Supplementary Figures for Isotopologue Consistency Paper

**Fig. S1** CO$_2$ 01101e-00001e analysis

**Fig. S2** CO$_2$ 01101f-00001e analysis
Fig. S3 CO₂ 00011e-00001e analysis

[Graphs and plots related to CO₂ analysis and TDM calculations]
Fig. S4 CO$_2$ 1001x-00001 analysis
Fig. S5 CO₂ 2001x-00001 analysis
Fig.S6 CO$_2$ 3001x-00001 analysis

Fig.S7 The Einstein-A coefficient $A_{21}$ and TDM (sum of transition dipole element squares) of 13 CO$_2$ isotopologues, for the R16e/f transitions in 0n11-0n01 ($\nu_2 + \nu_3 \leftrightarrow \nu_2$, n=0-3) bands.
**Fig. S8** The TDM_{SO}/TDM(626) ratios of 4 e/f→e/f branches of CO$_2$ 01111-01101 and 02211-02201 bands.

**Fig. S9** $K_a''$ dependence of SO$_2$ TDM ratios: a) $J''=32$; b) $J''=12$
**Fig. S10** The J dependence of the TDM ratios of “R transition in 9 bands: J”$^{14}K_a$,$K_c$ = 12,12,0 ($\Delta v$=0), or 12,0,12 ($\Delta v$>0). Note the 2v2 need more investigation / confirmation for higher J effects, where the TDM ratios of 6x6 isotopologues (x=3-6) spans from 1E-4 to 10. The local minima of TDM ratios are found at J=48,44,37,32 for x=3,4,5,6, respectively.
Fig.S11 Percentage of fitted lines in 30 EH(Ames) analysis, using Sym or Asym datasets, and four different uncertainty/error bar thresholds from 1E-4 to 1E-7 cm\(^{-1}\). (a) overview, by fitting accuracy level; (b) detailed comparison, by fitting accuracy, S and O isotopes.

Fig.S12 The K\(_C\) splits of Ames SO\(_2\) 668 ground state levels. Split \(\Delta = E(J, K_a, K_c=J-K_a) - E(J, K_a, J+1-K_a)\). Left side is overview; Right panel shows more details.
**Fig.S13.** D_κ analysis in the EH(Ames) SPFIT of 30 isotopologues, using Sym or Asym datasets, and four different uncertainty/accuracy level from 1E-4 cm\(^{-1}\) to 1E-7 cm\(^{-1}\). (a) overview, by fitting accuracy level, data are with respect to the value averaged from 8 D_κ; (b) the D_κ convergence with respect to fitting accuracy: δ(Asym-Sym) are black squares, and δ(D_κ(n+1) – D_κ(n)) are triangles, where n is the accuracy/uncertainty level, e.g. 5-4 means D_κ(1E-5) – D_κ(1E-4).

**Fig.S14** DJK in 30 EH(Ames) analysis, using Sym or Asym datasets, and four different uncertainty/error bar thresholds from 1E-4 to 1E-7 cm\(^{-1}\). (a) different DJK values, by fitting accuracy level and data type, with respect to the value averaged from 8 DJK’s; (b) the DJK convergence with respect to fitting accuracy: δ(Asym-Sym) are black squares, and δ(DJK(n+1) – DJK(n)) are triangles; (c) the DJK overview of 30 SO\(_2\) isotopologues.
Fig.S15 (a-b): Absolute (Left y axis, circles) and relative (Right y axis, squares) differences of the EH parameters acquired between fully relaxed fit (all vary freely) and the partially fixed fit (some parameters are fixed at 626 values): (a) 646 only; (b). the relative differences (%) of 646, 636, 628, 627, and 828, at 3 fitting accuracy levels.
Discussions in Ver.0, Section 4, for EH(Ames) and Line Positions

1. Isotopologue-consistency of EH model fits

To ensure the $\Delta$(Ames-Expt/CDMS) in Figs.16-17 are highly self-consistent, a uniform EH model has to be used for all the 30 isotopologues. We know the fitting residuals depend on the completeness of the EH model terms included in the fit. Adding more terms to describe higher order effects or inter-molecular couplings and perturbations usually can help reduce the fitting residuals. Experimentalists prefer the EH fitting residuals as small as, or even smaller than the data uncertainty. This is easy to understand, but sometimes it may lead to isotopologue-specific terms, which means EH model differences and introduces inconsistency in the $\Delta$ calculations. The 646 EH model from CDMS website is selected because the EH model was also used for 626, 628, 627, 636, and recent 828 experimental vibrational IR band studies. Extra terms (if any) accounting for the $^{33}$S and $^{17}$O related hyperfine splittings are excluded, which we assume not strongly coupled with remaining terms.

Given a uniform EH model on an isotopologue, its parameters can carry different least-squares weights. But the weight for any specific parameter should be a constant across all 30 isotopologues. However, the CDMS EH models of minor isotopologues had various higher order terms fixed at CDMS EH(626) values, i.e. their corresponding weights were set to essentially zero. It effectively breaks the consistency of EH(CDMS/Expt) data between 626 and the rest. If we ignore 626 EH parameters and make predictions with the $\Delta$s of minor isotopologues, the $\Delta$ consistency is still not very good, because different higher order terms were fixed in SFIT. This is part of the bottleneck damaging our prediction accuracy for general applications.

