Pseudo-Linelists derived from Laboratory Absorption Spectra

Geoffrey Toon
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, Pasadena, CA, 91109

I am an atmospheric scientist making measurements of more than 40 atmospheric gases using the JPL MkIV spectrometer by ground-based solar absorption spectrometry and by balloon-borne solar occultation spectrometry. I also analyze SWIR TCCON spectra to determine column-averaged amounts.

I get frustrated when I systematically get poor spectral fits to a particular region and felt compelled to investigate the cause.

I first search for a proper, quantum-mechanically-based linelist exists for a particular gas and spectral region of interest.

But if a proper linelist doesn't exist in HITRAN or GEISA, or is incomplete, I look for lab transmittance or cross-section spectra. If found, convert these into an empirical pseudo line-list (EPLL).

This talk will describe the process of generating these EPPLs and explain why I think that this is better than using the lab cross-section spectra directly. I will show examples of different EPLLs and their use for atmospheric remote sensing.

This EPLL saga started ~35 years ago when I received low-pressure measurement of CFC cross-sections from McDaniel and Varanasi. How could I make use of these to analyze ground-based spectra that were much smoother than the cross-secs?

What do *I* mean by “Empirical Pseudo-Line-List (EPPL)”

An EPLL is an old-HITRAN-format (100 CPL) linelist (at least the first 72 characters) founded entirely on measured lab transmittance or absorption spectra. Its sole purpose is to be able to represent the lab spectra (and hence also atmospheric spectra) to good accuracy using a Voigt lineshape. EPLLs have no quantum mechanical foundation; the fields in the HITRAN format that should contain quantum numbers, accuracy codes, and references are empty, or filled with comments. The Einstein A-coefficients are set to Zero.

EPPLs are developed for gases/bands for which no linelist exists in HITRAN, or the HITRAN list has major inadequacies (e.g. missing hot-bands). This is generally the case for heavier molecules with 5+ atoms

Below is an example of a small portion of an EPLL of C2H6 (gas 38 in ATM linelist, isotopolog 1)

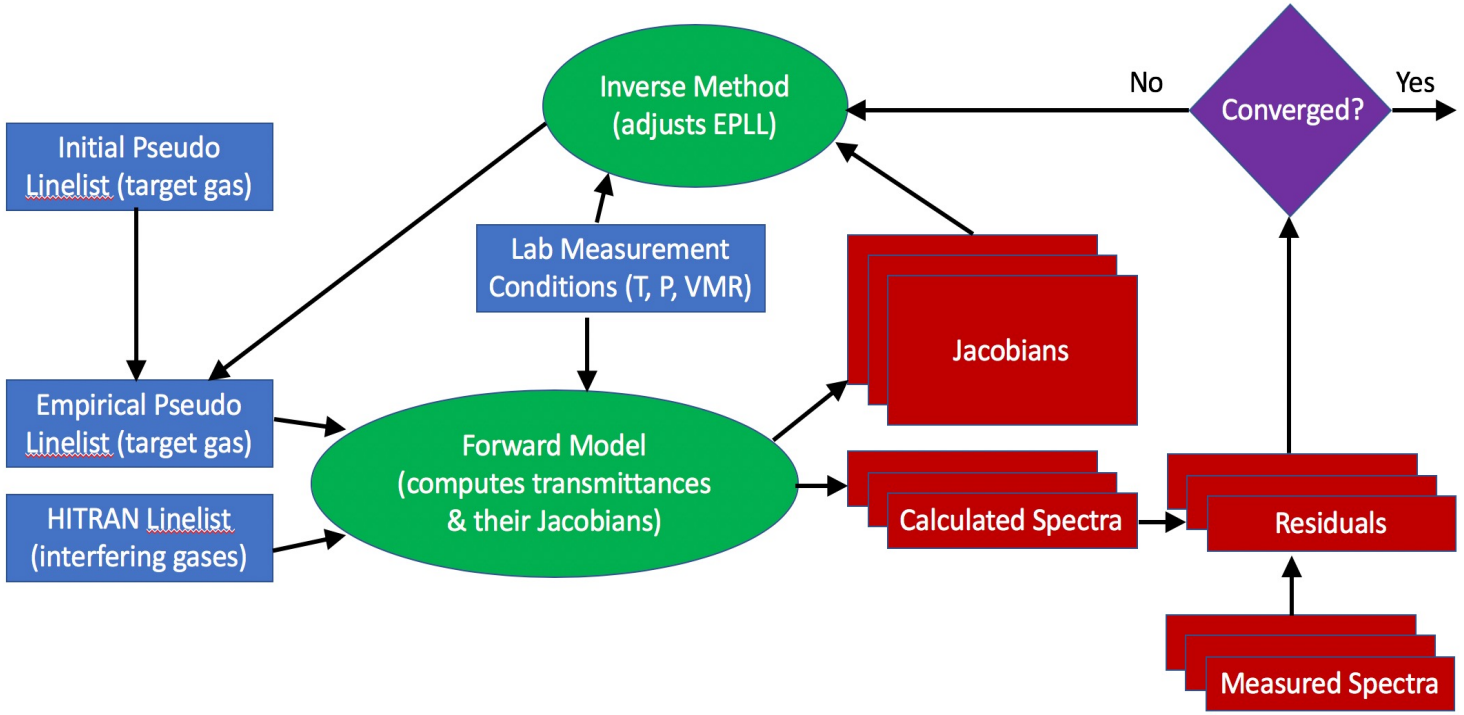
```
381 2770.809126 2.816E-23 0.000E+00.0727.1200  131.96550.80-.005000  PLL from J.Harrison lab spectra
381 2770.825741 3.238E-23 0.000E+00.0667.1200  396.19980.80-.005000  PLL from J.Harrison lab spectra
381 2770.828459 8.053E-24 0.000E+00.0783.1200    0.00000.80-.005000 generated by linefinder  extras
381 2770.829677 8.732E-23 0.000E+00.0669.1200  422.82280.80-.005000  PLL from J.Harrison lab spectra
381 2770.832533 2.219E-22 0.000E+00.0817.1200   34.53710.80-.005000 generated by linefinder  extras
381 2770.834721 1.474E-22 0.000E+00.0654.1200  536.33920.80-.005000  PLL from J.Harrison lab spectra
```

Generating an EPPL

In atmospheric remote sensing, the linelist is assumed known and the atmospheric spectra are calculated using a forward model (e.g., GFIT) and compared with the measured spectra. The assumed atmospheric composition is adjusted in order to produce better fits to a series of atmospheric spectra. This process is iterated until convergence.

In generating an EPPL, the contents of a gas cell are assumed known, and spectra are calculated (using Forward Model) based on the nominal cell conditions and compared with the measured lab spectra. Spectral lines are invoked and/or adjusted to achieve better fits to a series of lab spectra. Then the spectra are re-calculated using the updated linelist and the process is iterated until convergence.

Pseudo-Linelist Generation Flowchart. The GFIT Forward Model computes lab spectra based on the known measurements conditions (T, P, VMR, cell length, MOPD, FOV), the current EPLL, and the HITRAN linelist (for other gases). The Inverse Method looks at the differences (Residuals) between the measured lab spectra and the calculations and decides how best to adjust the EPLL. This entire process is iterated to convergence for each lab spectrum. Green ovals depict programs. Rectangles depict data files.



The benefits of fitting spectral lines to the lab cross-sections

Why not simply use the laboratory-measured cross-sections directly in the forward model when performing atmospheric remote sensing, interpolating in T/P and wavenumber?

There is absolutely nothing wrong with this if the cross sections are free from artifacts (zero offsets, channel fringes, contamination, ILS broadening), cover the wavenumbers of interest, fully bracket the atmospheric T/P range, and have accurate reported T/P information.

But in my experience it is rare that a set of cross sections are this perfect. And fitting an EPPL to the cross-sections provides an ***opportunity*** to identify and correct these defects.

Fully resolved lab spectra are rare because it is very time consuming. So sharp spectral features are always slightly broadened by the ILS of the laboratory spectrometer that measured the cross-sections. So if the cross-sections were to be used directly for atmospheric remote sensing, they get convolved with the ILS of the spectrometer that measures the atmospheric spectra, on top of that of the lab spectrometer. Generating an EPLL implicitly removes the ILS broadening, if the ILS is known from MOPD and FOV, because the infinite-resolution spectrum generated from the EPLL is convolved with the ILS before comparison with the measured cross-section spectra. So essentially, an implicit deconvolution of the ILS occurs.

Also, zero level offsets in the lab spectra can be determined and removed. This is especially important if any of the absorption features are close to saturation. And channel fringes in the lab spectrum can be fitted, and therefore won't propagate into the residuals, from which the EPLL is adjusted. Finally, the wavenumber calibration of each spectrum can be checked (and if needed, corrected) by comparing the positions of any sharp absorption features (including interfering H_2O lines).

Theoretical Basis - Summary

A more detailed explanation of the theoretical Basis can be found in the Supplemental Material. This is the 1-slide version.

In summary, after fitting a bunch of lab spectra with GFIT, we assume that the unknown strength $S^m_i(296)$ and E''^m_i of the i^{th} pseudo line are related to the measured an calculated transmittances by the equation

$$\ln\{VSF_j \ln[T^m_{i,j}]/\ln[T^c_{i,j}]\} = \ln[S^m_i(296)/S^c_i(296)] + hc(E''^m_i - E''^c_i)(1/k296 - 1/kt_j) \quad j=1,Nspec$$

where i is an index denoting the lab spectrum and j denotes the pseudoline, VSF_j is the retrieved VMR scaling factor.

Figure depicts plotting $y=\ln\{VSF_j \ln[T^m_{i,j}]/\ln[T^c_{i,j}]\}$ versus $x=hc(1/k296-1/kt_j)$, and fitting a straight line $y = a + bx$. results in new, better values for $S^c_i(296)$ and E''^c_i since

$$E''^m_i = E''^c_i + b$$

$b = (E''^m_i - E''^c_i)$, is the required adjustment to the currently assumed E".

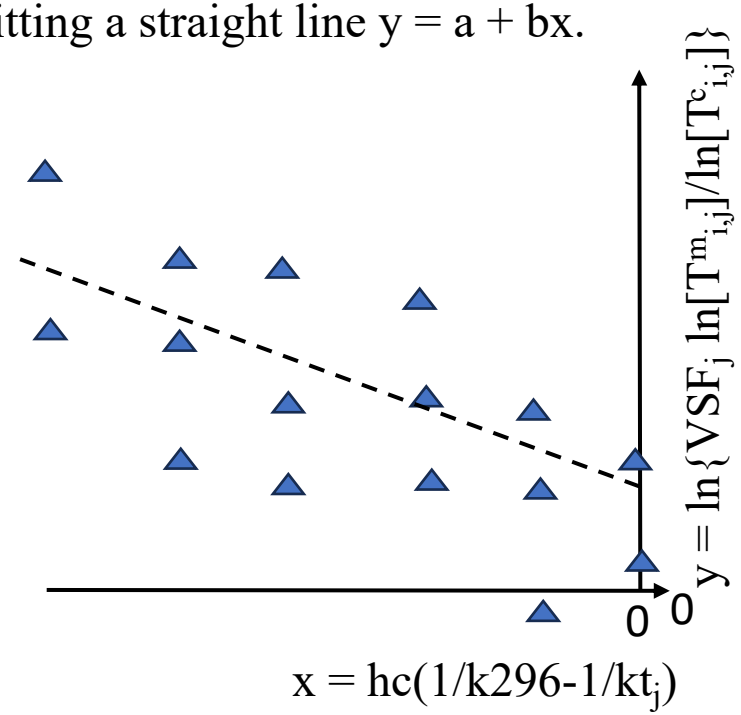
The x=0 intercept to is at $a = \ln[S^m_i(296)/S^c_i(296)]$, from which

$$S^m_i(296) = S^c_i(296)Exp[a]$$

x=0 corresponds top t_j = 296K; $x < 0$ corresponds to colder t_i

y=0 corresponds to $T^m=T^c$

Positive y-values mean $T^c_i > T^m_i$ and so the 296K intensity $S^c_i(296)$ needs increasing.



Case Study 1: CFH_2CF_3 (HFC-134a) Empirical Pseudo-Line-List

Another issue with using cross-sections directly in the Forward Model occurs when there are multiple datasets available. Which one to use? Or can they somehow be averaged or merged together first? Not easy if acquired at different temperatures or pressures or spectral resolutions, or if the cross-sections datasets have systematic biases.

In the case of CFH2CF3 (HFC-134a) six datasets of cross-sections have been acquired. An EPLL was developed by simultaneously fitting to ***all*** the available lab spectra from these datasets and adjusting the line intensities and E" to minimize the overall residuals. This process was repeated until convergence. Then the widths were changed slightly and the S and E" were re-derived. If the RMS fitting residuals and the spectrum-to-spectrum consistency of the retrieved gas amounts (VSF) improved, then the widths were changed again in same direction.

In this respect, an EPPL can be seen as a way of aggregating cross-section data from multiple data-sets. In doing this, systematic biases become apparent, that might not be obvious from the cross-sections themselves, due to their different T/P, resolutions.

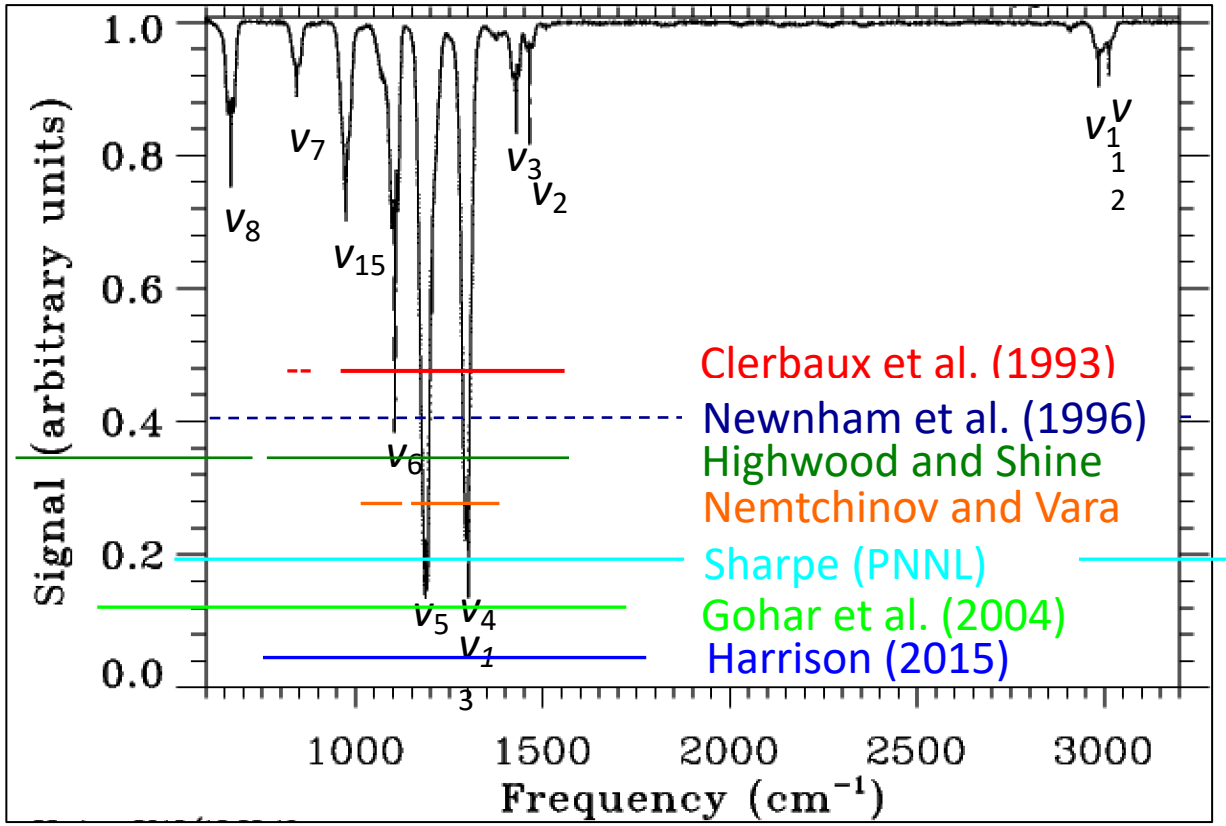
In 2018, six data-sets of laboratory spectra, downloaded from the HITRAN website, were used to generate an empirical pseudo-linelist. Transmission spectra, generated from the lab cross-section data, were fitted simultaneously in order to minimize residuals between the lab spectra and synthetic spectra calculated from the PLL. In some spectral regions up to 70 spectra were used.

