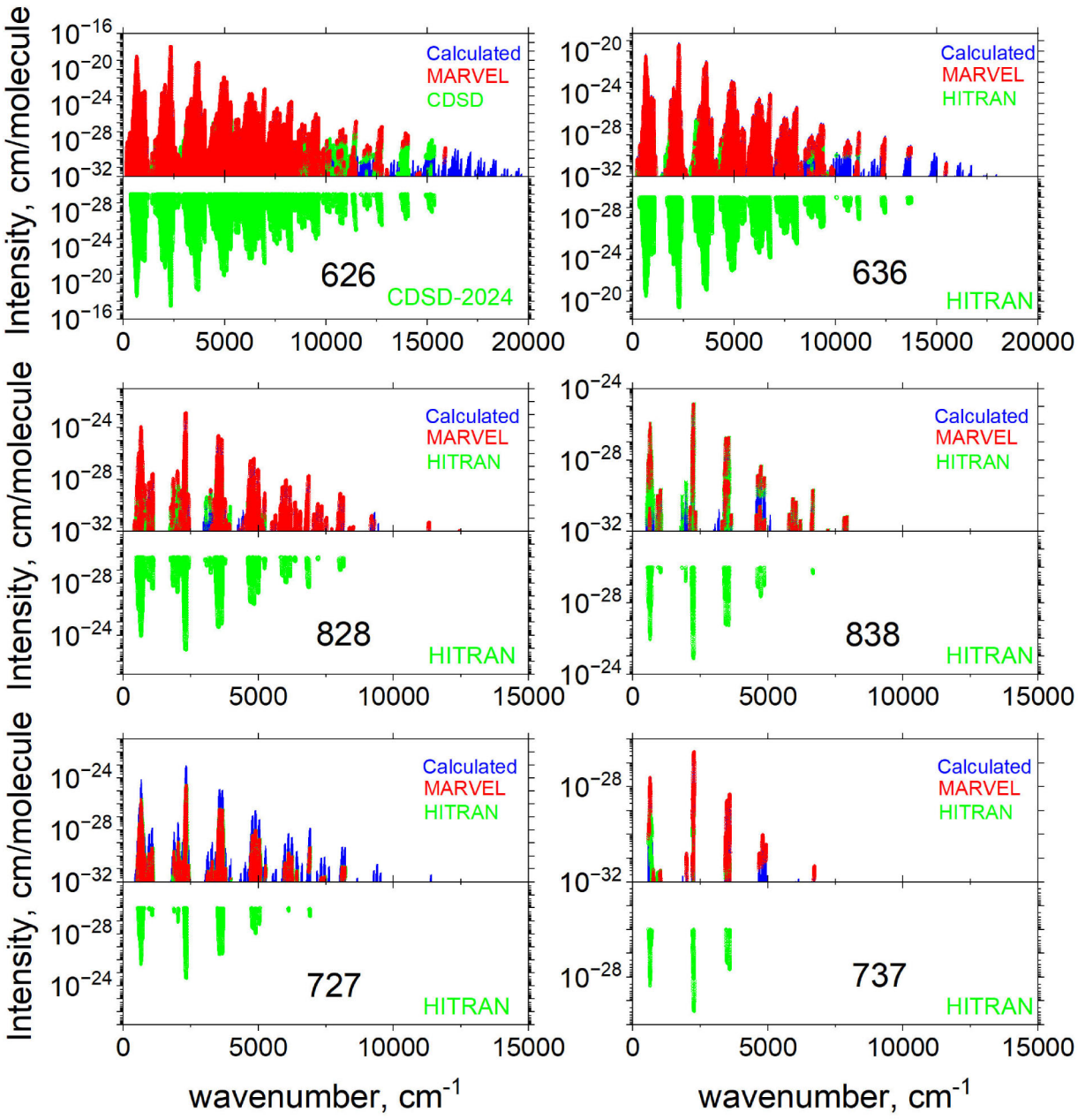

Figure 9. Stick spectra of six symmetric isotopologues of CO_2 at $T = 296$ K (absorption coefficients in cm molecule^{-1}). MARVELized transitions are highlighted, and comparisons are made with CDSD-2024 ($^{12}C^{16}O_2$) and HITRAN2020. Intensities are not scaled to isotopic abundances.

