Insights into the PHANGS-ALMA Catalogue of Giant Molecular Clouds

Feras Natsheh
ASTR-310: Stellar Astrophysics

May 3, 2024

Abstract

This study utilizes the Giant Molecular Cloud (GMC) [1] catalogue from the PHANGS-ALMA survey [2] to examine the physical properties of GMCs across different galaxies and their correlation with star formation. By analyzing GMC properties such as mass, size, and velocity dispersion, I hope to explore how these properties vary across different galactic environments and contribute to star formation processes. I utilize multiple Python libraries, mainly Astropy, pandas, NumPy, SciPy, Matplotlib and Seaborn to statistically analyze and visualize the GMCs across 10 galaxies from the catalogue. I also model Larson's scaling relations and test their validity on this catalogue. Finally, I try to relate this project to further our understanding of the role GMCs play in stellar astrophysics and star formation.

Introduction

Giant Molecular Clouds (GMCs) are the pillars of star formation. They consist of molecular gas predominantly composed of molecular hydrogen (H_2), helium, and of carbon monoxide (CO). GMCs represent the densest part of the interstellar medium, with masses typically ranging from 10^4 to 10^7 solar masses and radii spanning anywhere from a few parsecs to a couple of hundred parsecs. This section will introduce the physics underlying the study of GMCs, the chemical composition of GMCs, and Larson's Scaling Relations, which is a primary focus of this project.

The Physics of GMCs

Stellar Structure and Evolution by Ryden & Pinsonneault examines the instability of GMCs. [3] Jean's Mass and Length define the conditions of instability in GMCs. Jean's mass is the critical mass above which a cloud of gas and dust becomes gravitationally unstable and begins to collapse to form stars. It is expressed as

$$M_J = \frac{\pi}{6}\rho_0\lambda_J^3 \tag{1}$$

where ρ_0 is the assumed uniform density of a region of gas in a cloud, and λ_J is Jeans's length.

Jeans's length (or Jeans Radius) is the critical radius for a cloud or a part of a cloud at which gravitational forces overcome internal pressures, leading to collapse. It is given by the equation

$$\lambda_J = \frac{2\pi}{k_J} = \sqrt{\pi}\frac{c_s}{(G\rho_0)^{1/2}} \tag{2}$$

where c_s is the isothermal sound speed.

To study the pre-collapse stages of GMCs, we ought to take a look at their stable phases. The stability of GMCs is a balance of gravitational forces and internal kinetic pressures, where turbulence plays a crucial role in either fostering or hindering star formation. Turbulence in GMCs is the chaotic, random motions of gas molecules and dust within these clouds. Unlike orderly flow patterns, turbulent flows are characterized by random and unpredictable fluctuations in velocity and density at a range of scales. In GMCs, turbulence is characterized through velocity dispersion, σ, which is the measure of the statistical spread of velocities within the cloud, indicating how much the velocities of different parts of the cloud vary from the mean velocity. Velocity dispersion is a key factor in determining whether a cloud will collapse to form stars or if it will remain stable. It provides internal pressure that can counterbalance the gravitational pull trying to collapse the cloud. If the velocity dispersion is high relative to the cloud's size and mass, it can prevent collapse by providing enough internal kinetic energy to counteract gravitational forces. Furthermore, high velocity dispersion can indicate active or intense star formation processes within the cloud, as forming stars inject energy back into the cloud through stellar winds and outflows, increasing the local velocity dispersion.

Chemical Composition of GMCs

The structural complexity and extensive size of GMCs make them key subjects in the study of the molecular interstellar medium and star formation processes across various galactic environments. The most abundant molecules in GMCs are H_2, He and CO. [4] Carbon monoxide is critical for observational astrophysics because it serves as a reliable tracer for the more abundant but less directly detectable H_2. A study by Nishimura et al. (2017) provides a detailed analysis of the chemical composition in the giant molecular cloud W51. The study specifically discusses the abundance of CO, highlighting that the integrated intensity of ^{13}CO and contributions from 11 molecular lines across various sub-regions provide a broader perspective on the cloud's chemical heterogeneity. [5] Therefore, CO serves to probe a mass estimate of the GMC.

Another way to probe the estimated mass of the GMC is to consider the fact that the cloud is in virial equilibrium prior to collapse. GMCs are in virial equilibrium due to the balance between their gravitational self-attraction and their internal kinetic energies, including turbulence and thermal pressures. Astronomers derive virial mass estimates for GMCs by analyzing the balance between gravitational and kinetic energy, expressed through the virial theorem as:

$$2KE + U = 0 \tag{3}$$

where KE is the total kinetic energy and U is the gravitational potential energy. This approach uses observable properties like the size of the cloud and the velocity dispersion of its gas. The virial mass M_{vir} is then typically calculated from line width-size relationships or direct measurement of gas motions within the cloud [6].

