Machine Learning-Driven Determination of 3D Structures of Molecular Clouds in High-Resolution mm Images

Efe Öztaban

Sabancı University

MACHINE LEARNING FOR ASTROPHYSICS
2ND EDITION CATANIA, 8-12 JULY, 2024

Emrah Kalemci (Sabanci University)
Ahmed Abdullah Abbasi,
Atakan Saraçyakupoğlu (Sabanci University),

Maria Diaz-Trigo (ESO), Thomas Stanke (MPE)
Tom Maccarone (Texas Tech), John Tomsick (SSL, UC Berkeley)
Simone Migliari (ESA), James Miller-Jones (Curtin)

Sabancı Üniversitesi

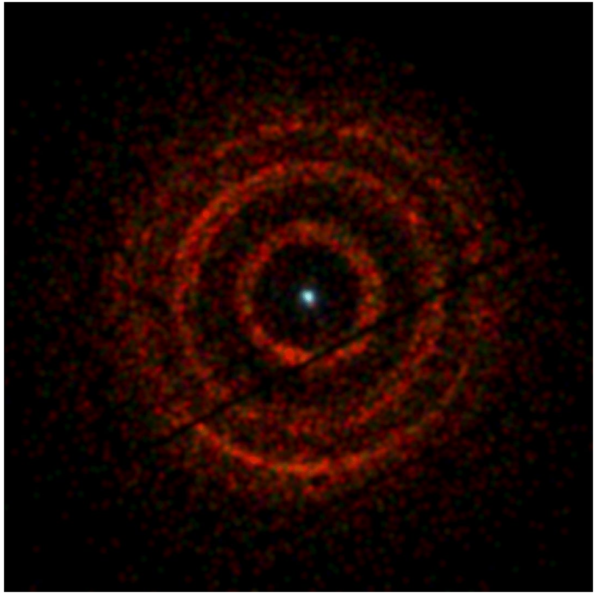
OVERVIEW OF MY TALK

- ➔ Dust Scattering Halos
- ➔ Our Method to Determine the Source Distance
- ➔ Our Test Source: 4U 1630-47 and Our Datasets
- ➔ Molecular Clouds in the Line of Sight

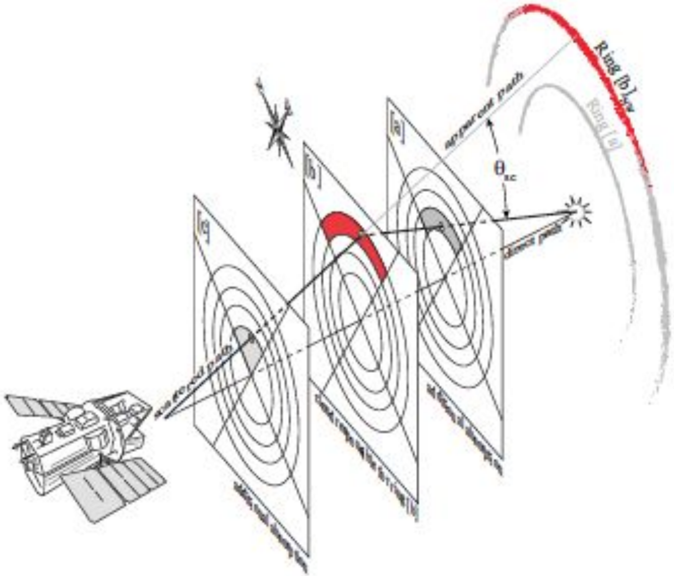
- ➔ Objectives and Constraints
- ➔ High-Resolution mm Images
- ➔ Finding the 3D Structure of the Clouds
  - ◆ Local Peak Detection
  - ◆ Data Segmentation
  - ◆ Removal of Non-Cloud Regions
  - ◆ Clustering
- ➔ Results

