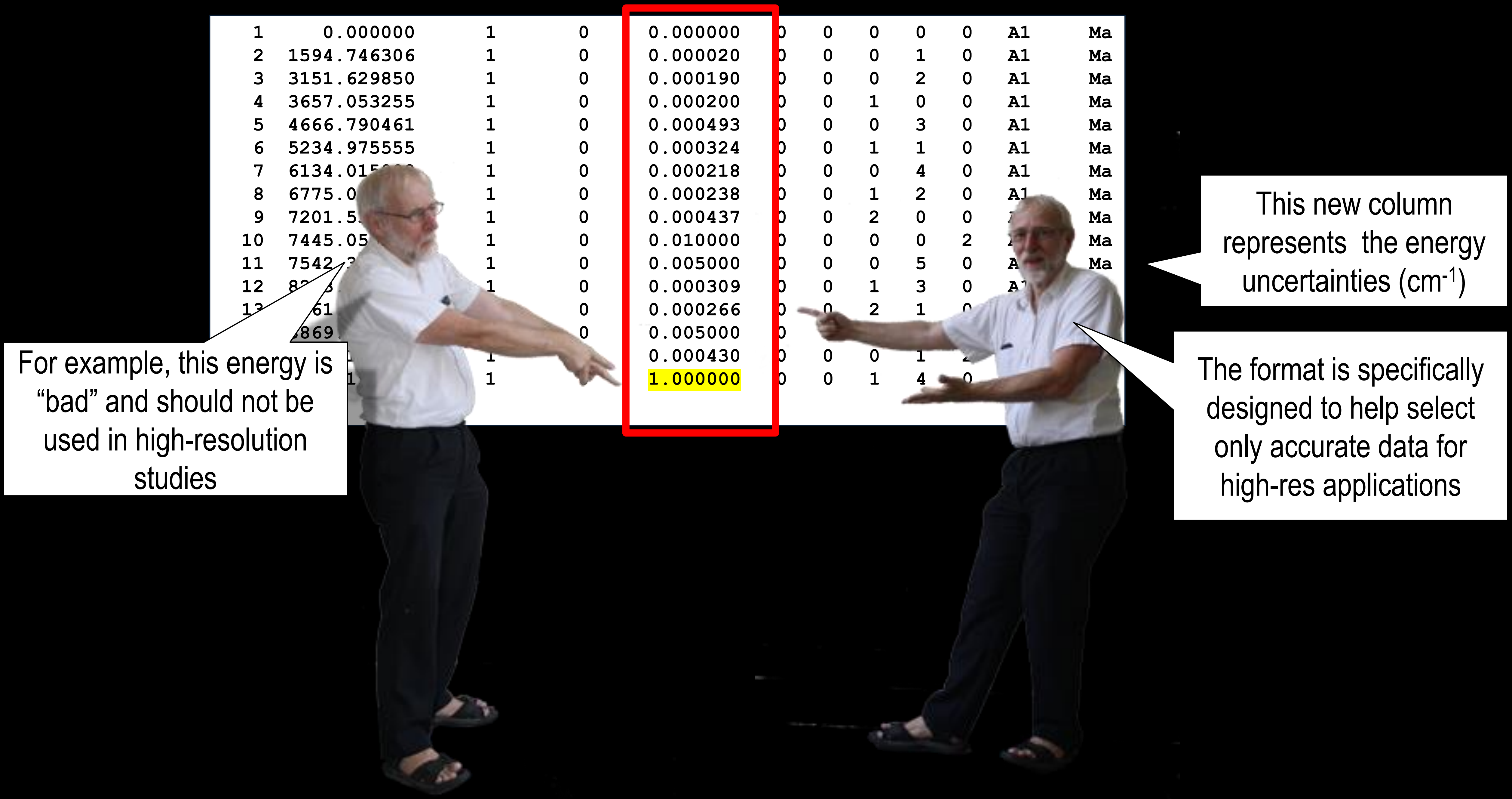
How to Characterise Uncertainties in Molecular Opacities: ExoMol Study

Sergey Yurchenko and Valentina Blokhina, UCL

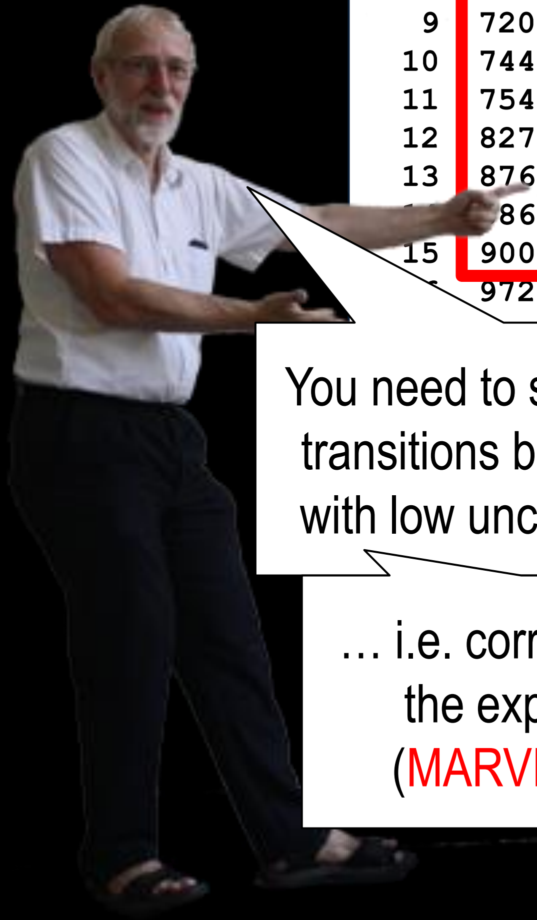
This is very important for High-Res applications

ExoMolHD format: We now have a column with energy uncertainties



0.000000
0.000020
0.000190
0.000200
0.000493
0.000324
0.000218
0.000238
0.000437
0.010000
0.005000
0.000309
0.000266
0.005000
0.000430
1.000000
This new column represents the energy uncertainties (cm-1)
The format is specifically designed to help select only accurate data for high-res applications
For example, this energy is “bad” and should not be used in high-resolution studies

1	0.000000	1	0	0.000000	0	0	0	0	0	A1	Ma
2	1594.746306	1	0	0.000020	0	0	0	1	0	A1	Ma
3	3151.629850	1	0	0.000190	0	0	0	2	0	A1	Ma
4	3657.053255	1	0	0.000200	0	0	1	0	0	A1	Ma
5	4666.790461	1	0	0.000493	0	0	0	3	0	A1	Ma
6	5234.975555	1	0	0.000324	0	0	1	1	0	A1	Ma
7	6134.015008	1	0	0.000218	0	0	0	4	0	A1	Ma
8	6775.093508	1	0	0.000238	0	0	1	2	0	A1	Ma
9	7201.539855	1	0	0.000437	0	0	2	0	0	A1	Ma
10	7445.056211	1	0	0.010000	0	0	0	0	2	A1	Ma
11	7542.372492	1	0	0.005000	0	0	0	5	0	A1	Ma
12	8273.975695	1	0	0.000309	0	0	1	3	0	A1	Ma
13	8761.581581	1	0	0.000266	0	0	2	1	0	A1	Ma
[illegible]	[illegible]869.950054	1	0	0.005000	0	0	0	6	0	A1	Ma
15	9000.136035	1	0	0.000430	0	0	0	1	2	A1	Ma
[illegible]	9724.195645	1	0	1.000000	0	0	1	4	0	A1	e



You need to select and use transitions between states with low uncertainties only

… i.e. corresponding to the experimental (MARVEL) values

… which are also indicated in the last column with Ma

MARVEL

Measured Active Rotational-Vibrational Energy Levels

MARVEL energies are obtained by inverting experimental frequencies

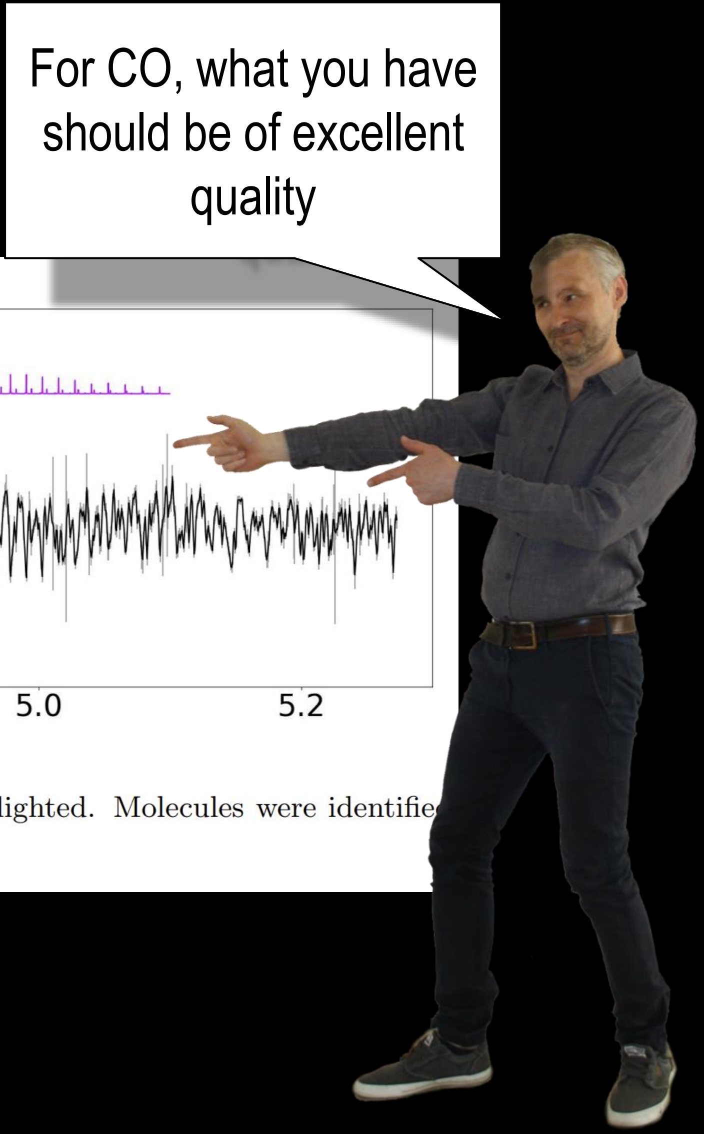
… as a solution of a system of linear equations

$$E_i^{\mathrm{Ma}} - E_j^{\mathrm{Ma}} = \nu_{ij}^{\mathrm{exp}}$$

How does this procedure affect the spectra (cross sections) and then also atmospheric retrievals?

Well, some data are already perfect
without MARVEL

You don’t need to worry about the lab data for CO



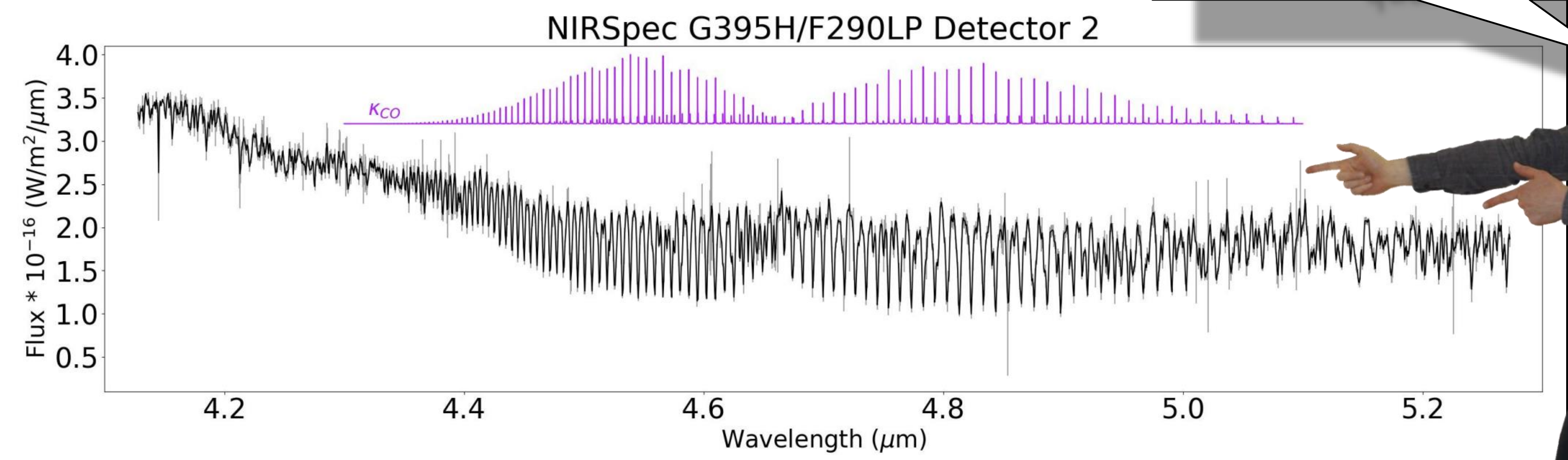
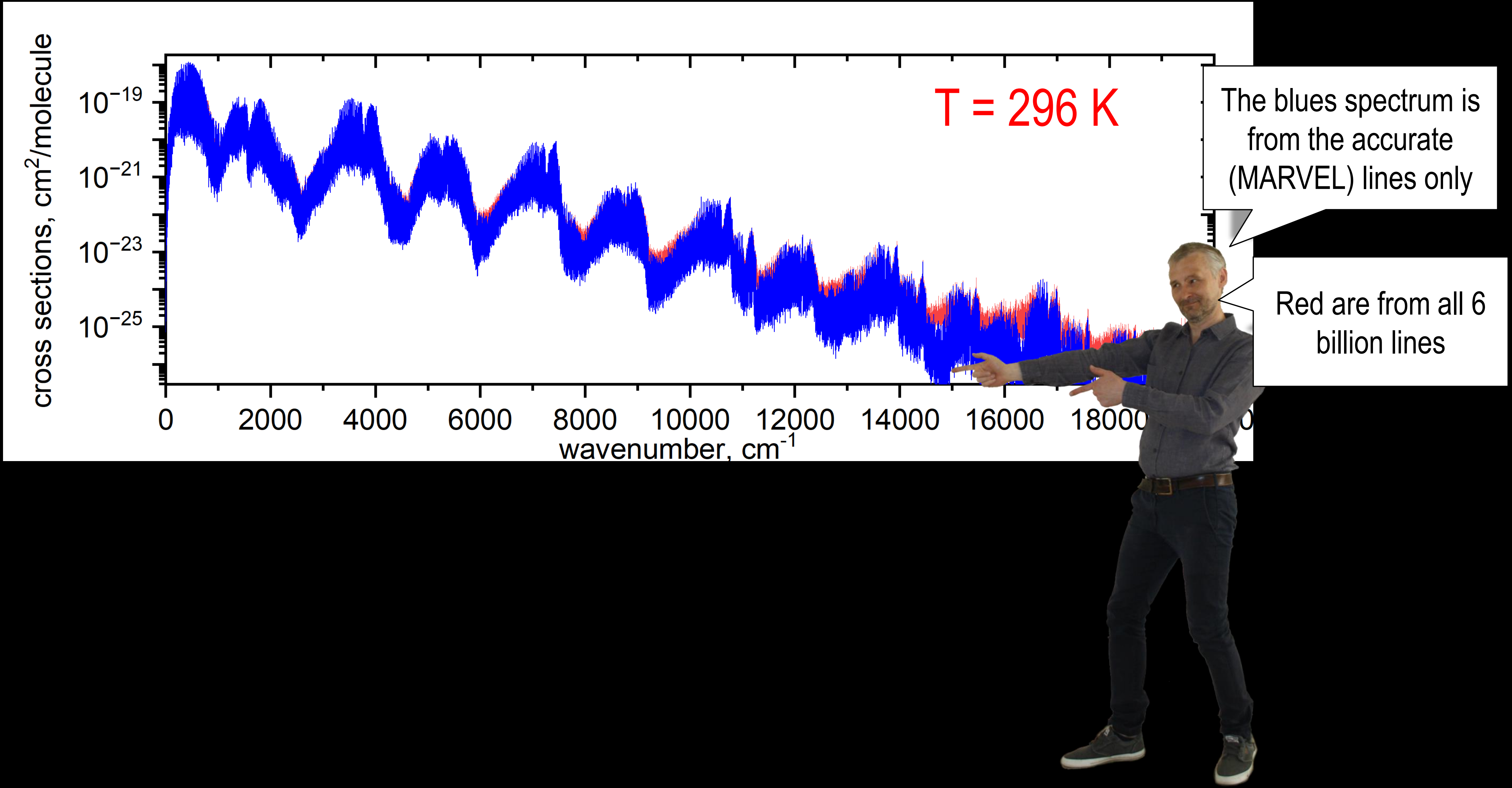
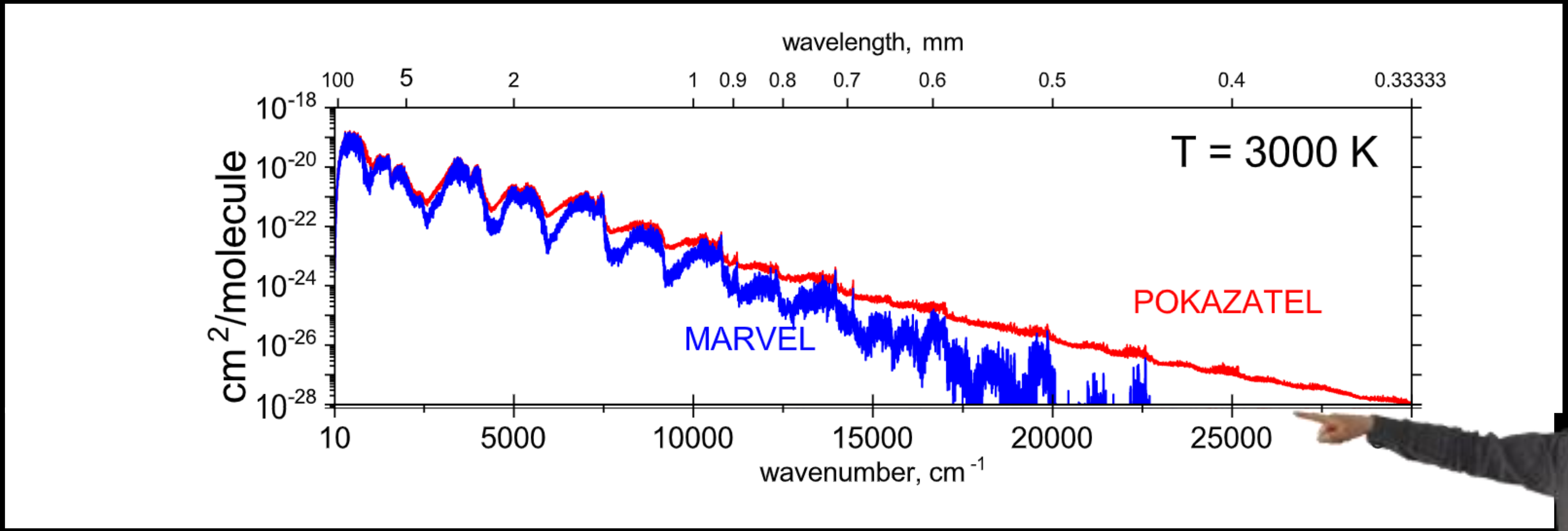


Figure 4. *JWST*/NIRSpec spectrum of VHS 1256 b, with important molecular gases highlighted. Molecules were identifie via visual comparison with template spectra.