Larson's Scaling Relations

Larson's Laws refer to three empirical relationships identified by Richard Larson in 1981 through observations of molecular clouds, primarily in our Milky Way galaxy. [7] These relationships describe some general properties of molecular clouds, particularly those relevant to their structure and dynamics. Although they are referred to as "laws," they are more accurately considered empirical relationships or rules because they are not universal physical laws derived from first principles.

1. Size-Line Width Relation: The velocity dispersion (σ) of a cloud increases with the size of the cloud. The relation is a power law, $\sigma \propto R^{\beta}$, where $\beta \approx 0.38$. This relation suggests that larger clouds have greater internal motions. It has been interpreted as evidence of turbulence within molecular clouds, with larger clouds exhibiting more complex, turbulent motion.
2. Virial Parameter Relation: The mass of a cloud is proportional to its size and velocity dispersion in a way that suggests most clouds are in approximate virial equilibrium ($M \propto R^{\gamma}$) where $\gamma \approx 0.2$. This relation suggests that the gravitational and kinetic energies within clouds are balanced, supporting the idea that GMCs are in a state of equilibrium when they are not actively collapsing.
3. Mass-Density Relation: The mean density of a GMC decreases as its size increases ($\rho \propto R^{-\alpha}$) where $\alpha \approx 1.1$. This indicates that larger clouds are less dense.

Data & Methods

PHANGS-ALMA Catalogue

The Physics at High Angular Resolution in Nearby GalaxieS (PHANGS) project created the catalogue used in this study. The catalogue is based on CO emission data from the PHANGS-ALMA survey, which includes 74 galaxies. For fair comparison across different galaxies, they selected galaxies that could be smoothed to a common resolution of 90 pc and a matched noise level. This reduced their catalogue to 10 galaxies.

The catalogue uses environment masks to categorize GMCs according to different galactic environments. This classification aids in understanding how the properties of molecular clouds vary with their galactic context. However, this categorization is beyond the scope of this study as it requires a more advanced analysis.

The catalogue also analyzes their properties like mass, radius, and line width (σ) across different environments. The catalogue has 11 variables, as listed in Table 1. The next subsection outlines which variables this study focuses on.

Variable	GALAXY	CLOUDNUM	XCTR_DEG	YCTR_DEG	VCTR_KMS	RAD_PC	ECCEN	SIGV_KMS	FLUX_KKMS_PC2	MLUM_MSUN	MVIR_MSUN
Meaning	Name of galaxy containing GMC	number of GMC within a galaxy	Right Ascension (RA) of the GMC center	Declination (Dec) of the GMC center	Velocity of the GMC center w.r.t Local Standard of Rest	Radius of the GMC	Eccentricity of the GMC (cloud's shape)	Velocity dispersion of the GMC	CO(2-1) flux from the GMC	CO-based mass estimate. Luminous mass of the GMC estimated from CO emissions	Virial-based mass estimate of the GMC
Units	N\A	N\A	degrees	degrees	km/s	parsecs	N/A	km/s	K km pc^2/s	Solar masses	Solar masses
Data Type	string	int64	float64	float64	float64	float64	float64	float64	float64	float64	float64

Figure 1: Varaibles in the ecsv Catalogue

Variables of interest

XCTR_DEG and **VCTR_DEG** are the right ascension and the declination of the GMC, respectively. These two variables are essential in plotting the spatial distribution of GMCs across galaxies.

RAD_PC is the radius of the GMC. Given the relatively high eccentricity of the GMCs (a mean of 0.78 across the 10 galaxies), we're making the assumption that this is the radius of the semi-major axis, though the catalogue does not specify.

SIGV_KMS is the velocity dispersion, σ, of the cloud, as explained earlier.

MLUM_MSUN is the CO-based mass estimate of the GMC. As explained above, this serves as one of the two mass estimates that this catalogue outlines.

MVIR_MSUN is the other mass estimate in the catalogue and it is the mass estimate derived from the virial equilibrium state of the cloud.

Together, these variables serve as the main focus of this study, as they help us understand the physical characteristics of GMCs.

