CO_2 Spectroscopy Evaluation: 670 to 7000 cm^{-1}

Geoff Toon
Jet Propulsion Laboratory
2019-03-17

Several CO_2 linelists including HITRAN 2008, 2012, and 2016 (two versions), have been evaluated by fitting laboratory spectra (mainly Kitt Peak) and atmospheric solar absorption spectra (MkIV & TCCON).

The 670-7000 cm^{-1} region of interest was divided into 41 windows, most encompassing at least one complete CO_2 absorption band or sub-branch. Regions with no discernable CO_2 absorption were skipped.

The GFIT spectral fitting algorithm was used in all cases assuming a Voigt lineshape and no line-mixing. This evaluation focusses on the RMS fitting residuals that were achieved and the window-to-window consistency of the retrieved CO_2 amounts.

Between evaluations of the different linelists, only the CO_2 linelist was changed. The spectroscopy of the interfering gases (e.g. H_2O, O_3, CH_4, etc.) was unchanged, so any difference in the RMS fitting residuals or the retrieved CO_2 amounts is entirely attributable to the CO_2 linelist under evaluation.

A new "greatest hits" linelist (ATM18) was subsequently developed by selecting from the best predecessor linelists. Ad hoc manual adjustments were then performed to fix obvious errors (e.g. bad line positions, pressure shifts, inconsistent retrieved CO_2 amounts). To keep this report concise, the new ATM18 linelist is presented in parallel with the evaluation of earlier linelists, even though it was developed much later.

Motivation

The main motivation for improving CO_2 spectroscopy is to make more accurate measurements of atmospheric CO_2, the second most important GHG (after H_2O) and the main driver of climate change [CO_2 has increased 60% since pre-industrial times whereas H_2O has increased by only 5-10%].

Being a simple linear molecule, the spectroscopy of CO_2 is already very good, in comparison with H_2O, O_3, CH_4 or even O_2. So the improvements embodied in HITRAN 2016 appear pretty modest in terms of the fitting residuals or the retrieved atmospheric CO_2 amounts. But due to the 20- to 200-year lifetime of atmospheric CO_2 (depending upon how you define lifetime) a large atmospheric concentration (400 ppm) has accumulated. This means that atmospheric CO_2 has a large DC component, upon which the AC variations of interest are superimposed. So a very high measurement accuracy (0.2%) is required to see ~1 ppm spatio-temporal variations of atmospheric CO_2 caused by source/sink imbalances. This does NOT mean that all spectroscopic line parameters of all CO_2 lines need to be measured to 0.2% accuracy. But certainly further improvements beyond HIT2016 are needed.

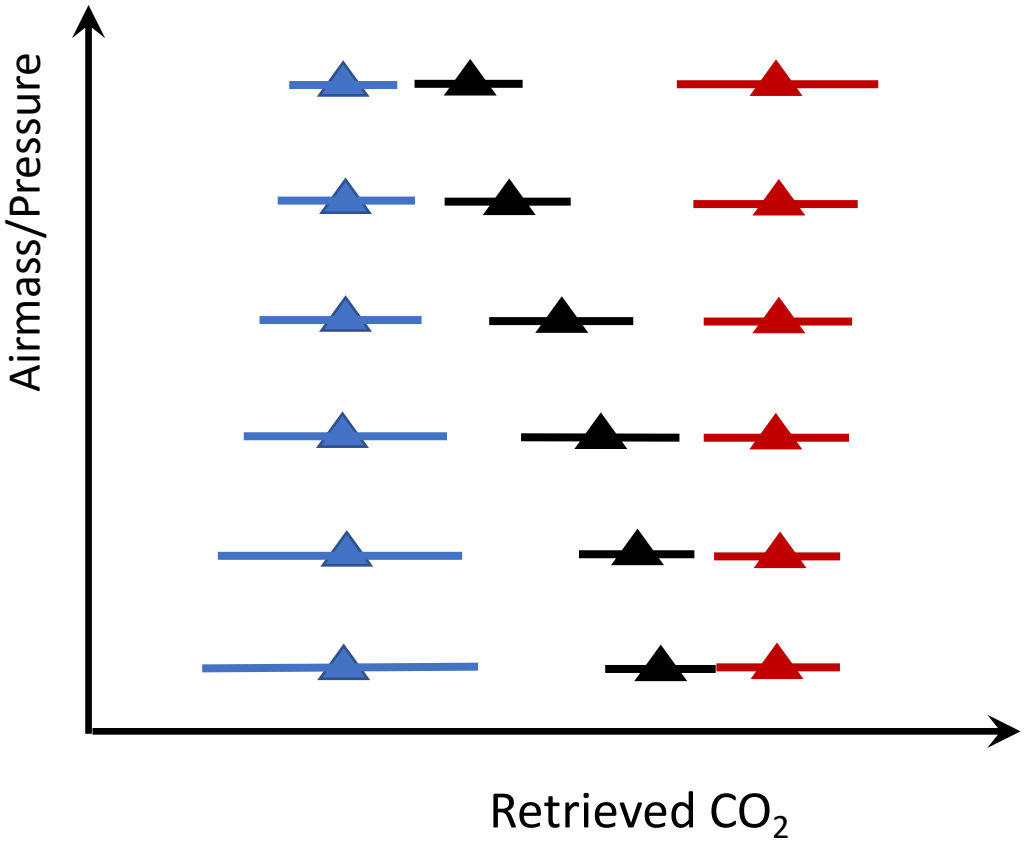
CO_2 has a highly predictable atmospheric abundance, founded on many highly accurate in situ measurements (mass spec. and NDIR) made on the surface, on aircraft, and balloons from a wide variety of locations and seasons. CO_2 can be predicted to better than 1 ppm (0.25%) anywhere in the Earth's atmosphere between the PBL and the mesosphere. This fact provides for the use of CO_2 for remote sensing of temperature, and as a proxy for the number of air molecules encountered along the light path. For example, CO_2 is used to determine the tangent altitude in limb viewing experiments (e.g. solar occultation). Also CO_2 measurements have been used to determine the light path in nadir measurements of reflected sunlight. In the Earth's atmosphere CO_2 also has a highly predictable isotopic composition, again founded on in situ measurements.

Avoiding Biases

Atmospheric CO_2 measurements are made over a wide range of conditions. For example, in limb viewing from balloon, the tangent pressures will vary from 3 mbar to 300 mbar, representing a 2 order of magnitude change in absorber amount.

For ground-based solar absorption measurements from mid-latitudes, the atmospheric airmass varies by more than an order of magnitude during a single day, from 1.5 at noon to 15 at 87º SZA, with the majority of the data acquired below 2 airmasses. For high-latitude sites the range of airmasses is smaller (3-15), but biased towards larger values, especially in the winter. Any airmass-dependent bias in the retrieved CO_2 will be mis-interpreted as a daily variation in CO_2 (at mid-latitudes) or a seasonal variation (at high latitudes), or a latitude gradient when comparing mid- and high-latitude sites.

At low airmasses or high altitudes the information comes mainly from the stronger (but still unsaturated) lines. At higher airmasses, or low altitudes in the case of limb viewing, the stronger lines saturate, so the information comes increasingly from the weaker lines. To avoid altitude- or airmass-dependent artifacts, it is important that the weak and strong CO_2 lines/bands give consistent results.



Combining retrievals from weak (blue) and strong (red) absorption lines results in a spurious airmass/altitude variation when bias is present, even though the individual windows are airmass-independent. This is because the uncertainties of the weak and strong lines have different behaviors, the former improving with increasing airmass, the latter worsening.

The Linelists Evaluated

HITRAN 2008: 314,919 CO_2 lines

HITRAN 2012: 471,847 CO_2 lines

ATM 2016: 450,493 CO_2 lines
Based mainly HITRAN 2012. Uses Toth (2009) for the 5740-6500 cm^{-1} region because it gave better fits (and still does). Empirical adjustments have been made throughout to fix obvious errors (mainly line position errors).

HITRAN 2016a: 554,183 CO_2 lines
Based on files the linelist that Iouli Gordon sent me June 27, 2017 (O2_hit16_first-9iso) and on June 30, 2017 (hit838corr). Before using, I fixed 19 lines with ABHW=0 and one with an intensity of zero (used HIT 2012 value).

HITRAN 2016b: 554,879 CO_2 lines
Downloaded from HITRAN-Online website on Nov 28, 2017 (5a1de32a.par). Includes isotopologs 11 & 12. A format of "f5.3" had been enforced for SBHW, which changes some lines that were previously "f5.4", e.g.

25 3.681760 1.943E-33 1.457e-12.0865.1155 in HIT16a

became

25 3.681760 1.943E-33 1.457e-12.08650.116 in HIT16b

ATM 2018: 524,724 CO_2 lines (new linelist)
Mostly HIT 2016b, except in regions where ATM 2016 or HIT 2008 were better. Some ad hoc empirical adjustments.

Absorber Amount Limitations

In the Earth's atmosphere, the total vertical column of all gases above sea-level is 2.15E+25 molecules.cm^{-2}. So CO_2, having a vmr of 400 ppm, will have a total column of 400E-06 x 2.15E+25 = 8.6E+19 molecules.cm^{-2}.

In a one-sided horizontal path, such as be obtained from an observer on the surface at sunset/rise, the airmass is 35 times that in a vertical path. So the CO_2 slant column in this case would be 3E+21 molecules.cm^{-2}.

Double-sided horizontal paths with an airmass of 70 are achievable from space or balloon (i.e., solar occultation) but not generally below 5 km altitude because cloud, aerosol, or mountains. Since the atmospheric number density at 5 km is roughly half that at the surface, in this geometry the CO_2 slant column is still 3E+21 molecules.cm^{-2}.

A CO_2 line of intensity S, will have an equivalent width of S x 3 x 10^{+21} cm^{-1}. At 5 km such a line will have a width of 0.03 cm^{-1}, so its depth will be S x 10^{+23}. At altitudes higher than 5 km the width will decrease, but so will the CO_2 slant column, so the depth will stay the same. The lines will simply narrow with increasing altitude, until their width drops below the spectrometer resolution. So a CO_2 line of strength 10^{-23} will have unit optical depth. If we can see lines down to 1% depth, this sets an intensity limit of 10^{-25} cm^{-1}/(molec.cm^{-2}) for evaluation of CO_2 lines in Earth atmospheric spectra.

In laboratory spectra the vmr can be increased to 1, gaining a factor of 2500. But the longest path lengths I have seen are only 384m (ignoring cavity ring-down). So lab spectra can far surpass the CO_2 amounts seen in the Earth's atmosphere. The largest slant column in the 148 lab spectra analyzed here was 10^{+23} molecules.cm^{-2} (160 Torr of CO_2 in a 192 m path). This spectrum covered only the 6700-9000 cm^{-1} region where CO_2 bands are very weak. These lines will have a width of 0.01 cm^{-1} and hence an equivalent width of S x 10^{+25} cm^{-1}. Assuming that we can see lines down to 1% depth, then this sets an intensity limit of 10^{-27} cm^{-1}/(molec.cm^{-2}) for CO_2 lines that can be evaluated.

HIT16 contains CO_2 lines as weak as 10^{-30} cm^{-1}/(molec.cm^{-2}), even weaker for heavy isotopologs. This study cannot evaluate lines weaker than 10^{-27} cm^{-1}/(molec.cm^{-2}) in lab spectra or 10^{-25} cm^{-1}/(molec.cm^{-2}) in Earth atmospheric spectra, although we recognize that for Mars or Venus such lines may be important.

The Laboratory Spectra of CO_2

Kitt Peak CO_2 lab spectra are available covering 600 to 12,000 cm^{-1}, although here we investigate 670 to 7000 cm^{-1}.

There are 148 spectra: 136 from Kitt Peak and 12 from JPL (Keeyoon Sung). Pressures range from 0.1 to 700 Torr.

All at room temperature (291-303 K) except for two Kitt Peak spectra:

- One at 268K and 14.2 Torr in a 30 cm cell covering 600 -1400 cm^{-1}
- One at 235K and 12.8 Torr in a 30 cm cell covering 600 -1400 cm^{-1}

23 Kitt Peak spectra are enriched in ^{13}C, giving the lab spectra a much higher sensitivity to spectroscopic errors in isotopologs 2, 5, 6, 10, 11, 12, than the other lab spectra (or atmospheric spectra).

Of the 6068 potential spectral fits (41 windows x 148 spectra), only 1816 (29.9%) could actually be performed for each linelist due to the limited spectral coverage of the individual spectra, most of which have < 1000 cm^{-1} of useful coverage.

This makes it difficult to compare intensities measured at low wavenumbers with those from high wavenumbers because these are seldom in the same spectrum. And on the rare occasions when they are, the SNR is poor.

Fitted Windows

Table summarizes the properties of the 41 windows in which lab spectra were fitted. Includes everywhere that there are discernable CO_2 lines. Collectively, these windows cover more than 3000 cm^{-1}.

A particular window covers the range: Center-Wid/2 to Center+Wid/2

NBF is the order of the polynomial fitted to the continuum level.

Iso_{bar} is the strength-weighted isotope number. Values of 1.0 mean that lines in the window are predominantly $^{12}C^{16}O_2$. Values of 2.0 likely indicate $^{13}CO_2$ or $OC^{17}O$. Values of 3.0 indicate ^{18}OCO.

S_{max} is the largest CO_2 line strength.
S_{tot} is the sum of the CO_2 strengths
S_{bar} is the mean CO_2 line strength
$ABHW_{bar}$: mean (S-weighted) ABHW
E''_{bar} : mean (strength-weighted) E"

Note that the windows in which S_{max} is largest have Iso_{bar} close to 1.0.

The tabulated properties depend only on the window, not the measured spectra.

#	Center	Wid	MI	A	NBF	Gases fitted	Iso_{bar}	S_{max}	S_{tot}	S_{bar}	$ABHW_{bar}$	E''_{bar}
1	693.2	48.7	20	2	10	co2 h2o	1.01	1.71E-19	2.93E-18	1.05E-19	0.0740	335.3
2	754.5	65.5	20	1	10	co2 h2o	1.01	2.96E-21	4.76E-20	1.83E-21	0.0742	998.8
3	827.2	68.6	20	1	15	co2 h2o	1.01	1.57E-23	2.55E-22	9.94E-24	0.0743	1597.6
4	925.0	72.1	20	1	8	co2 h2o	1.03	2.15E-23	3.35E-22	1.34E-23	0.0745	1686.9
5	980.0	42.1	20	1	5	co2 h2o nh3	1.01	2.24E-23	3.09E-22	1.60E-23	0.0744	1586.3
6	1056.4	93.8	20	1	9	co2 h2o	1.02	3.46E-23	9.95E-22	2.23E-23	0.0745	1553.0
7	1239.5	40.6	20	1	5	co2 h2o	3.05	5.65E-25	1.80E-23	3.59E-25	0.0745	274.2
8	1280.6	39.8	20	1	5	co2 h2o ch4	3.07	6.00E-25	1.81E-23	4.08E-25	0.0741	216.9
9	1367.4	51.8	20	1	5	co2 h2o ch4	3.07	6.56E-25	3.54E-23	4.50E-25	0.0753	196.2
10	1857.4	16.5	20	1	4	co2 h2o	1.13	5.76E-25	6.33E-24	4.28E-25	0.0691	1101.4
11	1906.5	71.0	20	1	5	co2 h2o	1.02	2.79E-23	6.32E-22	1.54E-23	0.0726	359.6
12	1958.0	28.6	20	1	5	co2 h2o	1.01	3.04E-24	3.54E-23	2.20E-24	0.0685	595.2
13	1982.5	17.0	20	1	4	co2 h2o	1.13	9.52E-25	5.43E-24	4.17E-25	0.0673	1196.4
14	2082.0	178.0	20	1	5	co2 h2o co n2o	1.02	2.02E-22	6.18E-21	1.06E-22	0.0737	378.0
15	2299.0	202.0	20	1	5	co2 h2o n2o co ch4	1.02	3.54E-18	1.01E-16	2.29E-18	0.0749	261.1
16	2432.0	67.0	20	1	7	co2 n2o ch4	1.03	2.50E-24	6.40E-23	1.80E-24	0.0754	1469.9
17	2502.0	70.0	20	1	4	co2 n2o hdo ch4	2.95	2.55E-25	1.51E-23	1.65E-25	0.0744	304.7
18	2601.0	96.0	20	1	5	co2 n2o hdo ch4	3.05	4.21E-25	2.49E-23	2.70E-25	0.0745	250.8
19	2760.0	62.0	20	1	4	co2 h2o hdo n2o ch4	3.05	6.34E-26	3.58E-24	4.22E-26	0.0748	213.7
20	3155.0	42.0	20	1	5	co2 h2o n2o ch4	1.03	3.09E-25	5.11E-24	1.89E-25	0.0701	530.9
21	3207.0	50.0	20	1	5	co2 h2o n2o	1.00	5.06E-25	7.59E-24	3.53E-25	0.0704	419.7
22	3309.0	49.0	20	1	5	co2 h2o n2o	1.05	1.71E-24	2.79E-23	1.02E-24	0.0696	580.3
23	3364.0	52.0	20	1	5	co2 h2o n2o	1.00	4.95E-24	8.64E-23	3.02E-24	0.0713	422.2
24	3496.3	62.6	20	1	5	co2 h2o n2o	1.95	2.09E-22	4.16E-21	1.05E-22	0.0741	574.8
25	3548.8	42.5	20	1	5	co2 h2o n2o	1.21	2.20E-21	4.36E-20	8.40E-22	0.0718	925.2
26	3618.5	97.2	20	1	5	co2 h2o	1.03	3.83E-20	1.08E-18	2.52E-20	0.0750	242.1
27	3712.6	93.2	20	1	5	co2 h2o	1.01	5.85E-20	1.64E-18	3.83E-20	0.0751	259.1
28	3811.0	74.0	20	1	5	co2 h2o	1.03	2.30E-24	5.98E-23	1.59E-24	0.0753	1480.4
29	3872.0	32.0	20	1	5	co2 h2o	1.88	3.70E-26	1.86E-24	2.26E-26	0.0750	1310.5
30	3992.0	82.0	20	1	5	co2 h2o	1.22	2.39E-25	6.21E-24	1.14E-25	0.0712	1017.9
31	4622.0	82.0	20	1	5	co2 n2o	2.81	2.75E-25	1.86E-23	1.62E-25	0.0744	290.4
32	4705.0	84.0	20	1	5	co2 h2o n2o	1.90	9.95E-25	2.83E-23	3.85E-25	0.0738	833.0
33	4825.0	157.0	20	1	5	co2 h2o n2o nh3	1.06	2.73E-22	8.27E-21	1.67E-22	0.0750	281.5
34	4962.0	116.0	20	1	6	co2 h2o n2o nh3	1.01	1.32E-21	3.71E-20	8.56E-22	0.0751	260.4
35	5096.0	152.0	20	1	6	co2 h2o n2o nh3	1.01	4.29E-22	1.22E-20	2.75E-22	0.0752	271.2
36	6072.0	98.0	20	1	5	co2	1.07	1.73E-24	5.05E-23	1.10E-24	0.0728	256.9
37	6211.0	128.0	20	1	6	co2	1.02	1.74E-23	4.97E-22	1.12E-23	0.0744	266.4
38	6338.0	124.0	20	1	6	co2	1.00	1.74E-23	4.84E-22	1.13E-23	0.0745	260.0
39	6506.0	124.0	20	1	6	co2	1.00	1.94E-24	5.78E-23	1.17E-24	0.0754	279.9
40	6769.0	82.0	20	1	6	co2 h2o	2.00	6.39E-25	1.76E-23	4.26E-25	0.0754	243.4
41	6940.0	120.0	20	1	8	co2 h2o	1.01	5.98E-23	1.69E-21	3.90E-23	0.0753	260.0

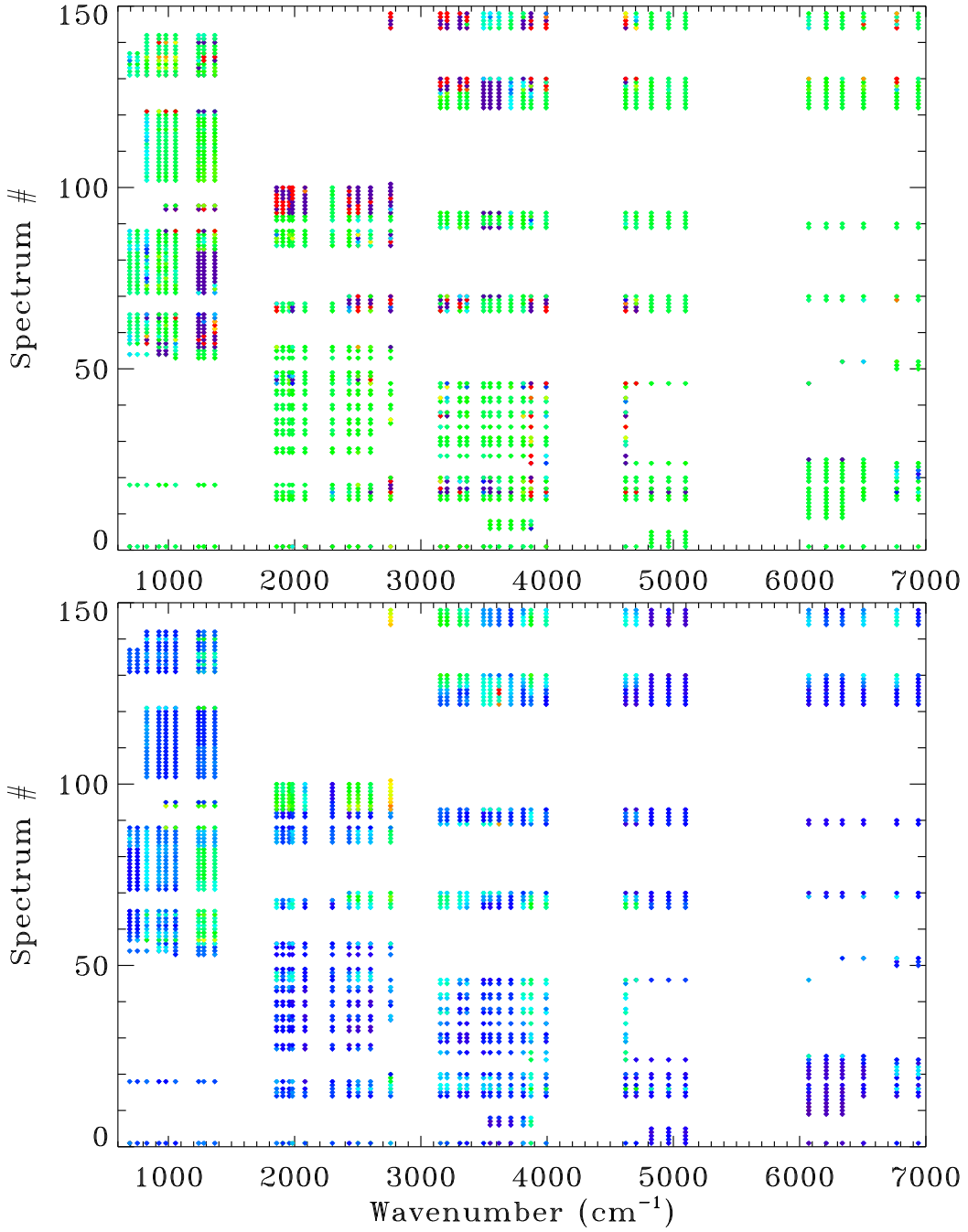
Retrieved CO_2 VMR Scale Factors

VMR Scale Factors (VSFs) are the ratio of the retrieved gas amount to that expected based on the measurement condition (cell length, T, P, VMR). In a perfect case, the VSFs should all be 1.0.