The full 24-page report can be found at: https://mark4sun.jpl.nasa.gov/data/spec/Pseudo/HFC-134a_PLL.compressed.pdf

Available CFH_2CF_3 (HFC-134a) Laboratory Spectra

Seven different CFH_2CF_3 spectral datasets have been measured. Six of these were down-loaded from the HITRAN website and used to derive the PLL. Preference was given to those in the supplemental folder, where the –ve absorption coefficients have not been zeroed. Figure illustrating main absorption bands of CFH_2CF_3 and the coverage of the laboratory spectra

#	Reference	Range (cm^{-1})	Temp (K)	Pres (Torr)	VSF	N_s
A	Clerbaux et al. [1993]	(815- 865) 935-1484	253-287		0.72	0/3 3/3
B	Newnham et al. [1996]	600-3500	203-294	38-760	-	0/3
C	Highwood and Shine, [2000]	75 - 590 600-1540	253 253	< 1 < 1	1.00	1/1 1/1
D	Nemtchinov and Varanasi [2004]	1035-1130 1135-1340	180-297	< 1 < 1	1.00	32/49 33/49
E	Sharpe et al. [2004]	600-5000	278-323	760	1.00	3/3
F	Gohar et al. [2004]	300-1550	296	700	0.95	1/1
G	Harrison [2015]	750-1600	191-296	24-760	1.00	27/27



Footnotes:

1. Newnham et al. are included in the table for completeness; the spectra are missing from the HITRAN website.
2. Only 32½ of 49 tabulated Nemtchinov spectra are available from HITRAN: the lowest-P spectra at each temperature are missing, as are the all the lowest-T (180K) spectra.
3. Clerbaux's spectra covering 815-865 cm^{-1} were too narrow and too noisy to be useful.

Twelve Fitted Windows

Defined 12 windows, each centered on an absorption band. Chose each window based on three criteria:
1) Boundaries in regions of weak absorption
2) Maximize utilization of the various datasets
3) Maximize spectral coverage

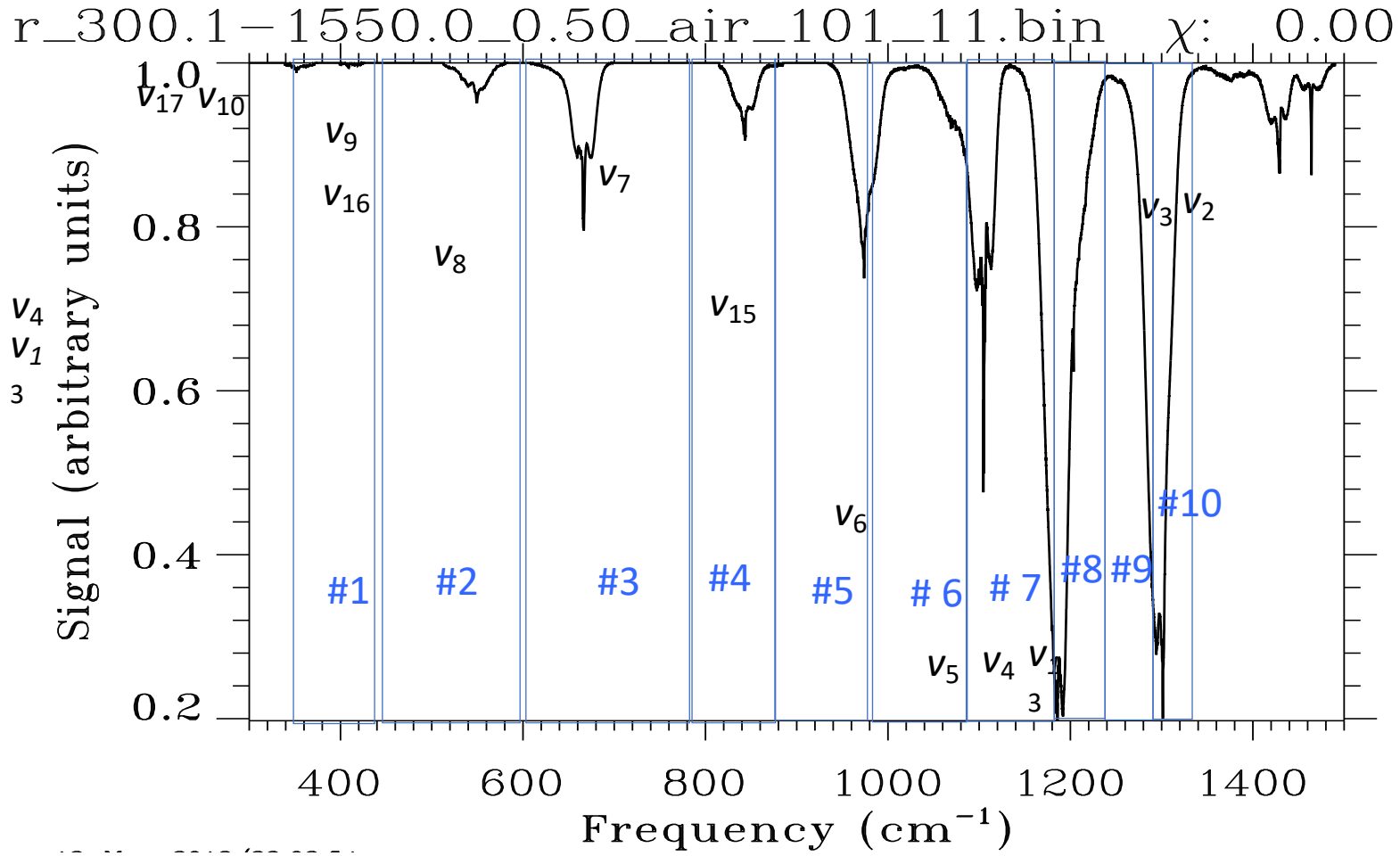
The right-most column (N_S) shows the total number of spectra that were fitted in each window. In total 412 spectral fits were performed out of a potential 103 x 12 = 1236

Although there are absorption bands visible below 500 cm^{-1} in the spectrum of Highwood and Shine, these are substantially weaker than the ν_9 band centered at 549 cm^{-1}. Based on their weakness and the very limited lab spectra, it was decided not to extend the PLL below 500 cm^{-1}.

Dataset B (Newnham) is currently missing.

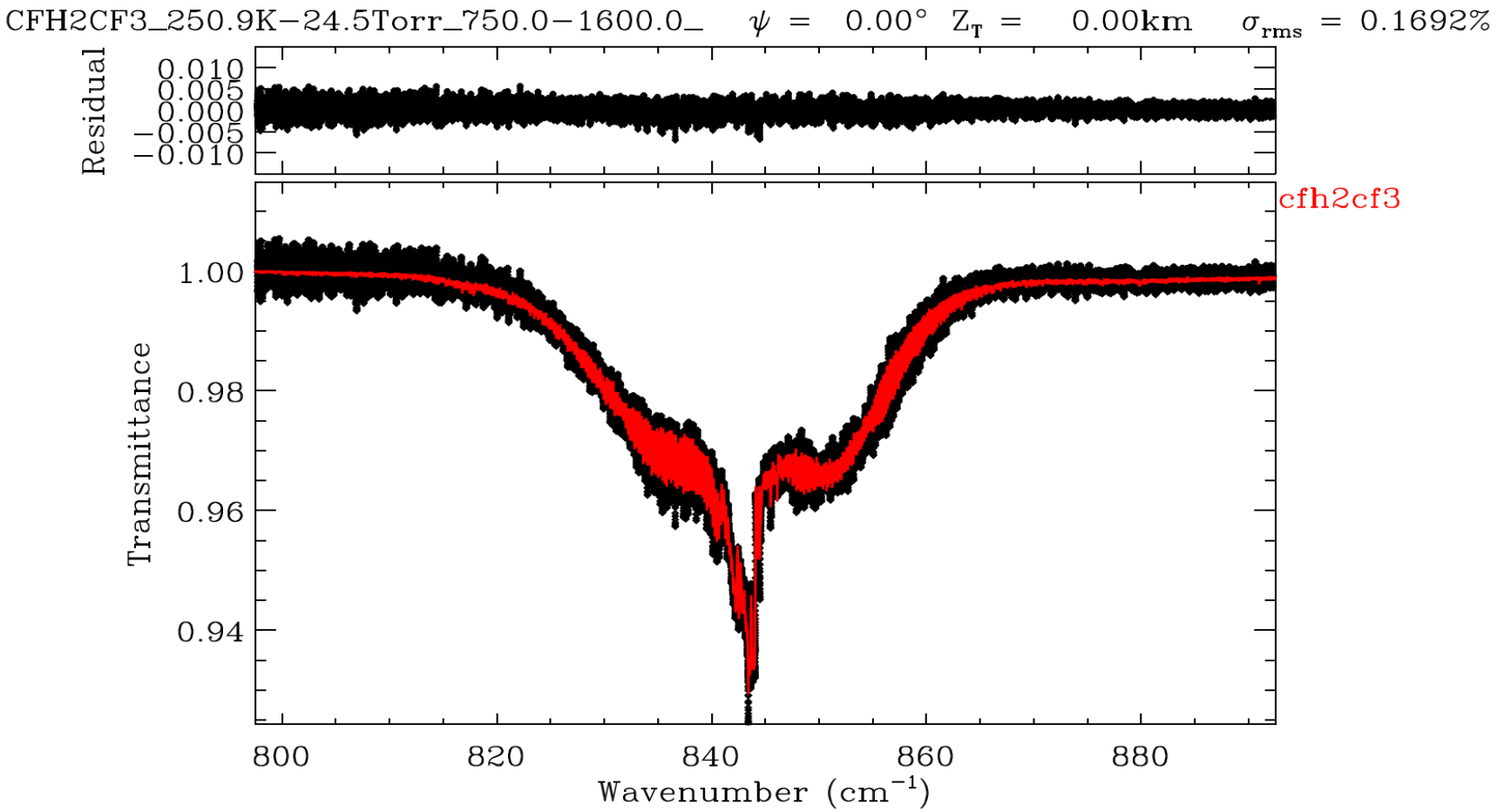
#	Center (cm^{-1})	Width (cm^{-1})	Datasets covered	N_{Spec}
1	548.00	83.00	C F	2
2	676.05	147.90	C EF	5
3	845.00	189.55	C EFG	34
4	987.50	95.00	A C EFG	37
5	1082.50	94.15	A CDEFG	68
6	1189.00	107.15	A CDEFG	70
7	1291.50	96.15	A CDEFG	69
8	1368.05	56.90	A C EFG	37
9	1421.60	50.20	A C EFG	37
10	1465.15	36.90	A C EFG	37
11	1511.75	55.50	C E G	33
12	2920.00	260.00	E	3

Gohar et al. 278K spectrum



Illustrating main absorption bands and fitted windows (#). Due to the weakness of the bands below 500 cm^{-1}, and the sparsity of the lab spectra, the PLL was not extended below 500 cm^{-1}.

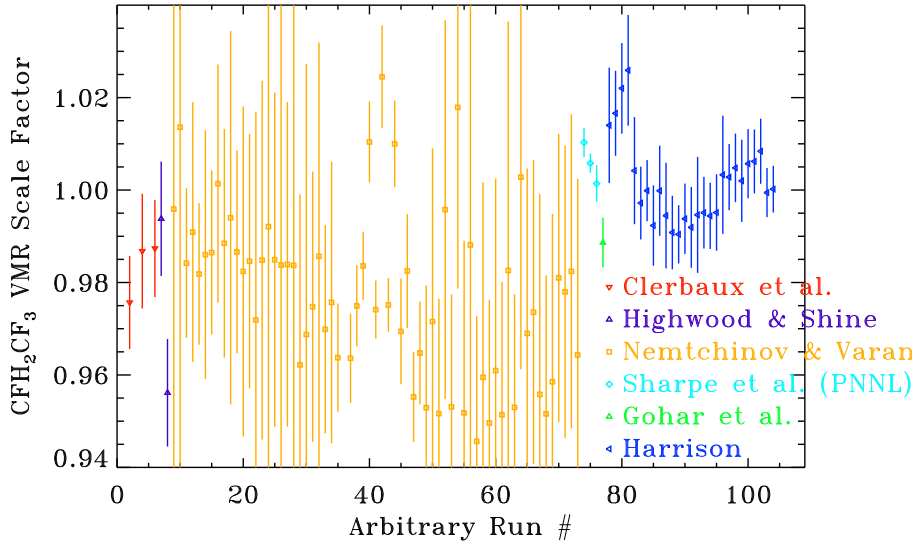
Example of Spectral fit to 251K, 25 Torr, Harrison CFH_2CF_3 lab spectrum

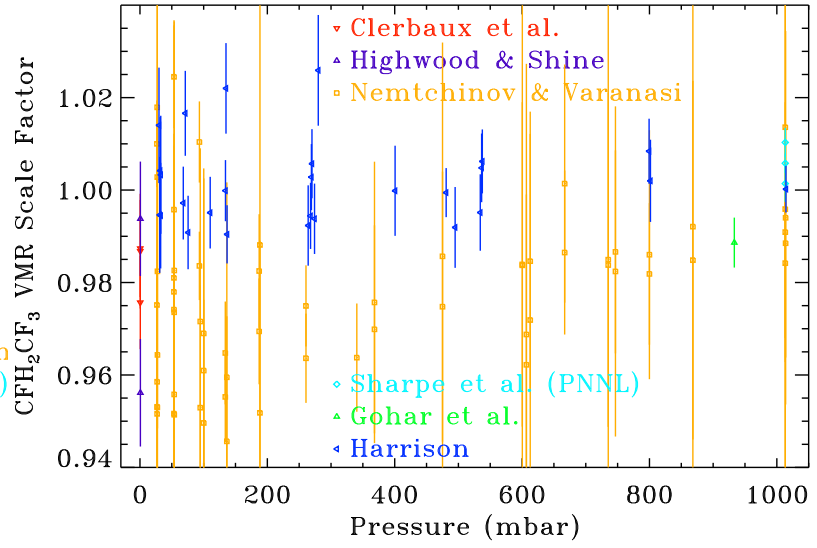


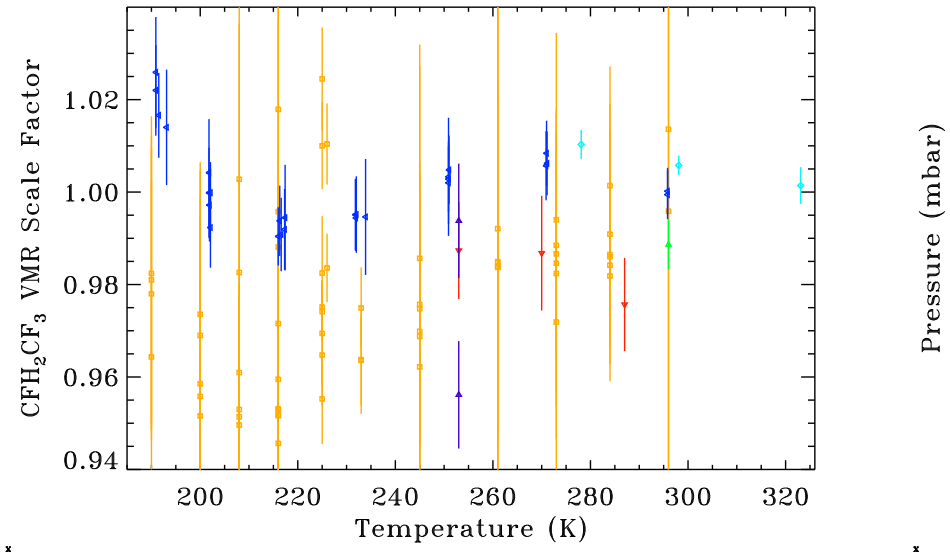
Note that I have converted the measured absorption cross sections into a transmittance spectrum, via knowledge of the cell conditions, before fitting for the pseudo-line parameter, because my analysis software assumes white noise, which is true in an FTS spectrum. In a cross-section spectrum, however, T = Exp[-L.D.Xsec] and so noise (and systematic errors) become proportional to the cross-section