Secondly, the datasets used in SFIT should be as consistent as possible, too. Ideally, they should have the same set of transition quantum numbers, so can be noted as “equivalent” datasets. This is trivial for EH(Ames) analysis because the all microwave transitions > 1E-36 cm/molecule at 296K have been computed. But it is highly unlikely for experimental analysis, if not totally unrealistic. Less abundant isotopologues naturally have less number of transitions observed in lab, and scientists always need the most reliable spectroscopic constants fit from the maximum number of reliable, identified transitions.

Even with “equivalent” transition datasets, in the least-squares fitting step, the experimental uncertainty associated with every transition is almost impossible to maintain 100% consistency for all isotopologues. Accordingly, the error bar set for rejection cannot be synchronized, either. For EH(Expt) fits, it is unwise to manually change the error bar (which in SFIT correlates with uncertainty), no matter larger or smaller. But the equivalent dataset extracted from Ames-296K IR lists have highly consistent uncertainties. In addition, we can also modify the uncertainty to check the convergence of fitted EH constants, see details in Sec.4.2.5.

In short, the best isotopologue-consistency should come from uniform model, identical fitting weights and error bar, and equivalent datasets (including uncertainties) for both EH(Ames) and EH(Expt) analysis on all involved isotopologues. Since the published or online EH parameters were usually fit from different models and datasets, one may need to extract the largest common transition dataset from the original experimental data, run least-squares fits independently but consistently, and use the results for $\Delta$ predictions. But less transitions could hurt the quality of EH(Expt) constants.

Now we can tell that the Fig.17 a) is partially wrong in the “higher-order terms” region, because 8 (12) higher order terms in the 646 (636) EH(CDMS) models were fixed at EH(CDMS) 626 values, they cannot work with any prediction formula. The values of the rest 6 sextic centrifugal distortion constants will be affected, too. Another choice is to re-fit the Ames datasets by fixing the parameters at EH(Ames) 626 values, too. But the results would still have broken consistency from fitting weight discrepancies. Third choice is to re-fit the experimental data of minor isotopologues with the original (fully relaxed) parameter set, but the transition sets do not contain enough information to determine those higher order terms reliably.

OK, it is clear both the dataset and EH model must be consistent for all isotopologues. Existing EH(CDMS/Expt) models for SO$_2$ isotopologues are not quite consistent. How about EH(Ames) then? Are we sure the Ames IR list based EH
parameters fitted from SPFIT are truly reliable, accurate and consistent?

2. Reliability of EH(Ames) constants: $A_0/D_K$

In Fig.16 b), the S isotope effect from $^{32}\text{S}$ to $^{36}\text{S}$ suggests that the $\Delta(32\text{S})$ in the middle is critical, if one wants to predict $\Delta(32\text{S})$ and $\Delta(36\text{S})$ with $\sim 0.01$ MHz accuracy. A 0.01 MHz error on $^{32}\text{S}$ may become 0.02 cm$^{-1}$ or larger at $^{36}\text{S}$. So less than 0.005 MHz deviations are preferred. In our tests, the $B_0/C_0$ in EH(Ames) usually are well converged, but the largest uncertainty lies in the $A_0$ constant of EH(Ames).

In the $A$-reduced Watson Hamiltonian, $A_0$ and $D_0$ are clearly correlated. The consequence is that, the SPFIT at the 1E-5 cm$^{-1}$ uncertainty level cannot guarantee $A_0$ convergence better than 0.01 MHz, at least impossible for the ~3000 transitions included in our EH(646) fit. The EH(Ames) reported in Fig.16 b) has the $A_0/D_0$ at 58990.81394 MHz / 2.440615 MHz. But there does exist another stable local minima where $A_0/D_0$ are 58990.88397 MHz / 2.441753 MHz. The corresponding changes on $B_0$ and $C_0$ are -0.0039 MHz and -0.0024 MHz, respectively. Obviously, such +0.07MHz / +0.00114 MHz differences are too large to ensure the prediction accuracy of EH(Predicted). For example, +0.07MHz on $A_0$ may cause deviations ~60MHz at J=30, and +0.001 MHz on $D_0$ can bring ~50 MHz errors at $K_s=15$.

If we fix the $D_0$ somewhere between the two minima, the fitted $A_0$ also comes between the two $A_0$ values at minima. You can imagine a “solution” curve of all possible $A_0/D_0$ combinations that give similar least-squares fits, e.g. same no. of failed transitions and total rms. We did locate two more local minima along the “solution” curve, with $A_0/D_0$ fit at 58990.8267 / 2.44082 MHz and 58990.8622 / 2.44147 MHz, respectively. These two EH(Ames) 646 fits have similar fitting quality, with 298 and 278 transitions rejected, respectively. The two fits are included in supplementary material. Interested readers can try SPFIT on your own.

The issue we describe above is all on EH(Ames) side. The solution is easy: no convergence at 5E-5 cm$^{-1}$? Try 5E-6 or 5E-7 cm$^{-1}$. At 5E-6 cm$^{-1}$, the $A_0$ comes out in the middle +0.07MHz $A_0$ variance is reduced to ~0.006 MHz. This seems good enough for linear approximations in Fig.16 b). The variance is further reduced at 5E-7 cm$^{-1}$ level.

