HI data from Arecibo Millennium Survey

See python scripts:
1. plot_specs.py
2. read_specs_data.py
3. nhi_heiles_uncertainty.py
4. plot_nhi_uncertainty.py

I. HI column densities N^*_{HI} under the optically thin assumption from Millennium Survey (MS) data

Aim:

- Compare the total N_{HI} from paper with the N^*_{HI} obtained from the optically thin assumption to do/understand how much HI could be underestimated.

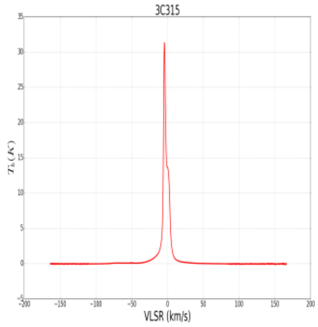
> *Note: Optically thin assumption means the optical depth $\tau << 1$ and no T(b) background.*
> *Eg: For $N_{HI}=2.10^{20}$ (cm^{-2}):*
> *CNM: $T_s \sim 100K$, $\tau \sim 1$: Optically thick*
> *WNM: $T_s \sim 10.000K$, $\tau << 1$: Optically thick*

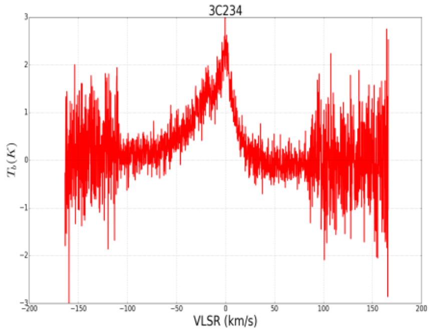
How?:

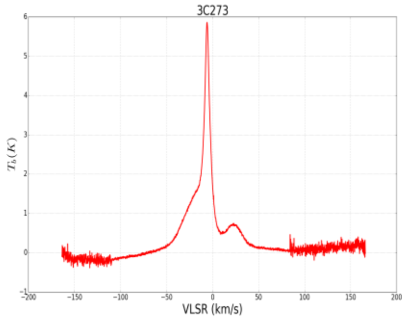
- *Step 1*: Use the "expected emission profiles" from Heiles & Troland (2003), which were derived for each of the MS sources, to compute HI column density N^*_{HI} under the optically thin assumption.
- *Step 2*: Compare optically thin HI column densities N^*_{HI} with the total HI column densities N_{HI} derived for each line-of-sight via the combination of emission and absorption spectra (table 2 in the paper).
- The difference gives us a measure of how much material would be missed under the optically thin assumption - i.e. an opacity correction, typical value is 10%.

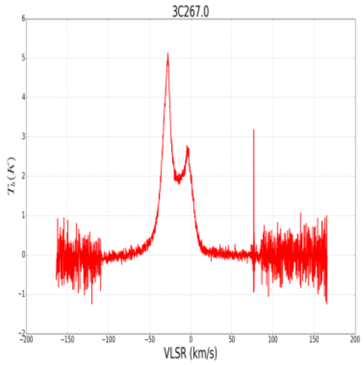
Process:

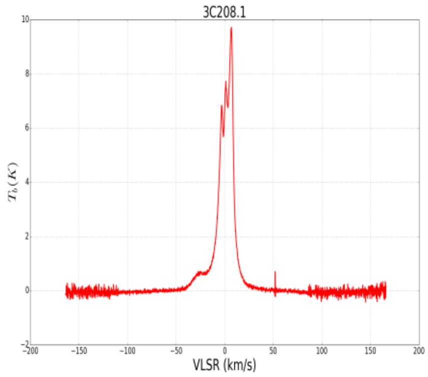
- Use Arecibo HI data from Heiles paper 2003 for 79 sources, download from http://vizier.cfa.harvard.edu/viz-bin/VizieR-3?-source=J/ApJS/145/329/table2&-out.form=%2bA
- (Show Sources in Galaxy Cordinates)
- Plot 79 spectra, *T_b (K) vs V_{LSR} (km/s)*, where T_b = 0.5*StokesI. This is Stokes I , which is TWICE the conventionally defined brightness temperature.











