DRAFT VERSION APRIL 12, 2023
Typeset using LaTeX **twocolumn** style in AASTeX631

Comparing Properties of Giant Molecular Clouds Observed in CO(1–0) and CO(2–1) emission lines

DHANANJHAY BANSAL AND ERIK ROSOLOWSKY[1]

[1]*Department of Physics, University of Alberta, CCIS 4-181, Edmonton, AB T6G 2E1, Canada*

ABSTRACT

We investigate the properties of giant molecular clouds (GMCs) in the molecular rich galaxy M83 (NGC 5236) in two different transitions of carbon monoxide, CO(1–0) and CO(2–1). We use the PYCPROPS algorithm to catalog molecular clouds and find 1508 molecular clouds in the CO(2–1) emission and 1674 clouds in the CO(1–0) emission. When properties like luminosity, radius, luminous mass and velocity dispersion of the same set of these extracted clouds are compared between the two emissions, they are found to deviate less than 25 % from the locus of equality. Possible reasons for deviation are inherent strong CO(1–0) emission from the clouds present in the nucleus of a galaxy and non-convergence of the slope while linear fitting data using the RANSAC algorithm. In order to improve the result, we propose including more galaxies in this cloud based analysis and running multiple iterations of the RANSAC linear regression algorithm.

Keywords: galaxies: individual (M83) — galaxies: ISM — ISM: clouds — ISM: structure — Molecular clouds — Giant molecular clouds

1. INTRODUCTION

Giant molecular clouds (GMCs) constitute the densest regions of the interstellar medium (ISM) and are representatives of sites of stellar nurseries in galaxies. Their properties determine the initial conditions for the star formation process and their life-cycle determines how stellar feedback regulates the overall evolution of a galaxy (Bolatto et al. 2008). Resolving these extragalactic entities does not only increase our understanding of their properties and distribution, but also, provides insight into how they influence the stellar initial mass function (McKee & Ostriker 2007). Despite variation in their population from galaxy to galaxy, GMCs exhibit a surprisingly uniform set of properties (Bolatto et al. 2008). When individual GMCs are resolved, their properties like luminous mass, velocity dispersion, radius and eccetera, are observed using the low rotational transitions of carbon monoxide (CO).

Observations of rotational line emission from carbon monoxide (CO) has proven to be a robust method for tracing the distribution, kinematics, and physical conditions in the molecular interstellar medium (ISM) in external galaxies (Bolatto et al. 2013). We don't use molecular hydrogen to trace GMCs even though it is the most abundant molecule in the ISM because it's difficult to excite its rotational emission lines at the temperatures found in most GMCs (Peñaloza et al. 2018).

Intensities of CO emission lines are critical for understanding the star formation process and the complicated morphologies of the ISM in our galaxy and beyond. CO emissions from galaxies are generally interpreted using ratios of the CO rotational line emissions,i.e., $R_{2-1/1-0}$, $R_{3-2/2-1}$, eccetera. This is because line ratios are expected to anti-correlate with the CO (1–0)-to-H_2 conversion factor (Leroy et al. 2022) and they serve as a great set of tools for investigating local physical conditions,i.e., density and temperature, in the molecular ISM (Peñaloza et al. 2018).

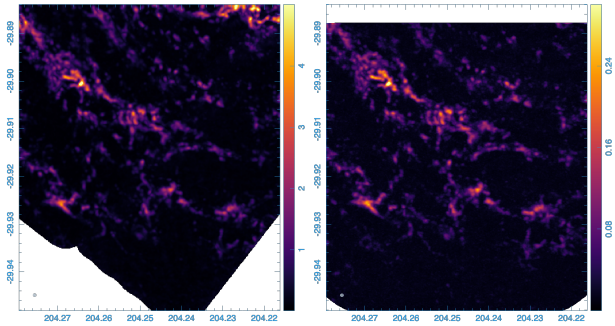

Figure 1. (A) Panel on the left depicts trace of Giant Molecular clouds in the CO(1–0) transition. (B) Panel on the right depicts the same but using the CO(2–1) transition.