Apparently, this issue and fix are not new to experimental spectroscopists. They have accumulated rich experiences in numerous EH model fits where some parameters cannot be accurately determined. In most cases, it is attributed to the dataset deficit, i.e. too few transitions or quantum numbers not high/low enough, or the EH model deficit, e.g. to add additional coupling terms or higher order terms. Note such multi-minima scenario did not occur when we reproduced the CDMS fit (using the CDMS dataset) or run SPFIT with the equivalent Ames set (CDMS quantum numbers + Ames-296K line positions). The ~3000 strongest MW transitions at 296K included in the EH(Ames) model fit are more than 10 times of EH(CDMS) fit. It is reasonable to suspect such multi-minima issue may only become significant for considerable number of transitions and/or relatively large uncertainty. For example, most transitions in CDMS EH(646) fit carry 0.05-0.20 MHz uncertainty, or 1.7-7E-6 cm$^{-1}$.

It raises a question on the rovibrational band EH analysis. In those experimental rovibrational IR work fitting thousands of unperturbed transitions carrying 1E-5~1E-4 cm$^{-1}$ accuracy, does the $A_0$ still have uncertainty as large as 0.05 MHz (or larger)? Many molecular rovibrational IR studies choose to report the observed line positions and fitting errors in unit of 1E-3 cm$^{-1}$ or at most 1E-4 cm$^{-1}$, which are limited by the sig figs of the frequency of the “standard reference” transition. Now we are telling people that even the fitting deviation $\sigma_{\text{rms}}$ as small as 5E-5 cm$^{-1}$ can not 100% guarantee the $A_0$ convergence better than 0.05 MHz. No need to mention EH fits with errors ~1E-3 cm$^{-1}$. In short, we seriously doubt the uncertainty range of published $A_0$ constants for many molecules, and bands. It could be larger than we had thought.

On the other hand, we have observed the $B_0/C_0$ are fairly stable. With less than 0.005 MHz variations, they probably can be safely adopted in the linear approximation formula (7) in Section 4.1.

In Fig.16 a) analysis, ~3000 transitions were fit with equally weighted EH parameters, while there were only 282 transitions in CDMS fit. How the rotational constants become if we use the equivalent Ames set of 282 lines? At 1E-5 cm$^{-1}$ level, the $A/B/C$ differences are -0.006, +0.004, and +0.001 MHz, respectively. But this test has limited value, because the results may rely on the specific $J/K_a$ data range, uncertainty level, as well as the molecular type and vibrational bands.
After checking the reliability and convergence of $A_0/D_0$ constants in EH(Ames) analysis, next we check the EH(Ames) consistency across all isotopologues, then discuss more about convergence and uncertainty.


This section explores how isotopologue-consistent the EH(Ames) analysis may become, if we trust the first principle calculations on the empirically refined PES. To ensure the rotational constants converged to better than 0.002 MHz, we need larger dataset. To check the convergence with respect to the uncertainty level and SPFIT error bars, we need more significant figures for line position. The Ames MW lists used in Sec.4.2.2 only carry 1E-5 cm$^{-1}$ resolution, now all the Ames-296K MW line sets are regenerated with 1E-8 cm$^{-1}$ (or 3E-4 MHz) resolution.

Asymmetric and 7x7 isotopologues have more IR active transitions than symmetric isotopologues, due to nuclei spin statistical weights. To maintain the best consistency, we created two sets of transitions for each isotopologue. One set is the asymmetric (Asym) isotopologue style, the other set is the symmetric (Sym) isotopologue style. For Asym and 7x7 isotopologues, it is straightforward to create a Sym dataset by filtering out those Sym-forbidden transitions from the Asym dataset. For other Sym isotopologues, those Asym-type transitions can only be created from the rovibrational energy levels we computed, with zero intensities. During this step, we found and fixed small inconsistency on a few isotopologue line lists, because their old IR lists were computed using slightly different parameters. At the end, the J=0-75 rovibrational energy levels of all 30 isotopologues have been re-computed, using an uniform set of VTET input parameters for Asym isotopologues. Sample input file and the rovibrational levels are included in supplementary files.

At the beginning, we chose those transitions with $J$≤60, $K_a$≤35, and $S$(intensity) > 1E-32 cm/molecule for each isotopologue. But the transition sets constructed in this way are not strictly consistent, because the intensity of a specific transition will vary by isotopologue. See Sec.3 for examples. As a result, the 30 datasets for SPFIT will have slightly different size and quantum numbers. To solve this problem, we randomly picked the 688 isotopologue Asym and Sym datasets as the “standard” transition sets for all SPFIT least-squares fitting analysis. We go through other 29 Ames-296K lists, and extract all equivalent transitions that form our final SPFIT sets. In this way, we avoid dataset contaminations on the isotopologue consistency.