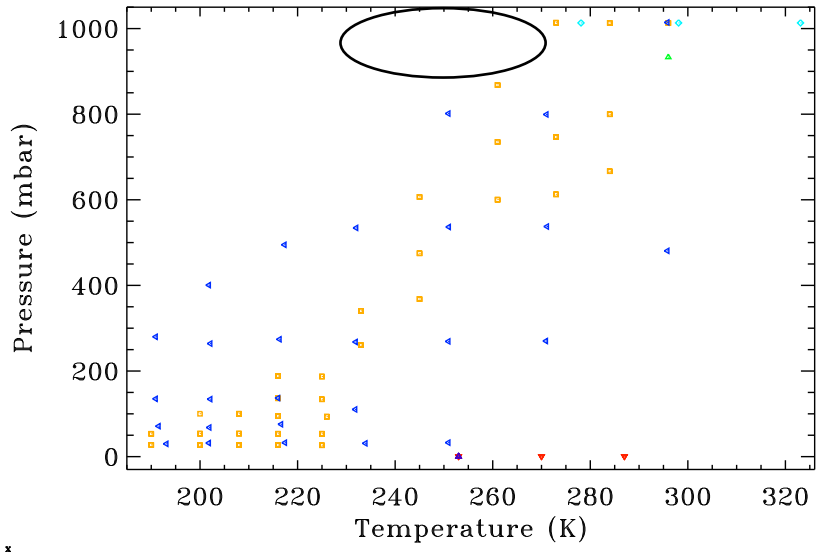
$$Xsec = -1/L/D \ln(T_m \pm \varepsilon) = -1/L/D \ln(T_m) \ln(1 \pm \varepsilon/T_m) \simeq -1/L/D \ln(T_m) \pm \varepsilon\, Xsec / T_m$$

Retrieved CFH_2CF_3 VMR Scale Factors (VSFs)









The VMR scale Factors (VSFs) are all close to 1.0, after scaling the Clerbaux cross-sections by 0.72 and Gohar's by 0.95.

There is little correlation with T or P, suggesting that the derived E" and width values are good.

Lower-Right: Note the dearth of measurements below 260K at 1 atm (empty oval). Such conditions are common over Canada, Russia and Antarctica in winter.

Yet it is rare for lab cross-sections to cover these conditions for any gas, requiring extrapoloation in T/P

CFH_2CF_3 Summary

Six different lab spectral datasets of CFH_2CF_3 absorption have been fitted simultaneously to generate a single PLL to represent their absorption.

103 spectra were fitted over 12 different windows. Out of the 1236 potential spectral fits, only 412 could actually be performed. This was mainly due to the limited coverage of many spectra and a few data quality issues.

The resulting pseudo-linelist contains:

- 209,201 lines covering 504-1550 cm^{-1} at 0.005 cm^{-1} spacing
- 26,001 lines covering 2790-3050 cm^{-1} at 0.010 cm^{-1} spacing

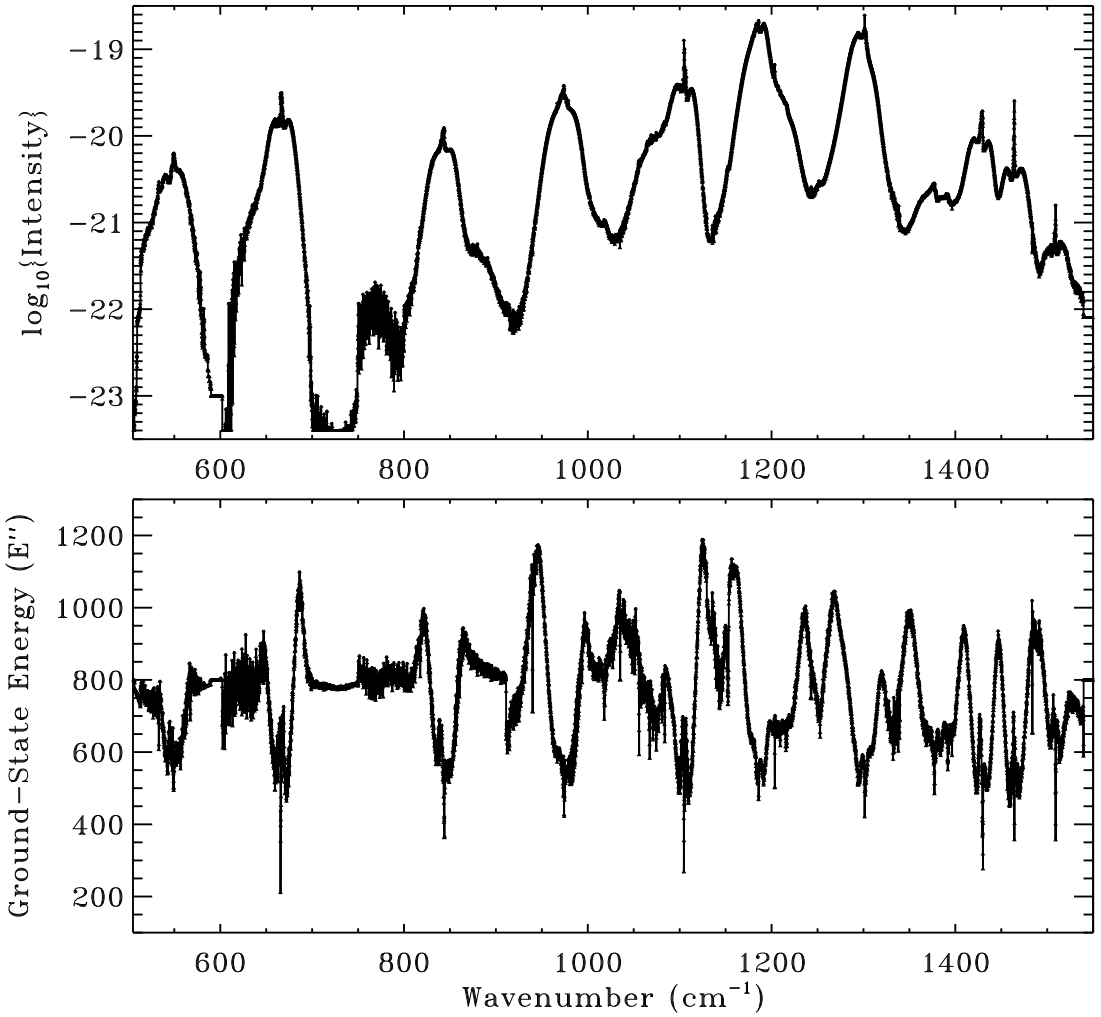
235,202 total lines

The linelist uses the old HITRAN 2000 format with 100 CPL.

Gas # 76 was arbitrarily assigned to CFH_2CF_3.

Top Panel: The derived intensities

Bottom Panel: The derived pseudoline E"s. Note that the lowest E" values correspond to Q-branches



Case Study 2: Propane (C_3H_8)

EPLL was generated in 2014 covering 2765 – 3080 cm^{-1} at 0.005 cm^{-1} spacing (63001 lines) based on Harrison's and Bernath's laboratory measurements.

[This is different from the pseudo-linelist covering 690-1550 cm^{-1} described by Sung et al. [2013], which was based on Sung's own lab measurements.]

Assumes:

- ABHW = 0.07
- SBHW = 0.14

Line intensities and E" were retrieved.

Assumed partition function following Sung et al. [2013]

- Vibrational: Used 25/27 vibrational modes (dropping torsional modes)
- Rotational: $(296/T)^2$

Journal of Quantitative Spectroscopy & Radiative Transfer 111 (2010) 1282–1288

Contents lists available at ScienceDirect

Journal of Quantitative Spectroscopy & Radiative Transfer

journal homepage: www.elsevier.com/locate/jqsrt

Infrared absorption cross sections for propane (C_3H_8) in the 3 μm region

Jeremy J. Harrison *, Peter F. Bernath

Department of Chemistry, University of York, Heslington, York YO10 5DD, UK

ARTICLE INFO

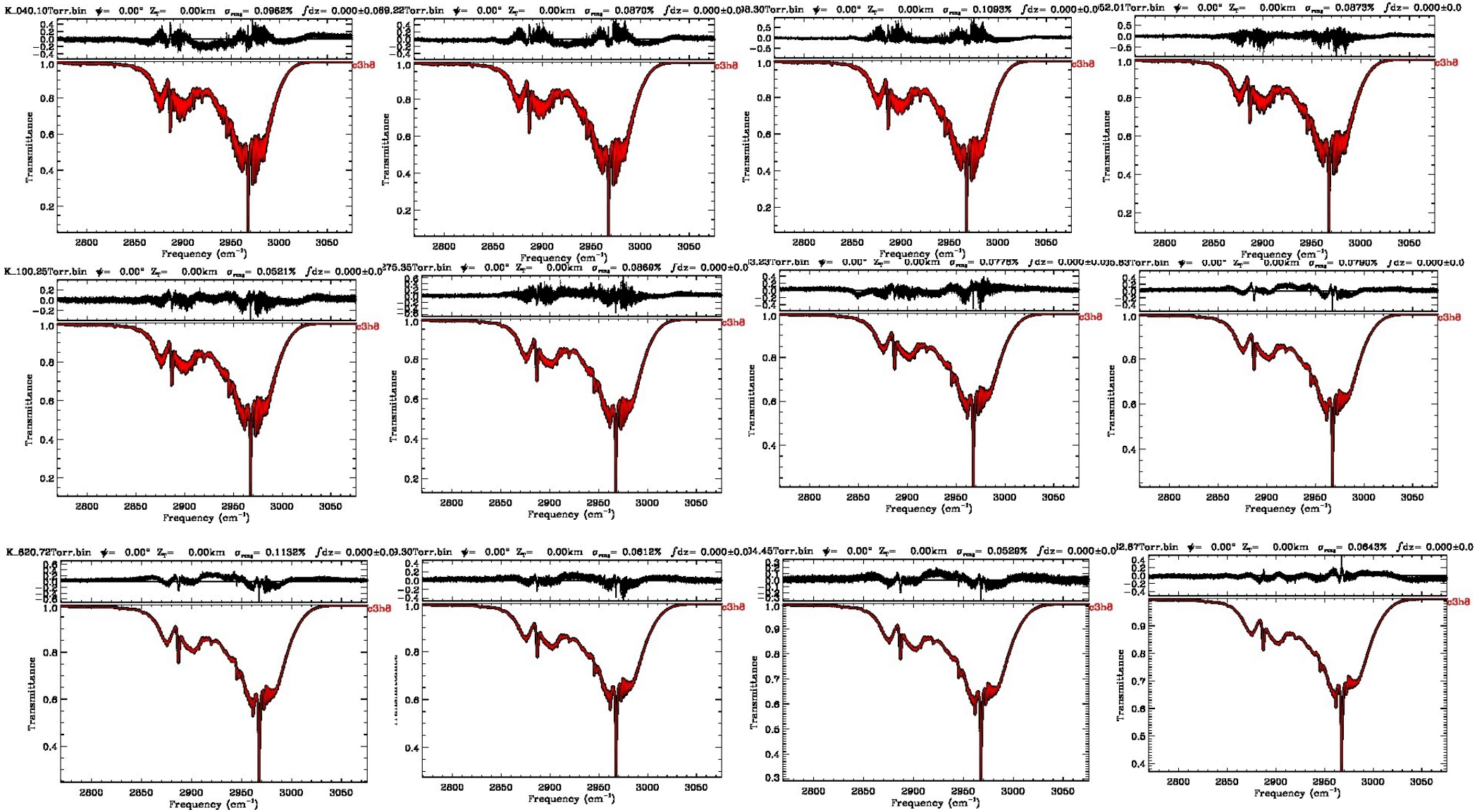
Article history:
Received 3 October 2009
Received in revised form
20 November 2009
Accepted 28 November 2009

Keywords:

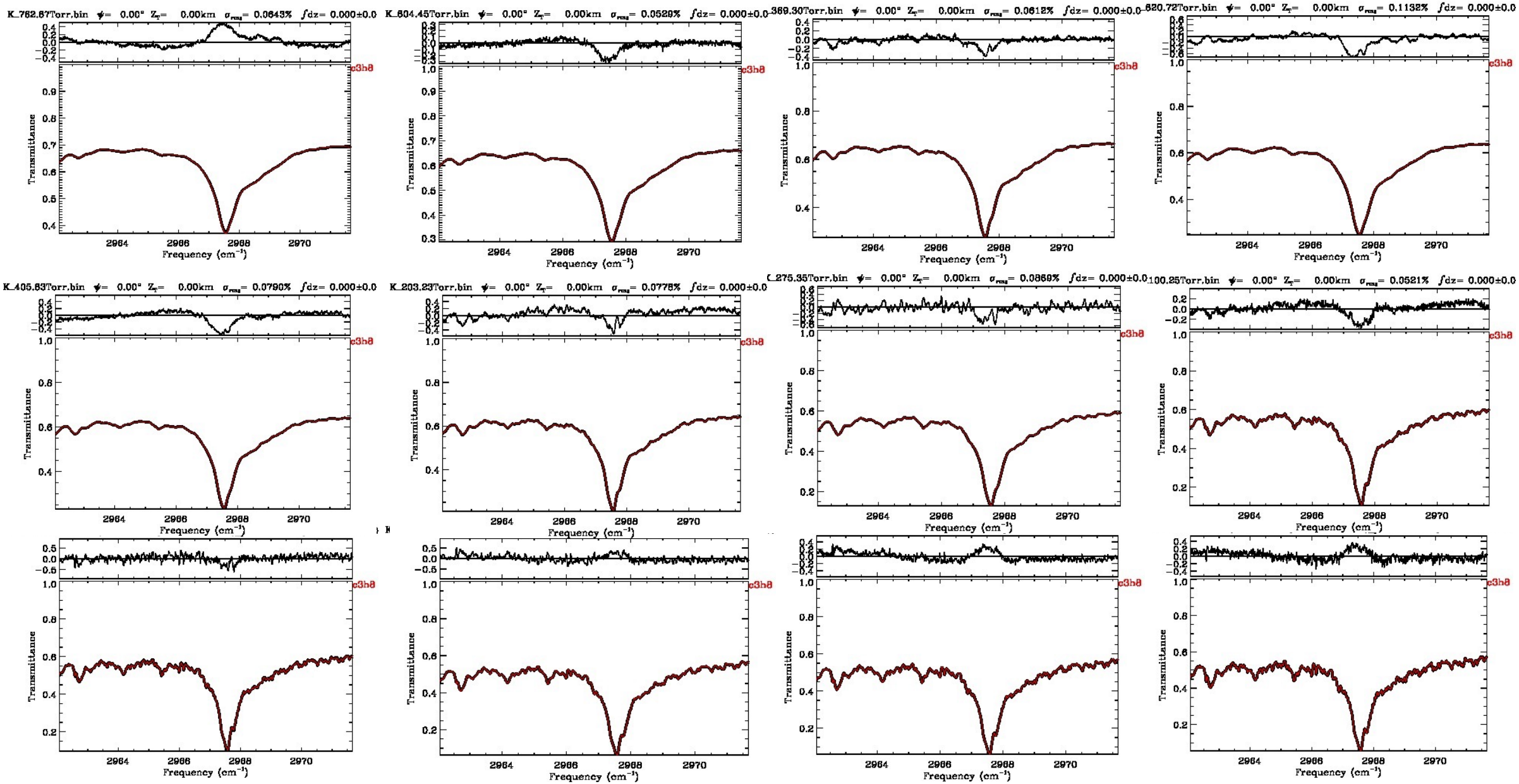
ABSTRACT

Infrared absorption cross sections for propane have been measured in the 3 μm spectral region from spectra recorded using a high-resolution FTIR spectrometer (Bruker IFS 125 HR). The spectra of mixtures of propane with dry synthetic air were recorded at 0.015 cm^{-1} resolution (calculated as 0.9/MOPD using the Bruker definition of resolution), at a number of temperatures and pressures appropriate for atmospheric conditions. Intensities were calibrated using two propane spectra (recorded at 278 and 293 K) taken from the Pacific Northwest National Laboratory (PNNL) IR database.