Upper Panel. The retrieved CO_2 VSFs values, for each window and each lab spectrum are color-coded (blue=0.5, green=1.0, red=1.5) and are plotted versus the window center wavenumber and an arbitrary spectrum #. Spectra 1-12 are from JPL (Sung), the remainder from Kitt Peak. Only one spectrum (#1) covers the full spectral range. Most cover less than 1000 cm^{-1}. Gaps in the wavenumber coverage (e.g. 4000-4600, 5200-6000 cm^{-1}) imply weak, undetectable CO_2 lines.

Lower Panel. The CO_2 VSF uncertainties are color-coded according to uncertainty (purple=0.2%; blue=1%; green=10%; yellow=100%; red=500%) and are plotted versus the window center wavenumber and the spectrum #. These uncertainties are based on the absorption depths of the CO2 lines and the fitting residuals.

On a subsequent slides the VSF values in the upper panel are averaged across each row (i.e., over windows) and down each column (i.e., over spectra). These averages are weighted using the uncertainties in the lower panel. Since regions with VSF values substantially different from 1 (blue or red) generally have large error bars, they don't adversely affect the average VSF values.

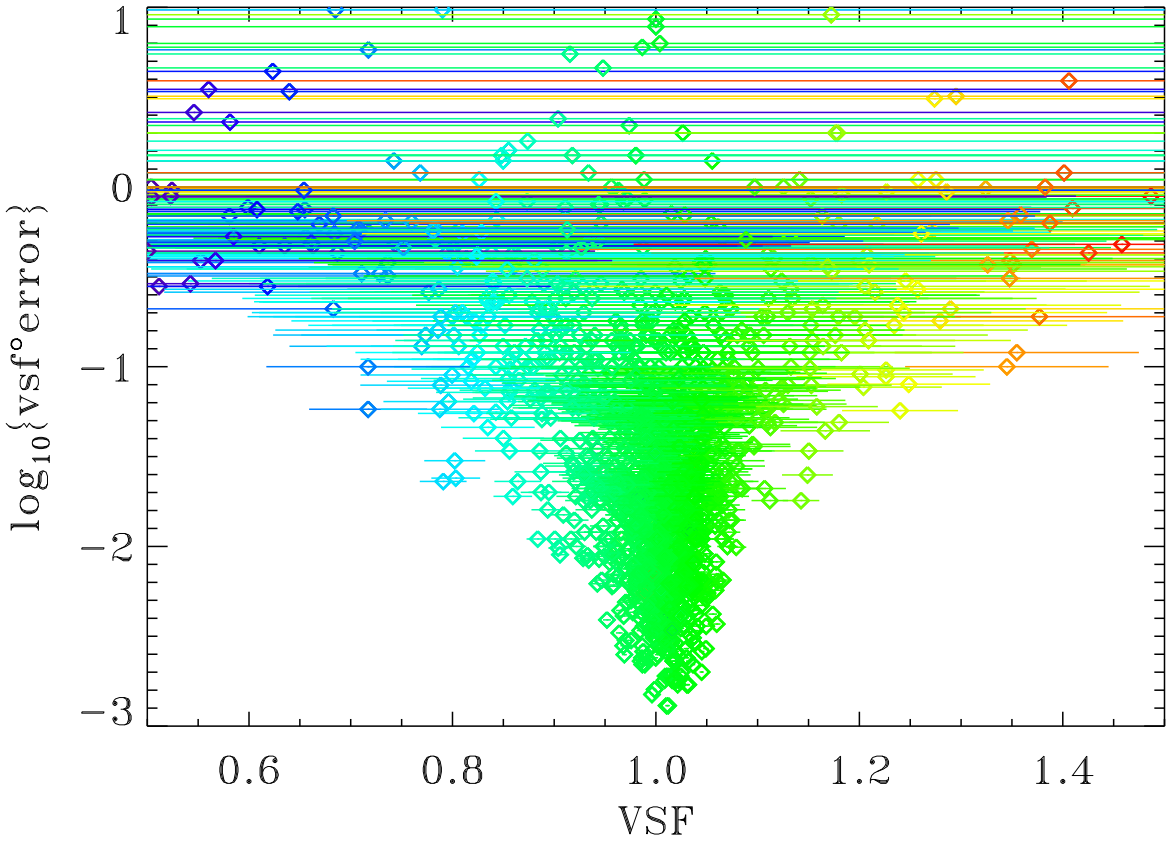


Investigating Outliers

The retrieved VMR scaling factors (VSF) are plotted against their uncertainties (VSF_error) for each of the 1816 spectral fits that were performed. For the ATM18 linelist, the smallest uncertainties are 0.2% but correspond to VSF values close to 1.0.

When the retrieved CO_2 VSF is far from 1.0, the error bar is usually large. Exceptions to this (i.e. points far from VSF=1 with small error bars) need investigation because they would have a large influence on the averaging over spectra and over windows.

This plot was much worse before I began investigating and correcting outliers. Several were discovered to be typographic errors in the entry of the cell measurements conditions. Others were due to zero-level offsets in the spectra. Others were due to unidentified contaminants within the cell. Many of the remaining outliers are associated with windows in which substantial CO_2 absorption is not from the parent isotopolog. In these cases errors in the assumed fractionation of the lab sample can cause erroneous retrieved CO_2 amounts, even though the spectral fits are excellent and therefore the error bars small.

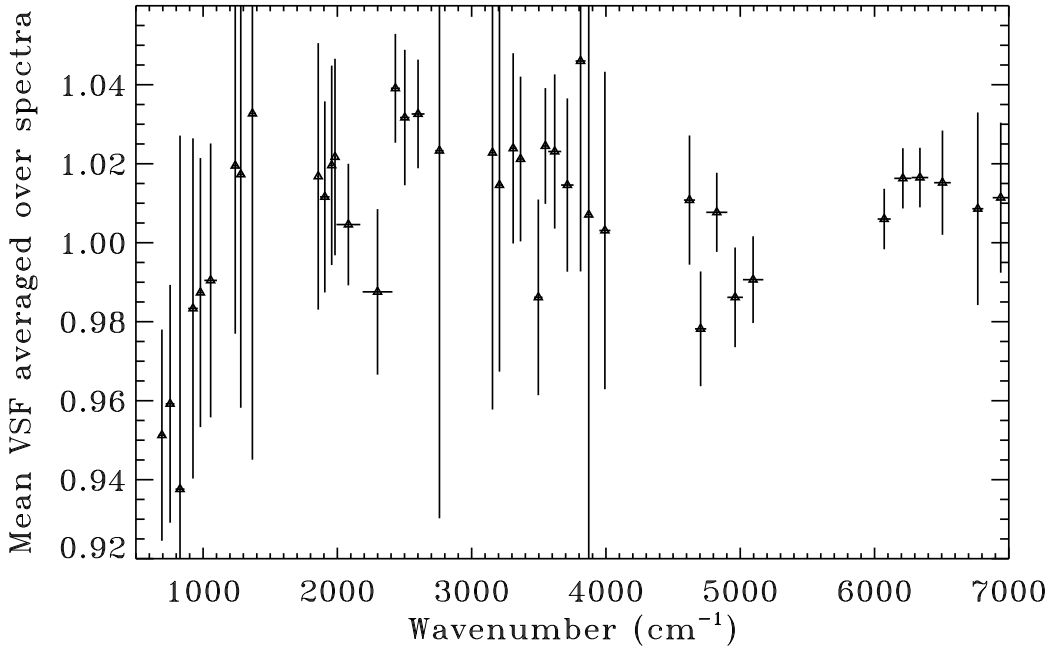


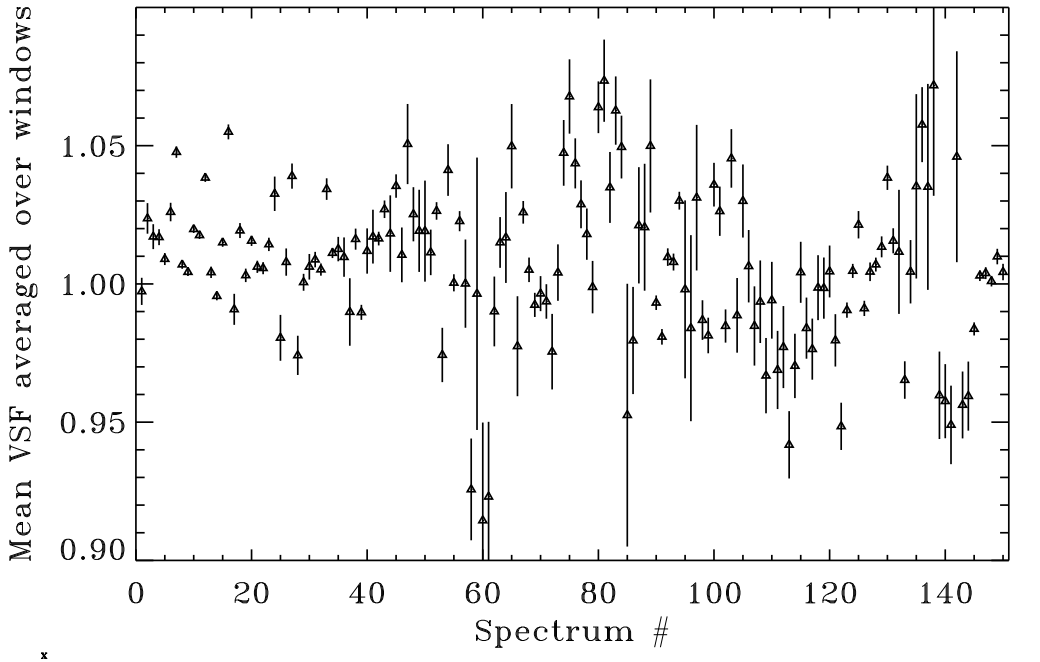
CO_2 VSFs averaged by window (top) and by spectrum (bottom)

Top Panel: VSF values obtained using the ATM18 linelist, averaged over the different lab spectra fitted in a particular window and plotted versus its center wavenumber. This exposes windows in which the retrievals are wrong due to factors common to the majority of the fitted spectra, e.g., spectroscopy.

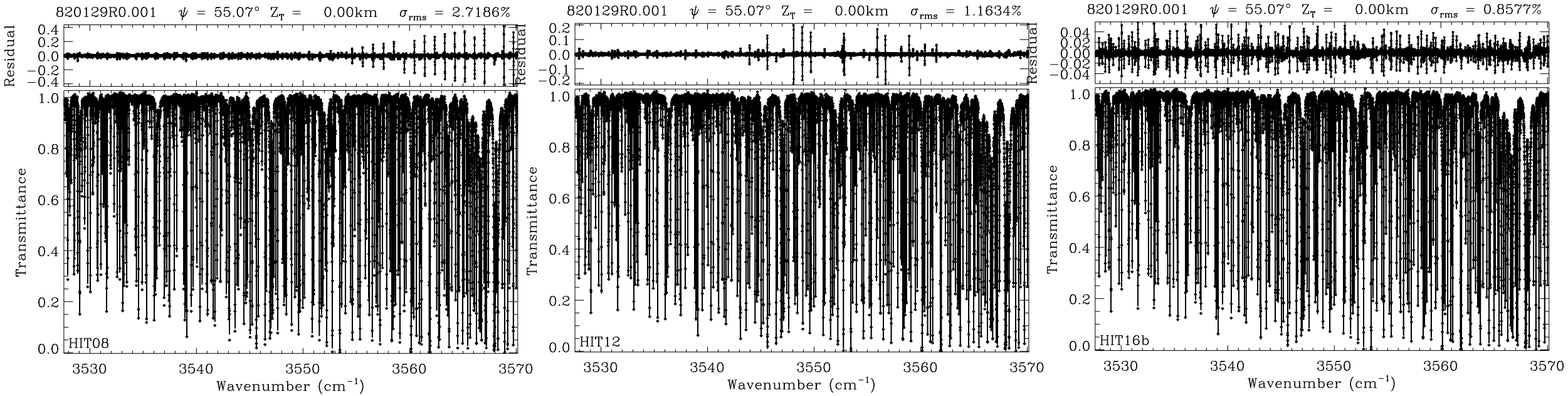
Bottom Panel: VSF values from a particular spectrum averaged over the fitted windows and plotted versus spectrum#. This exposes spectra in which the retrievals are wrong due to factors specific to that particular spectrum, e.g. the assumed VMR, Pressure, Temp, or path length may be wrong. Or the ILS might be mis-aligned. Or a large zero-offset is present.

These plots summarize the information presented 2 slides ago by averaging the columns and the rows. In general the error bars in the lower panel are smaller than those in the upper panel, which implies that spectrum-to-spectrum uncertainties in retrieved CO_2 are larger than window-to-window variations.



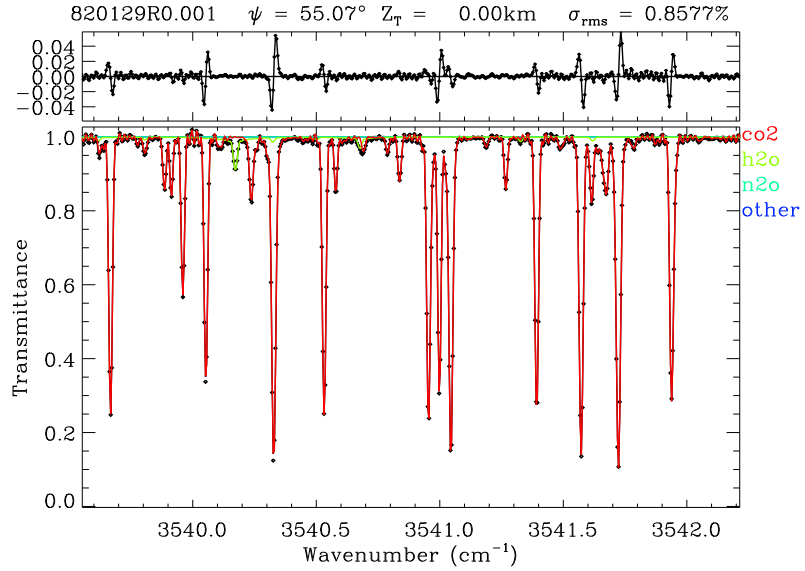


Spectral fits in the 3553 cm^{-1} window using recent HITRAN linelists

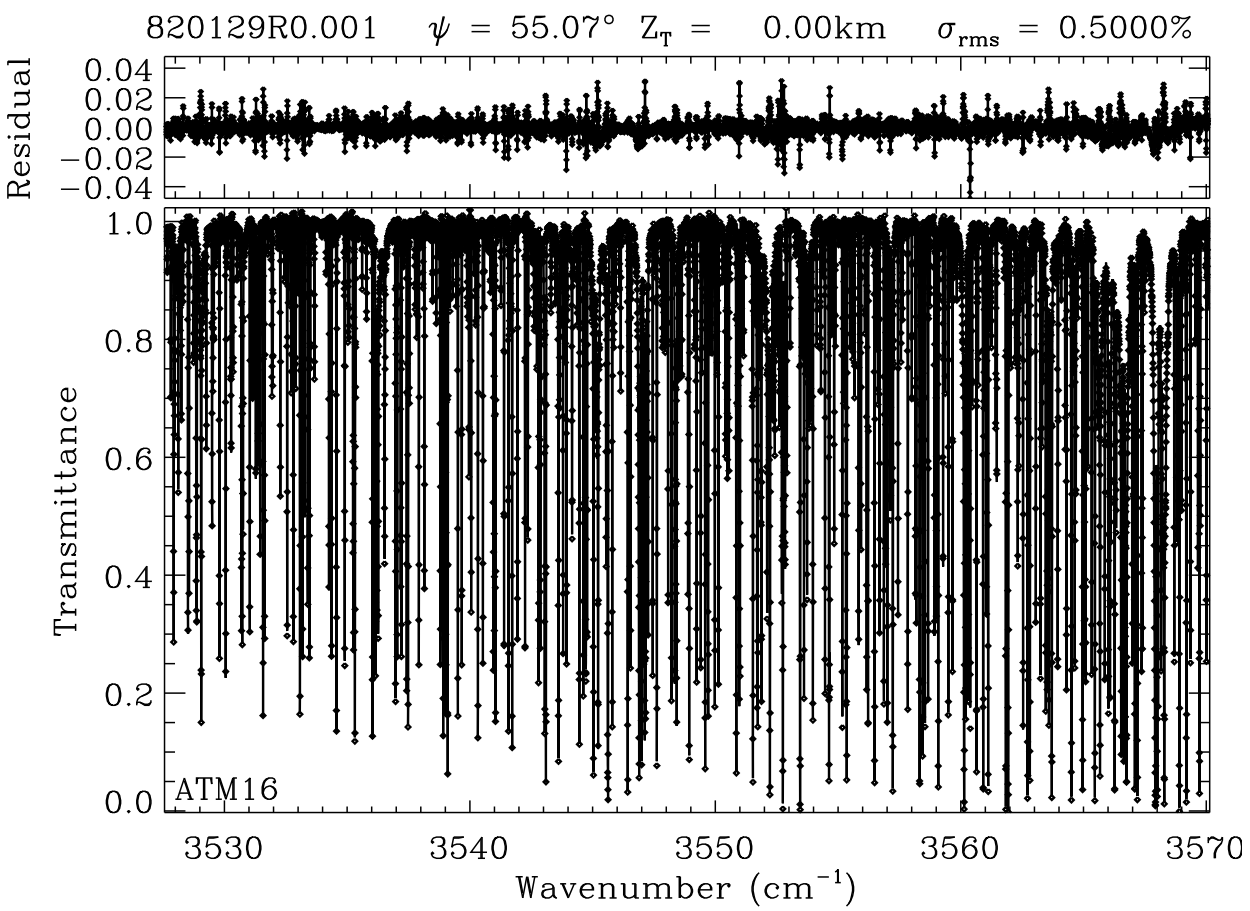


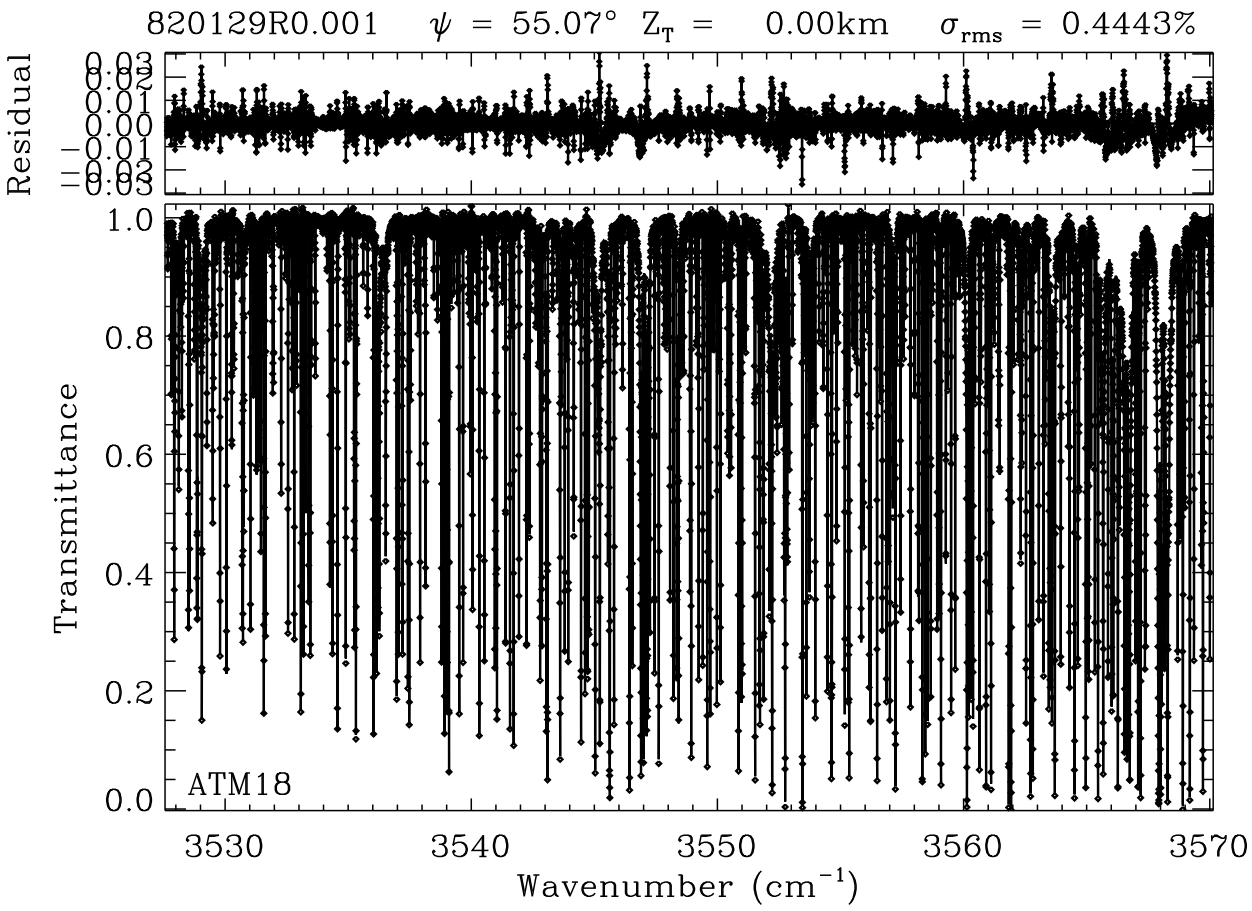
The panels above show the improvements to fits to a Kitt Peak lab spectrum (1.0 Torr of CO_2 at 297K in a 384 m path) in the 3548 cm^{-1} window. Peak residuals (due to line position errors) decrease from over 40% in HIT08 (left), to 20% in HIT12 (middle), to 6% in HIT16 (right).