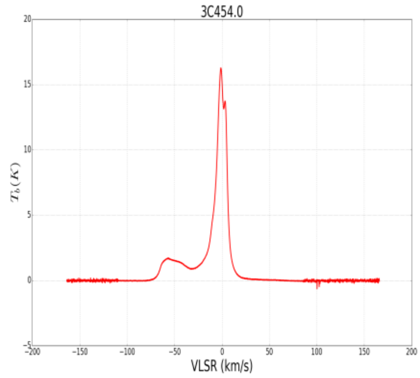
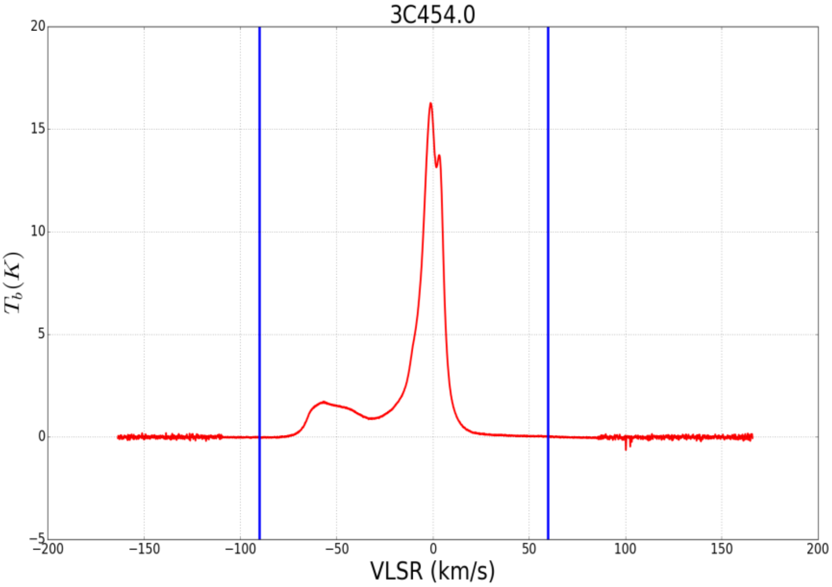


Figure ??: Example of HI spectra from Millennium Survey.

- HI spectra in general are clean, some with spike at a specific bin, some with noise, most with noise at 2 ends.
- For each spectrum, I define the velocity interval [v_1, v_2] containing the emission lines as illustrated in the figure below.



- Integrate the spectrum from v1 to v2 to get the HI integrated intensities W_{HI}.
- The optically thin HI column densities N^*_{HI} are then estimated as:

$$N^*_{HI} = 1.8224 \cdot 10^{18}\, W_{HI}\ (cm^{-2})$$

- The results for each source are listed in (Table 1).

#	V_{start}	V_{end}	Start index	End index	N_{HI} [1e20 cm^{-2}]	Source
0	-65.0	40.0	1434	782	5.08	3C18
1	-40.0	20.0	1279	906	2.74	3C33-1
2	-70.0	55.0	1465	689	2.77	3C33
3	-27.0	23.0	1198	888	2.82	3C33-2
4	-35.0	24.0	1248	882	5.96	3C64
5	-33.0	25.0	1236	875	7.39	3C75-1
6	-60.0	34.0	1403	819	7.31	3C75
7	-40.0	25.0	1279	875	7.64	3C75-2
8	-60.0	40.0	1403	782	8.89	3C78
9	-30.0	25.0	1217	875	8.47	3C79
10	-30.0	25.0	1217	875	8.65	CTA21
11	-38.0	27.0	1267	863	9.82	P0320+05
12	-63.0	30.0	1422	844	10.78	NRAO140
13	-65.0	30.0	1434	844	9.45	3C93.1
14	-30.0	28.0	1217	857	11.18	P0347+05
15	-40.0	40.0	1279	782	9.61	3C98-1
16	-30.0	35.0	1217	813	9.73	3C98
17	-30.0	40.0	1217	782	9.33	3C98-2
18	-30.0	33.0	1217	826	9.35	3C105
19	-40.0	34.0	1279	819	13.72	3C109

