

H₂O Spectroscopy Evaluation 700 to 12,000 cm⁻¹

Geoffrey C. Toon
Jet Propulsion Laboratory
California Institute of Technology
2020-05-07

The goal of this work was to evaluate existing H₂O linelists using measurements (lab & solar) acquired under conditions typical of the Earth's atmosphere. This was achieved by fitting laboratory and atmospheric spectra.

During the course of this it became clear that there wasn't a single linelist that was better than the others in all circumstances and spectral regions. Although the HITRAN linelists generally got much better with each new edition, there were always a few cases/regions in which older linelists was better ("Two steps forward, one step backward").

Attempting to capture the improvements without the lapses, we merged the best aspects of the existing linelists to create a new "cherry picked" linelist. We then "repaired" any obvious spectroscopic problems that remained in the cherry-picking, such as position, intensity, width, and shift errors,

So this report fulfills two functions: Firstly, to evaluate the linelists that existed in 2018. Secondly to report the creation of a "greatest hits" linelist that captured the best aspects of the predecessor linelists. For the sake of brevity and concision, these two function were merged into the one report, so that the new "greatest hits" linelist (ATM18) evaluation is shown in parallel with that of the earlier linelists, even though it happened 2 years later.

In this report H₂O, HDO and D₂O are considered to be different gases; they are evaluated separately. So when the term H₂O is used, we refer to the first three isotopologs of water vapor, and exclude HDO and D₂O.

HITRAN 2016 paper: Water Vapor Section: Introduction

“The HITRAN2016 edition has undergone a substantial revision and expansion of the database of water vapor. In HITRAN2012 a very large expansion of the dynamic range of the line intensities for **non-deuterated** isotopologues of water vapor was implemented, thanks to the ab initio calculations from the BT2 line list [15] and, in selected spectral intervals, Lodi et al. [16] for the principal isotopologue and from Lodi and Tennyson for H_2^{18}O and H_2^{17}O [17]. The ab initio results were replaced with high-quality experimental or semi-empirical data wherever possible. In HITRAN2016, we have done a similar expansion of the dynamic (and spectral) range of the **singly-deuterated** isotopologues.”

[15] Barber RJ, Tennyson J, Harris GJ, Tolchenov RN. A high-accuracy computed water line list. Mon Not R Astron Soc 2006;368:1087–94. <http://dx.doi.org/10.1111/j.1365-2966.2006.10184.x>.

[16] Lodi L, Tennyson J, Polyansky OL. A global, high accuracy ab initio dipole moment surface for the electronic ground state of the water molecule. J Chem Phys 2011;135:34113. <http://dx.doi.org/10.1063/1.3604934>.

[17] Lodi L, Tennyson J. Line lists for H_2^{18}O and H_2^{17}O based on empirical line positions and ab initio intensities. J Quant Spectrosc Radiat. Transf. 2012;113:850–8. <http://dx.doi.org/10.1016/j.jqsrt.2012.02.023>.

[18] Krasnopolsky VA, Belyaev DA, Gordon IE, Li G, Rothman LS. Observations of D/ H ratios in H_2O , HCl , and HF on Venus and new DCI and DF line strengths. Icarus 2013;224:57–65. <http://dx.doi.org/10.1016/j.icarus.2013.02.010>



Contents lists available at ScienceDirect

Journal of Quantitative Spectroscopy & Radiative Transfer

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

The HITRAN2016 molecular spectroscopic database

I.E. Gordon^{a,*}, L.S. Rothman^a, C. Hill^{a,b}, R.V. Kochanov^{a,c}, Y. Tan^a, P.F. Bernath^d, M. Birk^e, V. Boudon^f, A. Campargue^g, K.V. Chance^a, B.J. Drouin^h, J.-M. Flaudⁱ, R.R. Gamache^j, J.T. Hodges^k, D. Jacquemart^l, V.I. Perevalov^m, A. Perrinⁿ, K.P. Shine^o, M.-A.H. Smith^p, J. Tennyson^b, G.C. Toon^h, H. Tranⁿ, V.G. Tyuterev^q, A. Barbe^q, A.G. Császár^{r,rr}, V.M. Devi^s, T. Furtenbacher^r, J.J. Harrison^{t,tt,ttt}, J.-M. Hartmannⁿ, A. Jollyⁱ, T.J. Johnson^u, T. Karman^{a,v}, I. Kleinerⁱ, A.A. Kyuberis^{a,w}, J. Loos^e, O.M. Lyulin^m, S.T. Massie^x, S.N. Mikhailenko^m, N. Moazzen-Ahmadi^y, H.S.P. Müller^z, O.V. Naumenko^m, A.V. Nikitin^m, O.L. Polyansky^{b,w}, M. Rey^q, M. Rotger^q, S.W. Sharpe^u, K. Sung^h, E. Starikova^m, S.A. Tashkun^m, J. Vander Auwera^{aa}, G. Wagner^e, J. Wilzewski^{a,e}, P. Wcisło^{bb}, S. Yu^h, E.J. Zak^b

Atmospheric Water Vapor Abundances

Water vapor is extremely variable in the Earth's troposphere. From a mole fraction of 3 ppm at the tropical tropopause to 4% in the tropical PBL (4+ orders of magnitude). On top of that, the atmospheric pressure decreases by a factor 10 between the surface and the tropical tropopause. So a 5 order of magnitude range of $[H_2O]$. More for $[HDO]$ due to its strong fractionation.

Even for ground-based observations, the difference in H_2O slant column between a humid high-airmass spectrum and a dry low-airmass spectrum can still be 2-3 orders of magnitude.

Retrievals are most accurate when the absorption features are around unit optical depth. Weaker than this and random noise increasingly dominates; stronger than this and systematic errors dominate (e.g. zero level offset, line-shape). So in retrieving atmospheric H_2O , multiple windows with a large range of strengths are needed to accommodate its huge dynamic range. Hence the need for accurate spectroscopy with good line-to-line and window-to-window consistency.

Water vapor absorption is ubiquitous throughout the infra-red, making it a major interferent in retrievals of other gases (CO_2 , CO , O_3 , N_2O , CH_4 , NO , NO_2 , HCl , HF , etc.) for the purposes of NDACC, TCCON, ACE, MkIV, etc. Thus H_2O spectroscopy must also be good in the windows where we retrieve these other gases, not just in the windows where we retrieve H_2O itself.

Path Length and Slant Column Considerations

For open paths through the Earth's atmosphere, ground-based path lengths range from one scale height (~ 8 km) when zenith viewing to 35 scale heights (280 km) when SZA=90. Balloon-borne limb path lengths can reach 500 km

For laboratory FTIR spectra, the longest path lengths that I have encountered are only 0.4 km, which is 20-1000 times shorter than in the atmosphere. (Cavity Ring Down techniques can achieve much longer path lengths, but with caveats).

For gases whose atmospheric vmrs are < 0.001 (e.g. CO_2 , N_2O , CH_4 , CO) the absorption depths achievable in lab spectra can be made similar to those in atmospheric spectra by increasing the lab vmrs by a few orders of magnitude as compared with the atmosphere. But this is not an option for H_2O (or O_2 or N_2). For H_2O the attainable lab vmrs are severely limited by its low vapor pressure (condensation).

So lab H_2O slant column amounts do not get close to atmosphere values (although HDO and D_2O can by use of isotopically-enriched samples). So H_2O lines that appear fairly strong in high-airmass atmospheric spectra will not be discernable in lab spectra. Their spectroscopy, if present, will therefore be theoretically-based.

The maximum ocean temperatures on Earth are about 30C, at which the saturated vapor pressure of H_2O is 31 Torr. This is 4% of the Earth's sea-level pressure and represents the highest absolute humidity that might naturally be encountered.

In the Earth's troposphere, water vapor has a scale height of about 2 km in the daytime (1 km at night), which is much less than the 8 km density scale height. So when H_2O is 4% of the molecules at the surface, in terms of the vertical column abundance H_2O will only be 1% of the total. The vertical column of all molecules is $2.15\text{E}+25$ molecules. cm^{-2} . The largest conceivable H_2O vertical column is 1% of this or $2.15\text{E}+23$. In a horizontal path the H_2O slant column will be larger by a factor $\text{Sqrt}(\pi R/H/2)$ which for H_2O ($H=2$ km) is 70x larger = $1.5\text{E}+25$ molecules. cm^{-2} . The largest that TCCON has ever measured (sunny conditions) is $3\text{E}+24$ molecules. cm^{-2} , which is 5 times less than worst case limit. In such a path, lines that are 1.5% deep (clearly above noise level) and 0.1 cm^{-1} wide, will have an intensity of $0.015 \times 0.1 / 3\text{E}+24 = 5\text{E}-28\text{ cm}^{-1}/(\text{molec}.\text{cm}^{-2})$.

In the lab, the H_2O partial pressures are similar to those in the atmosphere i.e. 30 Torr. But instead of having an effective path of 140 km, it is only 0.4 km. So lab spectra do not come close to the H_2O amounts seen in the atmosphere. So H_2O lines weaker than $5\text{E}-27$ cannot be validated with lab spectra, at least the ones that I have. Thus, atmospheric spectra are needed to validate the spectroscopy of weaker H_2O lines.

The H₂O Linelists Evaluated

H₂O is represented by isotopologs 1-3, HDO by isotopologs 4-6, and D₂O by isotopologs 7-9.

D₂O cannot be seen in the Earth's atmosphere. The main purpose of the D₂O linelist is to facilitate analysis of highly D-enriched lab HDO spectra, in which D₂O absorptions can be strong. HIT16 includes a linelist for D₂¹⁶O (isotopolog #7) for the first time. Bob Toth had one in 2006 that never got into HITRAN, but was included in ATM16.

Whereas the HDO linelist in HIT08 and HIT12 extends to 22,708 cm⁻¹, in HIT16 it extends only to 19,935 cm⁻¹, despite containing 4x more lines.

The ATM16 H₂O is founded on Toth (2003), except for the 0-600 and 8100+ cm⁻¹ regions which are from HIT12. Many ad hoc empirical improvements have been made.

ATM18 is based on the HDO and D₂O from HIT16, except in the TCCON windows where the ATM16 HDO was retained. ATM18 H₂O is founded on ATM16

HIT08: 69,201 lines

54,177 H₂O lines covering 0 to 25,232 cm⁻¹

15,024 HDO lines covering 0 to 22,708 cm⁻¹

HIT12: 224,515 lines

209,492 H₂O lines covering 0 to 25,710 cm⁻¹

15,023 HDO lines covering 0 to 22,708 cm⁻¹

ATM16: 169,134 lines

147,636 H₂O lines covering 0 to 25,711 cm⁻¹

18,713 HDO lines covering 0 to 22,708 cm⁻¹

2,785 D₂O lines covering 2198 to 4255 cm⁻¹

HIT16: 304,225 lines

207,277 H₂O lines covering 0 to 25,711 cm⁻¹

73,460 HDO lines covering 0 to 19,935 cm⁻¹

23,488 D₂O lines covering 0 to 12,797 cm⁻¹

ATM18: 244,136 lines

148,486 H₂O lines covering 0 to 25,711 cm⁻¹

71,424 HDO lines covering 0 to 19,935 cm⁻¹

24,226 D₂O lines covering 0 to 12,797 cm⁻¹

The Fitted Spectra

Laboratory (154 spectra covering 650-11,000 cm^{-1})

140 from Kitt Peak (1983-1996)

- 58 of which are D-enriched
- 7 of which are ^{18}O -enriched
- 3 of which are ^{17}O -enriched

11 from Manfred Birk (2014-2015)

3 from Keeyoon Sung measured by the Bruker 125HR at JPL in 2008 (B0028 & B0030 series)

MkIV Balloon-borne (650-5650 cm^{-1})

One occultation of 34 spectral pairs (HgCdTe & InSb) covering 9 to 38 km altitude

MkIV Ground-based (650-5650 cm^{-1})

Subset of 121 spectral pairs (HgCdTe & InSb) covering 10 to 89° SZA and 0 to 3.8 km altitude

TCCON Ground-based (4000-15,500 cm^{-1})

26 spectral pairs (InGaAs & Si)

101 Fitted Lab Windows

Center	Width	MIT	A	I	F	Parameters to fit			Gases to fit							
709.00	54.1	15	2	1	0	ncbf=10	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
768.00	64.1	20	2	1	0	ncbf=10	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
831.50	59.1	20	2	1	0	ncbf=20	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
897.40	55.7	15	2	1	0	ncbf=11	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
961.20	69.8	15	2	1	0	ncbf=11	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
1033.10	74.1	15	2	1	0	ncbf=20	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
1103.00	66.0	15	2	1	0	ncbf=17	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
1182.15	90.3	15	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
1265.00	75.1	15	2	1	0	ncbf=18	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
1344.00	80.0	15	2	1	0	ncbf=14	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
1412.10	54.5	15	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
1466.60	54.1	15	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
1520.60	54.2	15	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
1573.60	52.0	15	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
1661.70	127.6	15	2	1	0	ncbf=16	fs	zo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
1762.79	74.0	15	2	1	0	ncbf=13	fs	zo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
1812.50	25.0	15	2	1	0	ncbf=11	fs	zo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
1837.50	25.0	15	2	1	0	ncbf=11	fs	zo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
1875.00	49.5	15	2	1	0	ncbf=11	fs	zo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
1936.00	72.0	15	2	1	0	ncbf=10	fs	zo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
2008.10	69.0	15	2	1	0	ncbf=10	fs	zo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
2094.50	99.0	15	2	1	0	ncbf=13	fs	zo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
2184.00	80.8	15	2	1	0	ncbf=11	fs	zo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
2312.60	176.7	15	2	1	0	ncbf= 9	fs	zo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
2447.50	95.2	15	2	1	0	ncbf=13	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
2541.00	91.2	15	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
2635.50	97.0	15	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
2729.00	90.0	15	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
2822.00	96.0	15	2	1	0	ncbf=13	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
2910.00	80.0	15	2	1	0	ncbf=14	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
2999.00	95.0	15	2	1	0	ncbf=13	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
3092.50	92.0	15	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
3181.00	86.0	15	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
3261.50	75.0	15	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
3343.70	89.4	15	2	1	0	ncbf=13	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
3433.40	90.0	15	2	1	0	ncbf=13	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
3517.40	78.0	15	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
3598.45	84.1	15	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
3707.85	134.8	18	2	1	0	ncbf=16	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
3843.50	136.5	18	2	1	0	ncbf=16	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
3956.00	88.5	18	2	1	0	ncbf=11	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
4041.90	82.7	18	2	1	0	ncbf=13	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
4122.75	79.0	18	2	1	0	ncbf=11	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
4212.00	100.0	18	2	1	0	ncbf=13	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
4299.00	74.0	18	2	1	0	ncbf=11	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
4398.00	124.0	18	2	1	0	ncbf=15	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
4500.00	80.0	18	2	1	0	ncbf=11	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
4570.50	61.0	18	2	1	0	ncbf=10	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
4643.50	85.0	18	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
4738.50	105.0	18	2	1	0	ncbf=13	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
4838.50	95.0	18	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
4932.50	93.0	18	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
5029.00	100.0	16	2	1	0	ncbf=13	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
5113.50	69.0	14	2	1	0	ncbf=10	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
5180.40	64.8	14	2	1	0	ncbf= 9	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
5257.20	88.8	16	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
5345.50	88.0	16	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
5436.40	93.8	16	2	1	0	ncbf=13	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
5535.50	105.0	16	2	1	0	ncbf=14	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
5636.00	96.0	16	2	1	0	ncbf=14	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
5732.00	96.0	16	2	1	0	ncbf=14	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
5827.00	94.0	16	2	1	0	ncbf=13	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
5935.00	122.0	16	2	1	0	ncbf=15	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
6051.00	110.0	16	2	1	0	ncbf=14	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
6168.00	124.0	16	2	1	0	ncbf=15	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
6285.00	110.0	16	2	1	0	ncbf=14	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
6380.00	80.0	16	2	1	0	ncbf=11	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
6470.00	100.0	16	2	1	0	ncbf=13	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
6579.00	118.0	16	2	1	0	ncbf=15	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
6703.00	130.0	16	2	1	0	ncbf=16	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
6826.00	116.0	16	2	1	0	ncbf=15	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
6937.50	107.0	16	2	1	0	ncbf=13	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
7039.50	97.0	16	2	1	0	ncbf=13	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
7137.75	99.5	16	2	1	0	ncbf=13	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
7228.50	82.0	16	2	1	0	ncbf=11	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
7318.40	97.6	16	2	1	0	ncbf=14	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
7424.30	114.2	16	2	1	0	ncbf=14	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
7514.80	66.8	16	2	1	0	ncbf=10	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
7586.10	75.8	16	2	1	0	ncbf=11	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
7661.50	75.0	16	2	1	0	ncbf=11	fs	xo	xf	:	h2o	hdo	d2o	co2	nh3	hcn
7760.00	122.0	16	2	1	0	ncbf=15	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
7878.00	114.0	16	2	1	0	ncbf=14	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
7990.50	111.0	16	2	1	0	ncbf=14	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
8118.00	144.0	20	2	1	0	ncbf=14	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
8225.00	70.0	20	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
8303.50	87.0	20	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
8398.50	103.0	20	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
8524.00	148.0	20	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
8653.00	110.0	20	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
8751.50	87.0	20	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
8858.00	126.0	20	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
8977.00	112.0	20	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
9111.50	157.0	20	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
9252.00	124.0	20	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
9387.00	146.0	20	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
9530.00	140.0	20	2	1	0	ncbf=12	fs	xo	cf	:	h2o	hdo	d2o	co2	nh3	hcn
9720.00	240.0	20	2	1	0</											

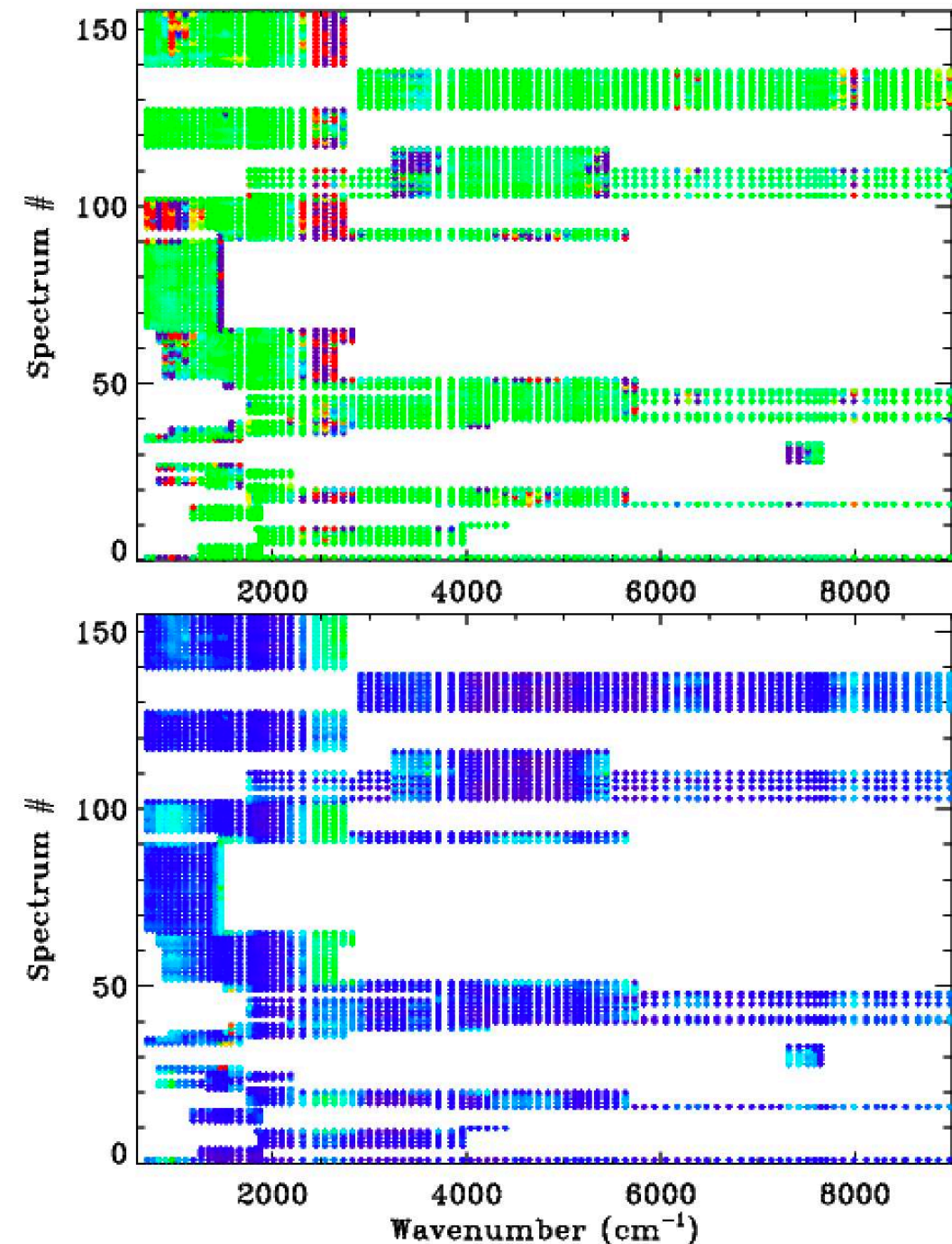
Retrieved H₂O VMR Scale Factors (lab)

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 H₂O 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 #. Each point represents a spectral fit to a ~100 cm⁻¹ wide window. Spectra 1-3 are from Keeyoon Sung, spectra 4-14 are from Manfred Birk, and the remainder from Kitt Peak. Most spectra cover less than 2000 cm⁻¹. Consequently only 29% of the 154x100=15400 potential spectral fits could be performed.

Lower Panel. The H₂O VSF uncertainties are color-coded according to uncertainty (purple=0.2%; blue=1%; green=10%; red=100%) and are plotted versus the window center wavenumber (x-axis) and the spectrum # (y-axis). These uncertainties are based on the absorption depths of the H₂O lines and the fitting residuals.

On 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.

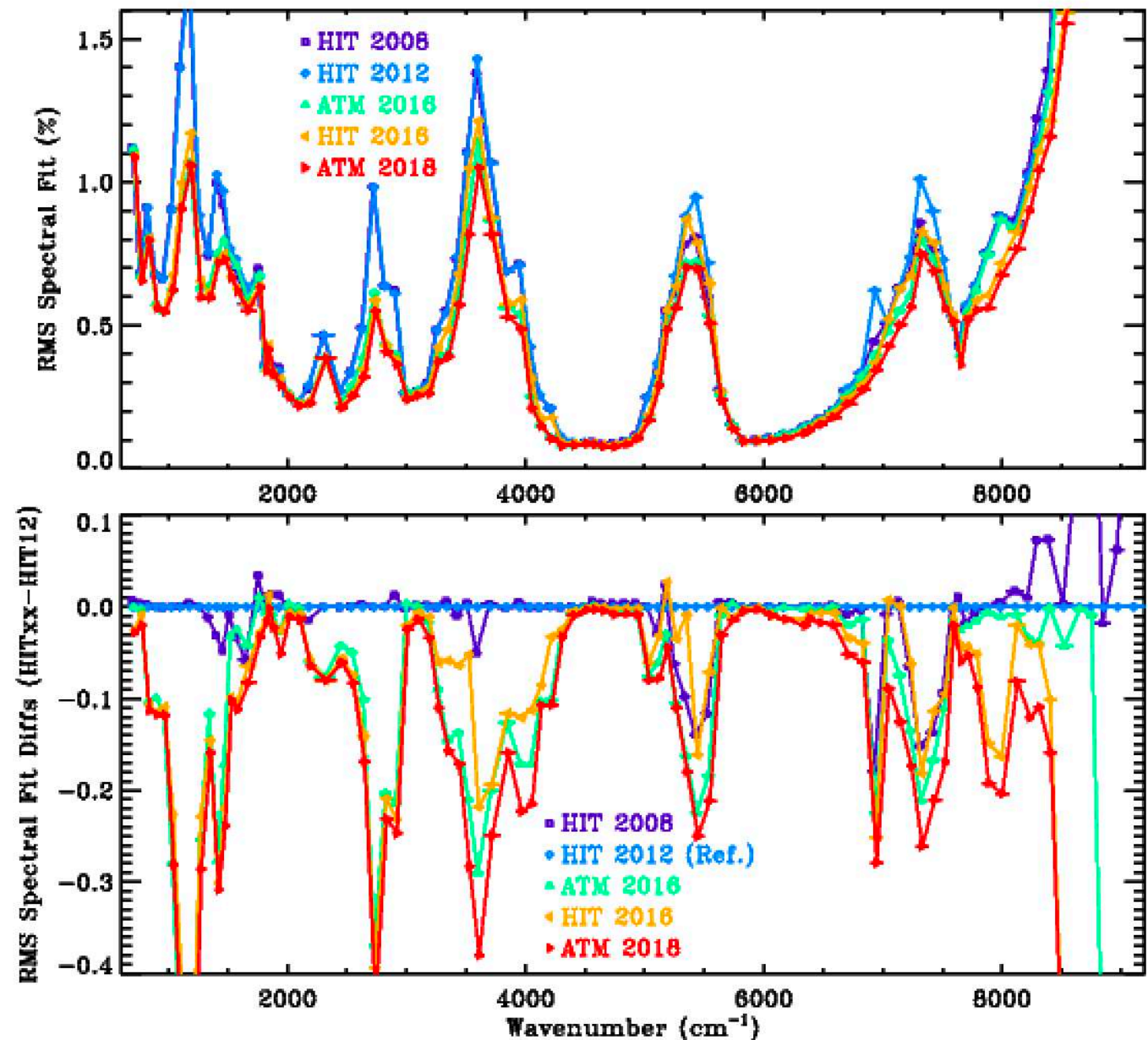


Lab Spectra: RMS Residuals

Upper panel shows the RMS spectral fitting residuals averaged over 154 lab spectra, mainly Kitt Peak. These tend to be large/poor in regions with strong H₂O absorption and smaller/better in regions of weak absorption. Above 6000 cm⁻¹ the available Kitt Peak lab spectra start to fall in signal, so the RMS expressed in transmittance, increases. **Lower panel** shows the difference in RMS relative to HIT12. Negative values imply improved performance.

Although HIT08 is marginally the worst linelist overall, it was much better than HIT12 around 5500 cm⁻¹ and around 1500 cm⁻¹. HIT12 is much better above 8000 cm⁻¹, giving it a slightly improved overall RMS. HIT16 is the worst linelist in just 3 windows.

The ATM18 linelist gives the best (or equal best) fits in every window. This is by virtue of being a “greatest hits” compilation culled from the best of the predecessor linelists. Also, ATM18 benefitted from ad hoc adjustments to fix obvious remaining errors, based on a subset of these same lab spectra (unfair). So need to look at other spectra, besides KP, to confirm its superiority.



Linelist	HIT08	HIT12	ATM16	HIT16	ATM18
Average % RMS:	0.8765	0.8739	0.7673	0.7359	0.6692

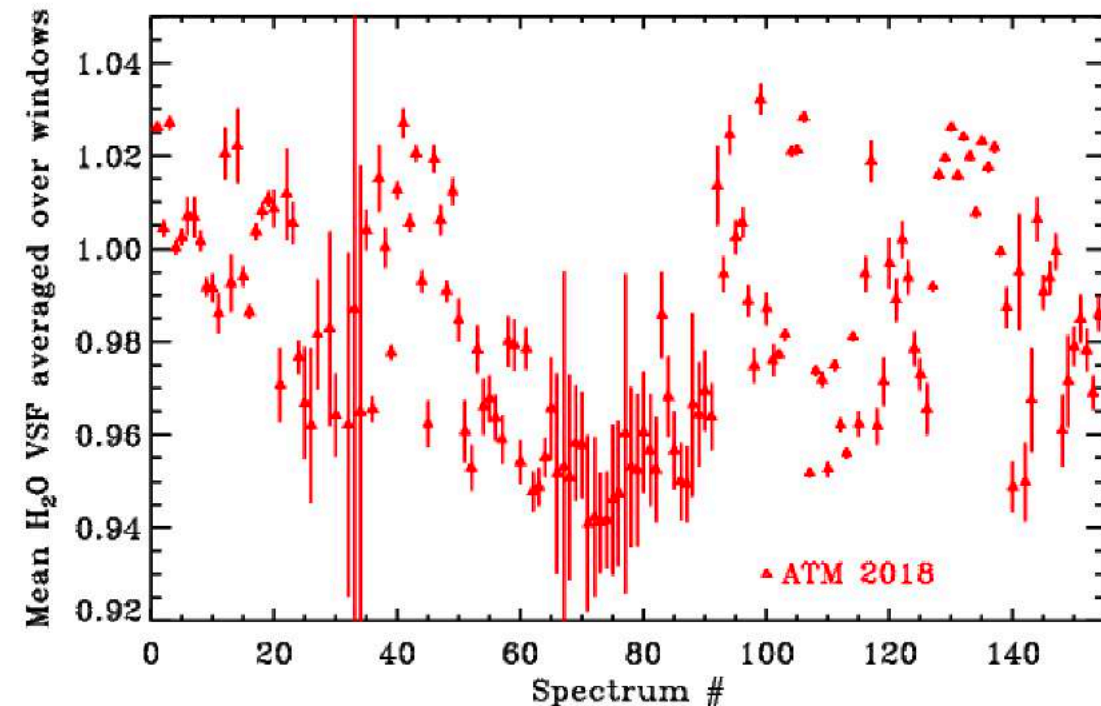
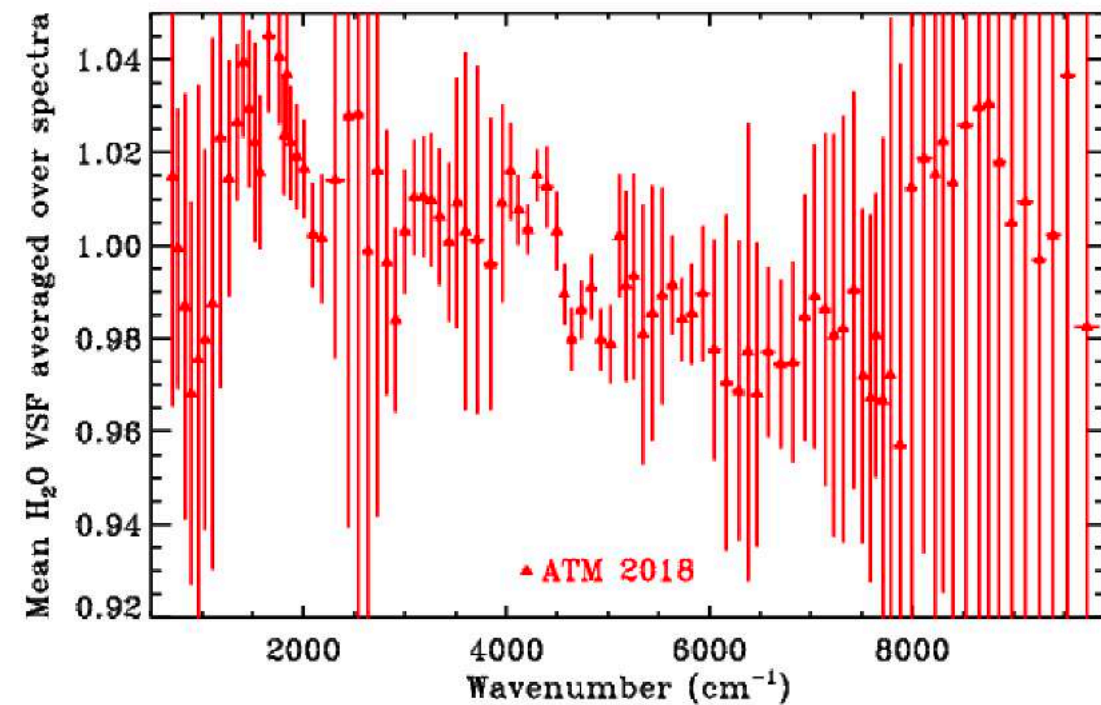
Summarizing Kitt Peak H₂O VSFs

Top Panel: VSF_H₂O 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. Over 2400-2700 cm⁻¹ errors are large due to the weakness of the H₂O lines. Above 7300 cm⁻¹ the lab spectra become increasingly noisy.