- ➔ DSH Simulations
- ➔ Future Work

DUST SCATTERING HALOS



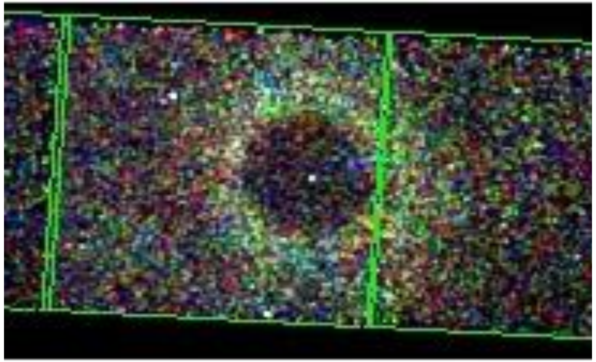
V404 Cygni, Andrew Beardmore (Univ. of Leicester) and NASA/Swift



Heinz et al. 2016

DUST SCATTERING HALOS

➔ Why to study DSH?
  - DSH intensity, shape and spectral properties depend on the distance to the source, its intrinsic spectrum and time history, the distribution of dust grains and their size and composition!
  - So, we can learn about these from studying DSHs
  - But, how can we determine the distance of the source object by studying it’s DSH?

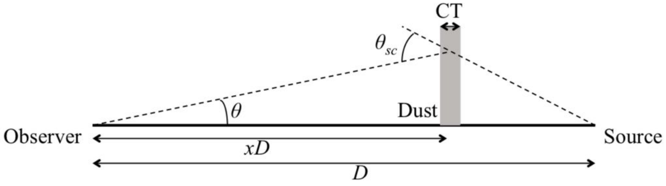


DSH of 4U 1630-47 from Chandra (2016)

OUR METHOD TO DETERMINE THE SOURCE DISTANCE

- ➔ Create 3D cloud shapes using the mm data from APEX.
- ➔ Determine cloud distances using a standard Galaxy rotation curve.
- ➔ Place each cloud at either the near distance or the far distance
- ➔ Using the flux history from the source and the methodology outlined in Kalemci. E, et al., regenerate the scattering halo
- ➔ A self absorption correction applied to take into account extinction for the clouds in front of the clouds that emitted X-rays.
- ➔ Best fit to observation data is selected among the simulated DSH images and distance of the source and clouds are determined by used parameters.

So, we need the 3D structure of the molecular clouds in the line of sight!



$$I_{\nu,r} = N_{H,r} \boxed{\frac{d\sigma_{sc,E}}{d\Omega}} \frac{F_\nu(t = t_{obs} - \boxed{\Delta t})}{(1-x)^2} \exp\left[-\sigma_{ph,E} \sum_{i=1}^{r} N_{H,r}\right]$$

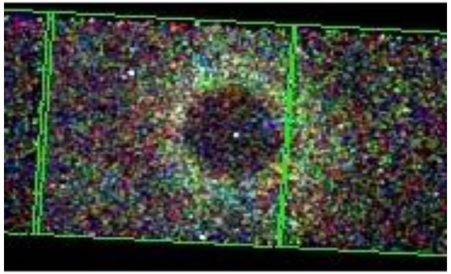
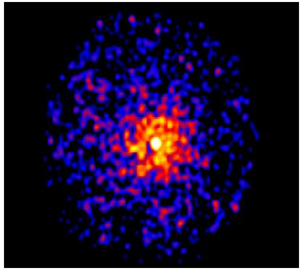
$$\boxed{\frac{d\sigma_{sc,E}}{d\Omega}} \sim C \left(\frac{\theta/(1-x)}{1000''}\right)^{-\alpha} \left(\frac{E}{1keV}\right)^{-\beta} \qquad \boxed{\Delta t} = \frac{xD\theta^2}{2c(1-x)}$$

Kalemci, E., Maccarone, T. J., & Tomsick, J. A. (2018). A dust-scattering halo of 4U 1630–47 observed with chandra and swift: new constraints on the source distance. The Astrophysical Journal, 859(2), 88.