20	-50.0	35.0	1341	813	17.34	P0428+20
21	-55.0	35.0	1372	813	9.25	3C120
22	-90.0	35.0	1589	813	15.49	3C123
23	-80.0	40.0	1527	782	21.94	3C131
24	-80.0	45.0	1527	751	19.66	3C132
25	-80.0	40.0	1527	782	23.16	3C133
26	-70.0	40.0	1465	782	17.17	3C138
27	-63.0	38.0	1422	795	40.16	3C141.0
28	-50.0	40.0	1341	782	30.5	T0526+24
29	-30.0	47.0	1217	739	16.91	3C142.1
30	-48.0	42.0	1329	770	21.86	P0531+19
31	-38.0	45.0	1267	751	45.52	T0556+19
32	-50.0	46.0	1341	745	40.07	4C22.12
33	-60.0	47.0	1403	739	31.26	3C154
34	-78.0	60.0	1515	658	44.36	T0629+10
35	-50.0	85.0	1341	503	42.21	3C167
36	-80.0	50.0	1527	720	7.41	3C172.0
37	-40.0	50.0	1279	720	2.39	DW0742+10
38	-40.0	60.0	1279	658	2.89	3C190.0
39	-80.0	54.0	1527	695	3.89	3C192
40	-65.0	65.0	1434	627	4.15	P0820+22
41	-40.0	50.0	1279	720	4.86	3C207
42	-50.0	55.0	1341	689	2.99	3C208.0
43	-50.0	55.0	1341	689	2.73	3C208.1
44	-60.0	35.0	1403	813	1.02	3C223
45	-85.0	50.0	1558	720	3.29	3C225a
46	-70.0	50.0	1465	720	3.2	3C225b
47	-50.0	45.0	1341	751	2.62	3C228.0
48	-100.0	50.0	1652	720	1.61	3C234
49	-80.0	52.0	1527	708	1.16	3C236
50	-80.0	80.0	1527	534	1.96	3C237
51	-65.0	40.0	1434	782	2.09	3C245
52	-80.0	55.0	1527	689	1.54	P1055+20
53	-78.0	20.0	1515	906	1.51	P1117+14
54	-90.0	26.0	1589	869	1.64	3C263.1
55	-80.0	20.0	1527	906	1.73	3C264.0
56	-77.0	34.0	1509	819	2.28	3C267.0
57	-75.0	27.0	1496	863	2.38	3C272.1
58	-73.0	100.0	1484	410	1.99	3C273
59	-67.0	50.0	1447	720	2.24	3C274.1
60	-50.0	20.0	1341	906	2.04	4C07.32
61	-60.0	40.0	1403	782	1.11	4C32.44
62	-75.0	72.0	1496	583	2.07	3C286
63	-72.0	50.0	1478	720	1.27	3C293
64	-50.0	25.0	1341	875	2.58	4C19.44
65	-80.0	30.0	1527	844	2.64	4C20.33
66	-80.0	40.0	1527	782	3.42	3C310
67	-45.0	42.0	1310	770	4.04	3C315
68	-85.0	35.0	1558	813	4.14	3C318
69	-80.0	40.0	1527	782	4.43	3C333
70	-40.0	65.0	1279	627	5.2	3C348
71	-75.0	65.0	1496	627	8.67	3C353

72	-60.0	70.0	1403	596	8.51	4C13.65
73	-107.0	85.0	1695	503	13.48	4C13.67
74	-100.0	55.0	1652	689	20.81	3C409
75	-105.0	50.0	1683	720	38.31	3C410
76	-80.0	60.0	1527	658	7.65	3C433
77	-80.0	55.0	1527	689	5.19	3C454.0
78	-80.0	55.0	1527	689	6.26	3C454.3

Table 1: HI column densities N^*_{HI} under the optically thin assumption for 79 MS lines-of-sigh.

- From the Heiles 2003 paper, calculate the total HI column density N_{HI} for each source by summing its cold N_{CNM} and warm N_{WNM} components. These values are to a certain extent accurate and reliable because they are derived from on-/off source observation method.

- The N^*_{HI} estimates are made using the assumption of opacity τ<<1 and should give an underestimate in cases where there is appreciable cold gas. From values in table 2, one should find some places where the true column density N_{HI} is underestimated by up to 50%, mostly where the HI columns are already high (see red lines in Table 2).

- Take the difference/ratio between N_{HI} and N^*_{HI} (Table 2)