The right hand panel zooms into a portion of the fit performed with HIT16b. The anti-symmetrical residuals are indicative of remaining line position errors.



Fits to Kitt Peak lab spectrum using ATM linelists at 3353 cm^{-1}





Showing the improvement between ATM16 (left) and ATM18 (right). Both are better than HIT16 by a factor of nearly 2 in terms of peak and rms residuals in this particular low-pressure spectrum, which has a high sensitivity to position errors. This improvement is due to ad hoc correction of line position errors such as those shown in the previous slide.

Average % RMS Fitting Residuals: Lab Spectra

10,896 spectral fits performed out of a potential 36,408 (148 spectra x 41 windows x 6 linelists) representing 29.9% coverage).

N_{used} is the number of spectra fitted in each window, out of the 150 total.

Values in the table are average rms spectral fitting residuals for a each window (averaged over fitted spectra).

Red values highlight the worst/largest pre-2018 RMS fits for each window; Blue values highlight the best/smallest.

Rectangles show the linelist portions on which ATM18 is based.

In "atm18" column, **bold typeface** indicates a better RMS than any predecessor linelist. Normal typeface means equal best. An "!" on far right means not the best (only 1 minor instance).

Iwin	v_{cen}	Width	N_{used}	N_{tot}	HIT08	HIT12	ATM16	HIT16a	HIT16b	ATM18	
1	693.5	24.7	39	148	0.4939	0.4771	0.4771	0.4736	0.4738	0.4738	!
2	754.5	32.7	39	148	0.5121	0.5037	0.5037	0.5033	0.5033	0.5033	
3	828.3	33.2	64	148	0.8084	0.7953	0.7953	0.7951	0.7951	0.7951	
4	925.0	36.0	66	148	0.4223	0.4222	0.4222	0.4216	0.4216	0.4216	
5	980.0	21.0	68	148	0.3112	0.3111	0.3111	0.3109	0.3109	0.3109	
6	1056.4	46.9	69	148	0.4007	0.3997	0.3997	0.3992	0.3992	0.3992	
7	1239.5	20.3	69	148	0.5158	0.5127	0.5127	0.5045	0.5044	0.5044	
8	1280.6	19.9	69	148	0.8634	0.8634	0.8634	0.8602	0.8601	0.8601	
9	1367.4	25.9	69	148	0.7182	0.7180	0.7180	0.7173	0.7173	0.7173	
10	1857.4	8.2	40	148	0.1666	0.1670	0.1670	0.1547	0.1547	0.1547	
11	1906.5	35.5	40	148	0.2151	0.2146	0.2146	0.2127	0.2127	**0.2121**	
12	1958.0	14.3	40	148	0.1737	0.1721	0.1721	0.1686	0.1687	**0.1681**	
13	1982.5	8.5	40	148	0.1368	0.1340	0.1340	0.1277	0.1277	0.1277	
14	2082.0	89.0	40	148	0.2531	0.2472	0.2472	0.2424	0.2424	**0.2410**	
15	2299.0	111.0	41	148	0.6587	0.5636	0.5627	0.5097	0.4940	**0.4749**	
16	2432.0	33.5	43	148	0.1337	0.1338	0.1338	0.1323	0.1323	**0.1312**	
17	2502.0	35.0	43	148	0.1194	0.1195	0.1195	0.1182	0.1182	**0.1139**	
18	2601.0	48.0	43	148	0.1217	0.1218	0.1218	0.1190	0.1190	0.1190	
19	2760.0	31.0	44	148	0.2941	0.2937	0.2937	0.2886	0.2886	0.2886	
20	3155.0	21.0	42	148	0.2980	0.2980	0.2980	0.2979	0.2979	0.2979	
21	3207.0	25.0	42	148	0.2670	0.2665	0.2665	0.2660	0.2660	0.2660	
22	3309.0	24.5	42	148	0.2413	0.2421	0.2415	0.2398	0.2398	0.2398	
23	3364.0	26.0	42	148	0.2663	0.2687	0.2648	0.2549	0.2549	0.2549	
24	3496.3	31.3	42	148	0.8246	0.7765	0.6154	0.6847	0.6841	**0.5706**	
25	3548.8	21.2	45	148	0.8267	0.6046	0.5538	0.6041	0.6027	**0.5047**	
26	3618.4	48.6	45	148	1.0603	0.7757	0.7655	0.7158	0.7151	**0.6737**	
27	3712.6	46.6	45	148	0.8178	0.6046	0.6046	0.6121	0.6106	**0.5954**	
28	3811.0	37.0	45	148	0.2886	0.2881	0.2881	0.2881	0.2881	0.2881	
29	3872.0	16.0	46	148	0.2763	0.2753	0.2753	0.2753	0.2753	0.2753	
30	3992.0	42.0	43	148	0.1555	0.1537	0.1537	0.1536	0.1536	0.1536	
31	4622.0	41.5	43	148	0.1576	0.1520	0.1520	0.1527	0.1527	**0.1511**	
32	4705.0	42.0	33	148	0.1143	0.1154	0.1119	0.1114	0.1115	**0.1110**	
33	4825.0	78.5	37	148	0.1852	0.1860	0.1840	0.1924	0.1926	**0.1829**	
34	4962.0	58.0	37	148	0.2904	0.2831	0.2806	0.2792	0.2796	**0.2727**	
35	5096.0	76.0	37	148	0.2121	0.2133	0.2125	0.2104	0.2106	**0.2103**	
36	6072.0	49.0	36	148	0.0898	0.0902	0.0894	0.0900	0.0900	**0.0886**	
37	6211.0	64.0	35	148	0.1302	0.1288	0.1236	0.1280	0.1281	**0.1215**	
38	6338.0	62.0	36	148	0.1374	0.1359	0.1299	0.1352	0.1354	**0.1268**	
39	6506.0	62.0	31	148	0.1278	0.1280	0.1278	0.1281	0.1281	0.1278	
40	6769.0	41.0	32	148	0.1847	0.1850	0.1850	0.1848	0.1858	0.1847	
41	6940.0	60.0	32	148	0.2820	0.2823	0.2823	0.2837	0.2838	0.2820	
	Average over all windows:				**0.3552**	0.3326	0.3265	0.3256	0.3251	**0.3170**	

RMS Residuals from fits to laboratory spectra

Showing the RMS spectral fits for 41 windows, averaged over the 148 lab spectra. This is done for 6 different linelists. These are the same data tabulated on the previous slide.

Upper panel shows absolute RMS residuals. Lower panel shows differences from HIT12.

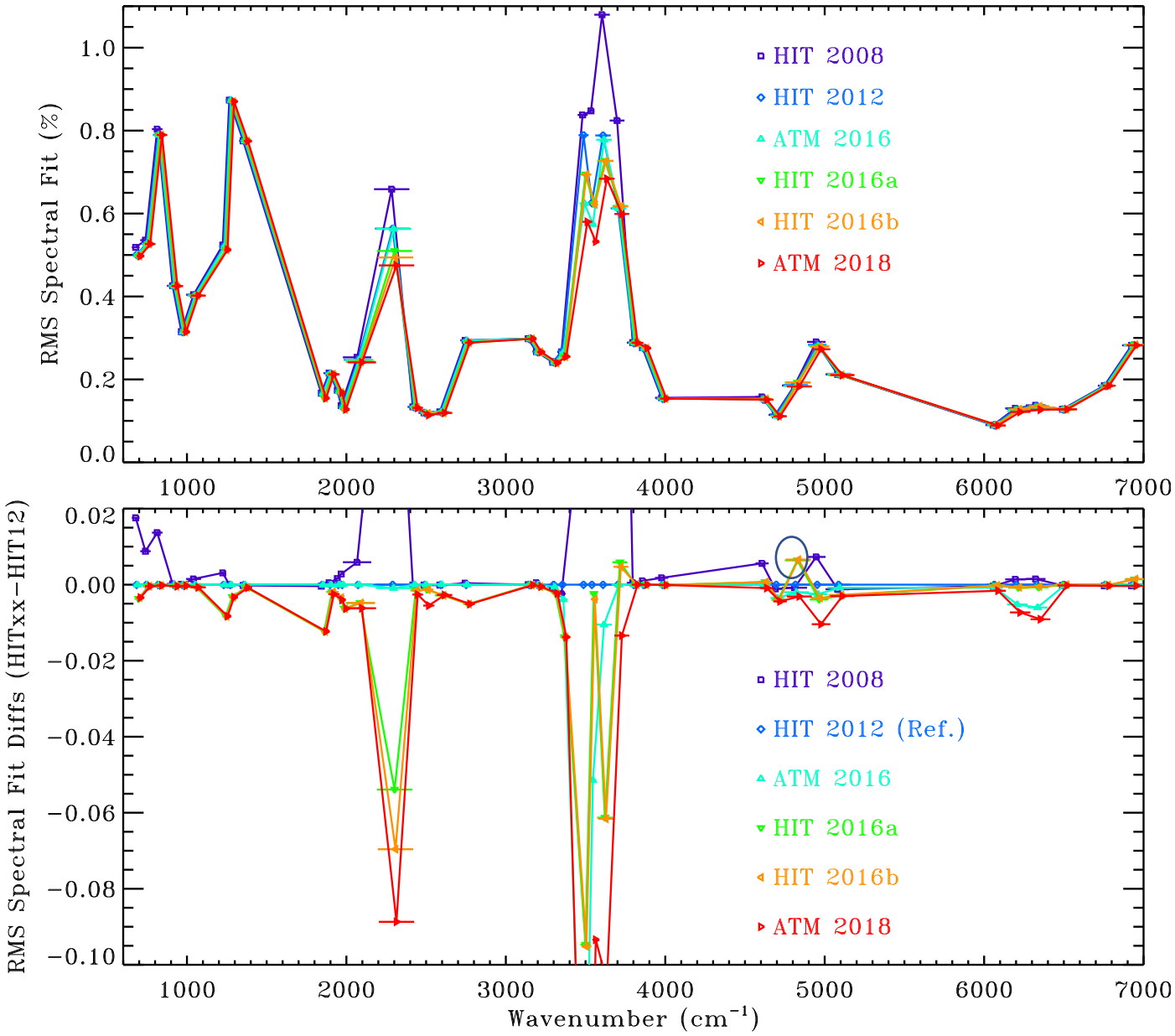
The absolute value of the RMS fit is unimportant. This is generally dominated by instrumental issues and interfering absorptions (e.g. H_2O).

Variation of RMS from linelist to linelist is entirely due to the CO_2 spectroscopy, since nothing else has been changed.

Big improvements are apparent for HIT16a,b in the 2300 cm^{-1} and the 3600 cm^{-1} regions.

HIT16a and HIT16b produce similar results. The largest difference is seen at 2300 cm^{-1} due to inclusion of isotopologs 11 & 12 i9nto HIT16b.

In the 4825 cm^{-1} window, HIT16 produces the worst fits (circled).



Retrieved Lab CO_2 VSFs

VMR Scale Factors (VSF) were obtained by averaging each retrieved single-spectrum VSF value over all spectra that were fitted for that window. The number of spectra averaged (Nrow) varies from 31 to 69.

The low VSF value of 0.7245 for HIT08 in the 3872 cm^{-1} window (circled) is the result of large ^{18}OCO line position errors, a major absorber in this window. This problem was fixed in later linelist editions.

At the foot of the table, results are shown after averaging over windows to obtain a retrieved CO_2 bias for each linelist, averaged over all windows and spectra.

Also calculated is the RMS deviation of the VSF from its mean value. These values represent the window-to-window variation in retrieved CO_2 amounts for each linelist.

The ATM18 linelist has the best window-to-window consistency (1.6%). Earlier linelists are all ~2.0%. Of course, this is a consequence of adjusting line intensities in ATM18 to correct windows (e.g. 1857, 1906, 6769 cm^{-1}) that were previously strongly-biased.

iwin	fcen	HWid	Nrow	Npp	HIT08	HIT12	ATM16	HIT16a	HIT16b	ATM18
1	693.3	24.3	39	148	0.9569	0.9459	0.9459	0.9518	0.9513	0.9513
2	754.5	32.7	39	148	0.9566	0.9482	0.9482	0.9594	0.9592	0.9592
3	828.3	33.1	64	148	0.9532	0.9309	0.9309	0.9378	0.9376	0.9376
4	925.0	36.0	66	148	0.9503	0.9535	0.9534	0.9835	0.9834	0.9834
5	980.0	21.0	68	148	0.9578	0.9560	0.9560	0.9876	0.9875	0.9874
6	1056.4	46.9	69	148	0.9626	0.9633	0.9633	0.9906	0.9905	0.9905
7	1239.5	20.3	69	148	1.0770	1.0908	1.0908	1.0195	1.0195	1.0195
8	1280.6	19.9	69	148	1.0988	1.1004	1.1004	1.0173	1.0173	1.0173
9	1367.4	25.9	69	148	1.0176	1.0255	1.0255	1.0336	1.0326	1.0327
10	1857.4	8.2	40	148	1.0102	0.9604	0.9604	1.0685	1.0684	1.0168
11	1906.5	35.5	40	148	1.0032	1.0103	1.0103	1.0634	1.0633	1.0116
12	1958.0	14.3	40	148	0.9923	1.0037	1.0037	0.9843	0.9842	1.0196
13	1982.5	8.5	40	148	0.9193	0.9514	0.9514	1.0217	1.0218	1.0217
14	2082.0	89.0	40	148	1.0087	1.0132	1.0133	1.0085	1.0085	1.0046
15	2299.0	111.0	41	148	1.0060	0.9890	0.9892	0.9876	0.9876	0.9876
16	2432.0	33.5	43	148	1.0279	0.9742	0.9742	1.0394	1.0393	1.0391
17	2502.0	35.0	43	148	0.9941	1.0194	1.0194	1.0316	1.0315	1.0317
18	2601.0	48.0	43	148	0.9865	0.9992	0.9992	1.0326	1.0326	1.0326
19	2760.0	31.0	44	148	1.0991	1.0552	1.0552	1.0235	1.0234	1.0233
20	3155.0	21.0	42	148	1.0013	1.0185	1.0185	1.0227	1.0228	1.0228
21	3207.0	25.0	42	148	1.0045	1.0036	1.0036	1.0146	1.0146	1.0146
22	3309.0	24.5	42	148	1.0272	1.0172	1.0174	1.0239	1.0239	1.0239
23	3364.0	26.0	42	148	1.0198	1.0300	1.0300	1.0212	1.0212	1.0212
24	3496.3	31.3	42	148	0.9802	0.9812	0.9843	0.9736	0.9736	0.9862
25	3548.8	21.2	45	148	1.0304	1.0238	1.0244	1.0137	1.0138	1.0242
26	3618.5	48.6	45	148	1.0344	1.0257	1.0253	1.0136	1.0131	1.0231
27	3712.6	46.6	45	148	1.0266	1.0209	1.0209	1.0059	1.0055	1.0146
28	3811.0	37.0	45	148	1.0365	1.0460	1.0460	1.0160	1.0160	1.0460
29	3872.0	16.0	46	148	0.7245	1.0069	1.0070	0.9922	0.9924	1.0071
30	3992.0	41.0	43	148	0.9952	0.9540	0.9540	1.0030	1.0030	1.0031
31	4622.0	41.5	43	148	1.0240	1.0280	1.0280	1.0110	1.0110	1.0110
32	4705.0	42.0	33	148	0.9829	0.9786	0.9780	0.9795	0.9797	0.9782
33	4825.5	78.5	37	148	1.0060	1.0050	1.0050	1.0070	1.0070	1.0080
34	4962.0	58.0	37	148	1.0050	1.0050	1.0050	0.9859	0.9859	0.9862
35	5096.0	76.0	37	148	1.0050	1.0070	1.0070	0.9911	0.9911	0.9906
36	6072.0	49.0	36	148	1.0200	1.0200	1.0160	1.0190	1.0200	1.0060
37	6211.0	64.0	35	148	1.0210	1.0200	1.0170	1.0140	1.0140	1.0160
38	6338.0	62.0	36	148	1.0210	1.0210	1.0180	1.0000	1.0000	1.0160
39	6506.0	62.0	31	148	1.0150	1.0160	1.0150	1.0150	1.0150	1.0150
40	6769.0	41.0	32	148	0.9685	0.9684	0.9684	0.9673	0.9675	1.0086
41	6940.0	60.0	32	148	1.0114	1.0114	1.0114	1.0023	1.0023	1.0114
Mean VSF (over windows)					**1.0103**	**1.0091**	**1.0081**	**1.0085**	**1.0085**	**1.0098**
RMS deviation from mean					**0.0192**	**0.0205**	**0.0207**	**0.0191**	**0.0191**	**0.0160**

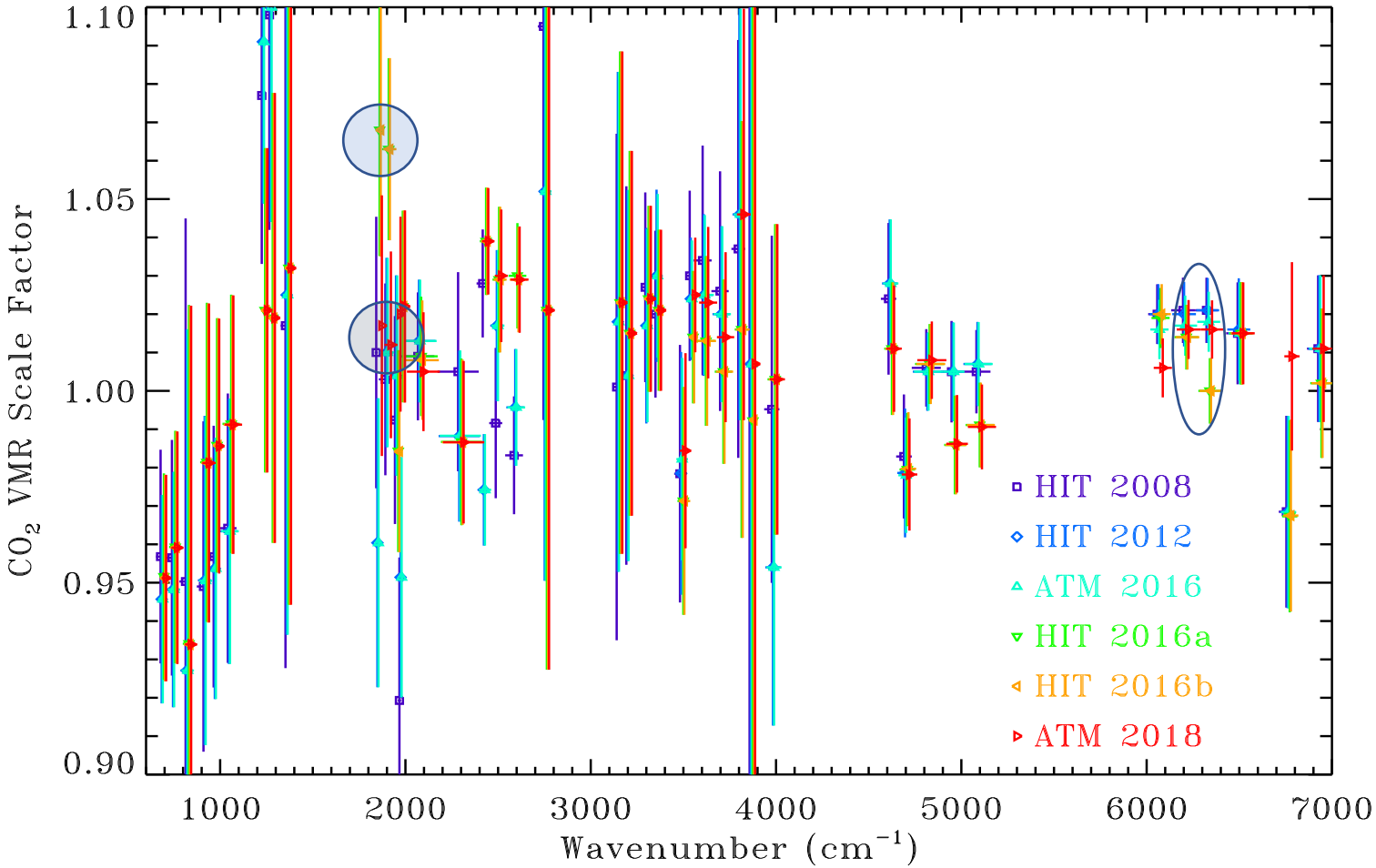
CO_2 VMR Scale Factors retrieved from lab spectra

Comparison of VSF values for all 6 linelists using the data tabulated in the previous slide, together with their untabulated uncertainties. Red points (ATM 2018) are identical to those shown in top panel of slide # 9.

VSF values greater than 1 mean that the line intensities, or the absorber amounts, need to be multiplied by the VSF value.

Width and line position errors can also contribute to an incorrect retrieved CO_2 amount, but in this case the relationship between the VSF value and the width/position error is more complicated.

A discrepancy is apparent between retrievals in the 6220 and 6338 cm^{-1} bands using HIT16



Chris Boone reported 5-10% larger retrieved ACE CO_2 amounts from 1915 cm^{-1} band than from 2050 cm^{-1} band using HIT16 with ACE data. Kitt Peak lab spectra confirm this (upper circle). This bias didn't exist with earlier linelist editions. It was fixed in ATM18 (lower circle). A 4% reduction to ATM18 CO_2 intensities in the 6740 cm^{-1} band is also apparent.

RMS Residuals from fits to lab spectra - Discussion

HITRAN 2008 is clearly the worst overall. But above 6400 cm^{-1}, it is the best.
Of the pre-2018 linelists, HIT16b is the best overall.
ATM18 is of course the best overall, being cherry-picked from the best parts of the earlier linelists.