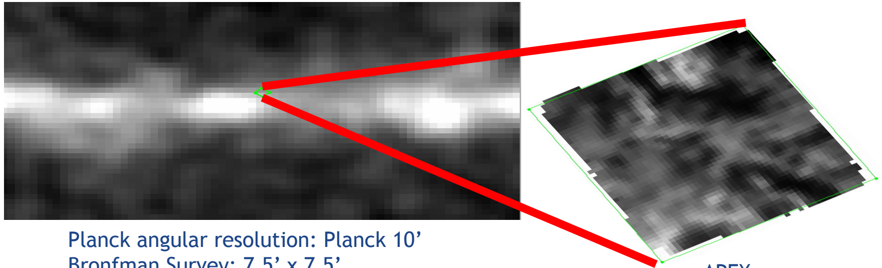
OUR TEST SOURCE: 4U 1630-47 and OUR DATASETS

- ➔ Source Object: 4U 1630-47
  - ◆ a Galactic black hole transient (GBHT)

- ➔ Our datasets:
- ➔ Chandra X-ray image from the outburst decay in 2016
- ➔ SWIFT XRT images from 2016 decay
- ➔ SWIFT XRT images from 2023
- ➔ MAXI and NICER light curves (required for flux history)
- ➔ APEX mm images 12CO, 13CO
- ➔ Other mm surveys (low resolution)

DSH of 4U 1630-47 observed with Chandra on the left and observed with SWIFT on the right



Planck angular resolution: Planck 10’
Bronfman Survey: 7.5’ x 7.5’ resolution
APEX resolution=0.5’

MOLECULAR CLOUDS IN THE LINE OF SIGHT

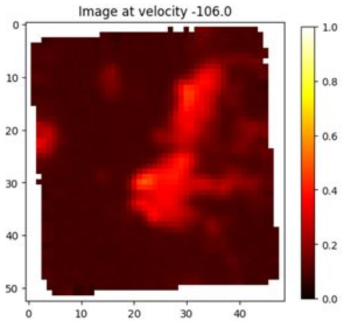
- ➔ There are 15 molecular clouds in the line of sight.
- ➔ Molecular emission of clouds (especially ^{12}CO and ^{13}CO) can be used to determine radial velocities through Doppler shifts.
- ➔ Each radial velocity implies a far and a near distance (based on galactic rotational curve model)
- ➔ So, there is a distance ambiguity for each molecular cloud.

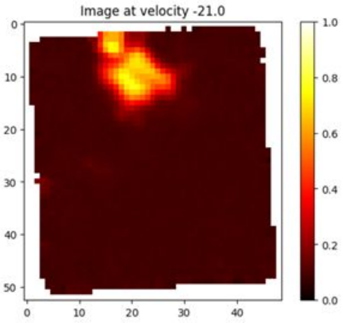
distance ambiguity

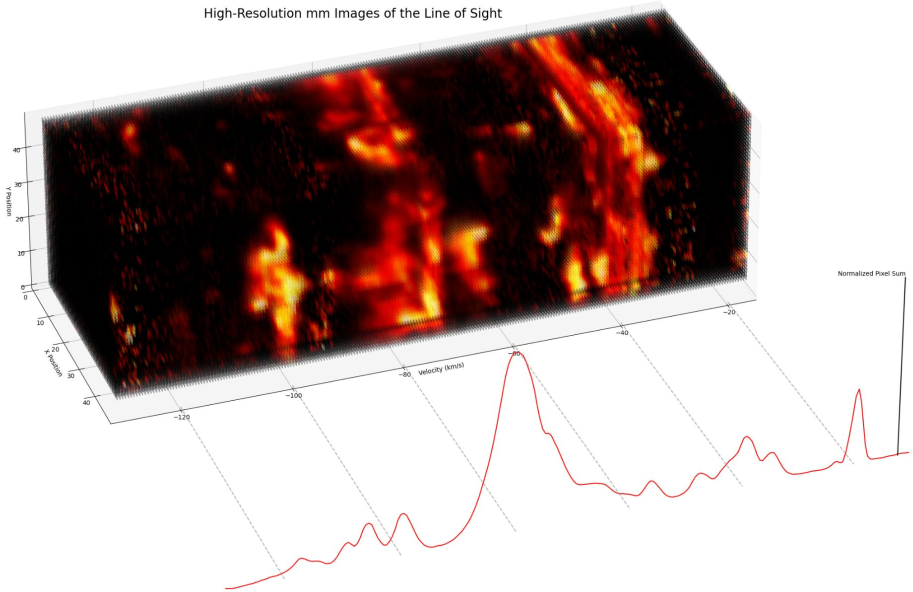
Name	V_{lsr} [a] (km/s)	ΔV(FWHM) (km/s)	Near Dist. (kpc)	Far Dist. (kpc)
MC −111	-111.0	9.0	5.75	9.70
MC −105	-104.9	3.6	5.53	9.92
MC −99	-98.6	4.7	5.32	10.15
MC −79	-79.4	9.6	4.63	10.86
MC −72	-72.3	4.1	4.37	11.14
MC −67	-66.6	10.1	4.14	11.38
MC −56	-56.1	3.9	3.71	11.84
MC −47	-47.0	4.0	3.30	12.27
MC −39	-39.3	6.4	2.92	12.67
MC −33	-33.5	3.2	2.62	13.01
MC −19	-19.1	2.8	1.76	13.94