Bottom Panel: VSF_H₂O 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, P, T, or path length may be wrong. Or the ILS might be mis-aligned. Or a large zero-offset is present.

Spectra #32-34 have large uncertainties, so why use them? These were recorded at KP in 1983 and are high quality but contaminated. They might be useful in future when contaminant is identified and fitted out. Until then they have little impact due to their large errors.

These plots summarize the information presented 2 slides ago: we have averaged over 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 H₂O are larger than window-to-window variations.



Kitt Peak Lab Spectra – VMR Scaling Factors (VSF) with different linelists

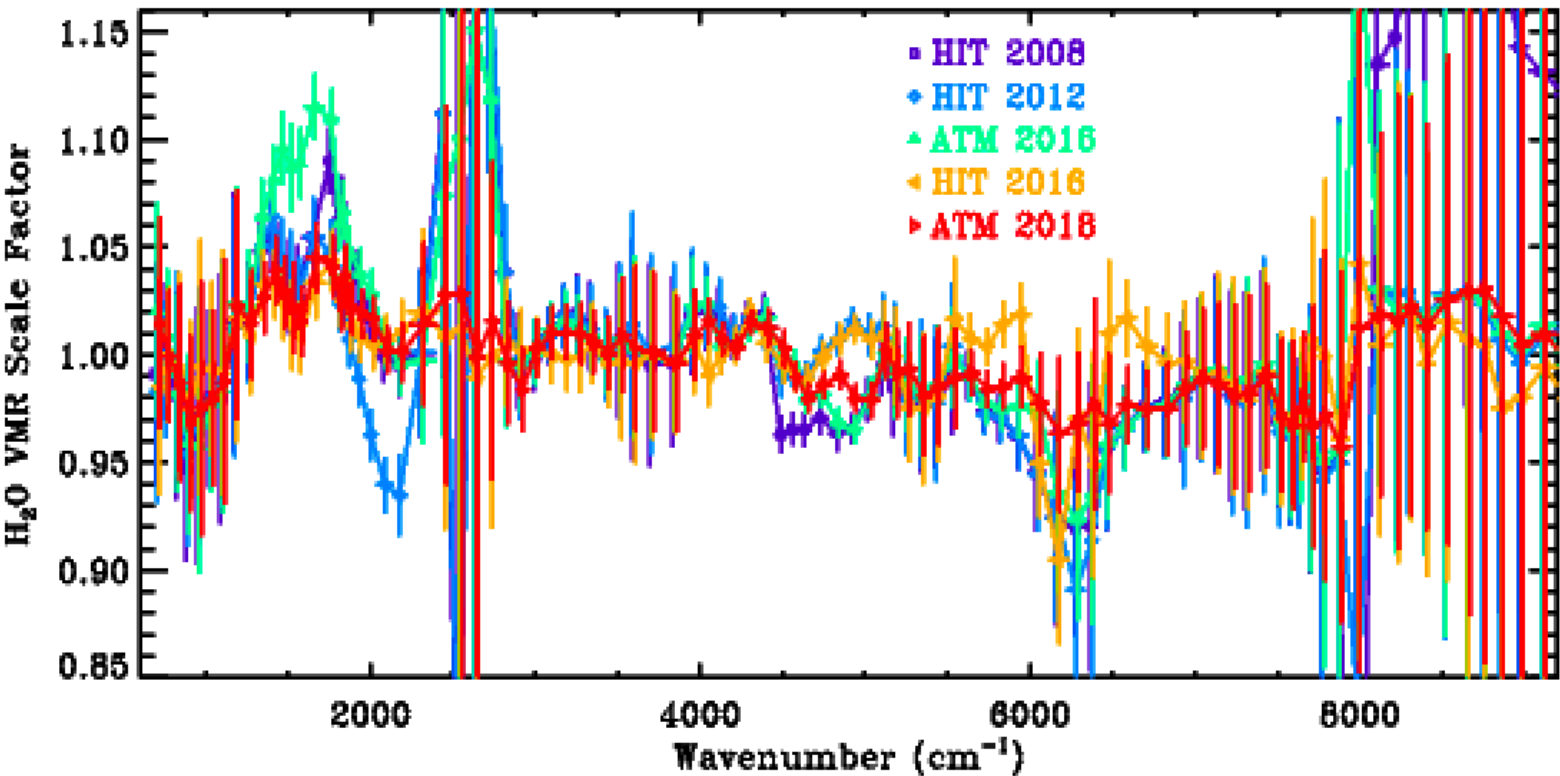
Figure shows VSFs for each window and each linelist, obtained by averaging over the spectra that could be fitted in this window. The red points are identical to those in upper right panel of previous slide.

The VSFs should be all 1.0. Error bars are large in regions where the H₂O lines are weak (e.g. ~2600 cm⁻¹ and 7900+ cm⁻¹).

All linelists have dips in their VSF around 900 cm⁻¹ and 6000-6400 cm⁻¹, and peaks in their VSFs around 1600-1800 cm⁻¹.

HIT 2008 lines were 15-20% too weak in the 8000-9300 cm⁻¹ region.

HIT16 VSFs are less variable than HIT12, but still suffer a dip around 6000-6400 cm⁻¹. Overall HIT16 has the least variable VSFs with an RMS deviation of only 1.41% RMS.



Linelist	HIT08	HIT12	ATM16	HIT16	ATM18
Weighted Mean VSF (over windows)	0.9960	1.0008	1.0036	1.0055	0.9997
RMS deviation from mean	0.0333	0.0243	0.0324	0.0141	0.0165

Summarizing Kitt Peak HDO VSFs

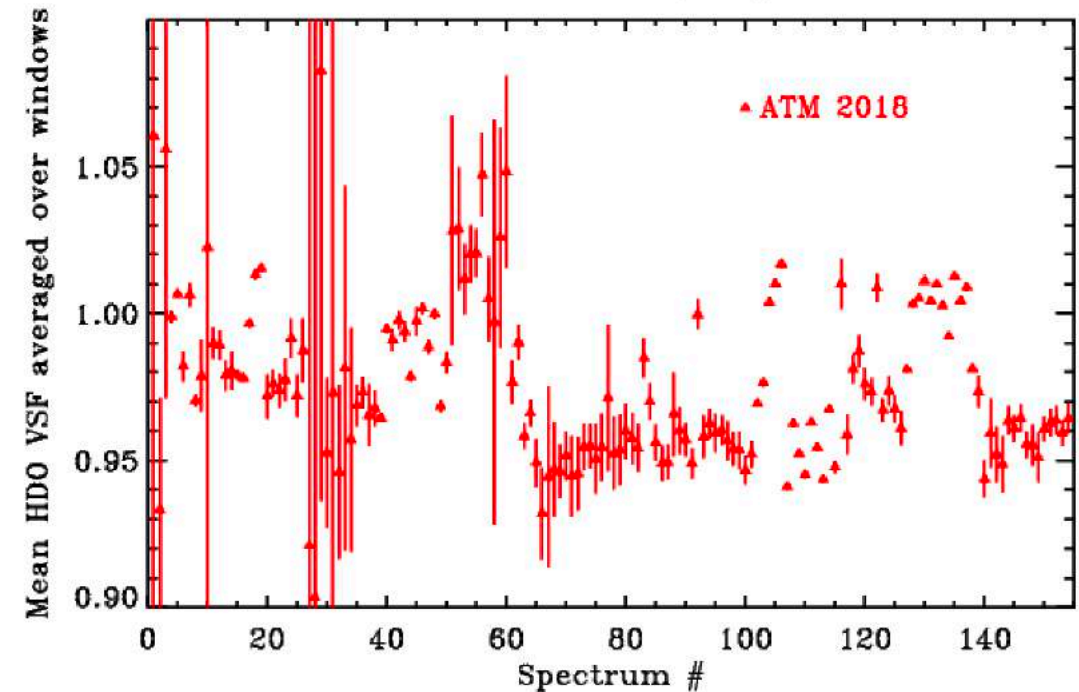
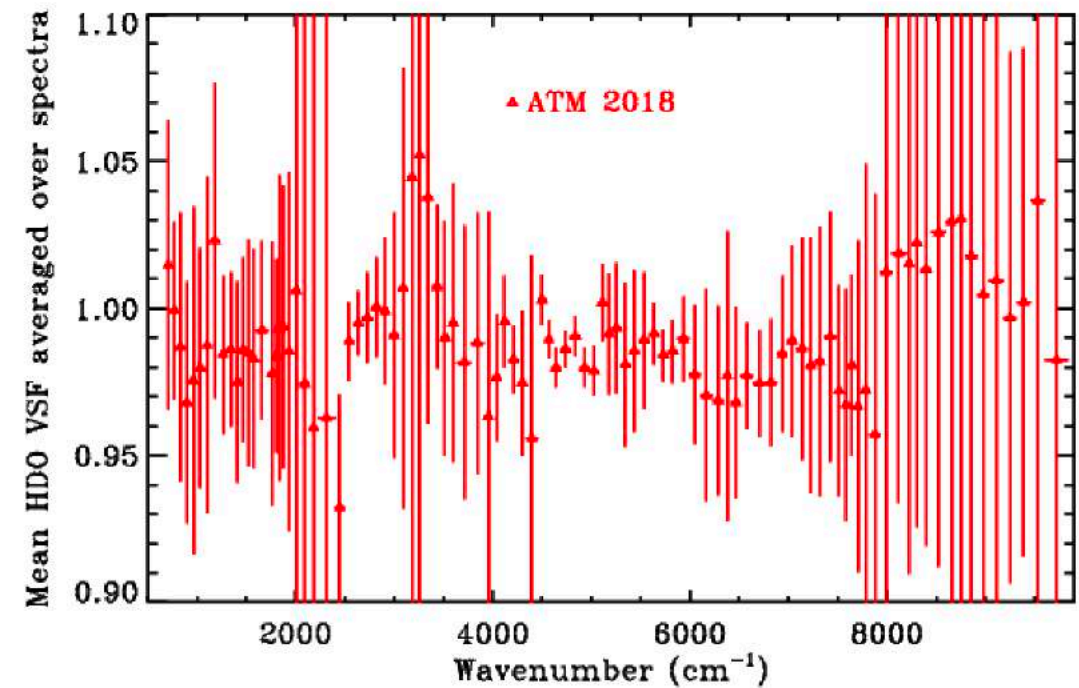
Same plot as two slides ago, but for HDO, not H₂O.

Upper Panel: VSF_HDO 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.

Lower Panel: VSF_H2O 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 a particular spectrum, e.g. the assumed VMR, Pressure, Temp, or path length may be wrong. Or the ILS might be misaligned. Or a large zero-offset is present.

Error bars are smallest in windows with strong HDO absorption, but weak interfering H₂O (e.g. 2500-2800 cm⁻¹; 4500-5000 cm⁻¹).

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 HDO are larger than window-to-window variations.



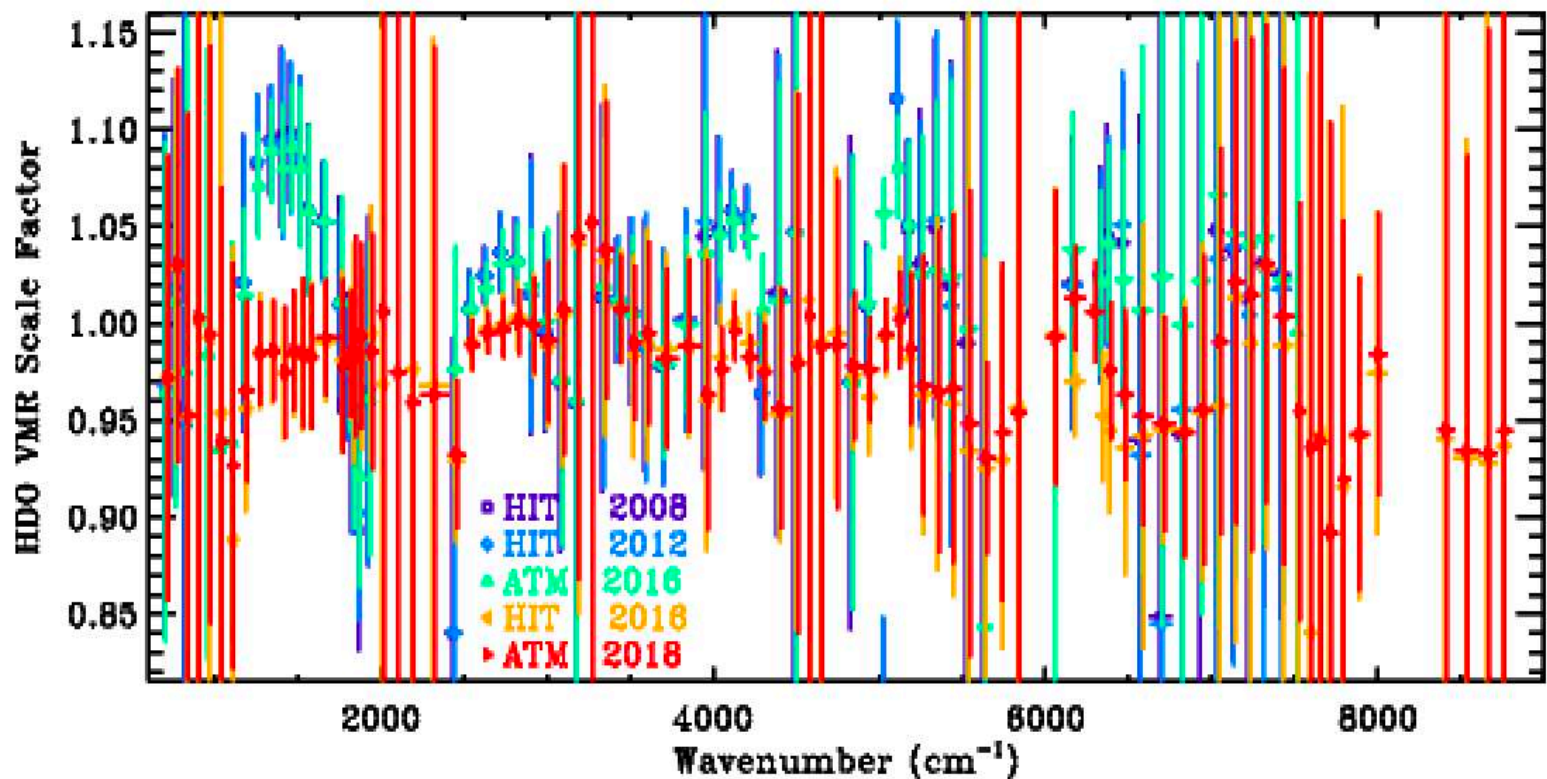
Comparing Kitt Peak HDO VSFs obtained using different linelists

VSF_HDO values, averaged over the different lab spectra fitted in a particular window, plotted versus center wavenumber. Red points (ATM18) are identical to upper right panel of previous slide.

Regions with large error bars contain weak HDO lines, and vice versa. Points with error bars exceeding 0.5 are omitted. For example, the HIT08, HIT12 and ATM16 linelists had no HDO lines 2000-2200 cm⁻¹, 5660-6100 cm⁻¹, and 7500-9600 cm⁻¹ and therefore huge uncertainties there.

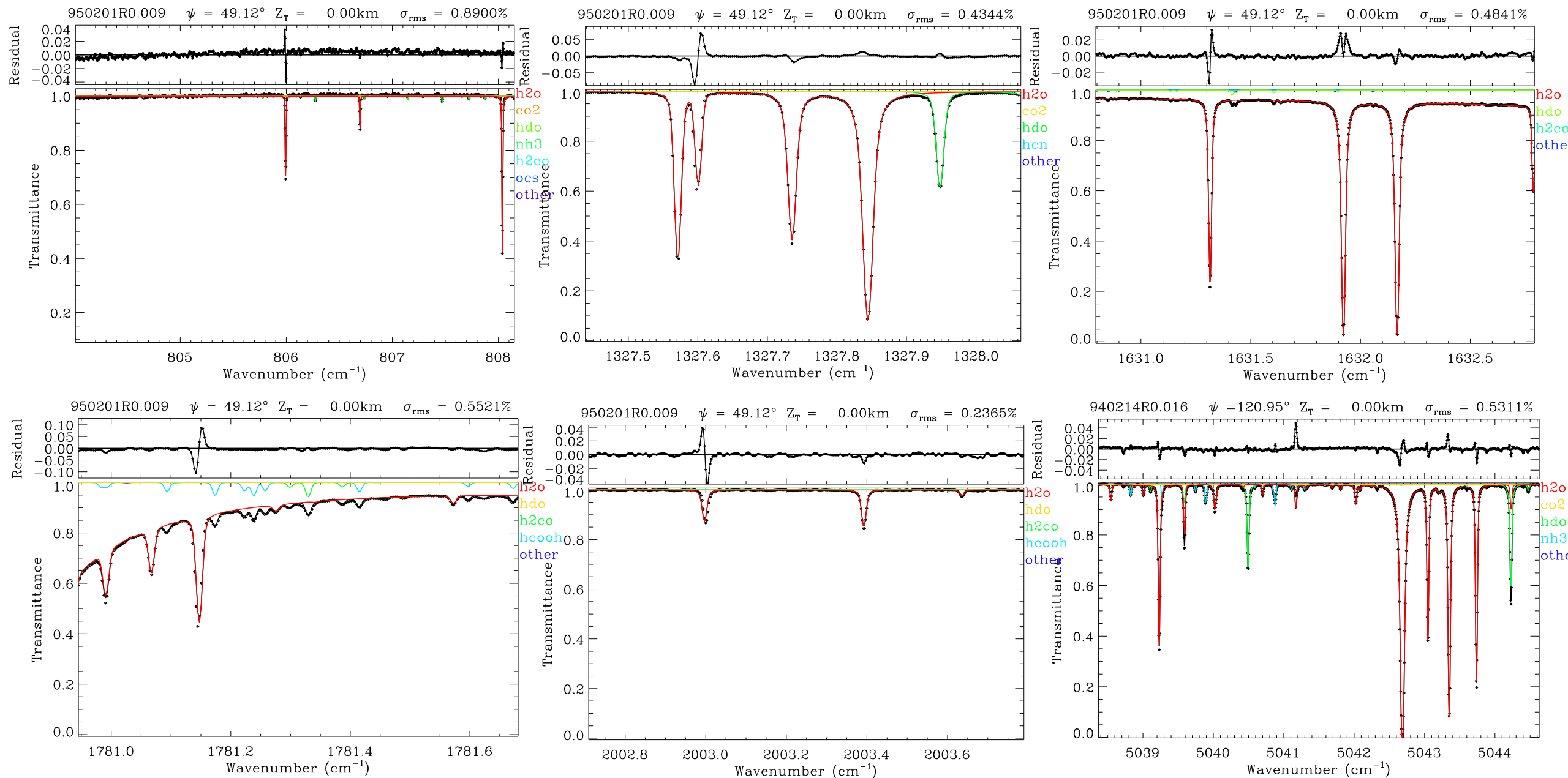
Smallest HDO error bars are 2600-2800 cm⁻¹ where the HDO lines are reasonably strong and with little H₂O interference. This plot exposes windows in which the retrievals are wrong due to factors common to the majority of the fitted spectra, e.g., spectroscopy.

The HDO RMS values are not plotted since they are identical to the H₂O RMS values shown earlier



Linelist	HIT08	HIT12	ATM16	HIT16	ATM18
Weighted Mean VSF (over windows)	1.0320	1.0306	1.0342	0.9976	0.9881
% RMS deviation from mean	0.0460	0.0457	0.0297	0.0174	0.0147

Examples of Line position error in HIT16 linelist in Kitt Peak lab spectra



Fits to KP lab spectra: 6180 to 6260 cm^{-1}

22 Kitt Peak lab spectra cover this band. Shown example is the one with the deepest H_2O absorptions, but only 10% deep despite 13.7 Torr of pure water vapor at 19.5C (80% RH) in a 0.4 km path.

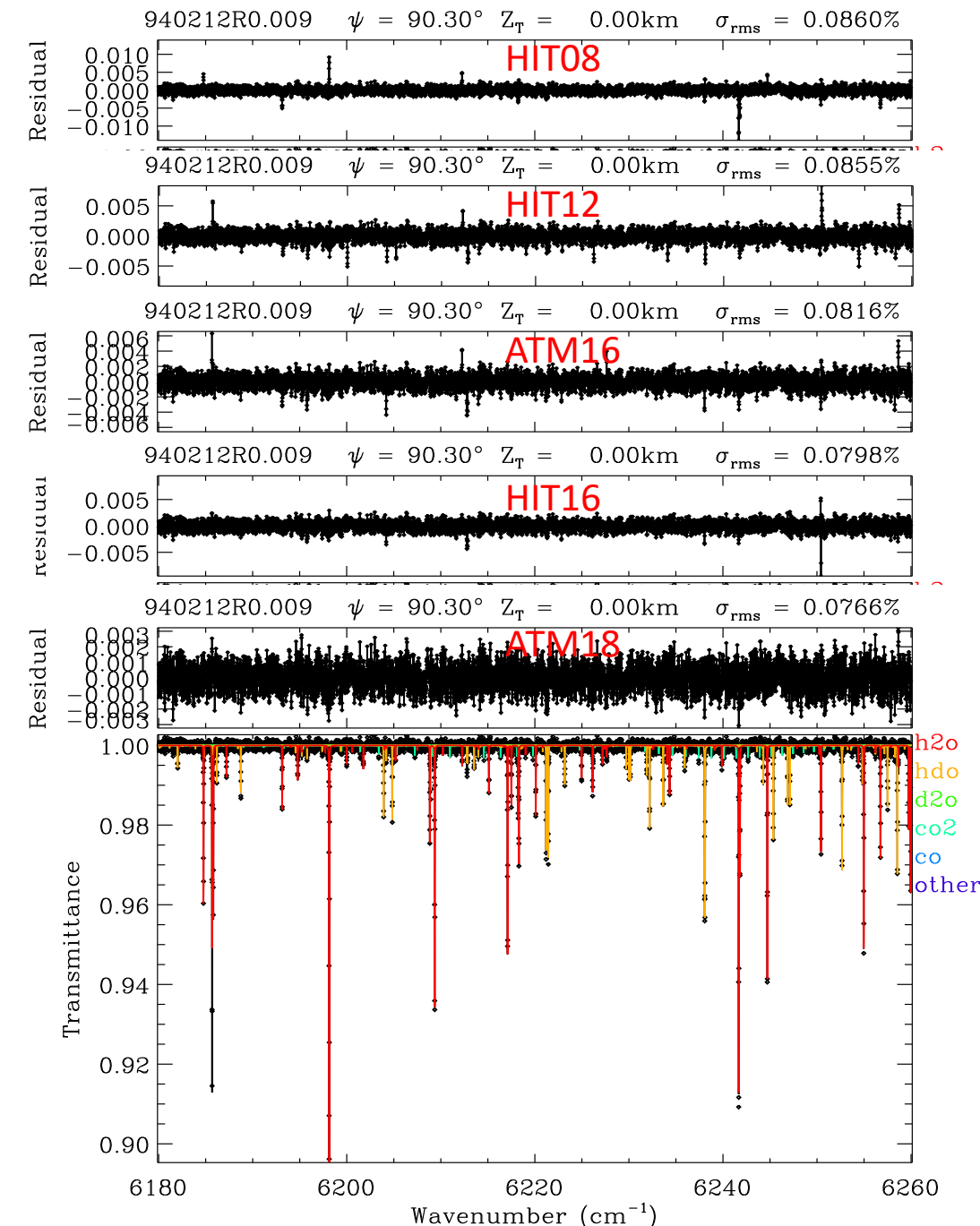
Note that the residual panels are y-auto-scaled: the top panel (HIT08) reaches 1.4% whereas the bottom panel (ATM18) reaches 0.3%. In the bottom panel the residuals become dominated by measurement noise.

Table shows average % rms fitting residuals for all 22 spectra, in the two windows used by TCCON to retrieve atmospheric CO_2 . A monotonic decrease/improvement from HIT08 to ATM18 is seen.

Window	HIT08	HIT12	ATM16	HIT16	ATM18
6180-6260	0.1311	0.1265	0.1261	0.1141	0.1120
6295-6380	0.1444	0.1443	0.1416	0.1307	0.1300

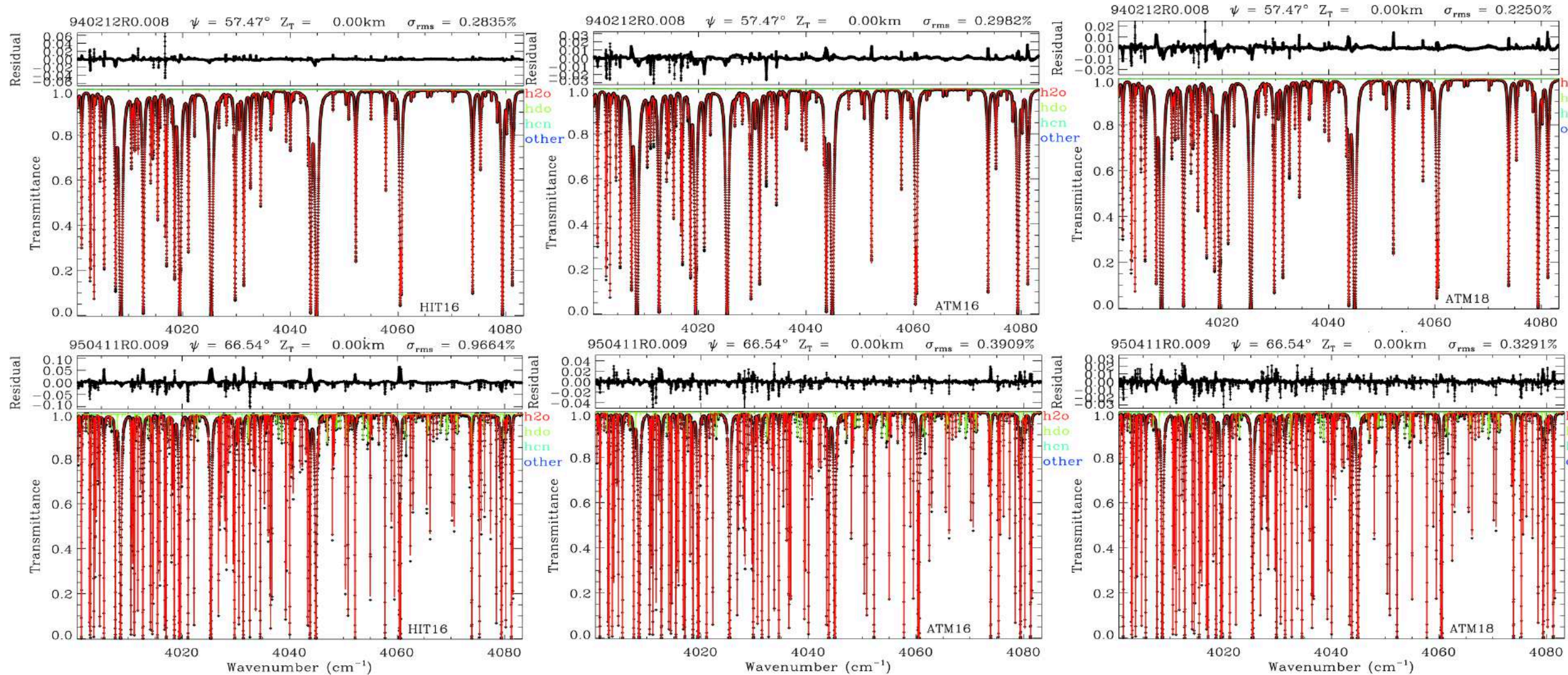
In air-broadened lab spectra in this region, most of the lines become so shallow that they submerge beneath the noise. So atmospheric spectra must be used to check the ABHWs in this region.