Comparing HITRAN 2012 with ATM 2016: They produce similar results

- In 14 windows ATM 2016 is better
- In 27 windows they produce equally good fits
- In 0 windows HITRAN 2012 is better

It is no surprise that HITRAN 2012 is nowhere better than ATM 2016. If it had been, I would have replaced the offending lines in ATM 2016 with those from HITRAN 2016. Additionally, empirical adjustments have been performed to the ATM linelist to fix obvious deficiencies (e.g., position errors).

Comparing HITRAN 16a and 16b: They produce very similar results. Small differences in fits to pure CO_2 spectra due to truncation of SBHW to "f5.3" format. Improvements seen in strong bands at 2300 cm^{-1} and ~3600 cm^{-1} in ^{13}C-enriched lab spectra due to addition of isotopologs 11 and 12 to HIT16b

- In 10 windows HIT16a is slightly better
- In 24 windows they produce the same rms fit
- In 7 windows HIT16b is slightly better

Since only 2/148 lab spectra used here was below 290K, these results don't really validate the T-dependent parameters.
In the 4825 cm^{-1} window the HIT16 linelists produces significantly poorer residuals than any predecessor.

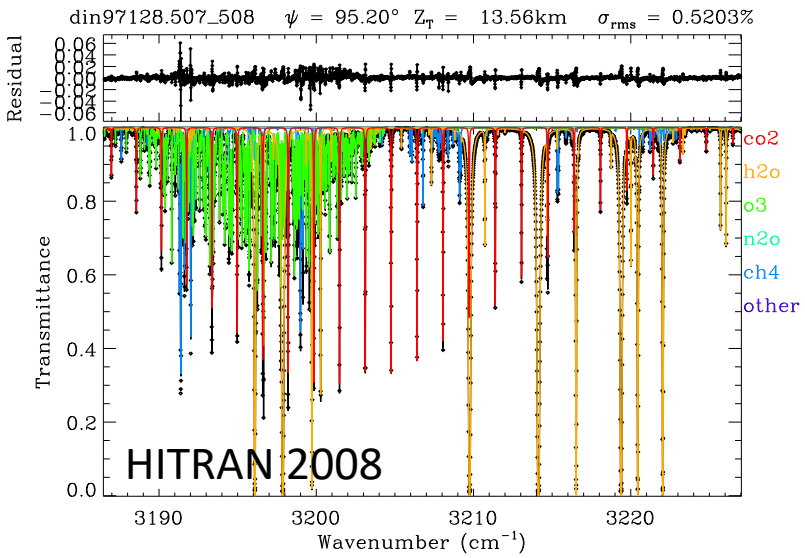
MkIV Balloon Spectra

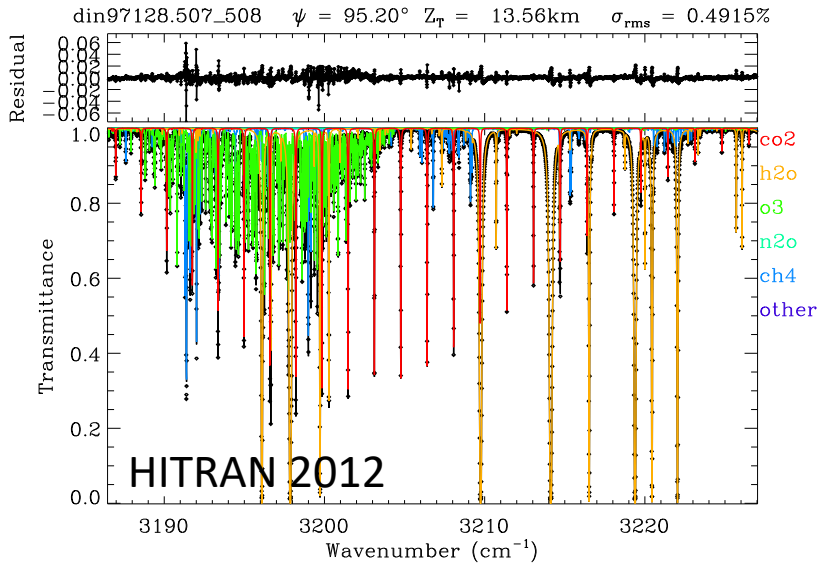
MkIV instrument observes direct sunlight from balloon, covering the entire 650-5650 cm^{-1} region simultaneously at 0.01 cm^{-1} resolution (60 cm OPD). Using the sun as a source allows a broad-bandwidth to be measured at high resolution and SNR. Broad simultaneous coverage is an important attribute when testing the band-to-band consistency of the spectroscopy. [In contrast, the majority of the Kitt Peak lab spectra have less then 1000 cm^{-1} of useful coverage, which means, for example, that the v_2 band is rarely in the same spectrum as the v_3]

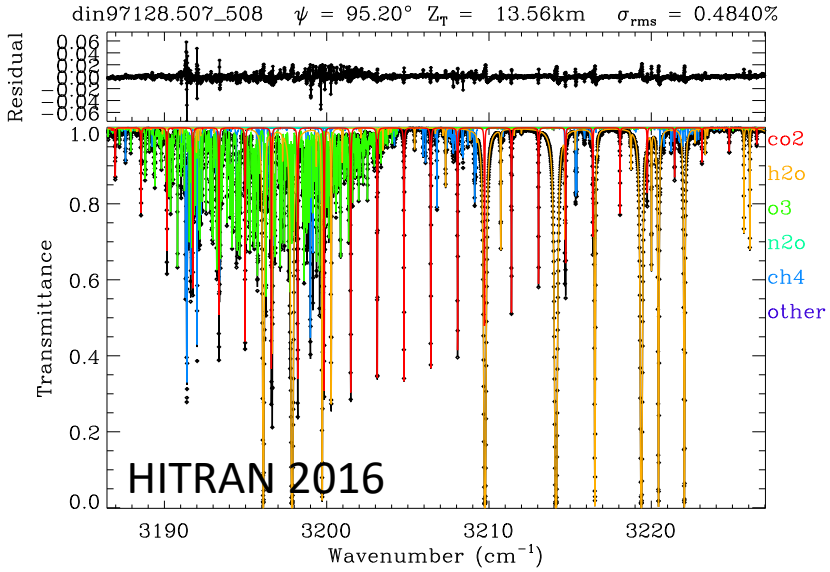
As the sun rises/sets, the ray path through the atmosphere passes through progressively lower/higher pressure. During the course of a 50 minute occultation, the tangent point varies from 8 to 40 km altitude encountering pressures from 3 mbar to 300 mbar and temperatures from 210 to 250K.

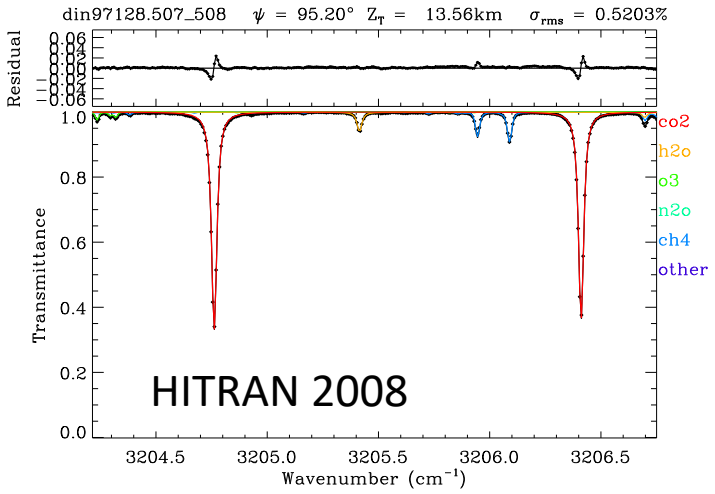
The CO_2 VSF (averaged over several good windows) is used to determine the viewing geometry (tangent altitude). It is therefore going to take an average value close to 1.0, by definition. The balloon spectra are therefore worthless for assessing absolute CO_2 band intensities. But the window-to-window biases are still valid. The average value of the VSF for the windows used in this work is not exactly 1.0 because an older linelist (pre-HITRAN_2K) was used in the determination of the tangent altitudes. Also, the windows used for the tangent point determination were only a small subset of those evaluated here.

MkIV balloon spectral fits at 14 km altitude in the 3200 cm^{-1} region

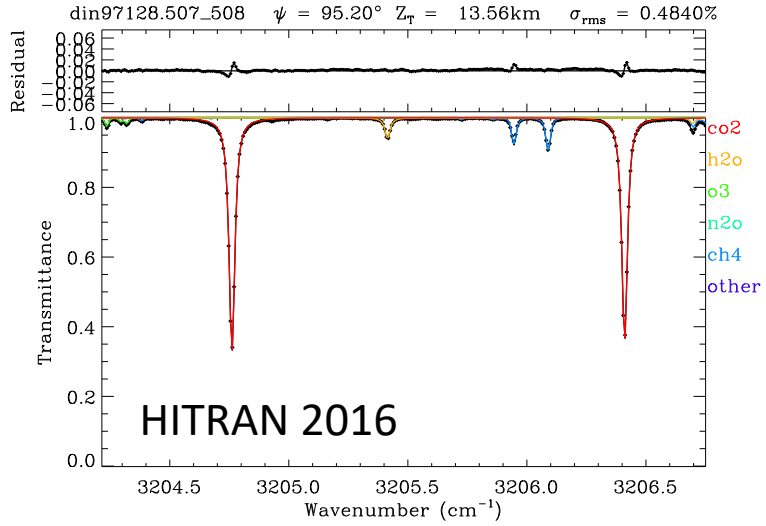








Illustrating the progressive improvement in the spectral fits from HIT08 to HIT16 in the 3205 cm^{-1} window where CO_2 lines are from $\nu_2+\nu_3$ combination band centered at 3182 cm^{-1}. Residuals are mainly due to H_2O, O_3 and CH_4, but these linelists were unchanged in this CO_2 study, so improvement in rms fitting residuals is due solely to CO_2. HIT16 still not perfect, but better than its predecessors.



RMS Fitting Residuals: MkIV Balloon

Table shows % rms fitting residuals for each window, averaged over N_{row} = 19 spectra covering tangent altitudes from 8 to 38 km

35 windows, 6 linelists, 19 spectra = 3990 fits

HITRAN 2012 is better than HITRAN 2008
HITRAN 2016 is better than HITRAN 2012

ATM16 is the best of the pre-2018 linelists, mainly due to the fact that (years ago) line position errors in the 3480-3570 cm^{-1} region were fixed manually.

For the ATM18 linelist, bold shading indicates a better RMS than any predecessor, normal shading indicates equal best, and "!" indicates not the best (only 5 instances). ATM18 is never the worst.

iwin	fcen	HWidth	Nrow	Npp	HIT08	HIT12	ATM16	HIT16a	HIT16b	ATM18	
1	694.6	23.6	19	19	0.6670	0.6597	0.6597	0.6571	0.6571	0.6571	
2	754.5	32.7	19	19	0.8274	0.7687	0.7687	0.7659	0.7659	0.7660	!
3	827.2	34.3	19	19	0.3490	0.3418	0.3418	0.3417	0.3417	0.3417	
4	925.0	36.0	19	19	0.2932	0.2927	0.2927	0.2926	0.2926	0.2926	
5	980.0	21.0	19	19	0.4610	0.4601	0.4601	0.4600	0.4600	**0.4599**	
6	1056.4	46.9	19	19	0.4219	0.4224	0.4224	0.4218	0.4218	0.4218	
7	1239.5	20.3	19	19	0.2529	0.2526	0.2526	0.2526	0.2526	0.2526	
8	1280.6	19.9	19	19	0.4311	0.4308	0.4308	0.4308	0.4308	0.4308	
9	1367.4	25.9	19	19	1.1334	1.1334	1.1334	1.1332	1.1332	**1.1329**	
10	1857.9	7.6	19	19	0.5769	0.5771	0.5771	0.5757	0.5757	0.5757	
11	1906.5	35.5	19	19	0.7008	0.7027	0.7027	0.7036	0.7036	0.7017	!
12	1958.0	14.3	19	19	0.5468	0.5451	0.5451	0.5498	0.5498	**0.5450**	
13	1982.5	8.5	19	19	0.4958	0.4884	0.4884	0.4879	0.4879	0.4879	
14	2082.5	88.5	19	19	0.7587	0.7543	0.7540	0.7530	0.7530	**0.7506**	
15	2299.3	125.1	19	19	0.5417	0.5195	0.5195	0.4839	0.4836	**0.4789**	
16	2430.6	32.1	19	19	0.1631	0.1628	0.1628	0.1626	0.1626	**0.1625**	
17	2501.9	34.8	19	19	0.2050	0.2044	0.2044	0.2047	0.2047	0.2045	!
18	2601.0	48.0	19	19	0.2565	0.2559	0.2559	0.2560	0.2560	0.2559	
19	2760.0	31.0	19	19	0.2641	0.2630	0.2630	0.2629	0.2629	0.2629	
20	3155.0	19.0	19	19	0.5220	0.5214	0.5214	0.5207	0.5207	0.5207	
21	3206.7	20.2	19	19	0.2183	0.2156	0.2156	0.2126	0.2126	0.2126	
22	3309.0	24.5	19	19	0.2037	0.1913	0.1901	0.1877	0.1877	0.1877	
23	3364.0	26.0	19	19	0.2435	0.2195	0.2120	0.2110	0.2110	0.2110	
24	3496.3	31.3	19	19	1.5552	1.2246	0.4571	0.7028	0.7028	**0.4302**	
25	3548.8	21.2	19	19	1.4110	0.8466	0.6371	0.7253	0.7253	**0.6163**	
26	3618.5	48.6	19	19	0.7801	0.5186	0.5143	0.4649	0.4649	**0.4595**	
27	3713.5	41.5	19	19	0.9043	0.6283	0.6283	0.6240	0.6240	**0.6228**	
28	3811.2	37.0	19	19	0.9409	0.9046	0.9046	0.9041	0.9041	0.9042	!
29	3871.9	15.8	19	19	1.1194	1.0621	1.0621	1.0621	1.0621	1.0621	
30	3991.0	42.0	19	19	0.4540	0.4535	0.4535	0.4535	0.4535	0.4535	
31	4622.1	41.4	19	19	0.2916	0.2798	0.2798	0.2793	0.2793	**0.2788**	
32	4705.0	42.0	19	19	0.2936	0.2820	0.2806	0.2852	0.2852	**0.2794**	
33	4825.0	78.0	19	19	0.5019	0.5052	0.5031	0.5498	0.5498	0.5026	!
34	4962.0	58.0	19	19	0.5863	0.5710	0.5688	0.5697	0.5697	**0.5419**	
35	5096.0	76.0	19	19	0.5282	0.5288	0.5275	0.5221	0.5221	**0.5189**	
Average over all windows:					**0.5686**	**0.5197**	**0.4912**	**0.4991**	**0.4991**	**0.4852**	

MkIV Balloon: RMS Spectral Fitting Residuals

Top Panel: Plot of the data tabulated on the previous slide. Shows RMS residuals for 35 windows using 6 different linelists. The absolute fitting residuals are dominated by interfering atmospheric absorptions, especially H_2O.

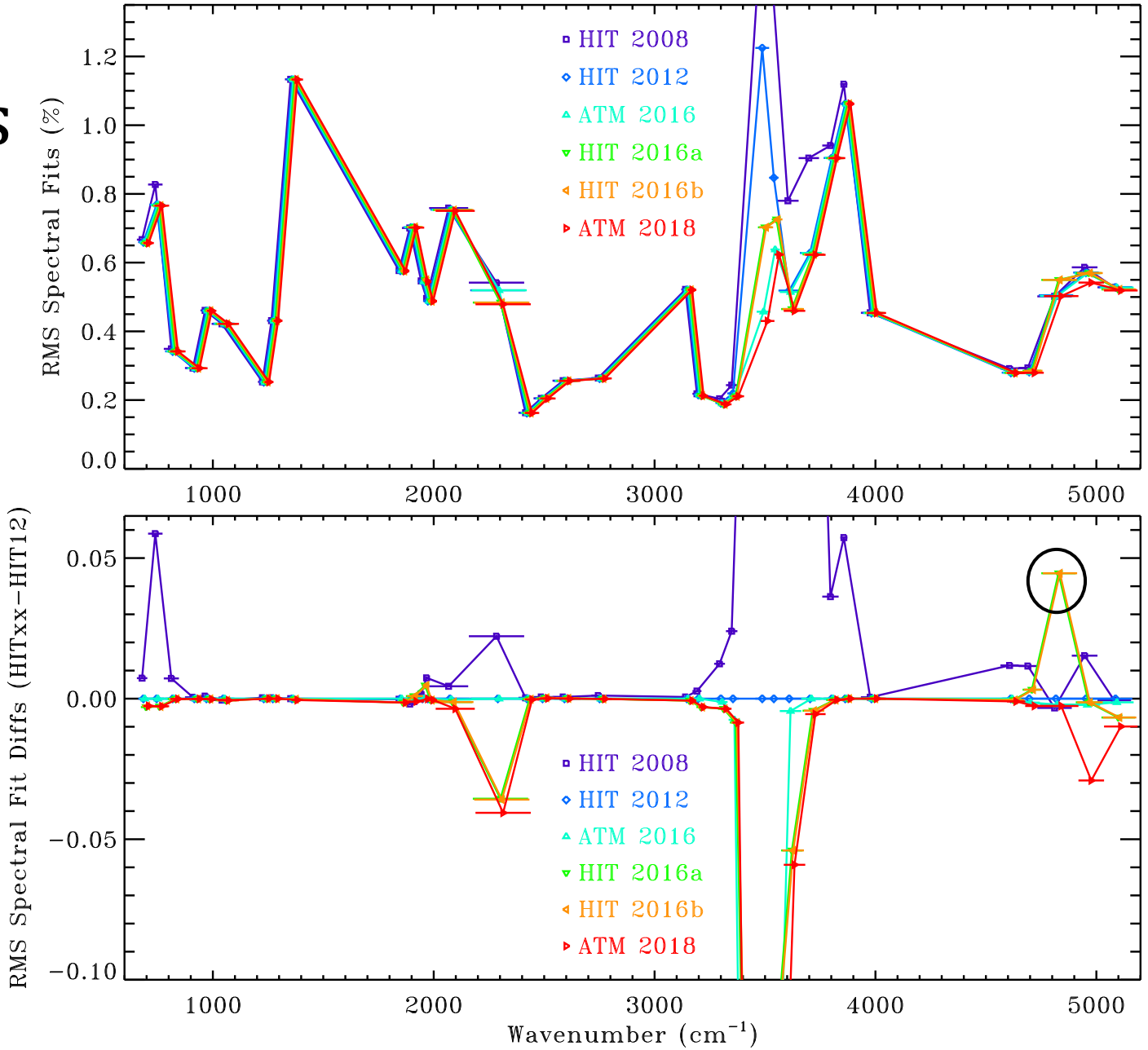
Bottom Panel: Differences from HIT12.

Difference between HIT16a and HIT16b are tiny because isotopologs 11 & 12 are not discernable in atmospheric spectra and because the rounding of the SBHW values doesn't matter in air-broadened spectra.

HIT16 shows improvements over HIT12 below 900 cm^{-1}, and in windows centered at 2290, 3496, 3548, 3623, 4962 and 5096 cm^{-1}.

In the 4825 cm^{-1} window, used by OCO & GOSAT, the HIT16 linelist achieves the worst fits (circled) and HIT08 the best, as for lab spectra.

The ATM18 linelist is always best, or close to.



Retrieved CO_2 VSFs from MkIV Balloon

The VMR Scale Factors (VSF) were obtained by averaging each retrieved single-spectrum VFS value over all 19 spectra that were fitted for each window.

At the foot of the table, results are shown for averaging over all windows and all spectra to obtain a retrieved CO_2 bias for each linelist.

Also calculated is the RMS deviation of this average VSF from its mean, representing the window-to-window variation in retrieved CO_2 amounts.

The HIT08 linelist has the best window-to-window consistency (1.68%), closely followed by ATM18 (1.74%). HIT12 has the worst window-to-window consistency (2.36%).

As in the case of lab spectra, the 3871.9 cm^{-1} window produces an abnormal VSF value of 0.106 for the HIT08 linelist (circled). This is because much of the CO_2 information in this window comes from ^{18}OCO absorption lines which were incorrectly positioned by 0.05 cm^{-1} in HIT08, a problem that was fixed in subsequent linelists. Since the CO_2 lines here are not particularly strong, the rms residuals (2 slides ago) are not severely impacted.

iwin	fcen	H-Width	Nrow	Npp	HIT08	HIT12	ATM16	HIT16a	HIT16b	ATM18
1	694.6	23.6	19	19	1.0865	1.0698	1.0698	1.0780	1.0780	1.0780
2	754.5	32.7	19	19	1.0333	1.0140	1.0140	1.0250	1.0250	1.0250
3	827.2	34.3	19	19	1.0054	0.9980	0.9980	1.0068	1.0068	1.0068
4	925.0	36.0	19	19	0.9990	0.9893	0.9893	1.0271	1.0271	1.0288
5	980.0	21.0	19	19	1.0223	0.9959	0.9959	1.0421	1.0421	1.0423
6	1056.4	46.9	19	19	1.0130	0.9997	0.9997	1.0352	1.0352	1.0352
7	1239.5	20.3	19	19	1.0420	1.0343	1.0343	1.0429	1.0429	1.0429
8	1280.6	19.9	19	19	1.0384	1.0274	1.0274	1.0437	1.0437	1.0437
9	1367.4	25.9	19	19	1.0315	1.0315	1.0315	1.0465	1.0465	1.0465
10	1857.9	7.6	19	19	1.0235	0.9741	0.9741	1.0578	1.0578	1.0074
11	1906.5	35.5	19	19	0.9900	0.9968	0.9968	1.0550	1.0550	1.0014
12	1958.0	14.3	19	19	0.9870	1.0004	1.0004	0.9646	0.9646	1.0030
13	1982.5	8.5	19	19	0.9569	0.9978	0.9978	0.9880	0.9880	0.9881
14	2082.5	88.5	19	19	1.0010	1.0115	1.0116	1.0120	1.0120	1.0063
15	2299.3	125.1	19	19	1.0466	1.0187	1.0187	1.0133	1.0132	1.0128
16	2430.6	32.1	19	19	0.9992	0.9436	0.9436	0.9994	0.9994	0.9994
17	2501.9	34.8	19	19	0.9860	1.0165	1.0165	1.0127	1.0127	1.0115
18	2601.0	48.0	19	19	0.9963	1.0142	1.0142	1.0169	1.0169	1.0169
19	2760.0	31.0	19	19	1.0415	0.9821	0.9821	1.0248	1.0248	1.0242
20	3155.0	19.0	19	19	0.9848	1.0008	1.0008	1.0052	1.0052	1.0052
21	3206.7	20.2	19	19	0.9948	0.9950	0.9950	1.0053	1.0053	1.0053
22	3309.0	24.5	19	19	0.9957	1.0004	1.0008	1.0044	1.0044	1.0044
23	3364.0	26.0	19	19	1.0009	1.0146	1.0147	1.0028	1.0028	1.0028
24	3496.3	31.3	19	19	0.9542	0.9792	0.9913	0.9829	0.9829	0.9902
25	3548.8	21.2	19	19	0.9781	0.9926	0.9969	0.9845	0.9845	0.9968
26	3618.5	48.6	19	19	0.9920	0.9772	0.9781	0.9737	0.9737	0.9715
27	3713.1	46.2	19	19	0.9923	0.9949	0.9949	0.9762	0.9762	0.9950
28	3811.2	37.0	19	19	1.0163	1.0334	1.0334	1.0042	1.0042	1.0335
29	3871.9	15.8	19	19	0.1060	1.0300	1.0300	1.0235	1.0235	1.0302
30	3991.0	42.0	19	19	0.9575	0.9281	0.9281	0.9725	0.9725	0.9739
31	4622.1	41.4	19	19	1.0106	1.0133	1.0133	0.9989	0.9989	0.9973
32	4705.0	42.0	19	19	0.9899	0.9864	0.9866	0.9890	0.9890	0.9869
33	4825.9	78.0	19	19	0.9948	0.9923	0.9924	0.9894	0.9894	0.9955
34	4962.0	58.0	19	19	0.9944	0.9943	0.9944	0.9809	0.9809	0.9768
35	5096.0	76.0	19	19	0.9783	0.9836	0.9836	0.9706	0.9706	0.9676
Mean VSF (over windows)					**1.0023**	**0.9986**	**0.9987**	**1.0088**	**1.0088**	**1.0080**
RMS deviation from mean					**0.0168**	**0.0236**	**0.0231**	**0.0191**	**0.0191**	**0.0174**

MkIV Balloon: Retrieved CO_2 VMR Scale Factors

Plotting the VSF values tabulated in the previous slide, along with their error bars.