Our water line list POKAZATEL
is quite good (at least at 296 K)


T = 296 K
cross sections, cm²/molecule
10−19
10−21
10−23
10−25
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
wavenumber, cm-1
The blues spectrum is from the accurate (MARVEL) lines only
Red are from all 6 billion lines

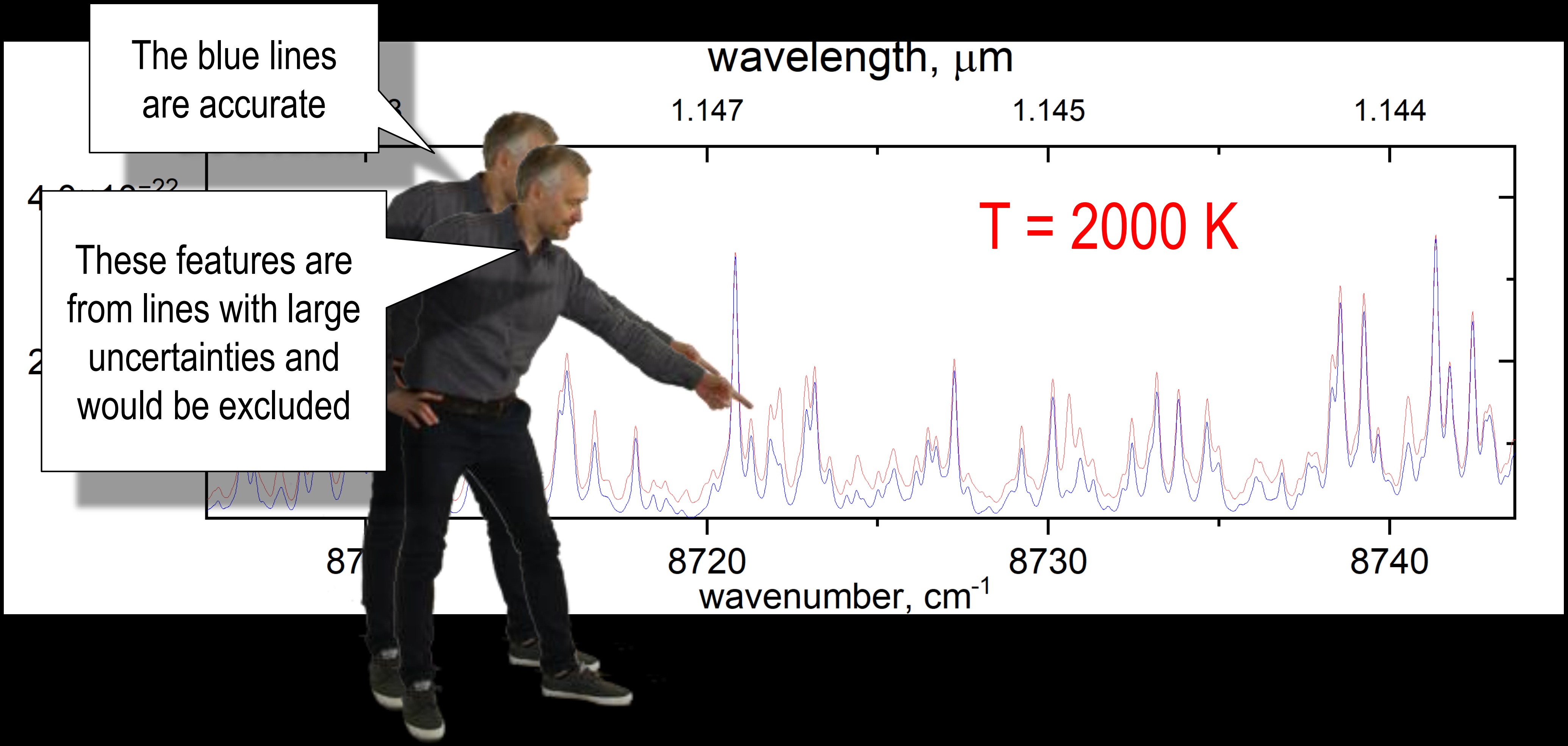
Let's increase temperature to $T=3000$ K



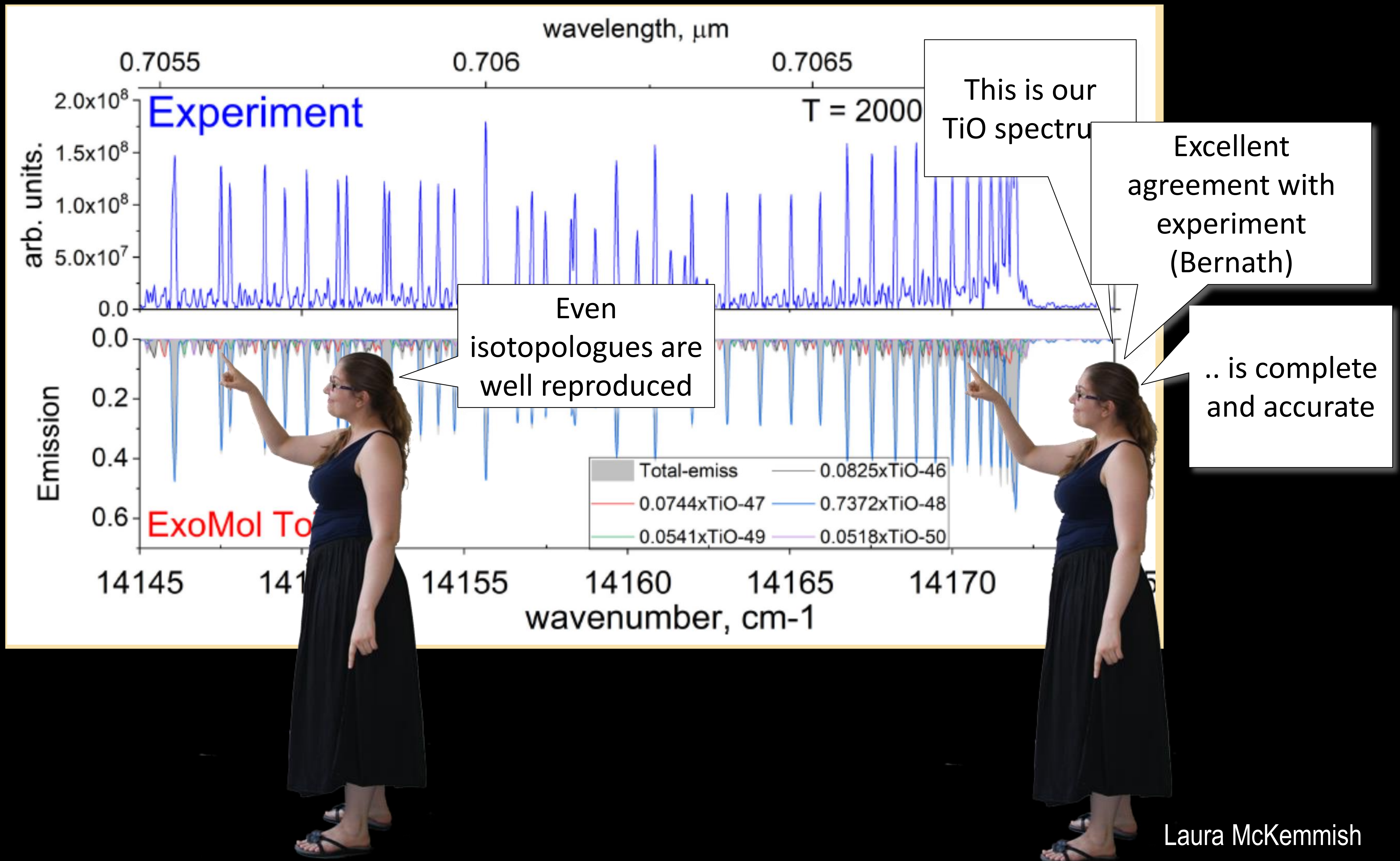
Red is complete data (theory) and blue is accurate data (experiment)

In HR applications, we suggest that the users select only accurate (Ma) lines and ignore the rest

I would like to look at how this affects the spectrum

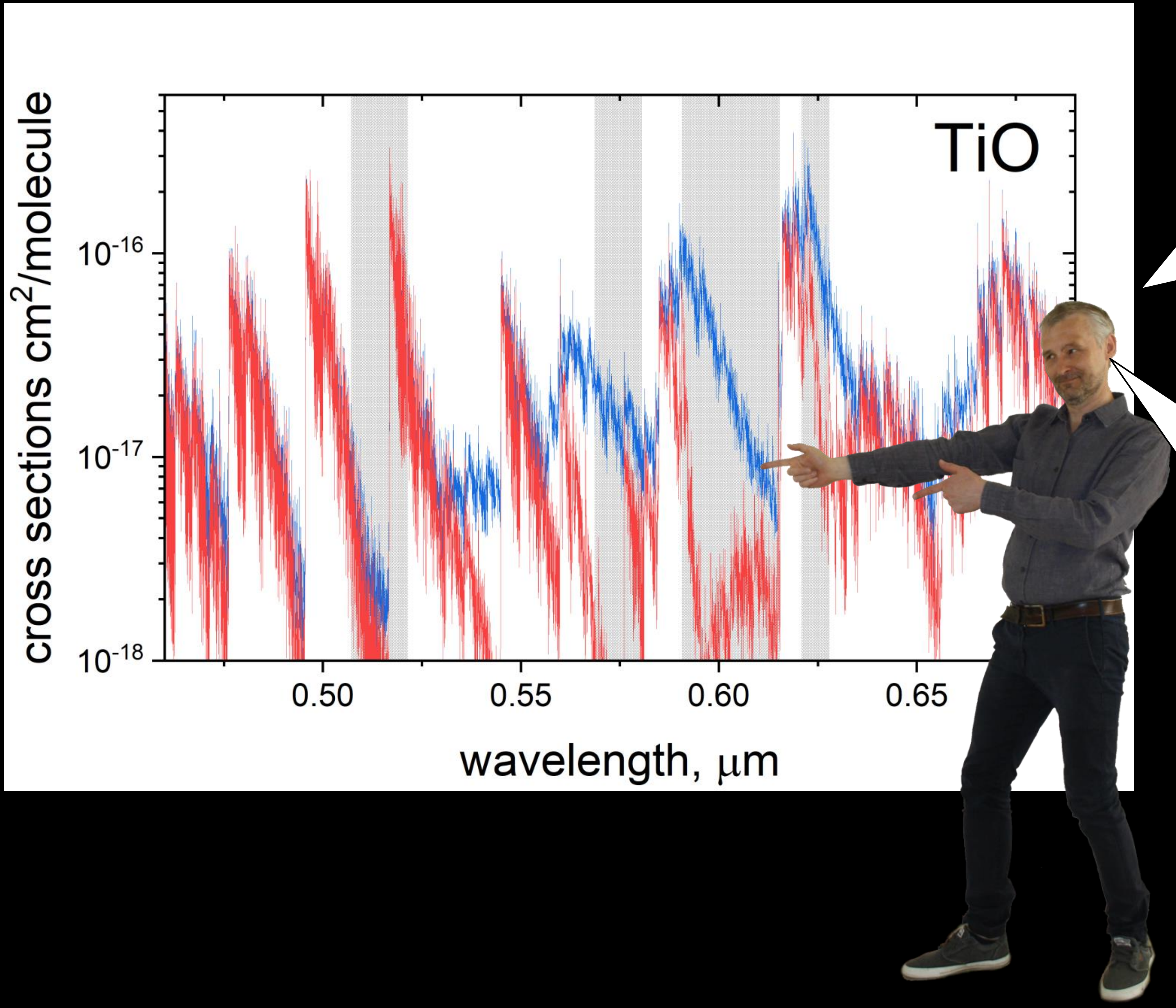

The blue lines are accurate
These features are from lines with large uncertainties and would be excluded
wavelength, μm
1.147
1.145
1.144
T = 2000 K
8720
8730
8740
wavenumber, cm-1

TiO in principle should be in a good shape



Laura McKemmish

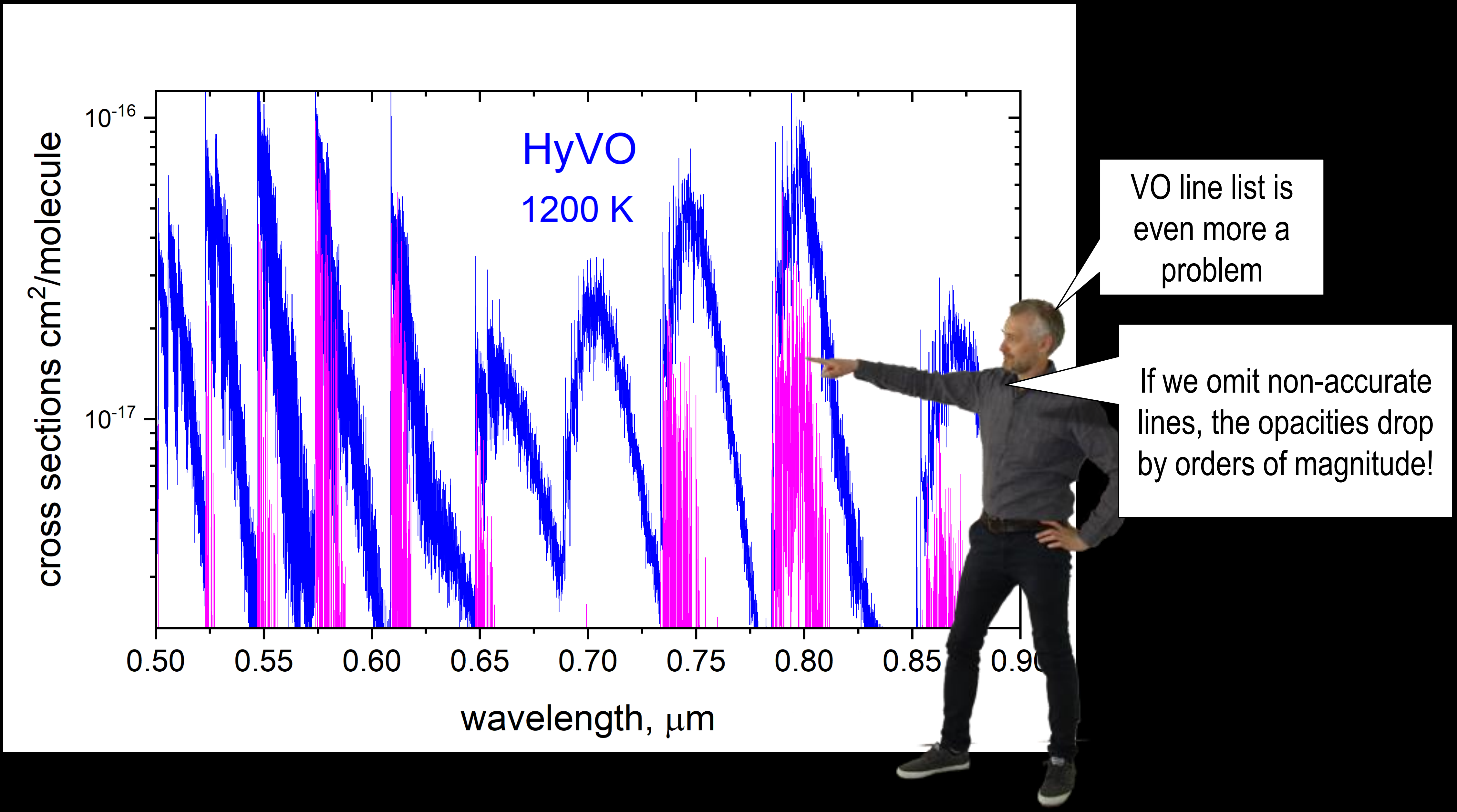
Let’s exclude lines with large uncertainties (non-MARVELised)



Now we see that some regions in the experimental spectrum (red) are heavily underrepresented

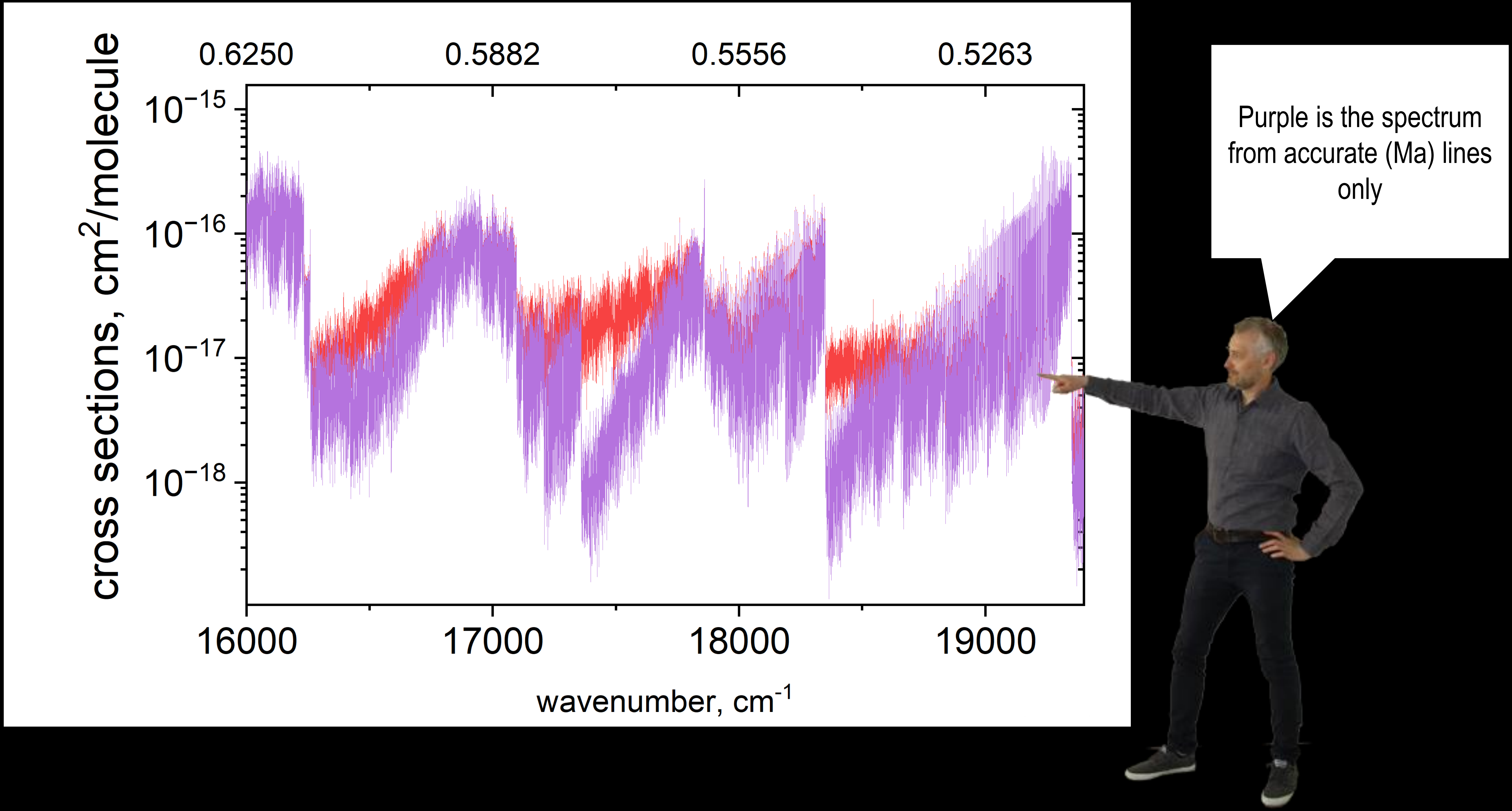
In grey I show the masks HighRes people commonly apply or TiO

Similarly for VO

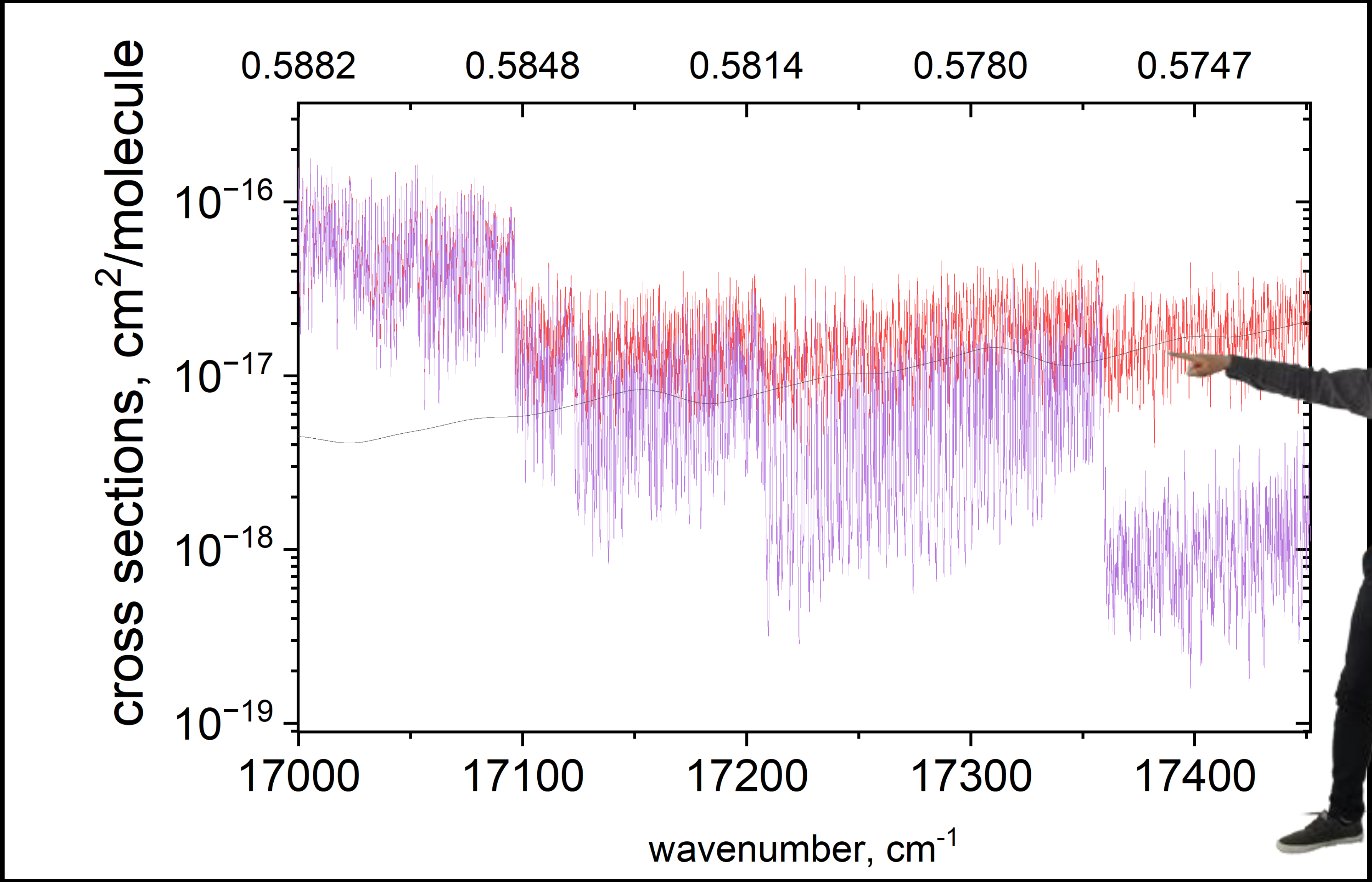

HyVO
1200 K
cross sections cm2/molecule
10-16
10-17
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
wavelength, μm
VO line list is even more a problem
If we omit non-accurate lines, the opacities drop by orders of magnitude!