accurate spectroscopic resource to date for CO_2 and its isotopologues. These data sets will be invaluable for atmospheric retrievals across the Solar system, brown dwarfs, and exoplanets, as well as for laboratory and terrestrial remote-sensing studies. Future work will focus on extending the spectral range into the ultraviolet and further refining high-temperature coverage.

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

This work was supported by the STFC Projects ST/Y001508/1 and UKRI/ST/B001183/1. The authors acknowledge the use of the Cambridge Service for Data Driven Discovery (CSD3) and the DiRAC Data Intensive service DIaL2.5 at the University of Leicester, managed on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). These DiRAC services were funded by BEIS, UKRI and STFC capital funding and STFC operations grants. DiRAC is part of the UKRI Digital Research Infrastructure. We thank Xinchuan Huang for help accessing AI-3000K and providing valuable suggestions. This work was also supported by the European Research Council (ERC)

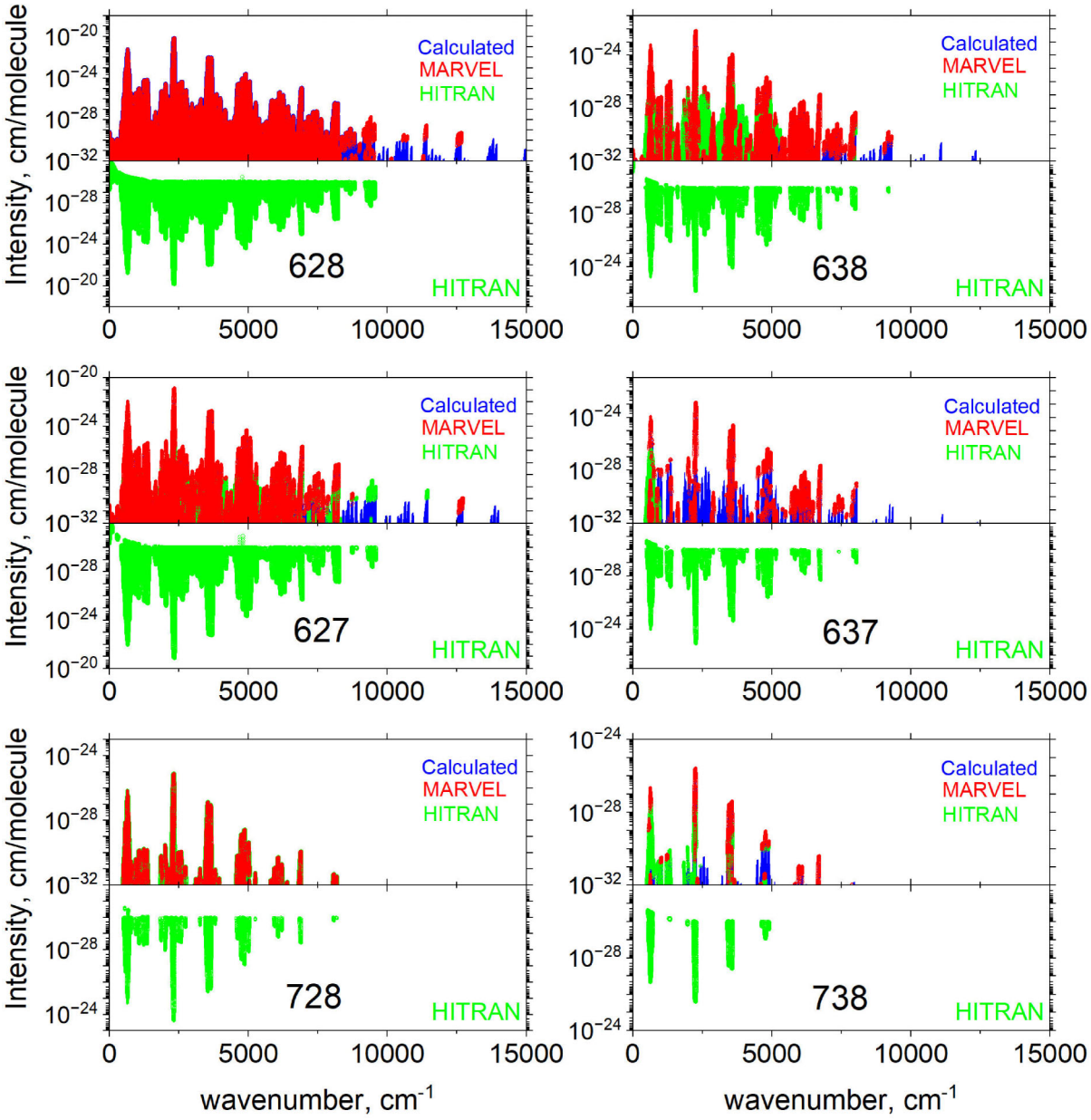

Figure 10. Stick spectra of six asymmetric isotopologues of CO_2 at $T = 296$ K (absorption coefficients in cm molecule^{-1}). MARVELized transitions are highlighted and compared with HITRAN2020. Intensities are not scaled to isotopic abundances.

under the European Union's Horizon 2020 research and innovation programme through Advance Grant number 883830.

DATA AVAILABILITY