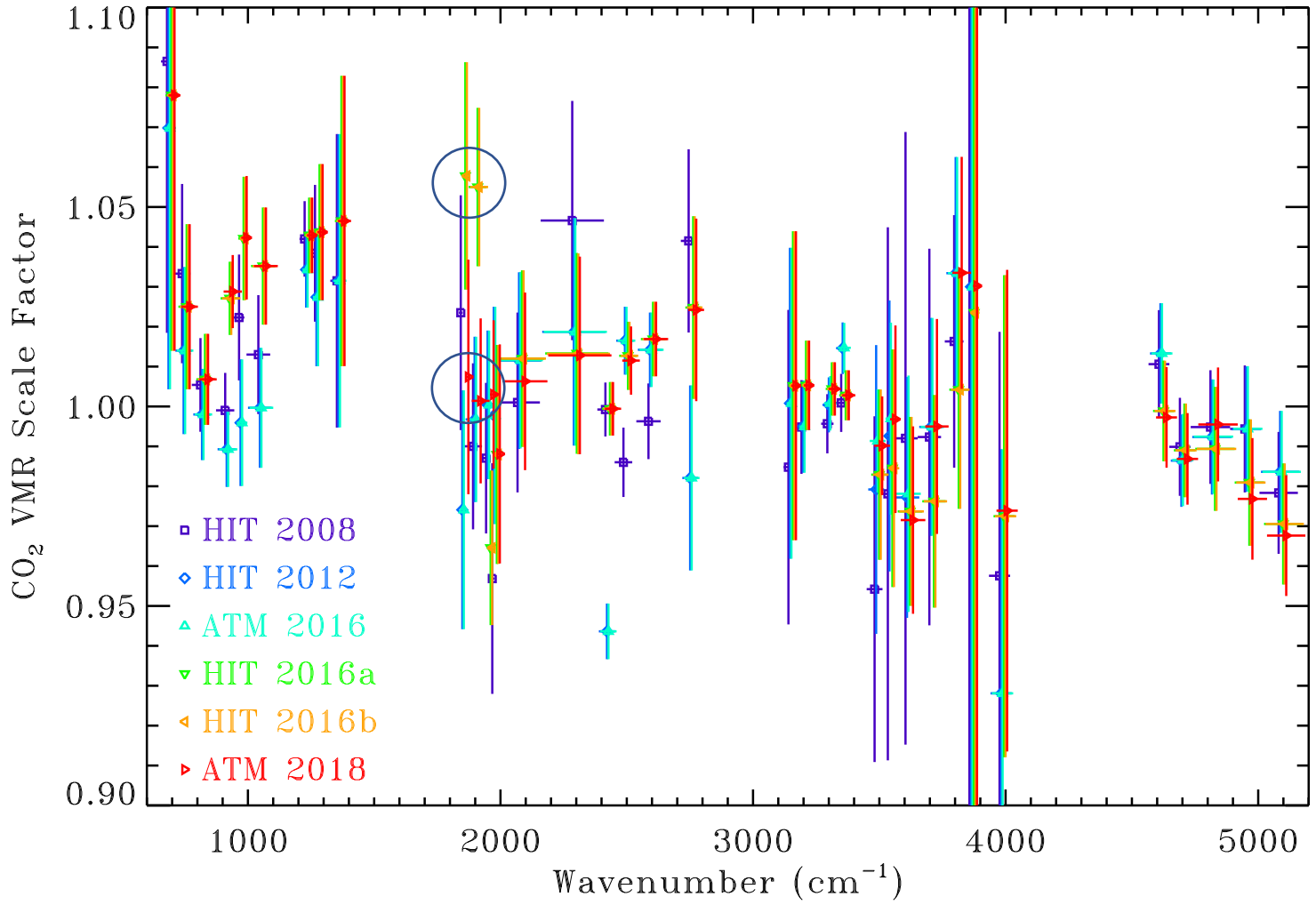
HIT12 points mostly buried under the ATM16 points, except around 3600 cm^{-1}.

HIT16a points mostly buried beneath HIT16b points. X-values offset for clarity.

MkIV instrument records 600-5650 cm^{-1} simultaneously, so derived VSFs should have good window-to-window consistency.

Upper circle indicate anomalously high HIT16 values in the 1800-2000 cm^{-1} region, which were also seen in lab spectra. Lower circle shows pre-2016 HITRAN & ATM18.

Somewhat high (~1.04) VSFs for all linelists are seen in 1200-1400 cm^{-1} region containing the v_1 band (symmetric stretch) of the ^{17}O and ^{18}O isotopologs. These are not definitive enough to warrant fixing.



MkIV Ground-Based RMS Fitting Residuals

34 windows, 122 spectra, 6 linelists = 24,888 spectral fits (100% completion).

Ground-based spectra are much more sensitive to line shape issues (widths, shifts, LM) than balloon spectra due to higher pressures (1 atm).

Fitted 122 spectra, each covering 650-5650 cm^{-1} simultaneously. Zenith angles from 20^{o} to 88^{o}. Surface temperatures of -30C to +35C. Altitudes from 0.0 to 3.8 km (650 to 1000 mbar).

In the ground-based case, the viewing geometry is computed from the zenith angle, not inferred from the CO_2 VSFs, which should therefore have good absolute accuracy, unlike the balloon case.

Blue/red shading implies the best/worst of the pre-2018 linelists.

For atm18, bold shading indicates a better RMS than any predecessor linelist, normal shading indicates equal best, and "!" indicates not the best. ATM18 is never the worst linelist.

iwin	Fcen	HWidth	Nrow	Npp	HIT08	HIT12	ATM16	HIT16a	HIT16b	ATM18	
2	755.2	32.0	122	122	1.1002	1.0775	1.0775	1.0830	1.0830	1.0827	!
3	827.2	34.3	122	122	0.5816	0.5799	0.5799	0.5781	0.5781	0.5781	
4	925.0	36.0	122	122	0.3415	0.3430	0.3430	0.3429	0.3429	0.3415	
5	980.0	21.0	122	122	0.3777	0.3816	0.3816	0.3808	0.3808	**0.3775**	
6	1056.4	46.9	122	122	0.6572	0.6540	0.6540	0.6536	0.6536	0.6536	
7	1239.5	20.3	122	122	0.7648	0.7649	0.7649	0.7649	0.7649	0.7649	!
8	1280.6	19.9	122	122	0.2640	0.2443	0.2643	0.2643	0.2643	0.2643	!
9	1367.4	25.9	122	122	0.0927	0.0927	0.0927	0.0927	0.0927	0.0927	
10	1857.9	7.6	122	122	0.2332	0.2332	0.2332	0.2332	0.2332	0.2332	
11	1910.3	29.7	122	122	0.6771	0.6782	0.6786	0.6807	0.6807	0.6801	!
12	1960.4	11.8	122	122	0.6336	0.6334	0.6334	0.6372	0.6372	0.6339	!
13	1981.5	8.5	122	122	0.6874	0.6868	0.6868	0.6867	0.6867	0.6867	
14	2082.5	88.5	122	122	1.1549	1.1647	1.1651	1.1657	1.1657	1.1550	
15	2299.3	125.1	122	122	0.4169	0.4183	0.4183	0.4185	0.4185	0.4185	!
16	2430.6	32.1	122	122	0.2102	0.2099	0.2099	0.2094	0.2094	0.2094	
17	2501.9	34.8	122	122	0.2462	0.2459	0.2459	0.2459	0.2459	0.2459	
18	2605.0	42.0	122	122	0.3710	0.3710	0.3710	0.3710	0.3710	0.3710	
19	2760.0	31.0	122	122	0.4228	0.4223	0.4223	0.4223	0.4223	0.4223	
20	3159.0	15.2	122	122	1.2236	1.2231	1.2231	1.2231	1.2231	1.2231	
21	3206.7	18.5	122	122	0.7805	0.7800	0.7800	0.7799	0.7799	0.7799	
22	3309.0	22.5	122	122	0.4706	0.4702	0.4702	0.4700	0.4700	0.4700	
23	3360.5	22.0	122	122	0.5317	0.5342	0.5339	0.5314	0.5314	0.5314	
25	3496.3	31.3	122	122	0.2577	0.2511	0.2493	0.2484	0.2490	**0.2264**	
25	3548.8	21.2	122	122	0.0919	0.0910	0.0912	0.0904	0.0904	0.0911	!
26	3618.5	48.6	122	122	0.1016	0.1079	0.1079	0.0995	0.0995	**0.0992**	
27	3713.5	41.5	122	122	0.0147	0.0147	0.0147	0.0148	0.0148	0.0147	
28	3811.2	37.0	122	122	0.0227	0.0227	0.0227	0.0227	0.0228	0.0227	
29	3871.9	15.8	122	122	0.1487	0.1487	0.1487	0.1487	0.1487	0.1487	
30	3991.0	42.0	122	122	0.6464	0.6438	0.6438	0.6440	0.6440	0.6440	!
31	4627.0	34.4	122	122	0.3436	0.3411	0.3411	0.3413	0.3413	0.3411	
32	4705.0	41.0	122	122	0.3082	0.3070	0.3067	0.3083	0.3083	**0.3064**	
33	4825.9	77.9	122	122	0.7303	0.7283	0.7276	0.7298	0.7298	**0.7171**	
34	4962.0	58.0	122	122	0.8648	0.8363	0.8363	0.7909	0.7909	**0.7820**	
35	5094.7	73.2	122	122	0.8106	0.8070	0.8069	0.8055	0.8055	**0.8051**	
Average over all windows:					**0.4894**	**0.4878**	**0.4878**	**0.4864**	**0.4864**	**0.4845**	

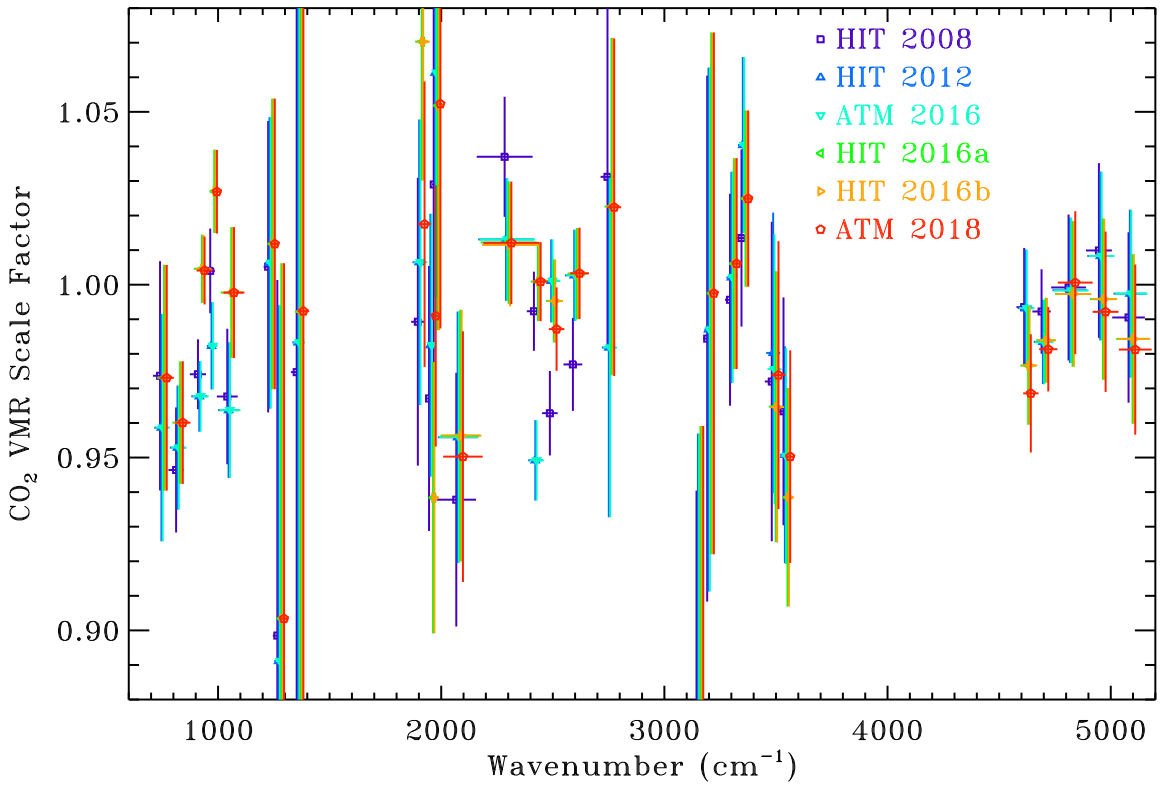
Retrieved MkIV Ground-based VSFs

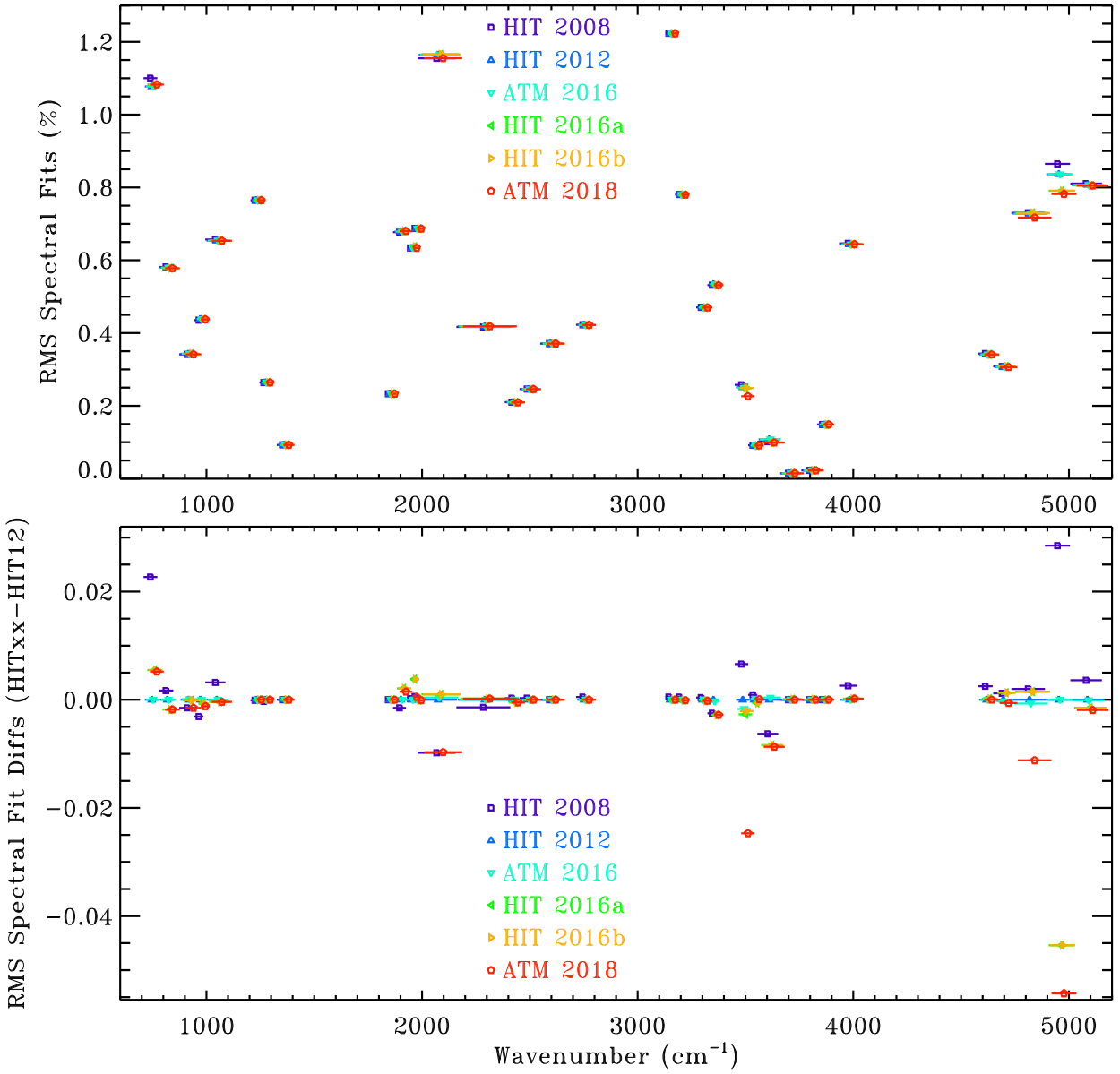
The ATM18 linelist has the smallest RMS deviation (2.02%) from the Mean VSF, indicating the the CO_2 retrieved from different windows is the most consistent. Its average VSF is 0.9961.

The 3700-4000 cm^{-1} windows (# 27 - 30) are so blacked out by H_2O from the ground that a meaningful estimate of the CO_2 VSF cannot be made.

iwin	fcen	H_Width	Nrow	Npp	HIT08	HIT12	ATM16	HIT16a	HIT16b	ATM18
2	755.2	32.0	124	124	0.9669	0.9521	0.9521	0.9664	0.9664	0.9664
3	827.2	34.3	124	124	0.9407	0.9472	0.9472	0.9543	0.9543	0.9543
4	925.0	36.0	124	124	0.9685	0.9622	0.9622	0.9988	0.9988	0.9984
5	980.0	21.0	124	124	0.9985	0.9770	0.9770	1.0214	1.0214	1.0213
6	1056.4	46.9	124	124	0.9613	0.9574	0.9574	0.9912	0.9912	0.9912
7	1239.5	20.3	124	124	1.0085	1.0097	1.0097	1.0152	1.0152	1.0152
8	1280.6	19.9	124	124	0.9012	0.8939	0.8939	0.9061	0.9061	0.9061
9	1367.4	25.9	124	124	0.9773	0.9854	0.9854	0.9946	0.9946	0.9947
10	1857.9	7.6	124	124	1.0321	0.9838	0.9838	1.0726	1.0726	1.0220
11	1910.3	29.7	124	124	0.9826	0.9998	0.9998	1.0634	1.0634	1.0106
12	1960.4	11.8	124	124	0.9603	0.9757	0.9757	0.9319	0.9319	0.9841
13	1981.5	8.5	124	124	1.0212	1.0528	1.0528	1.0441	1.0441	1.0446
14	2082.5	88.5	124	124	0.9322	0.9501	0.9502	0.9506	0.9506	0.9446
15	2299.3	125.1	124	124	1.0334	1.0095	1.0095	1.0085	1.0085	1.0085
16	2430.6	32.1	124	124	0.9869	0.9440	0.9440	0.9953	0.9953	0.9953
17	2501.9	34.8	124	124	0.9662	1.0046	1.0046	0.9988	0.9988	0.9906
18	2605.0	42.0	124	124	0.9804	1.0061	1.0061	1.0066	1.0066	1.0066
19	2760.0	31.0	124	124	1.0340	0.9845	0.9845	1.0253	1.0253	1.0252
20	3159.0	15.2	124	124	0.7954	0.8133	0.8133	0.8156	0.8156	0.8156
21	3206.7	18.5	124	124	0.9771	0.9797	0.9797	0.9901	0.9901	0.9901
22	3309.0	22.5	124	124	0.9888	0.9953	0.9953	0.9992	0.9992	0.9992
23	3360.5	22.0	124	124	1.0069	1.0336	1.0337	1.0182	1.0182	1.0182
24	3496.3	31.3	124	124	0.9750	0.9829	0.9783	0.9692	0.9683	0.9765
25	3548.8	21.2	124	124	0.9621	0.9502	0.9495	0.9375	0.9375	0.9495
26	3618.5	48.6	124	124	0.9729	0.9520	0.9518	0.9479	0.9479	0.9481
27	3713.5	41.5	124	124	–	–	–	–	–	–
28	3811.2	37.0	124	124	–	–	–	–	–	–
29	3871.9	15.8	124	124	–	–	–	–	–	–
30	3991.0	42.0	124	124	–	–	–	–	–	–
31	4627.0	34.4	124	124	0.9967	0.9963	0.9963	0.9797	0.9797	0.9717
32	4705.0	41.0	124	124	0.9966	0.9877	0.9876	0.9882	0.9882	0.9856
33	4825.9	77.9	124	124	0.9961	0.9954	0.9953	0.9944	0.9944	0.9976
34	4962.0	58.0	124	124	1.0051	1.0035	1.0035	0.9913	0.9913	0.9875
35	5094.7	73.2	124	124	0.9841	0.9909	0.9909	0.9779	0.9779	0.9738
Mean VSF (over windows)					**0.9857**	**0.9829**	**0.9828**	**0.9952**	**0.9975**	**0.9961**
RMS deviation from mean					**0.0218**	**0.0233**	**0.0233**	**0.0215**	**0.0215**	**0.0202**

MkIV Ground-based: RMS Fitting residuals (right) VMR Scale Factors (below)





The MkIV ground VSFs (above) are generally close to 1. For the 3150 cm^{-1} window, however, the VSF values are around 0.8 for all linelists and therefore off the bottom of the plot. But the tops of the error bars are visible, not quite reaching 1. In this window, there is strong absorption from H_2O and CH_4 which dwarfs that of CO_2.