No lab spectra of H_2O below 19C in this region. So atmospheric spectra are needed to evaluate the low-T performance of the linelists.



Examples of spectral fits to Kitt Peak lab spectra in 4000-4083 cm^{-1} window

Top Row: 1.55 Torr of H_2O in 360 m path air-broadened to 552 Torr. Bottom Row: 7 Torr of pure H_2O in 433 m path.
Left panels: HIT16 linelist. Middle panels: ATM16. Right panels: ATM18 linelist. Note change of Residual panel y-scales.



Fits KP lab spectra: 4780-4886 cm^{-1}

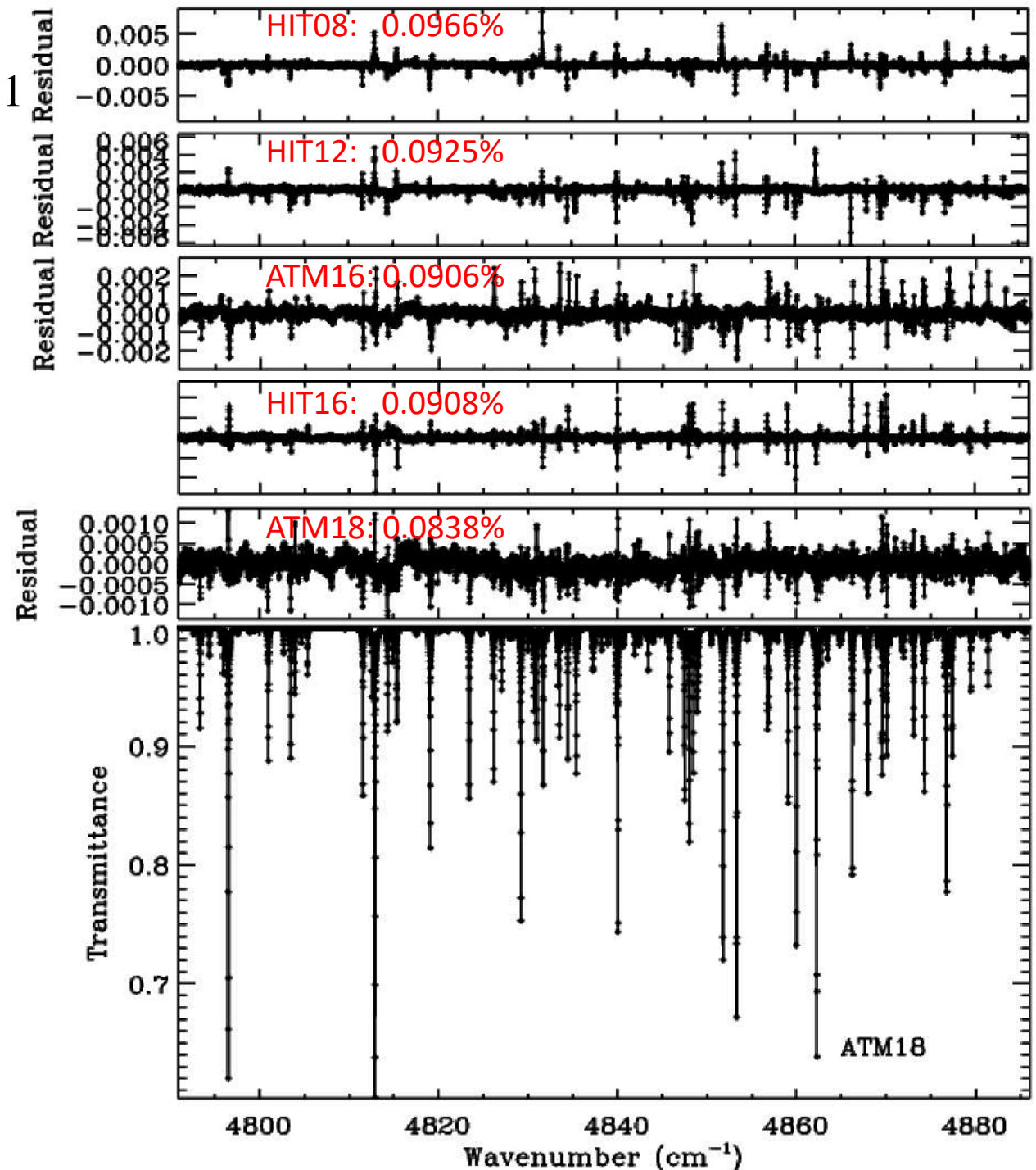
This region is used by OCO-2 and GOSAT: the strong CO_2 band. The residuals and spectra are an average over the 45 KP spectra that cover this band. Note that the residual panels are y-auto-scaled: the top panel (HIT08) reaches 0.9% whereas the bottom panel (ATM18) reaches 0.13%. In the bottom panel measurement noise starts to become noticeable.

HIT16 provides better fits than HIT12, which provides much better fits than HIT08. But the ATM16 was slightly better than HIT16 (in this particular window).

ATM18 is the best of all with peak residuals of 0.13% and RMS residuals of 0.0845%.

The large panel at the bottom shows the measured (points) and calculated (line) transmittance spectra. All significant absorptions are from H_2O .

No lab spectra of H_2O below 19C were available in this region. Atmospheric spectra are therefore needed to evaluate the low-T performance of the linelists.



MkIV Balloon Measurements – RMS Fitting Residuals

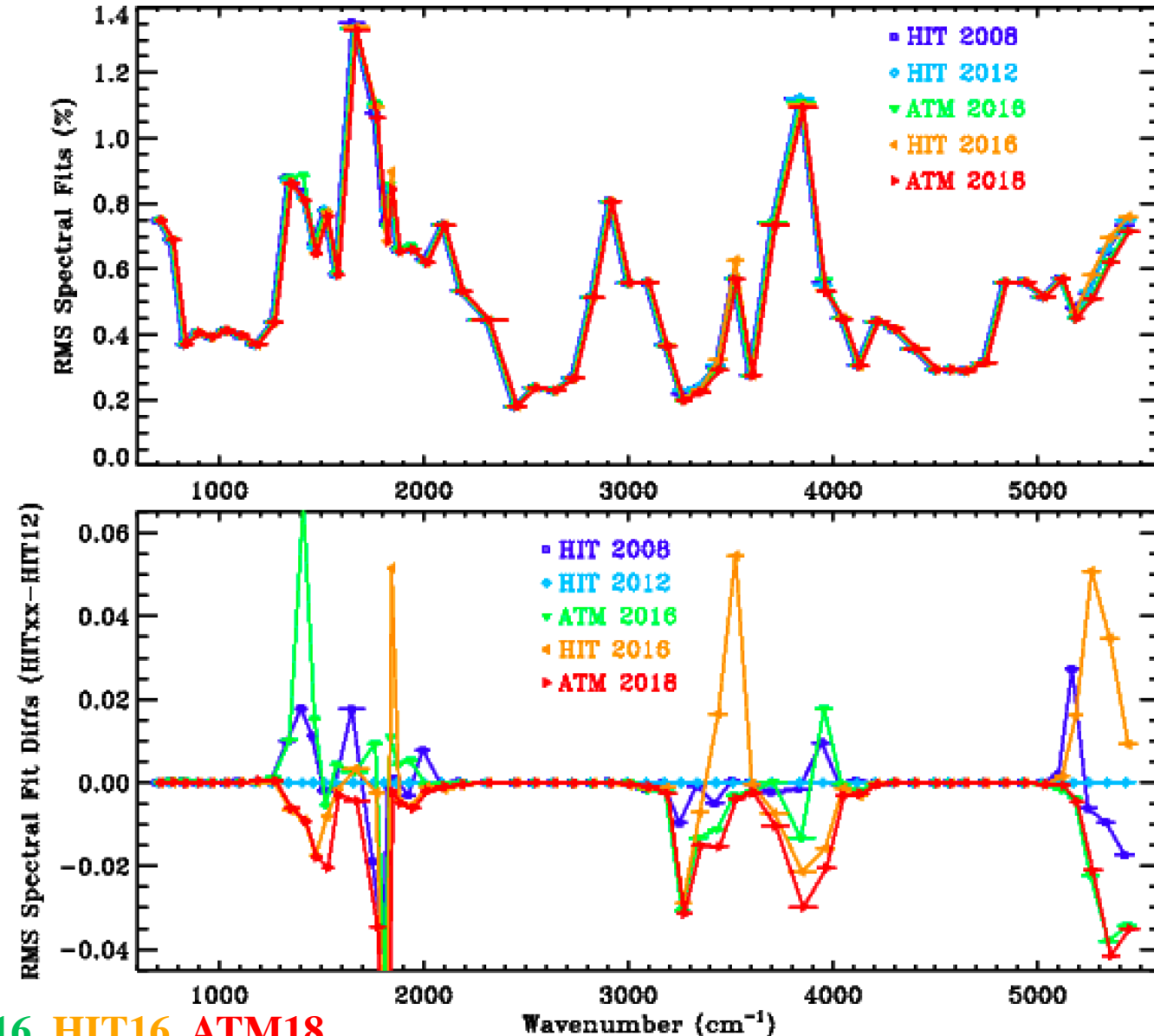
Balloon measurements cover the entire $650\text{-}5600\text{ cm}^{-1}$ region simultaneously, unlike lab spectra. They cover a wide range of T/P conditions (220-250K, 3-300 mbar). This represents the only studied sub-250K H_2O spectra.

Balloon spectra show less linelist-to-linelist variation in the RMS values because the non- H_2O gases, which were not changed, dominate the RMS spectral fit values.

Residuals are larger in regions with stronger H_2O absorption (e.g. $1500\text{-}1800\text{ cm}^{-1}$, $3700\text{-}3900\text{ cm}^{-1}$).

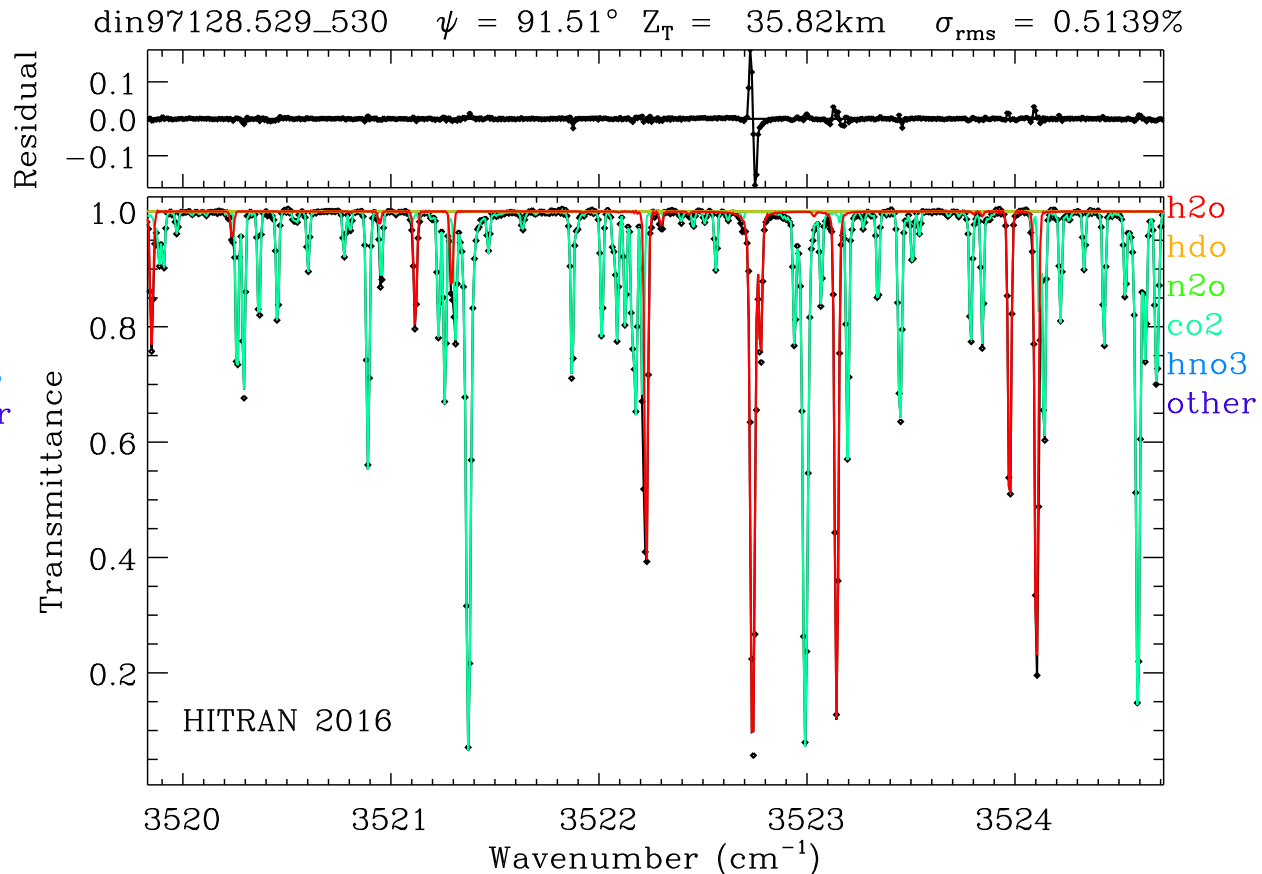
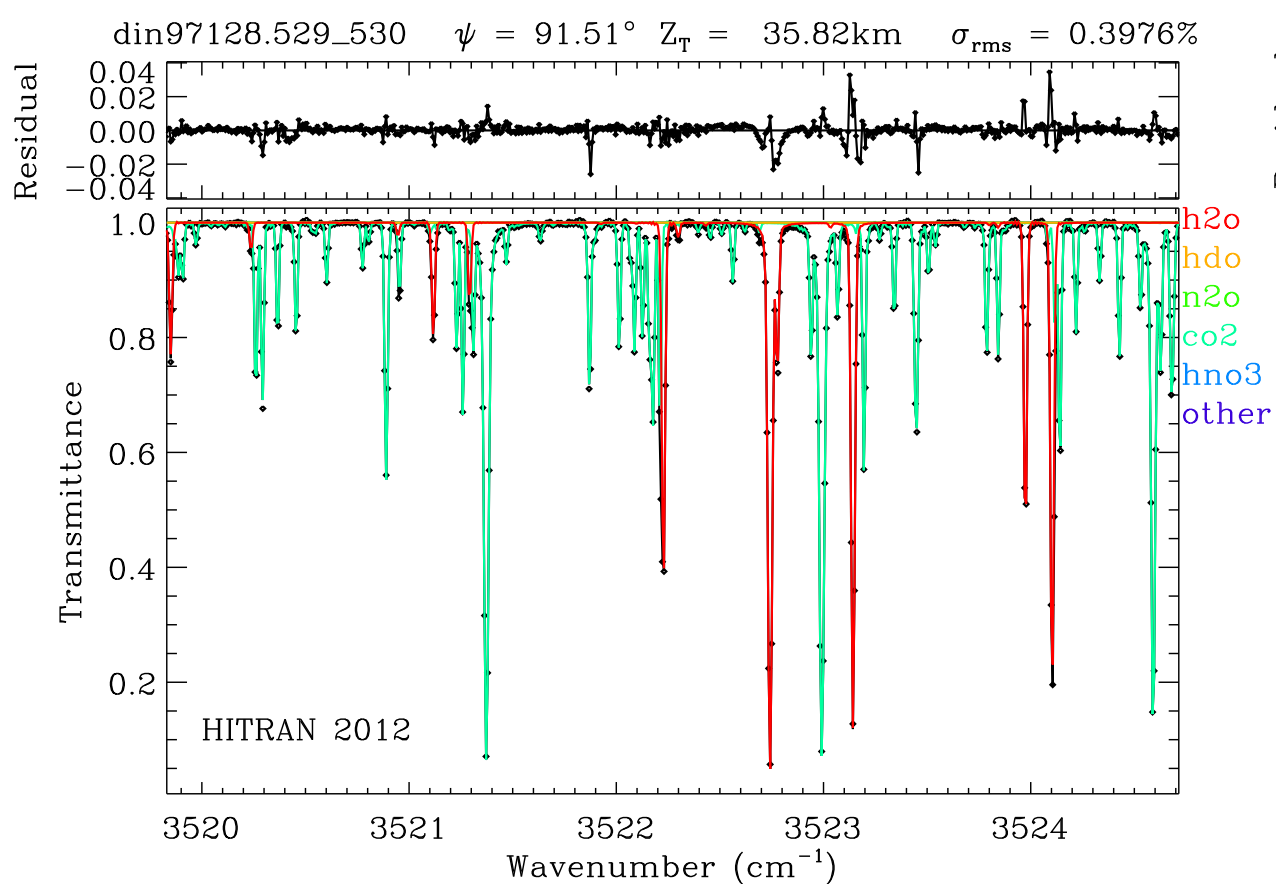
The orange trace shows that HIT16 is the worst of all linelists in just 6 windows (1800 cm^{-1} , $3400\text{-}3550\text{ cm}^{-1}$, & $5200\text{-}5500\text{ cm}^{-1}$).

The ATM linelists are the best around 2300 cm^{-1} , $3200\text{-}3600\text{ cm}^{-1}$, and $5000\text{-}5600\text{ cm}^{-1}$.



Linelist	HIT08	HIT12	ATM16	HIT16	ATM18
% RMS averaged over windows:	0.5375	0.5377	0.5362	0.5376	0.5300

Example of a fit to a MkIV balloon spectrum at 36 km tangent altitude



Illustrating the degradation of fitting residuals due to a mis-positioned H_2O line at 3522.74 cm^{-1} in HIT16.

HIT12: 11 3522.7403 4.276E-21 1.309E+00.10020.494 212.15640.64-.004000 1 0 0 0 0 0 2 1 2 3 2 1

HIT16: 11 3522.7371 4.392E-21 1.345E+00.10080.510 212.15640.76-.006259 1 0 0 0 0 0 2 1 2 3 2 1

There is a 3.2 mK difference in line position between HIT12 and HIT16, the former appearing to be correct.

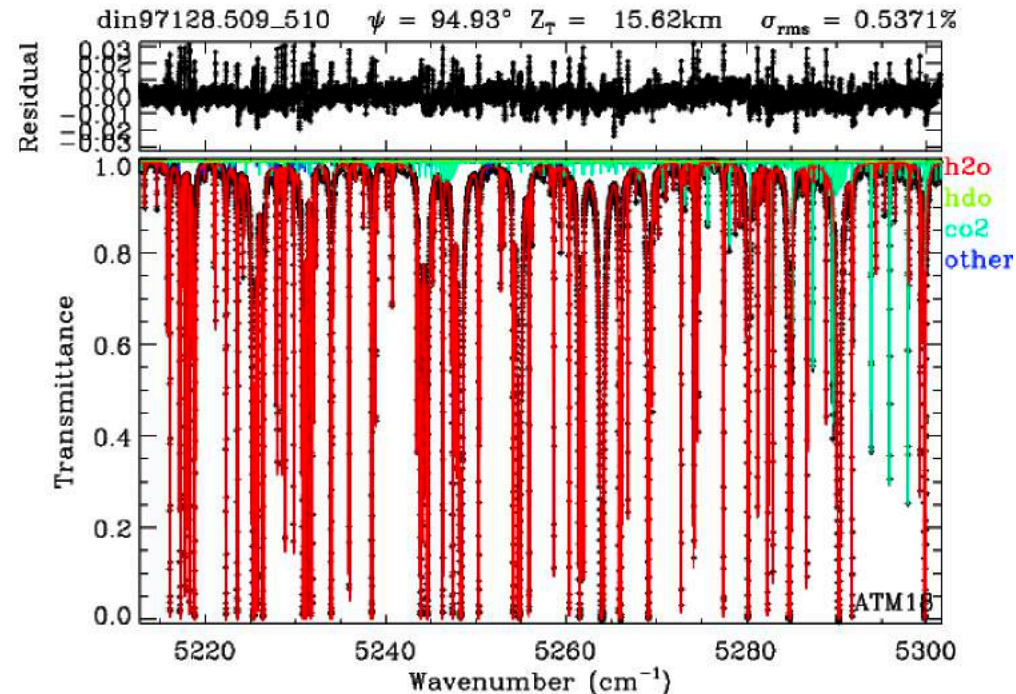
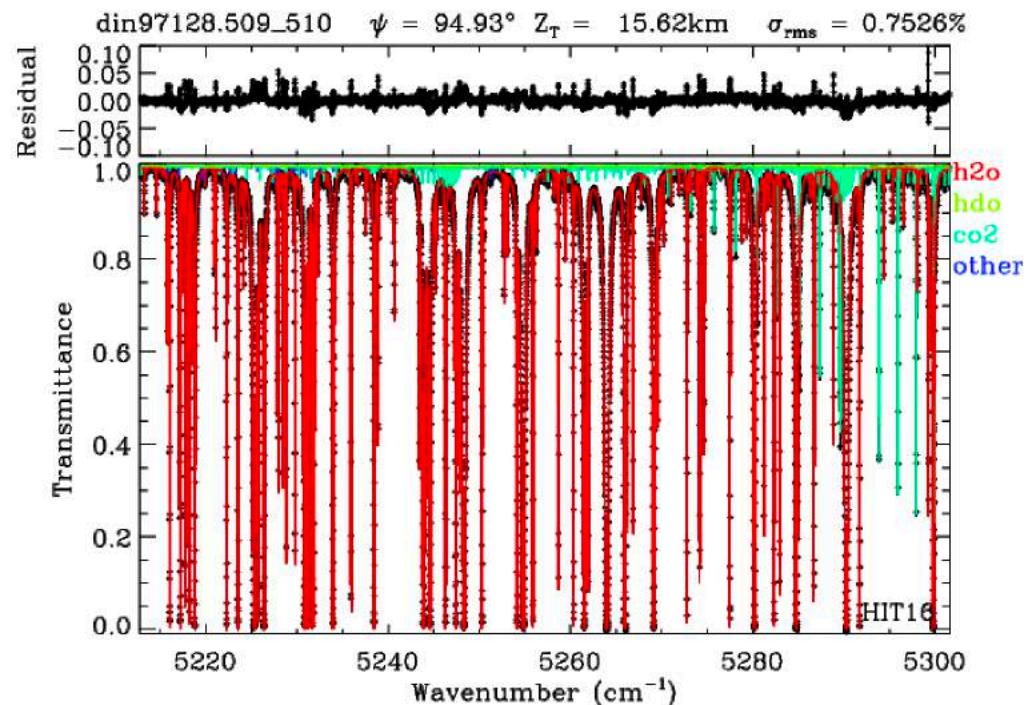
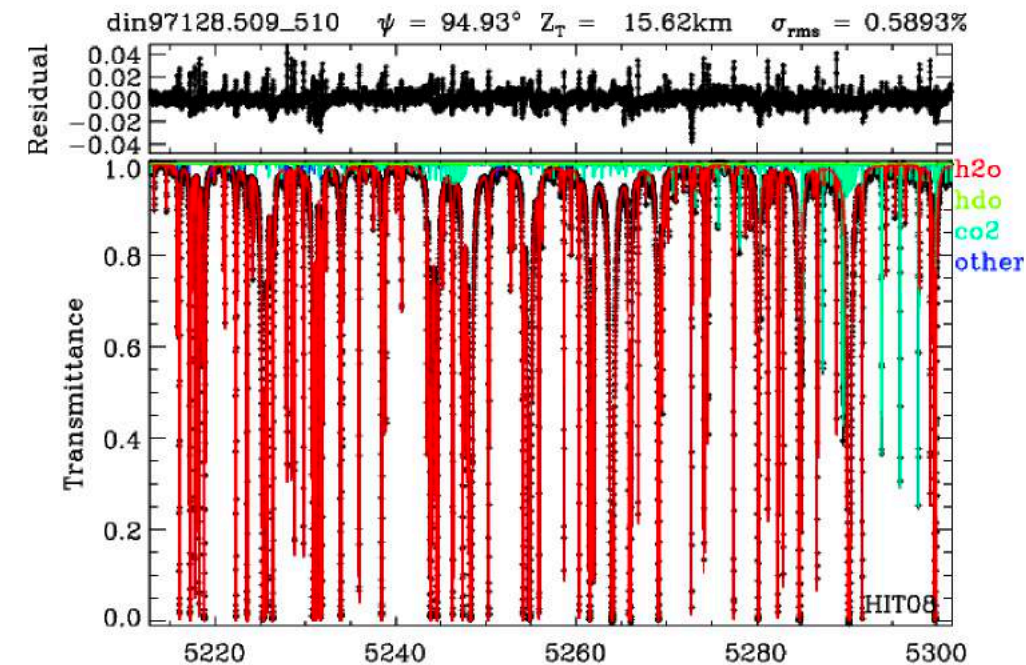
Spectral fits to MkIV balloon spectra

Example of spectral fits to 5257 cm^{-1} window at 15.6 km tangent altitude.

- Upper Left: HITRAN 2008
- Lower Left: HITRAN 2016
- Lower Right: ATM 2018

HITRAN 2016 is substantially worse than HITRAN 2008 in this particular window, and ATM18 is better than either.

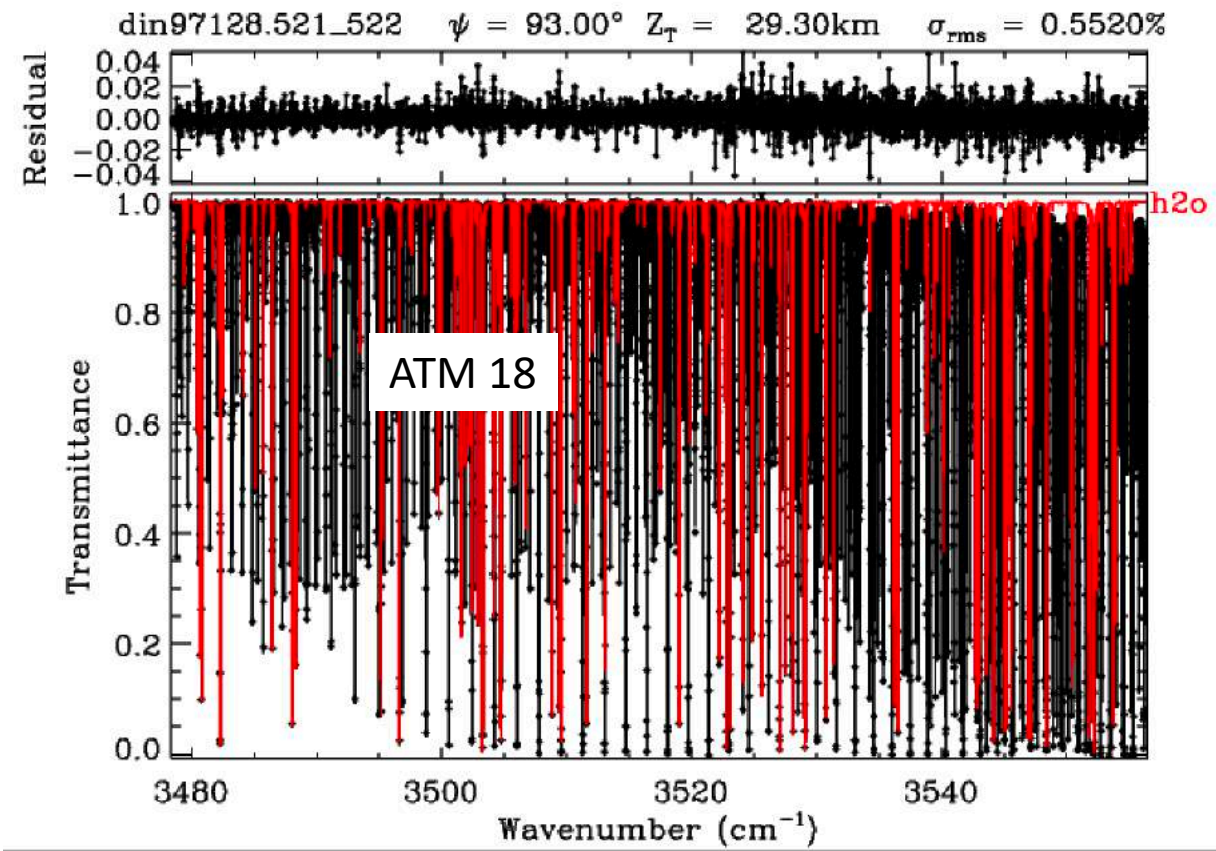
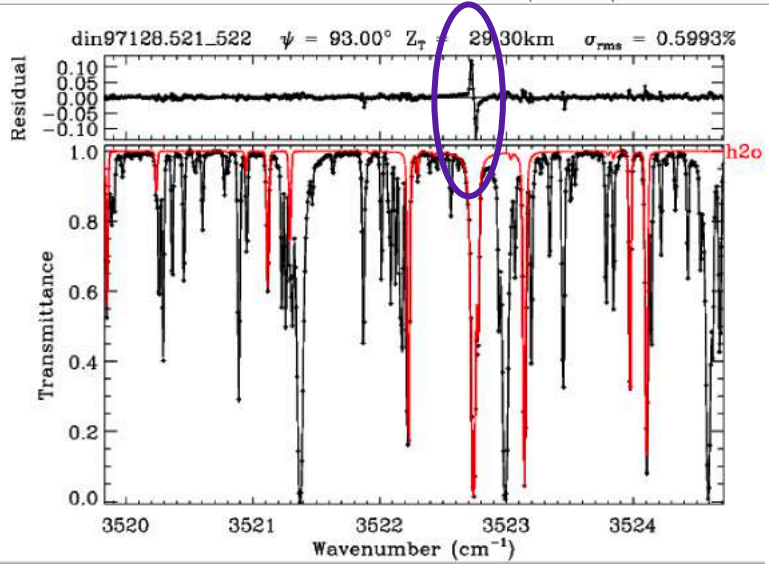
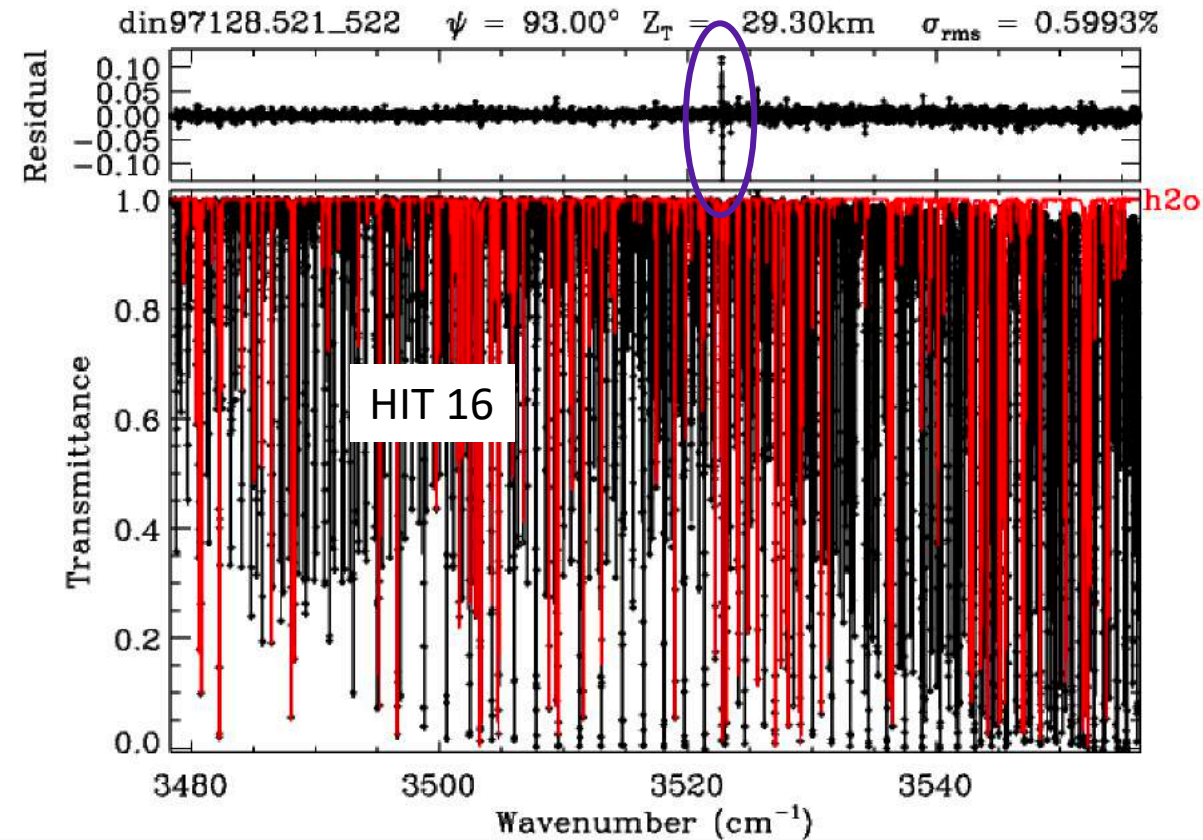
Note the y-scale changes of the residual: 0.05 for HIT08; 0.10 for HIT16; 0.03 for ATM18.