Fits to C_3H_8 Lab spectra of different P/T using derived EPLL : Full Band



Fits to C_3H_8 lab spectra of different P/T using derived EPLL: Zoom into Q-branch

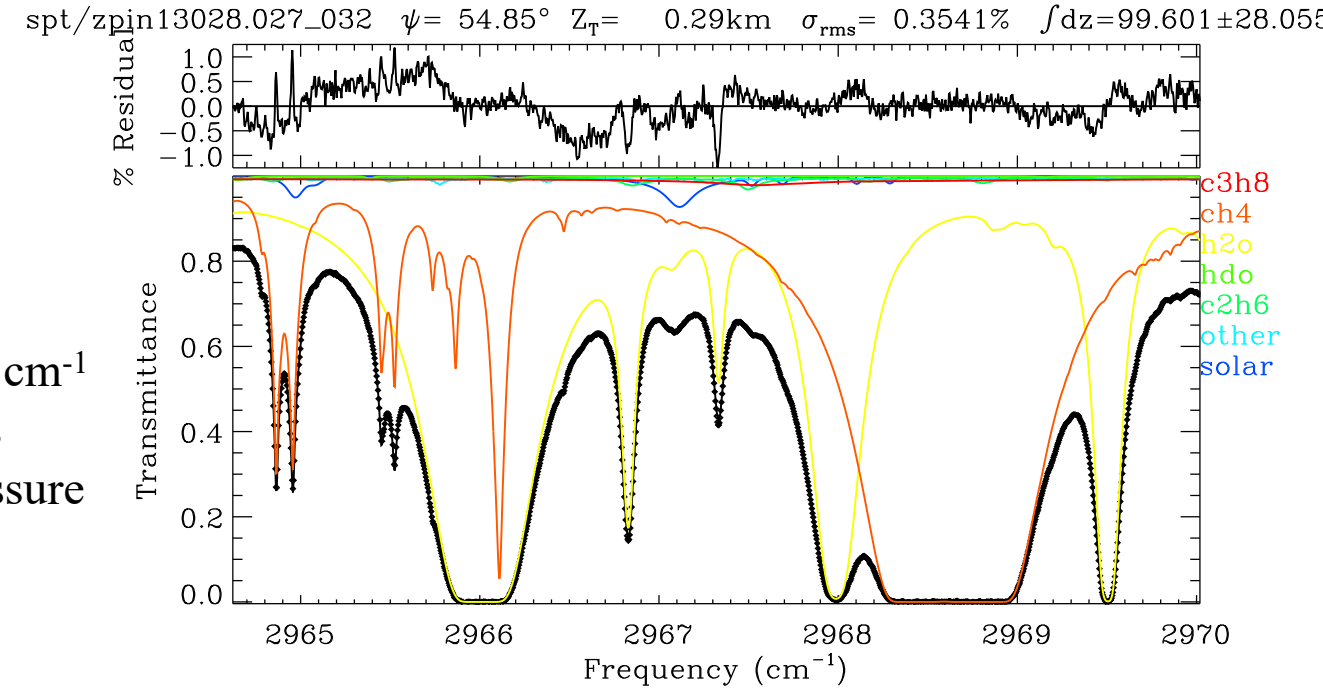


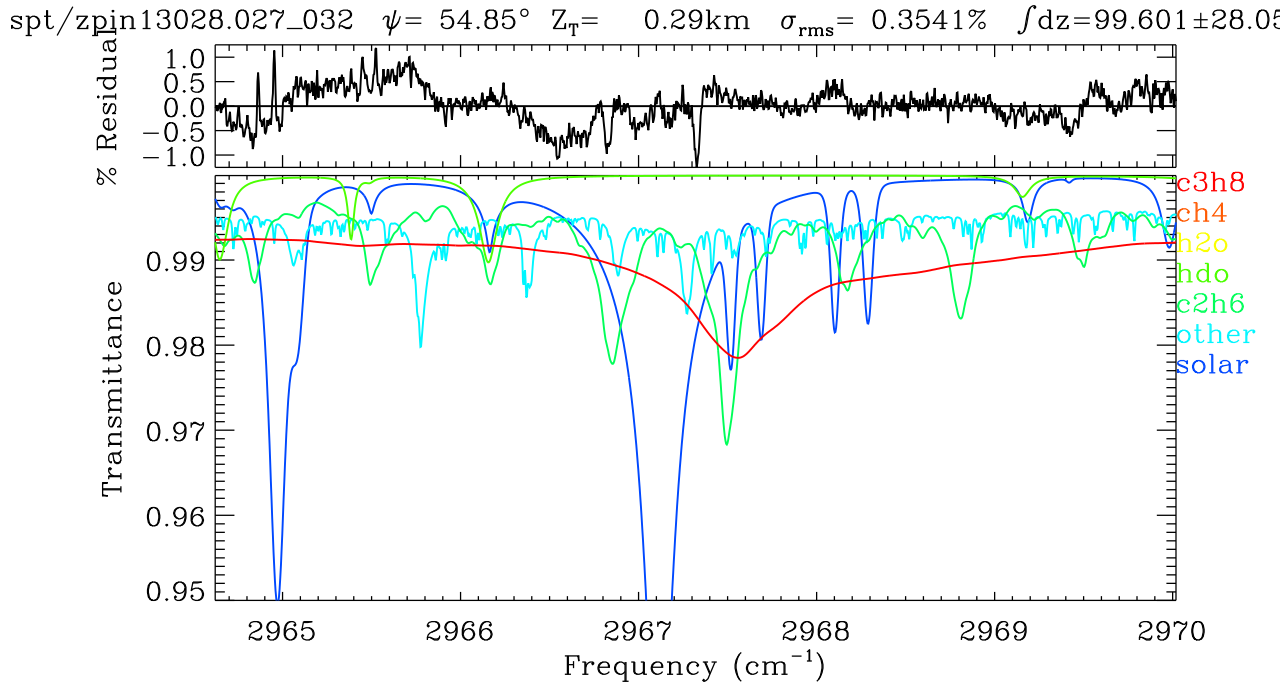
Fits to MkIV ground-based spectrum

C_3H_8 Q-branch lies in wings of strong CH_4 lines at 2968-2969 cm^{-1} and strong H2O lines at 2966.0 cm-1. So, the retrieved C_3H_8 is very sensitive to assumptions about CH_4 and H2O widths, pressure shifts, and line mixing.

Lower panel shows same data, but y-zoomed to see the weaker absorption contributions e.g. C_3H_8, C_2H_6, solar.

Also, C_3H_8 Q-branch overlaps a 3% deep C_2H_6 feature at 2967.5 cm^{-1}. Retrieved C_3H_8 will therefore be sensitive to errors to the C_2H_6 spectroscopy.





Case Study 3: C_2H_6 EPLL

In 2009 I created an ethane (C2H6) EPLL covering the 3000 cm^{-1} region. This was based mainly on Harrison's lab data.
Jeremy J. Harrison, Nicholas D. C. Allen, and Peter F. Bernath, Infrared absorption cross-sections for ethane (C2H6) in the 3 um region, JQSRT, 111, 357–363, 2010

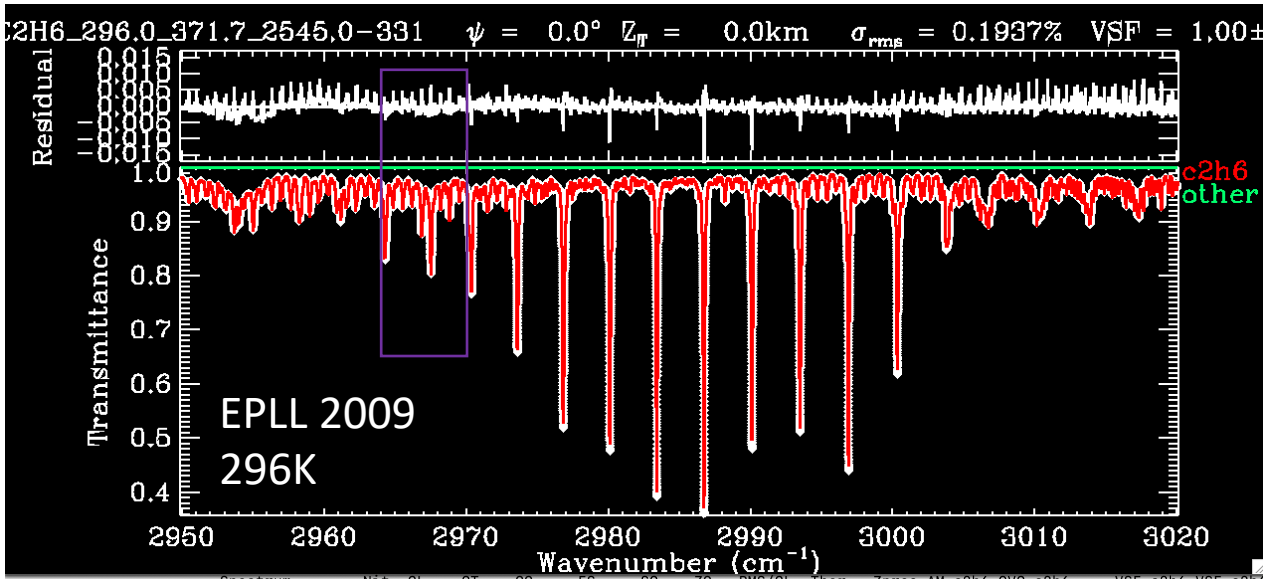
At that time, HITRAN C2H6 was very poor, with only the strongest PQ branches represented.
The derived EPLL is described here:
https://mark4sun.jpl.nasa.gov/report/C2H6_spectroscopy_evaluation_2850-3050_cm-1.compressed.pdf

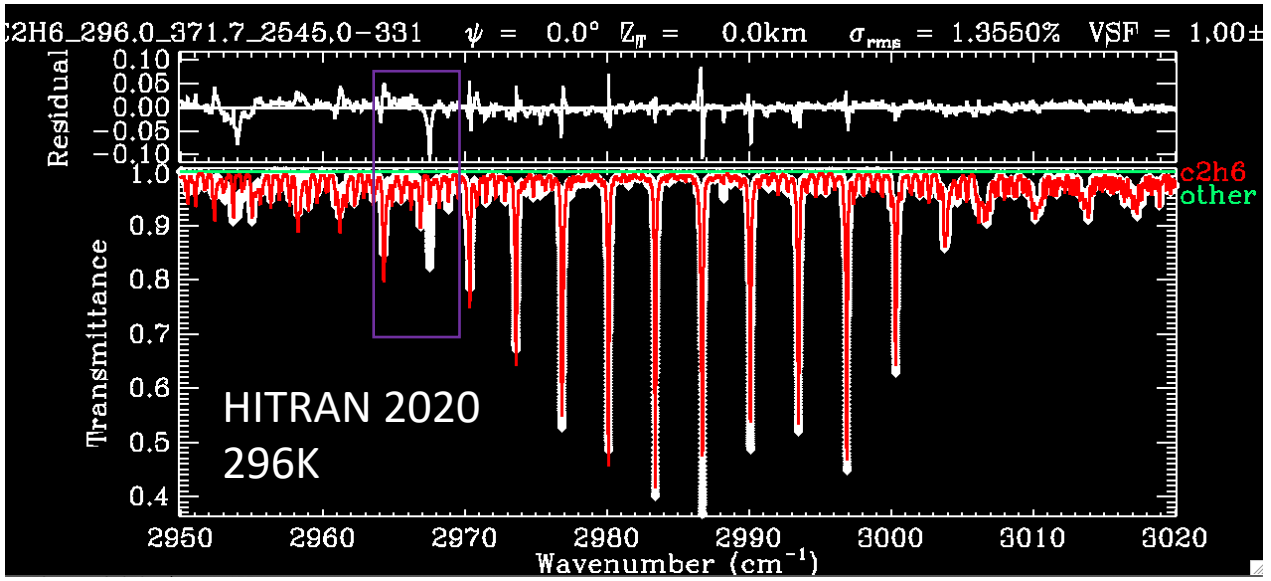
Gordon et al. (2022) described a new C_2H_6 linelist covering 2800 to 3071 cm^{-1} in HITRAN 2020.

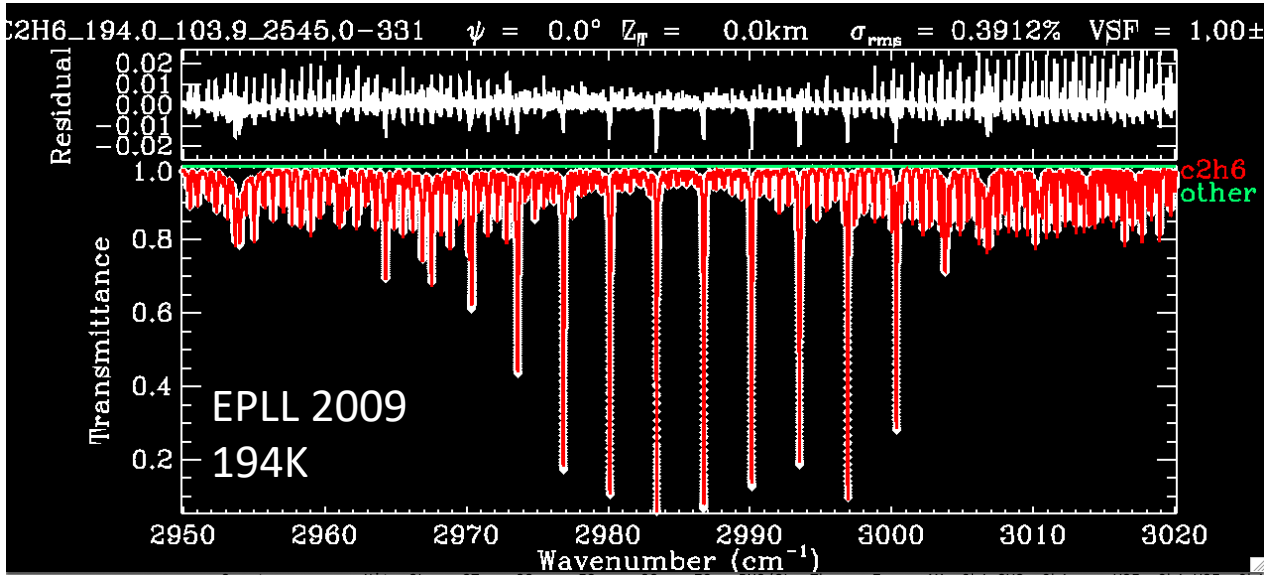
So I refitted the Harrison lab spectra using HITRAN 2020 and compared it with results using the EPPL. The latter gives much better fits.