TCCON Ground-Based RMS Spectral Fitting Residuals (%)

iwin	Fcen	Width	Nrow	Npp	HIT08	HIT12	ATM16	HIT16a	HIT16b	ATM18
31	4627.0	34.4	25	25	0.2041	0.1991	0.1991	0.1994	0.1994	0.1991
32	4705.0	41.0	25	25	0.1591	0.1556	0.1549	0.1583	0.1583	**0.1539**
33	4825.9	78.0	25	25	0.4546	0.4507	0.4495	0.4493	0.4493	**0.4333**
34	4962.0	58.0	25	25	0.6759	0.6386	0.6382	0.5800	0.5800	**0.5670**
35	5094.7	73.2	25	25	0.6874	0.6826	0.6825	0.6780	0.6780	**0.6758**
36	6071.7	48.7	25	25	0.2215	0.2245	0.2218	0.2198	0.2198	**0.2154**
37	6211.0	64.0	25	25	0.2042	0.2017	0.1932	0.2014	0.2014	**0.1913**
38	6338.0	60.5	25	25	0.2011	0.1992	0.1908	0.1991	0.1991	**0.1896**
39	6506.2	60.7	25	25	0.1780	0.1802	0.1780	0.1803	0.1803	0.1780
40	6769.0	41.1	25	25	0.8308	0.8309	0.8309	0.8309	0.8309	0.8308
41	6937.5	59.5	25	25	0.6651	0.6679	0.6679	0.6663	0.6663	0.6651
Average over all windows:					**0.4074**	**0.4028**	**0.4006**	**0.3966**	**0.3966**	**0.3908**

The Total Column Carbon Observing Network (TCCON) covers the 3950 to 9500 cm^{-1} region with an InGaAs detector at 45 cm OPD (0.02 cm^{-1} resolution). Some instruments also cover 9500 to 15500 cm^{-1} with a parallel Si detector, but these spectra were not used in this work. TCCON measures CO_2 using the two near-identical bands centered at 6220 and 6338 cm^{-1}. The OCO-2 and GOSAT satellite instruments also use the stronger CO_2 band contained in the 4825 cm^{-1} window.

In regions of weak absorption (e.g., windows 31, 32, 37, 38, 39), the fitting residuals approach the noise level (0.15%).

In 7/11 windows (**bold**) the ATM18 produces better fits than any predecessor. In the other 4 windows ATM18 is equal best. Nowhere does ATM18 produce inferior fits to any predecessor linelist.

TCCON Ground-Based Retrieved VMR Scale Factors (%)

iwin	Fcen	Width	Nrow	Npp	HIT08	HIT12	ATM16	HIT16a	HIT16b	ATM18
1	4627.0	34.4	25	25	1.0174	1.0170	1.0170	0.9999	0.9999	0.9912
2	4705.0	41.0	25	25	1.0048	0.9956	0.9955	0.9963	0.9963	0.9934
3	4825.9	77.9	25	25	1.0049	1.0043	1.0042	1.0030	1.0030	1.0066
4	4962.0	58.0	25	25	1.0186	1.0162	1.0161	1.0051	1.0051	1.0021
5	5094.7	73.2	25	25	1.0107	1.0178	1.0179	1.0042	1.0042	1.0009
6	6071.7	48.7	25	25	1.0163	1.0171	1.0123	1.0166	1.0166	1.0038
7	6211.0	64.0	25	25	1.0061	1.0059	1.0053	1.0003	1.0003	1.0049
8	6338.0	60.5	25	25	1.0046	1.0045	1.0043	0.9847	0.9847	1.0044
9	6506.2	60.7	25	25	1.0131	1.0141	1.0131	1.0138	1.0138	1.0131
10	6769.0	41.1	25	25	0.9647	0.9641	0.9641	0.9625	0.9625	1.0052
11	6937.5	59.5	25	25	0.9824	0.9808	0.9808	0.9713	0.9713	0.9824
Mean VSF (over windows)					**1.0091**	**1.0076**	**1.0066**	**1.0017**	**1.0017**	**1.0029**
RMS deviation from mean					**0.0061**	**0.0084**	**0.0076**	**0.0111**	**0.0111**	**0.0072**

The HIT08 has the largest mean VSF (1.0091) which means that CO_2 retrieved using this linelist will be biased 0.91% high compared with predictions from atmospheric models.

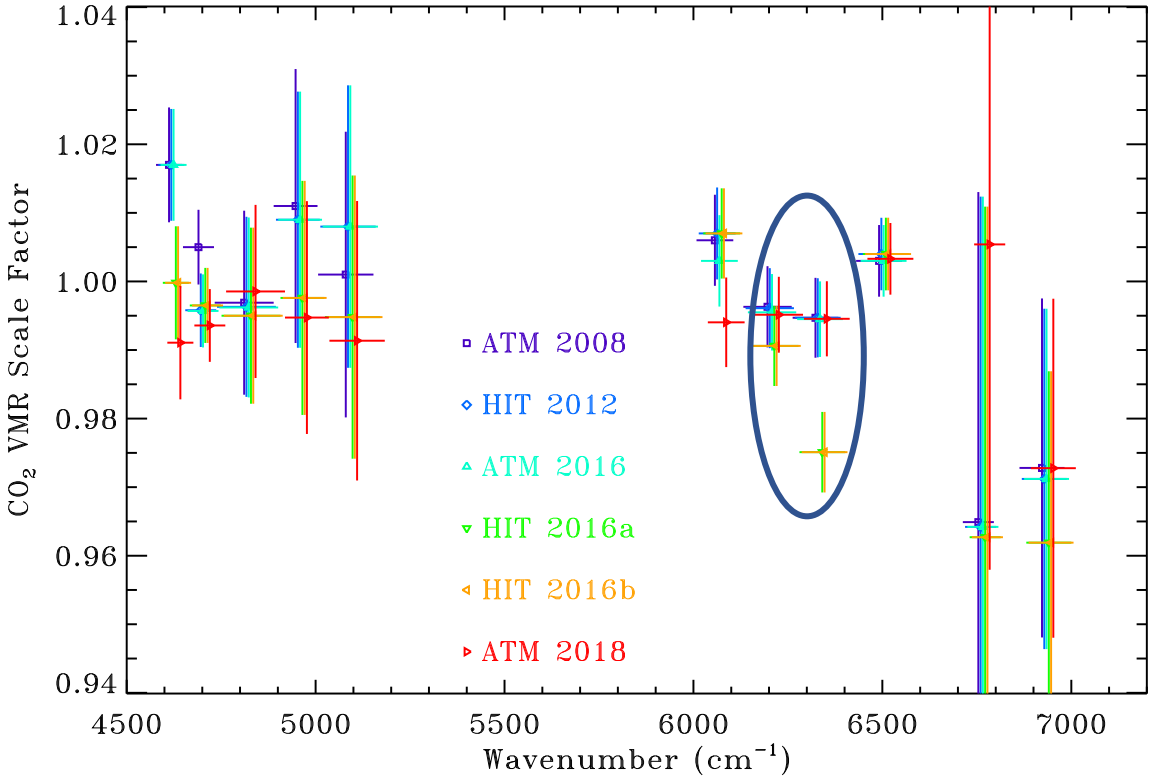
The HIT08 linelist has smallest RMS deviation from the mean (0.61%), hence the best window-to-window consistency.

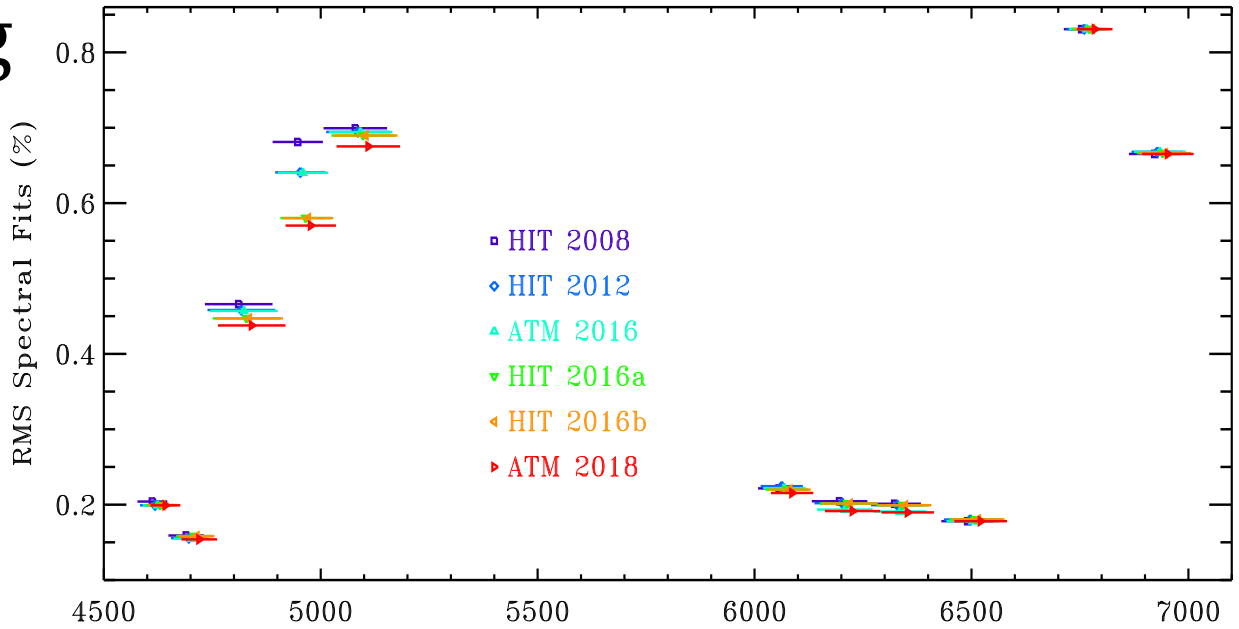
HIT16 linelists have the largest RMS deviation from mean (1.11%), hence the worst window-to-window consistency.

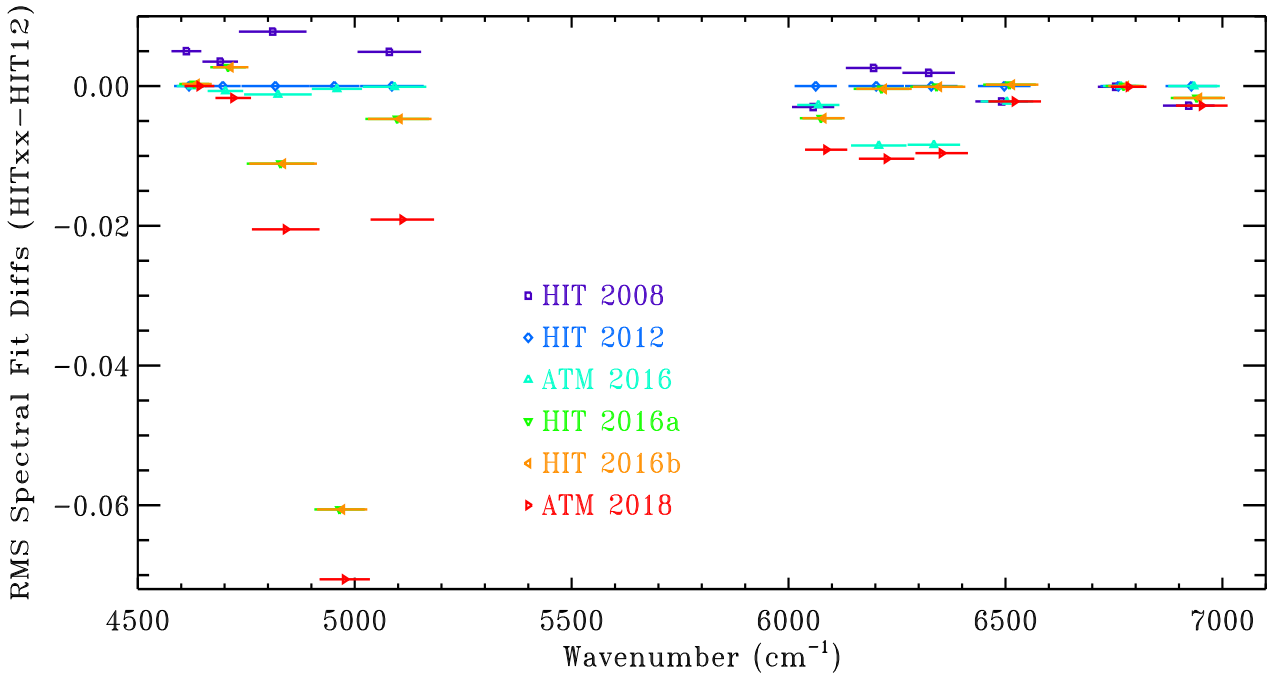
The HIT16 linelists retrieve 1.56% more CO_2 from the 6211 cm^{-1} window than the 6338 cm^{-1} window (circled).

TCCON Ground-Based RMS fitting residuals (right) & VSFs (below)

TCCON CO_2 windows circled below. The HIT16 linelists reduce the CO_2 retrieved from the 6221 and 6338 cm^{-1} windows by 0.5% and 1.5% respectively, as compared with the other linelists, introducing a new 1.5% inconsistency.







Effect of Line Mixing on RMS fits in 4825 cm^{-1} window

For speed of computation and simplicity we have so far ignored LM. To test its impact, we have included the effect of LM for a couple of test cases and compared the results with the non-LM case. The LM calculation was first-order implementation based on a subroutine signed F. Niro, 2005 and used in the paper: Hartmann, Tran, and Toon, 2009.

Type	fcen	Width	Nrow	Npp	HIT08	HIT12	ATM16	HIT16a	HIT16b	ATM18
Lab	4825.5	78.5	37	150	0.1852	0.1860	0.1840	0.1924	0.1926	**0.1829**
Lab LM	4825.5	78.5	37	150	0.1825	0.1839	0.1818	0.1915	0.1915	**0.1814**

For the lab spectra in the broad 4825 cm^{-1} window, the overall impact of LM is small since most of the spectra are at low pressure (only 30/148 spectra are above 250 Torr). But in every linelist, LM reduces the RMS residual by about 0.002% and has no effect on the linelist ranking. Linelist-to-linelist differences are nearly 0.01%.

Type	fcen	Width	Nrow	Npp	HIT08	HIT12	ATM16	HIT16a	HIT16b	ATM18
TCCON	4825.9	78.0	25	25	0.4546	0.4507	0.4495	0.4493	0.4493	**0.4333**
TCCON LM	4825.9	78.0	25	25	0.3331	0.3443	0.3442	0.3416	0.3416	**0.3420**

For ground based TCCON spectra the pressures are much higher and so the LM effect is larger, exceeding the linelist-to-linelist differences. In all cases, including LM reduces the residuals by about 0.1%. With or without LM, HIT16 is the best pre-2018 linelist and ATM18 is the best of all.

In both the Lab and TCCON cases, the improvement due to LM is larger using HIT2008 than any subsequent linelist. This might be because the HIT2008 linelist is closer to the HIT2000 linelist used by the Niro LM subroutine, and so the LM calculation is done more self-consistently.

We conclude that although LM has a significant effect on the rms and the VSF for spectra measured at higher pressures, it has little affect on the relative ranking of the linelists. The ATM16 remains the best of the pre-2018 linelists and HIT16b the worst. ATM18 is better than any predecessor, with or without LM.

Effect of Line Mixing on RMS Fits: 6211 cm^{-1} window

In the 6221 cm^{-1} window used by TCCON, OCO, and GOSAT, the lines are more than 10 times weaker than at 4825 cm^{-1} and so the benefits that line-mixing provide are much smaller.

	fcen	Width	Nrow	Npp	HIT08	HIT12	ATM16	HIT16a	HIT16b	ATM18
Lab	6211.0	64.0	35	150	0.1302	0.1288	0.1236	0.1280	0.1281	**0.1215**
Lab LM	6211.0	64.0	35	150	0.1297	0.1283	0.1234	0.1275	0.1276	**0.1211**

The HITRAN linelist improve their RMS fits by 0.0005% as a result of including LM. The ATM the improvements are smaller. Including LM makes no change to the ranking of the linelists. The LM effects in this window in lab spectra are ten times smaller than linelist-to-linelist differences in the CO_2 spectroscopy.

	fcen	Width	Nrow	Npp	HIT08	HIT12	ATM16	HIT16a	HIT16b	ATM18
TCCON	6211.0	64.0	25	25	0.2042	0.2017	0.1932	0.2014	0.2014	**0.1913**
TCCON LM	6211.0	64.0	25	25	0.2045	0.2017	0.1958	0.2017	0.2017	**0.1926**

For ground based TCCON spectra we found that at low airmasses including LM makes the RMS spectral fitting residuals worse, whereas at high airmasses they improve. The net result of this on a mixture of high and low-airmass spectra is small, with some linelists improving slightly (e.g. HIT12) as a result of LM, and others worsening (most notably ATM16).

We conclude that although LM has a significant effect on the rms and the VSF for spectral measured at higher pressures, it doesn't affect the relative ranking of the linelists. So choosing the right linelist has a much larger effect on the residuals than implementing LM. The ATM16 remains the best of the pre-2018 linelists and HIT16b the worst. ATM18 is better than any predecessor. With or without LM, HIT16 is the best pre-2018 linelist and ATM18 is the overall best.

Effect of Line Mixing on RMS fits in 6338 cm^{-1} window

In the 6338 cm^{-1} window used by TCCON, OCO, and GOSAT, the CO_2 lines are almost exactly the same strength width and E" as those in the 6221 cm^{-1} window.

Type	fcen	Width	Nrow	Npp	HIT08	HIT12	ATM16	HIT16a	HIT16b	ATM18
Lab	6338.0	62.0	36	150	0.1374	0.1359	0.1299	0.1352	0.1354	**0.1268**
Lab LM	6338.0	64.0	35	150	0.1367	0.1351	0.1295	0.1345	0.1347	**0.1263**

Including LM improves the RMS fits in all cases. The ATM linelist fits improve by only 0.0005% whereas the HIT linelists improve by 0.0007%. Including LM makes no change to the ranking of the linelists. The LM effects in this window are ten times smaller than linelist-to-linelist differences (0.005%) from the CO_2 spectroscopy.

Type	fcen	Width	Nrow	Npp	HIT08	HIT12	ATM16	HIT16a	HIT16b	ATM18
TCCON	6338.0	60.5	25	25	0.2011	0.1992	0.1908	0.1991	0.1991	**0.1896**
TCCON LM	6338.0	60.5	25	25	0.1987	0.1962	0.1917	0.1963	0.1963	**0.1879**

The inclusion of LM improves the fits for all linelists, except for ATM16, which worsens slightly. The ATM18 linelist still gives the best fits and HIT08 the worst, irrespective of whether LM is used or not.

[Effect of LM on VSF_CO2 should have been included, although this is strongly airmass-dependent].

Usefulness of spectra for Spectroscopy Evaluation

Type	Pros	Cons
Laboratory	- Well-known cell conditions (Leng, T, P, VMR) - VMR up to 1 are possible - Large isotopic enrichments possible	• Dim source, so narrow spectral coverage or poor SNR • Isotopic composition often uncertain
Occultation MkIV Balloon	- Bright source (sun) allows simultaneous coverage 650-5650 cm^{-1} at high resolution - Wide range of P/T conditions & slant colums - Solar and instrumental features removed - Long path lengths (~400 km)	• Inhomogenous atmospheric path • No control over P/T or VMR (400 ppm) • Interferences from other gases • CO_2 used to determine tangent altitude so no info on absolute CO_2 amounts
Ground-based MkIV / TCCON	- Bright Source (sun) - Broad simultaneous coverage - Long path lengths (~100 km) - Sensitive to lineshape (e.g. width, shifts, LM) - Accurate knowledge of airmass	• Inhomogeneous atmospheric path • No Control over P/T or VMR (400 ppm) • Wide regions blacked out: - H_2O (1350-1900; 3350-4000 cm^{-1}) - CO_2 (650-700; 2280-2390 cm^{-1})

Atmospheric spectra have better-known isotopic composition than lab spectra, unless the lab samples have been independently measured, e.g., by mass spectrometry. For example, for atmospheric CO_2, the $^{13}C/^{12}C$ ratio can be predicted anywhere to 0.1%. Atmospheric spectra contain no information on the SBHW.

In ground-based geometry, airmass is known to 0.1%, given a surface pressure measurement of 1 mbar accuracy.

Ad Hoc Linelist Corrections

After selecting lines from the best predecessor linelist for each spectra region, there would typically still be some large spectral fitting residuals. Sometimes these had an obvious cause (e.g. line position errors, pressure shifts).

Even though these defects were usually discovered in fits to atmospheric spectra (because I look at more atmospheric spectra than lab fits), their correction was always performed while fitting lab spectra. But the atmospheric spectra still play a role in deciding which fitting residuals to investigate.

Since the ad hoc correction process is highly error-prone, it is important that the spectra be refitted after the corrections have been made to ensure that the expected benefits materialized.

Are RMS fitting residuals a useful metric of linelist quality?

The RMS fitting residuals tell us about the consistency of the spectroscopy within a given window. They don't tell us whether it is right or wrong. For example, if all the CO_2 line intensities within a given window were 50% too high, we would still get a good spectral fit, after GFIT scales the assumed CO_2 vmr. Similarly, if all the CO_2 line positions within a given window were in shifted by 0.1 cm^{-1}, we would again get a good spectral fit because the GFIT code retrieves a frequency shift. In fact, these retrieved frequency stretches are a useful diagnostic, and are looked at but not reported here. Of course, if there are multiple gases present in a sample (e.g. CO_2 and H_2O) and the CO_2 lines have a position errors, but not the H_2O, then the position inconsistency will cause an increase in the RMS residuals because GFIT retrieves a single shift for the entire window, not one per gas.