Results

Analyzing the variables of interest in the catalogue, I was able to obtain multiple plots of various properties of the GMCs. This section outlines the main results and discusses their significance. Refer to the appendix for the full code used in this study.

Physical Properties

The team has catalogued a total of 5,758 clouds in the 10 galaxies, with an average of 436 clouds per galaxy. However, this count varies significantly across galaxies. Table 1 lists the 10 galaxies and the number of GMCs within each. Notably, NGC4826 contains far fewer GMCs than the other galaxies, which requires further investigation. This discrepancy could be due to the distinct galactic environment in NGC4826, which may affect the formation or stability of GMCs. The initial results of the PHANGS team indicate that this galaxy seems to be an anomaly. [2]

Table 1: Number of GMCs in Each Galaxy

Galaxy	Number of GMCs
ngc0628	472
ngc1637	275
ngc2903	810
ngc3521	1432
ngc3621	394
ngc3627	1048
ngc4826	48
ngc5068	74
ngc5643	695
ngc6300	510

Comparing the mass estimates of CO emissions and the virial equilibrium is another step in ensuring our catalogued GMCs are not collapsing or going abnormal activities. This helps ensure that this sample is representative. When the CO-based mass estimate and the virial-based mass estimate for a GMC are similar, it typically implies that the cloud is in a state of gravitational equilibrium and that the CO emission effectively traces the mass distribution within the cloud. [8] If the virial mass is significantly higher, it could be due to non-gravitational motions within the cloud, such as strong turbulence or external forces, which are not accounted for in the simple virial analysis.

Table 2: Mean CO Virial Mass Estimates of GMCs

Galaxy	MLUM_MSUN	MVIR_MSUN
ngc0628	2.817329×10^6	2.092842×10^6
ngc1637	1.729325×10^6	2.071854×10^6
ngc2903	3.902322×10^6	5.263166×10^6
ngc3521	4.141644×10^6	2.861519×10^6
ngc3621	2.111776×10^6	2.425354×10^6
ngc3627	5.213530×10^6	4.176494×10^6
ngc4826	6.182155×10^6	1.474709×10^7
ngc5068	1.369563×10^6	1.325697×10^6
ngc5643	2.571373×10^6	2.669681×10^6
ngc6300	2.737088×10^6	3.493979×10^6

Larson's Laws

Larson's first law, $\sigma \propto R^\beta$, seems to be applicable to this catalogue. With a β value of 0.33, the law seems to hold true in general, given how close this best fit value is to the empirically original value of 0.38. Figure 2 shows the best fit line. Note that the plot in figure 2 has a

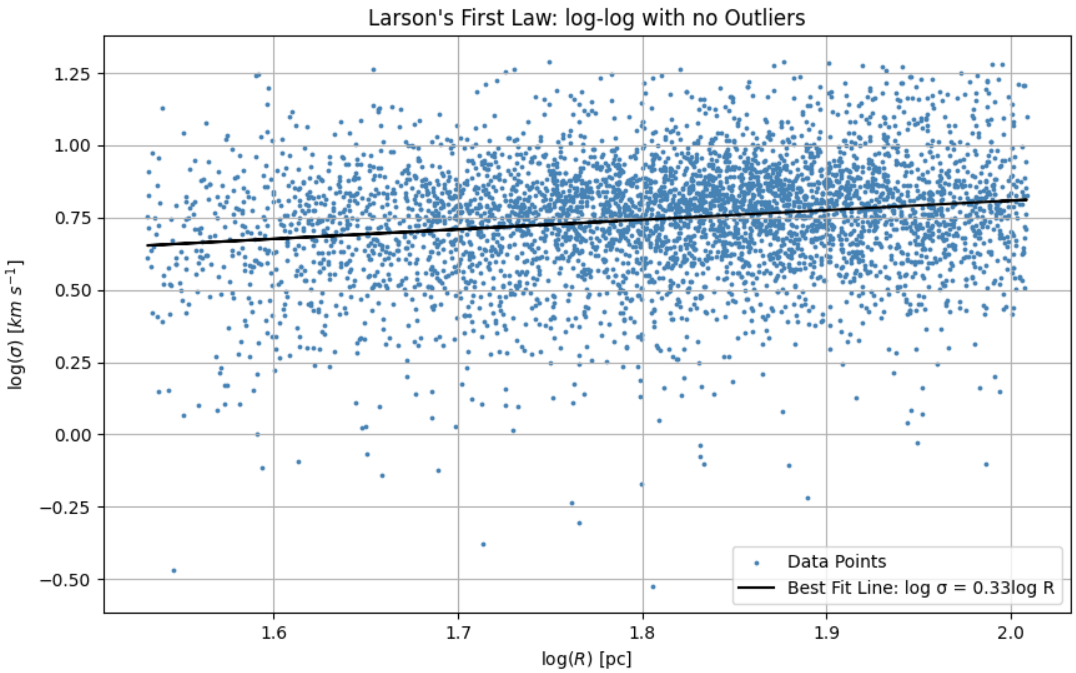


Figure 2: Larson's First Law

constrained range of (34, 102) parsecs to clean some of the outliers.