All data generated in this article are available in the article and its supplementary materials or via the the ExoMol website www.exomol.com. The HITRAN energies of 11 minor isotpologues generated using the MARVEL procedure underlying this article are available in the online supplementary material. The line lists and associated data are available from www.exomol.com. The codes used in this work, namely TROVE and EXOCROSS, are freely available via https://github.com/exomol.

REFERENCES

Abshire J. B. et al., 2010, in Singh U. N., Pappalardo G., eds, Proc. SPIE. Conf. Ser. Vol. 7832, Lidar Technologies, Techniques, and Measurements for Atmospheric Remote Sensing VI. SPIE, Bellingham, p. 72
Ahrer E.-M. et al., 2023, Nature, 614, 649
Al-Refaie A. F., Changeat Q., Waldmann I. P., Tinetti G., 2021, ApJ, 917, 37
Alatoom D., Ibrahim M. T. I., Furtenbacher T., Császár A. G., Alghizzawi M., Yurchenko S. N., Azzam A. A. A., Tennyson J., 2024, J. Comput. Chem., 45, 2558
Amat G., Pimbert M., 1965, J. Mol. Spectrosc., 16, 278
Arendas P., Furtenbacher T., Császár A. G., 2020, Sci. Rep., 10, 19489
Arendas P., Furtenbacher T., Császár A. G., 2024, Sci. Rep., 14, 794
Azzam A. A. A. et al., 2025a, Sci. Data, 12, 532

Azzam A. A. A., AlAlawin J. M. A., Tennyson J., Yurchenko S. N., Furtenbacher T., Császár A. G., 2025c, J. Quant. Spectrosc. Radiat. Transf., 343, 109485

Azzam A. A. A., Azzam S. A. A., Aburumman K. A. A., Tennyson J., Yurchenko S. N., Császár A. G., Furtenbacher T., 2024, J. Mol. Spectrosc., 405, 111947

Azzam A. A. A., Tennyson J., Yurchenko S. N., Furtenbacher T., Császár A. G., 2025b, J. Comput. Chem., 46, e27541

Balashov A. A., Raj A., Wójtewicz S., Ciuryłło R., Lisak D., Bielska K., 2024, J. Quant. Spectrosc. Radiat. Transf., 320, 108978

Barnfield M. G., Polyansky O. L., Yurchenko S. N., Tennyson J., 2026, J. Mol. Spectrosc. (under review)