For years, a particular transition of carbon monoxide i.e., CO(1–0), has been extensively used to study the Giant Molecular Clouds found in the ISM (Donovan Meyer et al. 2013), but recently, higher CO transitions, i.e., CO(2–1) & CO(3–2), are gaining popularity due to their high brightness. High brightness emissions are preferable because they lead to lower investment in a telescope's observing time and for the same value of resolution and sensitivity, telescopes can

dbansal@ualberta.ca

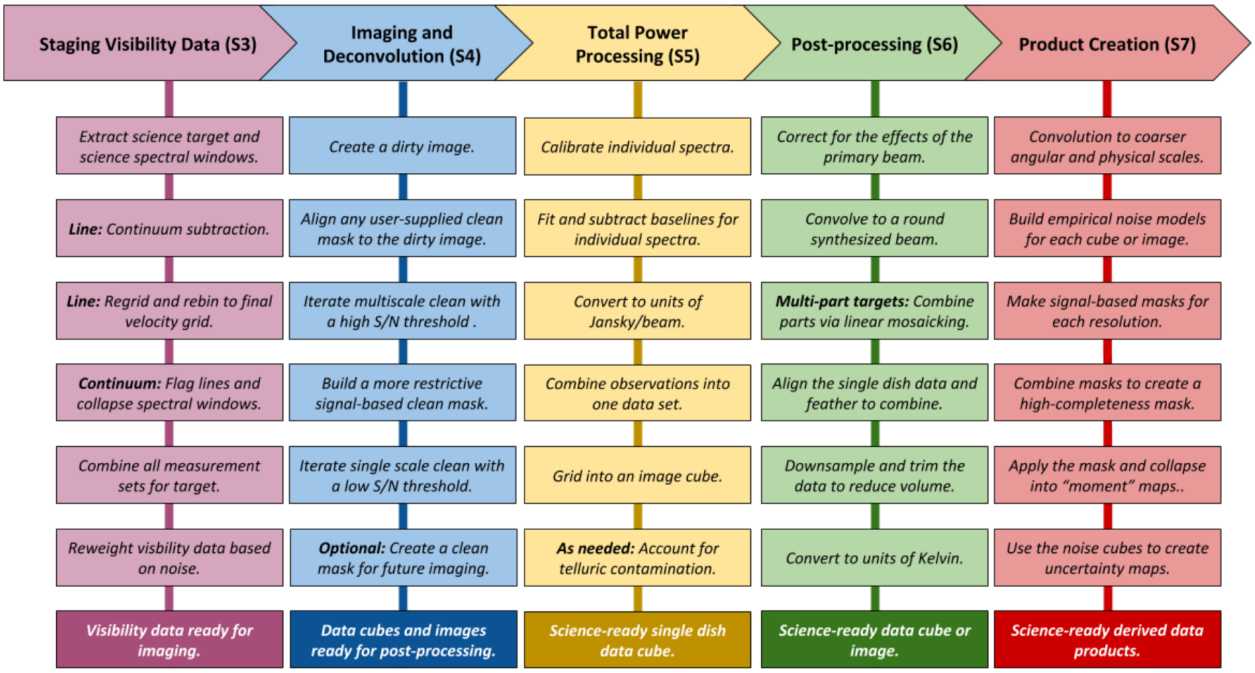

Figure 2. The PHANGS pipeline consists of 5 steps: Step 1. **Staging the visibility stage** - This step is responsible for extracting science targets and preparing u-v data for imaging. Step 2. **Imaging and Deconvolution** - After the visibility data is ready for imaging, this steps prepares data cubes for post-processing. Step 3. **Total Power Processing** - This step happens in parallel to step 2 and takes care converting the units of u-v data to Jansky/beam. Step 4. **Post-processing** - This is the second last step and is responsible for the production of spectral data cube with each point representing surface brightness temperature T(x,y,v). Step 5. **Product Creation** - Building of science ready derived data products corresponding to 'moment' maps is handled by this step (Reproduced from (Leroy et al. 2022)).

map higher CO transitions significantly faster than the basic transition of CO(1–0) (Leroy et al. 2022). Despite the advantages of mapping GMC with higher transitions of CO, many surveys still map molecular clouds using CO(1–0) emission lines (Saintonge et al. 2017) because CO(1–0) is the fundamental CO transition and higher transition emission lines from GMCs are harder to excite.