One way dealing with this problem is to add the non-accurate lines into a featureless baseline

For example, for TiO in the visible

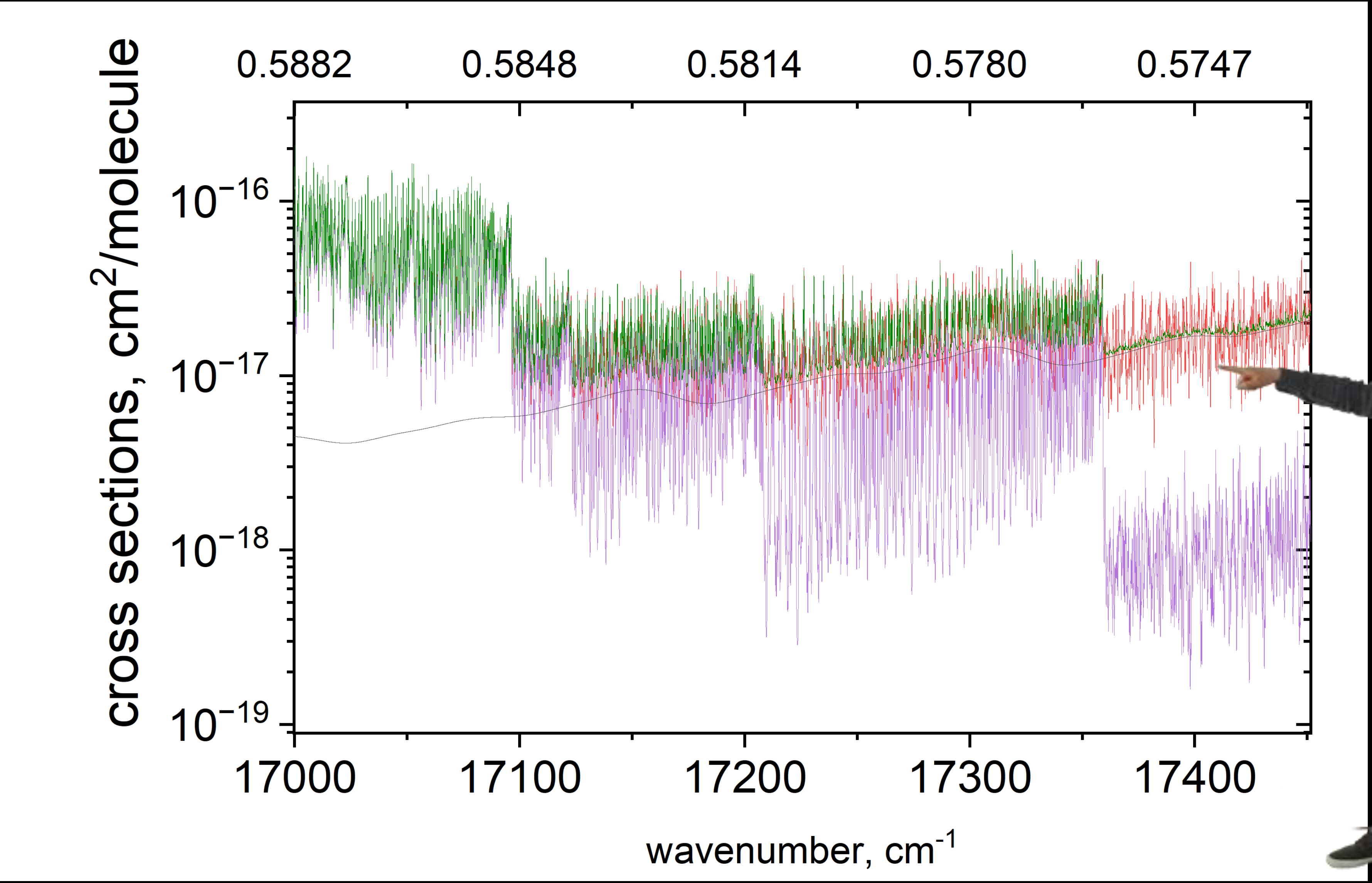


0.6250
0.5882
0.5556
0.5263
10^−15
10^−16
10^−17
10^−18
cross sections, cm^2/molecule
16000
17000
18000
19000
wavenumber, cm^-1
Purple is the spectrum from accurate (Ma) lines only

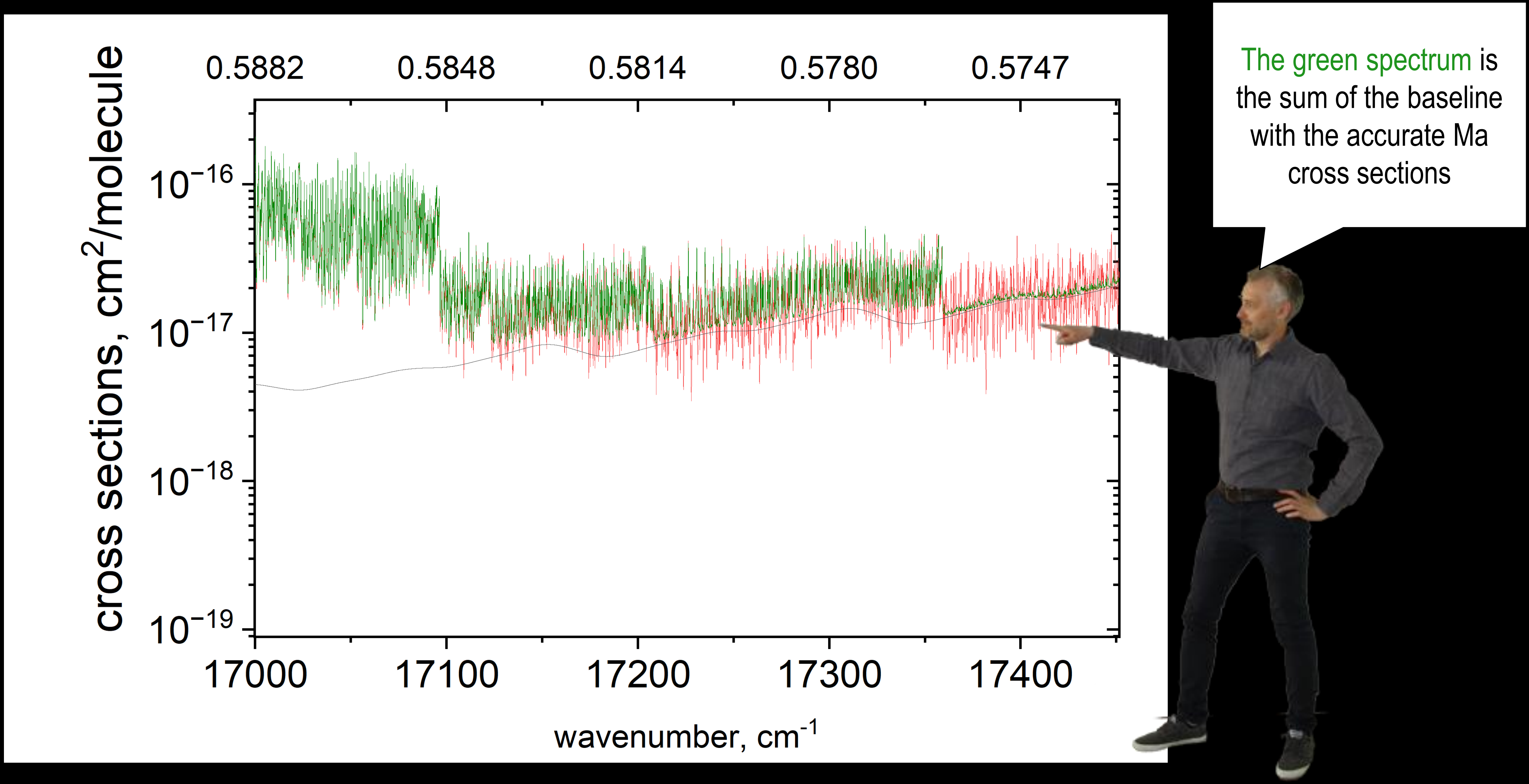


Here we add the missing not accurate transitions into the baseline (solid) binned at low resolution

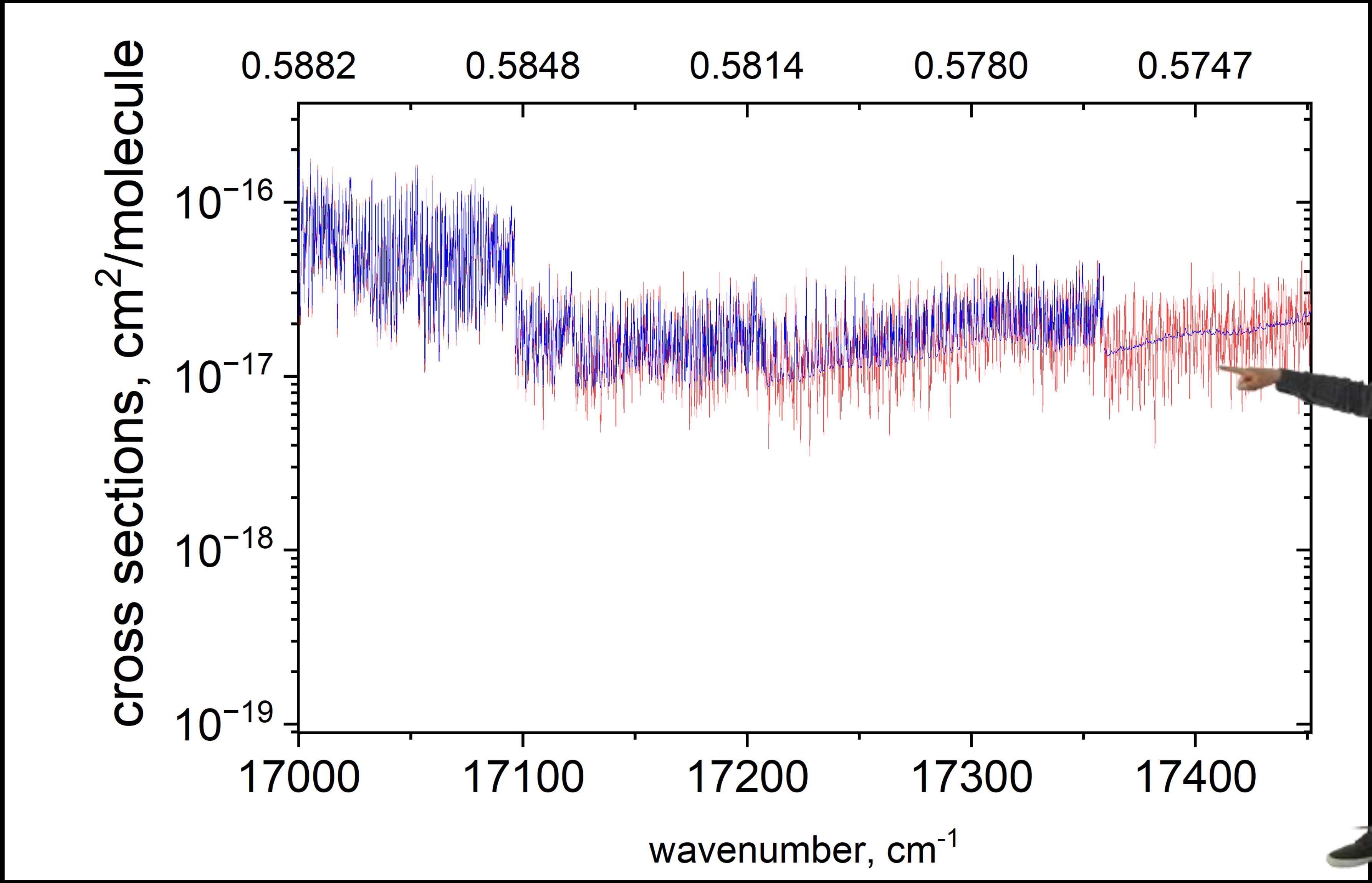
… and then add all accurate lines to this baseline


0.5882
0.5848
0.5814
0.5780
0.5747
10^{-16}
10^{-17}
10^{-18}
10^{-19}
cross sections, cm^2/molecule
17000
17100
17200
17300
17400
wavenumber, cm^{-1}

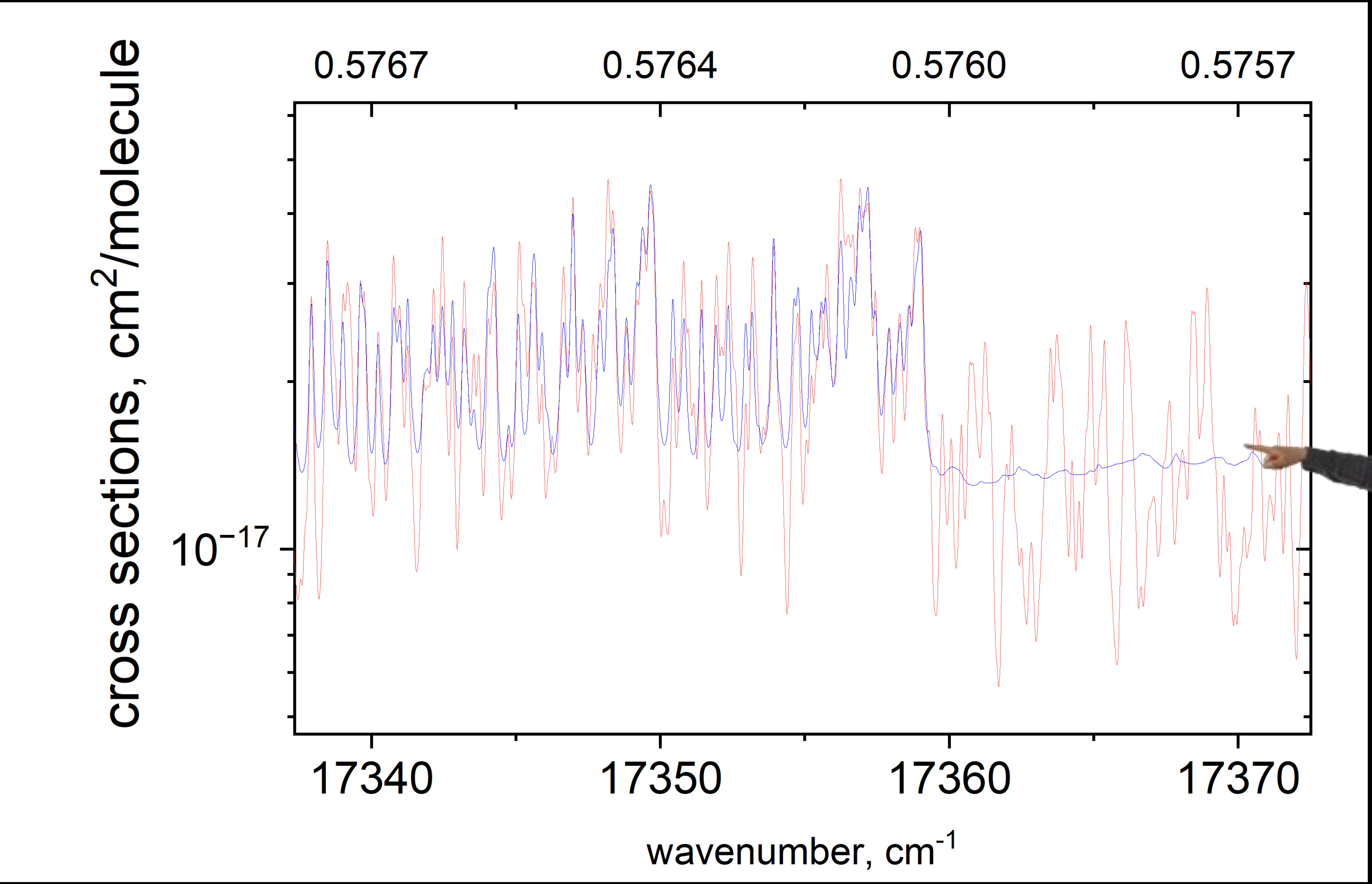

The green spectrum is the sum of the baseline with the accurate Ma cross sections



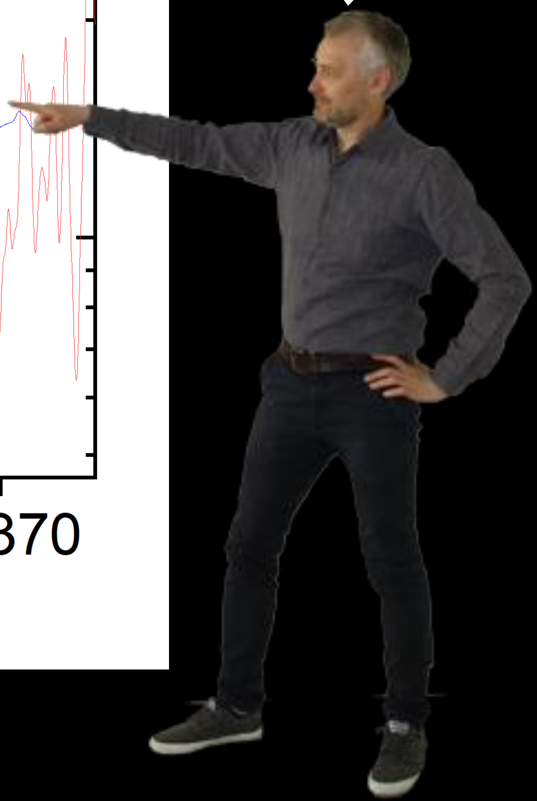
The green spectrum is the sum of the baseline with the accurate Ma cross sections
0.5882
0.5848
0.5814
0.5780
0.5747
cross sections, cm^2/molecule
10^{-16}
10^{-17}
10^{-18}
10^{-19}
17000
17100
17200
17300
17400
wavenumber, cm^{-1}



The blue spectrum is the spectrum to be used in the High Res studies



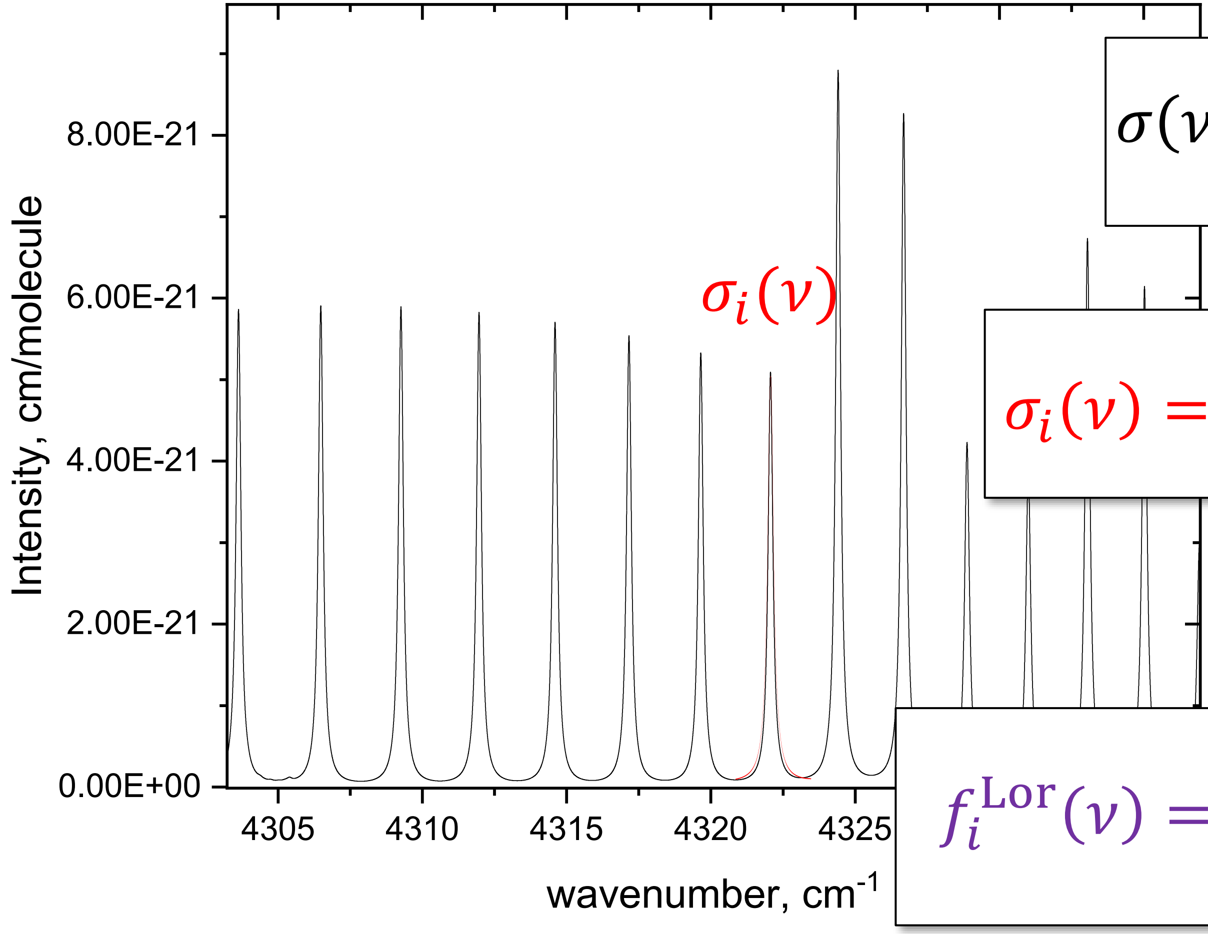
Let’s zoom-in: blue is the reduced spectrum which does not contain non-MARVALised lines



Does it make sense?

