

# Analysis of spectra from the SPIDER region using a multi-Gaussian decomposition

Mukesh Taank (mtaank)

May 8, 2020

# Abstract

This report covers the project of analyzing GHIGLS data for the "Spider" region, more specifically, the HI structures located at the disk halo interface that hint at the formation of molecular gas ( $\text{H}_2$ ). Through working on this project, I have explored extracting archival HI spectral data stored in data cubes and analyzing said cube. The data is in the form of a *PPV* cube, so the first axis is the velocity axis and the others are position coordinates. I focused on the region of the cube with coordinates  $x = 65$  and  $y = 154$ . From this, I identified the main features, which included a high peak at about 0 km/s (low velocity) and a smaller peak at about  $-100$  km/s (intermediate velocity) and lots of noise, fluctuating about zero. This spectrum can be represented by a combination Gaussian function, so the next step was to fit a Gaussian function to the data. Since this spectrum has two distinct peaks, it requires more than one Gaussian to fully represent the spectrum, so I followed this multi-Gaussian decomposition of the spectrum. All of this was just for one spectrum. It is important to look at a number of spectra and average them to get a good mean spectrum. I focused on a face of the cube, about  $20 \times 20$  pixels. One key feature compared to the single spectrum is that the noise was essentially negligible with the average spectrum. Another feature was the presence of another peak, at about  $-150$  km/s, representing a high velocity region (HVC). I then fitted this average spectrum with the Gaussian model and it was a much more accurate fit.

## Generating a Spectrum from the Cube

Once I downloaded the data cube for the SPIDER field, I was able to load the cube. I chose to focus on the region with coordinates  $x = 65$  and  $y = 154$ . From this I was able to generate one spectrum, seen in figure 1.

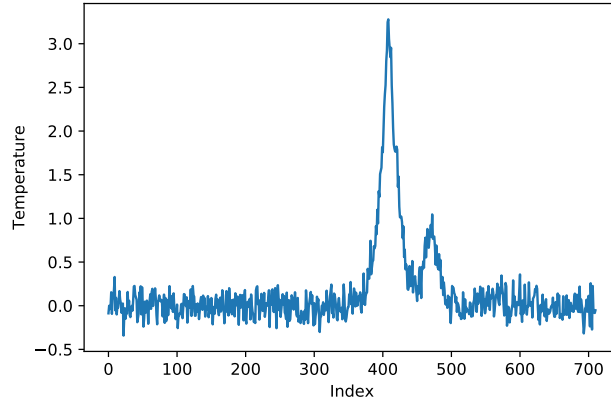


Figure 1: Plot of the temperature (K) with the index of the spectrum at position  $x = 65$  and  $y = 154$

Now that I have been able to extract a spectrum, I needed to fix the axis. Right now, the x-axis of the figure represents the index of each point in the spectrum. I need the x-axis to represent the velocity (the first axis of the cube). I used wrote a function that takes in the parameters from the header file and adjusts the velocity range so that is centers at zero (based on the reference). Now can re-graph this spectrum with the appropriate axis so that it can be properly analyzed.

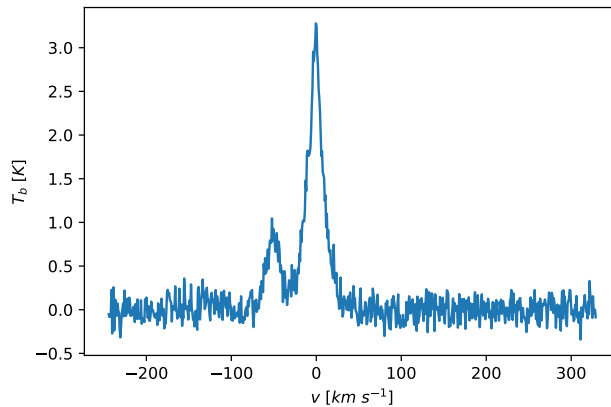


Figure 2: Plot of the temperature (K) along the velocity axis at position  $x = 65$  and  $y = 154$

From figure 2, we can see the axis has been changed to the appropriate parameter, the velocity. Now that the axis is correct, this spectrum can be properly analyzed. There are two distinct peaks present. One at about 0 km/s and one at about  $-60$  km/s. These peaks represent low velocity gas

and intermediate velocity gas respectively. The intermediate velocity gas has a negative velocity, meaning it is actually moving towards us. One other thing that I noticed in the spectrum was that on the left side, at around  $-150$  km/s, there looks to be another peak, which could represent the presence of a high velocity gas. The reason I think this is because the data does not fluctuate about zero at this point, it goes up in a wide "Gaussian shape", then turns back into the noise. One last thing I did was use ROHSA to generate the same spectrum. I called on the ROHSA function to plot the spectrum directly without having to adjust the axis. It already has the axis in units of velocity ( $\text{km s}^{-1}$ ). This step is just to show the usefulness of using ROHSA for the spectrum generation. It only took about three lines of code compared to doing it manually which requires a lot more lines.