Each Sym set has 13,744 lines, and each Asym set has 40,311 lines, in the range of 0 – 370~410 cm$^{-1}$. There are 4 levels of uncertainty: 1E-4 cm$^{-1}$, 1E-5 cm$^{-1}$, 1E-6 cm$^{-1}$ and 1E-7 cm$^{-1}$, and the SPFIT error bar is set at 5 times of uncertainty. Usually more data or smaller uncertainty requires more SPFIT cycles to find the “global minimum” in the phase space of EH parameters. Fig.S11 summarizes the percentage of lines successfully reproduced within the error bar. In general, heavier isotopes bring more lines fitted within the error bar, and the percentage of failed transitions rises on smaller uncertainty. For 1E-5 cm$^{-1}$ and 1E-4 cm$^{-1}$ cases, 96-100% lines can be reproduced, but some fitted EH parameters might not be tightly converged (Sec.4.2.2 and 4.2.5). 83-90% lines can be reproduced at 1E-6 cm$^{-1}$ level. Even for the most difficult 1E-7 cm$^{-1}$
level Sym fits, we still have 57-70% lines that can be successfully reproduced. It should be noted that the Asym sets have fitted percentages higher than those of the corresponding Sym sets. Does this suggest the EH(Expt) of asymmetric isotopologues are slightly more reliable than symmetric ones? Not necessarily, but it is an interesting topic to explore in future.

Due to the nature of over-determined linear equations, we cannot 100% guarantee the “best” EH parameters have been determined for all the 1E-6 and 1E-7 fits. However, we can confidently claim one more layer of consistency: the consistency of SPFIT fitted percentages. In Fig.S11b, detailed fitted % are presented by S isotope (horizontal axis), dataset Asym (square) vs Sym (circle), accuracy level 1E-6 (open) vs. 1E-7 (half filled), and O isotope combinations (6 colors). In all isotopologue groups, the 6x8 isotopologues consistently have the lowest fitted %. Looking at the figure, two general trends are mixed together: for symmetric isotopologues, heavier O isotopes means higher fitted %, i.e. 8x8 > 7x7 > 6x6; for asymmetric isotopologues, less O atom (relative) mass differences means higher fitted %, i.e. 7x8 > 6x7 > 6x8.

Which factors affect the fitted % at higher accuracy level, i.e. 1E-6 and 1E-7 cm$^{-1}$? We have identified part of them from the Kc split analysis, Ames vs. EH model. Use the 668 ground state rotational levels, we compute all Kc splits at selected Ka and J up to 70, $\Delta = E(K_c=J-K_a) - E(K_c=J+1-K_a)$. Results are plotted along J in Fig.S12. Kc=5 and Kc=10 splits shoot up at J~12 and J~45, respectively. For Kc=15 and up, the magnitude of Kc split oscillation in J=40-70 region keeps getting larger. The pattern is clear: the Kc splits of odd Kc (15,25,35,...) go negative, while the Kc splits of even Kc (20,30,40) go positive. See Fig.S12b. As far as we know, such Kc splits of 1E-5 to 1E-4 cm$^{-1}$ in high Kc region do not exist in the EH models we adopt in this work. For example, at J=60, all the Kc splits of Kc≥15 levels predicted by EH models are less than 1E-7 cm$^{-1}$, i.e. two orders of magnitude less than Ames Kc splits. Currently we are inclined to believe the higher Kc calculations may contain some numerical impurities that we should try to fix in future. Some EH(Ames) oscillations at higher J/Ka are larger than the rejection threshold, i.e. 5E-6 to 5E-7 cm$^{-1}$. This means the EH(Ames) models acquired at 1E-6 to 1E-7 cm$^{-1}$ accuracy level are fitted with less transitions at higher J/Ka region. It also suggests better reliability for the fitted EH(Ames) constants, including both lower order terms (A/B/C + quartic) and higher order terms.

Fig.S12 The Kc splits of Ames SO$_2$ 668 ground state levels. Split \( \Delta = E(J, K_a, K_c=J-K_a) - E(J, K_a, J+1-K_a) \). Left side is overview; right panel shows more details.

Next we check the convergence and quality of EH(Ames) constant, Dc.

4 Isotopologue Consistency Benchmark: Dc and Improved Prediction.

Except the rotational constants, the quartic centrifugal distortion constant Dc has the largest impact on prediction accuracy of rotational energy levels and line positions. The Dc values are plotted in Fig.13a. There are 8 different Dc for each isotopologue: two datasets (Asym vs. Sym) x 4 accuracy levels (1E-4 to 1E-7 cm$^{-1}$). The averaged Dc is taken as the zero line, all 8 values are adjusted with respect to it. Note the y axis unit is Hz, not MHz or kHz.

In Fig.S13a, from the 626 at left end to the 868 at the right end, the magnitude of Dc consistently decreases from 2.59 to 2.05 MHz, the spread of Dc values shrinks from ±20Hz to ±15Hz, while the ratio of spread/magnitude approximately remains the same. The spread is the difference between Dc(1E-4) and Dc(1E-7). It is evident that Dc monotonically drops
from 1E-4 cm\(^{-1}\) level to 1E-7 cm\(^{-1}\) level, which is probably the result of less higher J/Ka levels or transitions. An interesting observation is that Asym \(D_K\) are slightly higher than the Sym \(D_K\) while their differences drop from ~10 Hz (1E-4 fit) to 1-2 Hz (1E-6 fit), then negligible (1E-7 fit). This is an indication of convergence: Asym vs Sym dataset difference can be minimized by raising the fitting accuracy level. The trend of \(\delta(D_K(\text{Asym})-D_K(\text{Sym}))\) differences are clearly demonstrated by black squares in Fig.S13b.