In this paper, we have compared the properties of giant molecular clouds in M83 observed in the CO(1–0) and CO(2–1) emissions; traces of these GMCs in both the transitions, CO(1–0) and CO(2–1), can be seen Figure 1. M83 is a good target for such a study because of its proximity to the Milky Way galaxy, only 4.5 Mpc away, and its disk orientation with an inclination of 24 degrees. Moreover, it has already been extensively studied in the CO(1–0) transition via a cloud-based analysis (Freeman et al. 2017); the molecular clouds in M83 were found to be well-resolved in ALMA data and showed excellent correspondence with the scaling relation seen in other systems. In Sect. 2, we briefly describe the data reduction measures adopted to undertake our study. In Sect. 3, we present our results on the comparison between the properties of giant molecular clouds in the two emissions. In Sect. 4, we discuss the implication of these results and draw conclusions based on what we know from the previous studies. Finally, in Sect. 5, we present the closing remarks and conclude with the possible extensions on improving and corroborating this study further.

2. DATA REDUCTION

Here, we study the properties of giant molecular clouds in the galaxy M83, which is one of the 90 galaxies traced and mapped by the PHANGS-ALMA collaboration. PHANGS-ALMA is an ALMA survey which observes the CO(2–1) emission lines from nearby galaxies. ALMA has two arrays: a main 12-m array and the Morita Atacama Compact Array (ACA), which consists of the 7-m array and the total power antennas.

In the following paragraphs, we lay down a brief summary of how the data collected by the ALMA survey is reduced from being uncalibrated visibility data to cloud catalog assignments in both the CO(1–0) and CO(2–1) transitions. The first paragraph explains the working of the ALMA pipeline (Leroy et al. 2021) very briefly and lays the foundational steps to the beginning of the data reduction procedure. The second paragraph walks over how the PHANGS-ALMA pipeline transforms the output from the ALMA pipeline to a cleaned version of the data (Leroy et al. 2021). Finally, the third paragraph succinctly talks about the CPROPS algorithm (Rosolowsky et al. 2021) and concludes this section.

Both of the arrays of the ALMA telescope operate independently such that there is separate interferometric visibility

```
39
40      s = SpectralCube.read(datadir + '/' + cube_file)
41      mask = fits.getdata(datadir + '/' + mask_file)
42
43      if allow_huge:
44        s.allow_huge_operations = True
45
46      if noise_file is None:
47          print("No noise file found.  Calculating noise from Med. Abs. Dev of Data")
48          noise = s.mad_std().value
49          if delta is None:
50              delta = 2 * float(noise)
51      else:
52          noise = SpectralCube.read(datadir + '/' +noise_file)
53          if delta is None:
54              delta = 2 * noise.median().value
55
56      distance_Mpc = distance.to(u.Mpc).value
57
58      # Cast to boolean
59      nanmask = np.isnan(mask)
60      mask = mask.astype(np.bool)
61      mask[nanmask] = False
62
63      s = s.with_mask(mask)
```

Figure 3. Snippet of the working behind the PYCPROPS algorithm. Spectral cubes are transformed to cloud catalogs which are then used to study the properties of clouds (Reproduced from (Rosolowsky et al. 2021)).

data, or "u-v data", available for each ALMA target. However, this u-v data observed by each array is uncalibrated and needs to be processed to produce calibrated data with well understood properties and uncertainties. ALMA observatory provides calibrated version of the data by running it through the ALMA pipeline, which reduces the data by using the standard CASA tasks. Nonetheless, the ALMA pipeline does not provide derived data products beyond data cubes and images, which leaves the researchers the task of further processing the visibility and total power data into science ready products.