Let me suggest something different

But first, I would like to understand how the uncertainties of line positions affect the cross sections



$$\sigma(\nu) = \sum_i \sigma_i(\nu)$$

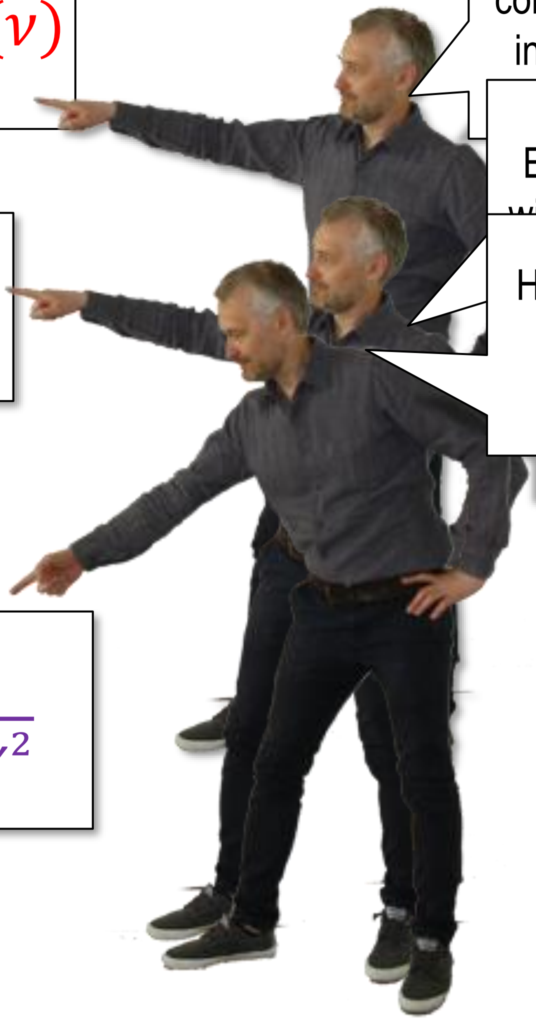
$$\sigma_i(\nu) = I_i \, f_i^{\mathrm{Lor}}(\nu)$$

$$f_i^{\mathrm{Lor}}(\nu) = \frac{\gamma}{\pi} \frac{1}{(\nu - \nu_i)^2 + \gamma^2}$$

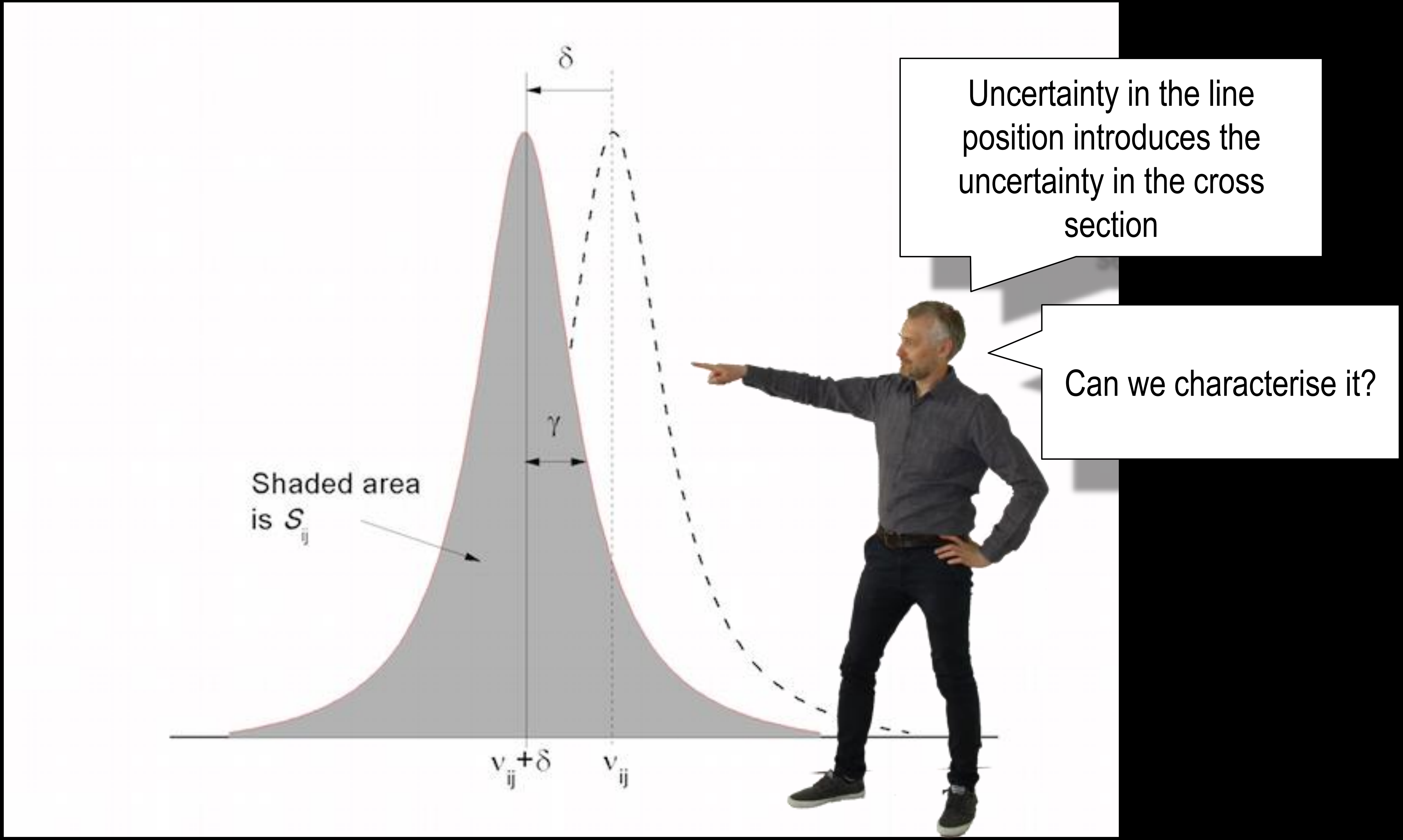
Cross sections are constructed as a sum of individual line profiles

Each line contributes

Here is a Lorentzian line profile (normalised to 1)



Uncertainty δ

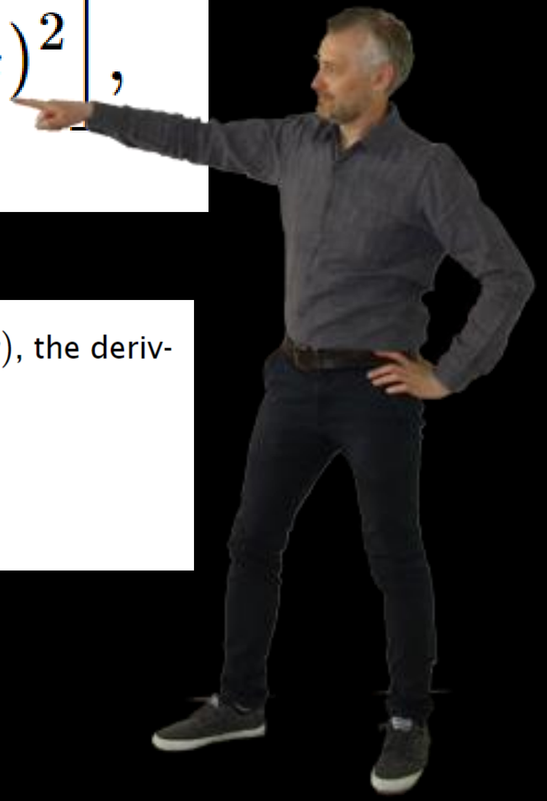


The obvious method is to propagate the error

Let’s propagate error through the standard error-propagation technique:

We need to evaluate the 1st derivatives of the cross sections wrt the line positions and combine them with the associated errors of each energy involved

$$(\Delta\sigma(\tilde{\nu}))^2 = \sum_{i,j}\left(\frac{\partial\sigma(\tilde{\nu})}{\partial\tilde{\nu}_{ij}}\right)^2\left[(\Delta\tilde{E}_i)^2+(\Delta\tilde{E}_j)^2\right],$$



where $\Delta\tilde{E}_i$ and $\Delta\tilde{E}_j$ are the uncertainties of the upper and lower states. Assuming a given line-profile $f(\tilde{\nu})$, the derivative wrt the energy is given by

$$\frac{\partial\sigma(\tilde{\nu})}{\partial\tilde{\nu}_{ij}} = I_{if}\frac{\partial f(\tilde{\nu})}{\partial\tilde{\nu}_{ij}}.$$

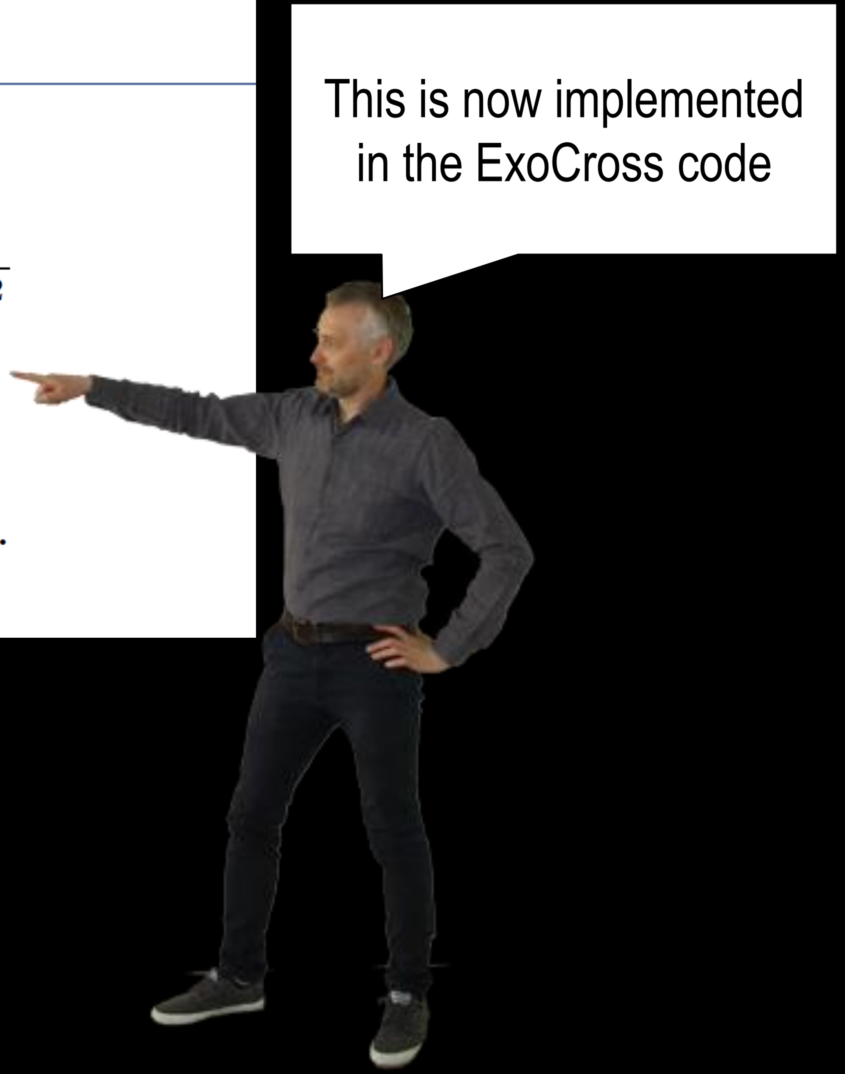
Error cross sections for the Lorentzian profile (Elorentz)

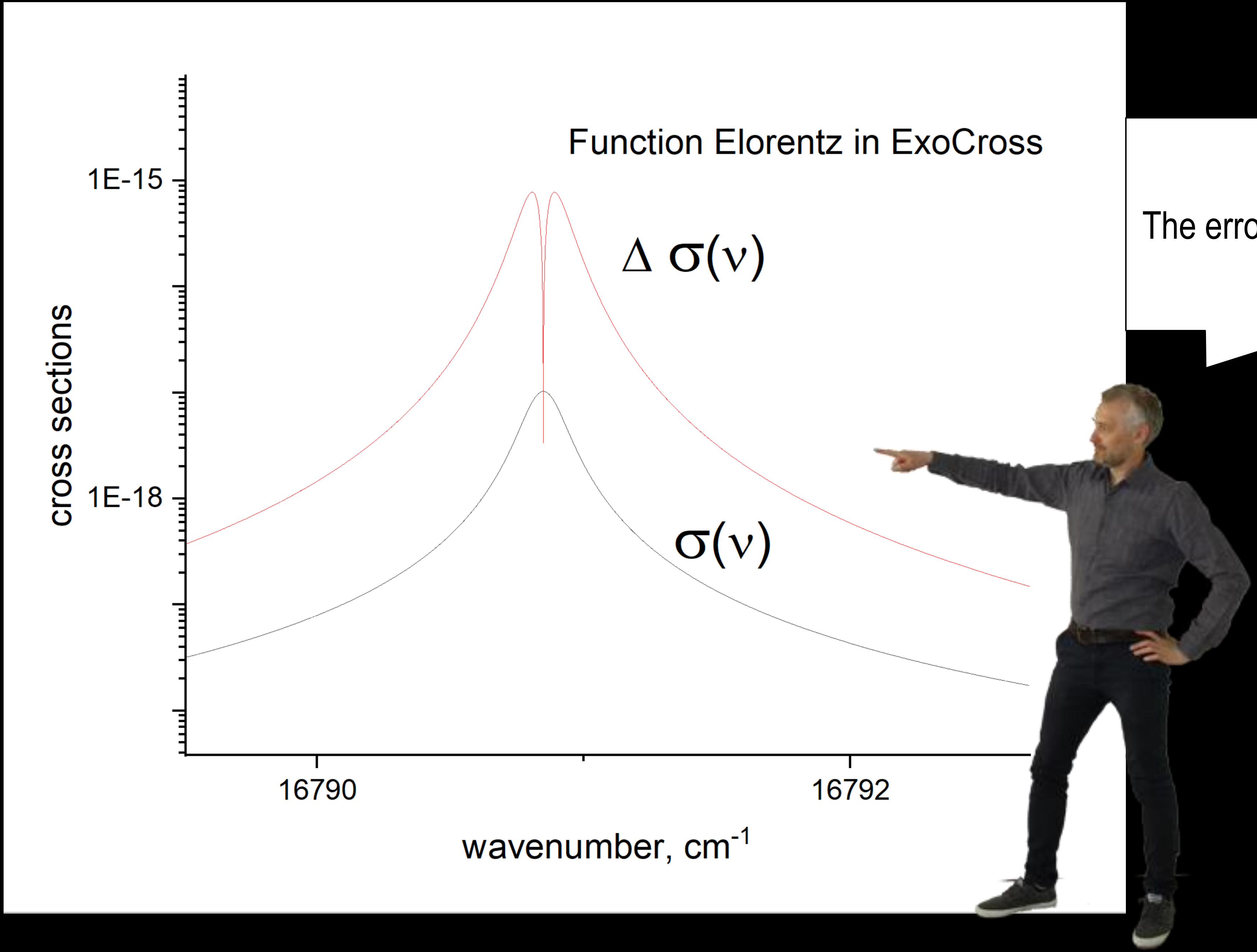
For the Lorentzian line profile centred at $\tilde{\nu}_{ij}$ with HWHM γ given by

$$f(\tilde{\nu}, \tilde{\nu}_{ij}, \gamma)_{\rm Lo} = \frac{\gamma}{\pi} \frac{1}{(\tilde{\nu} - \tilde{\nu}_{ij})^2 + \gamma^2}$$

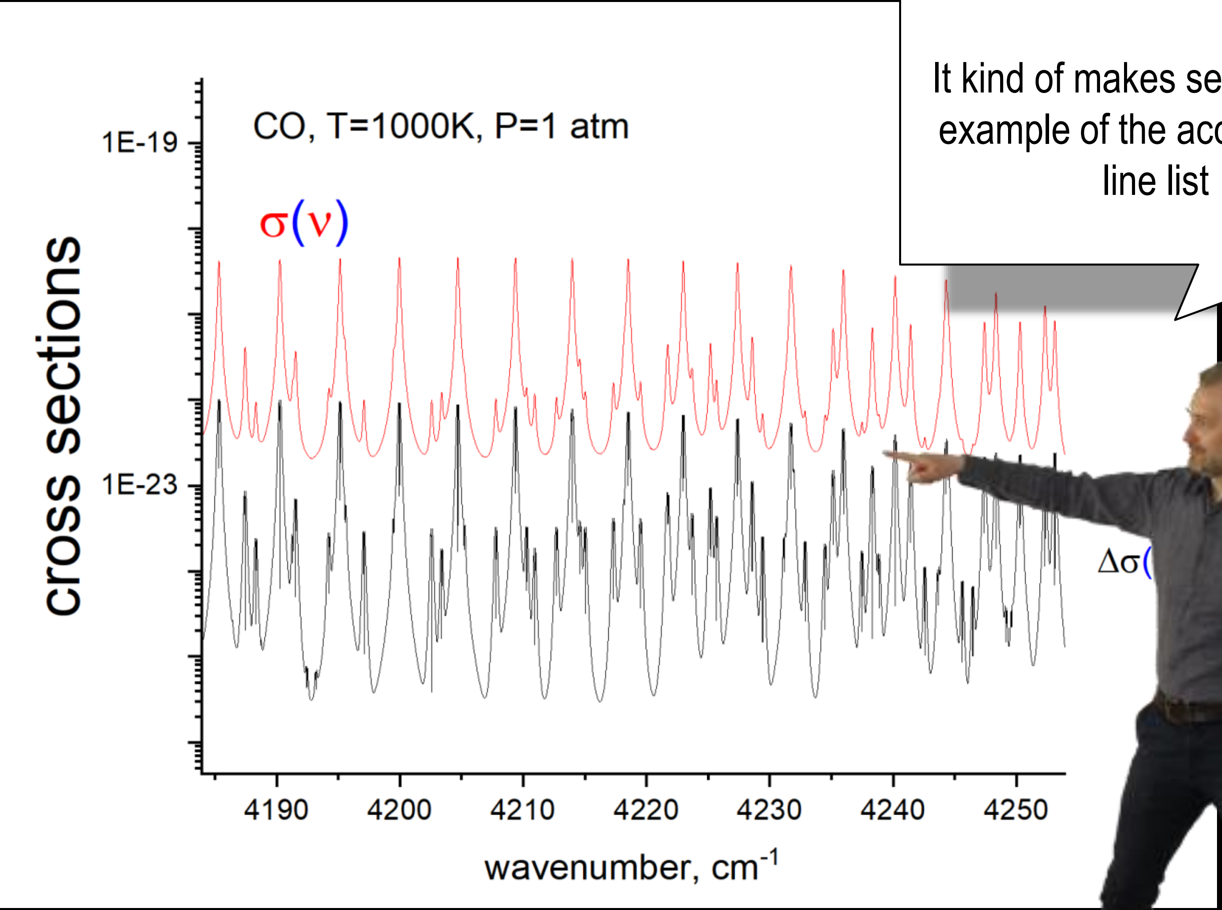
the corresponding derivative wrt $\tilde{\nu}_{ij}$ is given by

$$\frac{\partial f(\tilde{\nu})_{\rm Lo}}{\partial \tilde{\nu}_{ij}} = \frac{\gamma}{\pi} \frac{2(\tilde{\nu}_{ij} - \tilde{\nu})}{\left[(\tilde{\nu} - \tilde{\nu}_{ij})^2 + \gamma^2\right]^2}.$$