Spectral fits to MkIV balloon spectra at 29 km tangent altitude

Left HIT16: A single mis-positioned H₂O line at 3522.7 cm⁻¹ causes peak residuals of 14%.

Below ATM18: The peak residuals are only 4%.



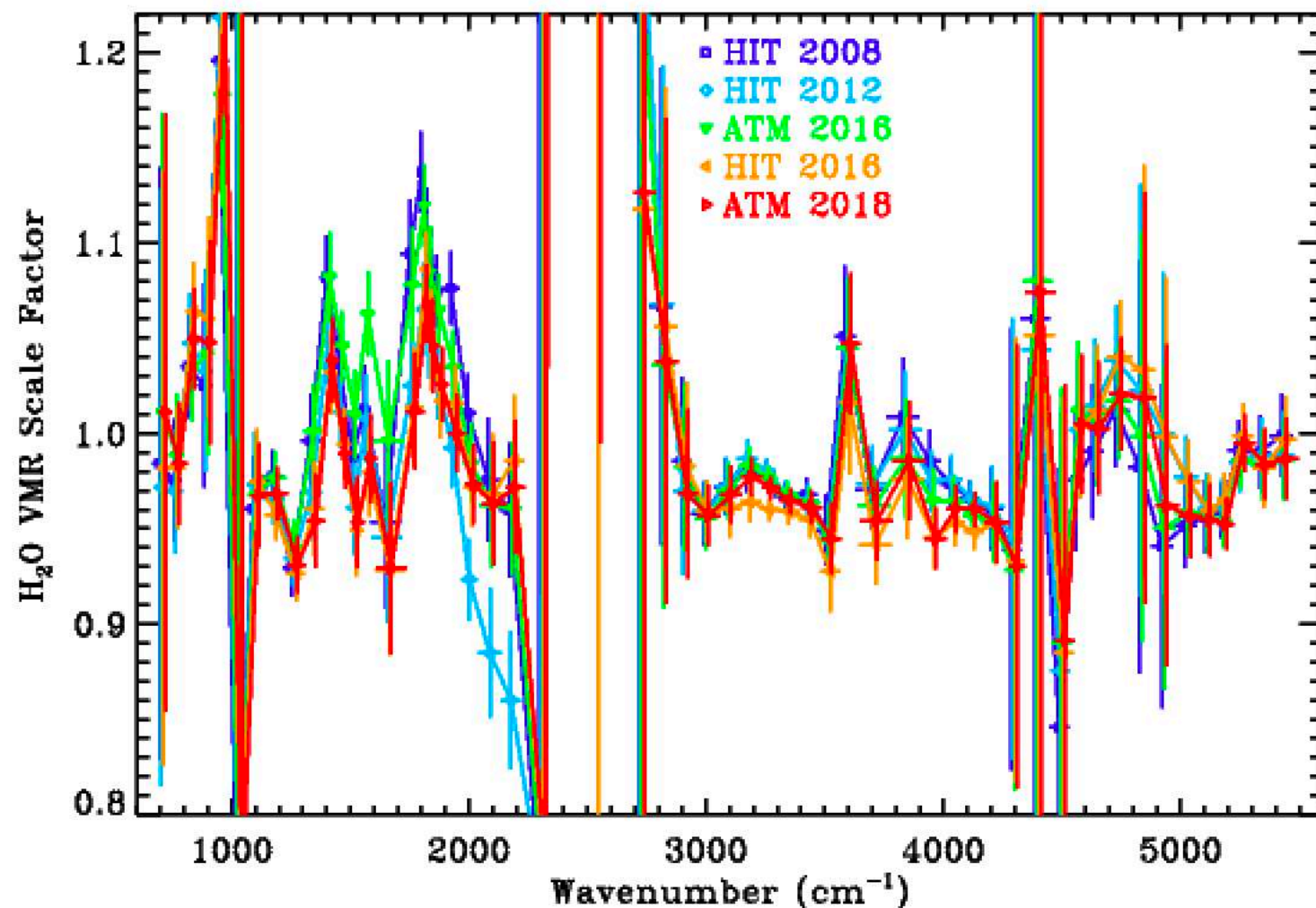
MkIV Balloon Measurements – H₂O VMR Scaling Factors

The assumed atmospheric H₂O profiles based on sondes, which are probably good to only 5-10%. So the absolute VSF values are not that accurate. But the window-to-window variations in VSF are much more precise since all windows were measured simultaneously in the same airmass.

In the 2200-2500 cm⁻¹ region the error bars are very large. This is because the limb path is blacked out by CO₂ and N₂ in the troposphere. And in the stratosphere the H₂O is only ~5 ppm.

In the 2500-2800 cm⁻¹ region the H₂O absorption lines are very weak. HDO is much stronger there.

ATM18 linelist has the best window-to-window consistency (smallest RMS deviation from mean) closely followed by HIT12. HIT08 has the worst window-to-window consistency.



Linelist	HIT08	HIT12	ATM16	HIT16	ATM18
Weighted Mean VSF (over windows)	0.9842	0.9769	0.9807	0.9682	0.9723
RMS deviation from mean	0.0395	0.0305	0.0372	0.0335	0.0297

MkIV Ground-Based - RMS Fitting Residuals

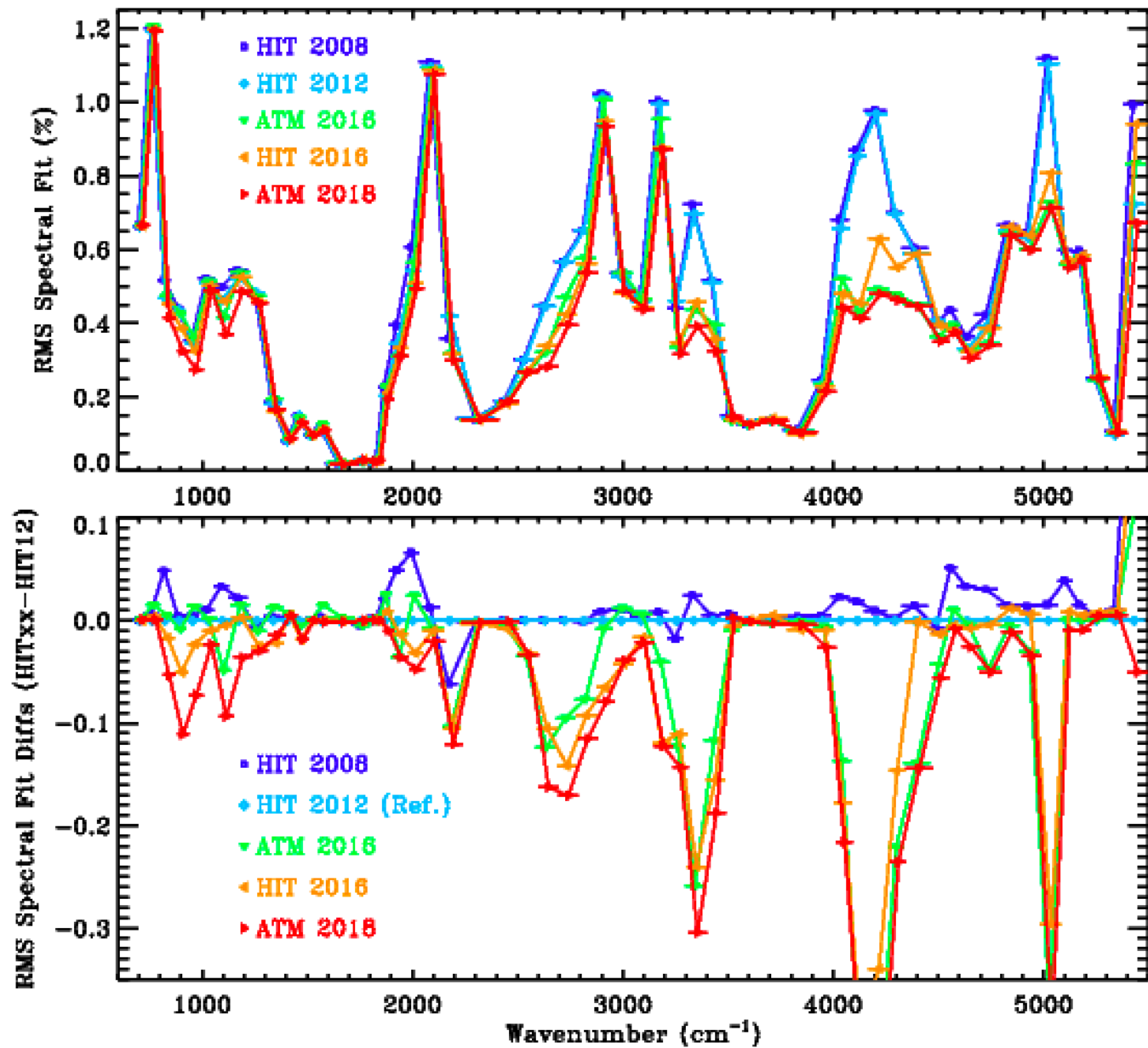
MkIV ground-based spectra are useful 650 to 5500 cm^{-1} .

In blacked-out regions absorption (e.g. 1300-1900 cm^{-1} , 3700-4000 cm^{-1} , 2200-2400 cm^{-1} , 3500-3700 cm^{-1}) retrievals tend not to converge reliably so residuals tend to be small and unstable.

Outside these blacked-out regions, the ATM18 linelist is best or equal best.

Overall, HIT12 is slightly better than HIT08. But around 2200 and 3200 cm^{-1} it is worse.

HIT16 was big improvement over HIT12, but did not surpass ATM16. Overall, ATM18 is by far the best.



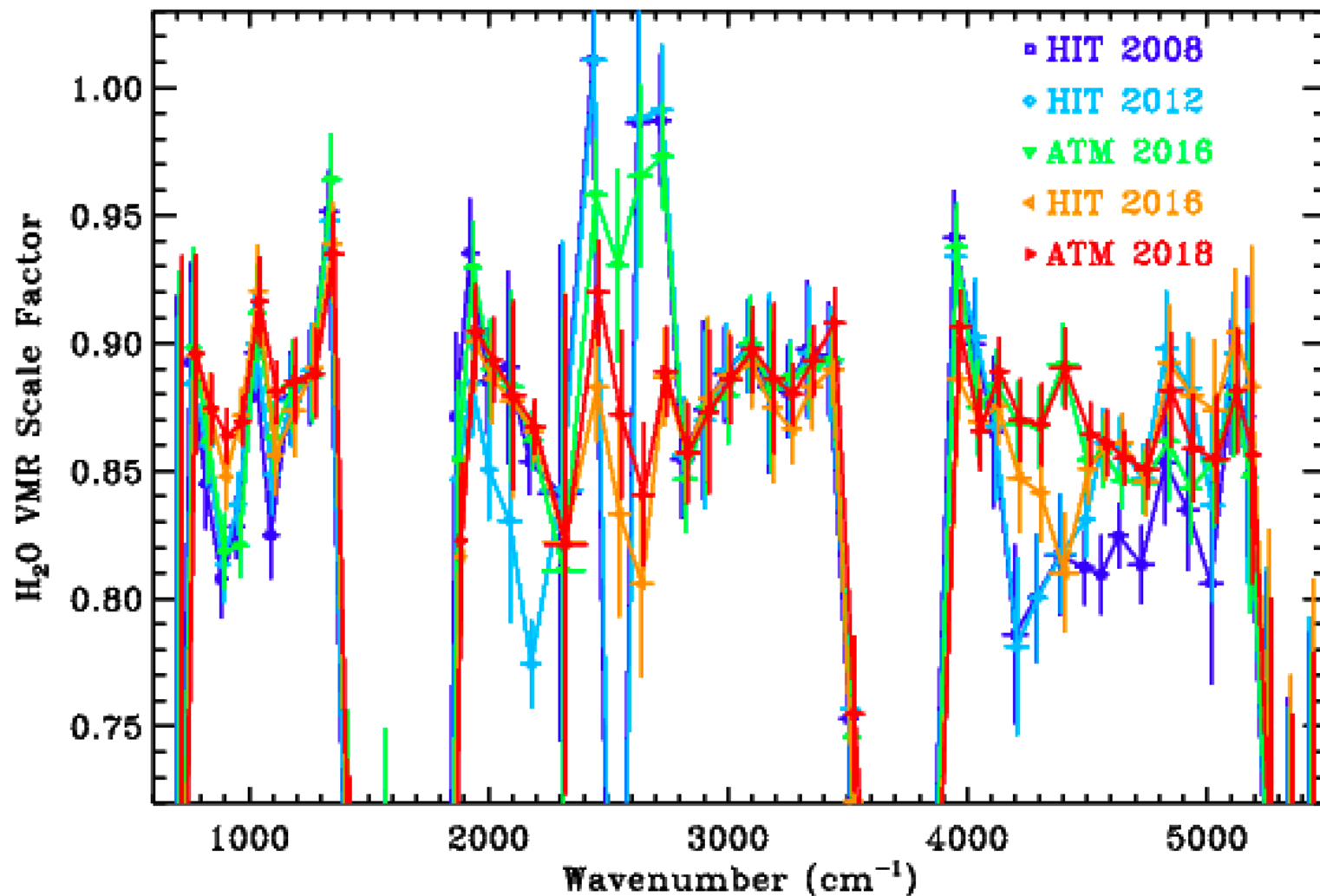
Linelist	HIT08	HIT12	ATM16	HIT16	ATM18
% RMS averaged over windows:	0.4630	0.4485	0.4014	0.4040	0.3752

MkIV ground-based H₂O VMR Scaling Factors

Assumed atmospheric H₂O profiles based on NCEP analysis, which are probably good to only 15%. So the absolute VSF values are not that accurate. But the window-to-window variations in VSF are precise since all windows were measured simultaneously in the same airmass.

In blacked out regions, e.g. 1400-1900, 2300-2400, and 3500-3900 cm⁻¹, the error bars are very large.

Window-to window consistency improves with each linelist version.



Linelist	HIT08	HIT12	ATM16	HIT16	ATM18
Weighted Mean VSF (/windows)	0.8587	0.8640	0.8724	0.8686	0.8755
RMS deviation from mean	0.0588	0.0544	0.0411	0.0357	0.0320

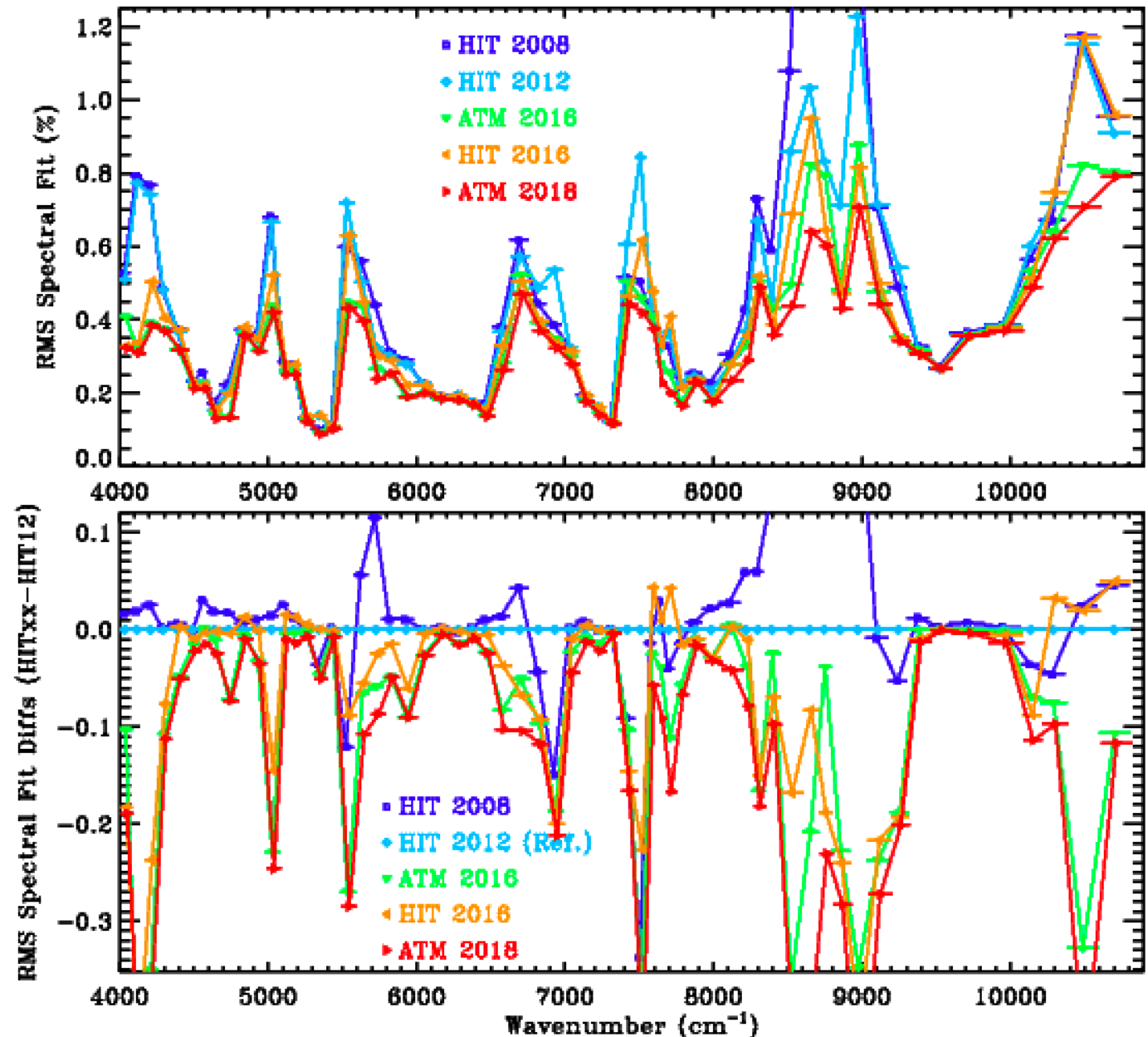
TCCON Ground-based: RMS Fitting Residuals

In terms of RMS fitting, HIT08 is by far the worst linelist, especially over the 8500-9000 cm^{-1} region, where its RMS residuals exceed 2%. Between 7500-7700 cm^{-1} and at 10,300 cm^{-1} , however, it is better than HIT2016!

Although HIT16 was much improved over HIT12, it was not better than ATM16.

HIT16 has the worst RMS in 7 windows: including 4850, 5200-5400, 7600-7700, and above 10200 cm^{-1} .

ATM16 has the best (or equal best) RMS in every window (if not, I would have replaced the defective lines from whatever linelist was best).



Linelist	HIT08	HIT12	ATM16	HIT16	ATM18
% RMS averaged over windows:	0.5050	0.4431	0.3507	0.3823	0.3209

TCCON Ground-based – H₂O VMR Scaling Factors

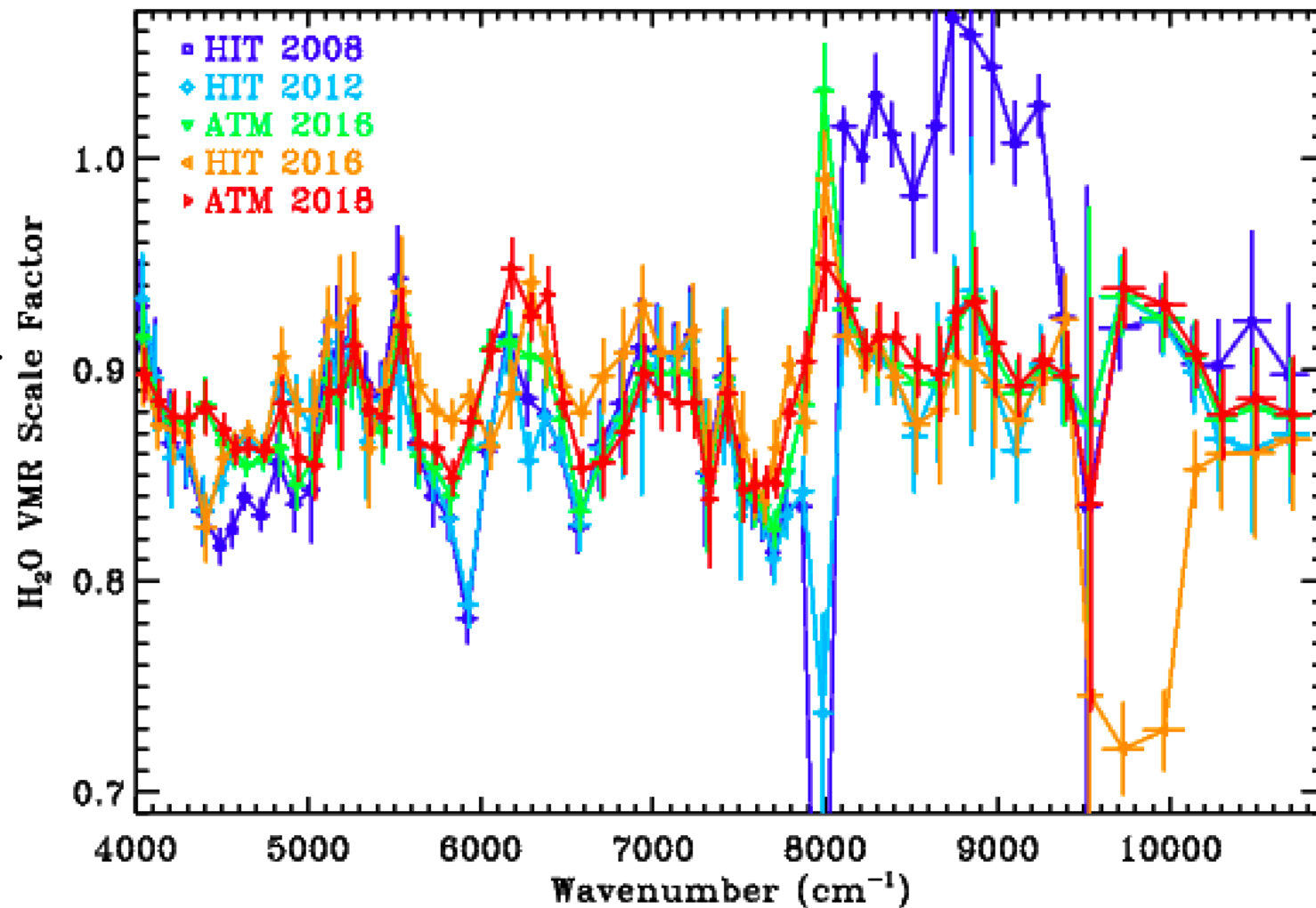
Assumed atmospheric H₂O profiles based on FPIT meteorological analysis (good to ~10%). So the absolute VSF values of 0.88 likely indicate error in assumed atmospheric H₂O, not spectroscopy.

Window-to-window variations in VSF are precise since all measured simultaneously in same airmass.

In the 8000-9300 cm⁻¹ region, HIT2008 produces much larger H₂O amounts than any other linelist.

In the 9500-10100 cm⁻¹ region, HIT2016 gives much lower H₂O amounts than any other linelist. This cannot be confirmed with the lab spectra at my disposal since the lines are too shallow.

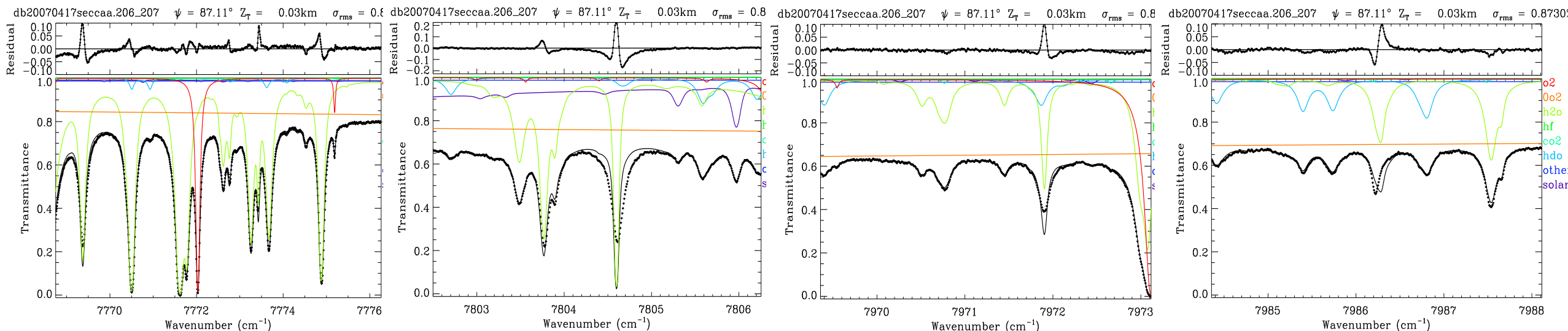
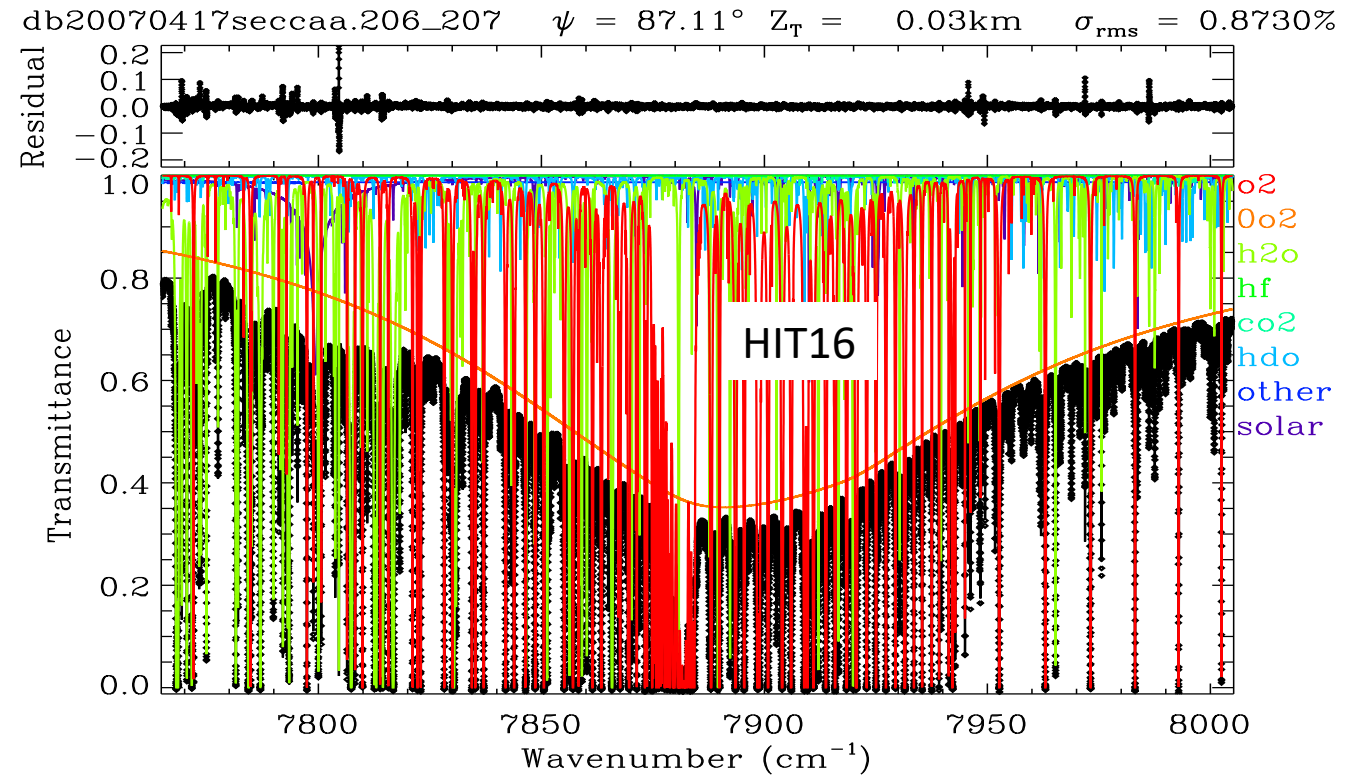
The ATM18 linelist has best window-to window consistency (smallest RMS deviation from mean).