## Fitting the Data using a Gaussian Model

I have generated a spectrum with the appropriate velocity axis, so now I can fit the data using a Gaussian model. The reason I am using a Gaussian model is because ROHSA is already equipped with computing a multi-Gaussian decomposition, so this will allow me to understand the process that goes on through ROHSA. The first step is to fit the spectrum above using only one Gaussian. Obviously, this will not be a good fit to the data, but it is a good starting point. Figure 3 shows a fit with one Gaussian function.

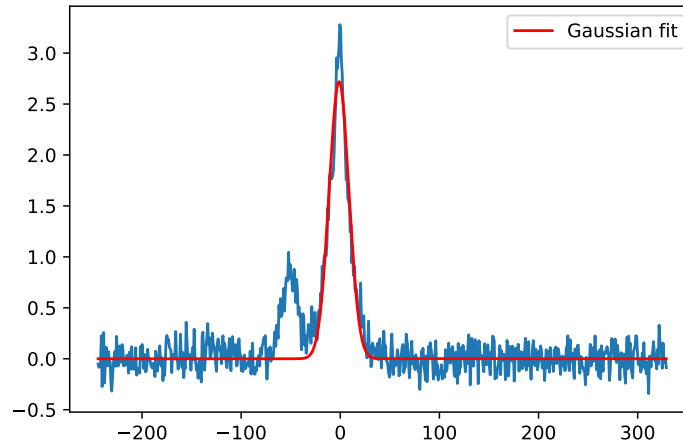


Figure 3: Plot of the spectrum with a one Gaussian fit to approximate the data.

From this plot, we see that this model completely misses the other peak. To check how well this model fits the data, I calculated the value of  $\chi^2$ , according to the formula:

$$\chi^2 = \sum_i \frac{[N_{exp} - N]^2}{\sigma^2} \quad (1)$$

where  $N_{exp}$  is the experimental data (the actual spectrum) and  $N$  is the model fitted to the data.  $\sigma$  is the standard deviation of the data. For a good fit, the value of  $\chi^2$  will be as close to the value of the number of points - number of parameters), as possible, which in this case is about 700. For this spectrum, the standard deviation is about 0.1353. For this model,  $\chi^2 = 1408$ . This is quite far from the expected 700, so this is not a great fit. As a next step, I can try plotting two Gaussian functions instead of one. I expect that this will be much better as it should account for the other peak. Figure 4 shows a fit of the combination of two Gaussian functions, as well as the two individual curves.

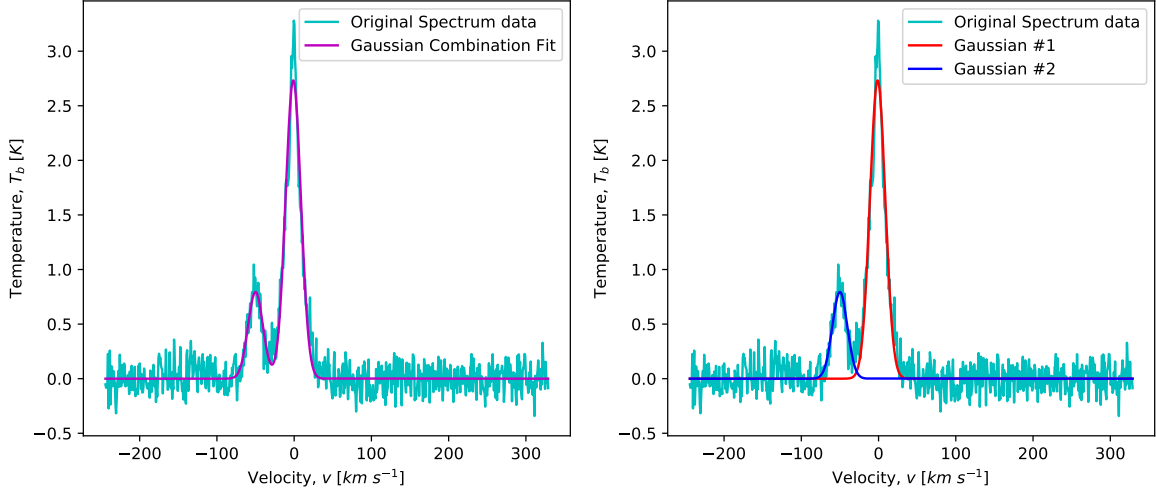


Figure 4: Plot of the spectrum with a two Gaussian fit to approximate the data The first is the combination function and the second is the individual curves.