The second law also seems to hold true in general. The power law has a coefficient that is close to the empirical one found by Larson.

The third law seems to hold true as a general relationship (radius increases, density decreases). However, the best fit power law coefficient value is somewhat divergent from the -1.1 that we expect. Figures 4 and 5 show Larson's third law based on the two mass estimates provided in the catalogue. This deviation from -1.1 could stem from the assumptions made during my calculations, which I will discuss in more detail in the conclusion section.

Distribution Across Galaxies

Figure 6 shows the distribution of GMC radii by galaxies. The mean cloud radius is 68 parsecs. Similar CO-based mass and virial-based mass. Moreover, the spatial distribution of GMCs across each galaxy is provided in Figure 7(a-j). NGC4826 stands out as anomaly in its minuscule number of GMCs, significant difference between the CO-based mass estimate and the virial-based mass estimate, and the spatial distribution of its GMCs.

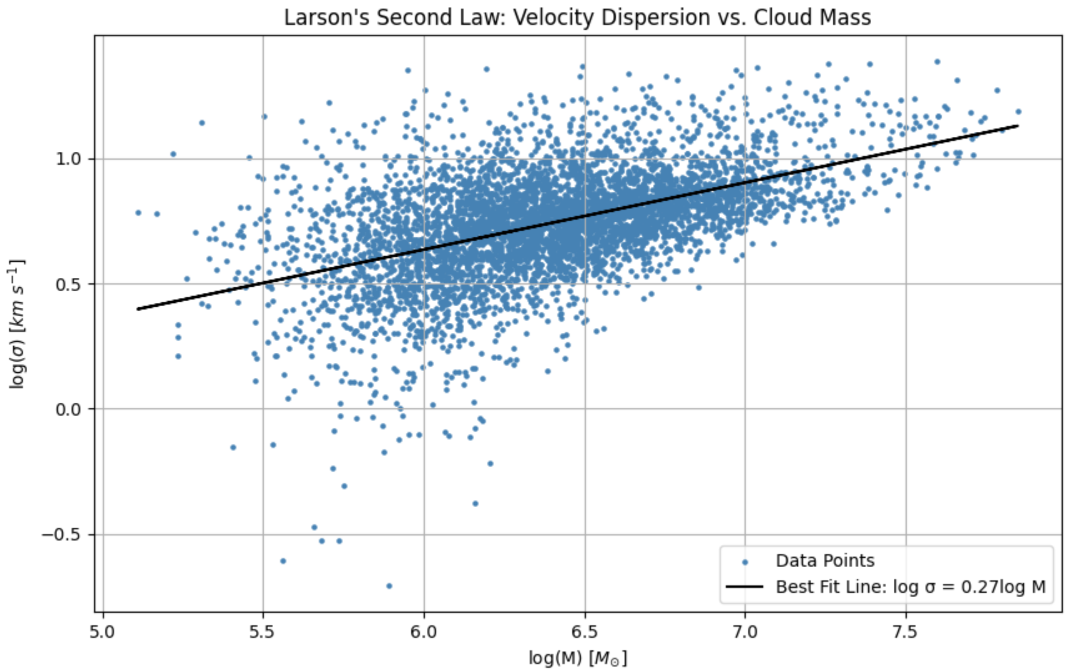


Figure 3: Larson's Second Law

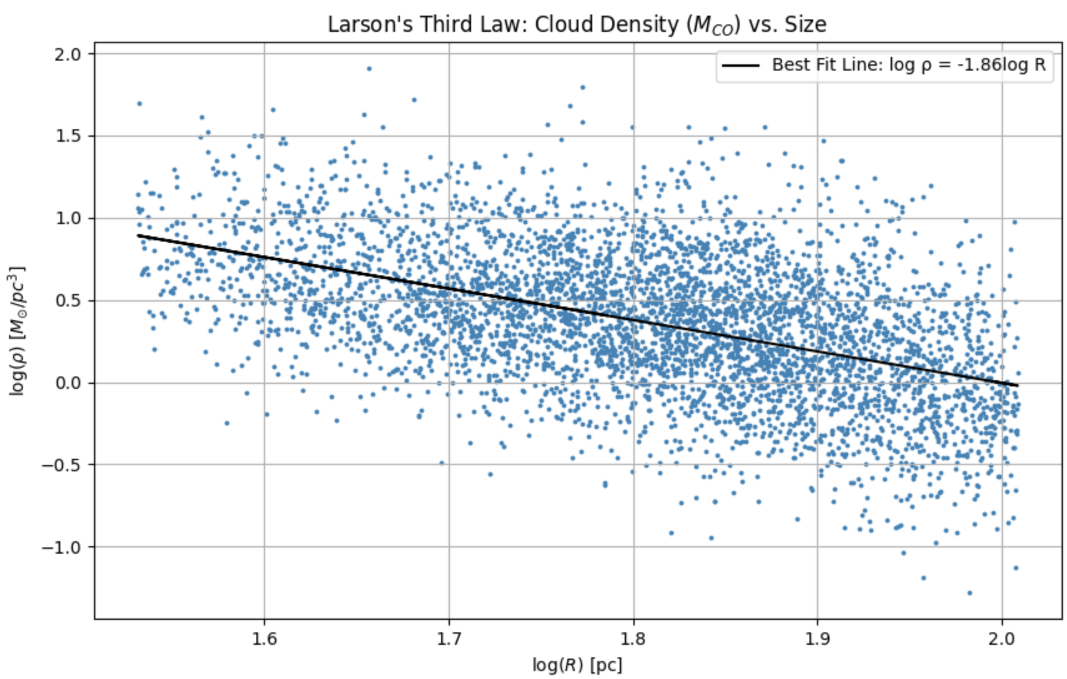


Figure 4: Larson's Third Law: M_{CO} Estimate

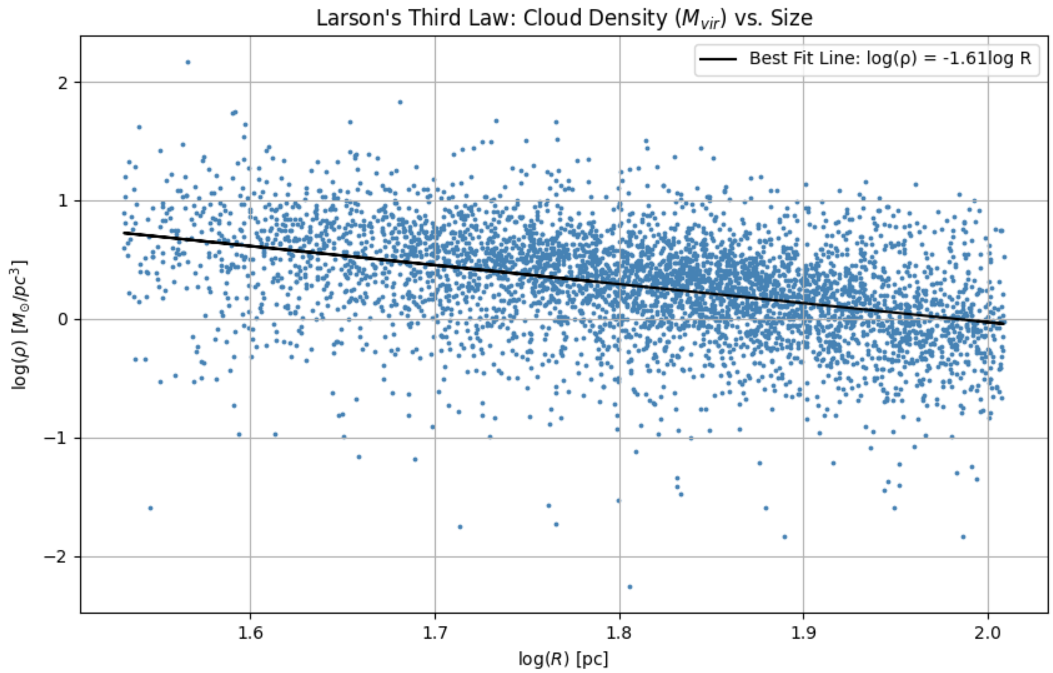


Figure 5: Larson's Third Law: M_{vir} Estimate

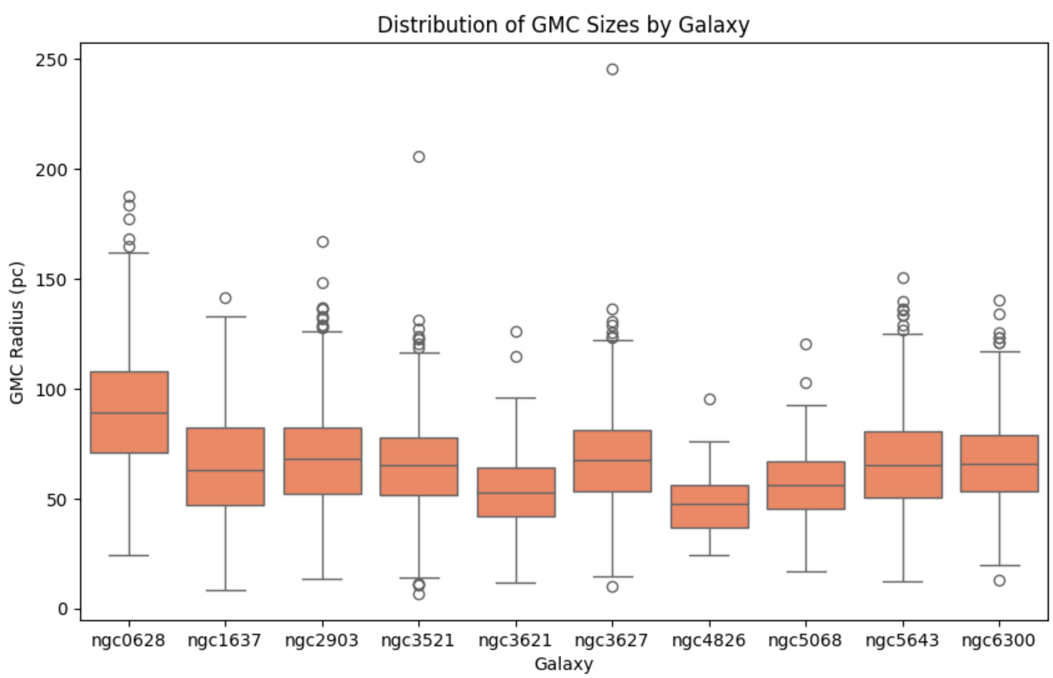
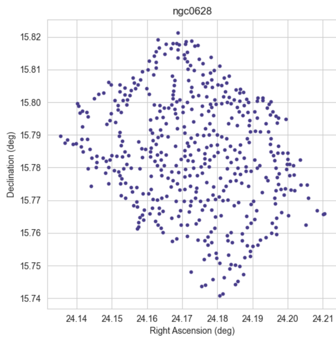
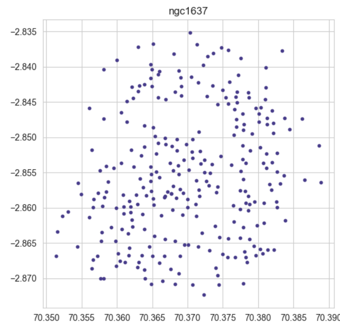


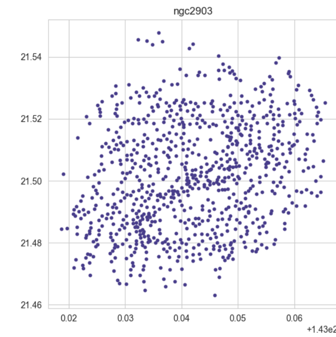
Figure 6: Distribution of GMCs by Radius Across the 10 Galaxies