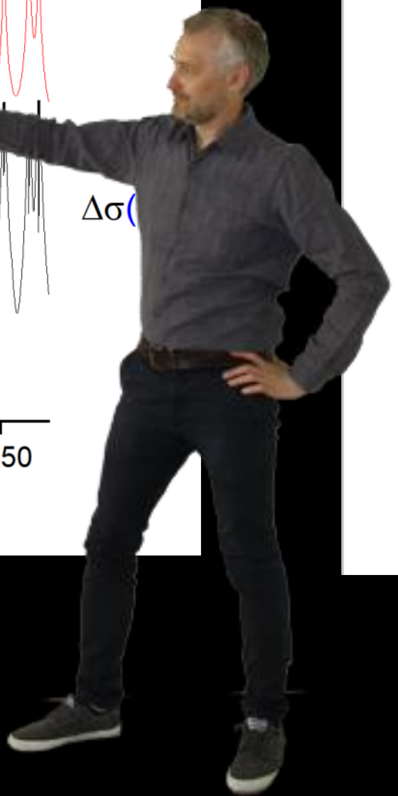


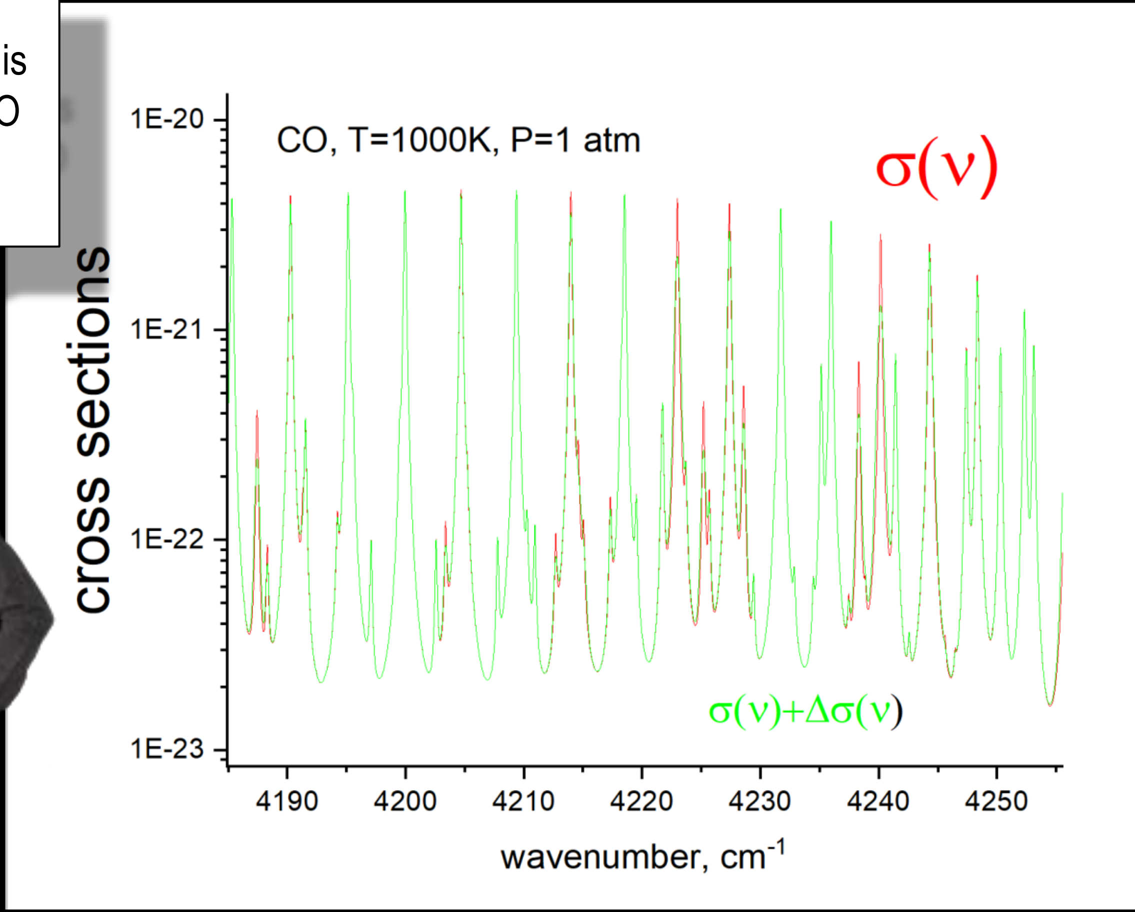


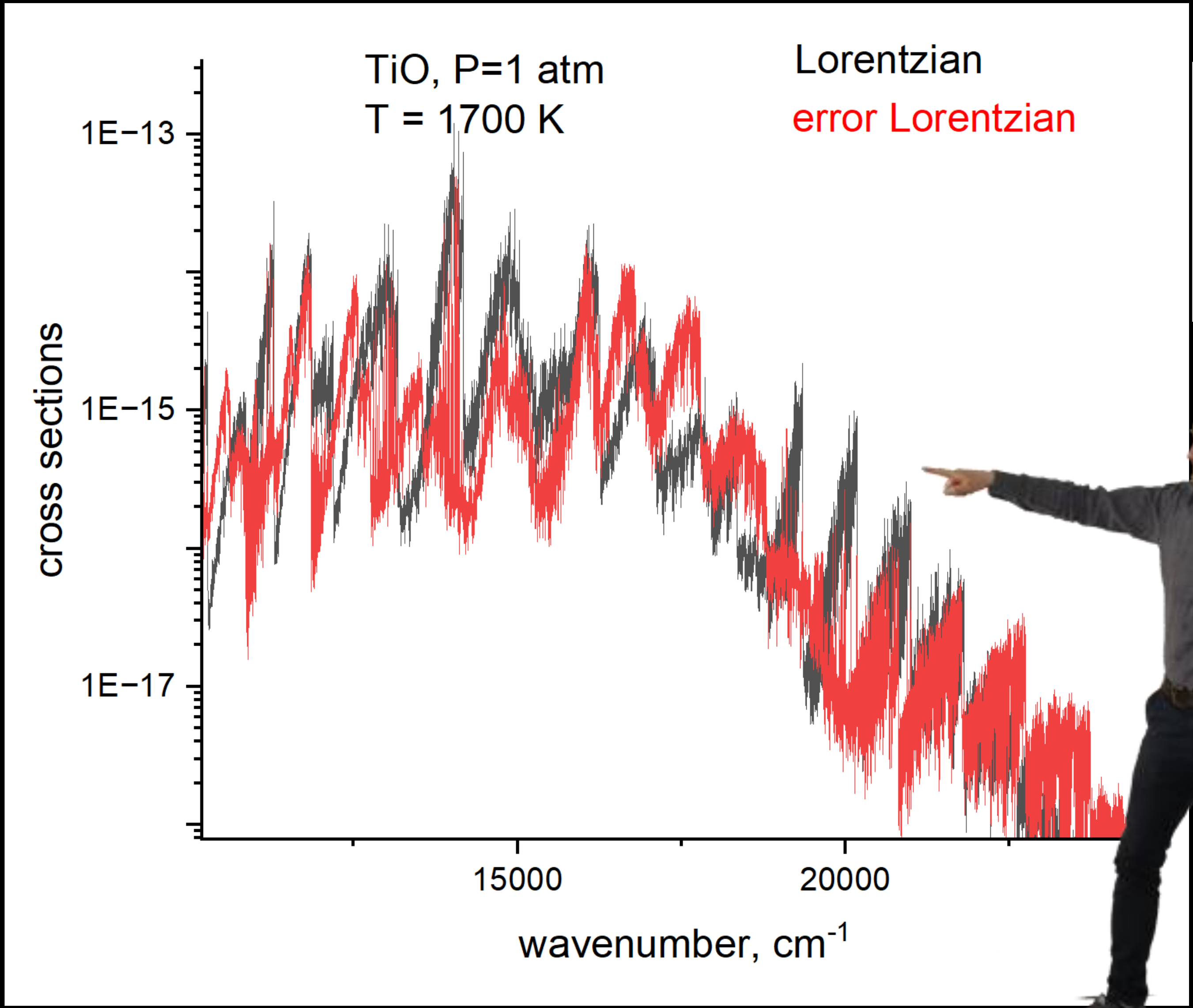
The error function looks strange



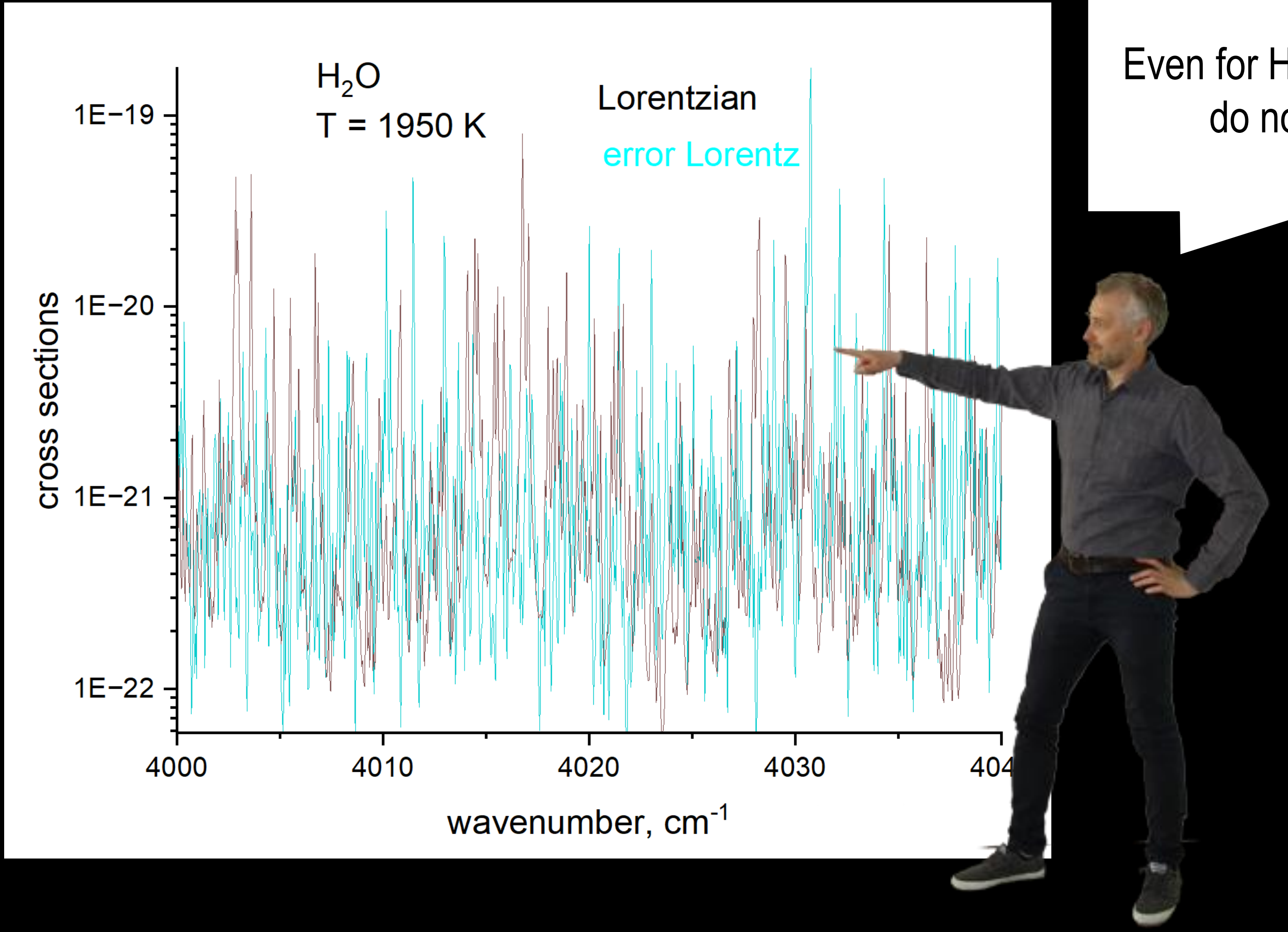
It kind of makes sense in this example of the accurate CO line list







Not so much for less accurate line list for TiO



Even for H_2O some error bars do not make sense

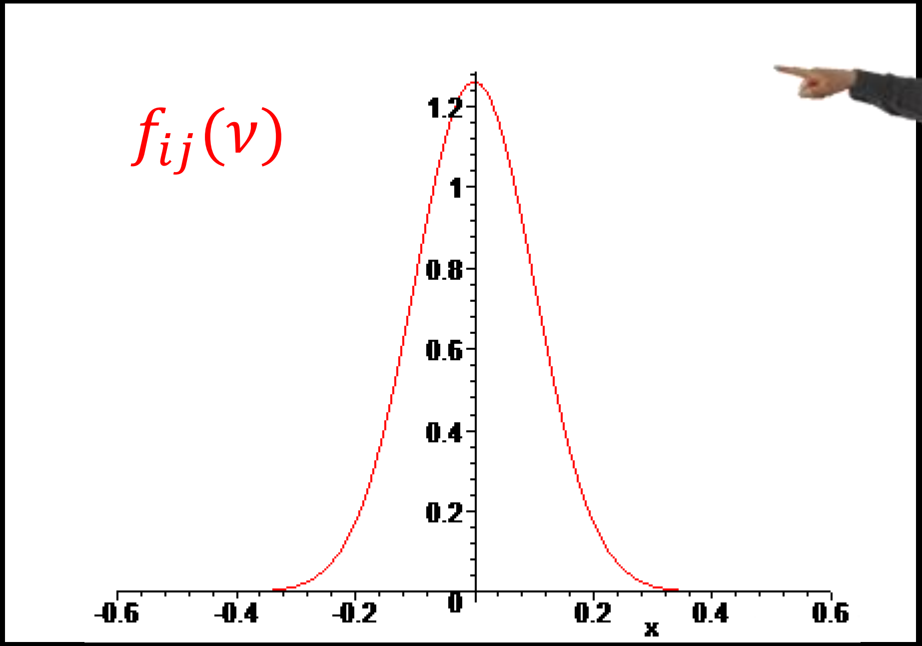
Well, it does not work because it is only valid for small errors

$$(\Delta\sigma(\tilde{\nu}))^2 = \sum_{i,j} \left(\frac{\partial\sigma(\tilde{\nu})}{\partial\tilde{\nu}_{ij}} \right)^2 \left[(\Delta\tilde{E}_i)^2 + (\Delta\tilde{E}_j)^2 \right],$$

Here is a different idea

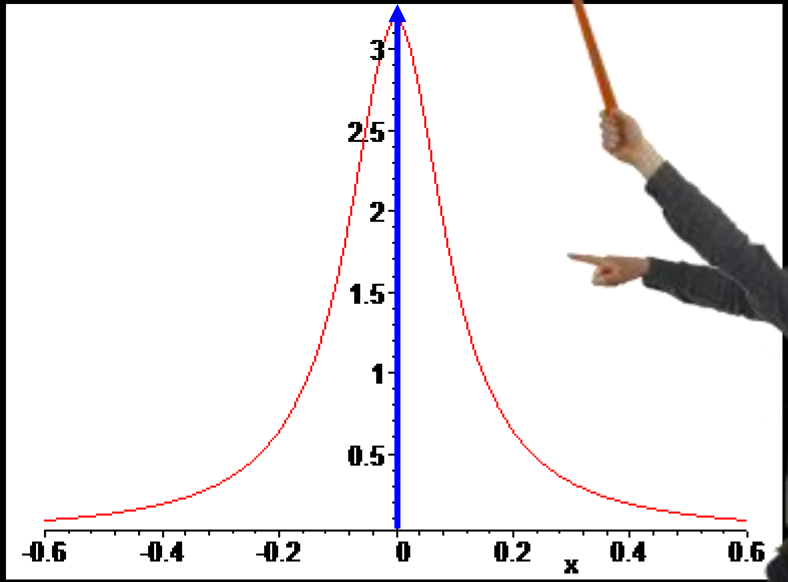
Let's assume that uncertainties σ in line positions represent a normal distribution of associated errors

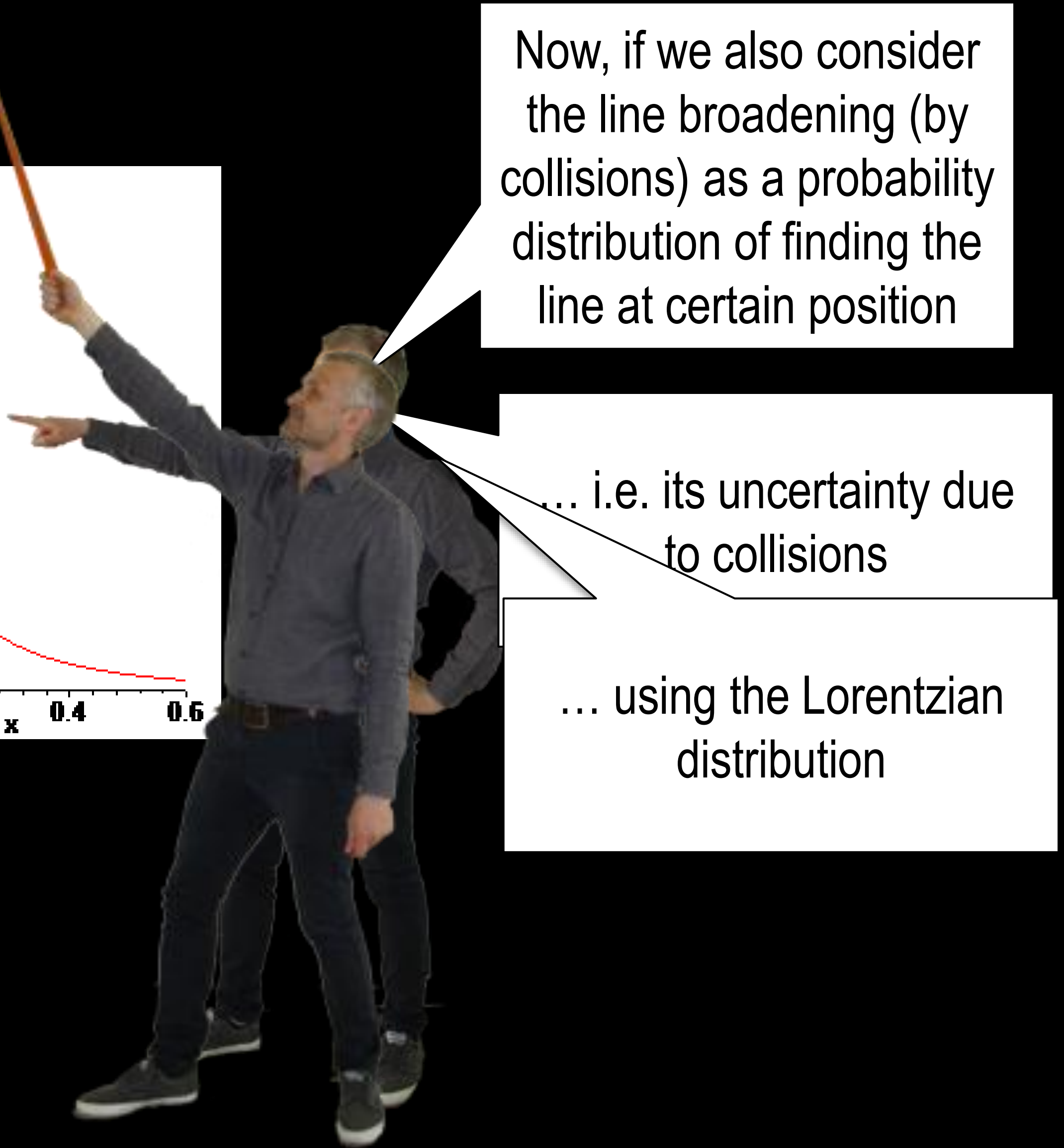
$$f_{\text{line}}(\nu, \nu_{ij}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(\nu_{ij}-\nu)^2}{2\sigma^2}}$$



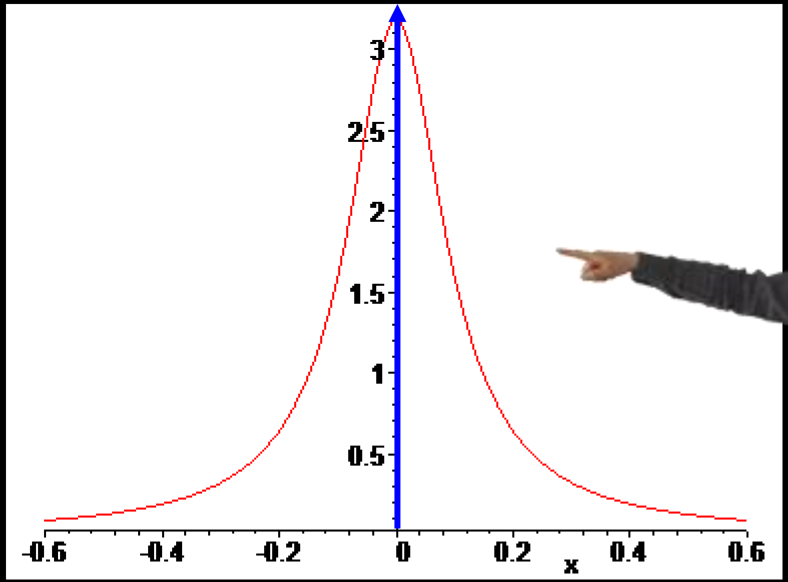
Uncertainty of our knowledge of the line position

$$f_i^{\mathrm{Lor}}(\nu) = \frac{\gamma}{\pi}\frac{1}{(\nu-\nu_i)^2+\gamma^2}$$





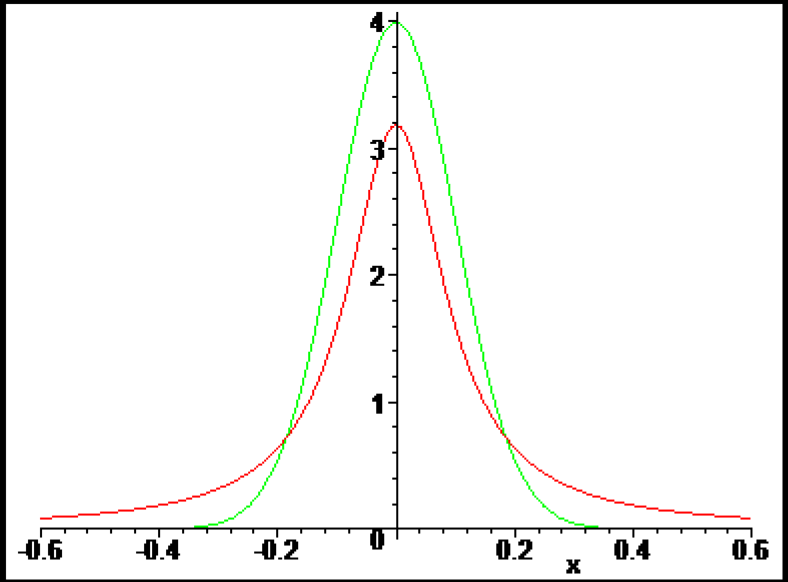
$$f_i^{\mathrm{Lor}}(\nu) = \frac{\gamma}{\pi}\frac{1}{(\nu-\nu_i)^2+\gamma^2}$$



Now remember that we are also not certain in our knowledge of the line position

The line can be anywhere around its predicted centre with an uncertainty σ

$$f_i^{\mathrm{Lor}}(\nu) = \frac{\gamma}{\pi} \frac{1}{(\nu - \nu_i)^2 + \gamma^2}$$