From this, right away the fit is much better. Both peaks are accounted for, but the noise is providing errors in the fit. For this fitting, I have that  $\chi^2 = 675$ . This is very close to the expected 700, so this fit is especially good. Now that I have the fit, I can use the Gaussian functions to find the value of the full width half maximum (FWHM), which is derived from the equation of the Gaussian:

$$f(x) = Ae^{\frac{-x^2}{2\sigma^2}}$$

Note: the value of the full width half maximum is independent of the position of the center. At the FWHM,  $x = h$  and  $f(h)$  is just half the amplitude.

$$\frac{1}{2}A = Ae^{\frac{-h^2}{2\sigma^2}}$$

Cancel the amplitudes and take the natural logarithm.

$$\ln\left(\frac{1}{2}\right) = \frac{-h^2}{2\sigma^2}$$

Now just solve for  $h$ .

$$h^2 = -2\ln\left(\frac{1}{2}\right)\sigma^2$$

$$h^2 = 2\ln(2)\sigma^2$$

$$h = \sqrt{2\ln(2)}\sigma$$

Now, we have an expression for  $h$ . The FWHM is just  $2h$ , so I multiply by 2 to get the FWHM (call this value  $\Gamma$ ).

$$\Gamma = 2\sqrt{2\ln(2)}\sigma \approx 2.355\sigma \quad (2)$$

Now I have an expression for the FWHM in terms of the dispersion,  $\sigma$ , which is one of the Gaussian parameters. I used this equation to find the FWHM of both peaks. Using equation (2), I was able to find the FWHM values. The FWHM of the highest peak (LVC) is about 28.9359 km/s. The FWHM of the lower peak (IVC) is about 28.2479 km/s. From this, I can use the kinetic theory of gases to convert this to temperature units, to find out how hot these gases are.

$$\frac{1}{2}mv^2 = \frac{f}{2}k_B T$$

where  $v$  is just  $\Gamma$ . Since we are only considering the  $z$  component of the velocity,  $f = 1$ . So an expression for the FWHM in temperature units is:

$$\Gamma = \sqrt{\frac{k_B T}{m_H}} \approx 2.355\sigma \quad (3)$$

Using equation (3), I converted the FWHM to Kelvin. The FWHM for the LVC expressed in units of temperature is: 101491 K. The FWHM for the IVC expressed in units of temperature is: 96723 K.

So far, all of this has been for a single spectrum. It is important to look at an average of many spectrum to get a good indication of the rest of the cube. I took a small face from the cube (20 x 20 pixels) around the same region of  $x$  and  $y$ . Figure 5 shows the average spectrum compared to the original single spectrum.

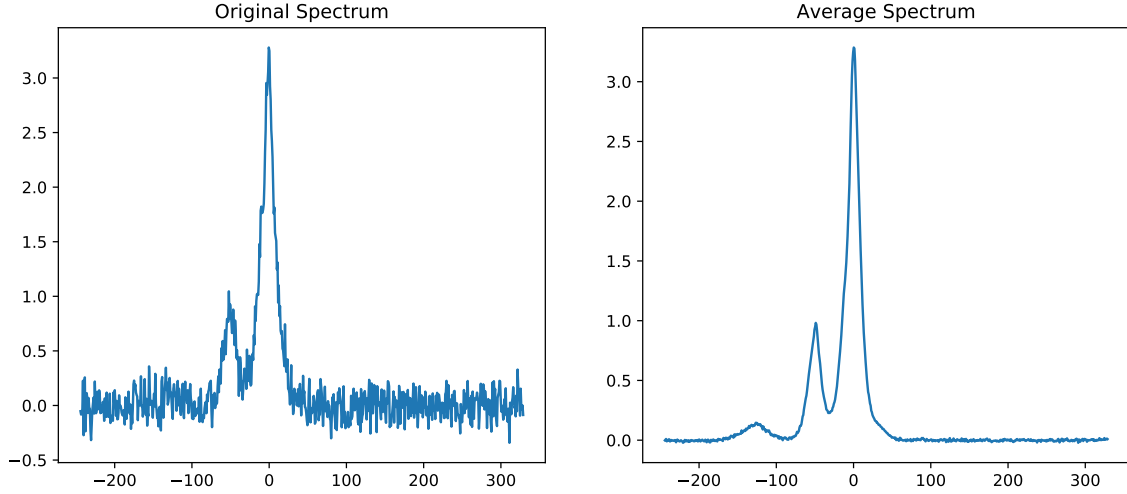


Figure 5: Plot of the single spectrum compared to the average spectrum.