HIGH-RESOLUTION mm IMAGES

- ➔ From APEX (Atacama Pathfinder EXperiment) telescope
- ➔ 500 mm wave images each with different velocities
- ➔ 2 type of the data (for ^{12}CO and ^{13}CO molecules)







PROJECT OBJECTIVES AND CONSTRAINTS

- **Objective**: Determining the 3D structure of interstellar molecular clouds in the line of sight of 4U 1630-47 from the high-resolution mm wave images.

- **Constraints:**
- There are spatial overlaps of clouds in the data
  - requires splitting some data points (pixels in the images) into several clusters
- Data has a very narrow field of view
  - images only capture a limited portion of the some clouds

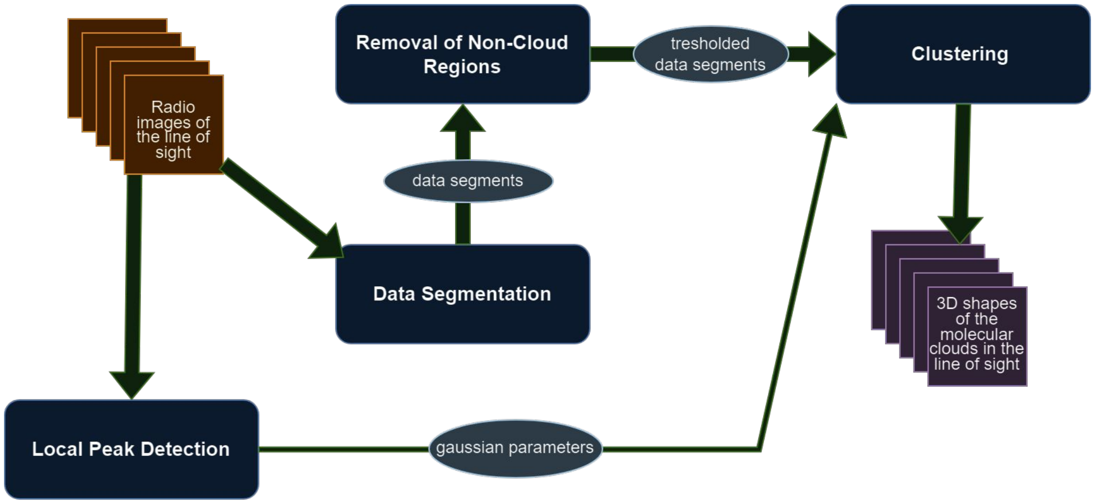
- **Literature:** There are methods in the literature:
  - Clumpfind (Stutzki et al. 1990)
  - Gauss Clumps (Williams et al. 1994)
  - FellWalker (Berry, D. S. 2015)
  - Denograms (Rosolowsky et al. 2008)
  - more...

- But, these algorithms did not create meaningful results for our data because of our constraints.