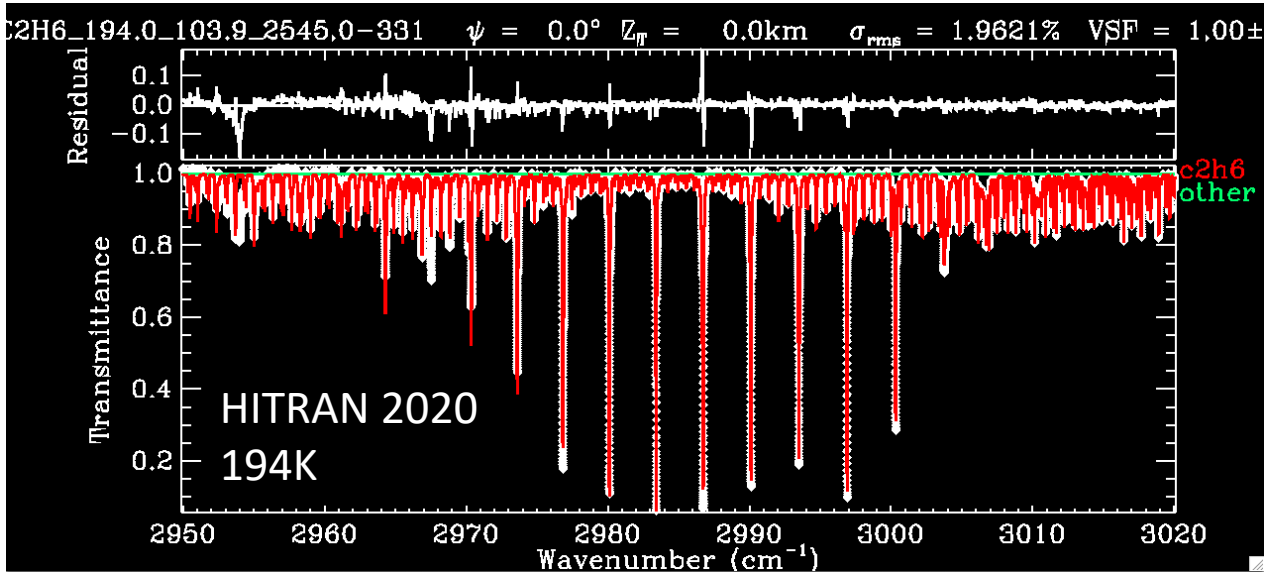
Of course, one could argue that this is an unfair comparison since I am using the same lab spectra (Harrison et al.) to evaluate the linelists as I used to generate the EPPL. I should really use an independent dataset. That said, for each pseudo-line, there will be 27 spectra covering it, so the process is heavily over-determined: We are trying to extract 2 pieces of information about each line (intensity, E") from 27 lab spectra.

Harrison et al. (2010) lab C_2H_6 spectra at 296K (Top) and 194K (Bottom) fitted using 2009 EPPL (Left); HITRAN 2020 (Right). Note the factor **8** scale change for the "Residual" panel between the EPPL fits and the HITRAN 2020 fits. Note the missing absorption 2967.6 cm^{-1}. The C_3H_8 window is denoted by purple rectangle.





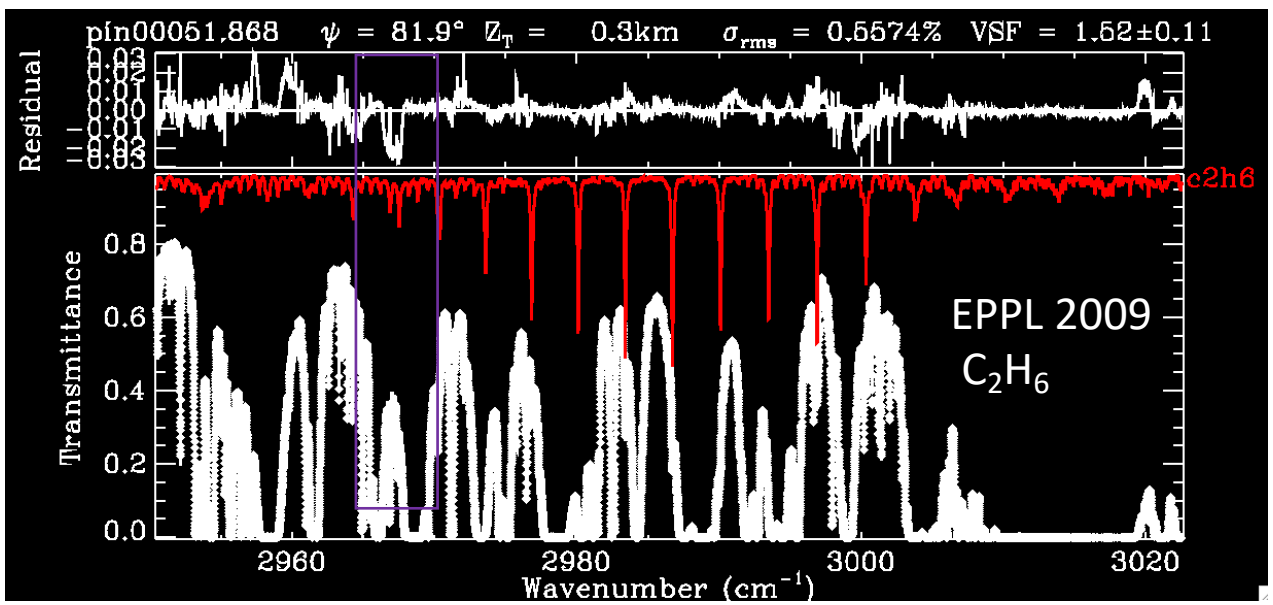


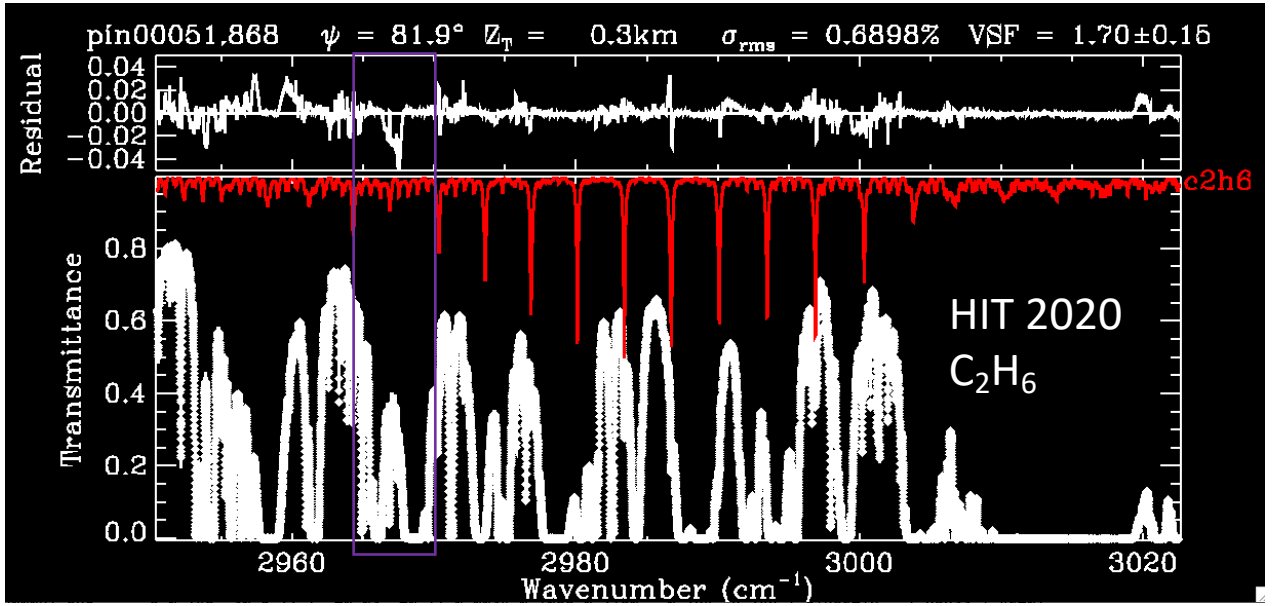


Fits to MkIV ground-based Atmospheric Spectra

Left: Example of fit using C_2H_6 EPLL. Right: Fit made using HITRAN 2020 C_2H_6 spectroscopy. Note change of residual scale from 3% on left to 5% on right.

Residuals are dominated by CH_4, mainly due to my neglect of line-mixing. But non-C2H6 spectroscopy is identical between the two panel, as are the other parameters (assumed atmospheric T, P, and VMR profiles). So the differences in fitting residuals are entirely due to C_2H_6 spectroscopy.

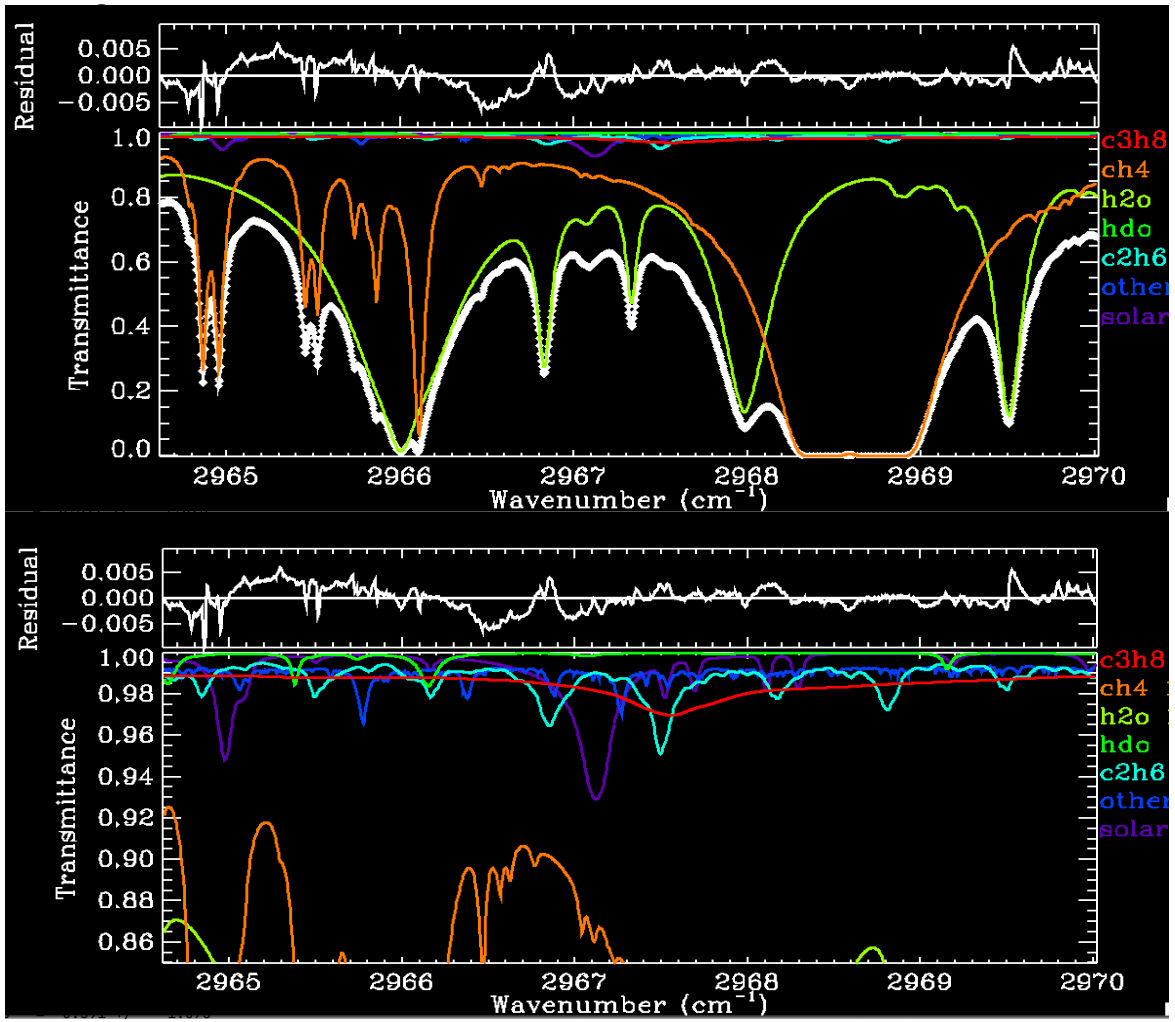


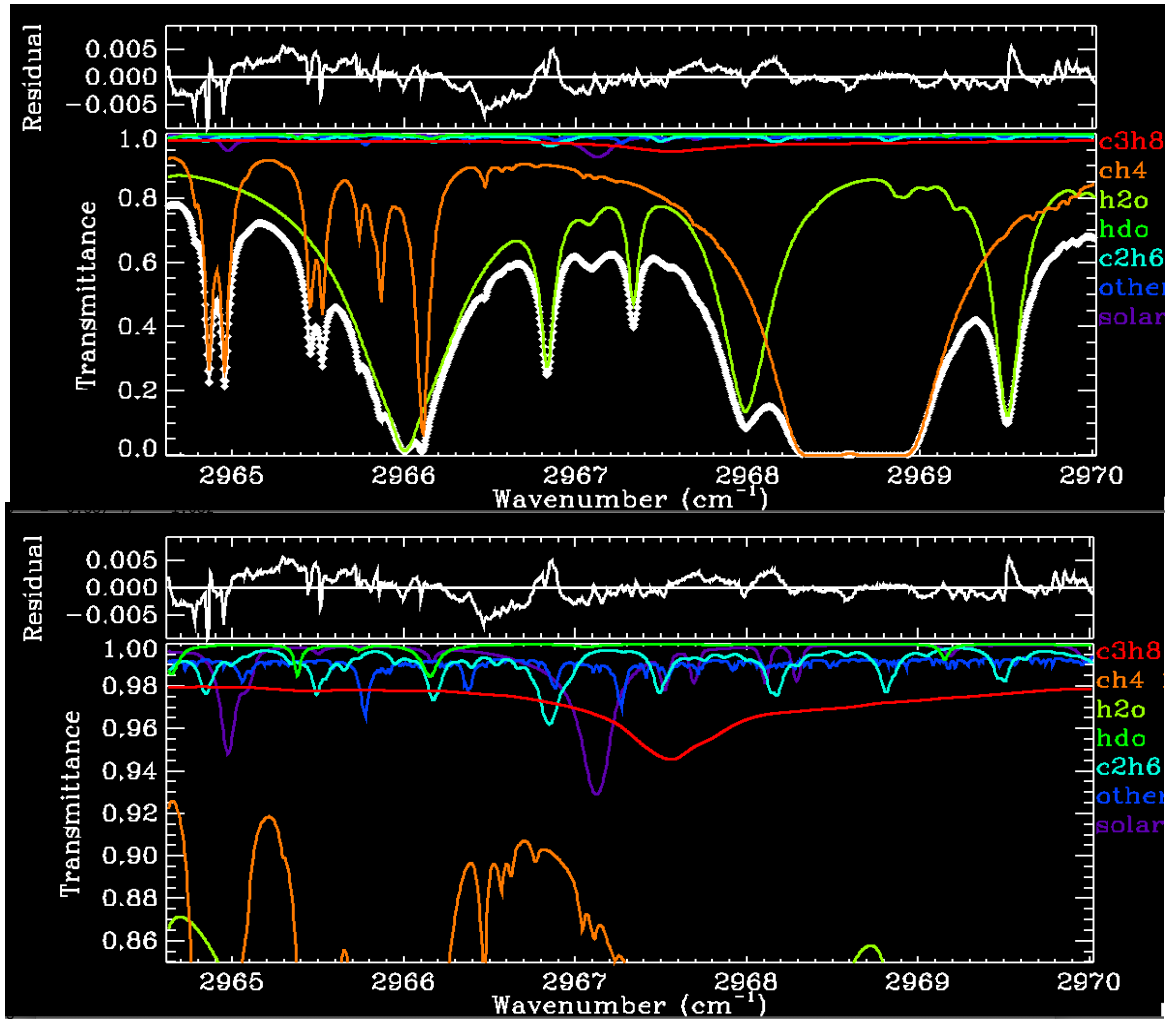


Fits to MkIV ground-based Atmospheric Spectra – in C_3H_8 window

Left: Example of fit using C_2H_6 EPLL.