In order to yield science-ready data cubes, the post-calibrated visibility data from the ALMA pipeline is run through the PHANGS pipeline which consists of automated methods to go from calibrated u-v data to data science ready products. In basic terms, the pipeline begins by imaging and deconvolving the output from the ALMA pipeline. Imaging is done because the u-v data samples the Fourier Transform of the emission from sky at every single frequency and it needs to be grided and Fourier transformed again at each frequency to produce data cubes. Deconvolution is required because the u-v plane is sampled incompletely by the ALMA interferometers and it is necessary to reassemble the true intensity distribution from the incomplete u-v data in order to produce accurate images of the sky. Imaging and deconvolving results in a reconstructed image, also known as a 'clean' image. In parallel, the total power data is reduced "via the calibration and imaging pipeline presented by (Herrera et al. 2020)". After imaging and deconvolution, the interferometric data is amalgamated with the single dish data to correct for the "interferometer's lack of sensitivity to extended emission". A more detailed step wise workflow from the beginning to end is shown in in Figure 2.

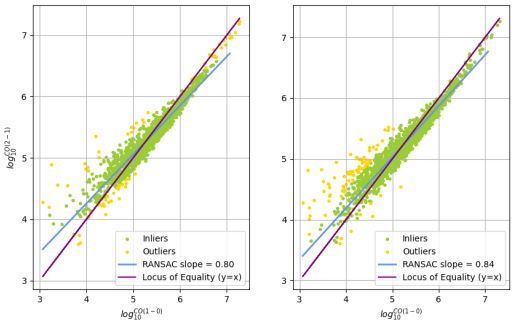

Figure 4. (A) Panel on the left compares the luminosity of clouds extracted in the CO(2–1) emission, i.e., 1508 clouds, and the same set of clouds extracted in the CO(1–0) emission using the CO(2–1) assignment. (B) Panel on the right compares the luminosity of clouds in the CO(2–1) emission extracted using the CO(1–0) assignment and the same set of clouds in the CO(1–0) emission, i.e., 1674 clouds.

The spectral line data cubes produced by the PHANGS pipeline are then run through the CPROPS algorithm which reduces them to catalogued molecular clouds. The python version of the CPROPS algorithm, PYCPROPS, achieves this GMC identification and characterization. In basic terms, this algorithm first identifies significant emission in the spectral data cube and assigns each pixel containing significant emission to a specific molecular cloud. Then, it proceeds onto characterizing the emission associated with each cloud to evaluate that cloud's properties, i.e., luminosity, velocity dispersion and, eccetera. A code snippet showing a taste of how the PYCPROPS algorithm works is shown in Figure 3.

The output from CPROPS marks the end of the data reduction procedure; we then step up and perform statistical analysis on the cloud catalogs to understand the co-relation between the properties of giant molecular clouds observed in the CO(1–0) emission line and the CO(2–1) emission line.

3. ANALYSIS

As briefly mentioned in section 2, we identify molecular clouds in the CO(1–0) emission line and the CO(2–1) emission line data using the CPROPS algorithm (Rosolowsky & Leroy 2006), and once the cloud catalogs with the relevant pixel assignments corresponding to both the rotational transitions are available, we generate a new cloud catalogue for CO(1–0) using the CO(2–1) assignment and a new cloud catalogue for CO(2–1) using the CO(1–0) assignment. This is done to make sure that the same set of clouds, i.e., 1508 clouds in the CO(2–1) emission and 1674 clouds in the CO(1–0) emissions, are compared between the two emissions.

A luminosity comparison between the clouds extracted in the CO(2–1) emission, i.e., 1508 clouds, and the same set of clouds in the CO(1–0) emission is displayed in the subplot A of Figure 4. In subplot B, we can see another luminosity comparison, but unlike before, this comparison is

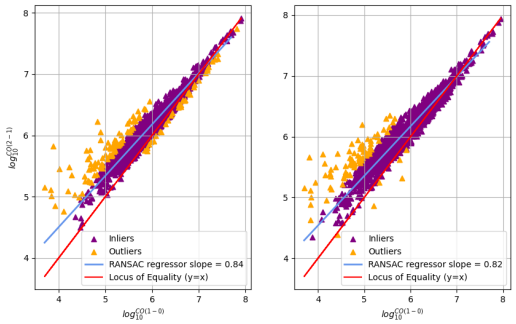

Figure 5. (A) Panel on the left compares the luminous mass of clouds extracted in the CO(2–1) emission, i.e., 1508 clouds, and the same set of clouds extracted in the CO(1–0) emission using the CO(2–1) assignment. (B) Panel on the right compares the luminous mass of clouds in the CO(2–1) emission extracted using the CO(1–0) assignment and the same set of clouds in the CO(1–0) emission, i.e., 1674 clouds.