Fig.S13. \(D_K\) analysis in the EH(Ames) SPFIT of 30 isotopologues, using Sym or Asym datasets, and four different uncertainty/accuracy level from 1E-4 cm\(^{-1}\) to 1E-7 cm\(^{-1}\). (a) overview, by fitting accuracy level, data are with respect to the value averaged from 8 \(D_K\); (b) the \(D_K\) convergence with respect to fitting accuracy: \(\delta(\text{Asym-Sym})\) are black squares, and \(\delta(D_K(n+1) - D_K(n))\) are triangles, where \(n\) is the accuracy/uncertainty level, e.g. 5-4 means \(D_K(1E-5) - D_K(1E-4)\).

Fig.S13b also reports the \(D_K\) convergence with respect to fitting accuracy, \(\delta(D_K(n+1) - D_K(n))\), where \(n\) is the accuracy/uncertainty level, e.g. \(\delta(6-5) = D_K(1E-6) - D_K(1E-5)\). Interestingly, the \(\delta(5-4)\) are small, ±5Hz; the \(\delta(6-5)\) are the largest, -25~15Hz; then the \(\delta(7-6)\) are reduced to -15~5 Hz. This may indicate that the 1E-5 results are not well converged, while the 1E-6 and 1E-7 are more reliable.

Fig.SS1. The linear and 2\(^{nd}\)-order fits of EH(Ames) \(D_K\) values (determined at 1E-6 cm\(^{-1}\) level) along the S and O isotopes. (a). S isotope; (b) O isotope; (c) S & O isotope effects combined. See text for averaged fitting residuals and \(\sigma_{\text{rms}}\).

Now it is time to re-check the validity of prediction in Fig.16. It turns out the simple approximation we used there was inadequate to distinguish 6x8 and 7x7 isotopologues. The x axis better uses the inverse of mass, instead of the mass itself. As shown in Fig.SS1, it works well. The S isotope effects and O isotope effects are in panel (a) and (b), respectively. The \(D_K\) values in panel (a) are with respect to the \(D_K(626)\), while panel (b) uses original \(D_K\). The linear fits (red) and 2\(^{nd}\)-order fits (blue) have fitting residuals similar to what we reported in Section 4.1. However, now we can separate 628 and 727 predictions, and both are reliable. Note Fig.SS1 and Fig.19 have same b and c panels.

Fig.SS1c tries to combine the S and O isotope effects. The scaling factor between \(1/m_S\) and \((1/m_{O1}+1/m_{O2})\), 3.9, is handpicked so that it can reproduce the linearity reasonably well. The linear fit has relative deviations from -0.11% to +0.33%, with mean ± \(\sigma\) = 4E-4 ± 0.11%. The 2\(^{nd}\) order fit has relative deviations in -0.038 ~ +0.024%, with mean ± \(\sigma\) = 0 ± 0.016%. It represents the isotopologue consistency of EH(Ames) \(D_K\) parameters, which can be utilized to predict or verify the experimental \(D_K\) of rare isotopologues. It is a better way to use the mass inverses.
On the other hand, not all EH constants can find such linear relations easily. The 2nd most important parameter at quartic level is $D_{JK}$. Fig.S14 reports the distribution of $\delta(D_{IK}-D_{JK})_{avg}$ in panel (a), the $\delta(\text{Asym-Sym})$ and $\delta(D_{IK}(n+1)-D_{IK}(n))$ in panel (b). Similarly, along the horizontal axis, larger $D_{IK}$ magnitude goes with larger $D_{JK}$ spread on vertical direction, see (a). Along the vertical axis in (a), higher fitting accuracy leads to more negative $D_{JK}$, and smaller $\delta(\text{Asym-Sym})$ differences. Note the $D_{JK}$ $\delta(\text{Asym-Sym})$ differences are one-order-of-magnitude smaller than $D_{IK}$. In Fig.S14b, the $\delta(D_{IK}(n+1)-D_{IK}(n))$ monotonically rises from $-12$ to $-2$ Hz. Even with such tight convergence, we still failed to find a $D_{IK}$—mass inverse relation to align all $D_{IK}$ on a single line, not even very approximately, see Fig.S14c. However, please remember what really matters is the pattern of the differences between EH(Expt) and EH(Ames), not either EH(Expt) or EH(Ames) only.

**Fig. S14 $D_{JK}$ in 30 EH(Ames) analysis, using Sym or Asym datasets, and four different uncertainty/error bar thresholds from 1E-4 to 1E-7 cm$^{-1}$. (a) different $D_{JK}$ values, by fitting accuracy level and data type, with respect to the value averaged from 8 D$_{JK}$’s; (b) the $D_{JK}$ convergence with respect to fitting accuracy: $\delta(\text{Asym-Sym})$ are black squares, and $\delta(D_{IK}(n+1) – D_{JK}(n))$ are triangles; (c) the $D_{JK}$ overview of 30 SO$_2$ isotopologues.**

### 5. Isotopologue Consistency Benchmark: Uncertainty of all 26 EH constants.

Follow the $D_{K}$ and $D_{JK}$ analysis, we check the uncertainty (or convergence) of other EH constants. In fact, there are two different “uncertainty” to discuss. The first kind is the width of data range for a specific EH constant from 8 SPFIT values, e.g. Fig.SS1a and Fig.S14a. The second kind of uncertainty refers to the numbers in the parentheses at the end of each constant, which is the uncertainty estimated for the specific over-determined least-squares problem. We call them the “constant uncertainty” and the “fitting uncertainty”, respectively.