Right: Fit made using HITRAN 2020





Only small differences here in RMS fitting residuals. But retrieved C_3H_8 changes by a factor 3 because C_2H_6 feature at 2967.5 cm^{-1} is much stronger in EPPL than in HITRAN 2020 so the C_3H_8 reduces to compensate. So, for accurate C_3H_8 retrievals, almost perfect C_2H_6 spectroscopy is needed. The CH_4 and H_2O need improving too!

Full report: https://mark4sun.jpl.nasa.gov/report/C2H6_spectroscopy_evaluation_2850-3050_cm-1.compressed.pdf

2.27. C_2H_6: *ethane* (molecule 27)

Ethane (C_2H_6) is the most abundant non-methane hydrocarbon (NMHC) in the atmosphere of the outer planets [534] and Titan [535], playing an important role as a tracer of atmospheric chemistry and dynamics. Ethane is also an important constituent of comets and their gaseous envelopes [536]. The relative abundance of isotopic species of ethane, such as D/H ratio from C_2H_5D/C_2H_6, can carry valuable information about the atmospheric formation and chemical evolution. In this work, we have expanded the ethane line list in HITRAN to include the ν_5, ν_7 and underlying combination bands of $^{12}C_2H_6$ and the ν_4, ν_{12}, and $2\nu_6$ bands of $^{12}C_2H_5D$ from recent model predictions validated through a laboratory study.

2.27.1. Region of ν_5 and ν_7 fundamentals (2800–3071 cm^{-1})

Until this present edition, HITRAN contained only strong Q-branch lines of the ν_7 band in the spectral region around 3.3 μm. Nevertheless, these lines alone are insufficient to correctly interpret atmospheric and planetary spectra and a better high-resolution spectroscopic model was needed. This spectral range is dominated by the C-H stretching fundamental of ν_5 (parallel band) and ν_7 (degenerate perpendicular band), and the $\nu_8+\nu_{11}$ combination band (ν_8 and ν_{11} are the degenerate antisymmetric and symmetric deformations of the two methyl groups, respectively). The characterization of rotational structure in this complex molecule is non-trivial because the ν_7 band is severely perturbed by overtones and combination states (with a low-frequency torsional mode, ν_4 at 289 cm^{-1}) that are in Fermi or Coriolis resonance with ν_7 [537].

For the HITRAN2020 edition, we expand and advance the ν_7 band at 3.3 μm based on Refs. [536,538], add a linelist for the ν_5 band of ethane at 3.4 μm based on Radeva et al. [539], and add combination bands that include the strong $\nu_8+\nu_{11}$ band based on Lattanzi et al. [538]. These references, and a summary of how their data were adapted to HITRAN, are described below.

2.27.2. Line list from Ref. [538]

Relying on a high-resolution FTS spectrum recorded at 229 K in Brussels and line positions measured in a Doppler-limited spectrum recorded at 119 K using a tunable difference-frequency laser spectrometer [538,540] , the 2860–3060 cm^{-1} region of ethane was re-investigated.

This work led to some progress in the understanding of the complex network of interacting vibrational levels occurring in this

energy range (see Fig. 2 of Lattanzi et al. [538]). In particular, 572 line positions belonging to ^PP and ^rR transitions in the ν_7 band (maximum $J=30$), $^rQ_0$, $^PP_1$ and $^rR_1$ transitions in the $\nu_8+\nu_{11}$ band, and $^PP_6$ transitions in the $\nu_3+2\nu_4+\nu_8$ band were least-squares fit to a Hamiltonian. The model involved the ν_7 degenerate vibrational level and four degenerate perturbers, i.e., the $\nu_8+\nu_{11}$, $\nu_3+2\nu_4+\nu_8$, $\nu_4+\nu_{11}+\nu_{12}$ and $\nu_3+3\nu_4+\nu_{12}$ vibrational levels. Although RMS deviations as large as 0.018 cm^{-1} were obtained, indicating that the analysis is far from complete, a line list was generated because it still provided a much improved description of the 3.3 μm region of the ethane spectrum. Positions, relative intensities, and lower-state energies of 4969 lines associated with transitions belonging to five perpendicular bands ($\nu_8+\nu_{11}$, $\nu_4+\nu_{11}+\nu_{12}$, $\nu_3+3\nu_4+\nu_{12}$, $\nu_8+\nu_{11}$ and $\nu_3+2\nu_4+\nu_8$) were calculated between 2900 and 3071 cm^{-1}, relying on the model and parameters involved therein and resulting from the least squares analysis. The content of the line list is summarized in Table 8 of Ref. [538]. As detailed in Ref. [538], incorrectly predicted line positions were recomputed using empirical upper-state energies. These altered positions are indicated by the HITRAN error code of 4 (see Table 2), while a conservative error code of 2 was assigned to the remaining predicted positions. The predicted relative line intensities were normalized by inspection of observed and calculated spectra (HITRAN error code = 2). The Lattanzi et al. [538] line list covers the 2900–3071 cm^{-1} region.

2.27.3. Line lists from Refs. [536,539]

These models of ν_5 and ν_7 were generated by characterizing the upper ro-vibrational states using linear progressions of J and K. For the ground vibrational state, spectroscopic constants from Pine and Lafferty [537] were used, with specific corrections for some J/K ladders (see details in Villanueva et al. [536]). For the ν_5 model, as explained in Radeva et al. [539], the upper-state rotational constants were not present in the literature. Therefore they were obtained by fitting experimental data given in Ref. [541] for each K ladder. For the band intensity of the ν_5 band, parameters reported in Ref. [542] were employed.

The ν_7 upper-state ro-vibrational structure was derived by fitting to experimental data as presented in Ref. [543], in which cross-sections for ethane in the 3-μm region at temperatures between 194 and 297 K and total pressures from 0.0689 Torr to 763.48 Torr were reported. Using this dataset, we identified 466 lines, which were consolidated with 122 lines reported in Ref. [542] and 66 reported in Ref. [541], ultimately deriving rotational constants for 30 K-ladders of the ν_7 band of ethane. Our model does provide good results for the selected lines (standard deviation of 0.005 cm^{-1} for the 654 lines), but because of the numerous perturbations, their validity is relatively uncertain.

Determining accurate band intensities from experimental data in this highly active spectral region can be complex, in particular for ethane at 3.3 μm, since multiple fundamental (e.g., ν_7 and ν_5), combination (e.g., $\nu_8+\nu_{11}$), and hot-bands (e.g., $\nu_7+\nu_4-\nu_4$) overlap at these wavelengths. As reported in Ref. [543], accurate absorption cross sections for ethane at these wavelengths were determined, with an overall uncertainty of 4%. Their cross-sections were calibrated against PNNL spectra [244]. Considering these new absorption cross-sections and taking into account the first torsional hot-band, we derived a band intensity of 301 $cm^{-2}atm^{-1}$ for the ν_7 band [536].

2.27.4. Combining the line lists based on validations against laboratory data

The three line lists described above were cross-evaluated against each other, HITRAN2016 data, and the experimental cross-sections from Refs. [543,544]. To that end, HAPI [52] was used to generate cross-sections under the same thermodynamic conditions and resolution as experimental data and the synthetic cross-sections were compared with the experimental ones. It was found that data from Refs. [536,538] both agree quite well with the experimental data near the ν_7 band center, with both line lists being superior to the HITRAN2016 data except for the region around the $^PQ_7$ manifold near 2976 cm^{-1} where HITRAN was based on purely empirical data from Pine and Lafferty [537]. As rotational quanta increase, Ref. [538] produced much better agreement with the experimental data. At around 3070 cm^{-1}, the deviations of the Ref. [536] line list from experimental data becomes so significant (up to 0.5 cm^{-1}) that it was decided to not use this list in the 3071-3100 cm^{-1} interval, which is not available in Ref. [538]. In summary, the ν_7 band and combination bands were taken from Ref. [538]. However in selected spectral windows where the residuals based on the Ref. [536] data were better, the latter line list was employed. In the small spectral window around 2976 cm,$^{-1}$ HITRAN2016 data were retained (although several lines had to be reassigned to the $\nu_8+\nu_{11}$ band).

Only the Ref. [539] line list is available for the ν_5 band. Validations have shown substantially larger disagreements than those observed with either of the line lists in the ν_7 band. Some notable modifications were therefore applied to the line list from Ref. [539]. First, intensities for all the lines have been reduced by 20% to better agree with both sets of experimental cross-sections. The line positions for many lines with $K>1$ appeared to strongly deviate from their observed values. For instance, the deviations from experimental values for lines with K=2 ranged from 0.004 cm^{-1} (for J=2) up to 0.17 cm^{-1} (for J=21). We therefore applied a third order polynomial correction in J to adjust the line positions of transitions with $K=2$ and 3, but further refinements are needed in the future. Considering the rapidly growing deviations (with rotational quanta), the Ref. [539] line list was also truncated by applying an intensity cutoff of 10^{-24} cm/molecule (as opposed to 10^{-33} cm/molecule used in the original line list). After these modifications the resulting line list produced satisfactory agreement with laboratory cross-sections. However, further improvements in this region, including addition of the hot bands, would clearly be beneficial.

Fig. 29 provides an overview of the ethane spectra in the 3.3-μm spectral region, showing experimental cross-sections from Ref. [544] in the lower panel, and those generated with HAPI using HITRAN2016 and HITRAN2020.

2.27.5. $^{12}CH_3{}^{12}CH_2D$

Mono-deuterated ethane is the third most abundant isotope of ethane, with a HITRAN abundance of 9.131×10^{-4} [57]. The deuterium substitution reduces the symmetry, which results in significantly more transitions being visible in the mid-infrared. It also slightly offsets the bright series of Q-branches around 2980 cm^{-1}, which are characteristic for C_2H_6, allowing the possibility of remote observations of the D/H ratio in this spectral range. Doney et al. [545] determined line positions and relative intensities of transitions in the C-D (centered around 2170 cm^{-1}) and C-H (2850–3030 cm^{-1}) stretches, capturing the ν_4 band (2170 cm^{-1}), the $2\nu_7$ band (2770 cm^{-1}), as well as a series of bands between 2850 and 3030 cm^{-1} ($2\nu_{14}$, $2\nu_6$, $2\nu_5$, ν_1, ν_2 and ν_{12}). The assignments were made by employing *ab initio* CCSD(T)/ANO1 calculations. The study was based on spectra recorded at high resolution using a Bruker IFS-125HR spectrometer equipped with a cryogenic Herriott cell at JPL [546,547]. For the assignments in Ref. [545], spectra were recorded at 85 K, at very low pressures below 0.0022 Torr with a pathlength of 20.941 m. The model includes transitions up to $J''\le22$, $K_a''\le10$ and $K_c''\le18$, with uncertainties of the order of ~0.05 cm^{-1}. Although the model captures most of the strong transitions, further work is needed to refine the description

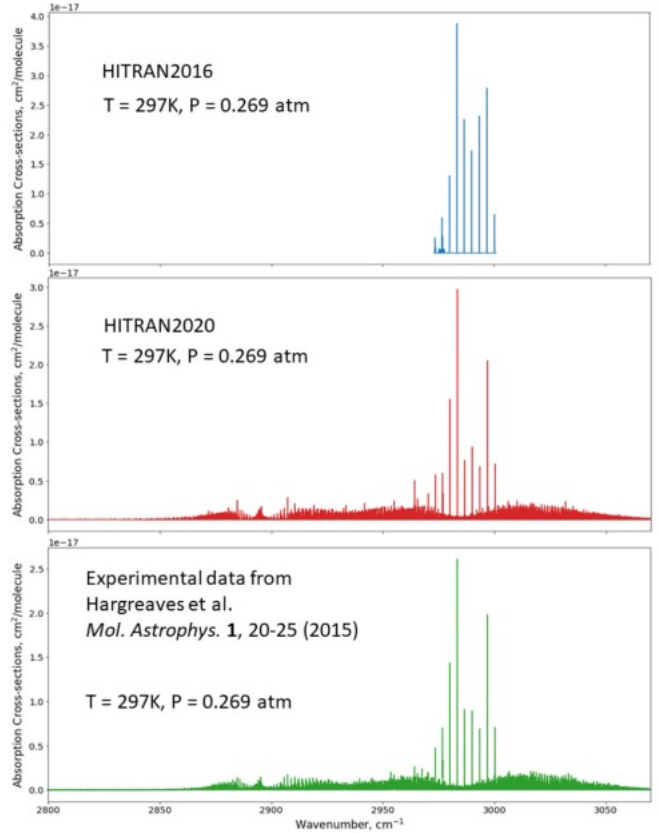


Fig. 29. The ethane spectra in the 3.3-μm spectral region, showing experimental cross-sections from Ref. [544] in the lower panel, and those generated with HAPI (under the same thermodynamic conditions) using HITRAN2016 and HITRAN2020 in the top and middle panels, respectively.

of weak transitions in the 2850–3030 cm^{-1} region. No hot bands are included in this line list.

For inclusion into HITRAN, the intensities of the C_2H_5D transitions have been calibrated against additional experimental spectra recorded using the same setup, but at higher pressures (2.023 and 0.1367 Torr), shorter path lengths (0.2038 and 0.1526 m), but at intermediate cold and room temperatures (130 and 298 K). This line list will be provided as one of the immediate updates to the official release of HITRAN2020.

2.27.6. Line-shape parameters

For all of the new bands of ethane (including the deuterated isotopologue) self- and air-broadening half-widths, and their temperature dependencees, were estimated using the expressions reported by Devi et al. [548,549] from measurements in the Q-branch of the ν_9 band near 822 cm^{-1}. The parameters involved in these expressions were applied from $K''=0$ to $K_{max}=3$ for the broadening coefficients and $K''=0$ to $K_{max}=7$ for their temperature dependence, while those provided for K_{max} were used for transitions with $K''>K_{max}$. The uncertainties for these pressure-induced coefficients are conservatively set (error code = 2, see Table 2) with the warning that the uncertainty is unknown for $J''>31$. Finally, a constant value of -0.004 $cm^{-1}atm^{-1}$ (error code = 1) was estimated for air pressure induced shifts, from the average of two air-broadening measurements at 296 K for $^rQ_0$ and $^PQ_3$ of the ν_7 band [541]. For pressure-shifts, we considered the N_2-broadened pressure-induced shifts of -0.004 $cm^{-1}atm^{-1}$ reported in Ref. [541] from $^rQ_0$ and $^PQ_3$.