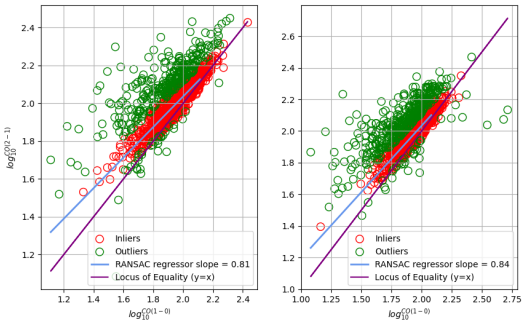

Figure 6. A) Panel on the left compares the radius of clouds extracted in the CO(2–1) emission, i.e., 1508 clouds, and the same set of clouds extracted in the CO(1–0) emission using the CO(2–1) assignment. (B) Panel on the right compares the radius of clouds in the CO(2–1) emission extracted using the CO(1–0) assignment and the same set of clouds in the CO(1–0) emission, i.e., 1674 clouds.

between the clouds extracted in the CO(2–1) emission using the CO(1–0) assignment, i.e., 1674 clouds, and the same set of clouds in the CO(1–0) emission. We use a machine learning algorithm, RANSAC regression, from the 'sci-kit' library from the python software package to achieve more accurate fitting results. Unlike ordinary linear regression algorithms like 'curve_fit' from the 'scipy' python library, RANSAC algorithm robustly fits faulty data and is insensitive to outliers. In basic terms, RANSAC regressor splits the data into outliers and inliers and only subjects the inlier data points to linear regression; this gives an accurate representation of the underlying relationship between the data points.

RANSAC coefficient, or slope, in subplots A & B of Figure 4 is found to be 0.80 and 0.84, respectively. Along with the RANSAC slope, Figure 4 also displays a locus of equality. Similar analysis is carried for the luminous mass,radius and velocity dispersion of the extracted clouds in the CO(2–1) and CO(1–0) transitions, and is shown in Figure 5, Figure 6 & Figure 7, respectively. The RANSAC slopes in the subplots A and B of Figure 5, Figure 6 & Figure 7 are found to be 0.81 & 0.84, 0.84 & 0.82, and 0.91 & 0.78, respectively. Table 1 numerically summarizes the distribution of the ratios of the properties that have been analyzed in Figure 4, Figure 5, Figure 6 and Figure 7.

4. DISCUSSION

The RANSAC slopes in subplots A and B deviate from the locus of equality by 20 % and 16 %, respectively. The same analysis concerning luminous mass, radius and velocity dispersion of the extracted clouds is noted when looking at Figure 5, Figure 6 and & Figure 7. In subplots A and B of Figure 5, the RANSAC slopes deviate from the locus of equality by 16 % and 18 % , in Figure 6 they deviate by 19 % and 16 %, and finally, in Figure 7, they deviate by 9 % and 22 %.

Table 1. Numerical summary of the ratio of properties of the extracted clouds in the CO(2–1) emission and CO(1–0) emission. Here, a label of 'Assigned' means extraction of clouds using the other emission assignment.

$\log_{10}$(Ratio)	Median	16^{th} %ile	84^{th} %ile
$\frac{L_{CO(2--1)}}{L_{CO(1--0),Assigned}}$	-0.052704	-0.152379	0.132561
$\frac{L_{CO(2--1)}}{L_{CO(1--0),Assigned}}$	-0.052704	-0.152379	0.132561
$\frac{R_{CO(2--1)}}{R_{CO(1--0),Assigned}}$	0.052595	0.003687	0.165898
$\frac{R_{CO(2--1),Assigned}}{R_{CO(1--0)}}$	0.054195	0.004955	0.178985
$\frac{LM_{CO(2--1)}}{LM_{CO(1--0),Assigned}}$	0.162910	0.042610	0.360549
$\frac{LM_{CO(2--1)}}{LM_{CO(1--0),Assigned}}$	0.162910	0.042610	0.360549
$\frac{VD_{CO(2--1)}}{VD_{CO(1--0),Assigned}}$	0.061130	0.005540	0.204777
$\frac{VD_{CO(2--1),Assigned}}{VD_{CO(1--0)}}$	0.065929	0.011449	0.198613