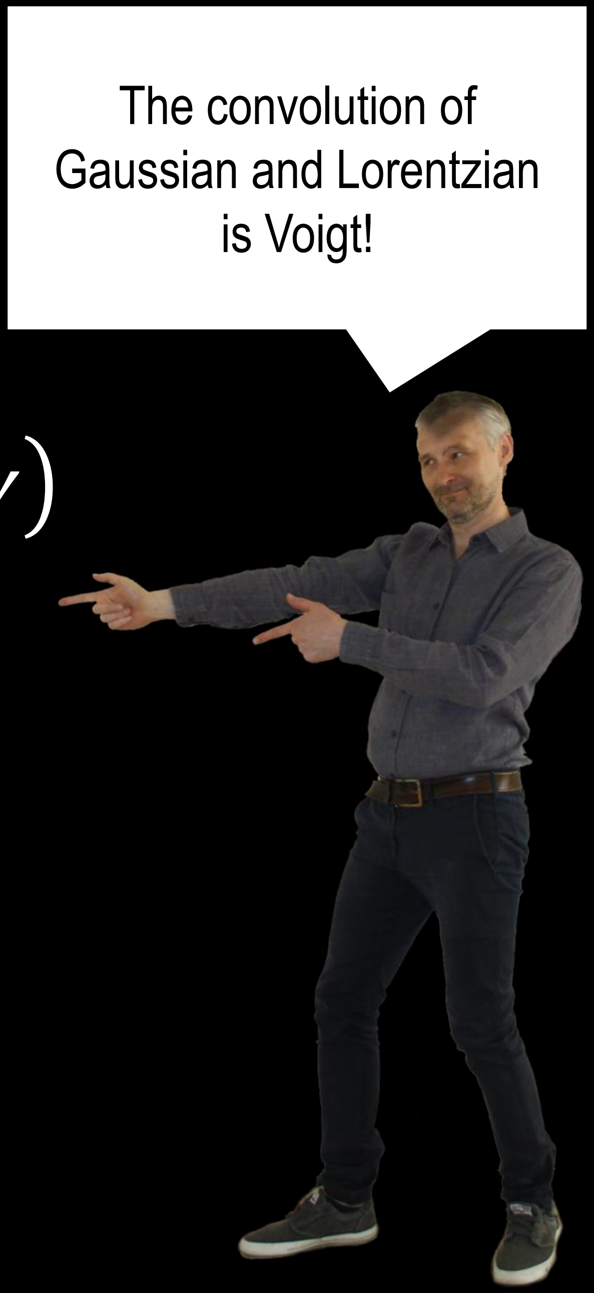


.. which satisfies the normal distribution

$$f_{\mathrm{line}}(\nu, \nu_{ij}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\, \mathrm{e}^{-\frac{\left(\nu_{ij} - \nu\right)^2}{2\sigma^2}}$$

Let’s convolve our uncertainties of the line positions due to our knowledge (Gaussian) of collision effects (Lorentz)

$$f_{\mathrm{tot}}(\nu, \nu_{ij}, \sigma, \gamma) = f^{\mathrm{G}}(\nu, \nu_{ij}, \sigma) \otimes f^{\mathrm{L}}(\nu, \nu_{ij}, \gamma)$$



All we need to do is to compute spectra using the standard Voigt profile with the collisional line widths γ and the uncertainty value σ, combined with the Doppler $\sigma_{\mathrm{Doppler}}$,

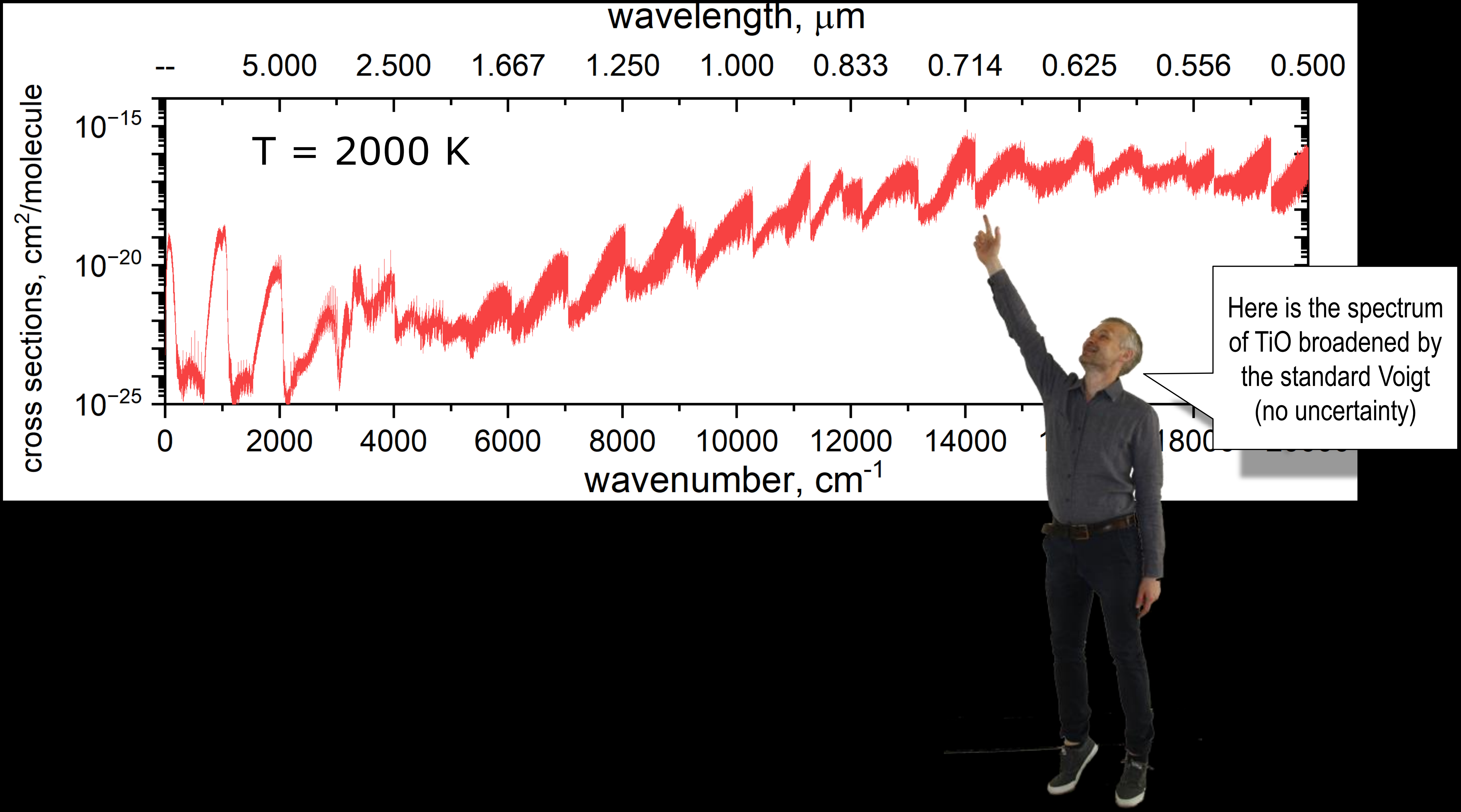
i.e. $\sigma + \sigma_{\mathrm{Doppler}}$

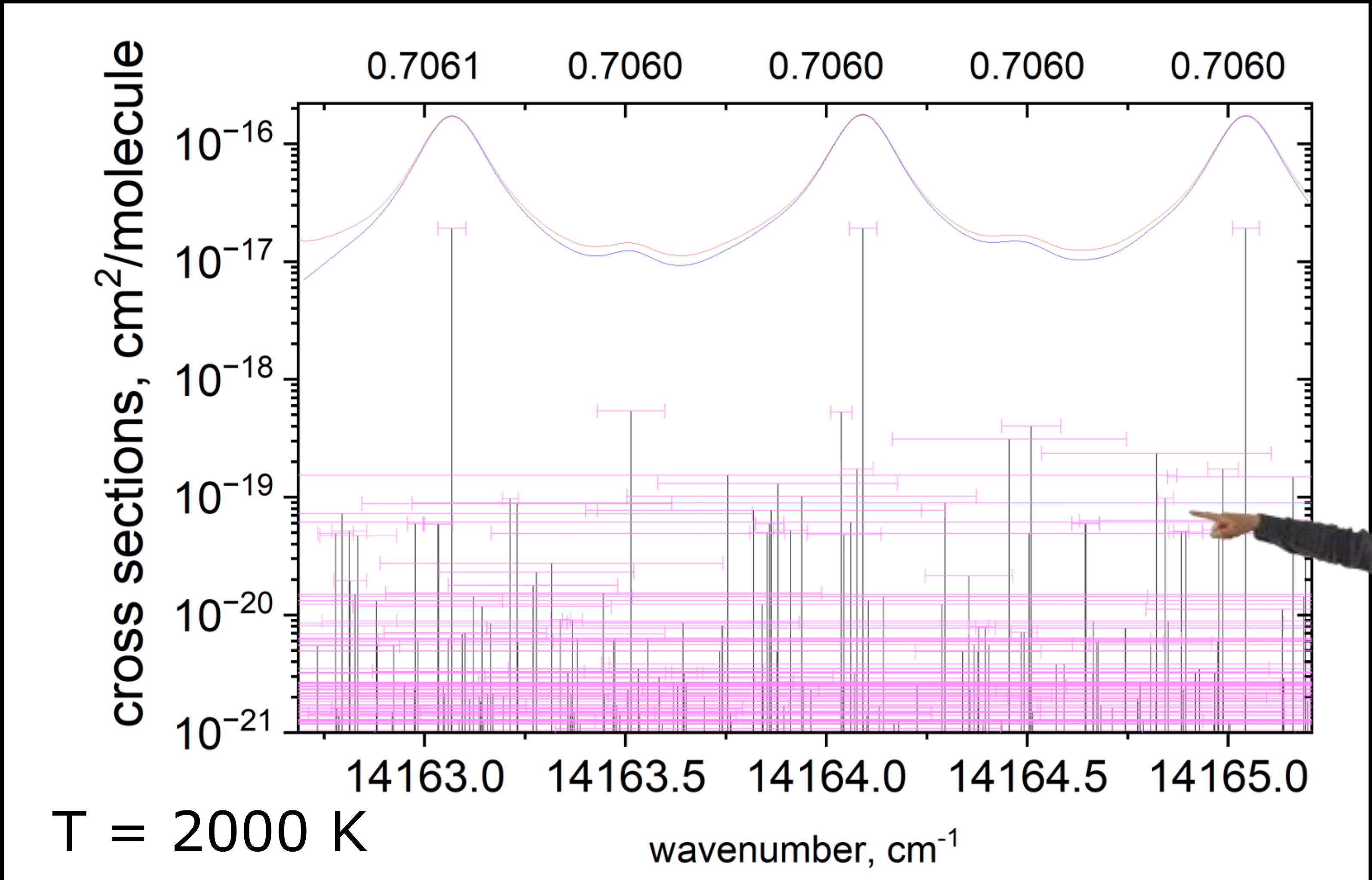
Cross sections with uncertainty is a convolution of the Lorentzian (uncertainty due to collisions) and Gaussian (uncertainty due to our knowledge)

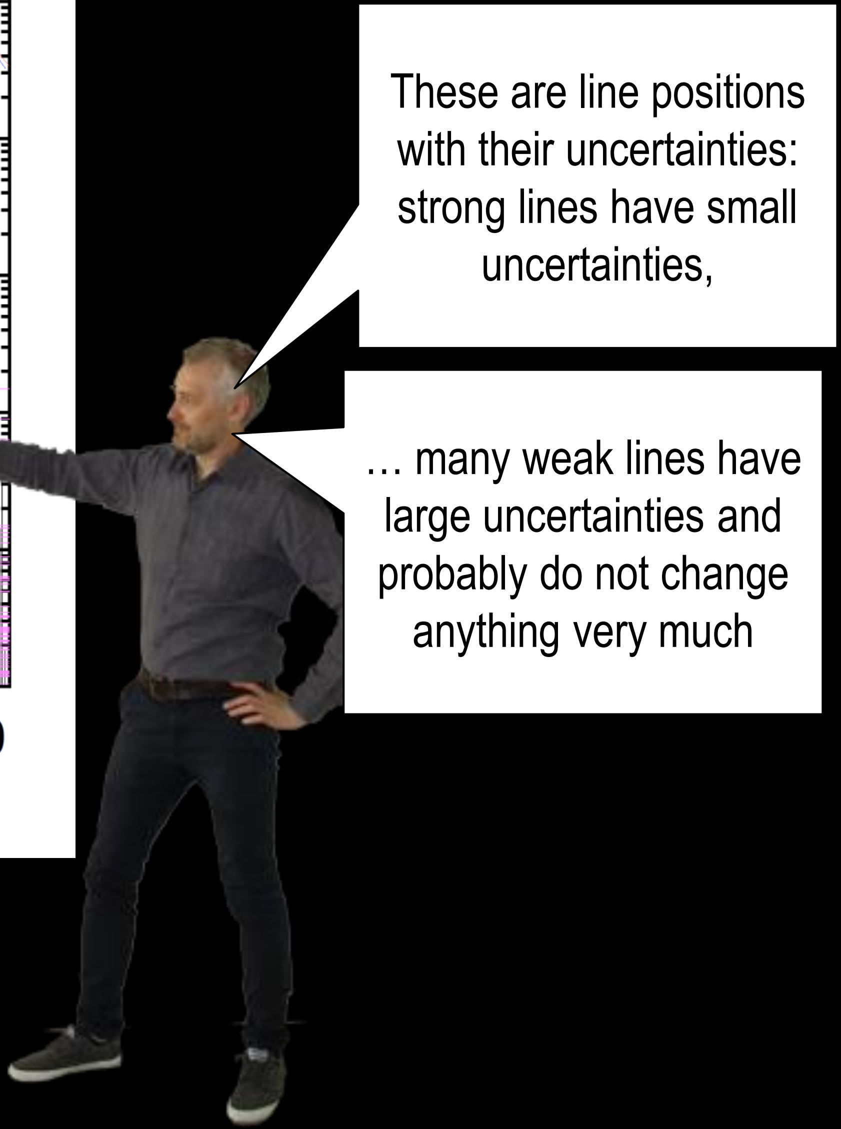
$$\sigma_{tot}(\nu) = \int_0^\infty f^{(\mathrm{Gauss-Err})}(\nu - \nu')\, f^{\mathrm{Lorentz}}(\nu') d\nu'$$

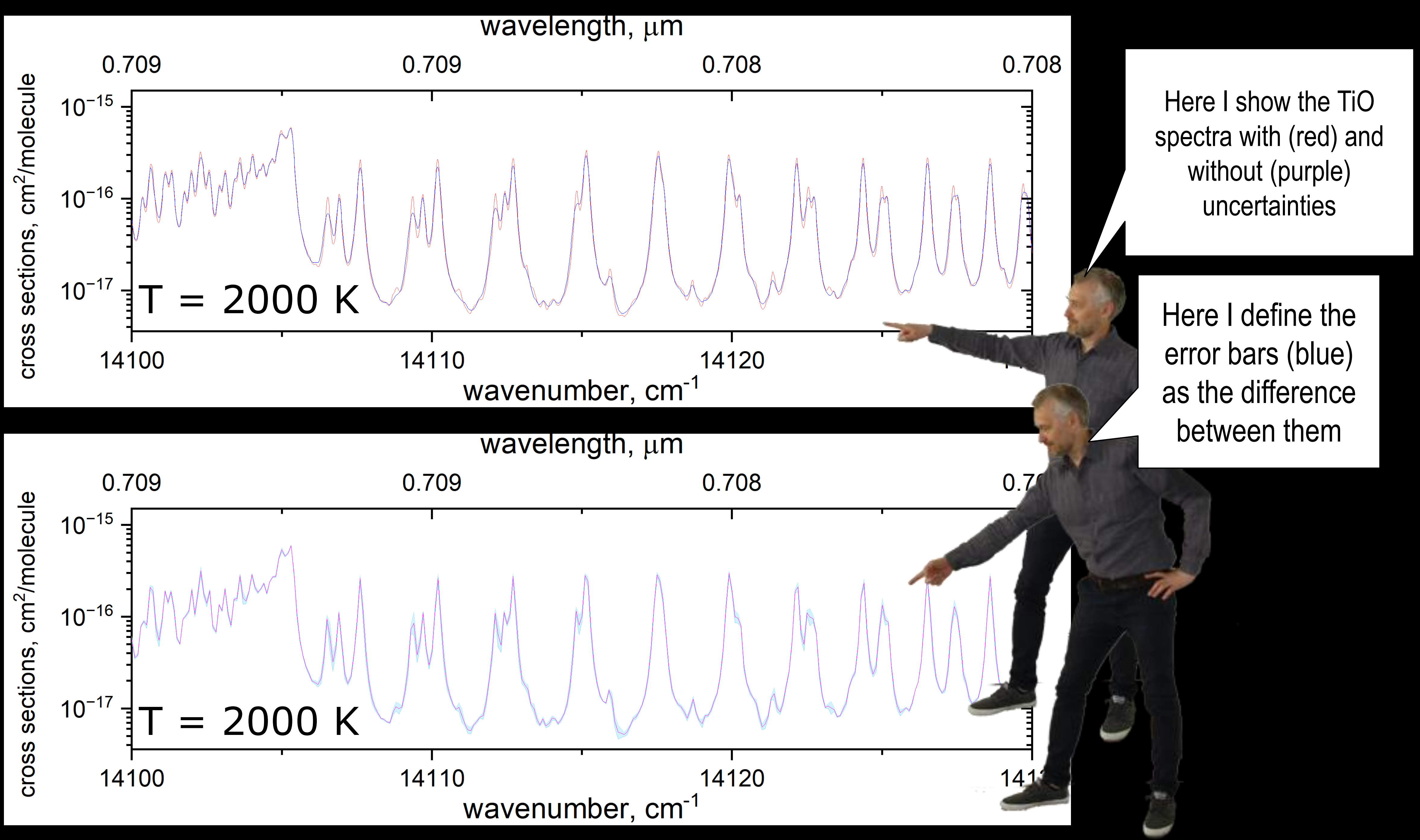
This is now implemented in ExoCross

Results: TiO


wavelength, μm
-- 5.000 2.500 1.667 1.250 1.000 0.833 0.714 0.625 0.556 0.500
T = 2000 K
cross sections, cm²/molecule
10^{-15}
10^{-20}
10^{-25}
0 2000 4000 6000 8000 10000 12000 14000
wavenumber, cm⁻¹
Here is the spectrum of TiO broadened by the standard Voigt (no uncertainty)

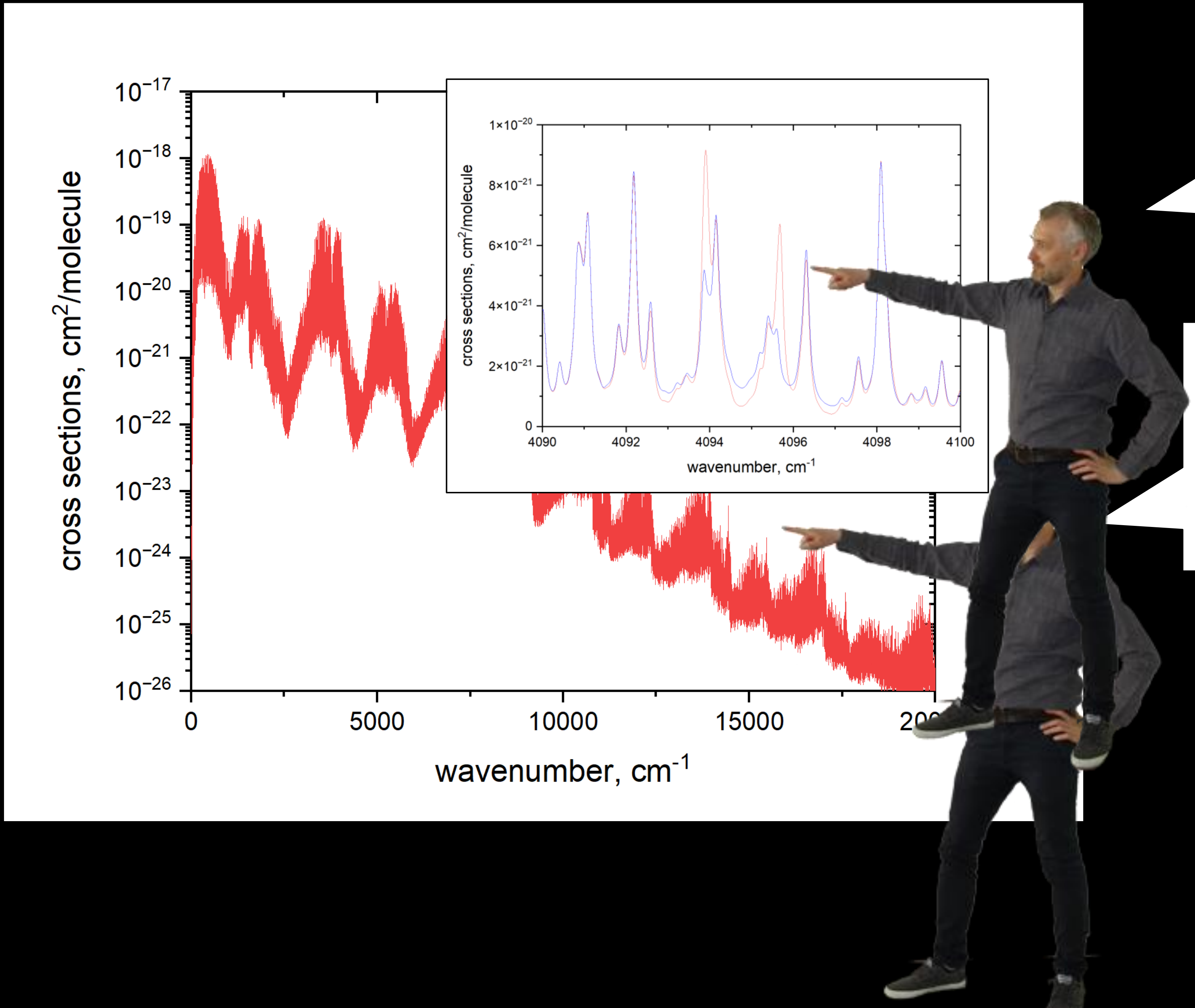

0.7061
0.7060
0.7060
0.7060
0.7060
cross sections, cm²/molecule
10⁻¹⁶
10⁻¹⁷
10⁻¹⁸
10⁻¹⁹
10⁻²⁰
10⁻²¹
14163.0
14163.5
14164.0
14164.5
14165.0
wavenumber, cm⁻¹
T = 2000 K