# id	V_{start}	V_{end}	StartID	EndID	N^*_{HI}[1e20]	N_{HI}[1e20]	N_{HI}_ratio(%)	source
0	-65.0	40.0	1434.0	782.0	5.08	5.99	15.19	3C18
1	-40.0	20.0	1279.0	906.0	2.74	2.81	2.49	3C33-1
2	-70.0	55.0	1465.0	689.0	2.77	2.78	0.36	3C33
3	-27.0	23.0	1198.0	888.0	2.82	2.92	3.42	3C33-2
4	-35.0	24.0	1248.0	882.0	5.96	6.33	5.85	3C64
5	-33.0	25.0	1236.0	875.0	7.39	7.97	7.28	3C75-1
6	-60.0	34.0	1403.0	819.0	7.31	7.89	7.35	3C75
7	-40.0	25.0	1279.0	875.0	7.64	8.23	7.17	3C75-2
8	-60.0	40.0	1403.0	782.0	8.89	10.06	11.63	3C78
9	-30.0	25.0	1217.0	875.0	8.47	9.37	9.61	3C79
10	-30.0	25.0	1217.0	875.0	8.65	9.56	9.52	CTA21
11	-38.0	27.0	1267.0	863.0	9.82	11.2	12.32	P0320+05
12	-63.0	30.0	1422.0	844.0	10.78	29.49	63.45	NRAO140
13	-65.0	30.0	1434.0	844.0	9.45	12.32	23.3	3C93.1
14	-30.0	28.0	1217.0	857.0	11.18	13.45	16.88	P0347+05
15	-40.0	40.0	1279.0	782.0	9.61	10.37	7.33	3C98-1
16	-30.0	35.0	1217.0	813.0	9.73	11.02	11.71	3C98
17	-30.0	40.0	1217.0	782.0	9.33	10.25	8.98	3C98-2
18	-30.0	33.0	1217.0	826.0	9.35	14.68	36.31	3C105
19	-40.0	34.0	1279.0	819.0	13.72	20.82	34.1	3C109
20	-50.0	35.0	1341.0	813.0	17.34	23.89	27.42	P0428+20
21	-55.0	35.0	1372.0	813.0	9.25	15.94	41.97	3C120
22	-90.0	35.0	1589.0	813.0	15.49	27.38	43.43	3C123
23	-80.0	40.0	1527.0	782.0	21.94	28.55	23.15	3C131
24	-80.0	45.0	1527.0	751.0	19.66	23.81	17.43	3C132
25	-80.0	40.0	1527.0	782.0	23.16	28.5	18.74	3C133
26	-70.0	40.0	1465.0	782.0	17.17	19.84	13.46	3C138
27	-63.0	38.0	1422.0	795.0	40.16	52.69	23.78	3C141.0
28	-50.0	40.0	1341.0	782.0	30.5	96.85	68.51	T0526+24
29	-30.0	47.0	1217.0	739.0	16.91	21.96	23.0	3C142.1

30	-48.0	42.0	1329.0	770.0	21.86	23.84	8.31	P0531+19
31	-38.0	45.0	1267.0	751.0	45.52	53.63	15.12	T0556+19
32	-50.0	46.0	1341.0	745.0	40.07	85.23	52.99	4C22.12
33	-60.0	47.0	1403.0	739.0	31.26	35.58	12.14	3C154
34	-78.0	60.0	1515.0	658.0	44.36	59.25	25.13	T0629+10
35	-50.0	85.0	1341.0	503.0	42.21	50.24	15.98	3C167
36	-80.0	50.0	1527.0	720.0	7.41	7.71	3.89	3C172.0
37	-40.0	50.0	1279.0	720.0	2.39	2.43	1.65	DW0742+10
38	-40.0	60.0	1279.0	658.0	2.89	2.82	2.48	3C190.0
39	-80.0	54.0	1527.0	695.0	3.89	3.97	2.02	3C192
40	-65.0	65.0	1434.0	627.0	4.15	4.23	1.89	P0820+22
41	-40.0	50.0	1279.0	720.0	4.86	5.25	7.43	3C207
42	-50.0	55.0	1341.0	689.0	2.99	2.99	0.0	3C208.0
43	-50.0	55.0	1341.0	689.0	2.73	2.76	1.09	3C208.1
44	-60.0	35.0	1403.0	813.0	1.02	0.98	4.08	3C223
45	-85.0	50.0	1558.0	720.0	3.29	3.4	3.24	3C225a
46	-70.0	50.0	1465.0	720.0	3.2	3.28	2.44	3C225b
47	-50.0	45.0	1341.0	751.0	2.62	2.61	0.38	3C228.0
48	-100.0	50.0	1652.0	720.0	1.61	1.61	0.0	3C234
49	-80.0	52.0	1527.0	708.0	1.16	1.21	4.13	3C236
50	-80.0	80.0	1527.0	534.0	1.96	2.2	10.91	3C237
51	-65.0	40.0	1434.0	782.0	2.09	2.03	2.96	3C245
52	-80.0	55.0	1527.0	689.0	1.54	1.57	1.91	P1055+20
53	-78.0	20.0	1515.0	906.0	1.51	1.57	3.82	P1117+14
54	-90.0	26.0	1589.0	869.0	1.64	1.68	2.38	3C263.1
55	-80.0	20.0	1527.0	906.0	1.73	1.73	0.0	3C264.0
56	-77.0	34.0	1509.0	819.0	2.28	2.33	2.15	3C267.0
57	-75.0	27.0	1496.0	863.0	2.38	2.39	0.42	3C272.1
58	-73.0	100.0	1484.0	410.0	1.99	1.93	3.11	3C273
59	-67.0	50.0	1447.0	720.0	2.24	2.36	5.08	3C274.1
60	-50.0	20.0	1341.0	906.0	2.04	2.11	3.32	4C07.32
61	-60.0	40.0	1403.0	782.0	1.11	1.05	5.71	4C32.44
62	-75.0	72.0	1496.0	583.0	2.07	2.04	1.47	3C286
63	-72.0	50.0	1478.0	720.0	1.27	1.28	0.78	3C293
64	-50.0	25.0	1341.0	875.0	2.58	2.65	2.64	4C19.44
65	-80.0	30.0	1527.0	844.0	2.64	2.69	1.86	4C20.33
66	-80.0	40.0	1527.0	782.0	3.42	3.71	7.82	3C310
67	-45.0	42.0	1310.0	770.0	4.04	4.77	15.3	3C315
68	-85.0	35.0	1558.0	813.0	4.14	4.75	12.84	3C318
69	-80.0	40.0	1527.0	782.0	4.43	5.09	12.97	3C333
70	-40.0	65.0	1279.0	627.0	5.2	5.69	8.61	3C348
71	-75.0	65.0	1496.0	627.0	8.67	10.85	20.09	3C353
72	-60.0	70.0	1403.0	596.0	8.51	9.16	7.1	4C13.65
73	-107.0	85.0	1695.0	503.0	13.48	16.72	19.38	4C13.67
74	-100.0	55.0	1652.0	689.0	20.81	25.81	19.37	3C409
75	-105.0	50.0	1683.0	720.0	38.31	48.22	20.55	3C410
76	-80.0	60.0	1527.0	658.0	7.65	7.89	3.04	3C433
77	-80.0	55.0	1527.0	689.0	5.19	5.38	3.53	3C454.0
78	-80.0	55.0	1527.0	689.0	6.26	6.53	4.13	3C454.3