For each isotopologue, the “fitting uncertainty” are extracted from the SPFIT results at 1E-6 cm$^{-1}$ or 1E-7 cm$^{-1}$ level, then divided by the averaged value of corresponding constants. This is relative fitting uncertainty in %. The “constant uncertainty” is estimated by $\delta|\text{EH}(1E-7)-\text{EH}(1E-6)|$ differences, which are divided by the corresponding constants to get the relative $\delta%$. For each EH constant, the relative constant uncertainty and relative fitting uncertainties are averaged across 30 isotopologues and shown as the red squares in Fig.20a, see left Y axis (%). The EH parameter values on the x axis are also averaged from 30 isotopologues. This is appropriate in the log scale. Obviously, the constant uncertainty (squares) is the major factor, which are 2-4 orders of magnitude larger than the fitting uncertainties.

The constant uncertainty is also averaged and reported as blue circles with crosses, see the right axis in unit of Hz. Obviously one may try some linear approximation to get the relation between constant uncertainty (in Hz) and parameter values. For example,

$$\log\delta_{HZ} = \log\delta_{EH \ \text{parameter in MHz}} \times 5/7 - 11.$$

The trend of relative constant uncertainty % can be approximated, too, but might not be quite useful. In general, we have 0.01% ~ 1E-4% for quartic constants, 0.1~10% for sextic terms, and 1-100% for higher order terms. For many constants below 1E-7 MHz, their relative constant uncertainties (red squares, half filled) are much smaller than the “prediction” errors in Fig.17a. Note we already know those “predictions” in Fig.17a were unreliable because some constants were fixed at 626 values. In Fig.20a, the largest $\delta%$ goes with: ~10% for $P_{JKK}$, $P_{KKK}$, and $\Phi_{JK}$, ~20% for $l_J$ and $l_K$, and 50-100% for $\phi_K$ and $l_{JK}$. Further investigations at higher $J$ and $K$ may help clarify if we have adequate information to determine these terms, e.g. the $\phi_K$ around 1E-8~1E-7 MHz. Some of those might be correlated with other EH parameters.
From another perspective, the averaged $\delta$ and $\delta%$ values in Fig.20a can be taken as an indicator for the compatibility between Ames datasets and the EH models, i.e. less compatibility leads to larger uncertainty. To have a more precise understanding about the consistency vs. uncertainty, Fig.20b reports the $\delta$ and $\delta%$ for all 26 EH constants of all 30 isotopologues. The spread of EH constants in panel (b) is affected by two factors: the isotopologue difference, and the EH model or data inadequacy. In other words, they are molecule specific and model/data specific, but some general trends may be still recognizable. For example, in the “lower order terms” which include rotational constants and quartic level constants, the $\delta%$ variation of $D_a$ and $D_k$ is as wide as more than 1 order of magnitude, and significantly wider than those of rest. Fortunately, even the largest $\delta%$ are less than 0.05% ($D_a$) or 0.005% ($D_k$), since our transition datasets are more than adequate for $D_a$ or $D_k$ determination. The EH(Ames) parameters are included in supplementary file and also available upon request. Interested readers can apply unique color & symbol to every EH constants and/or every isotopologue, for further explorations.

This uncertainty check sheds some lights on future EH analysis. For example, not all the “higher order terms” have similar reliability or relative variations. Even with ~40,000 transitions ($I/K_a$ up to 60/35), not all EH parameters can be precisely determined. For experimental data based EH model analysis using a few thousands or even just hundreds of transitions, the constant uncertainty may be larger than reported.


In CDMS analysis,[2] all 626 EH constants were fit with non-zero weights, then part of the fitted EH higher order terms were used for other minor isotopologues, including 636, 646, 627 and 628 (not for 828). Therefore, the higher order terms of other minor isotopologues are inconsistent from one to another. Since we have studied the isotopologue consistency of EH(Ames) constant uncertainties, we can compare the higher order terms, EH(Ames) vs. EH(CDMS), checking if there exists any pattern or trends.

To do that, the EH(Ames) lower order terms need to be replaced by EH(CDMS) values. We re-run the “prediction” procedure for 646, replace the A/B/C constants and quartic level constants, feed different EH sets to SPCAT and predict line positions, compare the predictions to the pure EH(CDMS/Expt) based reference predictions, then plot the line position differences in Fig.21a. The two terms before and after “+” refer to the source of “lower order” and “higher order” terms, respectively, while “A/B/C corrected” means the A/B/C constants in EH(Ames) models have been replaced with EH(CDMS/Expt) values. The EH(Ames) parameters set being used here was fit from Asym datasets with 1E-7 cm$^{-1}$ accuracy level, which we believe have been converged to better than 0.003 MHz, and more appropriate for the low $J/K_a$ line positions.