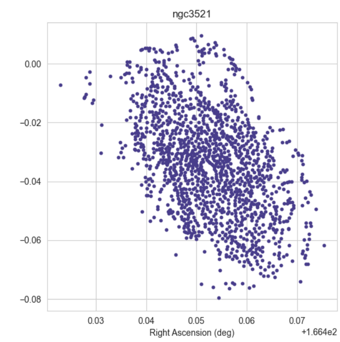
(a) NGC0628



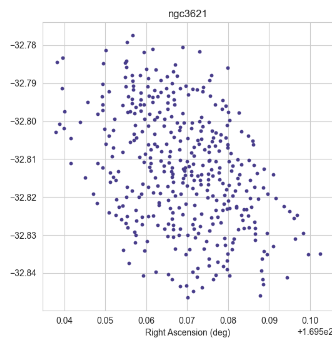
(b) NGC1637



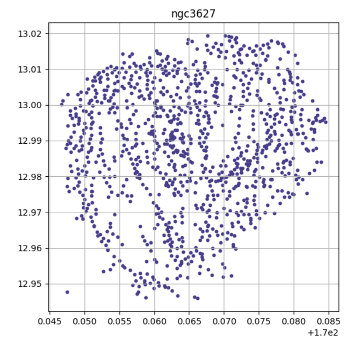
(c) NGC2903



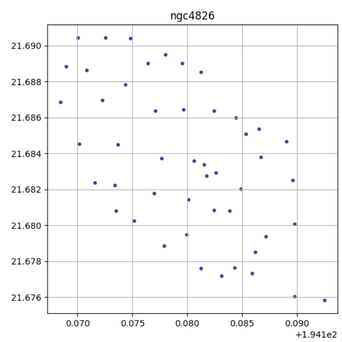
(d) NGC3521



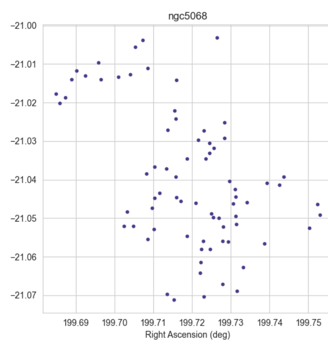
(e) NGC3621



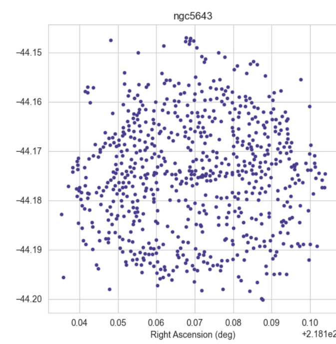
(f) NGC3627



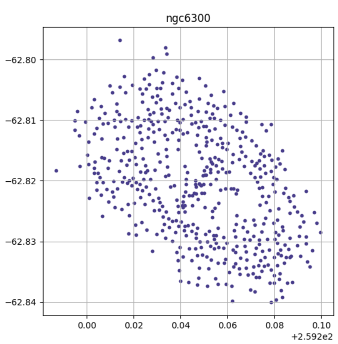
(g) NGC4826



(h) NGC5068



(i) NGC5643



(j) NGC6300

Figure 7: Spatial Distribution of GMCs Across Galaxies

Conclusion

Relevance to Stellar Astrophysics

The analysis of GMC properties and their correlation with star formation has important implications for stellar astrophysics. Understanding the size, mass, and velocity dispersion of GMCs helps reveal the conditions under which stars form, particularly in different galactic environments. The scaling relations presented by Larson continue to hold broadly true, reinforcing their significance in the theoretical framework of star formation. Additionally, discrepancies between the virial and CO-based mass estimates help in the study of turbulent dynamics and external forces affecting GMCs.

Source of Error

There are several potential sources of error in this analysis:

Sample Size and Selection Bias: The reduced catalogue contains only 10 galaxies, potentially limiting the study's generalizability. Selection bias could also arise from choosing only galaxies with specific characteristics.

Assumptions in Calculations: Assumptions made in estimating mass and density, particularly relying on certain empirical relationships, can introduce errors. For example, when computing density for Larson's third law, I assumed uniform spherical shape for the GMCs in order to apply the density relation, $\rho = \frac{M}{4/3\pi R^3}$. This assumption (an average GMC has a spherical shape) might be introducing error, given the high mean eccentricity (≈ 0.78) of the GMCs across the 10 galaxies. This could explain the resulting deviation in the best fit coefficient, α, in Larson's third law.

Outlier GMCs: Outliers in GMC properties may skew the statistical analysis. Such outliers could result from observational limitations or unique galactic environments. I tried to limit this source of error by cleaning the outliers (getting rid of "nan" entries) and choosing a range that excludes the outliers that don't have a null value.

References

[1] GMC_catalogue.ecsv.
https://www.canfar.net/storage/list/phangs/RELEASES/Rosolowsky_etal_2021

[2] Rosolowsky, E., Hughes, A., Leroy, A. K., Sun, J., Querejeta, M., Schruba, A., Usero, A., Herrera, C. N., Liu, D., Pety, J., Saito, T., Bešlić, I., Bigiel, F., Blanc, G., Chevance, M., Dale, D. A., Deger, S., Faesi, C. M., Glover, S. C., ... Williams, T. G. (2021). Giant molecular cloud catalogues for Phangs-Alma: Methods and initial results. Monthly Notices of the Royal Astronomical Society, 502(1), 1218–1245. https://doi.org/10.1093/mnras/stab085

[3] Pinsonneault, M. H., & Ryden, B. S. (2023). Stellar structure and evolution, 7.1: Molecular Clouds and Instability.Cambridge University Press.