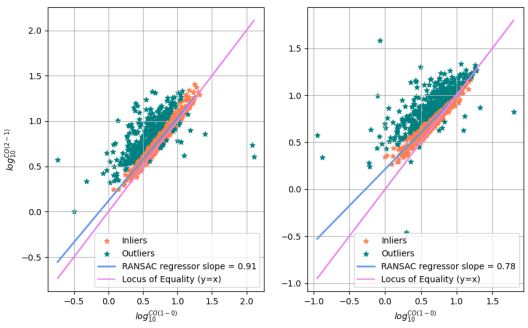

Figure 7. (A) Panel on the left compares the velocity dispersion of clouds extracted in the CO(2–1) emission, i.e., 1508 clouds, and the same set of clouds extracted in the CO(1–0) emission using the CO(2–1) assignment. (B) Panel on the right compares the velocity dispersion of clouds in the CO(2–1) emission extracted using the CO(1–0) assignment and the same set of clouds in the CO(1–0) emission, i.e., 1674 clouds.

The RANSAC slope fails short of aligning itself parallel to the locus of equality irrespective of the property being compared; we believe the reason is two fold. First, giant molecular clouds present around the nucleus of a galaxy strongly emit only in CO(1–0) emission (Nakai et al. 1986), which means there will always be inherent difference in the properties of the molecular clouds observed in the CO(1–0) and CO(2–1) emission. Second, we notice that the RANSAC algorithm coefficient, or slope, doesn't converge to a single value. The coefficient changes every time the algorithm is used, sometimes reporting less deviation from the locus of equality and sometimes more. We believe subjecting the RANSAC results to a statistical analysis where the algorithm is used more than a single time will help us estimate the corresponding slopes better. Also, in this study we couldn't determine the errors associated with the reported RANSAC slopes; it will be more helpful for discussing the implications of the results if we use the uncertainties associated with the slopes, rather than their percentage difference from the locus of equality.

The RANSAC slopes corresponding to each analyzed cloud property seem to follow the locus of equality with less than 25 % deviation, and with consideration of sources of deviation mentioned above, there is a good chance that the deviation can be reduced. For more robust results, we can extend our study to a few more galaxies from the group of 90 galaxies surveyed by the ALMA-PHANGS collaboration, however, we expect to see similar results. Considering the advantages of mapping galaxies in CO(2–1) over CO(1–0) emission, as mentioned in Sect. 1, along with the reported results, we believe it is justifiable to map a galaxy in just CO(2–1) emission and eliminate the need to map the same galaxy in CO(1–0) emission.

5. CONCLUSIONS

We present a study where we compare luminosity, radius, luminous mass and velocity dispersion of the same set of clouds in the CO(1–0) and CO(2–1) emission. Explicitly, this means comparing clouds in the CO(2–1) cloud catalog to the clouds in the CO(1–0) cloud catalog extracted using the CO(2–1) assignment, and the clouds in the CO(1–0) cloud catalog to the clouds in the CO(2–1) cloud catalog extracted using the CO(1–0) assignment. Table 1 summarizes the ratio of the properties of these clouds observed in both emissions. We report the following conclusions:

1. Properties observed in both emission lines roughly follow the locus of equality, a deviation of approximately 20 % is observed irrespective of the property being compared.

2. We believe there are two possible reasons behind the deviation. First, the clouds present in the nucleus of a galaxy strongly emit only in CO(1–0) emission lines, therefore, an inherent difference between the reported properties in the two emissions is expected and second, the slope from the RANSAC algorithm doesn't converge to a single value every time it is used, which makes some iterations favorable and some unfavorable to this study.

3. Considering a deviation of less than 25 % from the locus of equality, the sources behind this deviation, and the advantages of mapping GMCs using the CO(2–1) transition over CO(1–0) transition, i.e., less investment of a telescope's time, we believe CO(2–1) emission can substitute mapping of galaxies in CO(1–0) emission lines without a loss of critical information about the extracted molecular clouds. To corroborate our claim, we plan on extending this cloud-based study to more galaxies and performing statistical analysis on the RANSAC algorithm's outputs. We believe, these extensions will not only help us understand our result better but will also help us further justify them.