---

Fig.20 (a) The fitting uncertainty and constant uncertainty (convergence) of 26 EH(Ames) parameters, averaged from 30 SO$_2$ isotopologues. Left axis: the relative fitting uncertainty % from SPFIT (triangles), and $\delta%|EH(1E^-6) – EH(1E^-7)|$ (red squares, half filled). Right axis: $\delta|EH(1E^-6) – EH(1E^-7)|$ in Hz; (b) constant uncertainty of 26 EH(Ames) parameters in the 1E$^-6$ cm$^{-1}$ and 1E$^-7$ cm$^{-1}$ SPFITs of 30 isotopologues. Right axis: absolute uncertainty in Hz (blue squares); Left: relative constant uncertainty % (red circles).
As expected, the A/B/C constant is the primary source of deviations, and quartic centrifugal distortion constants are the 2$^{nd}$ most important factor with significant enough contributions. The “Expt + Ames” set based line positions do carry 0-5 MHz accuracy for K_a ≤ 15 transitions. Deviations may rise up to 5-15 MHz at K_a = 20. Note in Fig.21a the J goes up to 70, so we can see higher-J related deviations at both K_a=5 and K_a=20-25. If J is limited to ≤ 30, the deviations at K_a=15 would be only 1-2 MHz, see Fig.21b.

Same procedure is repeated for minor isotopologues for which EH(CDMS/Expt) model analysis are available. Results are all presented in Fig.21b, with J≤30 limit. Either 626 → 636 → 646 or 626 → 627 → 628, no obvious trend was found. If we multiply the 628 deviations by -1, the Δ would have apparent trend going from positive to negative, along with the O isotope substitution 626→627→628→828. Here we assume the EH(CDMS) is OK for 626, but we can notice the differences between the EH(CDMS) model and the improved EH(Expt) model from Ulenikov group. Please note those differences are beyond the scope of this study, as we just noticed a L_{JK} related discrepancy between the CDMS EH model [2] and Ulenikov EH model [1] for 626.[3]

Short conclusion: We have confirmed again the possibility to have 0-10 MHz prediction accuracy for isotopologue MW line positions, but current EH(CDMS/Expt) models do not have noticeable trend or consistency along the isotope substitutions. A more appropriate “apple-to-apple” comparison needs the EH(Ames) analysis to be done in the same way as in EH(Expt/CDMS) for minor isotopologues.


By fixing part of “higher order” terms at 626 values, we re-compute EH(Ames) constants for 636, 646, 627 and 628, denoted as “Fixed”, and compare the values of rest “not fixed”, i.e. relaxed terms with the original fully “Relaxed” EH(Ames) constants. The differences Δ are defined as X(Fixed) – X(Relaxed). For each isotopologue, we computed three sets of Δ at 1E-5, 1E-6 and 1E-7 levels. Then the Δ are divided by original “Relaxed” values to get Δ/EH ratios, or the relative Δ.

Fig.21a has results for 646 EH constants. In general, smaller constants have larger Δ differences, larger Δ/EH ratios, which indicate less reliability for the fitted EH parameters. Compare to 1E-5 cm$^{-1}$ results, the 1E-6 cm$^{-1}$ and 1E-7 cm$^{-1}$ parameters usually give smaller Δ/EH ratios. For A/B/C constants, there is no doubt the 1E-7 cm$^{-1}$ level fits have the least impact of fixing part of higher order terms, i.e. Δ% in 10$^{-8}$-10$^{-10}$. It is similar or slightly less than the constant uncertainty shown in Fig.20. For other terms, the order of Δ or Δ% is hard to predict along with higher accuracy from 1E-5 cm$^{-1}$ to 1E-7 cm$^{-1}$. For example, the $P_x$ at the left end of Fig.21a have “Fixed vs. Relaxed” values (x 1E+11) as: 1.50 vs. 1.45 (1E-5), 1.71 vs. 1.57 (1E-6), and 3.23 vs 1.65 (1E-7). The Δ% of $\Phi_x$ and $L_x$ also follow such reversed order, i.e. rising along 1E-5 → 1E-6 →
1E-7. Compared to the constant uncertainty we discussed in Sec.4.2.5, the “Fixed” effects of these three EH terms (P, Φ, and L) are notably larger, even by orders of magnitude. Another example is D. From 1E-5 cm⁻¹ to 1E-7 cm⁻¹, the D(Ames) increases from 2.44124 MHz to 2.44132 MHz, while the original D(Ames) decreases from 2.44129 MHz to 2.44126 MHz. Their trends are just opposite.

Fig.S15 (a-b): Absolute (Left y axis, circles) and relative (Right y axis, squares) differences of the EH parameters acquired between fully relaxed fit (all vary freely) and the partially fixed fit (some parameters are fixed at 626 values): (a) 646 only; (b) the relative differences (%) of 646, 636, 628, 627, and 828, at 3 fitting accuracy levels.

Fig.S15b shows the relative Δ% of 5 minor isotopologues at 3 fitting accuracy levels. In general, the effects on A/B/C constants are insignificant. The Δ% of 1E-7 cm⁻¹ level fits are usually the smallest, but with exceptions. Using higher J/K, dataset with more systematic SPFIT investigations may help find a path to optimize the EH model or determine certain parameters with higher confidence. It uses different filling pattern to separate the three fitting accuracy levels. Note the 828 fits in b panel were generated by fixing the same EH term set that was fixed in EH(628) fits. They are NOT the 828 fits we discuss in next section, which use a “prediction” scheme to estimate higher order terms.