Linelist	HIT08	HIT12	ATM16	HIT16	ATM18
Mean VSF (over windows)	0.8726	0.8669	0.8719	0.8805	0.8793
RMS deviation from mean	0.0631	0.0324	0.0284	0.0299	0.0255

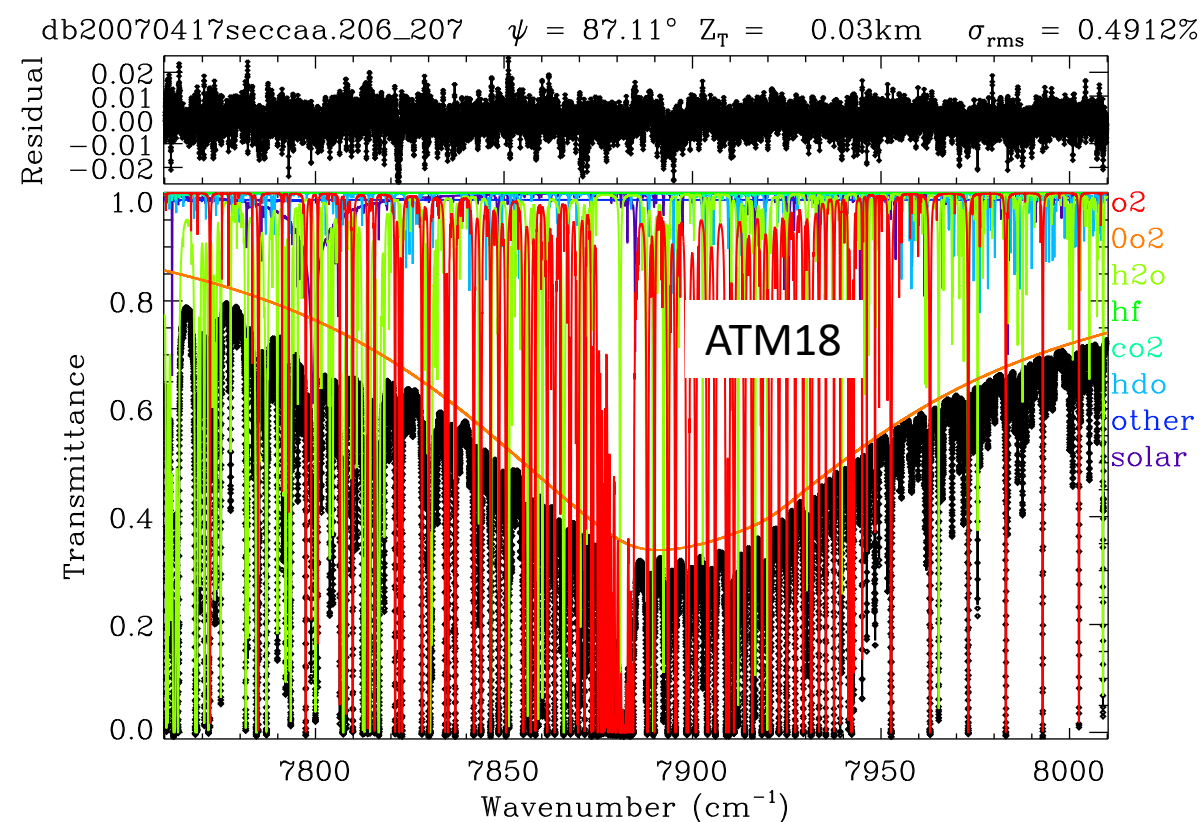
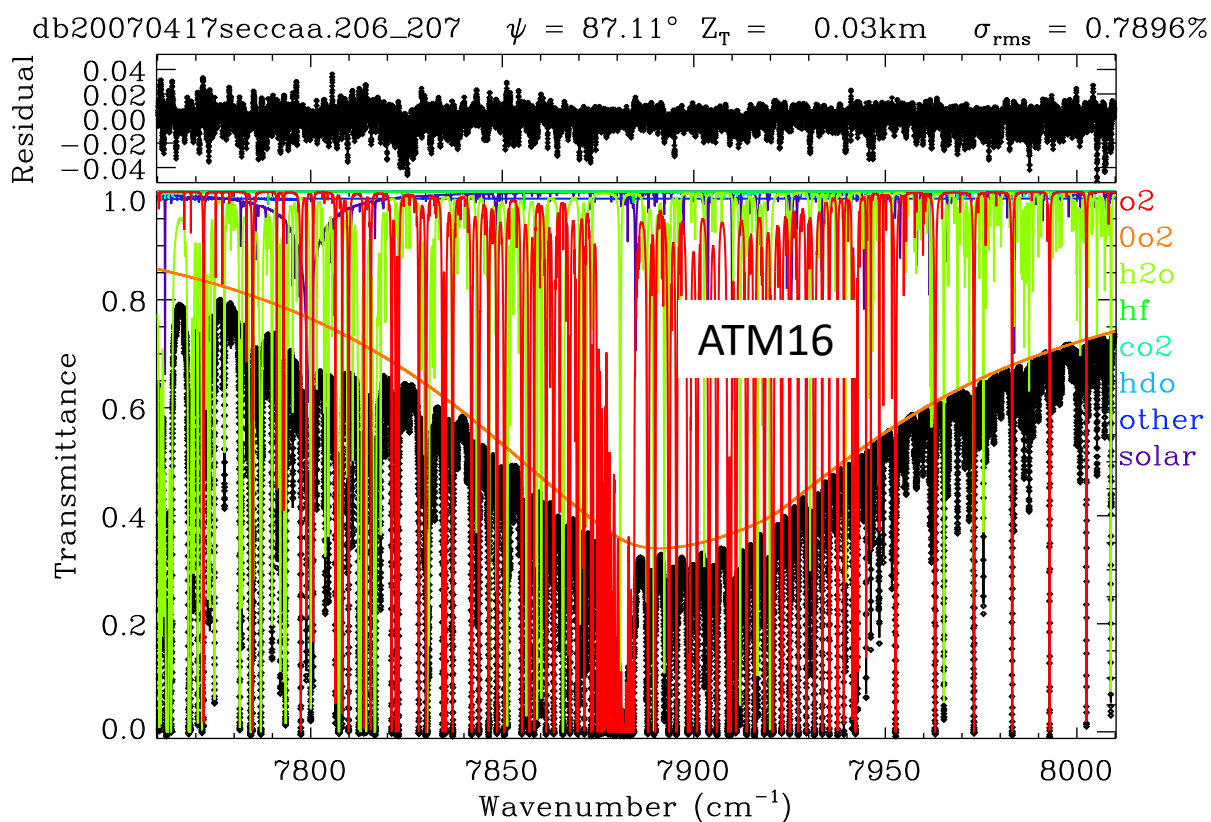
Fits to a humid high-airmass ground-based TCCON spectrum

Although fits with HIT16 H₂O linelist are generally good, in the 7765-8005 cm⁻¹ region, there are ~20 lines with poor widths and/or pressure shifts. These give rise to large residuals, which degrade the overall rms residual to 0.8730 %, as compared with 0.7896 % for the ATM16 linelist. Four examples are shown below. For this reason the ATM16 linelist was taken as the starting point for the ATM18 H₂O linelist in this window, which eventually achieved a RMS fitting residual of 0.491% by combining the best features of ATM16 and HIT16 (see next slide).



Fits to humid high-airmass ground-based TCCON spectrum (Darwin)

Illustrating fits to 7760-8010 cm^{-1} region, which is used by TCCON to retrieve O_2 . Left panel shows fits using ATM16 linelist. Fits on right uses ATM18 linelist, which results in the RMS residuals being reduced by 40%. This particular spectrum has a large H_2O slant column of $3\text{E}+24$ molecules. cm^{-2} , which makes it good at exposing deficiencies in the weak H_2O lines. In spectra acquired in less humid conditions or at lower airmass, the residuals are dominated by O_2 and the stronger H_2O lines, which don't change so much between linelist versions.



Examples of fits to TCCON ground-based spectra: 4790 to 4890 cm^{-1}

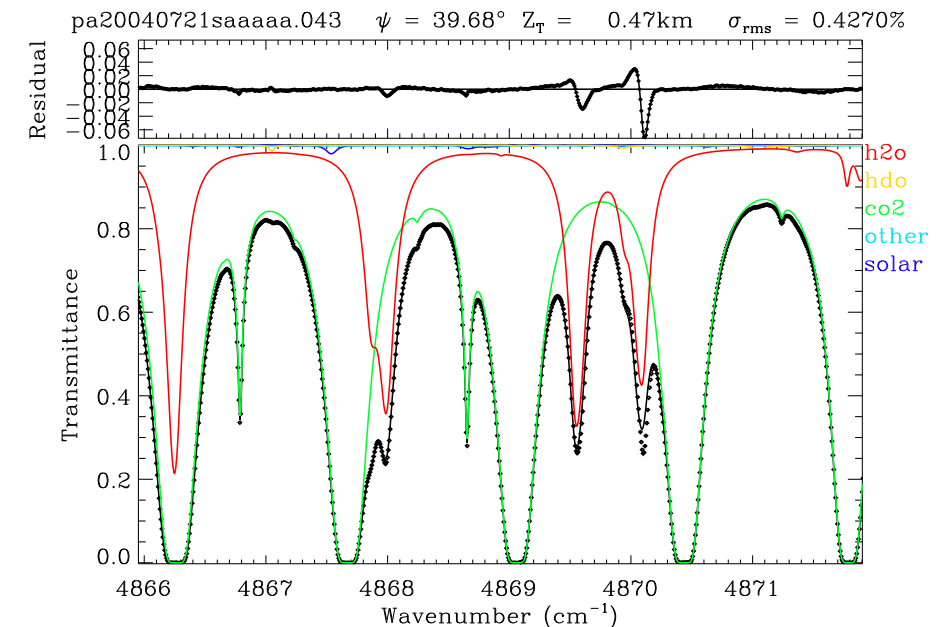
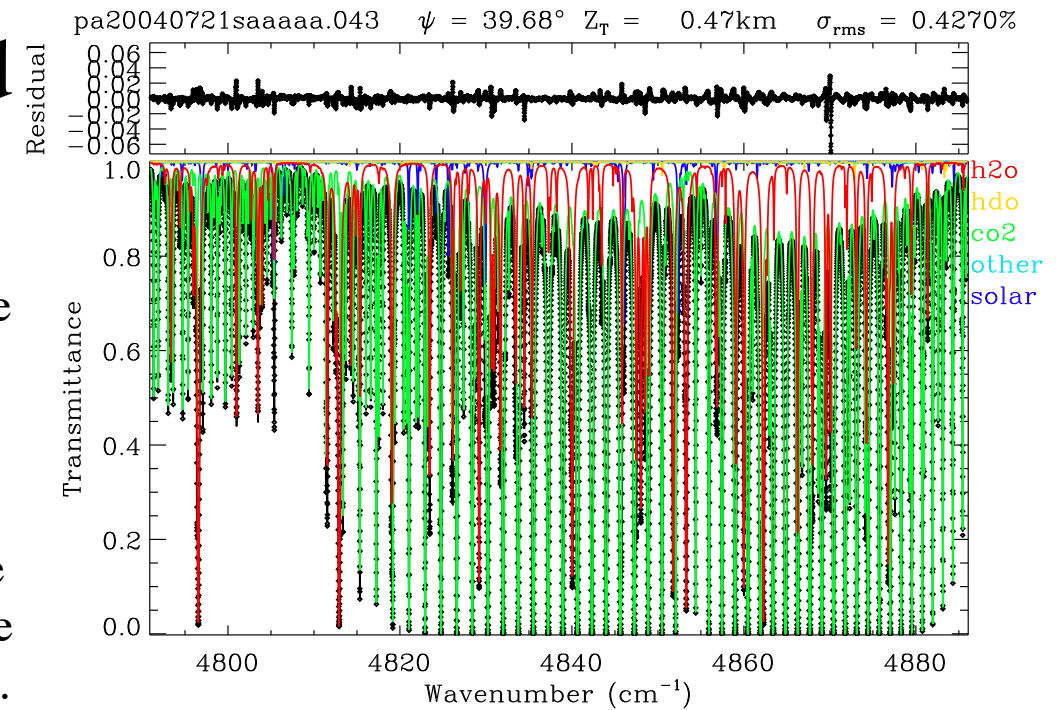
Under dry conditions, the dominant source of residuals in this region in the neglect of CO_2 LM. But since we use the same CO_2 linelist & lineshape in all cases, the small differences in the residuals are entirely due to H_2O .

That said, there is a danger in adjusting H_2O lines in this region when strong CO_2 absorption are overlapping them. You may adjust the H_2O line parameters to minimize CO_2 -related residuals (e.g., LM). This is why pure H_2O lab spectra are much preferred, if they exist under suitable conditions.

Table shows average % rms fitting residuals for all 27 spectra. HIT16 provides better fits than HIT12, which provides better fits than HIT08.

Window	HIT08	HIT12	ATM16	HIT16	ATM18
4795-4886	0.3744	0.3680	0.3628	0.3808	0.3581

Spectral fits with HIT16, Voigt lineshape and neglecting LM are shown (right) are for a low airmass Park Falls spectrum measured in July 2004. Large residuals seen for pairs of H_2O lines around 4869.6 and 4870.1 cm^{-1} . The maximum residuals in fits with ATM18 in this window in this spectrum is 1.4% (not shown) versus 7% for HIT16.

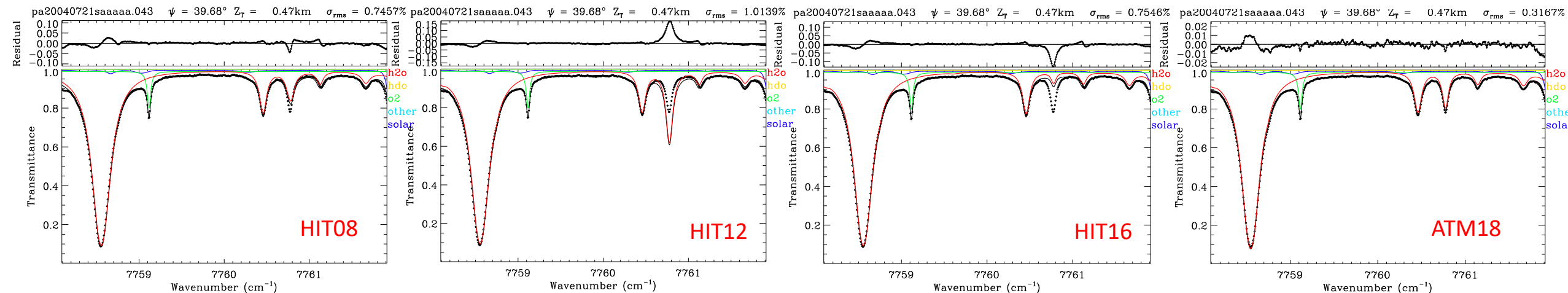
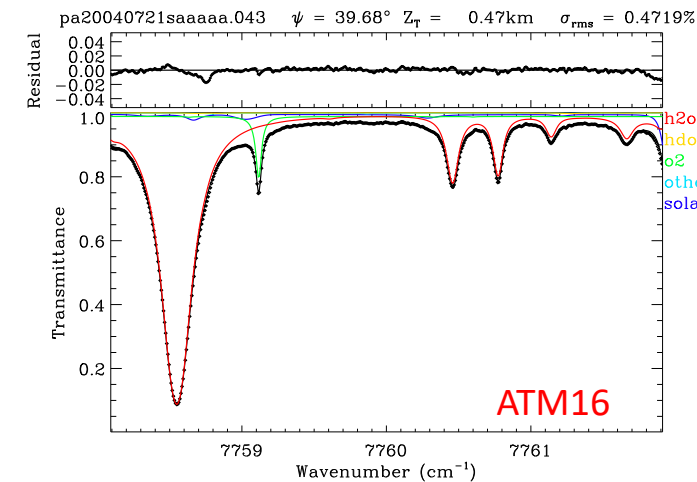


Fits to TCCON spectra over 7699 to 8046 cm^{-1} with different linelists

27 ground-based TCCON spectra were fitted over this region, which contains the O_2 band centered at 7885 cm^{-1} . The example shown is a low airmass summer spectrum from Park Falls. The H_2O line at $7760.7926 \text{ cm}^{-1}$ was pretty good in HIT08, was over-estimated in intensity by a factor 2 in HIT12, and under-estimated by a factor 4 in HIT16. This line has never been a problem in the ATM linelists.

Residual panels in spectral fits are y-auto-scaled: the HIT12 reaches 16%; ATM18 reaches only 2.5%. Table below shows average % rms fitting residuals over the 27 spectra.

Window	HIT08	HIT12	ATM16	HIT16	ATM18
7782	0.2161	0.2303	0.1738	0.2147	0.1638
7878	0.2541	0.2467	0.2366	0.2369	0.2303
7990	0.2296	0.2076	0.1787	0.1779	0.1753



Usefulness of types of spectra for Spectroscopy Evaluation

Type	Pros	Cons
Laboratory	<ul style="list-style-type: none"> - Well-known cell conditions (Length, T, P, VMR) - VMRs up to 1 are possible - Large isotopic enrichments possible 	<ul style="list-style-type: none"> • Dim source, so narrow spectral coverage or poor SNR • Isotopic composition often uncertain
Occultation MkIV Balloon	<ul style="list-style-type: none"> - Bright source (sun) allows simultaneous coverage 650-5650 cm^{-1} at high resolution - Wide range of P/T conditions & slant columns - Solar and instrumental features removed - Very long path lengths (~ 500 km) 	<ul style="list-style-type: none"> • Inhomogenous atmospheric path • No control over P, T, or VMR • Strong interferences from other gases • CO_2 used to determine tangent altitude so no info on absolute CO_2 amounts
Ground-based MkIV / TCCON	<ul style="list-style-type: none"> - Bright Source (sun) - Broad simultaneous coverage - Long path lengths (~ 100 km) - Sensitive to lineshape (e.g. width, shifts, LM) - Accurate knowledge of column from θ and P_s 	<ul style="list-style-type: none"> • Inhomogeneous atmospheric path • No Control over P/T or VMR • Wide regions blacked out: <ul style="list-style-type: none"> - 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 assayed, e.g., by mass spectrometry. For example, for atmospheric CO_2 , the $^{13}\text{C}/^{12}\text{C}$ ratio can be predicted anywhere to 0.1%. Atmospheric spectra contain no information on the SBHW.

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

Linelist Evaluation and Merging

Evaluation of linelists is relatively easy: Simply fit a bunch of high-quality spectra (lab and atmospheric) measured under well-known (and widely-ranging) conditions using different linelists for the target gas of interest (e.g. H₂O) but the same linelist for all the other gases. Compare the RMS fits and the consistency (window-to-window) of the retrieved gas amounts.

The latest linelist (e.g. HITRAN16) nearly always has the best results overall, but there are usually regions, conditions, or isotopes for which earlier linelists do better in terms of fitting quality or consistency of retrieved gas amounts. For example,

- Linelist A may be better than B below 4000 cm⁻¹, but worse above.
- Linelist A may be better than B for low-pressure lab water samples, but worse at higher pressures.
- Linelist A may be better than B for natural H₂O samples but worse for D-enriched lab samples.

How best to generate a new linelist that captures the best features of the predecessor linelists?

There's a range of possibilities ranging from:

- Determine which linelist is the best overall (usually the latest), and then use that one, making no attempt to capture the good parts/features of the other (usually earlier) linelists.
- Average the spectroscopic parameters from all linelists, using the fitting residuals as weights
- For each line, determine which linelist gives the best fits in its vicinity and copy those parameters to the new linelist. This is “cherry-picking” on a line-by-line basis. You could also cherry-pick on a band-by-band or window-by-window basis.

Spectroscopists don't like Cherry Picking because it damages the coherence and self-consistency of the original linelists.

Linelist users don't like or Cherry Picking or empirical adjustments because there is no easy way to cite the resulting linelist.

Summary and Conclusions

The use of multiple spectral datasets, including lab and atmospheric, provides a more stringent test of spectroscopic linelists than lab data alone, by exposing types of inadequacies in the spectroscopy that might not be apparent in the lab spectra. Atmospheric spectra provide long paths (up to 500 km from balloon) and access a wide range of temperatures and pressures (220-310 K). This is especially important for H₂O whose low VP at low temperatures limits the absorber amounts in a cell.

Although atmospheric spectra provide poor absolute accuracy (~10%) because we rely on meteorological models for H₂O column, their broad bandwidth (10,000 cm⁻¹ for TCCON) confers good band-to-band precision.

The ranking of the linelists in terms of RMS and VSF varies from window to window and from dataset to dataset. Lab spectra are generally at lower pressure than atmospheric and so the results are sensitive to line intensities, positions and, for pure gas samples, the SBHW. Ground-based observations are relatively more sensitive to lineshape (ABHW, pressure shifts, and Line Mixing) and to the weaker lines. Balloon measurements encounter cold temperatures and so provide a good validation of the T-dependent spectroscopic parameters.

HIT16 water vapor spectroscopy is better than any previous linelist, mainly due to the upgrade to the HDO, which now has many more weaker lines. So for humid conditions where weak lines are observable in atmospheric spectra, HIT16 does much better than previous linelists, which are missing those weak HDO lines. The stronger HDO line positions in HIT16 are empirically-based and therefore haven't changed much since HIT12.

Although the HIT16 HDO linelist is far better than any predecessor in terms of line positions and intensities, the widths and pressure shifts are sometimes in error by a factor 2 or more.

D-enriched lab spectra are very useful for distinguishing H₂O and HDO absorption lines, which cannot be determined from lab samples of un-enriched H₂O. Ditto for H₂¹⁸O and H₂¹⁷O. D-enrichment generates D₂O.

Supplemental Material

Additional Comments on Methodology

In retrospect, I made a mistake by using ATM16 as the starting point for ATM18. I should have used HIT16 instead, despite it providing worse RMS fits than HIT16. The reason is that the overall RMS fits are governed largely by the stronger lines, and these were better in ATM16 (thanks to earlier empirical adjustments). But the weak lines were much better in HIT16 than ATM16 (where many were missing or badly wrong), and the work involved in copying thousands of weak lines from HIT16 to ATM18 was greater than re-fixing relatively few strong lines based on ATM16. The end result is the same, but it would have been less work getting there.

In atmospheric composition measurement (NDACC/TCCON), the selected windows for retrieving gases of interest (e.g. CO₂, O₃, CO, N₂O, CH₄, HF, HCl, NO, NO₂, HNO₃) avoid regions that may black out under humid conditions, and hence avoid the stronger H₂O lines. So it is usually the weaker H₂O lines that are most relevant to NDACC, despite these lines not having much impact on the overall rms fitting residuals.

It is important that adjacent windows be consistent in terms of the retrieved H₂O amounts, otherwise fits will be degraded in broader regions than span multiple windows. For example, TCCON retrieves O₂ from a 240 cm⁻¹ wide region; 7765 to 8005 cm⁻¹. This spans 3 of the windows used in this H₂O evaluation.

```
7760.00 122.0  16 2 1 0  ncbf=15  fs  xo  cf  :  h2o  hdo  d2o
7878.00 114.0  16 2 1 0  ncbf=14  fs  xo  cf  :  h2o  hdo  d2o o2
7990.50 111.0  16 2 1 0  ncbf=14  fs  xo  cf  :  h2o  hdo  d2o
```

Without good window-to-window consistency, in parts of the fitted region, H₂O lines will be too weak causing dips in the residuals, while in other parts of the region the H₂O lines will be too strong, causing bumps in the residuals.

HITRAN 2016 paper: H₂O section (1/2)

2.1. H₂O (molecule 1)

The HITRAN2016 edition has undergone a substantial revision and expansion of the database of water vapor.

In HITRAN2012 [1] a very large expansion of the dynamic range of the line intensities for non-deuterated isotopologues of water vapor was implemented, thanks to the *ab initio* calculations from the BT2 line list [15] and, in selected spectral intervals, Lodi et al. [16] for the principal isotopologue and from Lodi and Tennyson for H₂¹⁸O and H₂¹⁷O [17]. The *ab initio* results were replaced with high-quality experimental or semi-empirical data wherever possible. In HITRAN2016, we have done a similar expansion of the dynamic (and spectral) range of the singly-deuterated isotopologues. In addition, the D₂¹⁶O isotopologue makes its debut in the database. D₂O has a very low natural abundance on Earth but is measurable in the atmospheres of other planets, notably Venus which has an enhanced deuterium content [18] raising the importance of D₂O. Also, for fitting D-enriched HDO laboratory spectra, a good D₂O line list is essential. For this reason, D₂O transitions have been included using a lower intensity cutoff of 10⁻³² cm⁻¹/(molecule·cm⁻²) once isotopic abundance is accounted for.

A dedicated paper detailing the update of the water-vapor dataset in HITRAN2016 along with atmospheric and laboratory validations is planned. Here we summarize only the most important details.

2.1.1. Line positions and intensities

This update is informed by the work of an International Union of Pure and Applied Chemistry (IUPAC) task group which produced systematic sets of empirical energy levels (and hence transition frequencies) for all the stable isotopologues of water [19–22]. In HITRAN2012 non-deuterated isotopologues already benefited from the availability of the IUPAC dataset of empirically-derived energy levels (and transition wavenumbers derived from them), and we extend this to the deuterated species here. However, some important caveats of that compilation have to be noted.

- The accuracy of some of the transitions generated from the IUPAC set may be inferior to individual high-accuracy experiments. Therefore, just like in HITRAN2012, we have given

preference to the experimental line position data from previous HITRAN editions if the uncertainty code was 5 (0.00001–0.0001 cm⁻¹) or higher. Comparisons with atmospheric spectra have shown that this was a correct choice.

- The datasets constructed in Refs. [19–22] do not include experimental data from the papers published after them. Quite a few new levels have become available over the years and some reassignments were in order (see for instance Mikhailenko et al. [23] and Liu et al. [24]). We partially updated the IUPAC datasets here. In particular, for the HD¹⁷O and HD¹⁸O species, levels from Kyuberis et al. [25] were used. from Mikhailenko et al. [23, 26] and Liu et al. [24] were used. Unfortunately, the IUPAC and Mikhailenko et al. quantum assignments often differ and complete matching would require a substantial amount of time in the future. An update of the IUPAC energy levels is in progress which will endeavor to resolve these assignment issues.

Many of the updated water vapor lines use line positions generated from the IUPAC energy levels and theoretical transition intensities, based on a high-accuracy *ab initio* dipole moment surface [16]. The methodology developed by Lodi and Tennyson [17] involves using several calculated line lists to identify reliable

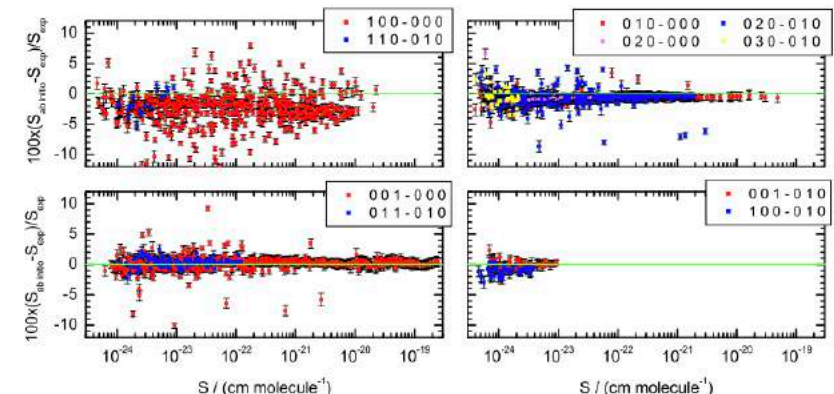


Fig. 1. Percentage *ab initio*-experimental vs. intensity for bands in the 1850–2280 cm⁻¹ and 2390–4000 cm⁻¹ regions of H₂¹⁶O. The error limit of the experimental data is 1% [31].

theoretical predictions. Lodi and Tennyson's intensity data were used in their entirety for H₂¹⁷O and H₂¹⁸O in HITRAN2012. Subsequent measurements and analysis by Regalia et al. [27,28] gave good agreement with these intensities. The present update relies heavily on Lodi-Tennyson style calculations for all isotopologues, supplemented with high-quality experimental data where available. This approach has proven to work well in application to atmospheric and laboratory spectra (see for instance Campargue et al. [29] and Ponomarev et al. [30]).