FINDING THE 3D STRUCTURE OF THE CLOUDS

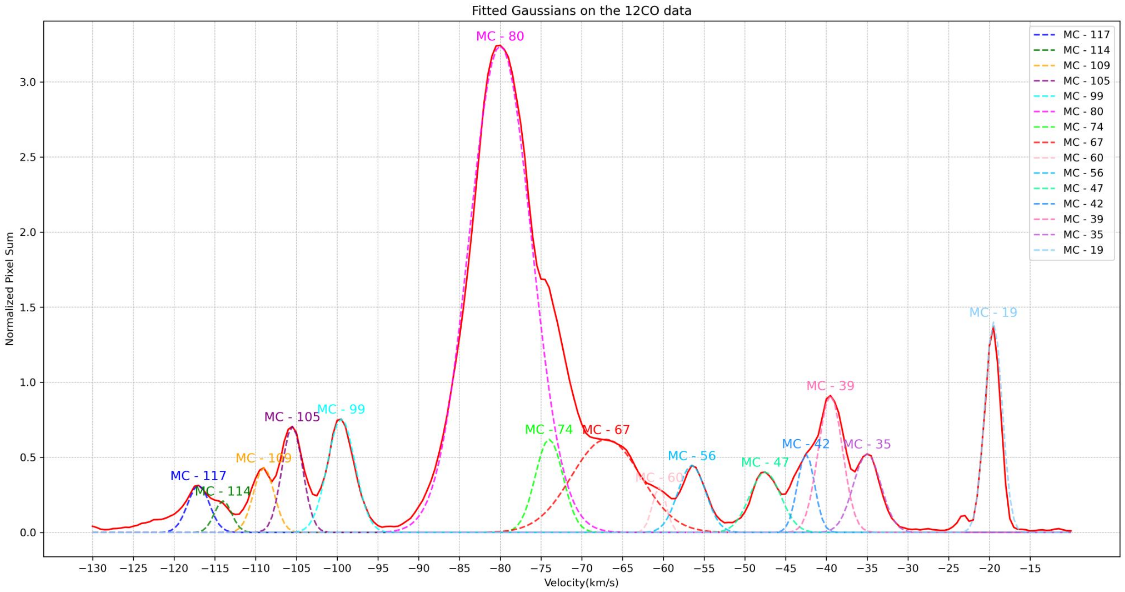
Aim: Constructing 3D shapes of the clouds from 500 APEX mm images