Inconsistent Lineshape Assumptions

The good performance of the HIT08 linelist at higher wavenumbers (e.g. the improved consistency of TCCON retrieved CO2 VSFs) might be partly because a Voigt line-shape was used in its derivation, the same as was used in this study, whereas in recent HITRAN versions more complex lineshapes were employed. Long et al. (2011) showed that differences of up to 2% in CO_2 line area could arise at 70 mbar due to inconsistent lineshape assumptions (see fig on right). At higher pressures, collisional broadening completely dominates and so the impact diminishes to 0.5% at 340 mbar.

So for ground-based measurements of the total column, this should cause errors of less than 0.5%. For retrieval algorithms that attempt to retrieve atmospheric profile information, however, this lineshape inconsistency would be much more important.

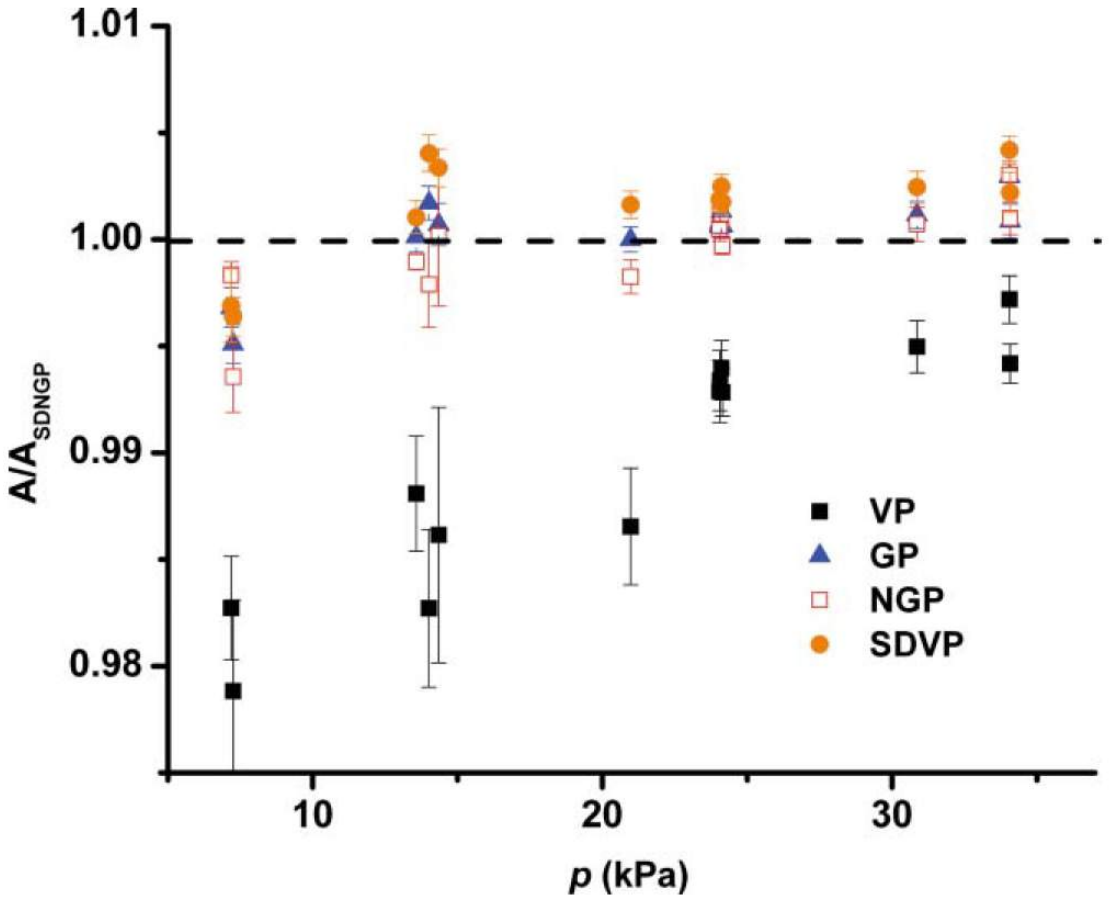


FIG. 3. Pressure dependence of the ratio of the spectrum area, A, obtained by fitting various line profiles (VP, GP, NGP, SDVP) to that obtained with the SDNGP (denoted by A_{SDNGP}). These data correspond to the air-broadened

Retrieved CO_2 VSFs: All 4 datasets

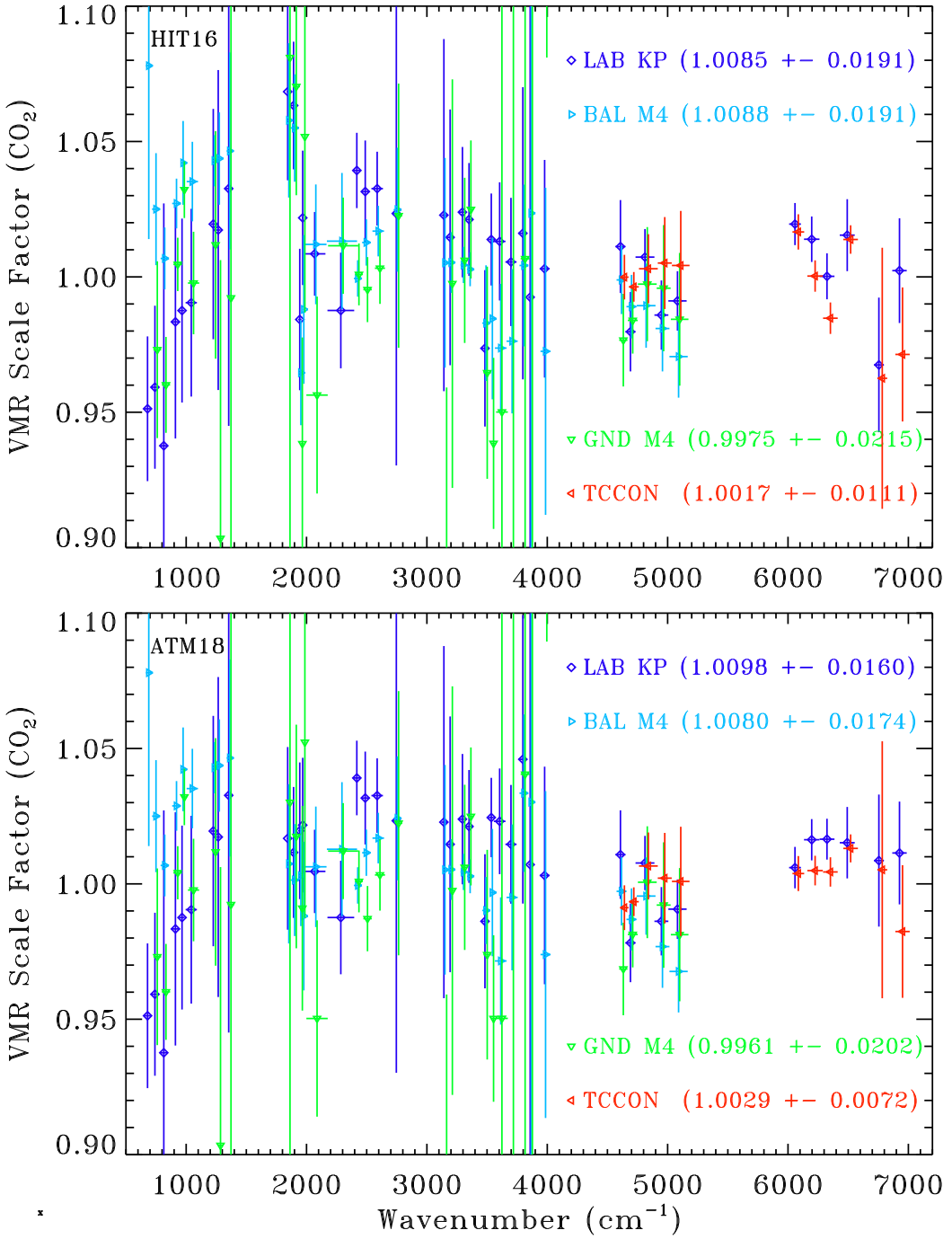
Plot show CO_2 VSFs for the HIT16b linelist (top panel) and the ATM18 linelist (bottom panel). The data are color-coded by the measurement data type (not the linelist). The ATM18 has better window-to-window consistency for all four measurement types, due mainly to the explicit adjustments made in the 1800-1993 cm^{-1} and 6720-6800 cm^{-1} regions.

In general, the error bars overlap between the four different datasets, in terms of the bias in the retrieved CO_2 amounts.

Lab data cover the entire wavenumber range. The M4_GND and M4_BAL datasets cover 700 to 5500 cm^{-1}. The TCCON covers 4000+ cm^{-1}. Only in the 4000 to 5600 cm^{-1} interval do all four datasets overlap.

MkIV GND VSFs are generally lower than those from the other datasets. Below 1200 cm^{-1} the lab data and MkIV GND show VSFs below 1.0, but the MkIV BAL shows values above 1.0 and with smaller error bars.

MkIV ground also have the worst window-to window consistency (2.02%). TCCON has the best (0.72%) but only covers well-behaved, unsaturated windows.



Summary & Conclusions

Four spectral datasets (Kitt Peak lab, MkIV balloon, MKIV ground, and TCCON ground) have been used to evaluate six different CO_2 linelists (HIT08/12/16a,b and ATM 16/18) over 670 to 7000 cm^{-1}.

Spectral fitting was performed with the GFIT code using a Voigt lineshape. The linelists were evaluated in terms of: the rms fitting residuals; and the window-to-window consistency of the retrieved gas amounts. There was no analysis of separate isotopologs. They were all lumped together as CO_2 which makes it important to know the fractionation. Analyzing the twelve CO_2 isotopologs separately is beyond the scope of this work.

RMS Spectral Fitting Residuals

Results show progressive overall RMS fit improvements in each HITRAN version, but there have been some regions where the HITRAN 2016 fits have regressed. For example, in the 4825 cm^{-1} window used by OCO-2 and GOSAT, HIT16 produces the worst fits to lab and MkIV balloon spectra (low-P) but the best fits to ground-based spectra (high-P), suggesting the positions and/or relative intensities in HIT16 are worse than predecessors, but that the widths/shifts are better.

Window-to-Window Consistency of Retrieved CO_2 Amounts

Retrieved CO_2 in the 1900 cm^{-1} region with HITRAN 2016 is biased 5% larger than in the 2050 cm^{-1} region, as pointed out by Chris Boone from ACE data. In previous HITRAN version the 1900 cm^{-1} window produced no significant bias. This problem was fixed for ATM18.

In the 6200-6400 cm^{-1} region used by TCCON the consistency of the retrievals between the 6220 and 6338 cm^{-1} bands has degraded from better than 0.1% to 1.5%. This is a serious problem because TCCON performs an weighted average of the CO_2 retrieved from these two windows. With the existence of a bias, anything that affects the uncertainties of one window relative to the other will perturb the weighted average.

ATM18 CO_2 Linelist

A new linelist (ATM18) was generated, based primarily on HITRAN 2016b, except for:

- Replacing the 3419 - 3923 cm^{-1} and 5750 - 6598 cm^{-1} sections with ATM16.
- Replacing the 6715 - 7000 cm^{-1} section with HIT08
- For isotopologues 10,11,12, using HIT16b throughout.
- Scaling all CO_2 line intensities in the 1800 to 1993 cm^{-1} interval by 1.05
- Scaling all $^{12}CO_2$ line intensities in the 6720 to 6800 cm^{-1} interval by 0.96
- Scaling all $^{12}CO_2$ line widths by 0.99 over 960 to 1000 cm^{-1}

On top of this, ad hoc corrections (mainly position adjustments) were applied, where beneficial.

There are very few windows where the ATM18 linelist doesn't produce the best (or equal best) RMS fits. In all four datasets, the ATM18 linelist produces the best average rms fits. In 2/4 datasets the ATM18 linelist produces the best window-to-window consistency in retrieved CO_2 amount, the exceptions being that the **HIT08** linelist produces the best consistency for the MkIV balloon and TCCON ground datasets.

The main weakness of this evaluation is that there were very few low-temperature lab measurements. So the selection of predecessors lines for inclusion into ATM18 might be different with more low-T lab spectra. In future, obtain additional low-T lab spectra (air-broadened) to test T-dependence of ABHW.

I recommend the ATM18 linelist for use by the NDACC and TCCON FTIR networks.

Supplemental Material: HITRAN Paper (Gordon et al., 2017)

2.2. CO_2 (molecule 2)

Accurate and comprehensive line lists for all naturally abundant isotopologues of carbon dioxide are required by remote-sensing missions dedicated to monitor the concentration of carbon dioxide in Earth's atmosphere. The recently launched OCO-2 mission [64–66], together with several other space and ground based projects (GOSAT [67], AIRS [68], ASCENDS [69], TCCON [32], NDACC [70]) are dedicated to explicitly monitor the atmospheric

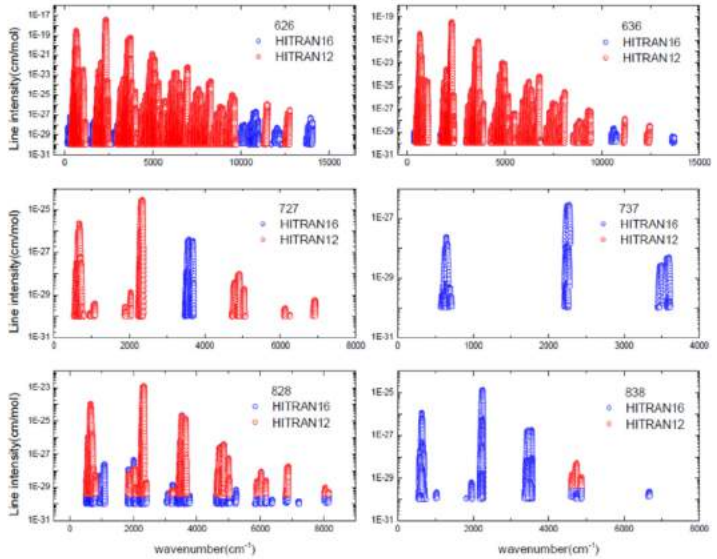


Fig. 5. Overview of the line lists of stable symmetric isotopologues of carbon dioxide in HITRAN2012 and HITRAN2016.

CO_2 content. These experiments aim not only to look at overall CO_2 concentration and its variation, but also wish to pinpoint where CO_2 is being produced (sources) and where it is absorbed (sinks). This activity is clearly vital to monitoring and essential for eventually controlling the CO_2 content of the atmosphere [71]. A successful retrieval of CO_2 concentration requires validated line lists with transition intensities given at sub-percent accuracy, line positions accurate to 0.0001 cm^{-1} or better, and beyond-Voigt-profile line-shape models [65,72,73].

Determination of isotopic ratios of carbon in Earth's samples and astrophysical objects remains crucial for modeling geophysical processes. For example, quantification of ^{14}C in fossil fuels can provide information about the sources of human-related contribution to the total CO_2 concentration in the terrestrial atmosphere. This can be done with recently developed cavity-enhanced laser spectroscopic techniques in the IR [74–76]. However, these measurements require a priori simultaneous knowledge of reliable line intensities of many isotopologues. Precise determination of $^{13}C/^{12}C$ and $^{16}O/^{17}O/^{18}O$ ratios is also vital, for instance, in understanding processes of formation of radiation fields in the Martian atmosphere, which is 96% rich in carbon dioxide [77].

A summary of the carbon dioxide line list in the HITRAN2012 database and comparison to HITRAN2016 is given in Table 4. The HITRAN2012 database was considerably improved with respect to its previous 2008 edition. However, several issues related to spectral completeness, inconsistency of multiple data sources, and insufficient accuracy of line intensities, still remained unsolved. The majority of entries in the 2012 version of the HITRAN database were taken from the effective Hamiltonian calculations included in the 2008 edition of the CDSD-296 database [78].

For less abundant isotopologues, obtaining high-quality experimental data is not trivial. Therefore fits of the effective Hamiltonian or the effective dipole moment [78], were based on only four major isotopologues $^{12}C^{16}O_2$, $^{13}C^{16}O_2$, $^{16}O^{12}C^{18}O$ and

Table 4
Comparison of HITRAN2016 and HITRAN2012 line lists for isotopologues of carbon dioxide.

ISO/abundance	HITRAN2012			HITRAN2016		
	Number of lines	Spectral region (cm^{-1})	Q(296 K)	Number of lines	Spectral region (cm^{-1})	Q(296 K)
626/0.984204	169,292	345.936–12,784.056	286.94	173,024	158.301–14,075.298	286.094
636/1.1057×10^{-2}	70,611	406.834–12,462.046	578.41	70,577	332.649–13,734.963	576.644
628/3.9470×10^{-3}	116,482	0.736–9557.398	609.48	127,850	1.472–12,677.181	607.713
627/7.3399×10^{-4}	72,525	0.757–9599.317	3552.70	77,941	0.757–12,726.562	3542.610
638/4.4345×10^{-5}	26,737	489.678–6744.158	1229.10	43,782	2.945–9212.609	1225.270
637/8.2462×10^{-6}	2953	583.593–6768.643	7162.90	25,175	9.086–8061.741	7140.024
828/3.9556×10^{-6}	7118	491.688–8160.439	324.21	10,522	482.813–8162.743	323.424
728/1.4718×10^{-6}	821	626.438–5046.875	3776.40	15,878	491.181–8193.172	3766.044
727/1.3685×10^{-7}	5187	535.384–6932.980	11,002.00	6518	535.383–6932.693	10,971.91
838/4.4440×10^{-8}	121	4599.239–4887.290	653.50	2916	2245.898–4750.068	652.242
837/1.653×10^{-8}	N/A	N/A	7615.20	4190	549.472–4914.496	7593.900
737/1.5375×10^{-9}	N/A	N/A	N/A	1501	575.852–3614.084	22,129.96
646/radioactive	N/A	N/A	N/A	41,610	426.445–7928.788	2033.353

Note: ISO is the AFGL shorthand notation for the isotopologue, abundance is the terrestrial value assumed by HITRAN, and Q(296) is the partition sum at 296 K.

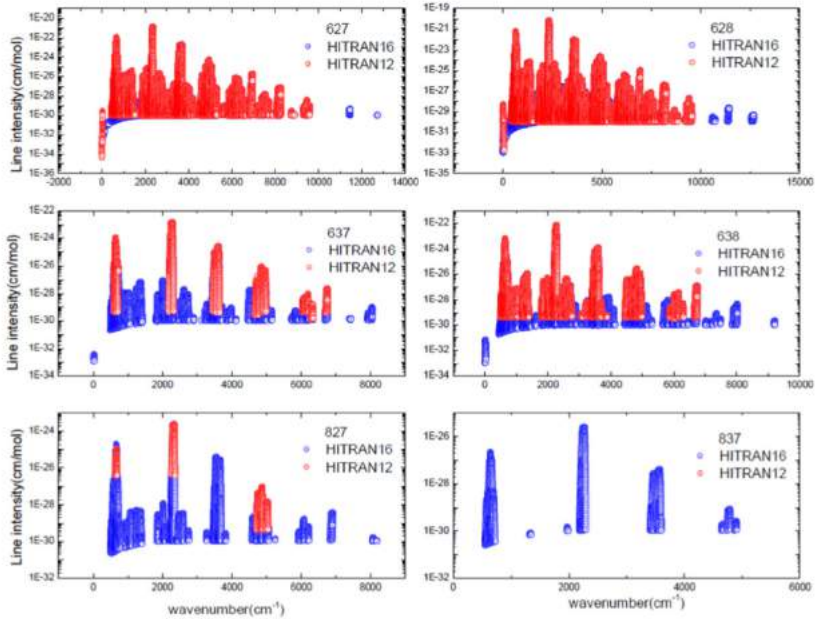


Fig. 6. Overview of the line lists of stable asymmetric isotopologues of carbon dioxide in HITRAN2012 and HITRAN2016.

$^{16}O^{12}C^{17}O$, for which measured spectroscopic parameters were available. As a result, several spectral gaps were present in HITRAN2012 (see for instance discussion in Refs. [79,80]) which represent regions where experimental data were unavailable. For similar reasons, no entries were included in the database for the $^{18}O^{13}C^{17}O$, $^{17}O^{13}C^{17}O$, and $^{16}O^{14}C^{16}O$ isotopologues (837, 737, and 646 in old AFGL notation). An overview of this problem is displayed in Figs. 5 and 6, where CO_2 ro-vibrational spectra from HITRAN2012 and HITRAN2016 are compared for different isotopologues.

Wherever possible, the effective Hamiltonian fits were extrapolated to the trace isotopologues, using a method of isotopic substitution [81]. In the 2012 edition, multiple data sources caused sporadic discontinuities in intensity patterns of ro-vibrational lines [82–84]. Furthermore, a high percentage of line intensities in HITRAN2012 have stated uncertainty of 20% or worse (HITRAN uncertainty index equal to 3). Although this assessment has been proven to be overly pessimistic in many cases [82,83,85–87], the uncertainty budget, especially for the Effective Hamiltonian calculations, was still too high for precise measurements of atmospheric CO_2 concentration. The most accurate entries in HITRAN2012 were taken from NASA JPL measurements by Toth et al. [88–90] and covered the 1.6-μm and 2.06-μm spectral regions, which are used in remote-sensing measurements. The stated 1–5% accuracy of these experimental line intensities (HITRAN uncertainty index equal to 7 and 6), has been confirmed by a number of comparisons [82,83,91]; nonetheless the rigorous requirements for part-per-million resolution in measurements of CO_2 atmospheric concentration were not achieved.

Since the 2012 release of the HITRAN database, a large number of experimental and theoretical studies have been devoted to improve the knowledge of line positions, line intensities and line shapes of CO_2 isotopologues. For a comprehensive review of measurements and theoretical models see Ref [92]. and references therein.