Now that we have the average spectrum, we can directly compare it to the single spectrum. The first thing is that there is very little noise throughout. From the single spectrum, the standard deviation of the noise in the end channels is about 0.1353. Since the latter is an average of 400 spectra, the standard deviation should have gone down by a factor of  $\sqrt{400}$  (20). The standard deviation of the average spectrum end channels is about 0.006471, which is 20 times less. Also in the spectrum, we can still easily identify the LVC and IVC from the peaks, however, now that the noise has been minimized, there is a distinct peak appearing at about -120 km/s. This can represent the presence of a high velocity gas (HVC). In the first spectrum, we observed that there could possibly be an HVC peak, but the noise was so strong that it was very difficult. In the latter, it is very clear that this is not just noise, but a distinct peak. From this, I can again use the Gaussian model to fit the spectrum, but now, I will be able to account for the HVC peak.

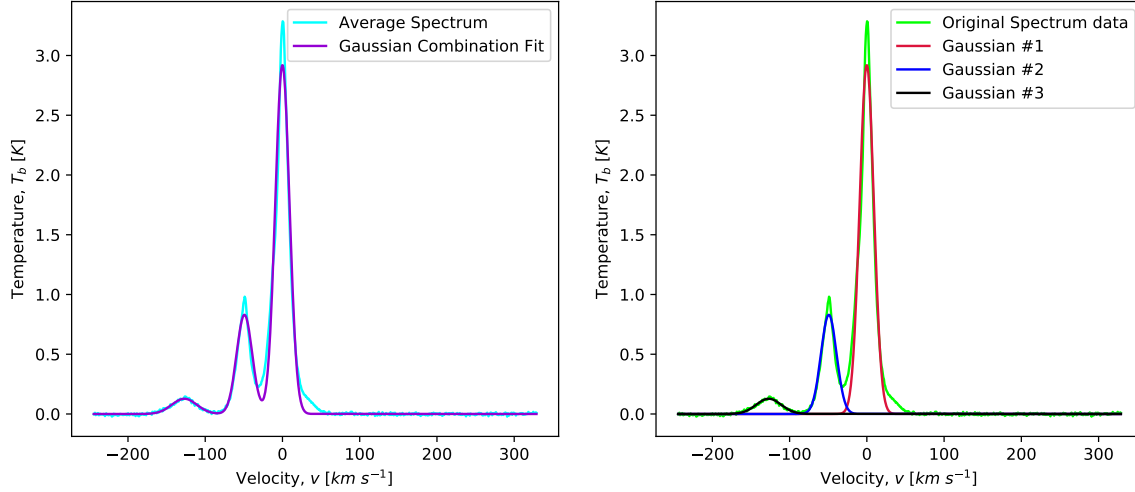


Figure 6: Plot of the average spectrum with a multi-Gaussian fit to approximate the data. The first is the combination function and the second is the individual curves.

From this plot, we can see that the multi-Gaussian fit is very accurate. Since there is a lack of noise, the fit is much better towards the ends. I am able to find the FWHM for the three peaks to get some more information about the gases. The FWHM of the LVC is about  $26.9033 \text{ km/s} \approx 87734 \text{ K}$ . The FWHM of the IVC is about  $29.4938 \text{ km/s} \approx 105442 \text{ K}$ . The FWHM of the HVC is about  $47.3698 \text{ km/s} \approx 271994 \text{ K}$ . From these numbers, we see that the FWHM's are very consistent with what was calculated earlier. Now, however, we can learn some information about the HVC. moving about twice as fast and it is almost 2.5 times hotter than the IVC. Overall, the average spectrum gives us lots of information about the gas in more than just one spot in the cube. It allows us to understand how to gases are changing in different areas of the sky, and that is very valuable.

## Conclusion

Throughout this project, I learned many things about how to analyze and understand information from a data cube. I was able to load the GHIGLS data and use the cube to create a spectrum. From that I was able to plot it against the velocity axis to be properly examined. After that I used a Gaussian fitting model to represent the spectrum. We can check the accuracy of the model by finding  $\chi^2$ . Ideally, it would be 700 (from the numbers of parameters minus the number of variables). For the one Gaussian fit,  $\chi^2 \approx 1400$ . For a two Gaussian fit, the  $\chi^2 = 675$ . This leads to the thinking that using more Gaussian functions would produce an even better fit. I then produced an average spectrum, which is the mean of 400 spectra (region of  $20 \times 20$  pixels). The main feature for the average spectrum is lack of noise. When comparing the standard deviation of the spectrum to the average spectrum, the values are 0.13532 and 0.00647 (less by a factor of  $20 (\sqrt{400})$ ). Also



from the average spectrum, one main thing that is noticeable is a small peak at about  $-150$  km/s. This shows the presence of an HVC. Overall, this project has been very informative in showing me the process of analyzing a data cube. Once I complete all this work manually, I was able to use ROHSA to do most of this in many less steps. The ROHSA allows us to fit multiple Gaussian functions as accurately as possible.