Method: Forming clouds by clustering the data with a ML algorithm

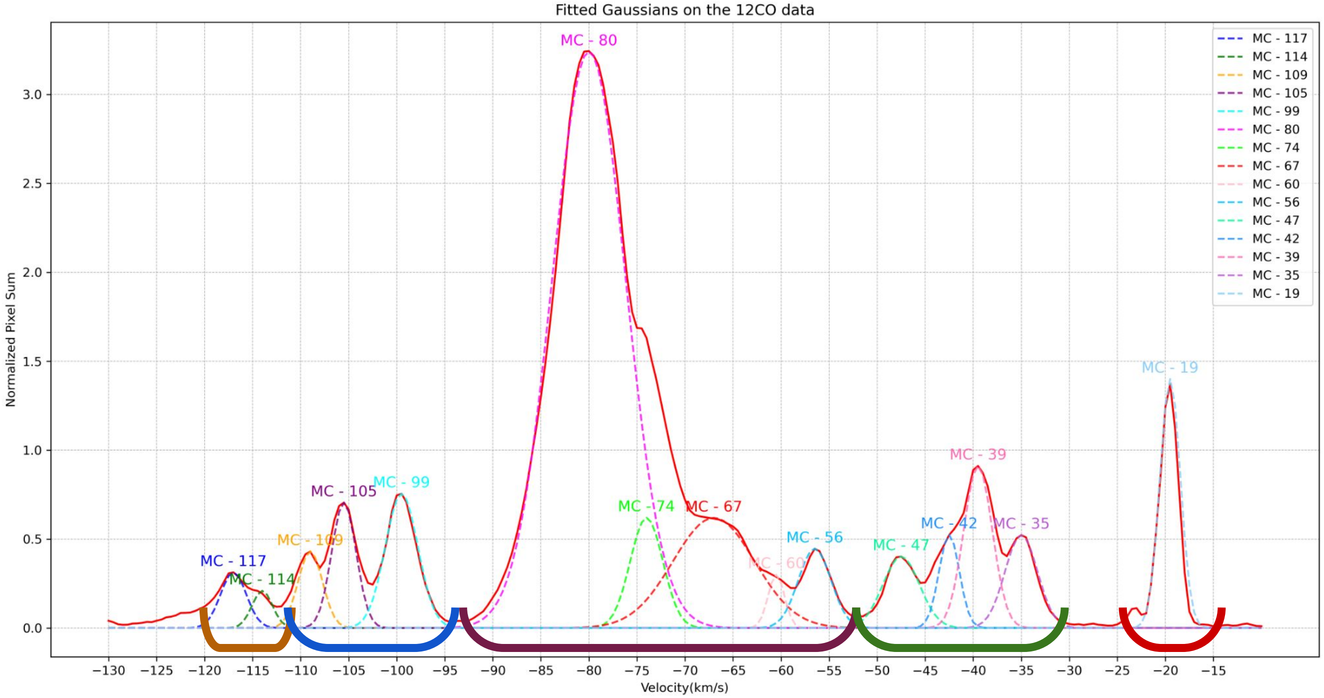


1) LOCAL PEAK DETECTION

- ➔ Gaussian decomposition to identify local density peaks in the spectral data
- ➔ Number of the clouds (clusters) and peak velocity of the clouds (cluster centers) are detected

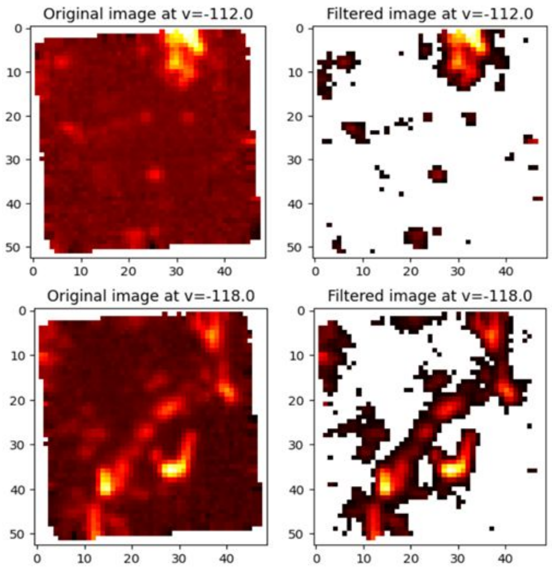


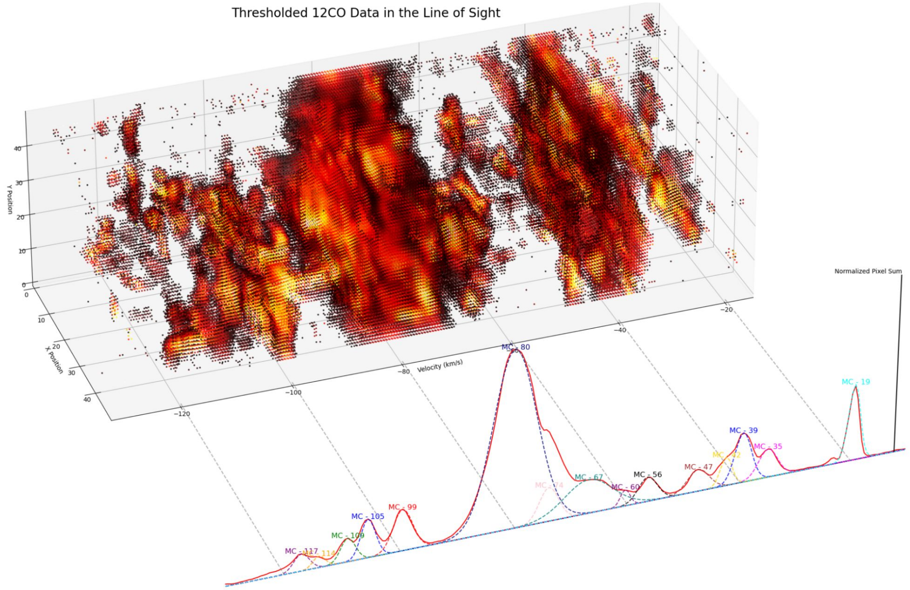
2) DATA SEGMENTATION



3) REMOVAL OF NON-CLOUD REGIONS

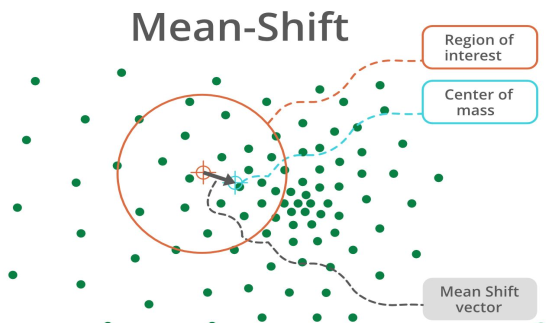
- ➔ Image processing methods are used
- ➔ OTSU thresholding applied for every segment
- ➔ Adaptive filtering for every region