REFERENCES

Bolatto, A. D., Leroy, A. K., Rosolowsky, E., Walter, F., & Blitz, L. 2008, ApJ, 686, 948, doi: 10.1086/591513

Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207, doi: 10.1146/annurev-astro-082812-140944

Donovan Meyer, J., Koda, J., Momose, R., et al. 2013, ApJ, 772, 107, doi: 10.1088/0004-637X/772/2/107

Freeman, P., Rosolowsky, E., Kruijssen, J. M. D., Bastian, N., & Adamo, A. 2017, MNRAS, 468, 1769, doi: 10.1093/mnras/stx499

Herrera, C. N., Pety, J., Hughes, A., et al. 2020, A&A, 634, A121, doi: 10.1051/0004-6361/201936060

Leroy, A. K., Hughes, A., Liu, D., et al. 2021, ApJS, 255, 19, doi: 10.3847/1538-4365/abec80

Leroy, A. K., Rosolowsky, E., Usero, A., et al. 2022, ApJ, 927, 149, doi: 10.3847/1538-4357/ac3490

McKee, C. F., & Ostriker, E. C. 2007, ARA&A, 45, 565, doi: 10.1146/annurev.astro.45.051806.110602

Peñaloza, C. H., Clark, P. C., Glover, S. C. O., & Klessen, R. S. 2018, MNRAS, 475, 1508, doi: 10.1093/mnras/stx3263

Rosolowsky, E., & Leroy, A. 2006, PASP, 118, 590, doi: 10.1086/502982

Rosolowsky, E., Hughes, A., Leroy, A. K., et al. 2021, MNRAS, 502, 1218, doi: 10.1093/mnras/stab085

Saintonge, A., Catinella, B., Tacconi, L. J., et al. 2017, ApJS, 233, 22, doi: 10.3847/1538-4365/aa97e0

APPENDIX

Data reduction steps that were briefly mentioned in section 2 are discussed in much further detail here. The first section describes the PHANGS-ALMA pipeline and how it processes calibrated visibility data into science-ready data product (Leroy et al. 2021), and the second section expounds on the work of PYCPROPS algorithm in reducing and selecting data (Rosolowsky et al. 2021).

.1. *PHANGS-ALMA*

We get a well calibrated visibility data after it goes through the ALMA interferometric calibration pipeline. This calibrated u-v data of galaxies is then processed by the PHANGS pipeline in the following way:

1. **Staging of the u-v data** - The pipelines first extracts the relevant science targets and spectral windows from the calibrated u-v data coming from the ALMA pipeline, followed by subtracting the continuum signal from this extracted data. To use the extracted data for imaging, the pipeline then re-grids and re-bins it onto a common velocity grid. After, it flag lines and collapses windows in the spectrum from the original measurements to produce "a continuum-only u-v dataset".

1.1 Starting Point

Using the same version of CASA that was used by the ALMA pipeline for processing uncalibrated visibility data to calibrated visibility data, the PHANGS pipeline applies "calibration and flagging produced by the ALMA observatory interferometric pipeline to the data".

1.2 Manual quality assessment of PHANGS-ALMA visibility data

After applying calibration in step 1, the PHANGS pipeline performs a bunch of tests to inspect the effect of additional flagging on the final image.

1.3 Staging and continuum subtraction

The data with modest amount of flagging is then subjected an extraction process where the pipeline extracts the calibrated data pertaining to each galaxy and spectral line by implementing the CASA task 'split'. For the continuum subtraction process, multiple spectral windows are selected, or all, if needed. This is followed by subtracting the continuum by using the CASA task 'uvcontsub'. Note, for CO(2–1), the observations used a spectral window with a width of $\approx 1200\ kms^{-1}$.

1.4 Spectral regridding and rebinning

"After continuum subtraction, for each spectral line of interest", the PHANGS pipeline now creates a "line-specific measurement set. This is done to combine all data on the common velocity grid to be used for imaging".