Biassoni F., Borsa F., Haardt F., Rainer M., 2024, A&A, 691, A283

Bunker P. R., Jensen P., 1998, Molecular Symmetry and Spectroscopy, 2nd edn. NRC Research Press, Ottawa

Butz A. et al., 2011, Geophys. Res. Lett., 38, L14812

Carleo I. et al., 2022, A&A, 164, 101

Chubb K. L. et al., 2021, A&A, 646, A21

Chubb K. L. et al., 2024, RASTI, 3, 636

Crisp D. et al., 2004, Adv. Space Res., 34, 700

Devi V. M., Benner D., H. Smith M. A., Rinsland C. P., 1998, J. Quant. Spectrosc. Radiat. Transf., 60, 771

Fedorova A., Bézard B., Bertaux J.-L., Korablev O., Wilson C., 2015, Planet Space Sci., 113, 66

Frediani J. et al., 2025, A&A, 701, A14

Freedman R. S., Lustig-Yaeger J., Fortney J. J., Lupu R. E., Marley M. S., Lodders K., 2014, ApJS, 214, 25

Furtenbacher T., Császár A. G., Tennyson J., 2007, J. Mol. Spectrosc., 245, 115

Gal Y., Ghahramani Z., 2016, in Balcan M. F., Weinberger K. Q., eds, Proceedings of Machine Learning Research Vol. 48, Proceedings of The 33rd International Conference on Machine Learning. PMLR, New York, USA, p. 1050

Gamache R. R., Lamouroux J., 2013, J. Quant. Spectrosc. Radiat. Transf., 130, 158

Gamache R. R., Lamouroux J., Blot-Lafon V., Lopes E., 2014, J. Quant. Spectrosc. Radiat. Transf., 135, 30

Gamache R. R., Vispoel B., Tennyson J., Yurchenko S. N., Polyansky O. L., Gordon I. E., Hargreaves R. J., Huang X., 2025, J. Quant. Spectrosc. Radiat. Transf., 345, 109568

Glidden A., Seager S., Petkowski J. J., Ono S., 2023, Life, 13, 2325

Gordon I. E. et al., 2022, J. Quant. Spectrosc. Radiat. Transf., 277, 107949

Hargreaves R. J., Gordon I. E., Huang X., Toon G. C., Rothman L. S., 2024, J. Quant. Spectrosc. Radiat. Transf., 333, 109324

Hargreaves R. J., Gordon I. E., Rey M., Nikitin A. V., Tyuterev V. G., Kochanov R. V., Rothman L. S., 2020, ApJS, 247, 55

Hashemi R. et al., 2020, J. Quant. Spectrosc. Radiat. Transf., 256, 107283

Heng K., Lyons J. R., 2016, ApJ, 817, 149

Holmberg M., Madhusudhan N., 2024, A&A, 683, L2

Huang X., Freedman R. S., Tashkun S. A., Schwenke D. W., Lee T. J., 2013, J. Quant. Spectrosc. Radiat. Transf., 130, 134

Huang X., Freedman R. S., Tashkun S., Schwenke D. W., Lee T. J., 2023, J. Mol. Spectrosc., 392, 111748

Huang X., Gamache R. R., Freedman R. S., Schwenke D. W., Lee T. J., 2014, J. Quant. Spectrosc. Radiat. Transf., 147, 134

Huang X., Schwenke D. W., Freedman R. S., Lee T. J., 2017, J. Quant. Spectrosc. Radiat. Transf., 203, 224

Huang X., Schwenke D. W., Freedman R. S., Lee T. J., 2022, J. Phys. Chem. A, 126, 5940

Huang X., Schwenke D. W., Tashkun S. A., Lee T. J., 2012, J. Chem. Phys., 136, 124311

Ibrahim M. T. I., Alatoom D., Furtenbacher T., Császár A. G., Yurchenko S. N., Azzam A. A. A., Tennyson J., 2024, J. Comput. Chem., 45, 969