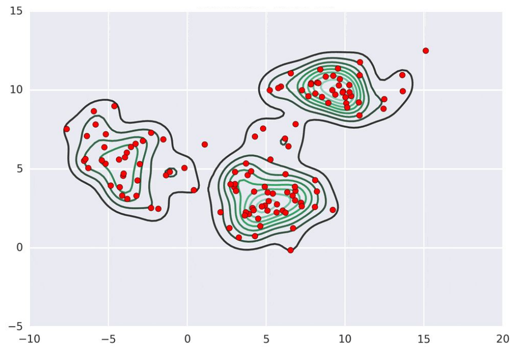


4) CLUSTERING

- ➔ Mean-Shift Clustering Algorithm is used.
- ➔ Mean-Shift algorithm is selected due to its effectiveness in image segmentation.
- ➔ Detected local peak velocities are used as cluster seeds.



- ➔ Mean-Shift algorithm is modified as:
  - ◆ Modified to soft clustering
  - ◆ Datapoints (pixels) are splitted into different clusters with soft assignment results
- ➔ Mean-Shift algorithm is modified in order to overcome the clustering in the overlapping regions.



4) CLUSTERING

➔ Parameters of the Clustering algorithm:

- Start velocity, End velocity
- Seeds

fixed values for the each cloud group determined from Gaussian fits

- Otsu factor

adjusted for every cloud group

- Weighting method
- Alpha, beta (soft clustering thresholds)

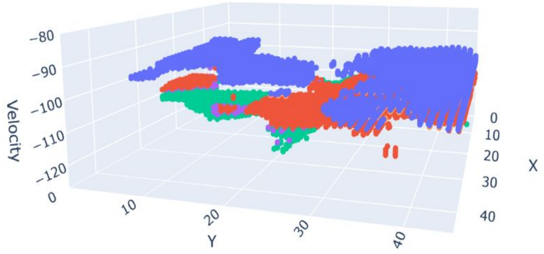
fixed values as gaussian kernel weighting and (0.7,0.2)

- Bandwidths (for every cluster)
- Feature importance factors

Hyperparameters to tune

➔ Hyper-parameter tuning is applied.

➔ Firstly, Silhouette coefficient (cohesion and separation) is used to eliminate most of the parameter combinations.



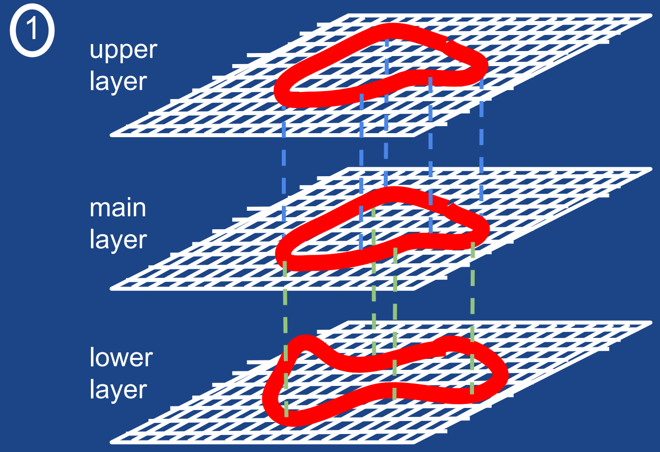
4) CLUSTERING

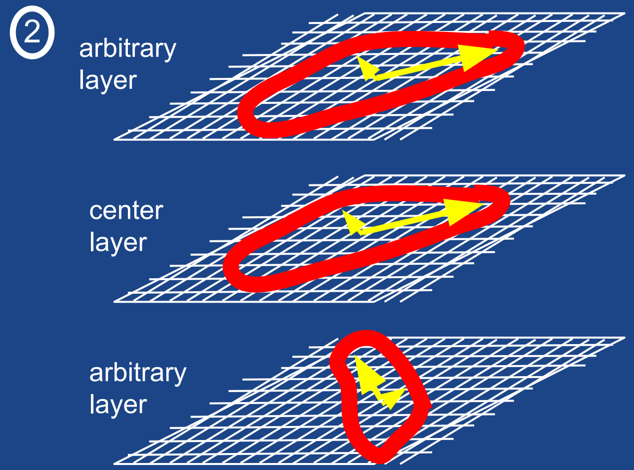
- Then, best parameters are selected with specific error functions.
- 1) Shape Smoothness Error
  - Aim: To ensure that the shapes of the clouds are smooth and there are no abrupt changes in shape.