Table 2: HI column densities N^*_{HI} under the optically thin assumption and total HI column densities N_{HI} in comparison for 79 MS lines-of-sigh.

- Correlation between the total HI column densities derived for each sightline, N_{HI}, and the HI column densities under the optically thin assumption, N^*_{HI}, is shown in Figure ?? and Figure ??. While figure ?? is the histogram of the ratio $f=N_{HI}/N^*_{HI}$ in percentage, figure ?? displays the ratio f as a function of $log_{10}(N^*_{HI}/10^{20}\ cm^{-2})$. Clearly, the ratio f increases with N^*_{HI}.

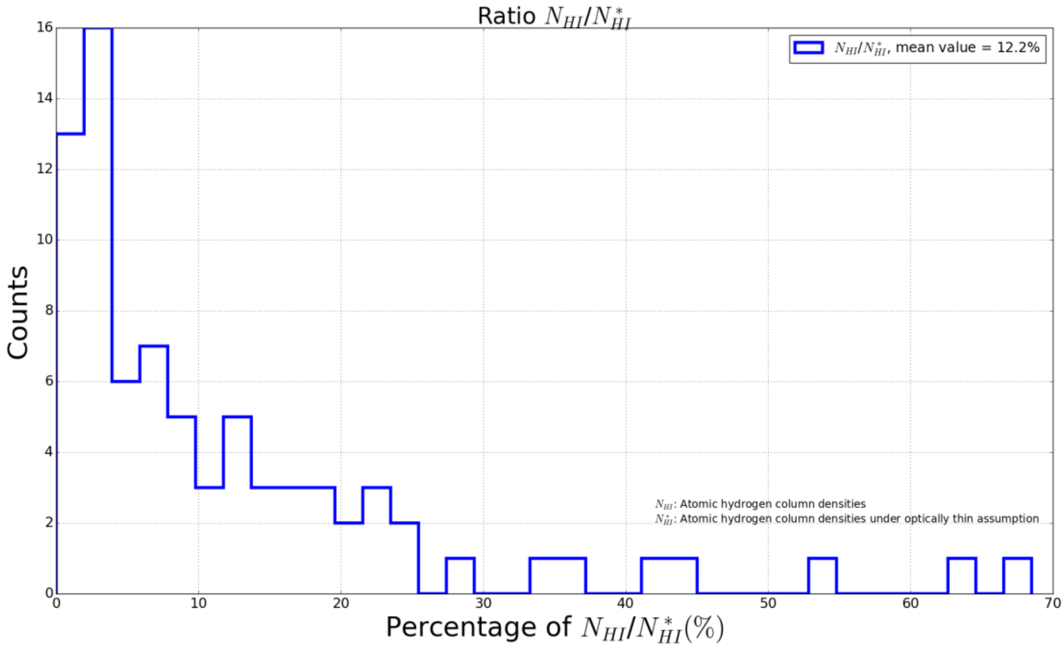


Figure ??: Histogram of N_{HI}/N^*_{HI} in percentage

- The difference gives us a measure of how much material would be missed under the optically thin assumption - i.e. an opacity correction.