These are line positions with their uncertainties: strong lines have small uncertainties,
… many weak lines have large uncertainties and probably do not change anything very much

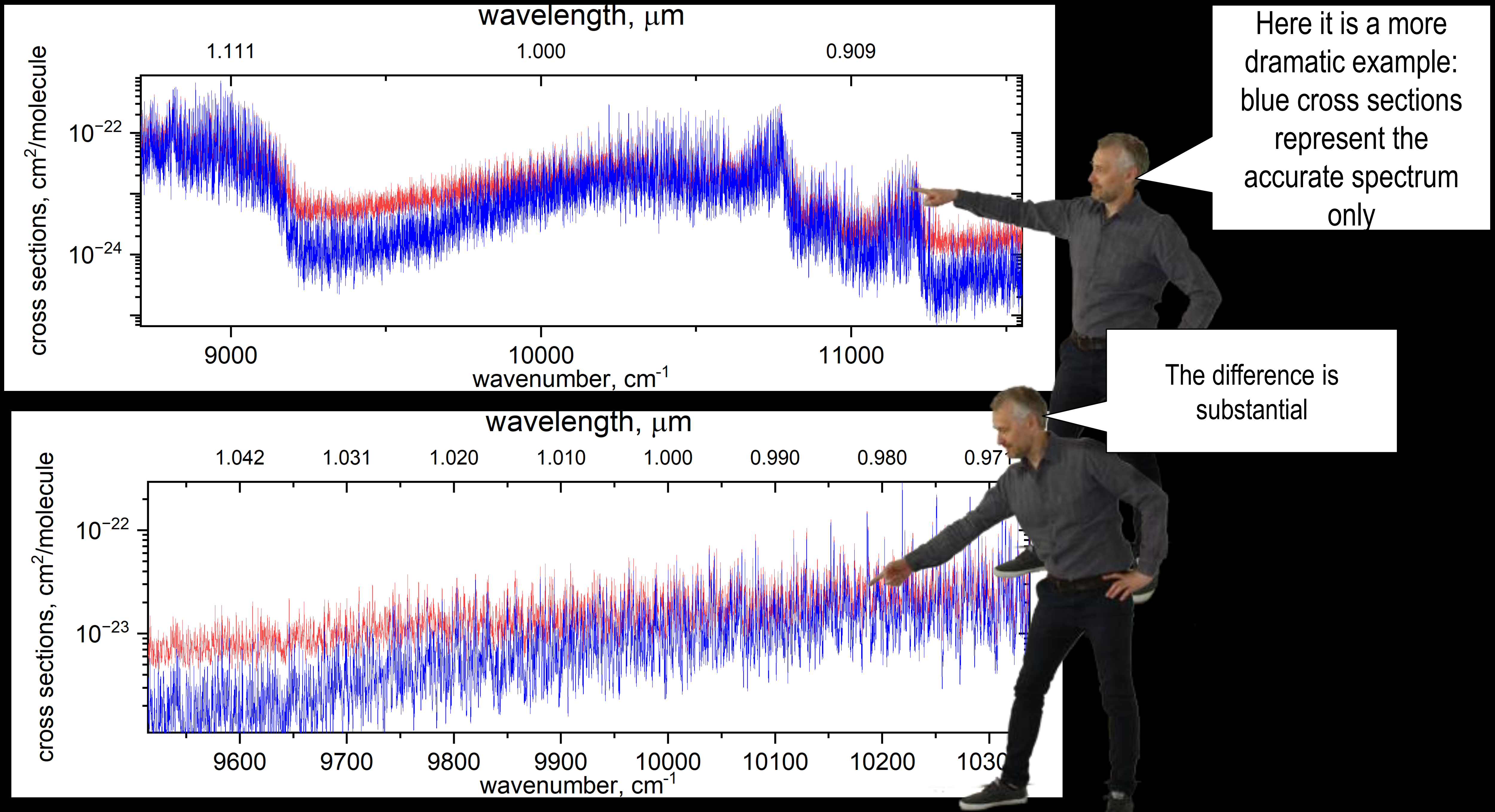

wavelength, μm
0.709
0.709
0.708
0.708
cross sections, cm²/molecule
10⁻¹⁵
10⁻¹⁶
10⁻¹⁷
T = 2000 K
14100
14110
14120
wavenumber, cm⁻¹
wavelength, μm
0.709
0.709
0.708
cross sections, cm²/molecule
10⁻¹⁵
10⁻¹⁶
10⁻¹⁷
T = 2000 K
14100
14110
14120
wavenumber, cm⁻¹
Here I show the TiO spectra with (red) and without (purple) uncertainties
Here I define the error bars (blue) as the difference between them

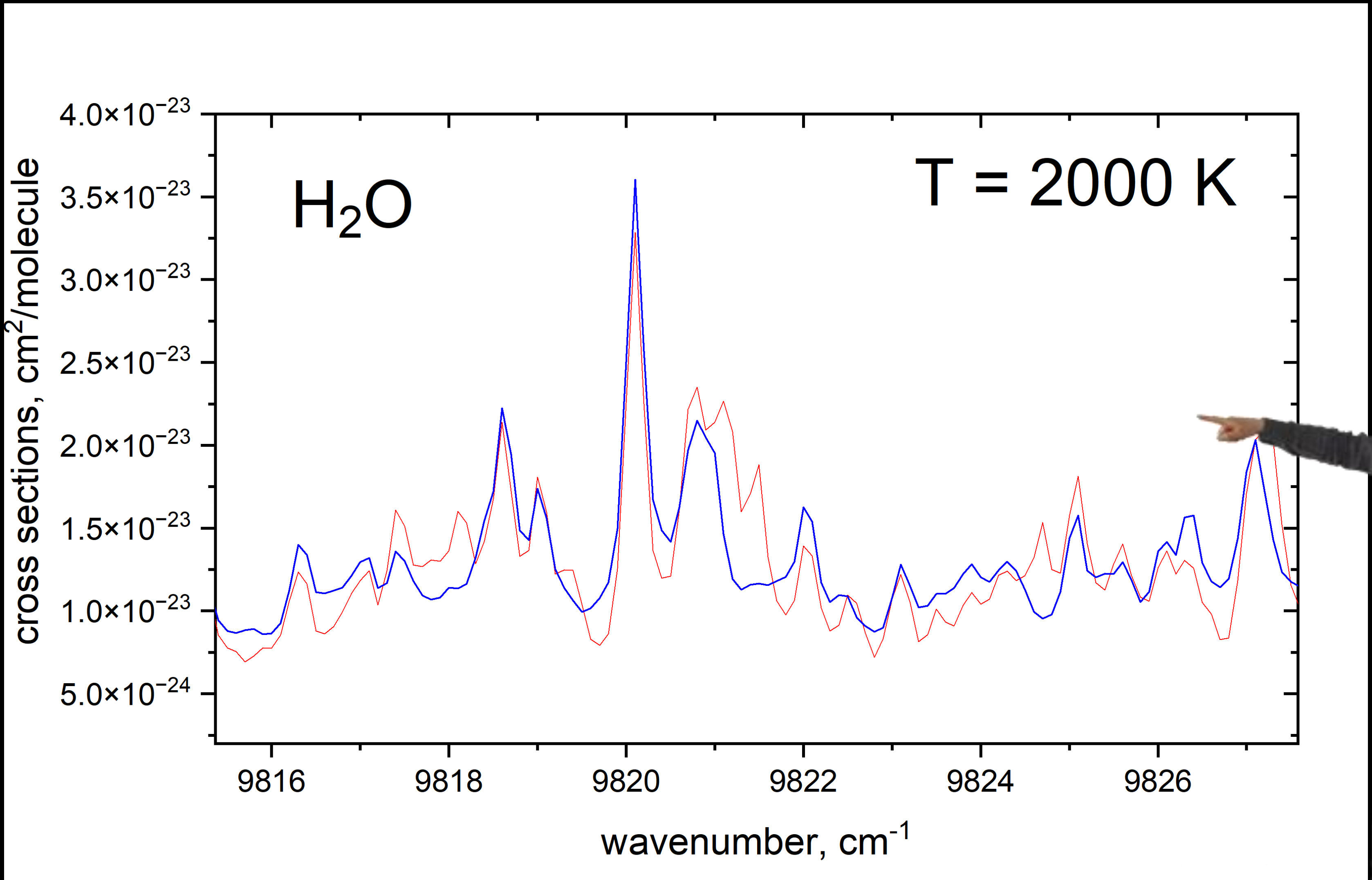
Results: Water

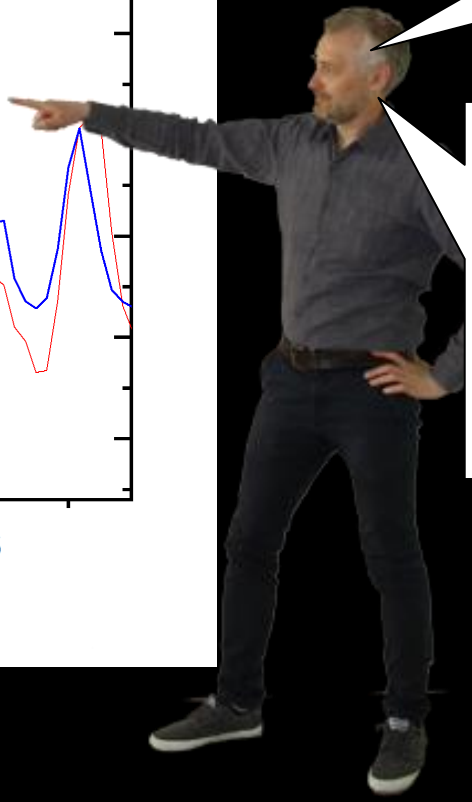
… and focus on a small region containing a few lines. Here red are the full spectrum and blue the accurate (Ma) contributions



Let's take the water spectrum at T=2000 K

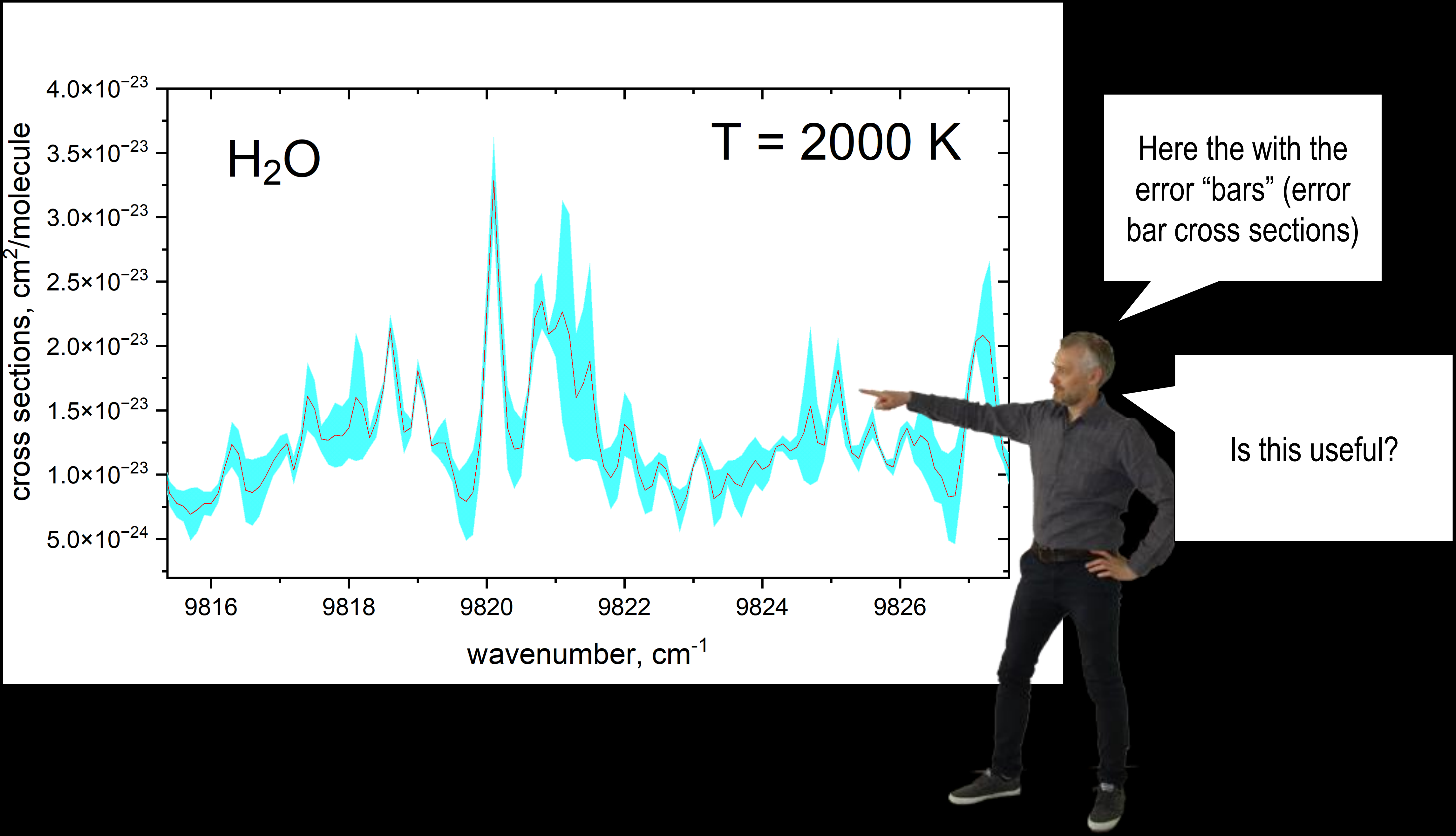

wavelength, μm
1.111
1.000
0.909
10−22
10−24
cross sections, cm2/molecule
9000
10000
11000
wavenumber, cm-1
Here it is a more dramatic example: blue cross sections represent the accurate spectrum only
wavelength, μm
1.042
1.031
1.020
1.010
1.000
0.990
0.980
0.971
10−22
10−23
cross sections, cm2/molecule
9600
9700
9800
9900
10000
10100
10200
1030
wavenumber, cm-1
The difference is substantial





Here the spectrum in blue is with the uncertainty and red is without

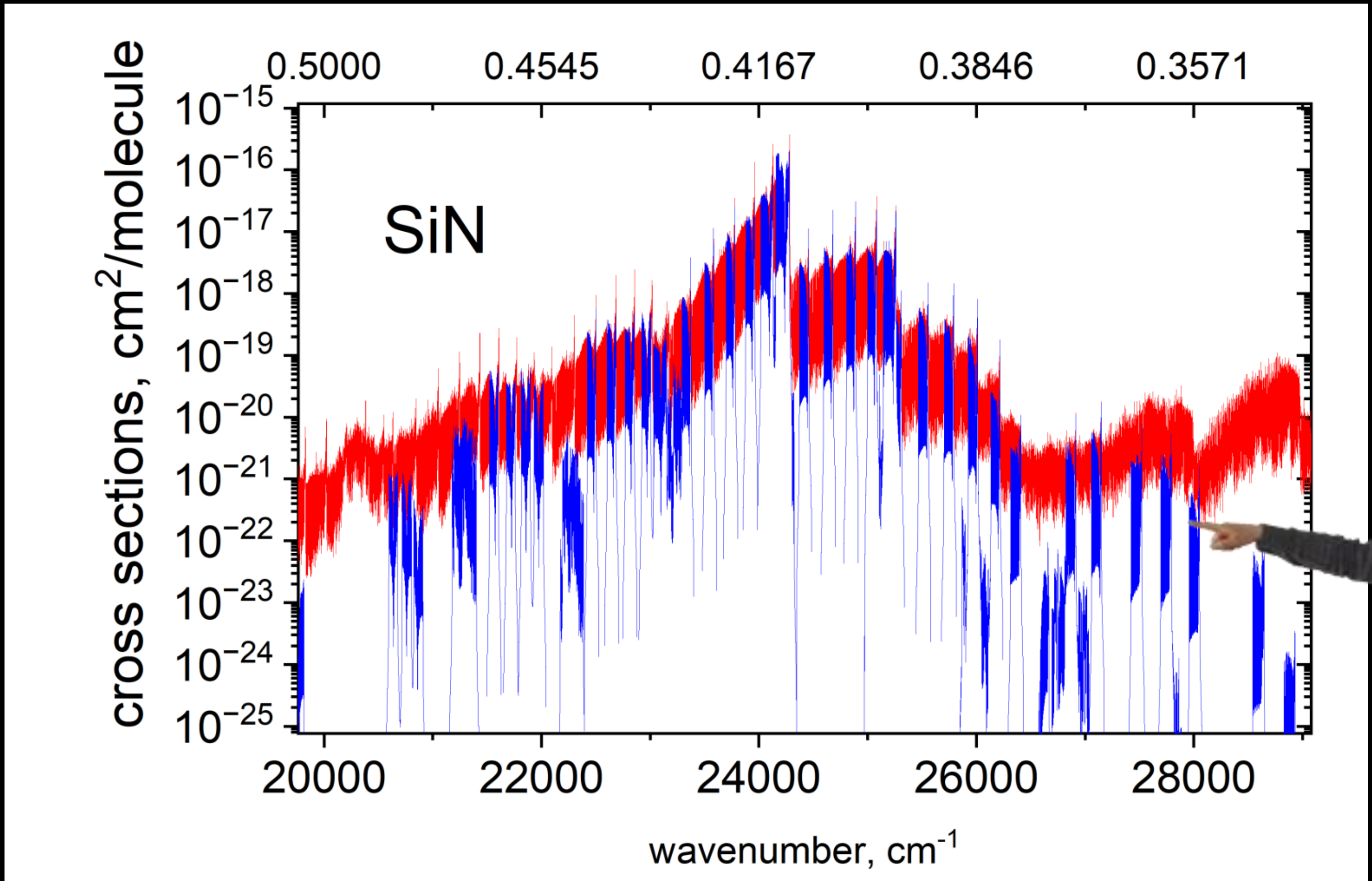
I would suggest using the blue spectrum in high resolution retrievals



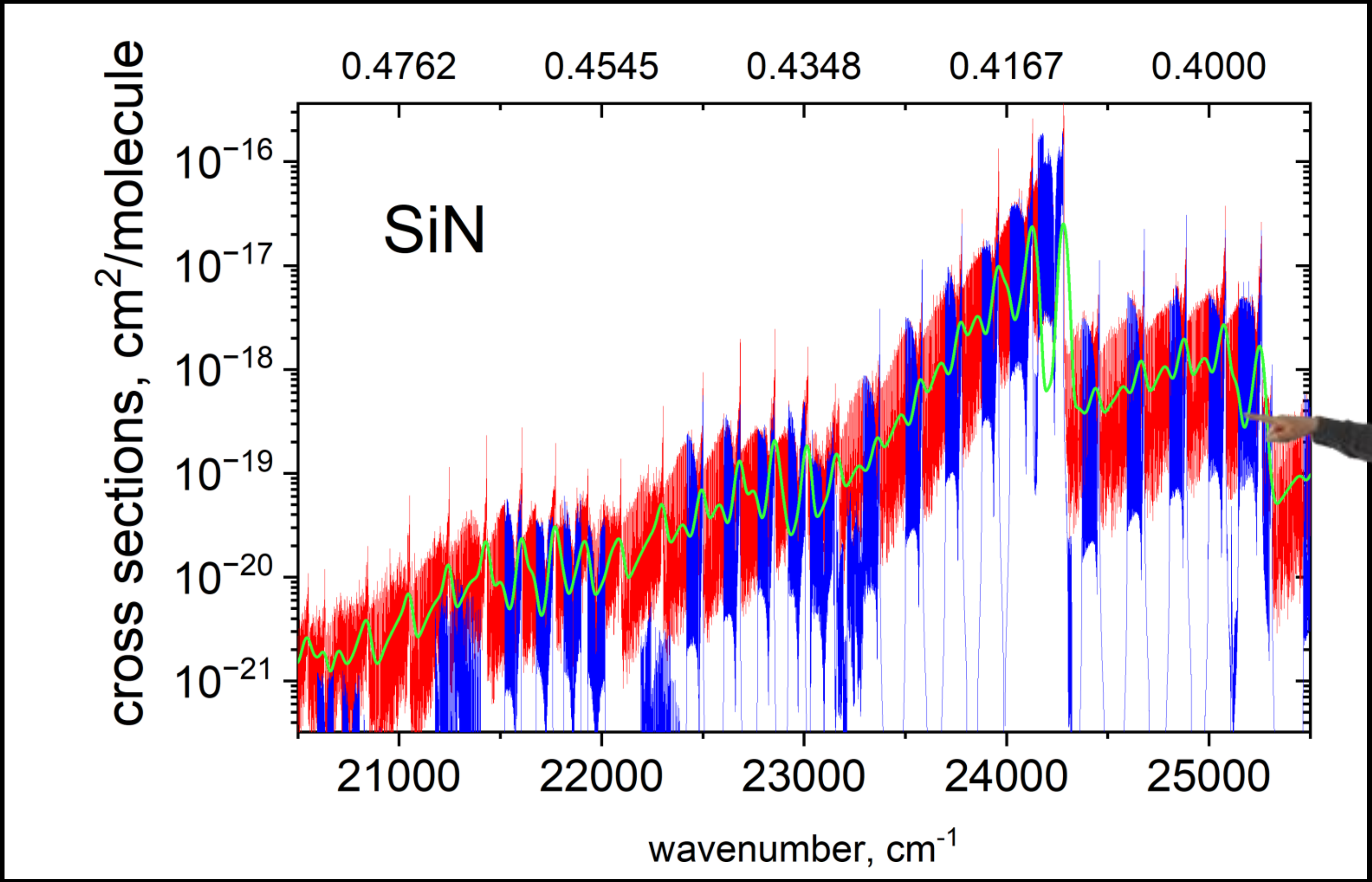
4.0×10^-23
3.5×10^-23
3.0×10^-23
2.5×10^-23
2.0×10^-23
1.5×10^-23
1.0×10^-23
5.0×10^-24
cross sections, cm²/molecule
H₂O
T = 2000 K
9816
9818
9820
9822
9824
9826
wavenumber, cm-1
Here the with the error “bars” (error bar cross sections)
Is this useful?