For the ν_4 torsional band at 35 μm region [550], the temperature dependence exponent of the air-broadened line half-widths, n_{N_2}, listed in HITRAN2016 [16] had a truncation error which removed the integer part when the exponent is greater than 1. This issue has been fixed for HITRAN2020. In addition, the self-broadening values in that band were previously given as a con-

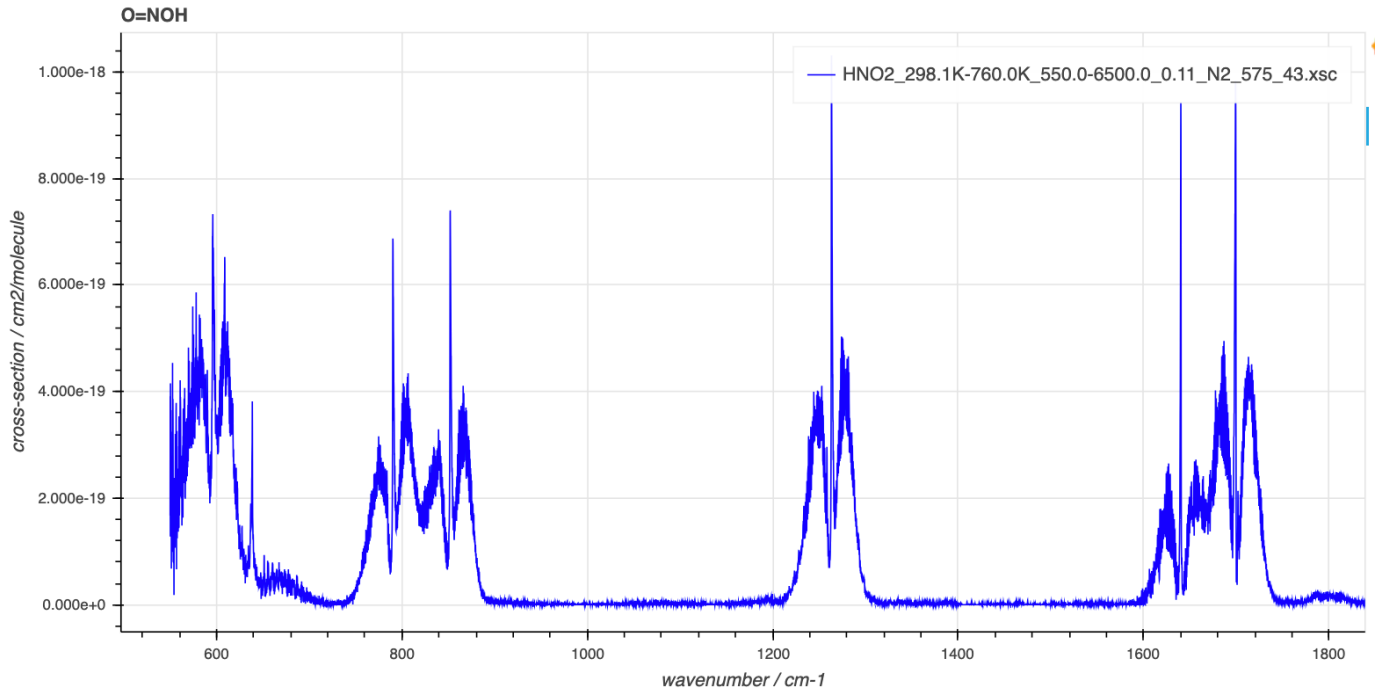
Case Study 4: HONO (Nitrous Acid)

Although GEISA has a HONO linelist covering 750 to 880 cm-1, I was asked if I could produce a HONO EPLL for the stronger band at 1220-1310 cm^{-1}. HITRAN has no discrete linelist for any HONO bands. It has just one cross-section spectrum at 298K at 1 atm of N_2. I initially replied that I couldn't make an EPLL because there is no T-dependent cross-section information from which to determine the E"s.

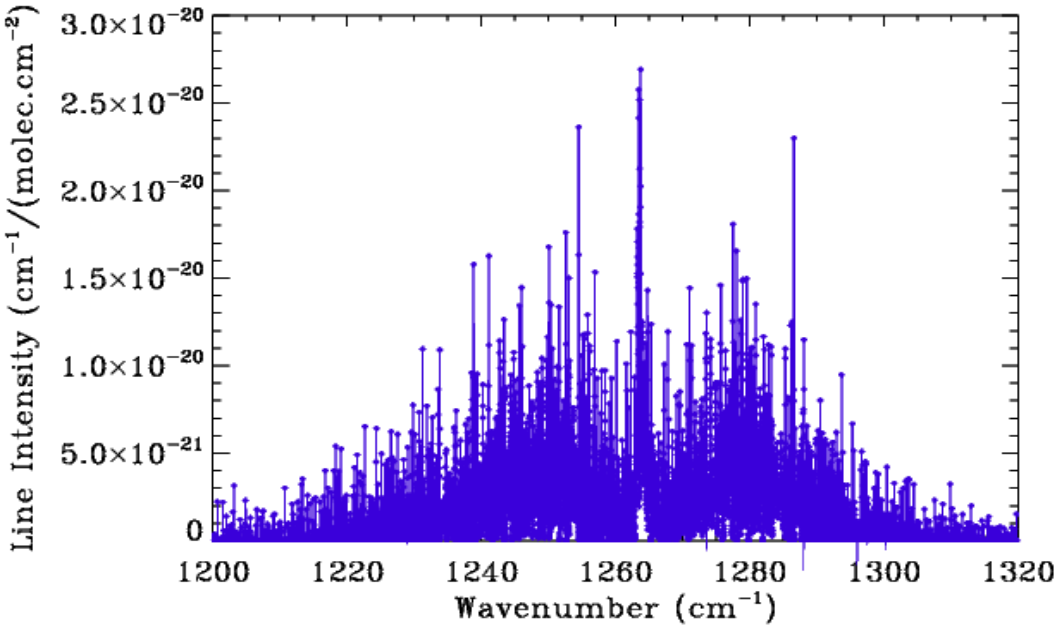
Since the band has a very simple P-Q-R-branch structure. I wondered if I could use a priori E" values based on the assumptions (1) that E" is zero at band center, (2) increases quadratically with wavenumber from band center, and (3) the integrated band intensity is independent of T since it looks like a fundamental band.

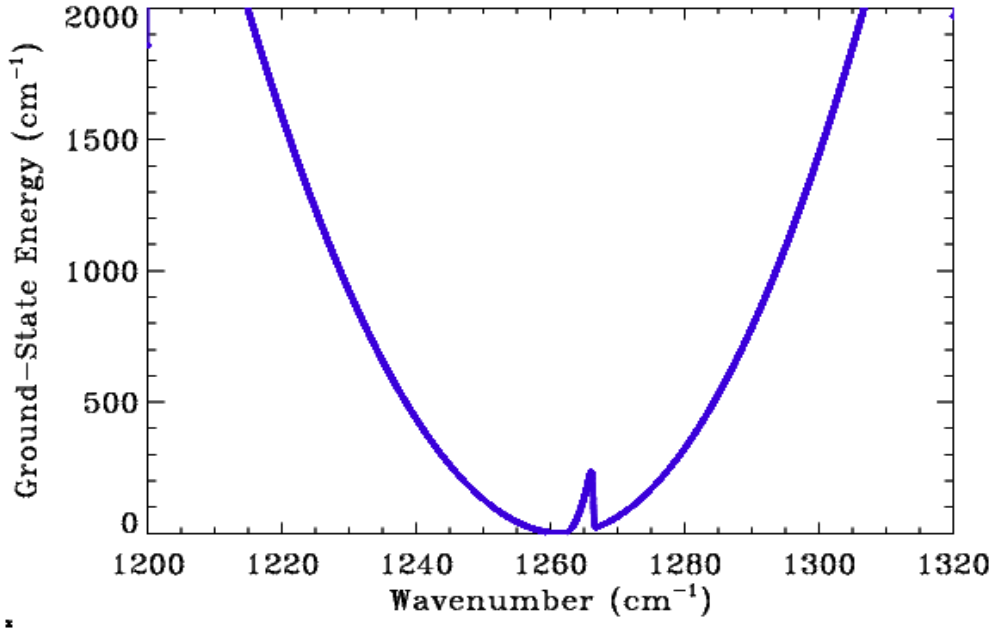
HNO2_298.1K-760.0K_550.0-6500.0_0.11_N2_575_43.xsc [974.0 KB] Cross section file (HITRAN .xsc format)
648af86c.bib.html [3.0 KB] List of sources (HTML format)
648af86c.bib [894 bytes] List of sources (BibTeX format)

ZIP archive: 648af86c.zip



HONO Empirical Pseudo-Line-List 1200-1320 cm^{-1}

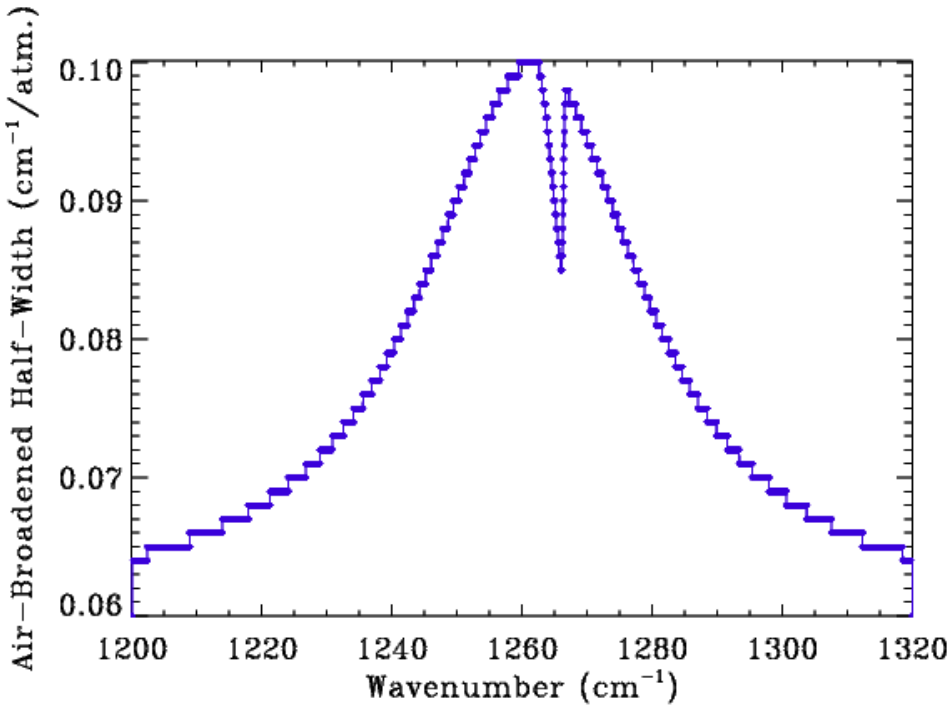




Left: Derived intensities of individual pseudo-lines

Below: Assumed E" and ABHW

The parameterization of E" and width will hopefully extend the range of conditions (temperatures and pressures) over which this linelist produces good results. But there is no way of testing this, at present, in the absence of Low-T lab measurements.

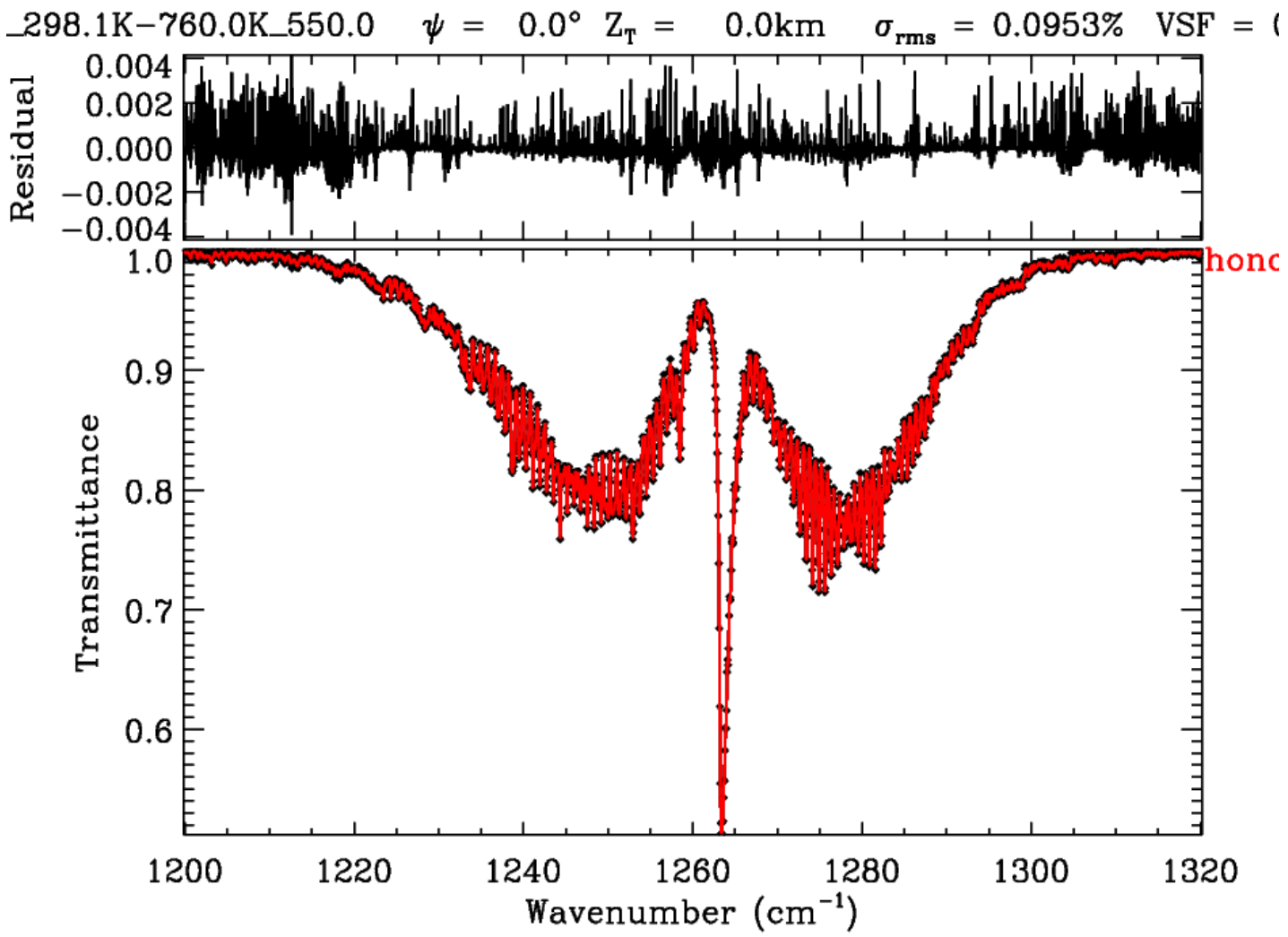


Spectral Fit to 298K PNNL spectrum from which the EPLL was derived

As a self-consistency check, the transmittance spectrum that was computed from the PNNL cross-sections, was refitted using the final EPPL

The fit is not perfect, despite there being only 1 spectrum, because the lines are too broad at 760 mbar to be able to fit the noise in the measurements. A perfect fit is possible only by allowing the intensities to go negative, which was prevented. This is particularly noticeable in the wings where the lines are weaker.

Still waiting to hear whether anyone was able to detect atmospheric HONO using this new EPLL.



mark4sun.jpl.nasa.gov/pseudo.html

YouTube Maps Weather 91011 JPL: MkIV FTIR Int... vpl.astro.washingt...