1.5 Continuum extraction

In basic terms, a line free continuum measurement set is extracted by data has been subjected to spectral regridding and rebinning.

1.6 Staged u-v data

Steps 1.1 to 1.5 finally culminates in a "single, combined u-v measurement set for each combination of target, spectral product, and array combination."

2. **Imaging and Deconvolution** - This step is performed by iterating over CASA's tclean task, which creates a dirty image and aligns it with any user supplied mask, while simultaneously creating masks "that guide the deconvolution" operation. During this step, the algorithm uses a mixture of multi-scale clean with high signal to noise ratio threshold and single scale clean with low signal to noise ratio threshold calls to create signal based clean masks.

3. **Reduction of Total Power Data** - Here, the total power data is processed. The individual spectra are fit and have their baseline removed. Then, the spectral data units are converted to Jansky/beam and all the observations are combined into a single dish data cube.

4. **Post processing** - The data cubes and the imaged data from step 2 and the reduce total power data from step 3 are post processed here. The data is subjected to primary beam corrections and convolved to a round synthesized beam, and the interferometric and the total power data are combined to convert the data units to Kelvin. This allows exportation of images to science-ready FITS cubes.

.2. *PYCPROPS*

The spectral line fits cubes from the PHANGS-ALMA pipeline are made of stacks of velocity planes consisting of data in units of Kelvin, and they are passed through the PYCPROPS algorithm for giant molecular cloud identification and categorization. Cataloguing molecular cloud takes place in 4 distinct phases.

Signal identification - Starting with the spectral line fits data cube described above, "a new three dimensional estimate of the RMS brightness temperature noise, σ_T(x,y,v) following the same procedure as described in section 2.2 (Rosolowsky et al. 2021)" is constructed. With the knowledge of T(x,y,v) at each pixel and a local noise estimate, σ_T(x,y,v), a Boolean mask, M(x,y,v), is established that indicates "which pixels are likely to contain significant emission". Just like in paper by Rosolowsky and Leroy (Rosolowsky & Leroy 2006), two masks: a high significance mask and a low significance masks are constructed; "all contiguous

regions in the low significance mask that do not contain any pixels in the high mask significance mask are rejected". This filtered low significant mask is the final Boolean emission mask M(x,y,v) that indicates which pixels are likely real emission. "Cloud identification and decomposition is restricted to lie within this emission mask".

Cloud decomposition algorithm - Once the final emission Boolean mask has been established, each cloud is linked with a local maximum of emission in the spectral fits cube data. The algorithm starts decomposing clouds by first identifying the local maxima and then checking whether these maxima are significant compared to noise fluctuation expected from the noise in the data. Then, it also checks whether the identified local maxima are distinct from the neighbouring local maxima. "Finally, it assigns emission to each local maximum". Local maxima are identified using the advance algorithm provided by the ASTRODENDO package and only those maxima are selected that are likely to be significant. "The PYCPROPS algorithm rejects local maxima based on four criteria: significance, number of associated pixels, separation, and uniqueness".

First, the algorithm demands only significant local maxima to be selected and it implements this by requiring that intensity at local maximum, T_{max}, "be an interval δ above the highest contour of emission containing at least one other neighbour". Second, it requires "a minimum number of cube pixels to be uniquely associated with the local maximum". This makes sure that small poorly defined objects in the spectral cube are rejected. Third, local maxima are required to be separated from each other, spatially by a minimum distance of d_{min} and spectrally by a minimum distance of v_{min}. This ensures all the "maxima are reasonably resolved from one another". Finally, the algorithm makes sure the selected local maxima are unique by merging them with another object and testing whether their properties change significantly.

Once a set of unique, significant local maxima has been identified, they are then used "as seeds to assign emission to molecular clouds". A watershed algorithm is used to associate "each pixel in the emission mask to a molecular cloud, generating a label cube L(x,y,v)".

Cloud decomposition in PHANGS-ALMA - To cloud decompose the PHANGS-ALMA data, the following conditions are imposed: $d_{min} = 0$ and $v_{min} = 0$.

Property estimation - After clouds are assigned with emission, "CPROPS calculates properties of the identified clouds." Some of the properties include luminosity, luminous mass, radius, velocity dispersion, eccetera.