Results: SiN (less accurate system)

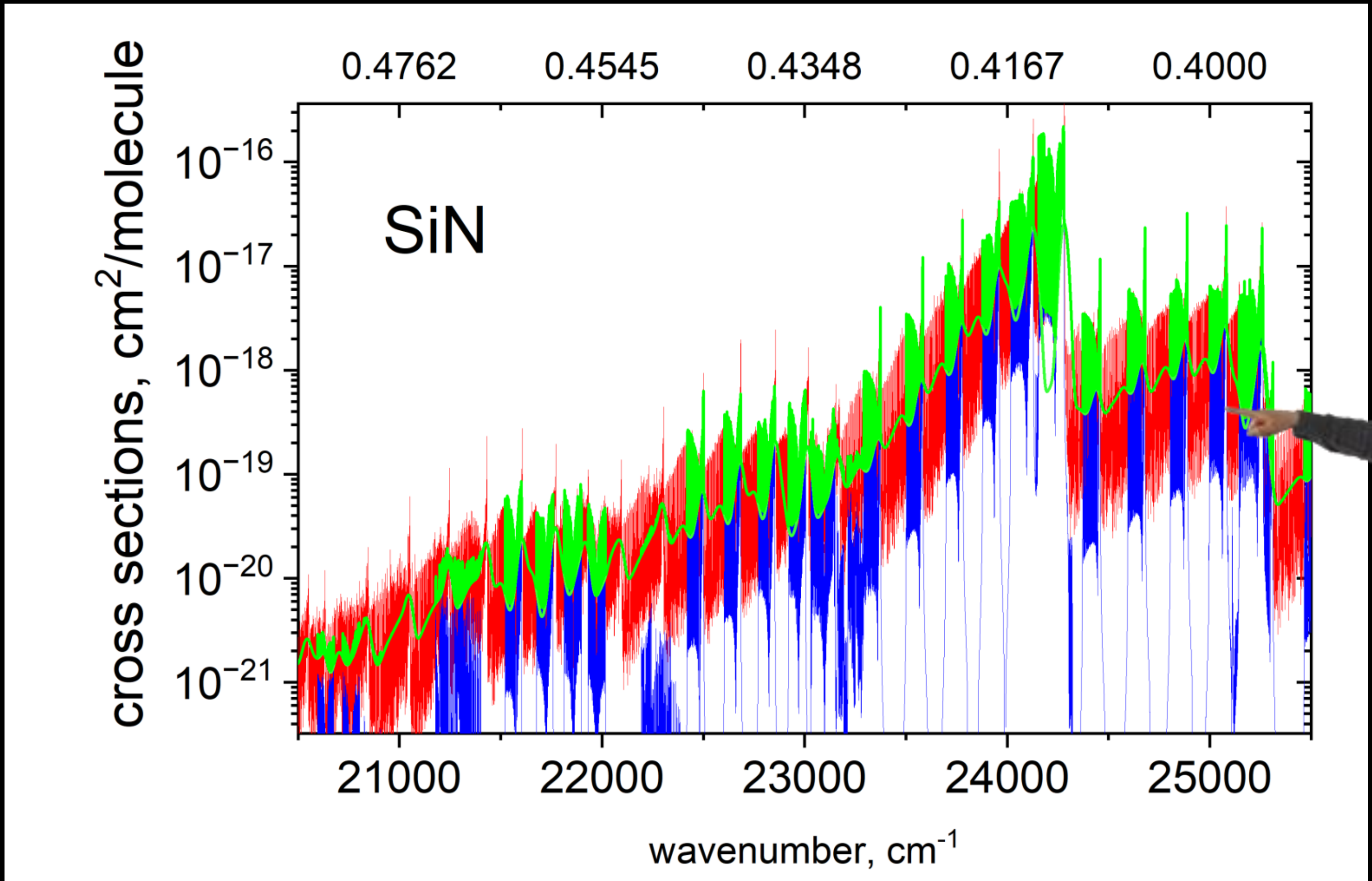
Option 1: Use MARVELised lines


SiN
0.5000
0.4545
0.4167
0.3846
0.3571
cross sections, cm²/molecule
10⁻¹⁵
10⁻¹⁶
10⁻¹⁷
10⁻¹⁸
10⁻¹⁹
10⁻²⁰
10⁻²¹
10⁻²²
10⁻²³
10⁻²⁴
10⁻²⁵
20000
22000
24000
26000
28000
wavenumber, cm⁻¹


Blue spectrum is from the MARVEL (accurate) lines only

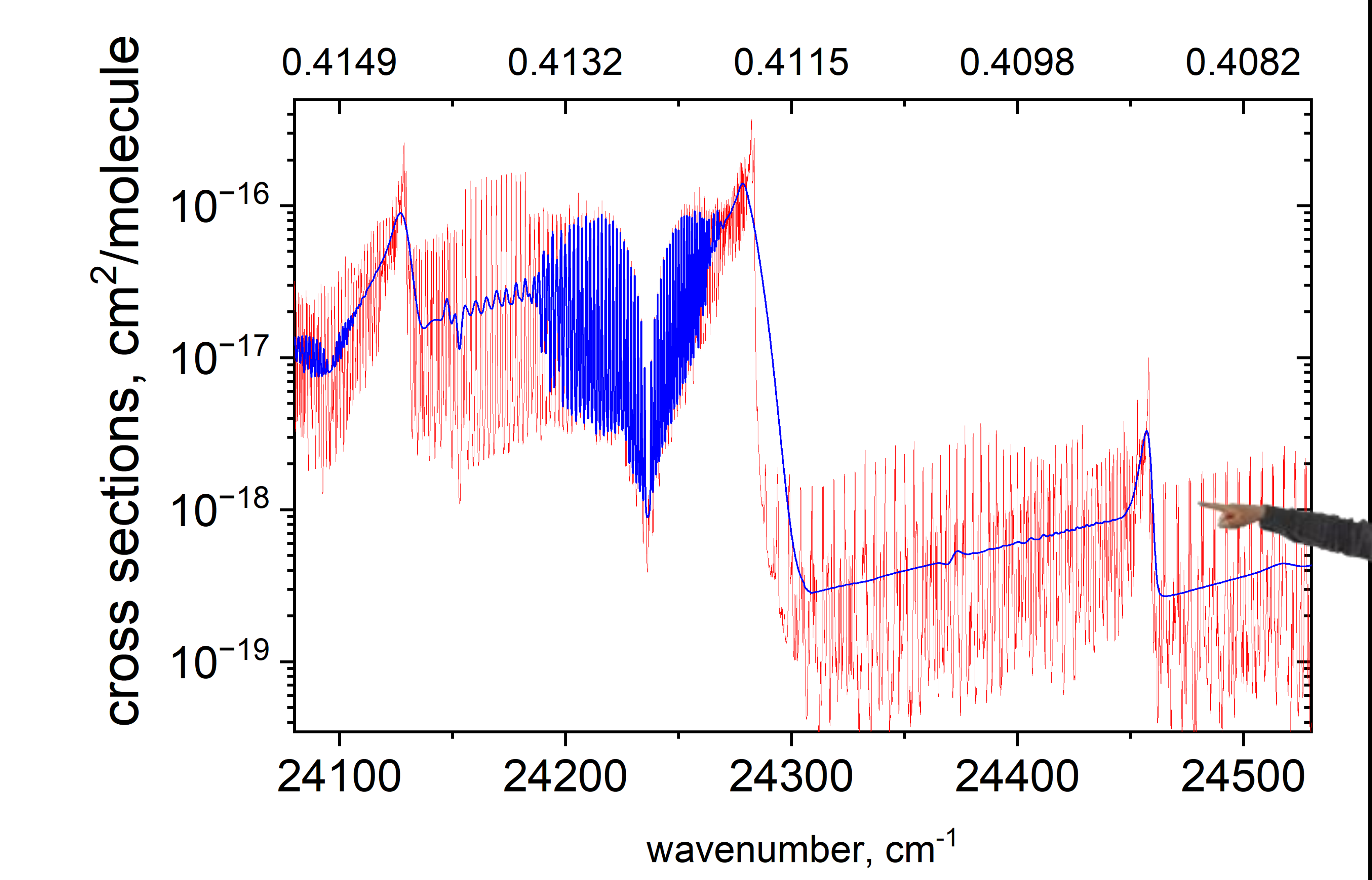

SiN
0.4762
0.4545
0.4348
0.4167
0.4000
10−16
10−17
10−18
10−19
10−20
10−21
cross sections, cm²/molecule
21000
22000
23000
24000
25000
wavenumber, cm⁻¹


Let's create the continuum from the non-accurate lines


0.4762
0.4545
0.4348
0.4167
0.4000
SiN
cross sections, cm²/molecule
10⁻¹⁶
10⁻¹⁷
10⁻¹⁸
10⁻¹⁹
10⁻²⁰
10⁻²¹
21000
22000
23000
24000
25000
wavenumber, cm⁻¹


… and combine with
the MARVELised lines
(green spectrum)

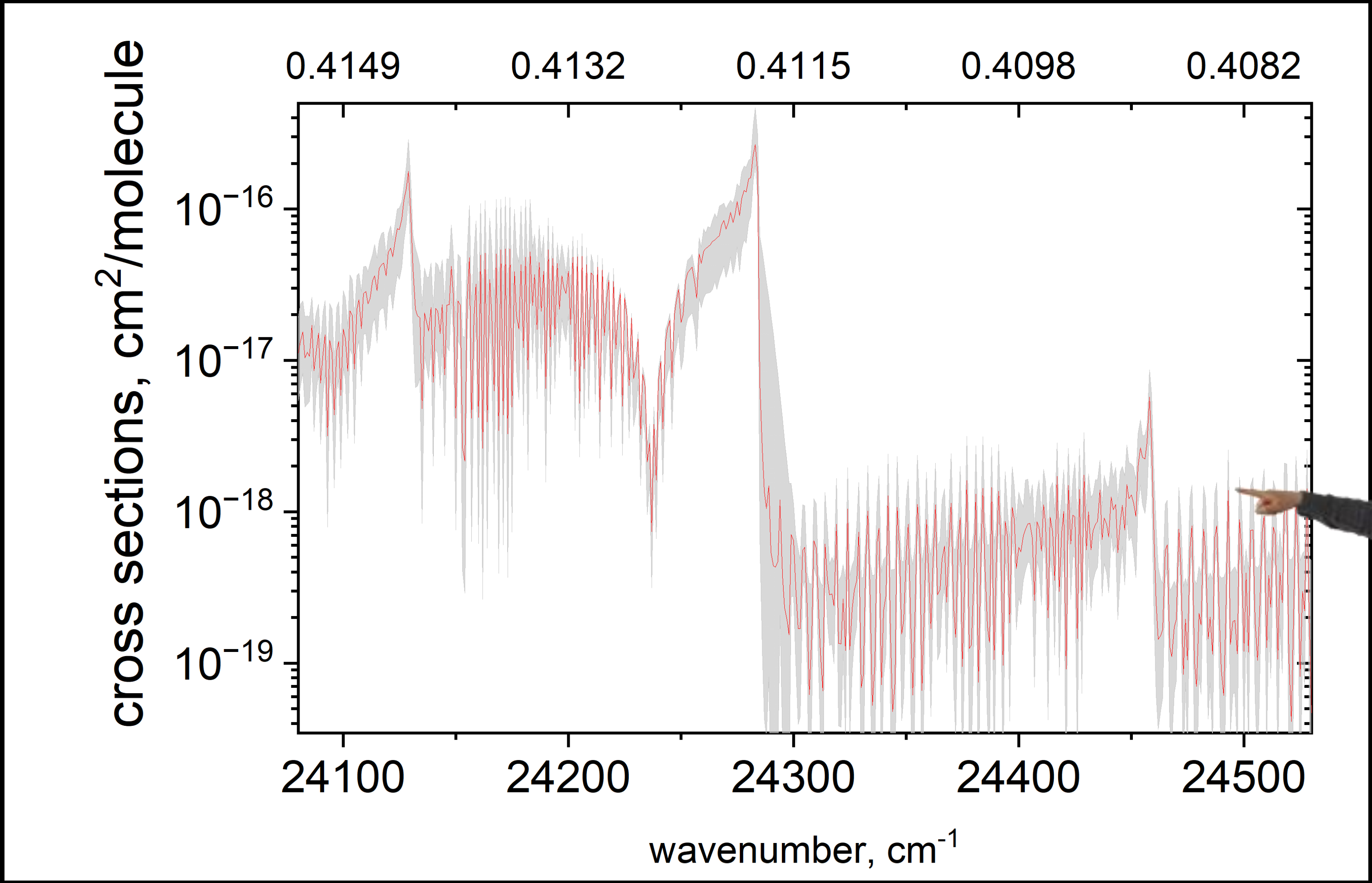
Option 2: include the uncertainty of our knowledge into absorption cross sections



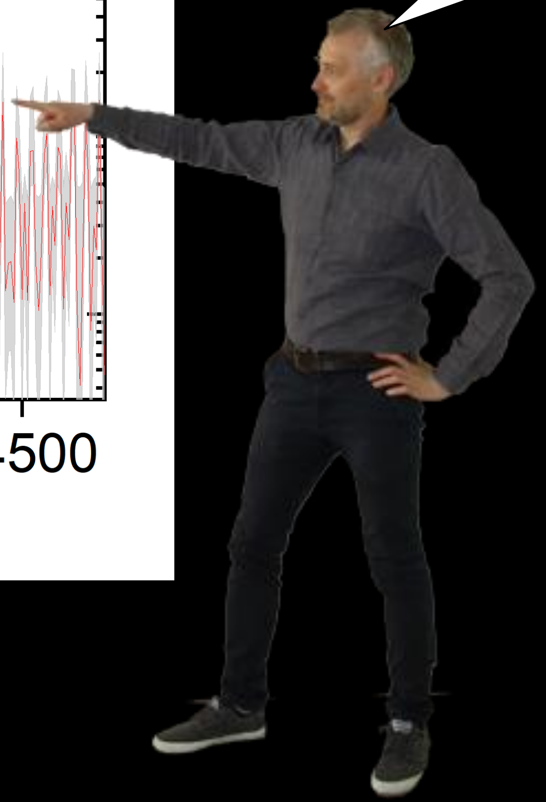
This is the blue spectrum which is the convolution of the two uncertainties compared to the standard spectrum (collisions)

How about using these blue cross sections in HR cross correlations and retrievals?

$$f_{\text{tot}}(\nu, \nu_{ij}, \sigma, \gamma) = f^{\text{G}}(\nu, \nu_{ij}, \sigma) \otimes f^{\text{L}}(\nu, \nu_{ij}, \gamma)$$



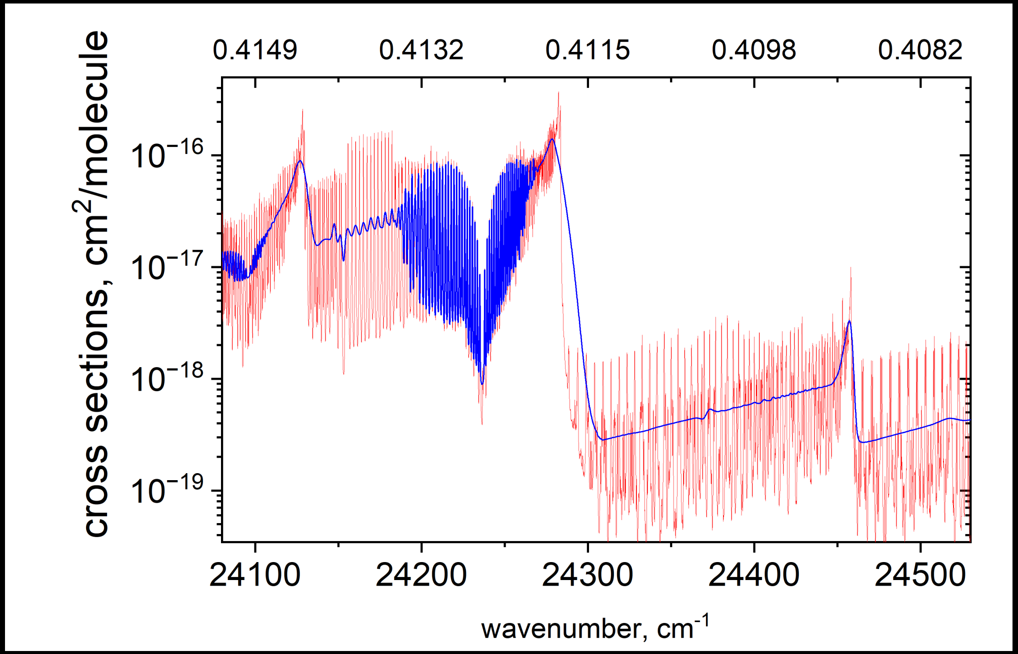
We can use the uncertainty spectrum to define the error bars as $\pm\,\Delta\sigma$



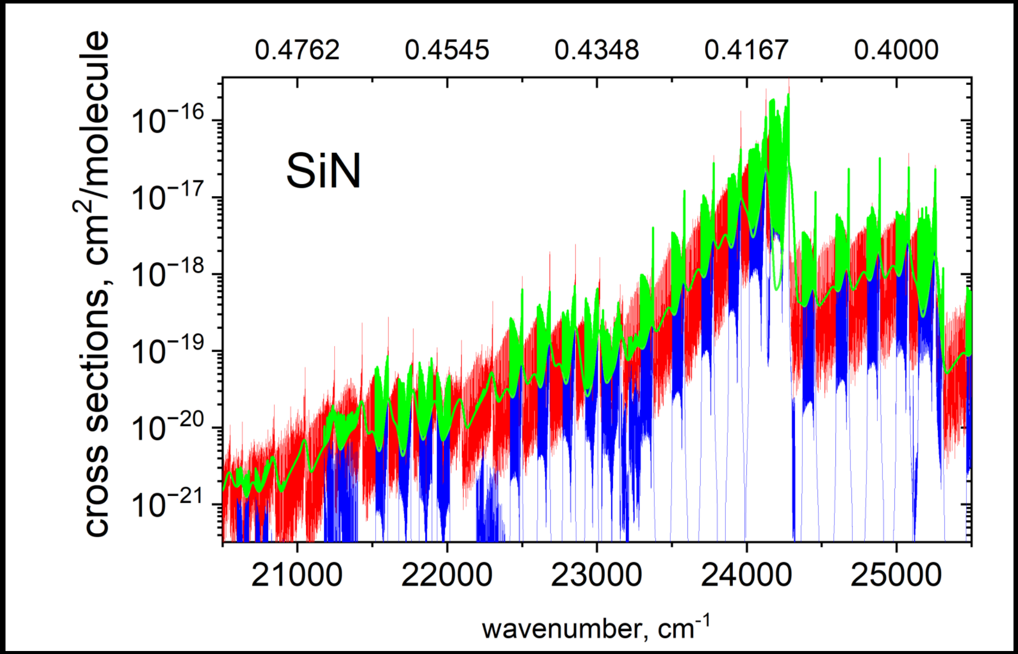
$$\sigma_{\mathrm{range}}(\nu) = \sigma \pm \,\Delta\sigma$$

Here are two versions back-to-back

From uncertainties



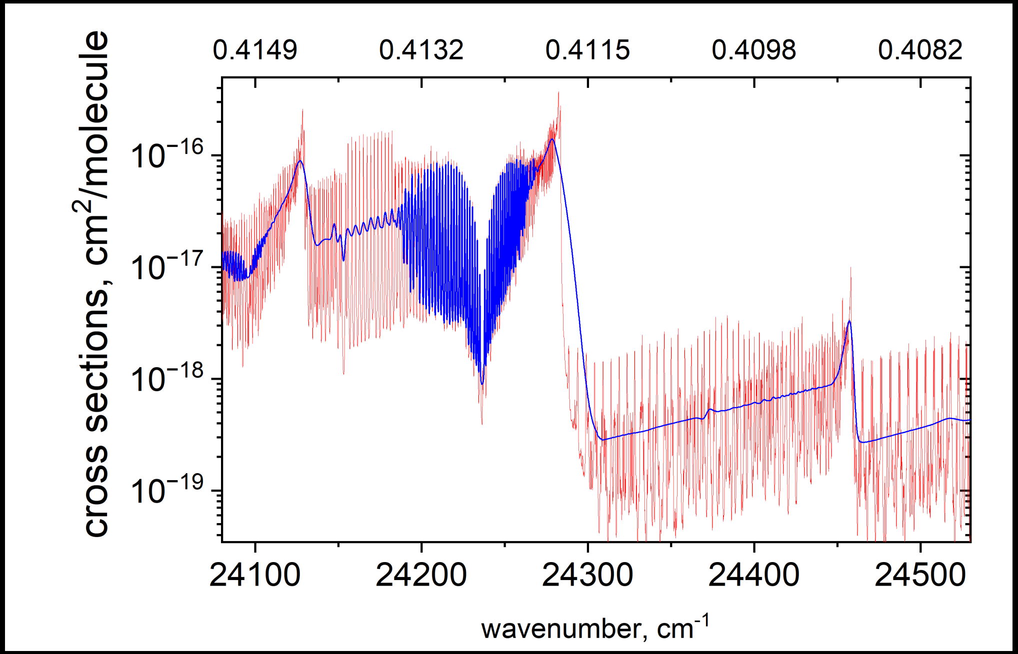
Combined with the baseline



Let me know if this does not make any sense!

Thanks!

From uncertainties



Combine