→ Result:

ratio $R = N^*_{HI}/N_{HI}$

- 32/79 sources have ratio values <= 5%.
- 47/79 sources have ratio values <= 10%.
- 32/79 sources could be considered as optically thin with ratio values <= 5% ??

- The HI mass increases by ~12% compared to the optically thin assumption (as illustrated in Figure ??). This is consistent with the typical value (~10%) throughout the Galaxy.

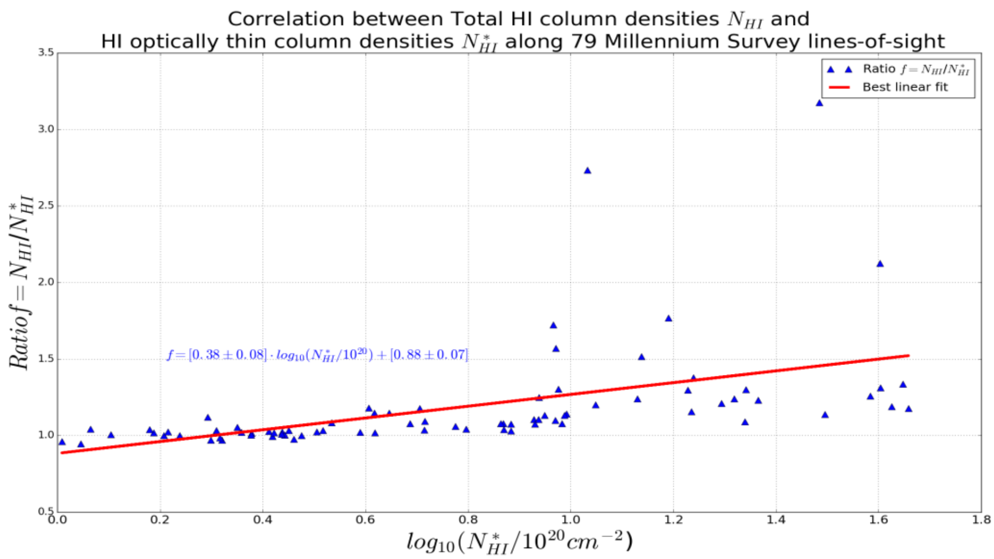


*Figure ??: Correlation between NHI and N*HI along 79 MS sightlines.*

II. Uncertainty of the total HI column densities N_{HI}

- In the paper, the uncertainties on HI column density are not estimated. The value of N_{HI} for each Gaussian component is computed given opacity τ, excitation temperature T_s and the line-width Δv using the standard equation. Since the MS paper gives uncertainties on each of those values, one just needs to propagate those errors through in the usual way.

1. For WNM:

I estimate the uncertainties by using the optically thin assumption, that is:

$$\frac{N(\mathrm{HI})}{\mathrm{cm}^{-2}} = 1.8224 \times 10^{18} \int_{-\infty}^{\infty} T_{\mathrm{b}}(v)\, \mathrm{d}\left(\frac{v}{\mathrm{kms}^{-1}}\right)$$

HI column density in the optically thin limit with no background source

$y = N_{HI} = const*\int T_b*exp[-(v-v_0)^2/\sigma^2]dv$

where:

const = $1.8224*10^{18}$ [$cm^{-2}K^{-1}km^{-1}s$]

σ = FWHM/(2$\sqrt{\ln 2}$), given FWHM = Δv in the paper.

Consequently, $\Delta\sigma = \Delta$(FWHM)/(2$\sqrt{\ln 2}$)

Have: $\Delta^2 y = (\Delta y/\Delta T_b)^2 \Delta T^2_b + (\Delta y/\Delta v_0)^2 \Delta^2 v_0 + (\Delta y/\Delta\sigma)^2 \Delta^2\sigma$

$\Delta y/\Delta T_b = const*\int exp[-(v-v_0)^2/\sigma^2]dv = const*\sigma*\sqrt{\pi}$

$\Delta y/\Delta v_0 = (2*const*T_b/\sigma^2)\int x*exp[-x^2/\sigma^2]dv = 0$

$\Delta y/\Delta\sigma = const*T_b*\sqrt{\pi}$

$\rightarrow \Delta y = \Delta N_{HI} = const*\sqrt{[\pi*(\sigma^2\Delta T_b^2 + T^2_b*\Delta^2\sigma)]}$