Given the reliance on these theoretical intensities, it is important to understand the systematic errors which *ab initio* calculations may show. In the same issue of this journal, Birk et al. [31] give an extensive intercomparison of *ab initio* calculations with high-quality experimental data. The experimental intensities used in the intercomparison are all included in the HITRAN2016 update. Much of the data show agreement between *ab initio* and experiment within 2%. However, for some bands, notably those involving excitation of the stretching modes, there are larger offsets of up to 8% which can be attributed to systematic errors in the *ab initio* dipole moment calculations. Furthermore, in the ν_1 fundamental, differences show a systematic dependence on wavenumber, ΔK_a , and ΔJ . Other vibrational bands involving the changes in ν_1 quanta also showed larger systematic errors. The theoretical calculations also have problems in predicting local resonances accurately, although these resonances start to appear

more frequently in the NIR part of the spectrum. The information on systematic differences between the *ab initio* theory and experiment for H₂¹⁶O can be used to estimate uncertainties for minor isotopologues for which no high-accuracy experimental intensities are available.

An important finding of this work is that for an individual vibrational band an intensity-independent offset appears in graphical representations of differences between experiment and theory vs. intensity. This information can be used to validate experiment as well as theory. The constant offset can also be used to predict weak line intensities where experimental data are unavailable by scaling *ab initio* values.

Fig. 1 compares theory and experiment for the 1850–2280 cm⁻¹ and 2390–4000 cm⁻¹ regions. The agreements for the fundamental bands (001)-(000), (010)-(000), the overtone (020)-(000), the hot bands (030)-(010), (020)-(010), (011)-(010), (001-010), (100)-(010) are all excellent, mainly within 2%. The bands involving changes in the ν_1 quanta, (100)-(000) and (110)-(010), however, show large scatter and an offset around -2%. The scatter includes the systematic differences of +5% to -13%, which becomes obvious when plotting the differences against wavenumber or lower-state energy and color coding ΔK_a and/or ΔJ [31].

The 2016 update provides comprehensive line lists for the six main isotopologues of water: H₂¹⁶O, H₂¹⁸O, H₂¹⁷O, HD¹⁶O, HD¹⁸O and HD¹⁷O, as well as the newly-added isotopologue D₂¹⁶O. The

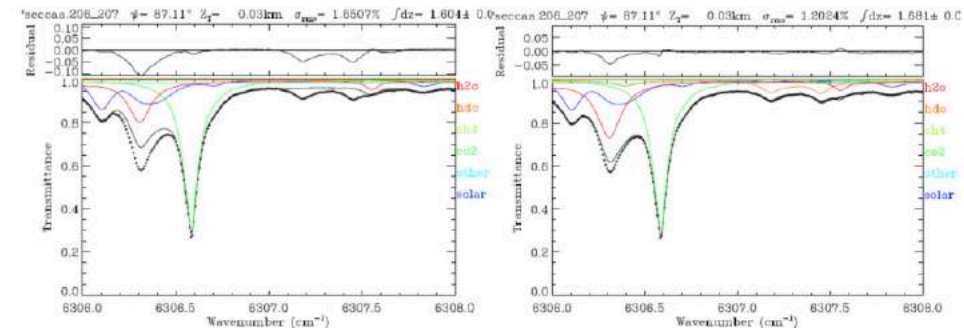


Fig. 2. Spectral fits to the high-H₂O Darwin (Australia) spectrum using the HITRAN2012 line list (left panel) and the new line list (right panel). The improvement in the residuals is due to the addition of HDO lines (three in this plot) denoted by the orange trace. Note the slight change of scale in the residuals on the left and right sides.

HITRAN 2016 paper: H₂O section (2/2)

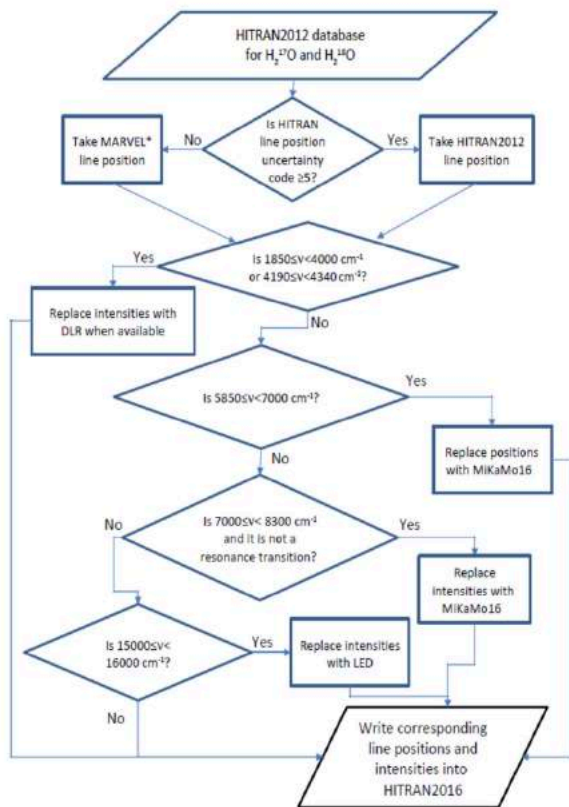


Fig. 4. Flow diagram for the construction of line positions and intensities for H₂¹⁶O and H₂¹⁷O. The asterisks indicate that the IUPAC levels were slightly updated in this work. LED refers to the experimental data from Ref. [26]. See text for details.

intensity data are described below.

2.1.1.1. DLR experiments. An extensive work in the spectral ranges 1850–2280 cm⁻¹ and 2390–4000 cm⁻¹ was carried out by Loos et al. [33,34] and in the range 4190–4340 cm⁻¹ by Birk et al. [31]. In Fig. 3 we will refer to these, and the experiments from the same authors that are already in HITRAN, as DLR. In Loos et al. [33,34] line positions, intensities, self- and air-broadened line-shape parameters, their temperature dependence as well as Rosenkranz line coupling coefficients were retrieved from numerous Fourier-Transform transmittance measurements of self- and air-broadened water vapor at 296 K as well as air-broadened water vapor measurements at low and high temperatures. During the analysis, a large effort was undertaken to give consolidated error bars. In the analysis, a multispectrum fitting approach applying a quadratic speed-dependent hard collision model based on the Hartmann-Tran profile [7,8] and extended to account for line mixing in the Rosenkranz first order perturbation approximation [35] was used. Line positions, intensities and self-broadening, self-speed-

dependence and self-shift parameters, as well as in some cases self-line coupling coefficients, were retrieved from pure water-vapor measurements of total pressures up to 20 mbar. Air-broadening, speed-dependence, pressure shift parameters, Dicke narrowing and line-mixing coefficients as well as temperature-dependence parameters were obtained from air-broadened measurements at ambient temperature and total pressures from 30 to 1000 mbar as well as low and high-temperature measurements at 100 mbar total pressure. The intensities of lines with retrieved line parameters range from 3×10^{-26} to 3×10^{-19} cm²/(molecule cm²). In the 4190–4340 cm⁻¹ region, several Fourier-Transform transmittance spectra of pure and air-broadened water vapor at 296 K as well as low and high temperatures were measured and analyzed by Birk et al. [31]. These measurements were dedicated to water vapor parameters to be used in TROPOMI/SP-P [36] retrievals. The analysis was also based on a multispectrum fit using the HT profile. Line positions, intensities, self- and air-broadened line-shape parameters including speed-dependence and Dicke narrowing parameters as well as their temperature dependence

use of variational calculations to provide the underlying line lists guarantees the completeness of the lists for the intensity cutoff employed. This completeness leads to a significant expansion of the number of important transitions of the deuterated isotopologues into the NIR (near-infrared) region. These line lists are presented and analyzed in articles published in this issue [25]. Fig. 2 shows an example where atmospheric retrieval from a Total Carbon Column Observing Network (TCCON) [32] spectrum

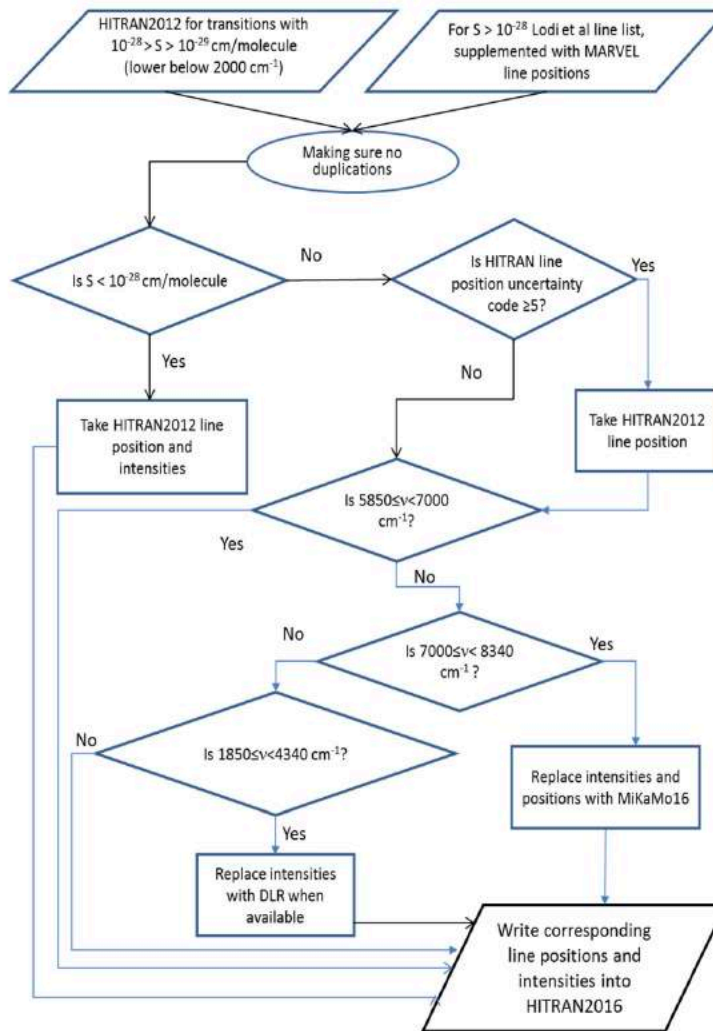


Fig. 3. Flow diagram for the construction of line positions and intensities for H₂¹⁶O below 8340 cm⁻¹. See text for details.

benefits from inclusion of HDO lines in the NIR. It is also important to note a significant increase in coverage for HD¹⁷O and HD¹⁸O which were poorly represented in previous HITRAN releases.

The variational (calculated) intensities described above were then replaced with intensities from available high-quality experiments, many of which were already in the HITRAN2012 database. The details will be given in a dedicated paper, but some of the experiments that represent a substantial bulk of new experiments

were retrieved in the analysis.

2.1.1.2. MiKaMo16 experimental set. Recently, Mikhailenko et al. [23] have compiled a database of measurements in 5850–8340 cm⁻¹ region from the Grenoble-Tomsk collaboration reported in Refs. [37–41] supplemented with measurements from Sironneau and Hodges [42] and for strong lines from Toth [43] and for very weak lines from ab initio work of Partridge and Schwenke [44,45]. This database is here referred to as MiKaMo16. Unfortunately, due to some differences in assignments with the variational Lodi-Tennyson-like line lists described above, it was hard to take full advantage of this line list. However, tests against TCCON spectra have shown that this line list is superior to any other line list in the region of 7000–8340 cm⁻¹. It was therefore used in HITRAN2016 in that region as is almost everywhere in this spectral region. This, however, creates some consistency issues with assignments of energy levels throughout the database, but substantial improvement of atmospheric retrievals justifies this approach.

As examples of the complexity of the update, Figs. 3 and 4 show the flow diagrams of the construction of position-intensity line lists of the principal isotopologue (only up to 8340 cm⁻¹ for simplicity) and for H₂¹⁸O/H₂¹⁷O respectively. Note that the line list from Lodi et al. [16] for the principal isotopologue is limited to 10⁻²⁸ cm²/(molecule cm²); therefore intensities for weaker lines still originate from HITRAN2012/BT2.

Whenever a rotational quantum number could not be determined unambiguously, the index of symmetry (1, 2, 3, and 4 as defined in the BT2 work [15]) accompanied with a negative sign was used. Note that 1 and 2 indicate para states, whereas 3 and 4 indicate ortho states. For the case of unassigned vibrational quanta, a “-2” label has been adopted.

2.1.2. Line-shape parameters for the H₂¹⁶O, H₂¹⁸O and H₂¹⁷O line lists

The air-broadened half widths (in Voigt profile formalism) for the first three isotopologues of water in HITRAN2008 and, with some modifications, in HITRAN2012 were derived using the procedure “Diet” described by Gordon et al. [46]. This procedure continues to be used in this edition with experimental outliers identified and removed. One notable example is the removal of a large portion of the half widths measured in Jenouvrier et al. [47] from the experimental data that feeds the “Diet”. Indeed, some of the recent evaluations of atmospheric spectra showed that many data in Jenouvrier et al. are questionable (see for instance Armante et al. [48]). However, although the removal of this extensive dataset from the updated Gordon et al. [46] procedure, as was done in the GEISA database [49] and the MiKaMo16 line list, have improved the residuals in some of the lines, some of the residuals became worse than in HITRAN2012. The main problem here is due to the fact that, while there are definite outliers in the values reported in Jenouvrier et al. [47], it is the most extensive study of the air-broadened line widths in the 4200–6600 cm⁻¹ region. In fact, for many transitions no other measurements exist. Removal of these measurements from the experimental dataset has invoked the use of the approximations of different levels (some are very crude) from the work of Jacquemart et al. [50] for many of the transitions. The only solution in this case was to filter the Jenouvrier et al. dataset. This was done through direct comparisons with the TCCON spectra for some of the transitions and cross-comparison with similar rotational transitions but from other bands, accounting for vibrational dependence from Eq. (1) of Jacquemart et al. [50]. After this filtering procedure, about 1400 measurements (out of almost 5000) from Jenouvrier et al. [47] were retained.

It is also important to emphasize that measurements by Birk and Wagner [51] in the ν₂ band region were given a priority and

were written into the database directly as was done in HITRAN2012.

Since the modeling of water vapor absorption lines using a Voigt model is no longer sufficient in many applications and the HT profile provides considerably higher accuracy, experimental HT profile parameters of Birk et al. and Loos et al. [31,33,34] in the spectral ranges 1850–2280 cm⁻¹, 2390–4000 cm⁻¹ and 4190–4340 cm⁻¹ were accommodated wherever possible. The experiments and analysis procedures applied are described briefly in Section 2.1.1.

2.1.3. Line-shape parameters for the deuterated isotopologues

The addition of the line-shape parameters for D₂O follows a similar procedure as the other isotopologues; however, the available data for air-broadening of D₂O are considerably less complete than for the other isotopologues. There are some measurements for air-broadening of D₂O [52–55] as well as Complex Robert-Bonamy calculations [56]. These data amount to several hundred lines from the measured data and just over 550 lines from the theoretical calculations. Using these data, average values as a function of J' were determined and extrapolation to $J' = 50$ was done using a third-order polynomial fit.

Using these data, the half widths and some line shifts have been added to the D₂O lines in the database using a priority scheme. First, the measured values were added; if the measurement did not exist for the transition in question, the theoretical value was used. If neither exists, the J' -averaged value was used. This algorithm allowed half widths to be added to all D₂O lines in the HITRAN database.

For self-broadening parameters of D₂O, the work of Gamache et al. [57] was used. These data are an extensive set of calculations for D₂O rotational band transitions extended to 6536 lines using the partner transition rule [58]. Using these data, a set of J' -averaged half widths was determined for $J' = 0$ –50 [57]. These data have been added to the algorithm that adds line-shape data to the water lines in the HITRAN database.

2.1.4. Future plans

Recently, Lampel et al. [59] have identified important absorption features in atmospheric spectra due to water vibration-rotation transitions in the near ultraviolet around 363 nm. Laboratory measurements of water vibration-rotation transition intensities only extend to 25,470 cm⁻¹ [393 nm] [60] although multiphoton spectra do provide some information on energy levels in the region of question [61,62]. Future updates will look to extend coverage to water absorption in the near ultraviolet; these data are important for upcoming satellite missions, including TEMPO [63] which will record spectra at these wavelengths.

At the intensity limit assumed for D₂¹⁶O, the even rarer isotopologues D₂¹⁷O and D₂¹⁸O should also be visible. Line lists for these two species have also been prepared as part of the work on deuterated water in Ref. [25]; these data will be added in the forthcoming updates to HITRAN2016.

We will be populating the database of the HT line-shape parameters after evaluating existing literature values and new measurements as they become available.

2.2. CO₂ (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

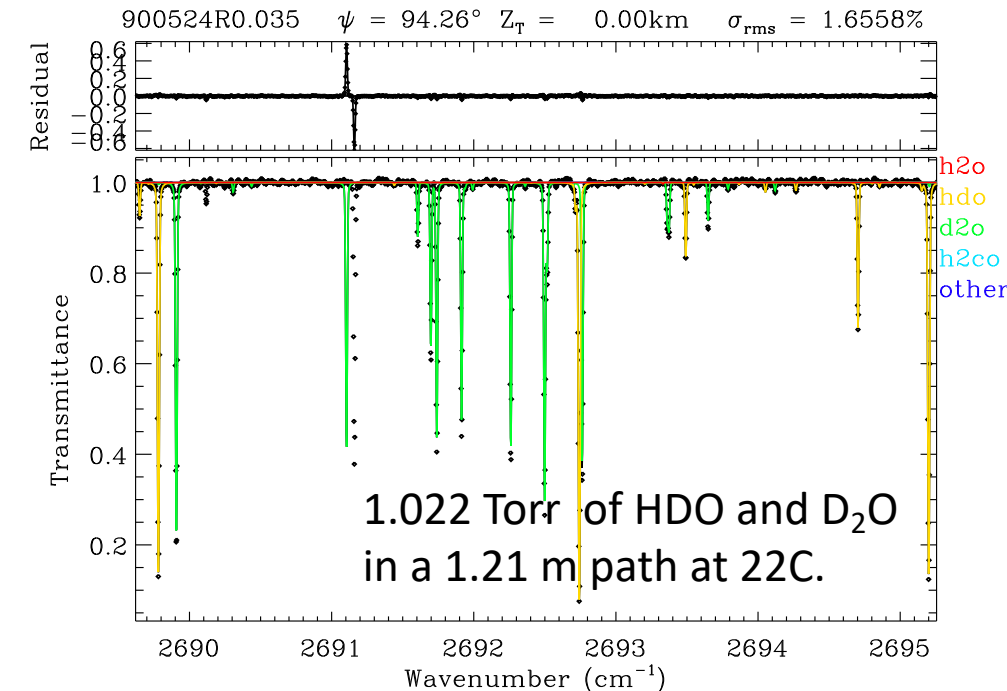
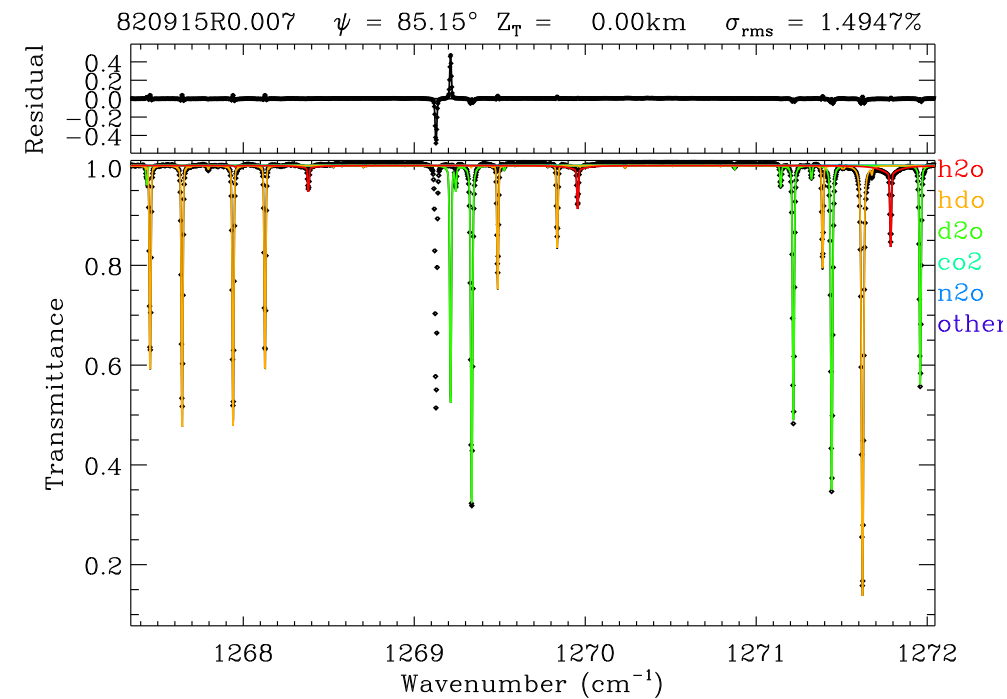
D₂O – Doubly Deuterated Water Vapor

In the Earth's atmosphere D₂¹⁶O is ~8 orders of magnitude less abundant than H₂O and therefore can never be seen. So why the sudden appearance of D₂O lines in HITRAN16 with 24391 lines covering 0 to 12,797 cm⁻¹ ?

We need D-enriched lab spectra of HDO in order to characterize the weaker HDO lines, that can be seen in high-airmass atmospheric spectra, but not in lab spectra of un-enriched water vapor. In D-enriched lab spectra of HDO, the D₂O lines can be completely saturated. Without correctly representing the D₂O absorptions, the weak HDO lines cannot be accurately characterized.

Figures (right) show examples of spectral fits to a D-enriched Kitt Peak lab spectra using HIT16. Green shows D₂O absorptions; orange shows HDO. The position of D₂O line at 1269.21 cm⁻¹ is off by nearly 0.1 cm⁻¹ and the 2691.16 cm⁻¹ D₂O line position is off by 0.055 cm⁻¹.

In 2006 Bob Toth generated a D₂O linelist covering 2198 to 4255 cm⁻¹, presumably to facilitate analysis of D-enriched lab spectra. Toth's D₂O linelist still gives slightly better fits to this window than the HIT16 linelist, even after fixing dozens of the largest residuals in the latter. We nevertheless opted to use the (corrected) HIT16 D₂O linelist due to its broader wavenumber coverage and much improved E'' values.



Appendix: D₂O Linelist Anomalies

The HITRAN 2016 D₂O linelist was evaluated using D-enriched Kitt Peak lab spectra, acquired primarily for HDO purposes.

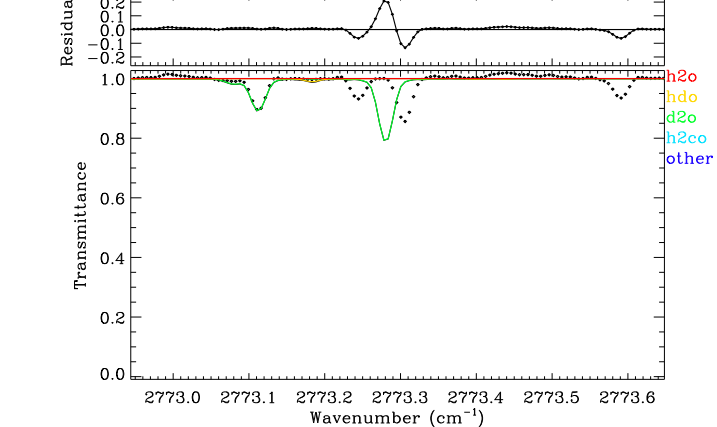
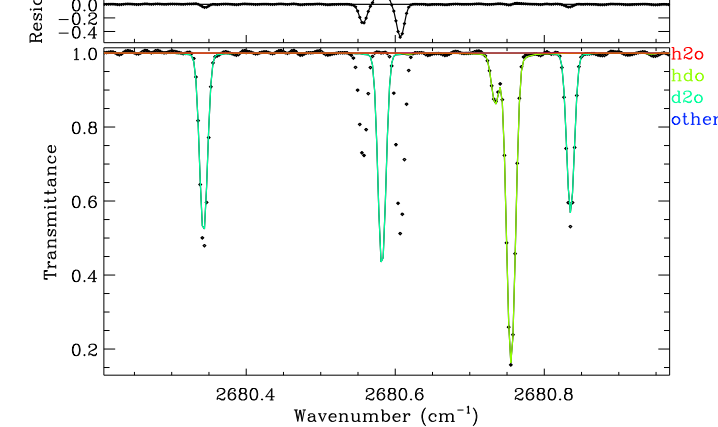
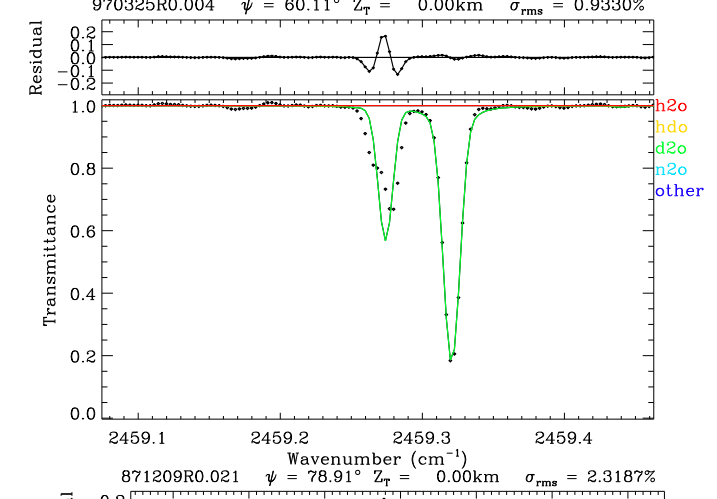
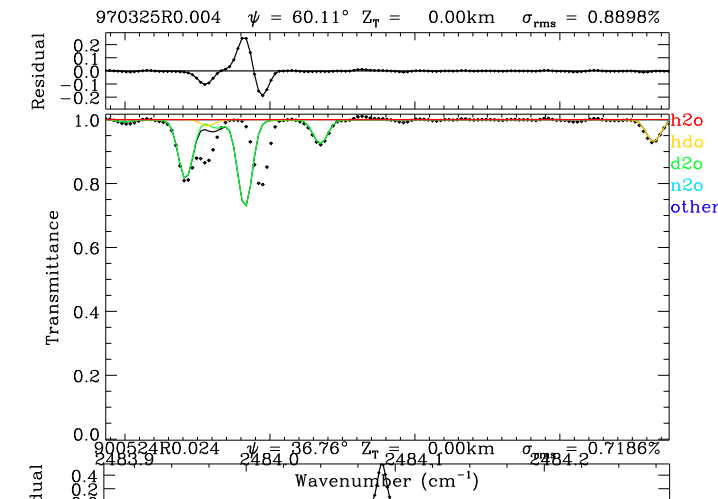
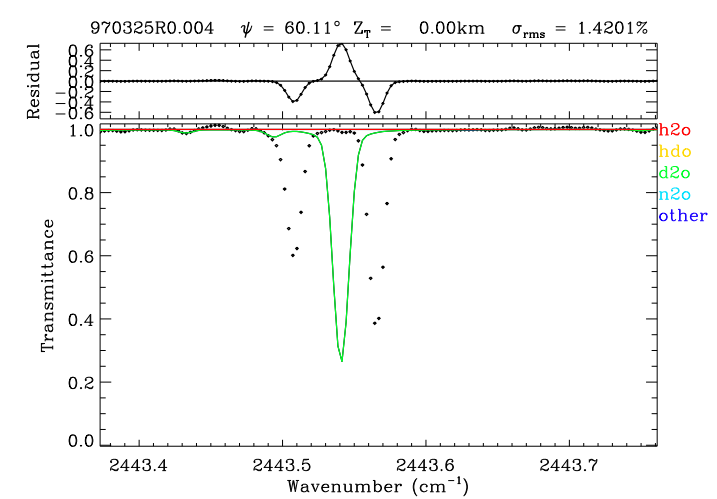
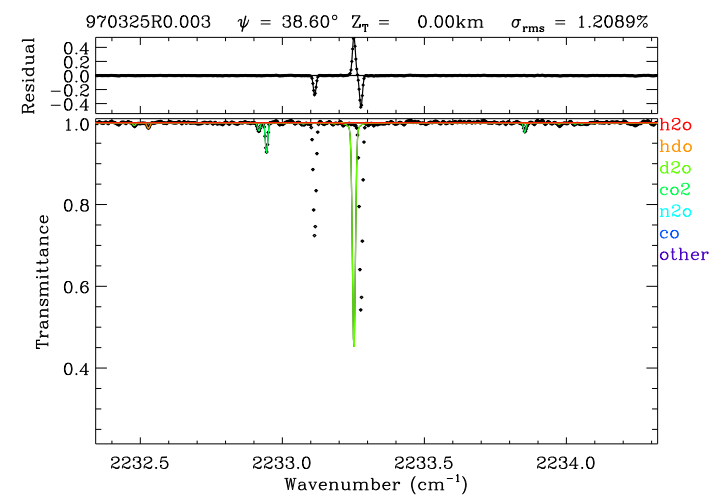
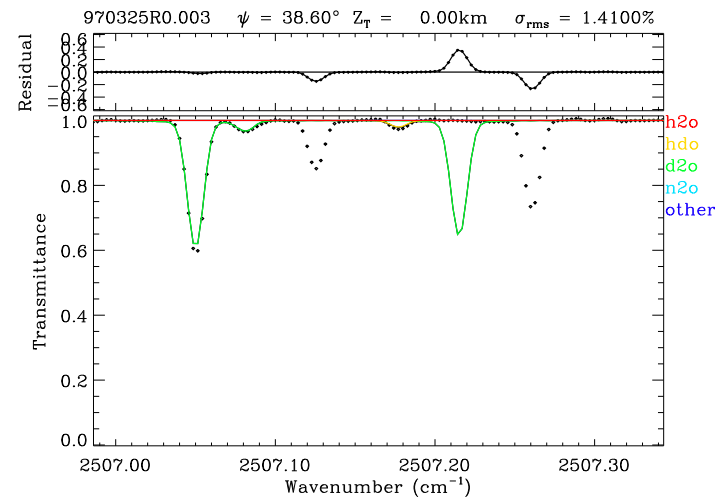
Large residuals in D₂O bands, which scale with the other D₂O lines, fall into one of four categories:

- 1) Position error of singlet line. This is the easiest to fix. Simply adjust the line position.
- 2) Doublet (two lines with identical positions and with intensities in a 2:1 ratio) is not split in HIRAN16. This is fixed by moving one line to the left and the other to the right.
- 3) Missing absorption line with no D₂O lines in the vicinity. Added fake line labeled “GCT fudge”.
- 4) Found 293 instances of duplicated lines: two identical lines (including quantum numbers) causing a factor 2 too much absorption. Deleting one of the two lines completely fixes the bump in the residual.