Irwin P. G. J. et al., 2008, J. Quant. Spectrosc. Radiat. Transf., 109, 1136

Jiménez-Monferrer S. et al., 2021, Icarus, 353, 113830

Karlovets E. V. et al., 2021, J. Quant. Spectrosc. Radiat. Transf., 276, 107896

Kochanov R., Perevalov V., 2025, J. Quant. Spectrosc. Radiat. Transf., 341, 109428

Line M. R. et al., 2017, AJ., 848, 83

Madhusudhan N., Sarkar S., Constantinou S., Holmberg M., Piette A. A. A., Moses J. I., 2023, ApJL, 956, L13

Mansour M. H. I., Azzam A. A. A., Tennyson J., Yurchenko S. N., Furtenbacher T., Császár A. G., 2025, Molecular Physics, An International Journal at the Interface Between Chemistry and Physics. Taylor & Francis, New York, p. e2550568

Mant B. P., Yachmenev A., Tennyson J., Yurchenko S. N., 2018, MNRAS, 478, 3220

Mayo A. W., Fortenbach C. D., Louie D. R., Dressing C. D., Turtelboom E. V., Giacalone S., Harada C. K., 2025, ApJ, 170, 44

McKemmish L. K., Bowesman C. A., Kefala K., Perri A. N., Syme A. M., Yurchenko S. N., Tennyson J., 2024, RASTI, 3, 565

Medvedev E. S., Meshkov V. V., Stolyarov A. V., Ushakov V. G., Gordon I. E., 2016, J. Mol. Spectrosc., 330, 36

Medvedev E. S., Ushakov V. G., 2022, J. Quant. Spectrosc. Radiat. Transf., 288, 108255

Medvedev E. S., Ushakov V. G., Conway E. K., Upadhyay A., Gordon I. E., Tennyson J., 2020, J. Quant. Spectrosc. Radiat. Transf., 252, 107084

Min M., Ormel C. W., Chubb K., Helling C., Kawashima Y., 2020, A&A, 642, A28

Mollière P., Wardenier J. P., van Boekel R., Henning T., Molaverdikhani K., Snellen I. A. G., 2019, A&A, 627, A67

Mondelain D., Campargue A., Gamache R. R., Hartmann J.-M., Gibert F., Wagner G., Birk M., Röske C., 2025, J. Quant. Spectrosc. Radiat. Transf., 333, 109271

Obaidata S. A. M., Azzam A. A. A., Tennyson J., Yurchenko S. N., Furtenbacher T., Császár A. G., 2025, J. Mol. Spectrosc., 340, 109444

Oppenheimer B. R. et al., 2013, AJ., 768, 24

Oyafuso F. et al., 2017, J. Quant. Spectrosc. Radiat. Transf., 203, 213

Paszke A. et al., 2019, in Wallach H. M., Larochelle H., Beygelzimer A., d'Alché-Buc F., Fox E. A., Garnett R., eds, Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, PyTorch: An Imperative Style, High-Performance Deep Learning Library. Vancouver, BC, Canada, p. 8024

BPollack J. B. et al., 1993, Icarus, 103, 1

Polyansky O. L., Bielska K., Ghysels M., Lodi L., Zobov N. F., Hodges J. T., Tennyson J., 2015. Phys. Rev. Lett., 114, 243001

Refaat T. F., Singh U. N., 2024, IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium, Profiling of Mars Atmospheric Carbon Dioxide Isotopologues Using 2-Micron Orbiting Lidar. IEEE, Athens, Greece, p. 930

Roche S. et al., 2021, AMT, 14, 3087

Rothman L. S. et al., 2010, J. Quant. Spectrosc. Radiat. Transf., 111, 2139

Rothman L. S., Young L. D. G., 1981, J. Quant. Spectrosc. Radiat. Transf., 25, 505

Snels M., Stefani S., Grassi D., Piccioni G., Adriani A., 2014, Planet Space Sci., 103, 347