2. For CNM

I use the function of T_s and τ_v to compute N_{HI} namely:

$$\boxed{\frac{N(\mathrm{HI})}{\mathrm{cm}^{-2}} = 1.8224\times10^{18} \left(\frac{T_s}{\mathrm{K}}\right) \int_{-\infty}^{\infty} \tau_v \, \mathrm{d}\left(\frac{v}{\mathrm{kms}^{-1}}\right)}$$

Where (from the paper):

- T_s is a variable with its uncertainty
- tau is a Gaussian, $\tau = \tau_0 * e^{[-(v-v0)^\wedge 2/\sigma^\wedge 2]}$, τ_0 is the central opacity of the component.
- τ *has* its uncertainty
- v_0 with its uncertainty
- σ with its uncertainty, note that in the paper, Δv is FWHM, so σ = FWHM/(2*√ln2)

Use the Gaussian integral to calculate the derivatives:

$y = N_{HI} = \mathrm{const}*T_s*\int\tau_0*\exp[-(v-v_0)^2/\sigma^2]dv$, where const = $1.8224*10^{18}$ [$cm^{-2}K^{-1}km^{-1}s$]

Have: $\Delta^2 y = (\Delta y/\Delta T_s)^2\Delta^2 T_s + (\Delta y/\Delta\tau)^2\Delta^2 v_0 + (\Delta y/\Delta v_0)^2\Delta^2 v_0 + (\Delta y/\Delta\sigma)^2\Delta^2\sigma$

$\Delta y/\Delta T_s = \mathrm{const}*\tau_0*\int\exp[-(v-v_0)^2/\sigma^2]dv = \mathrm{const}*\tau_0*\sigma*\sqrt{\pi}$

$\Delta y/\Delta\tau = \mathrm{const}*T_s\int\exp[-(v-v_0)^2/\sigma^2]dv = \mathrm{const}*T_s*\sigma*\sqrt{\pi}$

$\Delta y/\Delta v_0 = (2*\mathrm{const}*\tau*T_s/\sigma^2)\int x*\exp[-x^2/\sigma^2]dv = 0$

$\Delta y/\Delta\sigma = \mathrm{const}*\tau_0*T_s*\sqrt{\pi}$

$\rightarrow \Delta y = \Delta N_{HI} = const*\sqrt{[\pi*(\sigma^2\Delta T_s^2*\tau^2_0 + T^2_s*\Delta^2\sigma*\tau^2_0 + \sigma^2 T^2_s\Delta^2_\tau)]}$

Note: σ = FWHM/(2√ln2), given FWHM = Δv in the paper.

And $\Delta\sigma = \Delta(FWHM)/(2\sqrt{ln2})$

→ Then combine the two.

$N_{HI} = N_{CNM} + N_{WNM}$

Therefore, $\sigma_{N(HI)} = \sqrt{[\Sigma_i\sigma^2_{N(CNM)\_i} + \Sigma_k\sigma^2_{N(WNM)\_k}]}$

The figure shows the histogram of total N_{HI} uncertainties for 78 lines-of-sight. 3 sources with high uncertainties are in Galactic plane or with low latitude:

T0556+19   *l = 190.0893, b = -2.1665*

T0526+24   *l = 181.3551, b = -5.1933*

NRAO140   *l = 159.0002, b= -18.7646*

The mean value is ~ 6.6%

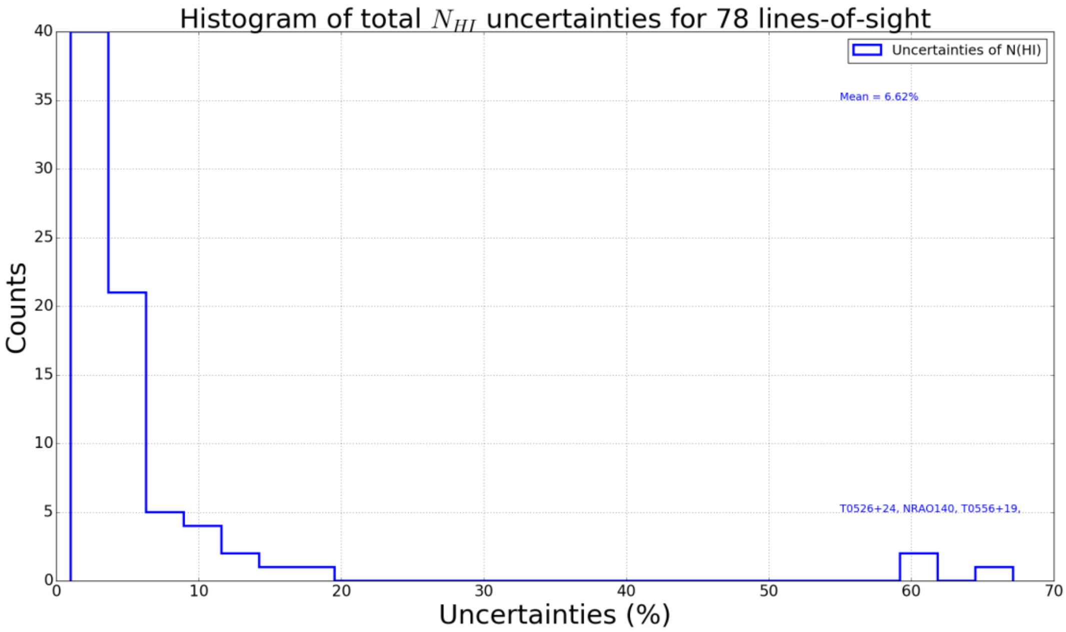


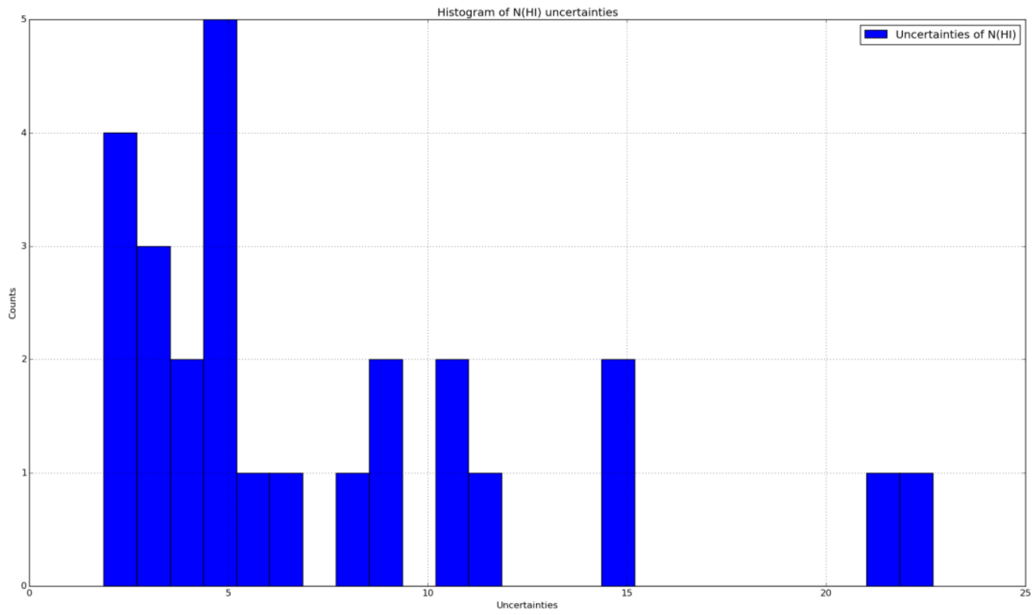
Figure ??: Histogram of N_{HI} uncertainties for 78 MS lines-of-sight

============= END ===============

I had been concerned about how the assumed ordering of emission/absorption components affected the uncertainties. Having read the paper more carefully, I realise that the errors quoted on tau, Ts and the linewidth already take that effect into account. i.e. they varied the order, recalculated the parameters in each case, and used a weighted average of all trials to derive their uncertainties.

- The problem is:
For CNM components, it is (tau_0;n) from equation, derived directly from the least-squares fit to the opacity profile. For WNM components it is the upper limit to peak opacity (tau_0;k) , estimated by eye (paper session 3.6), and has a very large error.
Here, for CNM components, the uncertainty on tau_o is set to Zero.
Overall, the obtained uncertainties for N(HI) are NOT accurate.

For 26 sources without CO,
- 19/26 sources with the uncertainties < 10%
- The largest uncertainty is from 3C18 (21.74%) because the contribution of tau (0.077 +/- 0.019) of a cold component.



However, [Prof.] For WNM components, tau is always small, so it's fine to calculate N(HI) from the optically thin assumption. i.e. in the optically thin limit, N(HI, WNM) is actually not dependent on tau or Tex at all, just on Tb(peak) and delta-v. Integrating that Gaussian and multiplying by 1.8 x 10^18 should give you your WNM column density, and the uncertainties on tau and Tex do not factor in.