[4] Georgia State University. (n.d.). STAR FORMATION. Star Formation. http://www.astro.gsu.edu/w̃iita/a3010lecture_7.html

[5] Nishimura, Y., Watanabe, Y., Harada, N., Shimonishi, T., Sakai, N., Aikawa, Y., Kawamura, A., & Yamamoto, S. (2017). Molecular-cloud-scale chemical composition. II. mapping spectral line survey toward W3(OH) in the 3 mm band. The Astrophysical Journal, 848(1), 17. https://doi.org/10.3847/1538-4357/aa89e5

[6] Goldbaum, N. J., Krumholz, M. R., Matzner, C. D., & McKee, C. F. (2011). The global evolution of giant molecular clouds. II. the role of accretion. The Astrophysical Journal, 738(1), 101. https://doi.org/10.1088/0004-637x/738/1/101

[7] Plunkett, A. (2013, November 6). Astrophysical classics: Larson's "Laws." astrobites. https://astrobites.org/2012/11/18/astrophysical-classics-larsons-laws/

[8] Lada, C. J., & Dame, T. M. (2020). The mass–size relation and the constancy of GMC surface densities in the milky way. The Astrophysical Journal, 898(1), 3. https://doi.org/10.3847/1538-4357/ab9bfb

[9] ECSV Format: https://docs.astropy.org/en/stable/io/ascii/ecsv.html

[10] Pandas documentation(agg, groupby, describe, dropna): https://pandas.pydata.org/docs/

[11] Seaborn documentation (boxplots, histograms): https://seaborn.pydata.org/

[12] SciPy Linear Regression:
https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.linregress.html

Appendix [9], [10], [11], [12]

```
from astropy.table import Table
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
from scipy.stats import linregress

table = Table.read('GMC_catalogue.ecsv', format='ascii.ecsv')
df = table.to_pandas()
print(df.head(10))
df.describe()
print(df.isnull().sum())

MCO_Mvir_group = df.groupby('GALAXY').agg({'MLUM_MSUN': 'mean', 'MVIR_MSUN': 'mean'})
print(MCO_Mvir_group)

#get number of GMCs in each galaxy
grouped = df.groupby('GALAXY')['CLOUDNUM']
cloud_count = grouped.count()
print(cloud_count)
print(np.sum(cloud_count))
```

```
plt.figure(figsize=(10, 6))
sns.boxplot(x='GALAXY', y='RAD_PC', data=df, color='coral')
plt.title('Distribution of GMC Sizes by Galaxy')
plt.xlabel('Galaxy')
plt.ylabel('GMC Radius (pc)')
plt.savefig('BOXPLOT Distribution of GMC Sizes by Galaxy')
plt.show()

### I am only providing code for one of the Larson's law plots,
the rest follow a very similar format ###
# Larson's first law: Velocity dispersion is proportional to cloud size (R)

#get rid of all nans
df_clean = df.dropna(subset=['RAD_PC', 'SIGV_KMS'])

radius = df_clean['RAD_PC']
velocity_dispersion = df_clean['SIGV_KMS']

#create best fit
slope, intercept, r_value, p_value, std_err = linregress(np.log10(radius),
    np.log10(velocity_dispersion))

log_radius = np.log10(radius)
line = slope * log_radius + intercept

plt.figure(figsize=(10, 6))
plt.scatter(radius, velocity_dispersion, alpha=1, s=3, color='steelblue',
    label='Data Points')
plt.title('Velocity Dispersion vs. Cloud Radius')
plt.xlabel('$R$ (pc)')
plt.ylabel('$\\sigma$ ($km \\ s^{-1}$)')
plt.legend()
plt.grid(True)
plt.savefig('Larsons First Law')
plt.show()

gal = 'ngc----'
gal_data = df[df['GALAXY'] == gal]

#get coords
x = gal_data['XCTR_DEG']
y = gal_data['YCTR_DEG']

plt.figure(figsize=(6, 6))
sc = plt.scatter(x, y, s=10, color='darkslateblue')
plt.xlabel('Right Ascension (deg)')
plt.ylabel('Declination (deg)')
```

```
#plt.xscale('log')
#plt.yscale('log')
plt.title(f'{gal}')
plt.grid(True)
plt.savefig(gal)
plt.show()

print(f"Slope: {slope:.2f}")
```