Sokolov A., Yurchenko S. N., Tennyson J., Gamache R. R., Vispoel B., 2025, J. Quant. Spectrosc. Radiat. Transf., 330, 109225

Swain M. R. et al., 2009b, ApJ, 704, 1616

Swain M. R., Vasisht G., Tinetti G., Bouwman J., Chen P., Yung Y., Deming D., Deroo P., 2009a, ApJL, 690, L114

Tan Y., Kochanov V R., Rothman L. S., Gordon I. E., 2019, J. Quant. Spectrosc. Radiat. Transf., 124, 11580

Tan Y., Skinner F. M., Samuels S., Hargreaves R. J., Hashemi R., Gordon I. E., 2022, ApJS, 262, 40

Tashkun S. A., Perevalov V. I., 2011, J. Quant. Spectrosc. Radiat. Transf., 112, 1403

Tashkun S. A., Perevalov V. I., Gamache R. R., Lamouroux J., 2015, J. Quant. Spectrosc. Radiat. Transf., 152, 45

Tashkun S., Perevalov V., Teffo J.-L., Bykov A., Lavrentieva N., 2003, J. Quant. Spectrosc. Radiat. Transf., 82, 165

Tennyson J. et al., 2024b, J. Quant. Spectrosc. Radiat. Transf., 326, 109083

Tennyson J., Furtenbacher T., Yurchenko S. N., Császár A. G., 2024a, J. Quant. Spectrosc. Radiat. Transf., 316, 108902

Tennyson J., Hill C., Yurchenko S. N., 2013, in 6th international conference on atomic and molecular data and their applications ICAMDATA-2012. AIP, New York, p. 186
Tobias R. et al., 2020, Nature Comms., 11, 1708
Toth R. A., Brown L. R., Miller C. E., Devi V. M., Benner D. C., 2008, J. Quant. Spectrosc. Radiat. Transf., 109, 906
Vispoel B., Gamache R. R., 2024, J. Quant. Spectrosc. Radiat. Transf., 316, 108896
Wang R., Balciunaite U., Chen J., Yuan C., Owens A., Tennyson J., 2023, J. Quant. Spectrosc. Radiat. Transf., 306, 108617
Wattson R. B., Rothman L. S., 1986, J. Mol. Spectrosc., 119, 83
Webster C. R. et al., 2013, Science, 341, 260
Yurchenko S. N., Barber R. J., Tennyson J., Thiel W., Jensen P., 2011, J. Mol. Spectrosc., 268, 123
Yurchenko S. N., Barber R. J., Yachmenev A., Thiel W., Jensen P., Tennyson J., 2009, J. Phys. Chem. A, 113, 11845
Yurchenko S. N., Mellor T. M., 2020, J. Chem. Phys., 153, 154106
Yurchenko S. N., Mellor T. M., Freedman R. S., Tennyson J., 2020, MNRAS, 496, 5282
Yurchenko S. N., Mellor T., Tennyson J., 2024, MNRAS, 534, 1364
Yurchenko S. N., Thiel W., Jensen P., 2007, J. Mol. Spectrosc., 245, 126
Yurchenko S. N., Yachmenev A., Ovsyannikov R. I., 2017, J. Chem. Theory Comput., 13, 4368
Zak E. J., Tennyson J., Polyansky O. L., Lodi L., Zobov N. F., Tashkun S. A., Perevalov V. I., 2017a, J. Quant. Spectrosc. Radiat. Transf., 189, 267
Zak E. J., Tennyson J., Polyansky O. L., Lodi L., Zobov N. F., Tashkun S. A., Perevalov V. I., 2017b, J. Quant. Spectrosc. Radiat. Transf., 203, 265
Zak E., Tennyson J., Polyansky O. L., Lodi L., Tashkun S. A., Perevalov V. I., 2016, J. Quant. Spectrosc. Radiat. Transf., 177, 31

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

HITRAN-MARVEL.zip

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a TEX/LATEX file prepared by the author.