Pseudo Linelists

Pseudo-Linelists are available on-line along with reports and links to any papers

Readme

Isotopomers	Fortran code	Data	
n-butane / C4H10 (660-1538 cm^{-1})	n-butane_PLL.pdf	n-butane_660_1538.101.gz	
trans-2-butene / C4H8 (800-1120 + 1120-1534 cm^{-1})	trans-2-butene_PLL.pdf	t2C4H8_pll8x.mw1_0800x1120.101.z3.par.gz	t2C4H8_pllx7.mw2_1120x1534.101.z3.par.gz
HCFC-22 / CHClF2 (560-1375 + 2980-3070 cm^{-1})	HCFC_22_Pseudo_Line_List.pdf	chclf2_560_3070.101.gz	
Isocyanic Acid / HNCO (501-1020 + 2097-2338 cm^{-1})	HNCO_EPLL.pdf	stripped_g80_ia_pll.101.gz	
C_5H_8 / Isoprene (800-1870 + 2700-3200 cm^{-1})	Readme.C5H8	C5H8_isoprene.101.gz	
CFH_2CF_3 / HFC-134a (504-1550 + 2790-3050 cm^{-1})	Readme.CFH2CF3	CFH2CF3.101.gz	
C_3H_6 / Propene (800-1121 cm^{-1})	Readme.C3H6	C3H6_800_1121.101.gz	
C_3H_6 / Propene (1320-1524 cm^{-1})	Readme.C3H6	C3H6_1320_1524.101.gz	
CH_3COCH_3 / Acetone (700-1910 + 2615-3250 cm^{-1})	Readme.CH3COCH3	CH3COCH3.101.gz	
C_6H_6 / Benzene (630-735 cm^{-1}) + H_2/He	Readme_C6H6	C6H6+H2He_pll.630-735.102	Sung_et_al.C6H6.PLL.2016.pdf
C_6H_6 / Benzene (630-1535 cm^{-1}) + N_2	Readme.C6H6	C6H6+N2_pll.0630-1535.102	Sung_et_al.C6H6.PLL.2016.pdf
C_3H_8 / Propane (2560-3280 cm^{-1})	Readme.c3h8	c3h8_pll_2560_3280.101	
C_3H_8 / Propane (670-1550 cm^{-1})	Readme.c3h8	c3h8_pll_670_1550.101	
CH_3OH / Methanol	Readme.ch3oh	ch3oh_pll.101	
CHF_3 / HFC-23 (2011)	Readme.chf3	chf3_2011.101	
CHF_3 / HFC-23 (2020)	CHF3_PLL_Update.pdf	chf3_2020.101	
$C_2H_3Cl_2F$ / HCFC-141b	Readme.f141b	f141b.101	
CH_3CHO / Acetaldehyde	Readme.ch3cho	Linelist	
$C_2H_3NO_5$ / PAN	Readme.pan	Linelist	
CH_3CN	Readme.ch3cn	Linelist	
CH_3CHO / Acetaldehyde	Readme.ch3cho	Linelist	
$C_2H_3NO_5$ / PAN	Readme.pan	Linelist	
CH_3CN	Readme.ch3cn	Linelist	

mark4sun.jpl.nasa.gov/pseudo.html

YouTube Maps Weather 91011 JPL: MkIV FTIR Int... vpl.astro.washingt...

CH_3CN	Readme.ch3cn	Linelist	
HNO_3 (all PLL regions)	Readme.hno3	hno3_pll_all.101	HNO3_Spectroscopy_Evalua
C_2H_6 (1350-1496 cm^{-1})	Readme.c2h6_1350_1496	c2h6_1350_1496.101	
C_2H_6 (2720_3100 cm^{-1})	Readme.c2h6_2720_3100	c2h6_2720_3100.101	C2H6_spectroscopy_evaluati
N_2O_5		Linelist	
$ClNO_3$	Readme.clno3	Linelist	
NF_3	Readme.nf3	Linelist	
CF_4		Linelist	
CCl_2F_2 / CFC-12		Linelist	
CCl_3F / CFC-11		Linelist	
CCl_4		Linelist	
$CHClF_2$ / HCFC-22		Linelist	
$COCl_2$ / Phosgene		Linelist	
SF_6		Linelist	
$C_2Cl_3F_3$ / CFC-113		Linelist	
$C_2H_3ClF_2$ / HCFC-142b	Readme.f142b	Linelist	
CH_3COOH / Acetic Acid		Linelist	
Foreign collision-induced absorption	Readme.cia	O2_CIA_7885.pdf	FCIA-Linelist
Self collision-induced absorption		SCIA-Linelist	

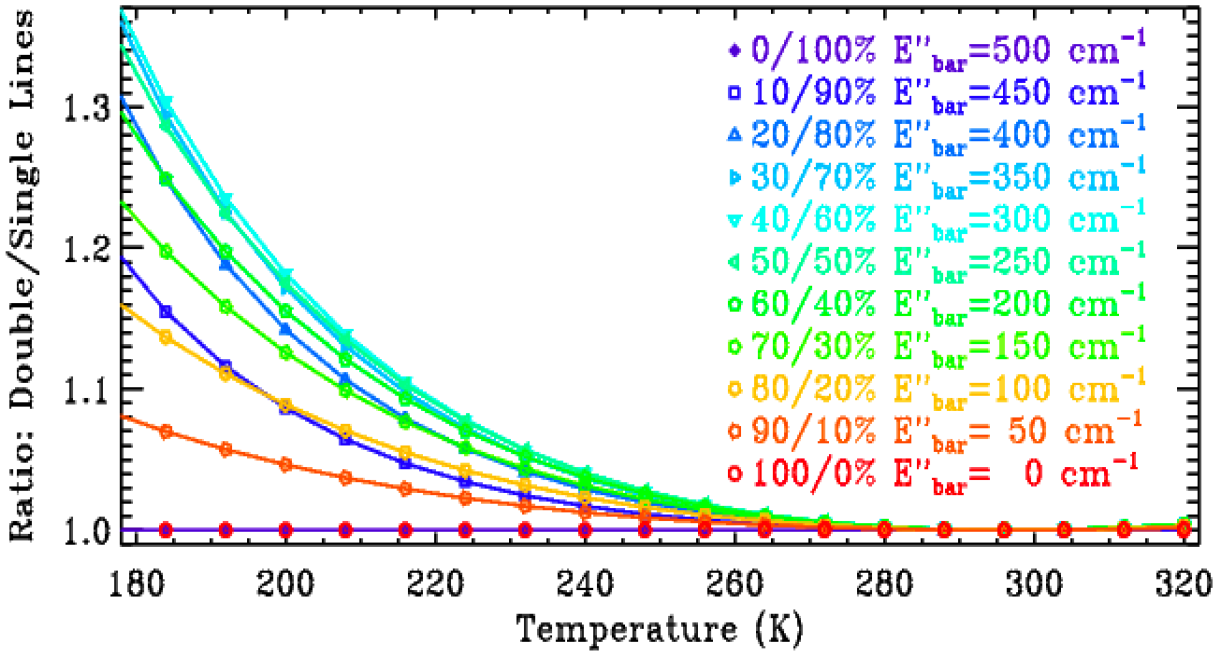
Summary and Conclusions

Fitting a EPLL to the lab spectra provided a means of:

- Aggregating multiple cross-section measurements made under different condition (spectral coverage and resolution, T/P) to yield a single EPLL.
- Test the self-consistency of the ingested absorption measurements, leading to identification of biases that can be corrected. or identification of outlier spectra which can then be rejected from the EPLL process or.
- Opportunity to remove artifacts from the lab spectra (ILS, zero offsets, channeling, contaminant absorptions)
- Physics-constrained extrapolation in P and T to conditions that were not sampled/encompassed by the original lab measurements.
- Smoother interpolation in P and T than is possible by mathematical interpolation of the lab measurements. This minimizes artifacts in the retrieved atmospheric vmr profiles arising from roughness in the cross-sections as a function of P and T (due to P and T being functions of altitude).
- If all the pseudo-lines in a given band are assumed to have the same PBHW and Doppler widths, only one evaluation of the Voigt lineshape per atmospheric level is necessary to compute the absorption spectrum resulting from all the pseudo-lines (provided that this lineshape is stored). Thus, reducing the speed of using the pseudo linelists.
- Finally, EPLLs are preferred by atmospheric scientists to cross-section spectra for convenience – all the machinery to handle HITRAN-format linelists already exists in their analysis codes. They don’t need to implement new computational machinery to perform bi-linear interpolation (T,P) into the cross-sections.

Deficiencies with EPLL Approach

If there are two overlapping lines at a particular wavenumber, with similar depths but a large difference in E", their T-dependence cannot be accurately represented by a single pseudoline of intensity $I=I_1+I_2$ and $E"=(I_1E"_1+I_2E"_2)/(I_1+I_2)$. This results in the relationship between $hc(1/k296-1/kt_j)$ and $\ln\{VSF_j \ln[T^m_{i,j}]/\ln[T^c_{i,j}]\}$ becoming curved rather than linear.



The figure (right) shows the fractional error made by representing two overlapping lines having E"=0 and E"=500 by a single line with the sum of strengths and an intensity-weighted E". Note that the first (purple) and last (red) lines fall on top of each other with a ratio of 1.0 at all temperatures. Although a 30% error (40%/60% at 180K) is a lot, it probably doesn't happen that often, and so averaged over the whole absorption band or fitting window the effect will be smaller. Typically the T-dependence of the EPLL matches the lab measurements to ±3% and often better. For the vast majority of gases for which I have made EPPLs, an atmospheric measurements to within 5% would be a break-through. I also point out that many atmospheric polluteant originate near the Earths surface at 250K or more, so the T-dependent error in this case would be <4%. But for planetary applications (e.g. Titan) much larger erros are possible, especially if the lab measurements were made under warmer conditions

More Potential Deficiencies with EPLL Approach

Another possible deficiency relates to the curve of growth. If the pseudo-lines are too widely spaced and the absorption is strong, in the lab or atmospheric spectra, then at low pressures the pseudo-lines may become saturated and the absorption spectrum resembles a comb. So this is only an issue for gases with significant absorption in the stratosphere (e.g. N2O5, ClONO2) and can be mitigated by use of a sub-doppler pseudo-line spacing (very large linelists), or by artificially inflating the pseudo-line doppler widths.

This choice must be made at the beginning before the widths are chosen and before I and E" values are derived snce inflating the doppler widths will change the values.

Most continuum absorbing gases are in the linear part of their curve of growth in the atmosphere. But in the low-pressure laboratory spectra the lines are frequently growing non-linearly, especially the stronger absorption features favored for atmospheric remote sensing of trace gases. Since the lab spectra rarely fully resolve individual lines in low-pressure spectra, this creates ambiguity. You can't tell from a single spectrum whether a particular under-resolved absorption feature is a single saturated line, or several adjacent weaker lines. In the former case the absorption feature will grow non-linearly with absorber, whereas in the latter case it will be grow linearly. You can only tell by looking at multiple spectra.

Theoretical Basis (1/2)

By assuming that the absorption at the center frequency of a particular line comes only from that line (i.e., negligible contribution from the overlapping neighboring lines), we can solve for S and E" at each frequency separately.

So at each frequency we would have only 2 unknowns and N=20+ pieces of information; one from each lab spectrum. So when adjusting the intensity and ground-state energy of line i (S_i and $E"_i$), the influence of errors in S_{i+1} and $E"_{i+1}$ is ignored. So our methodology is optimal only when the lines overlap minimally. This is what we want anyway – it would be wasteful to do otherwise.

The transmittances of the measured and calculated lab spectrum are given by

$$T^m_{i,j} = \mathrm{Exp}(-S^m_i X^m_j L^m_j) \quad (1)$$

$$T^c_{i,j} = \mathrm{Exp}(-S^c_i X^c_j L^c_j VSF_j) \quad (2)$$

i is an index over spectral lines/wavenumber

j is an index over spectra

S^m_i and S^c_i are the unknown true and the currently-assumed strength of line i ($cm^{-1}/(molec.cm^{-2})$)

L^m_j and L^c_j are the true and assumed path length for the j'th spectrum (cm)

X^m_j and X^c_j are the true and assumed absorber number density in the j'th spectrum ($molec.cm^{-2}$)

VSF_j is the factor by which X_j was scaled in fitting the j'th spectrum (unitless)

Assume that the path lengths and number densities are known perfectly ($L^m_j = L^c_j$; $X^m_j = X^c_j$)

Taking the natural logarithms:

$$\ln[T^m_{i,j}] / \ln[T^c_{i,j}] = S^m_i/S^c_i/VSF_j \quad (3)$$

Taking another logarithm

$$\ln\{ \ln[T^m_{i,j}] / \ln[T^c_{i,j}] \} = \ln[S^m_i/S^c_i] - \ln[VSF_j] \quad (4)$$

Theoretical Basis (2/2)

The line strengths at a temperature t_j are the product of several factors:

$$S^m_i(t_j) = S^m_i(296).IRPF^m(t^m_j).IVPF^m(t^m_j).SE^m(t^m_j).Exp[hcE"^m_i(1/k296-1/kt^m_j)]$$

$$S^c_i(t_j) = S^c_i(296). IRPF^c(t^c_j) . IVPF^c(t^c_j) . SE^c(t^c_j) . Exp[hcE"^c_i(1/k296-1/kt^c_j)]$$

$S^m_i(296)$ and $S^c_i(296)$ are the unknown true and currently-assumed, 296K, line intensities

$E"^m_i$ and $E"^c_i$ are the unknown true and currently-assumed ground-state energies

$IRPF^m$ and $IRPF^c$ are the true and assumed Inverse Rotational Partition Functions for spectrum j

$IVPF^m$ and $IVPF^c$ are the true and assumed Inverse Vibrational Partition Function for spectrum j

t^m_j and t^c_j are the true and assumed temperatures of the j'th spectrum

SE^m and SE^c are the true and assumed Stimulated Emission terms at a temperature t_j

If the assumed temperatures (t^c_j) are correct **and** if our assumed inverse partition functions are also correct, then they cancel

$$S^m_i(t_j)/S^c_i(t_j) = S^m_i(296)/S^c_i(296). Exp[(E"^m_i - E"^c_i)(q/296-q/t_j)] \quad (5)$$

Taking the natural logarithm:

$$\ln[S^m_i(t_j)/S^c_i(t_j)] = \ln[S^m_i(296)/S^c_i(296)] + hc(E"^m_i - E"^c_i)(1/k296-1/kt_j) \quad (6)$$

Substituting (6) into equation (4) yields

$$\ln\{VSF_j \ln[T^m_{i,j}]/\ln[T^c_{i,j}]\} = \ln[S^m_i(296)/S^c_i(296)] + hc(E"^m_i - E"^c_i)(1/k296-1/kt_j) \quad (7)$$

So plotting $\ln\{ VSF_j \ln[T^m_{i,j}]/\ln[T^c_{i,j}] \}$ versus $hc(1/k296-1/kt_j)$ and fitting a straight line for each pseudo-line should yield:

$$gradient = E"^m_i - E"^c_i$$

$$offset = \ln[S^m_i(296)/S^c_i(296)]$$

Thus, the currently-assumed values of $S^c_i(296)$ and $E"^c_i$ can be moved toward the unknown correct values ($S^m_i(296)$, $E"^m_i$)

$$E"^c_i = E"^c_i + gradient \quad (8)$$

$$S^c_i(296) = S^c_i(296).Exp[offset] \quad (9)$$

Pet Peeves about Lab Cross-Section Measurements

Too much gas of interest. Sometimes I see spectra in which the strongest features of the gas of interest are 95% deep, while in the atmosphere they are <10% deep. This makes the stronger absorption coefficients sensitive to zero level offsets in the original spectra and can put them into a non-linear part of the curve of growth. So please try to replicate the atmospheric conditions, including gas column, for the cell measurements. Remember, it is the strongest absorbing features that atmospheric scientist focus on.

In some spectra, the scientist has tried to eradicate contaminant absorption by ratioing against an empty cell spectrum. Often this results in a partially cancelled contaminant line which is impossible to model because P or T has changed.

I have also seen spectra in which some of the contaminant lines have been removed by interpolating the target gas cross-sections across the base of the contaminating lines. But this has to be done for all contaminating lines, not just the strongest.

Lab cross-section measurements often omit low temperatures at 1 atm pressure. Over Canada or Russia in winter, daytime temps can be -30C, Also over Antarctica. So, for ground-based measurements from these sites, cross-sections have to be extrapolated.