Theoretical line lists (denoted as "Ames") for 12 stable and one radioactive (646) isotopologue of CO_2 were published by Huang et al. in the 0–20,000 cm^{-1} spectral region and for temperatures below 1500 K [79,93,94]. Room-temperature line lists (denoted as "UCL-IAO") for 13 CO_2 isotopologues were also calculated by Zak et al. [82,83,85,91] in the 0–8000 cm^{-1} spectral region. Both of these latter studies contained intensities computed with *ab initio* dipole moment surfaces and semi-empirical line positions, based on a fitted potential energy surface for the Ames effort and on the effective Hamiltonian calculations for UCL-IAO. A major advantage of the variational approach used in the Ames and UCL-IAO line lists is that it should give similar accuracy for all isotopologues. This allows coverage of spectral regions currently not probed by experiments for rare isotopologues. UCL-IAO also provides uncertainty estimates of line intensities, based on a purely theoretical methodology [17]. Such a reliable analysis allows for the detection of ro-vibrational resonance interactions, which significantly lower the accuracy of theoretical line positions and intensities. Using this method, the lines identified as unreliable have been replaced with the data from CDSD-296 [92] and, in several cases of interpolyad resonance interactions (asymmetric isotopologues), with the experimental data from Lyulin et al. [95], Karlovets et al. [96,97] and Campargue et al. [98].

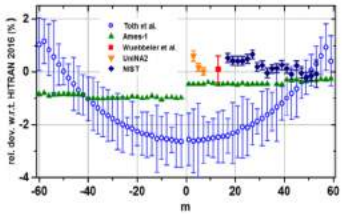


Fig. 7. Comparison of line intensities between HITRAN2016, HITRAN2012 (Toth et al. [88]), and other accurate experimental and theoretical sources for the 20012 - 00001 band (2-μm band) of $^{12}C^{16}O_2$: measurements Wübbeler et al. [102], NIST [104] and UniNA2 [103], and theory Ames-1 [79]. The zero relative deviation line corresponds to HITRAN2016 line intensities (in this case from Zak et al. [82]). The running index *m* equals -*J*, *J*, and *J* + 1 for the P, Q, and R branches, respectively.

Recently, there have been a number of high-precision near-IR spectroscopic measurements which provide rigorous tests of theoretical line intensities based on effective dipole moment surface and ab initio calculations [84–87,99]. Particularly, in the 1.6-μm and 2.0-μm spectral regions, the UCL-IAO line lists have been experimentally verified as accurate to the sub-percent level. Fig. 7 compares the UCL-IAO and Ames line lists to HITRAN2012 (Toth et al. blue open circles [100, 101]) for the 20012 – 00001 band and to state-of-the-art experiments including advanced high-resolution laser measurements [101–104]. A number of comparisons here suggest that the UCL-IAO study models line intensities more accurately than the Ames study. Note that more recent results from Ames, which are available from their website (www.huang.seti.org), give closer agreement with UCL-IAO. From Fig. 7 it is clear that there is a 1–3% average difference in line intensities between the new and the previous version of HITRAN for this band. The independent experiments from the National Institute of Standards and Technology [104], and the University of Naples II [103] confirm, however, a conservative 0.5% accuracy of line intensities for this band in HITRAN2016. This level of accuracy could potentially satisfy even the most stringent requirements of current remote-sensing missions. Interestingly, although line intensities for this band and the 20013 – 00001 and 30013 – 00001 bands probed by OCO-2 originate from the same source (Toth et al. [100]), the agreement between UCL-IAO and HITRAN2012 is substantially better for the OCO-2 bands.

For wavenumbers greater than 8000 cm^{-1}, the majority of the line parameters were taken from CDSD-296 [92]. As we have already mentioned above, both HITRAN2012 and CDSD-296 have several spectral gaps, in particular in the wavenumber region greater than 8000 cm^{-1}. Recently, several experimental studies of the carbon dioxide spectra in the high-frequency region have been performed [105–108]. The measured line intensities allowed determining the absent effective dipole moment parameters for several series of transitions. Using these effective dipole moment parameters and an effective Hamiltonian from Ref. [109], the line positions and intensities for the principal isotopologue were generated and included into HITRAN2016 covering the 9800–10,500 cm^{-1} and 11,600–12,400 cm^{-1} wavenumber gaps.

Line positions were updated with respect to the previous version of the database. The majority of lines come from the effective Hamiltonian calculations included in UCL-IAO line lists, which are based on the fits to the observed line positions collected from the literature and published in the latest, 2015 release of the CDSD-296 database [92]. These fits were completed and updated with recent, accurate measurements performed on isotopically-enriched samples of CO_2. Uncertainties in the fitted line positions depend on the quality of the experimental data and vary from 0.001 cm^{-1} to 10^{-9} cm^{-1}. For asymmetric isotopologues, a number of bands are affected by strong interpolyad anharmonic resonance interactions. The effective Hamiltonian model does not include this type of interaction for the asymmetric isotopologues. Hence in such cases, line positions were taken directly from measurements [80,95–98,110,111].

The uncertainty codes for the line positions were transferred from CDSD-296 to HITRAN2016. The uncertainty code 3 (0.001–0.01 cm^{-1}) was given for the line positions in the 9800–10,500 cm^{-1} and 11,600–12,400 cm^{-1} wavenumber regions. Partition functions in the current release of the database are based on the direct summations taken from the variational calculations of Huang et al. [79]. On average, the new partition functions agree excellently with those of CDSD-296 [83]; however they do not agree perfectly with those in HITRAN2012 (from TIPS [112]) and three previous editions of the database and differ at 296 K by about −0.3%. Although this difference is marginal, it could have an effect in the applications where sub-percent accuracy is required.

As stated above, the radioactive isotopologue $^{14}CO_2$, 646, has been added to the database. This is the first edition of HITRAN where radioactive species have been incorporated (also for CO, see Section 2.5). All lines of the 646 isotopologue were taken from the UCL line lists given in Ref. [83]. Due to issues with what constitutes a so-called natural terrestrial abundance of radioactive species (which is part of the traditional definition of intensities in HITRAN, see Eq. (1) of the Definitions and Units documentation in HITRAN*online*), line intensities for these type of species are given for unit abundance; a 10^{-27} $cm^{-1}/(molecule \cdot cm^{-2})$ cut-off value for the intensity has been applied. This cut-off produced 41,610 lines in the *J* range 0 to 114. Vibrational assignments for the 646 isotopologue were based on isotopic shifts of energy levels and respective assignments for the 626 and 636 isotopologues, and hence should be regarded as provisional. An abundance-scaled intensity cut-off of 10^{-30} $cm^{-1}/(molecule \cdot cm^{-2})$ is used for all stable isotopologues. Note that, for the time being, data for the radioactive isotopologues are provided as static files rather than through the HITRAN*online* interface.

Uncertainties of line intensities were informed by theoretical error analysis, which classified lines as reliable, intermediate, or unreliable. Bands with reliable lines stronger than 10^{-23} $cm^{-1}/(molecule \cdot cm^{-2})$ (for unit abundance) were assigned HITRAN uncertainty code 8 (i.e. accuracy of 1% or better). Line intensities of reliable parallel bands weaker than 10^{-23} $cm^{-1}/(molecule \cdot cm^{-2})$ were given an uncertainty code 7 (i.e. accuracy 1-2%). Reliable perpendicular bands weaker than 10^{-23} $cm^{-1}/(molecule \cdot cm^{-2})$ and intermediate lines were marked with HITRAN uncertainty code 6 (i.e. accuracy 2-5%). So-called unreliable lines were taken from the effective dipole moment calculations [92] and experiments. Typical intensity uncertainties for these entries range between 5 and 20%.

It is important to point out that an intensive study of the 1.6-μm and 2.06-μm bands that includes non-Voigt lineshapes and line mixing has been published by the OCO-2 spectroscopy support group ABSCO (ABSorption COefficient tables for the OCO-2 mission) [84,86]. The data were fit using a multi spectrum fit procedure which, among other things, enables retrieval of the line-shape parameters using the speed-dependent Voigt (SDV) profile as well as line mixing. These are very good experiments and it is debatable whether to use them for the strong and weak bands in place of UCL-IAO parameters described above. Indeed the ABSCO team have validated (using TCCON spectra) the cross-sections generated using results of Refs. [84,86] and found them to be the most efficient [113]. However, achieving high-precision results in nuanced correlations, with line mixing and model assumptions that can create discontinuities in inter-band comparisons, is difficult. At the moment, HITRAN cannot provide users with tools that can be used to generate cross-sections from the works of Devi et al. [84] and Benner et al. [86]. The usable products of the ABSCO effort are absorption coefficients (available upon request from the ABSCO group) rather than spectral parameters, which are available in the publications. Moreover, these absorption coefficients are empirically scaled by the factors of 0.6% and 1.4% for the 1.6-μm and 2.06-μm bands respectively, due to lingering data and/or model biases (the use of partition function HITRAN 2012/TIPS is up to 0.3% of this factor). After these studies, an update of the multi spectrum fitting code with CDSD partition functions was done. Additional methods to adjust the intensity distribution closer to the UCL list by scaling experimental conditions within the uncertainties are under evaluation. The intensities of the band at 2.06 μm are already within 0.7% of Zak et al. [82], indicating that the additional 0.7% scaling of ABSCO cross-sections may be unrelated to intensities. These issues will be considered for future editions of the database.

For wavenumbers greater than 8000 cm^{-1}, two sources of the line intensities are used: CDSD-296 [92] and the newly-generated line list in the 9800–10,500 cm^{-1} and 11,600–12,400 cm^{-1} regions based on the new measurements [105–108]. The uncertainty codes of the CDSD-296 line intensities were transferred to HITRAN2016. Based on the uncertainties of the line intensity measurements in the 10,700–10,860 cm^{-1} region [106], we use uncertainty code 5 (5%−10%) for the line intensities of the 3003i-00001 (i= 1,2,3,4) series of bands and based on the uncertainties of the line intensity measurements in the 10,000–10,300 cm^{-1} and 11,600–12,400 cm^{-1} [107] wavenumber regions we use uncertainty code 3 (> 20%) for the line intensities of the 4003i-00001 (i= 1,2,3,4,5) and of the 6001i-00001 (i= 1,2,3,4,5,6,7) series of bands.

The Voigt line-shape parameters throughout the entire database were calculated using the predictive routine of Gamache and Lamouroux explained in Refs. [114–116]. For the line mixing, we now provide a code from Lamouroux et al. [117] which has been updated to operate with HITRAN2016. We note that Lamouroux et al. [117] line mixing coupled with the HITRAN2012 data has worked really well and in fact produced residuals hardly exceeding 1% when applied to the TCCON data in Ref. [113], although slightly inferior to the ABSCO cross-sections in the 2.06-μm region.

Effect of Line Position Errors on Retrieved Gas Amounts (1/4)

Although line position errors can cause prominent anti-symmetric residuals, do they significantly affect the retrieved gas amounts? The equivalent width (EW) of the absorption line is unchanged by the position error, so a robust EW-matching retrieval algorithm should be unaffected. Sadly, the EW-matching spectral fitting algorithm is an unattainable ideal. In the real world, least-squares spectral fitting algorithms are used, which minimize the sum of the squares of the residuals, weighted by the measurement uncertainties

$$\chi^2 = \sum_i [(T^m_i - T^c_i(x))/\varepsilon_i]^2$$

where T^m_i is the measured transmission spectrum, ε_i is its uncertainty, T^c_i is the calculated transmission spectrum, and x is the VMR scale factor that we are trying to determine. Subscript i represents different spectral points. To minimize the residuals, differentiate wrt x and set to zero yielding

$$\sum_i J_i(T^m_i - T^c_i(x))/\varepsilon_i^2 = 0$$

where $J_i = \partial T^c_i(x)/\partial x$ is the Jacobian vector. If the measurement noise is white, as is usually the case in FTIR spectra, then the ε_i term can be dropped. The quantity $\sum_i [(T^m_i - T^c_i(x)]$ is the difference in the EWs of the measured and calculated absorption lines, so a robust EW matching retrieval algorithm would set this to zero. But least squares fitting does not do this. Instead, it zeros $\sum_i J_i^T[T^m_i - T^c_i(x)]$. Here the residuals, $R_i = T^m_i - T^c_i(x)$, are multiplied by the Jacobian, J_i, so EW is not matched. J^T is symmetric about the **line centers** in the calculated spectrum, whereas the residuals, are anti-symmetric about the **mid-point** of the measured and calculated line positions. So if there is a position error, the dot product of J_i^T and R_i won't be zero, and so x will be adjusted to an incorrect value in order to set $\sum_i J_i^T[T^m_i - T^c_i(x)]$ to zero. The adjustment to x is Δx, which is given by the unconstrained least-squares equation

$$J^T J \Delta x = J^T R$$

Effect of Line Position Errors on Retrieved Gas Amounts (2/4)

The measured transmittance of a single isolated line of strength s_m in a cell path containing c_m molecules.cm^{-2} is

$$T^m_i = \mathrm{Exp}[-s_m.c_m.f(\nu_i-\nu_m)]$$

where ν_m is the true line center position, s_m is the true line intensity, c_m is the true absorber amount, and f is the unit-area lineshape function. Measurement noise has been ignored. We try to model this absorption line with the equation

$$T^c_i = \mathrm{Exp}[-x.s_c.c_c.f(\nu_i-\nu_c)]$$

where s_c and ν_c are the assumed line intensity and position (from HITRAN), x is a dimensionless scaling factor that adjusts for any error in the assumed line intensities (if we are analyzing lab spectra with c assumed perfectly known) or in the a priori absorber slant column (if we are analyzing atmospheric spectra with the s assumed perfect).

The residuals are $R = T^m_i - T^c_i = \mathrm{Exp}[-s_m.c_m.f(\nu_i-\nu_m)] - \mathrm{Exp}[-x.s_c.c_c.f(\nu_i-\nu_c)]$

The Jacobian is $J^T = \partial T^c_i(x)/\partial x = -s_c.c_c.f(\nu_i-\nu_c).T^c_i$

If the only error is the line positions being wrong by δ , then $s_c=s_m$ and $c_m=c_c$, x=1, but $v_c=v_0+\delta/2$, $v_m=v_0-\delta/2$, then

$$R_i = \mathrm{Exp}[-s_m.c_m.f(\nu_i-\nu_0+\delta/2)] - \mathrm{Exp}[-s_m.c_m.f(\nu_i-\nu_0-\delta/2)]$$

For small $\delta/2$, $f(\nu_i-\nu_0\pm\delta/2) = f(\nu_i-\nu_0) \pm \delta/2.(\partial f/\partial \nu)$ where $\partial f/\partial \nu$ is evaluated at $v=v_0$

$$R_i = \mathrm{Exp}[-s_m.c_m.f(\nu_i-\nu_0)] \cdot [\mathrm{Exp}[\delta/2.s_m.c_m.(\partial f/\partial_v)] - \mathrm{Exp}[-\delta/2.s_m.c_m.(\partial f/\partial_v)]]$$

If $\delta/2.s_m.c_m.(\partial f/\partial_v) << 1$, i.e., small position errors or weak lines, we can approximate

$$R_i \cong \mathrm{Exp}[-s_m.c_m.f(\nu_i-\nu_0)].\delta.s_m.c_m.(\partial f/\partial \nu)$$

Effect of Position Errors on Retrievals: Gaussian case (3/4)

For many line shape functions (e.g. Gaussian), analytic expressions can be developed for T^C, R, J, and hence Δx. For a Gaussian lineshape $f(v_i\text{-}v_0) = w^{-1}\pi^{-½}\,\mathrm{Exp}[-((v_i\text{-}v_0)/w)^2]$ where w is the line width (half-width at 1/e)

So $\partial f/\partial v_i = w^{-2}\pi^{-½}\ \text{-}2(v_i\text{-}v_0)\cdot\mathrm{Exp}[] = \text{-}2w^{-1}(v_i\text{-}v_0)\cdot f(v_i\text{-}v_0)$

$$R_i \cong T^0{}_i\cdot\delta\cdot s_m\cdot c_m\cdot(\partial f/\partial v) \cong \text{-}2\delta\cdot s_m\cdot c_m\cdot w^{-1}\cdot(v_i\text{-}v_0)\cdot f(v_i\text{-}v_0)\cdot T^0{}_i$$

So the residuals are anti-symmetrical about v_o

Since $J^T = \text{-}s_c.c_c.f.T^c{}_i$, the retrieved Δx value that results from the misplaced line center is

$$\Delta x = (J^TJ)^{-1}\,J^TR = \sum_i \text{-}s_c.c_c.f(v_i\text{-}v_c).T^c{}_i\,R_i \,/\, \sum_i s_c{}^2.c_c{}^2.f^2(v_i\text{-}v_c).T^c{}_i{}^2$$

$$\Delta x = (J^TJ)^{-1}\,J^TR = \sum_i \text{-}s_c.c_c.f(v_i\text{-}v_c).\text{-}2\delta.s_m.c_m.(v_i\text{-}v_0)/w.f(v_i\text{-}v_0).T^c{}_i\,.T^0{}_i \,/\, \sum_i s_c{}^2.c_c{}^2.f^2(v_i\text{-}v_0).T^c{}_i{}^2$$

$$\Delta x = (J^TJ)^{-1}\,J^TR = \text{-}2\delta/w\sum_i f(v_i\text{-}v_c).(v_i\text{-}v_0).f(v_i\text{-}v_0).T^c{}_i\,.T^0{}_i \,/\, \sum_i f^2(v_i\text{-}v_0).T^c{}_i{}^2$$

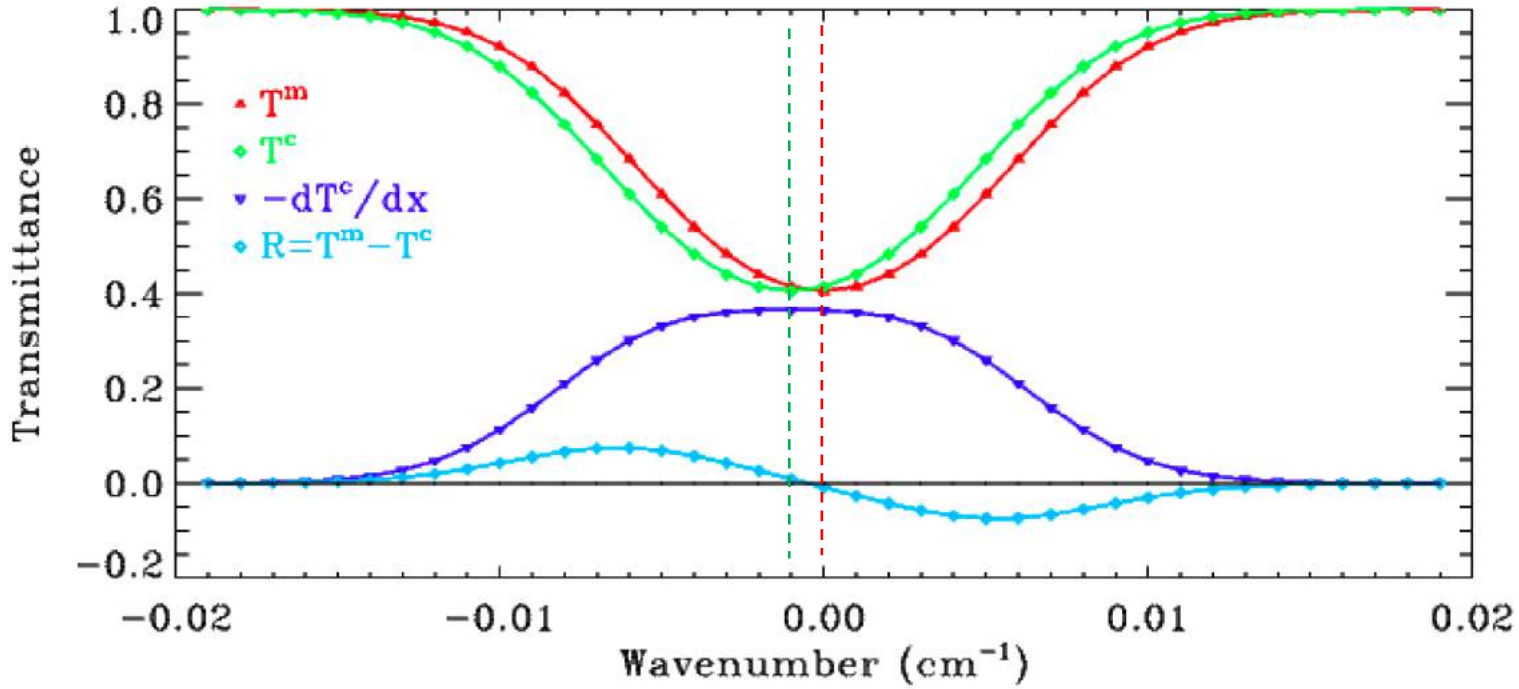
Assuming that $f(v_i\text{-}v_0)$ is very similar $f(v_i\text{-}v_0)$ and hence . So Δx is the weighted average value of $2\delta(v_i\text{-}v_0)/w$ over the spectral points. The weights being $f^2(v_i\text{-}v_c).T^c{}_i{}^2$, which peak at the calculated line center v_c. *If the line is weak, such that T_i=1, then the weights are f2 and so the integrals can be performed analytically to yield*

$$\Delta x = (J^TJ)^{-1}\,J^TR = \text{-}\,2^{-½}\,(\delta/w)^2$$

Similar expressions can be derived for the Lorentzian lineshape.

Effect of Position Errors on Retrievals: Gaussian case (4/4)

Red curve shows a measured transmittance spectrum T^m centered at 0 cm^{-1}. A Gaussian line shape is assumed with a HWHM of 0.005 cm^{-1}, corresponding to a stratospheric CO_2 line at 4850 cm^{-1}. The green trace (T^c) shows a computed spectrum with everything correct (strength, width, absorber amount) except the line position, which is off by 0.001 cm^{-1} (1/5 of the HWHM). The cyan trace shows the fitting residuals which are anti-symmetric about (i.e. cross zero at) the mid-point of the line centers. The blue trace shows the Jacobian (dT^c/dx) which has been inverted to better fit on the figure. The Jacobian peaks at, and is symmetrical about, the calculated line center (green dash line).



The numerically computed Δx value corresponding to this plot is -1.2% so the retrieved gas amount would be under-estimated by this amount. This is within a factor 2 of the $2^{-½}(\delta/w)^2$ = 2.9% predicted from the equation on the previous slide for the weak line limit. If the position error were doubled to 0.002 cm^{-1}, Δx would quadruple to 4.8%.

Of course, most instruments won't fully resolve a doppler line (MKIV has a resolution of 0.010 cm^{-1}) and will therefore impart additional broadening, possibly reducing the impact of position errors (haven't tested this).