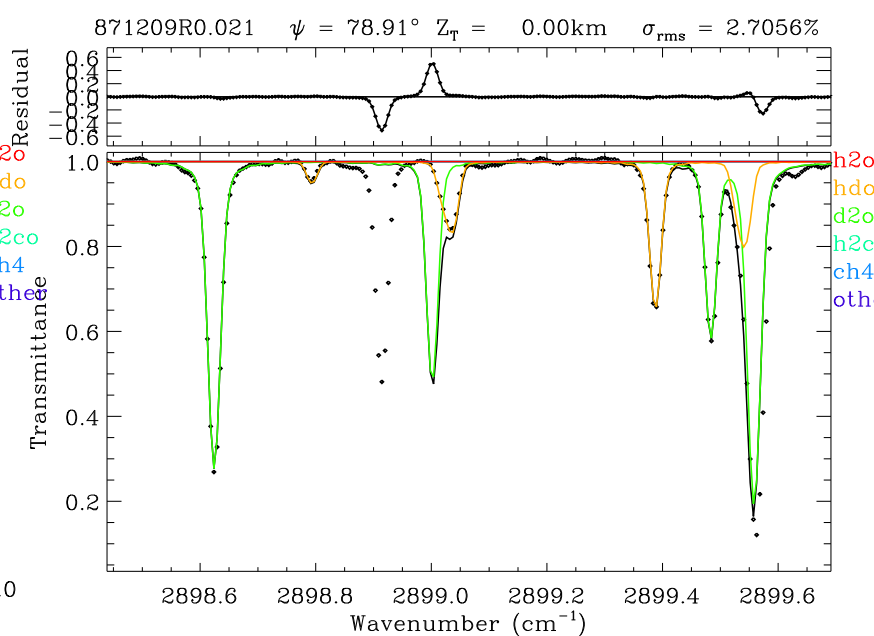
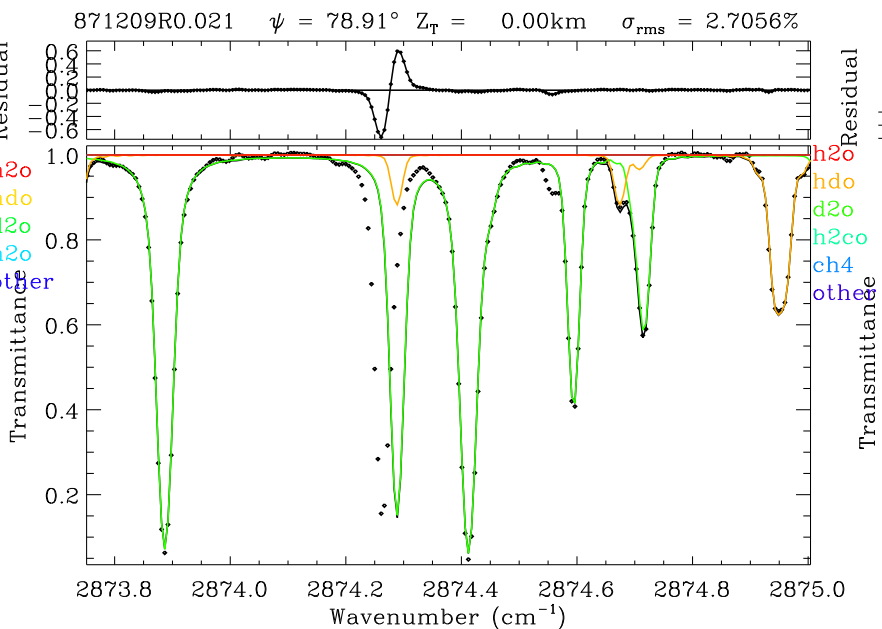
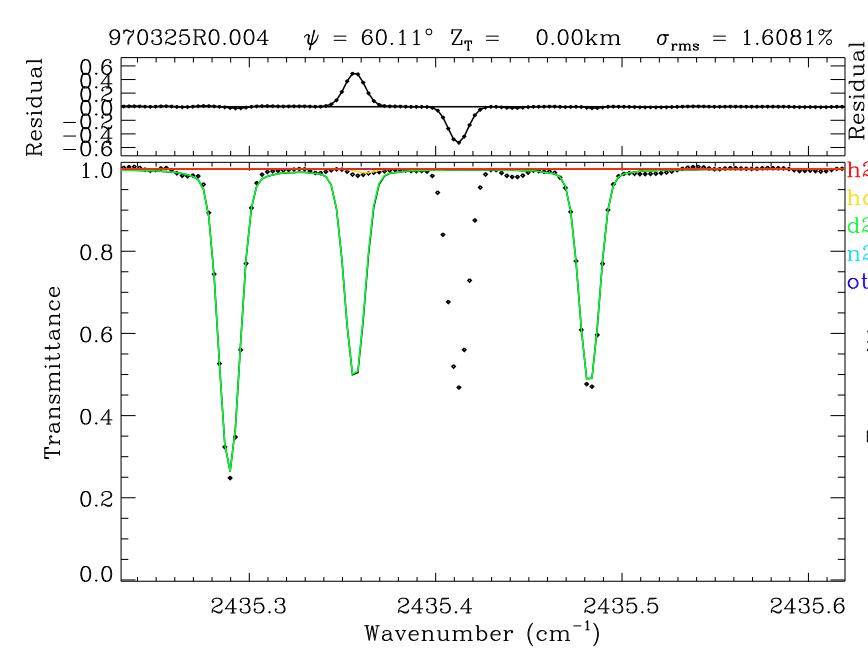
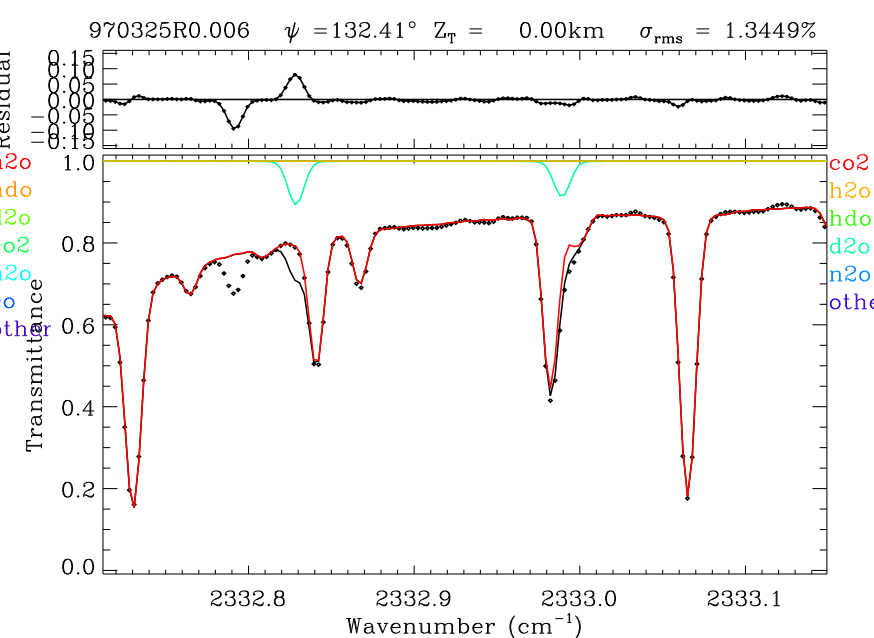
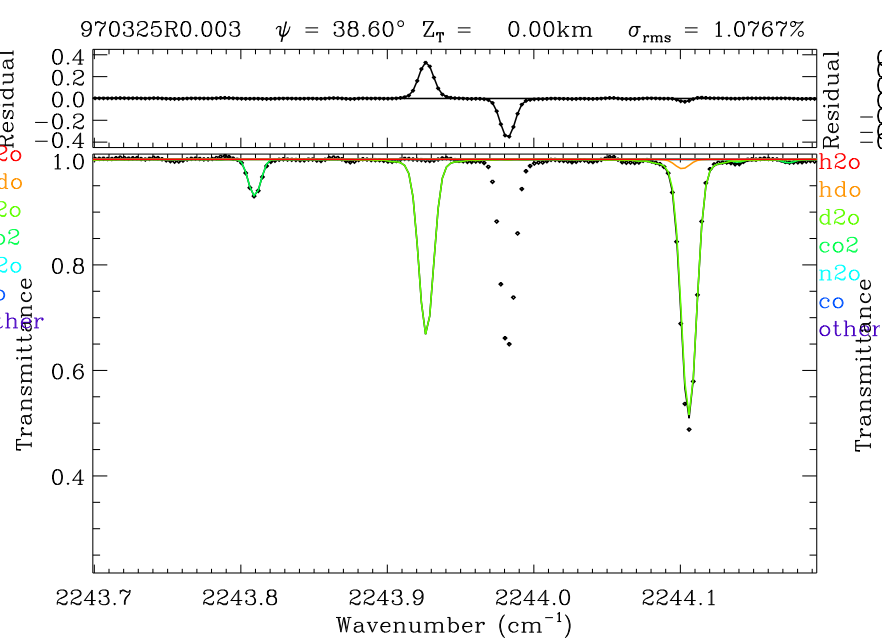
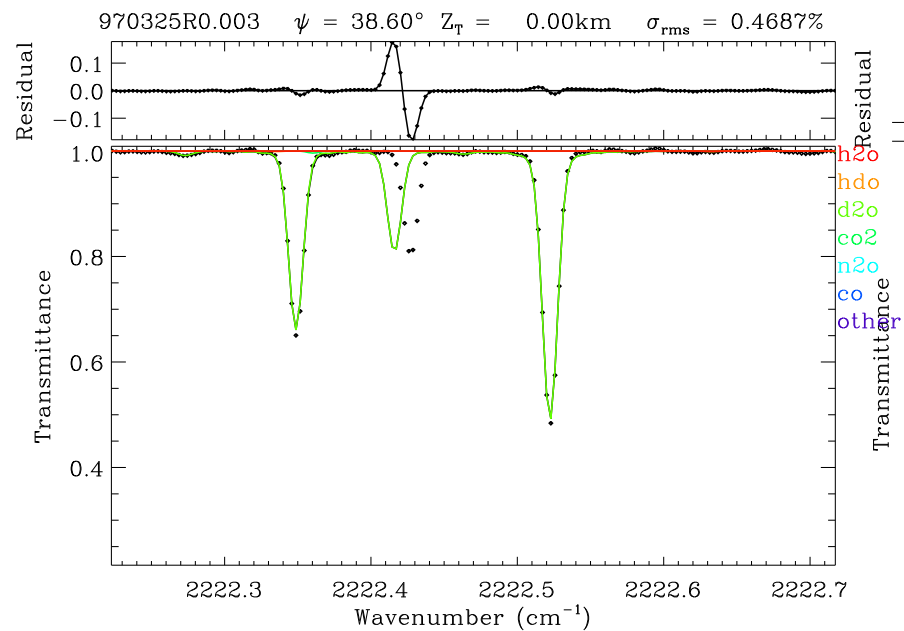
Examples of D₂O-related residuals are shown in the following slides.

8 examples of D₂O doublets with intensities in a 2:1 ratio, which have the same position in the HIT16 linelist, but not in reality.

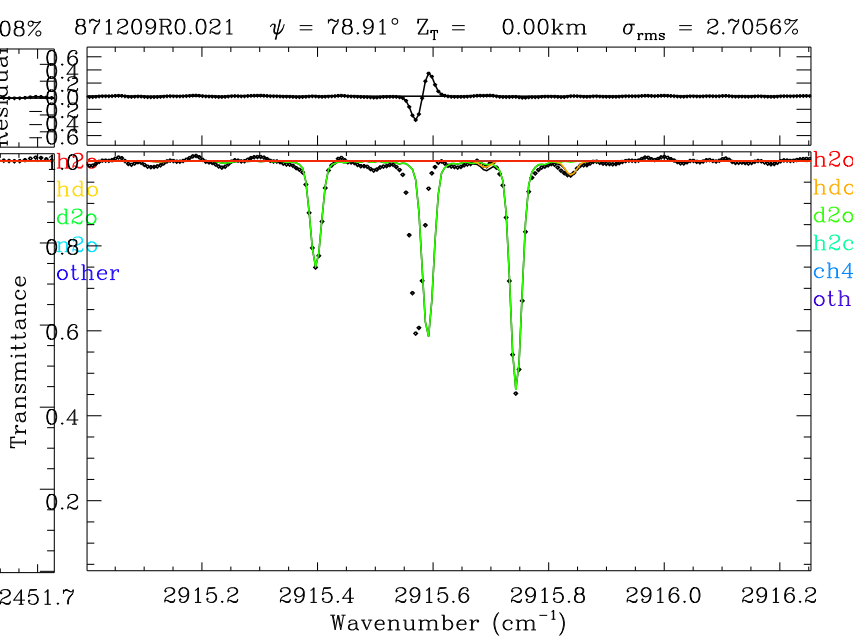
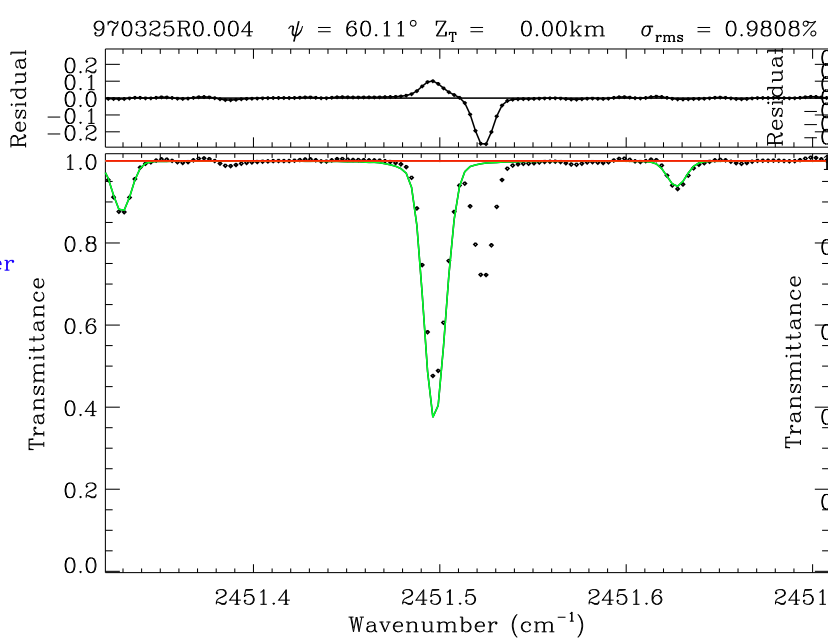
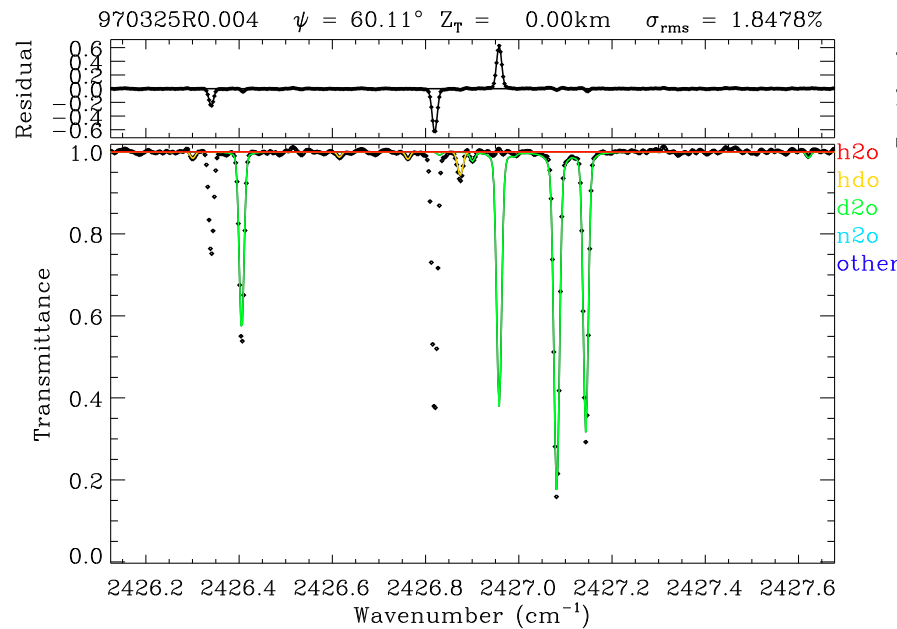
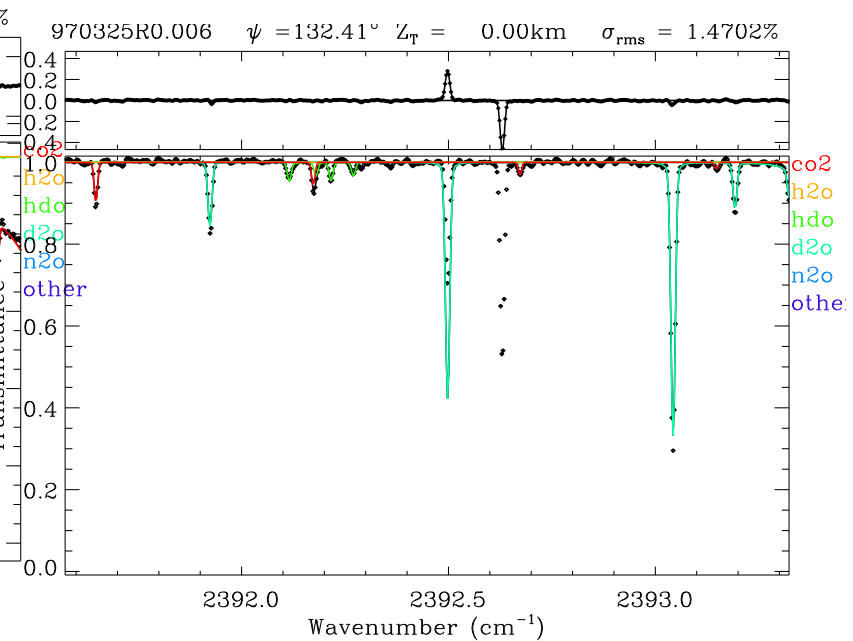
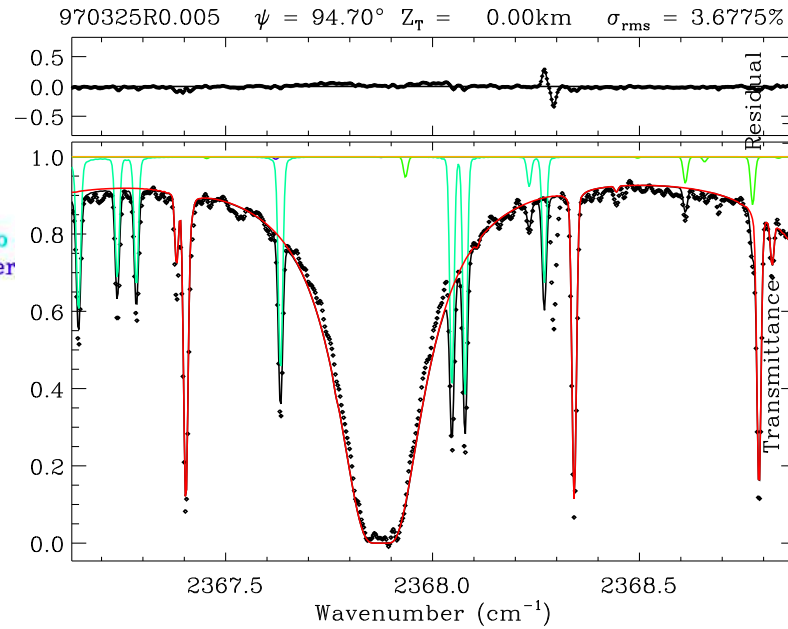
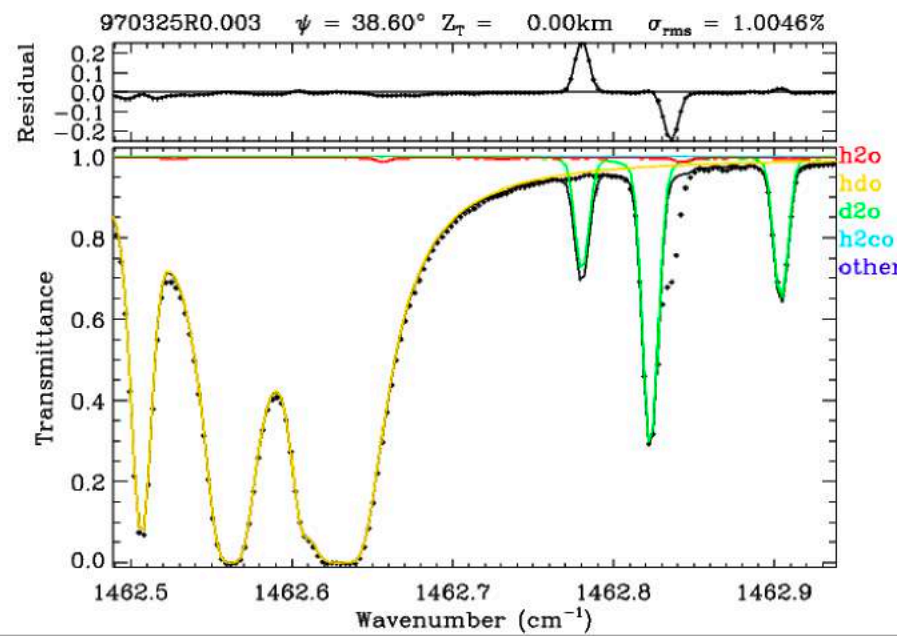
In each case the stronger of the two lines has to be moved to the right, and the weaker line to the left



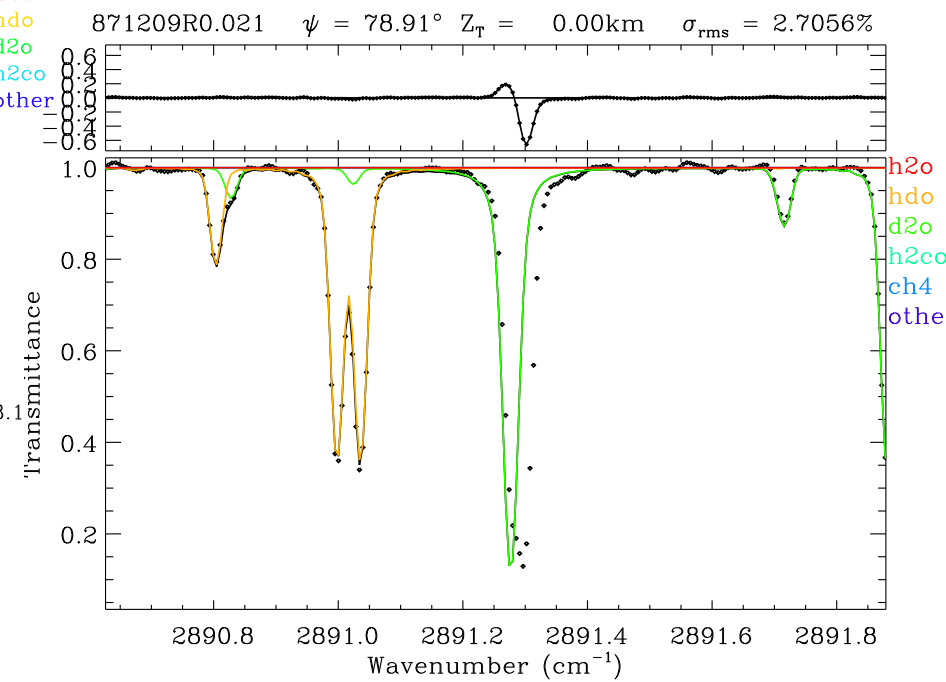
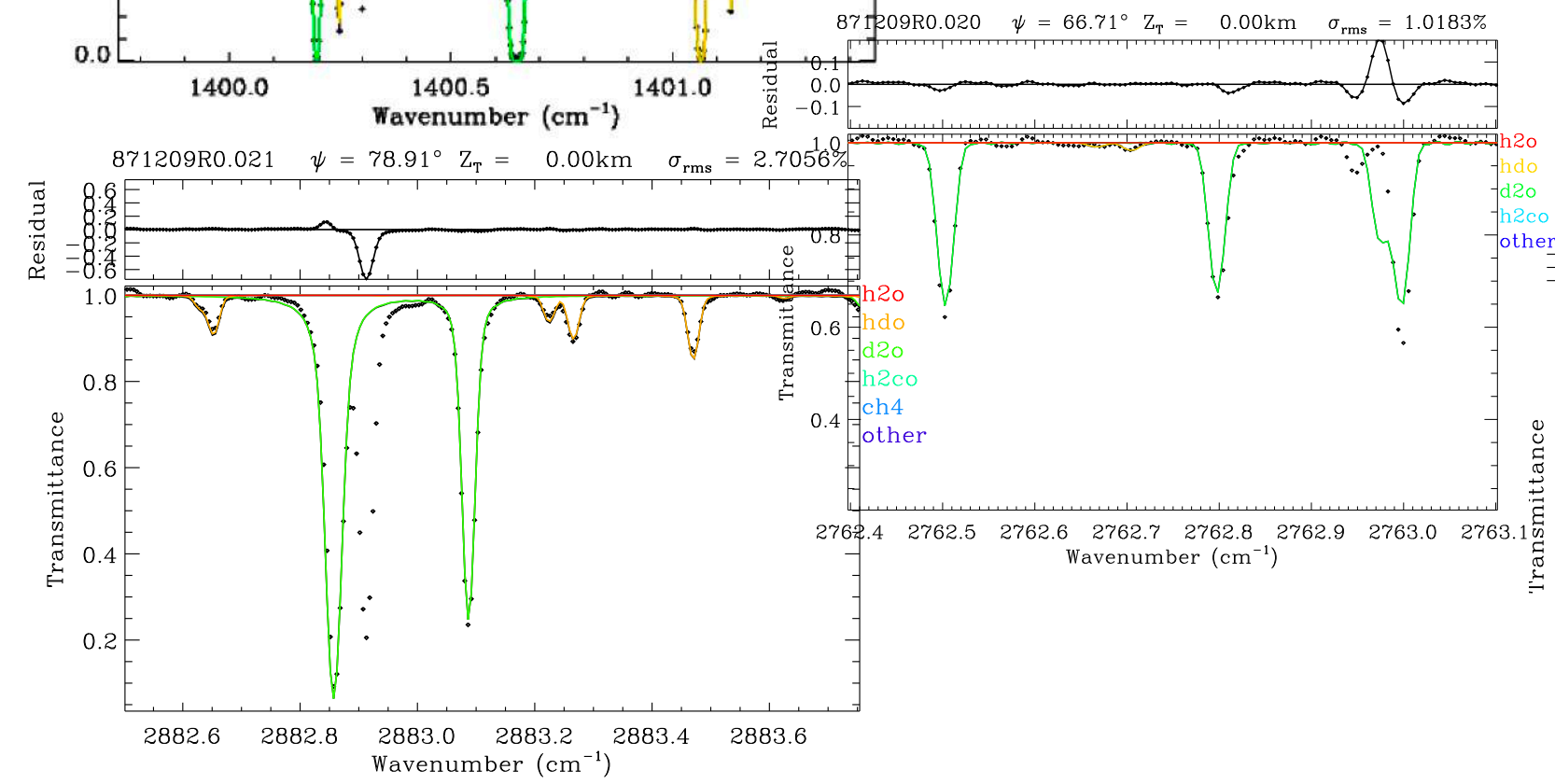
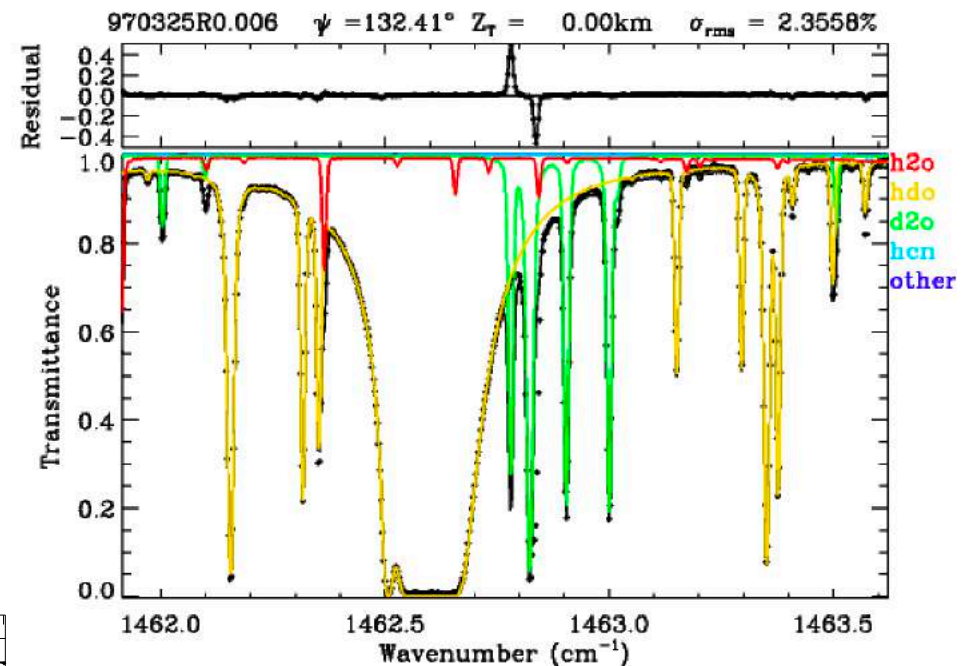
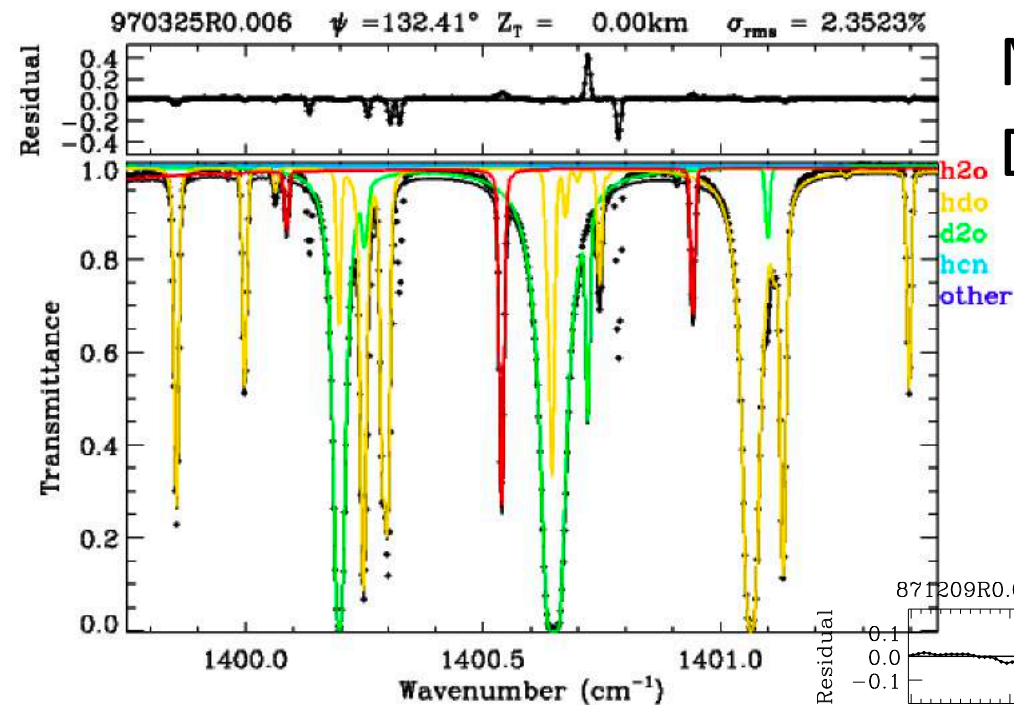
Six Examples of D₂O singlet position errors