8. Isotopologue Consistency Benchmark: Higher order term prediction.

The EH(828) analysis in this section follows the procedure given in the section 3.2 of Ulenikov et al [1], which utilizes the 828 vs. 626 ratios of A and C constants to estimate other 828 EH constants. The isotope substitution theory [4,5,6] predicts the centrifugal distortion constants as:

\[ \frac{X_{828}}{X_{626}} = \frac{\alpha_A}{\alpha_C} \] \[ \alpha_A = \frac{A_{828}}{A_{626}} \] \[ \alpha_C = \frac{C_{828}}{C_{626}} \]

For a specific EH constant, if its total power is m, the power of its J/ part is i, and f = m/2 - i.

In Ref.1, this rotational-constant dependent prediction formula is applied using the established 626 A/B/C, and the 828 A/B/C constants fit from J=0-4 transitions (to get αA and αC). In their EH(828) least-squares fit, all the higher order terms are fixed at the predicted values, except \( H_i \) and \( K_i \).

Since part of our goal is to report our capability to achieve high prediction accuracy on isotopologues, it is definitely necessary to run parallel check and determine which prediction is more accurate. Compare to experimental EH models, our biggest advantage is we have 30 isotopologue IR lists which are computed with a consistent VTET [7] parameter set and fitted to a uniform EH model with accuracy control and satisfactory constant uncertainties (Fig.20). The Ames isotopologue consistency is far better than experimental EH models which were constructed independently from each other and with various limitations. We can take EH(Ames) model fits of any two symmetric isotopologues (the formula used in Ref.1 is for symmetric C2V type), predict one set from the other set, compare to the original EH(Ames) constants that we take as reliable reference, calculate relative errors, and collected in Fig.22a. This plot may combine with Fig.20 to get more interesting results, while Fig.20 includes both symmetric and asymmetric isotopologues.
Look at the performance of prediction formula in Fig.22a. Except for $D_K$, quartic term predictions have relative deviations $|\delta|$% in 0.1 ~ 10%. For sextic and higher order terms, the deviations are noticeably higher, $|\delta| = 0.1 ~ 100%$. Apparently, they all spread in a wide range of 2-3 orders of magnitude. But the log scales is not an ideal choice if we want to find out how the deviations vary from one isotopologue to another. So in Fig.22b, $|\delta(D_4)|$ terms are plotted with linear scale on horizontal axis. Obviously, there are two groups of $|\delta(D_4)|$: the 0.1-0.3% group for O isotope substitutions, and 0.0003-0.007% group for S isotope substitutions. In second group, heavier O isotopes are associated with smaller $|\delta|$, e.g. 868. This probably relates to the effective mass changes between two isotopologues, which vary by specific EH constants.

In panel (a), we can only tell the $|\delta|$ of $D_K$ in the range of 0.002~10%. By changing the vertical axis to linear scale, and using original $\delta$ instead of $|\delta|$, the Fig.22c clearly demonstrates how the $\delta$ correlates with S and O isotope substitutions. The isotope effects on the prediction deviations are linear again in the range of 0-5%, similar to the patterns we observed in Section 3 or Fig.SS1 (and Fig.19b,c). With such EH(Ames) based deviation patterns, it is possible to further calibrate the $D_K$ predictions, if $D_K$ can be accurately determined for a few isotopologues. But many other higher order terms do not have such systematic $|\delta|$ patterns.

Compare Fig.20 and Fig.22, it is clear our EH(Ames) fits do provide far better results for quartic level centrifugal distortion constants. For higher order terms, both the average magnitude and the spread of EH(Ames) constant uncertainties in Fig.20 are usually smaller than those of $|\delta|$ here. In principle, the performance of the $\alpha_A/\alpha_C$ based formula is not bad at all, with mass-difference dependent accuracy. However, the formula requires knowledge of $A/C$, and our test is done with “accurate” rotational constants. If one use roughly estimated $A/B/C$, the scaling factor $\alpha_A/\alpha_C$ will inherit and magnify the $A/C$ deviations, which will add up to the method-dependent deviations we report in Fig.22. Even with the “exact” $A/C$ constants, the quartic level constants from EH(Ames) are still 2-4 orders of magnitude more accurate than those predicted by the $A/C$ dependent formula.

In principle, we are inclined to believe the formula predictions can serve as useful tool to check the noise level of some higher order EH constants. But in reality, there is a bottleneck which obviously requires much more thorough investigation, i.e. the quality of higher order EH(Expt) terms which scientists need to take as standard reference. They must come from reliable least squares fitting of another isotopologue, to which our findings in Fig.20 and related discussions are surely applicable. The SPIFIT fitting residuals of our $1E-7$ cm$^{-1}$ EH(Ames) fits are less than $5E-7$ cm$^{-1}$ or 15 kHz, and $\sigma_{\text{RMS}} \sim 10$ kHz. We believe this is close to the accuracy level of some experimental MW studies. The EH(expt) constant uncertainty will add to the method deviation patterns, if it does not ruin the latter first. In our most recent studies on 626 [3] and 636 [8], if one higher order term was wrong but used in EH analysis of any isotopologue, it would cause collateral damages on other correlated parameters or even noticeably alter the value of rotational constants.

2 Online EH(CDMS) model data https://www.astro.uni-koeln.de/site/vorhersagen/daten/SO2/ : 34SO2/s34.par for 646; SO2/alt.1/so2.par for 626; 33SO2/s33.par for 636; SO18O/o18.par for 628; SO17O/17.int for 627.

3 Huang X, unpublished work and private communications (2018)


8 Huang X, unpublished work and private communications (2017)