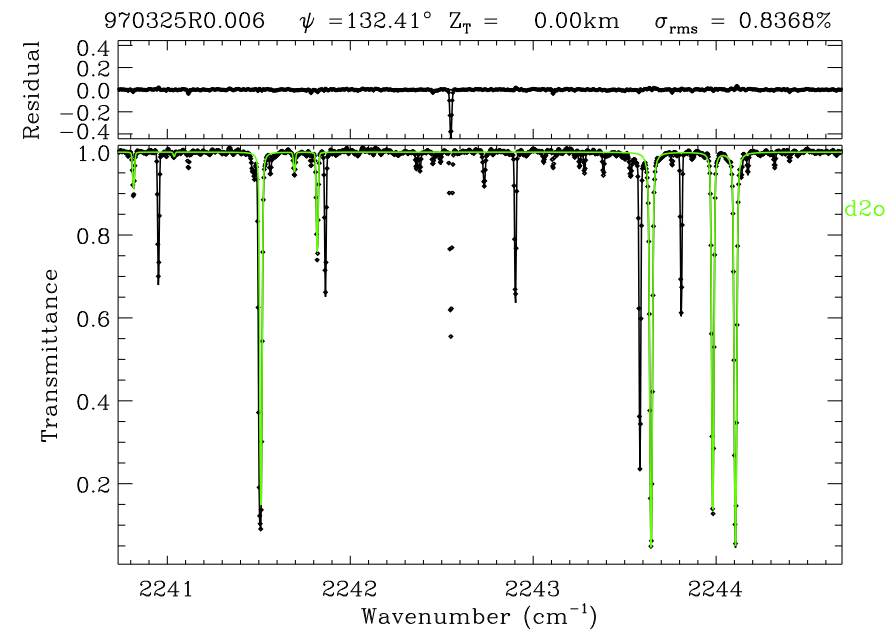


Six More Examples of singlet position errors

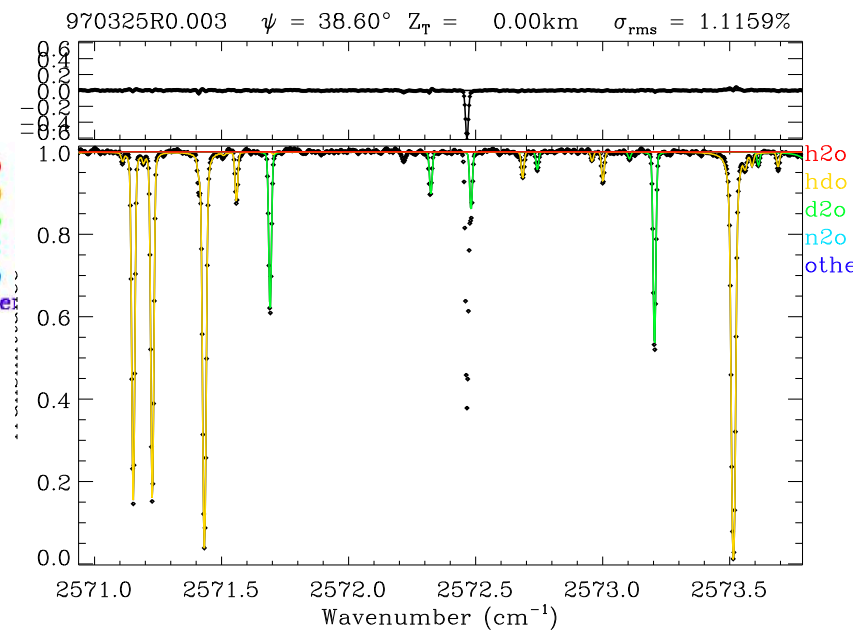
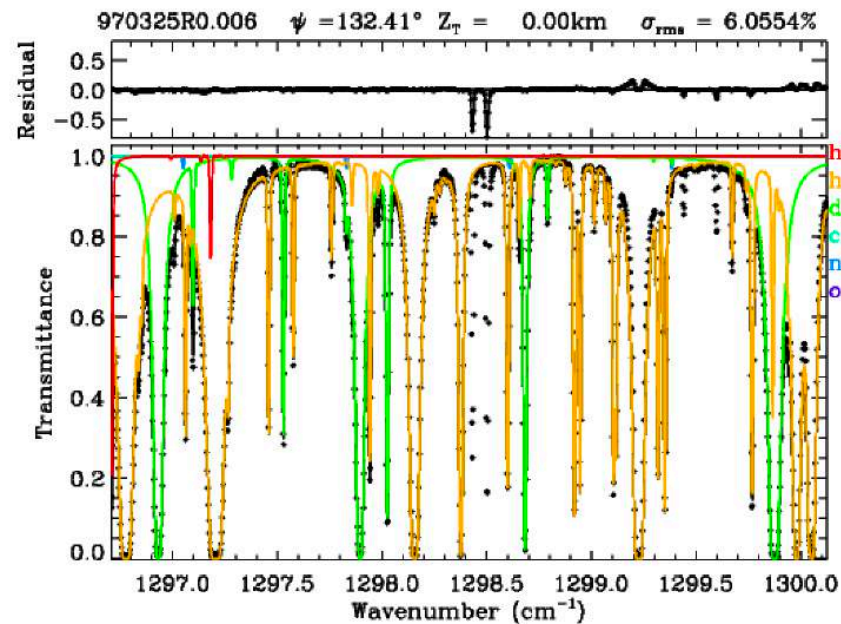
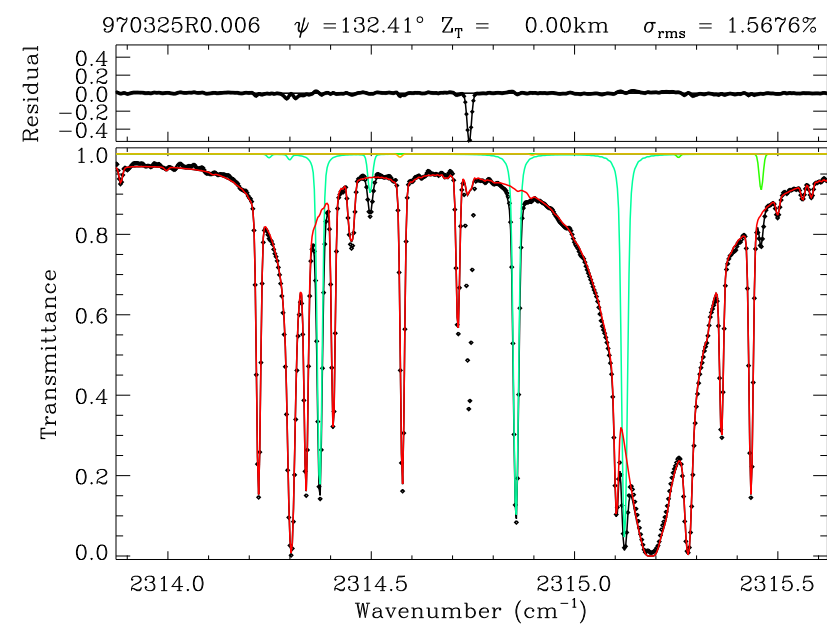
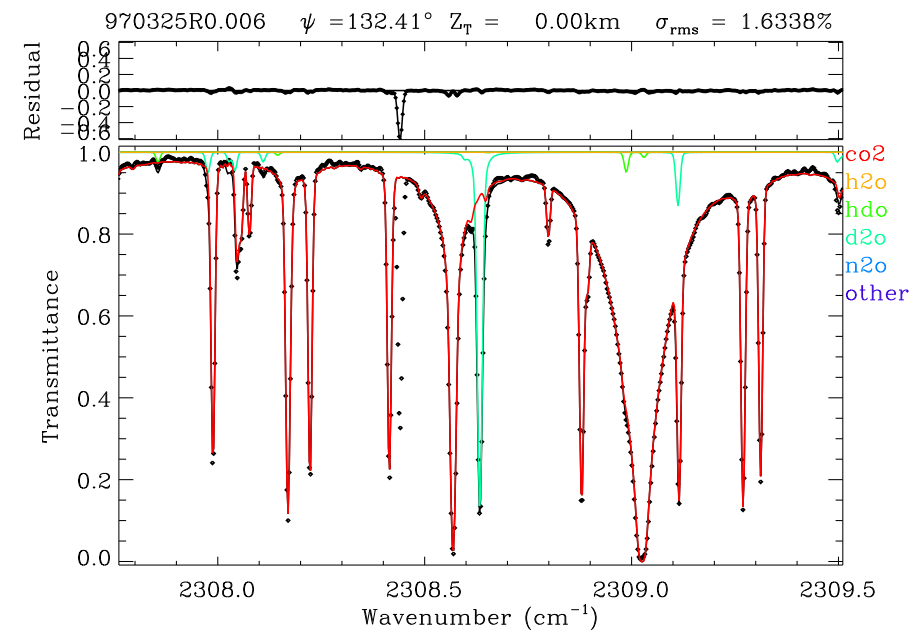


More examples of D₂O position errors

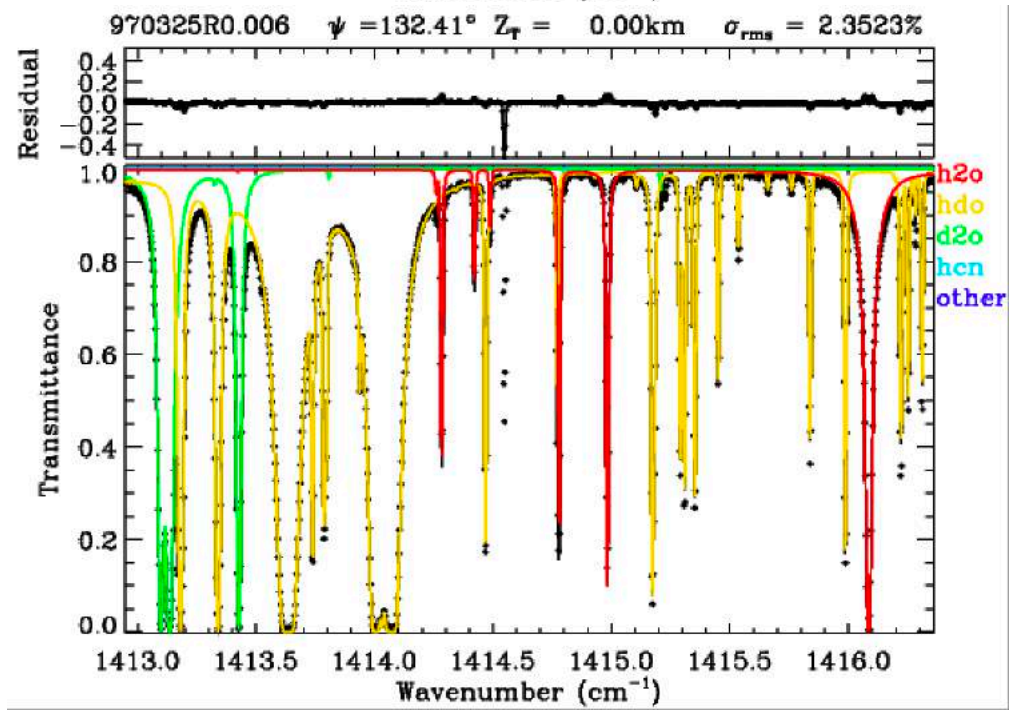
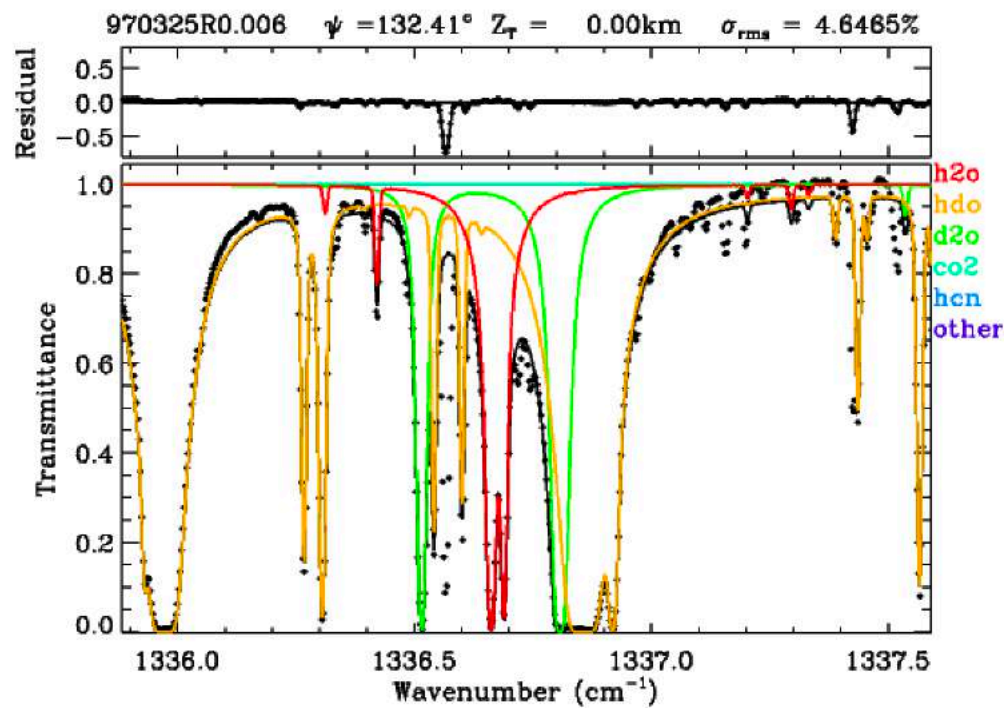
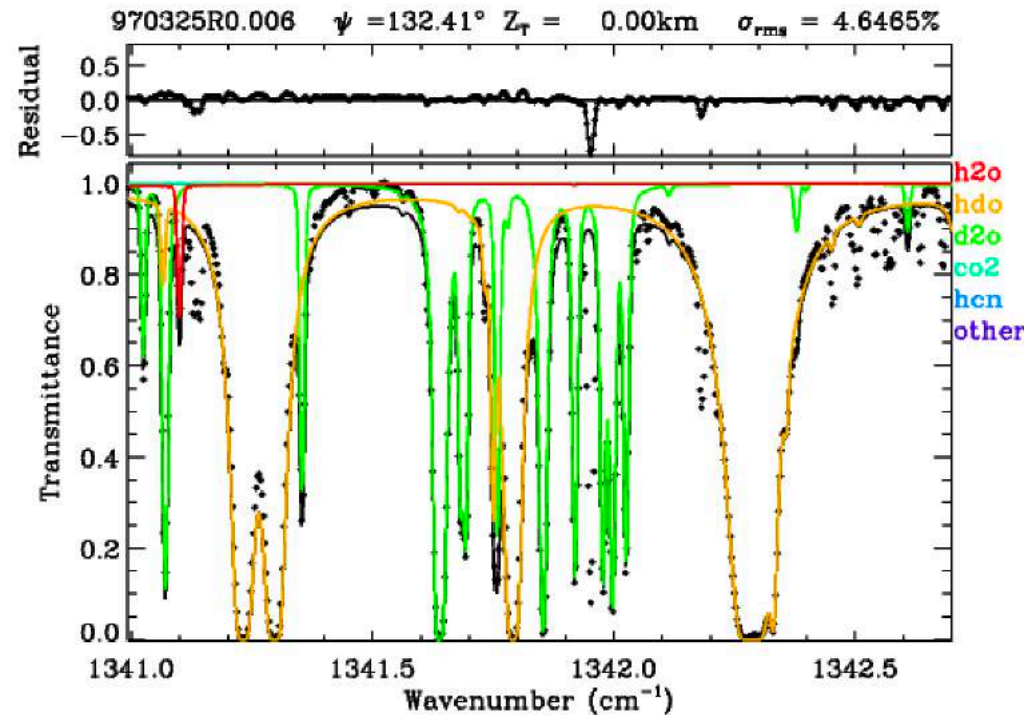
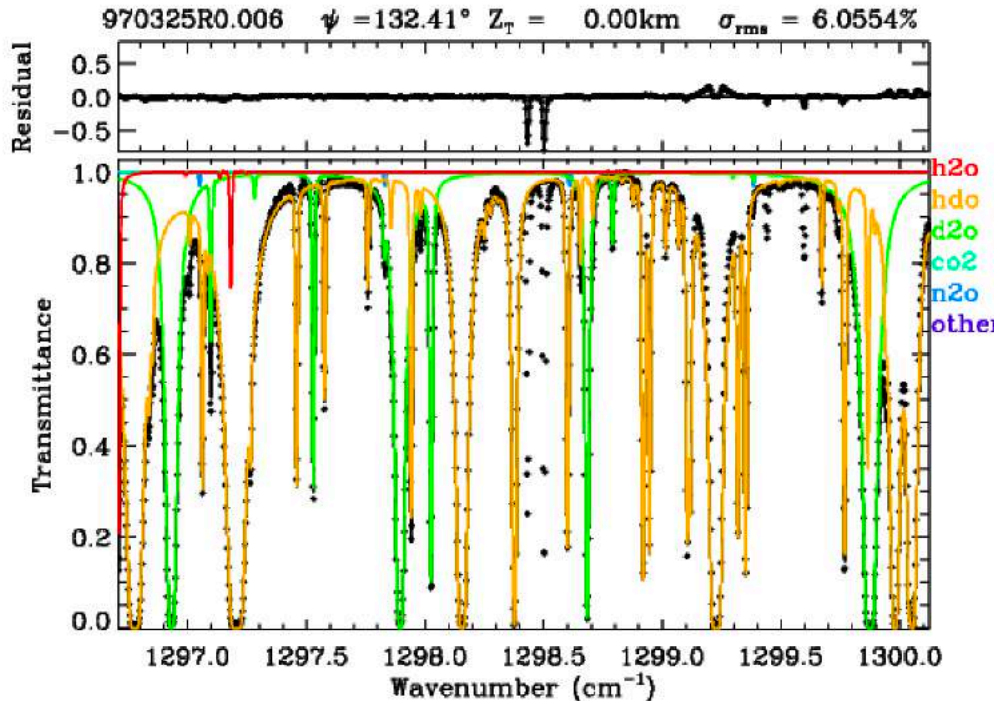




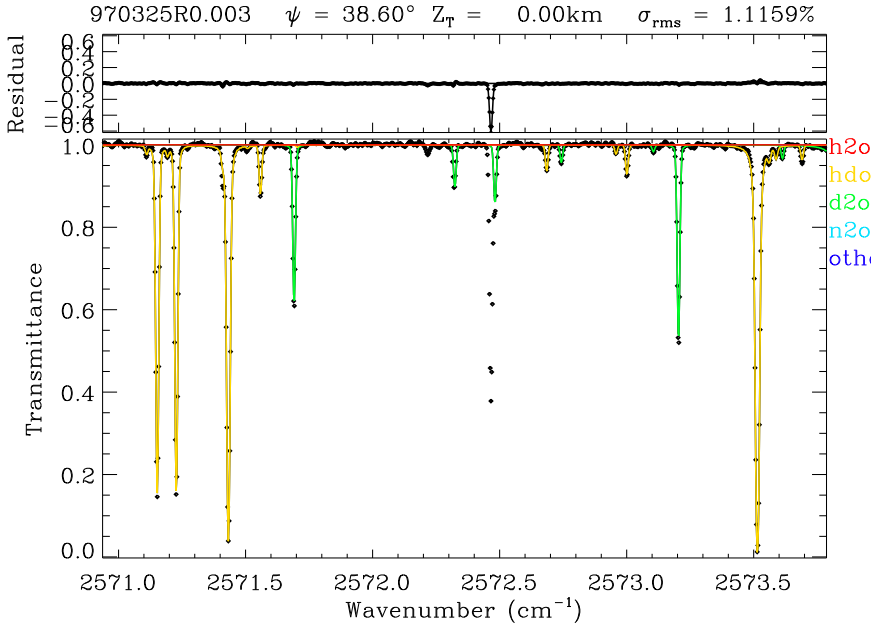
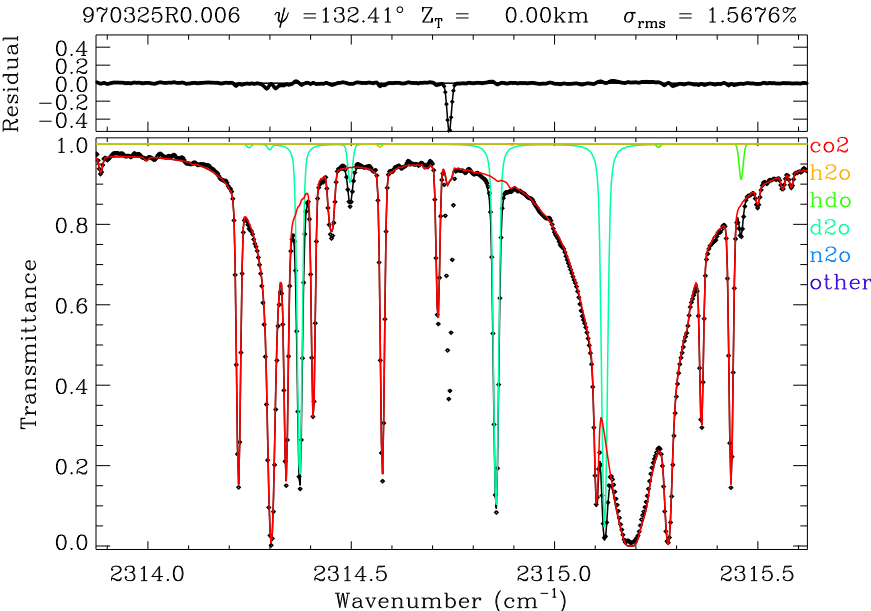
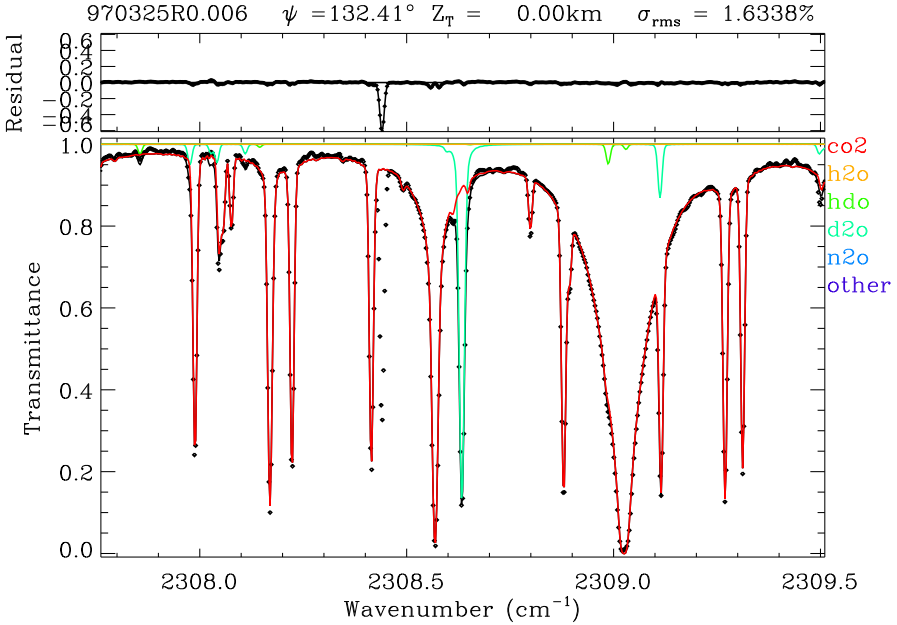
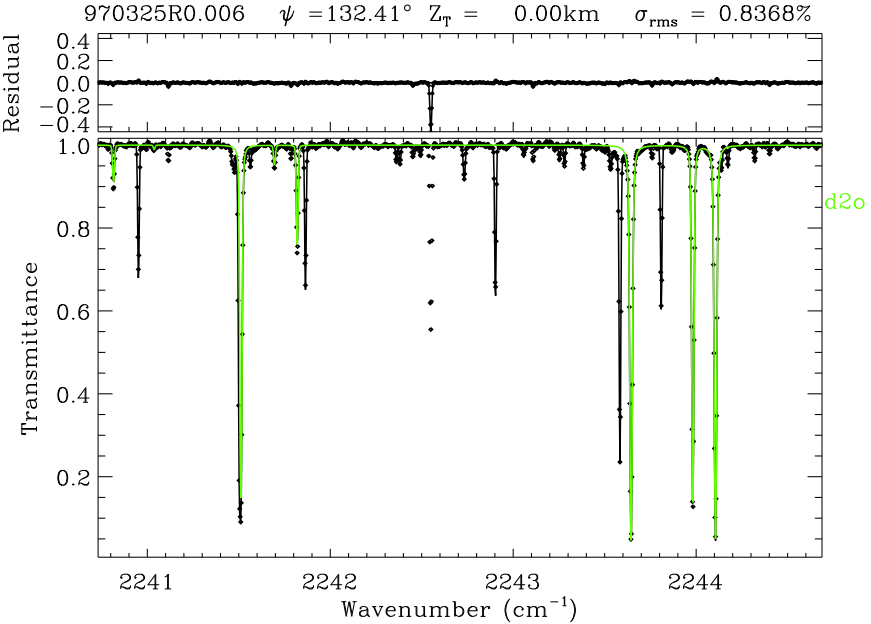
Missing D₂O lines:
2200-2600 cm^{-1} ?



Missing D₂O lines:
1200-1500 cm⁻¹



Missing D₂O lines 2200-2600



Examples of mis-positioned D_2O lines in the region of the 7885 cm^{-1} O_2 band.