$$error = \frac{1}{num\ of\ clouds} \sum_{c \in clusters} [\frac{1}{size_c} \sum_{v \in velocity} (2 * image_v - image_{v+0.5} - image_{v-0.5})]$$

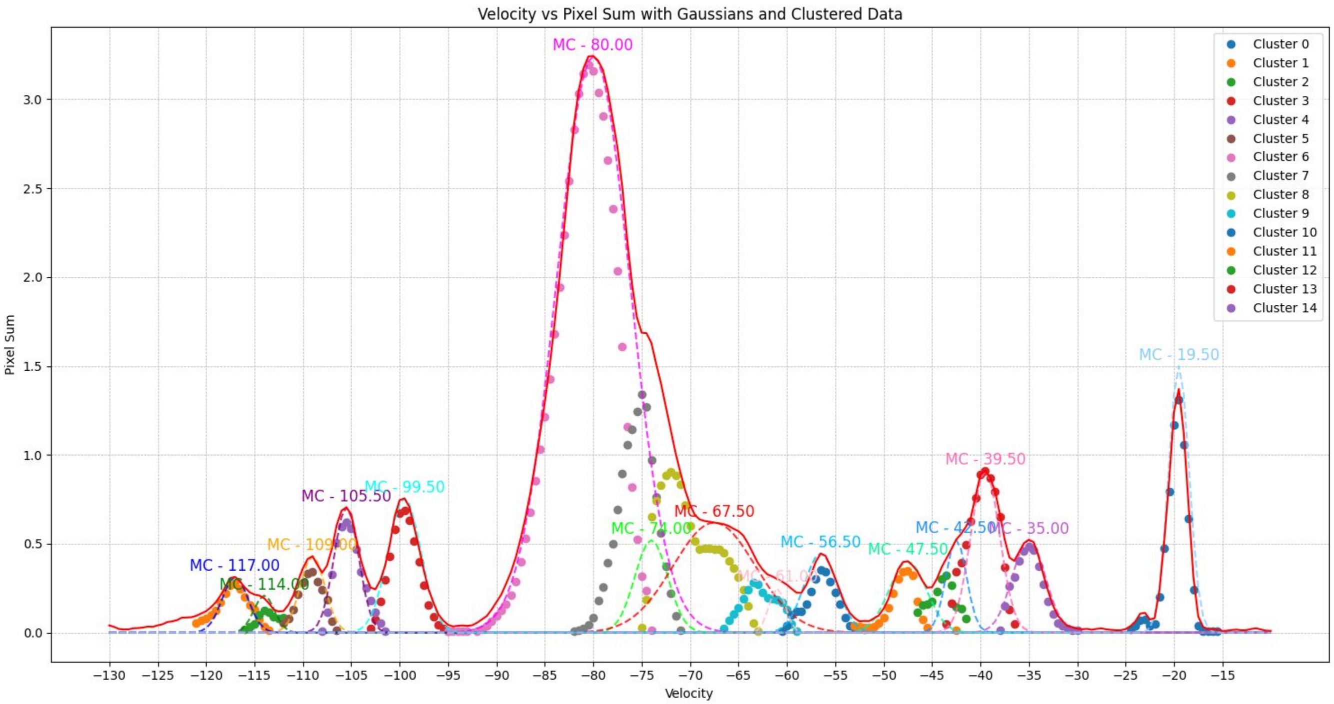
- 2) Eccentricity Consistency Error
  - Aim: To determine how consistent the edges of the clouds are with their central portion.

$$error = \frac{1}{num\ of\ clouds} \sum_{c \in clusters} [\frac{1}{size_c} \sum_{v \in velocity} (center\ eigen\ ratio - eigen\ ratio_v)^2]$$



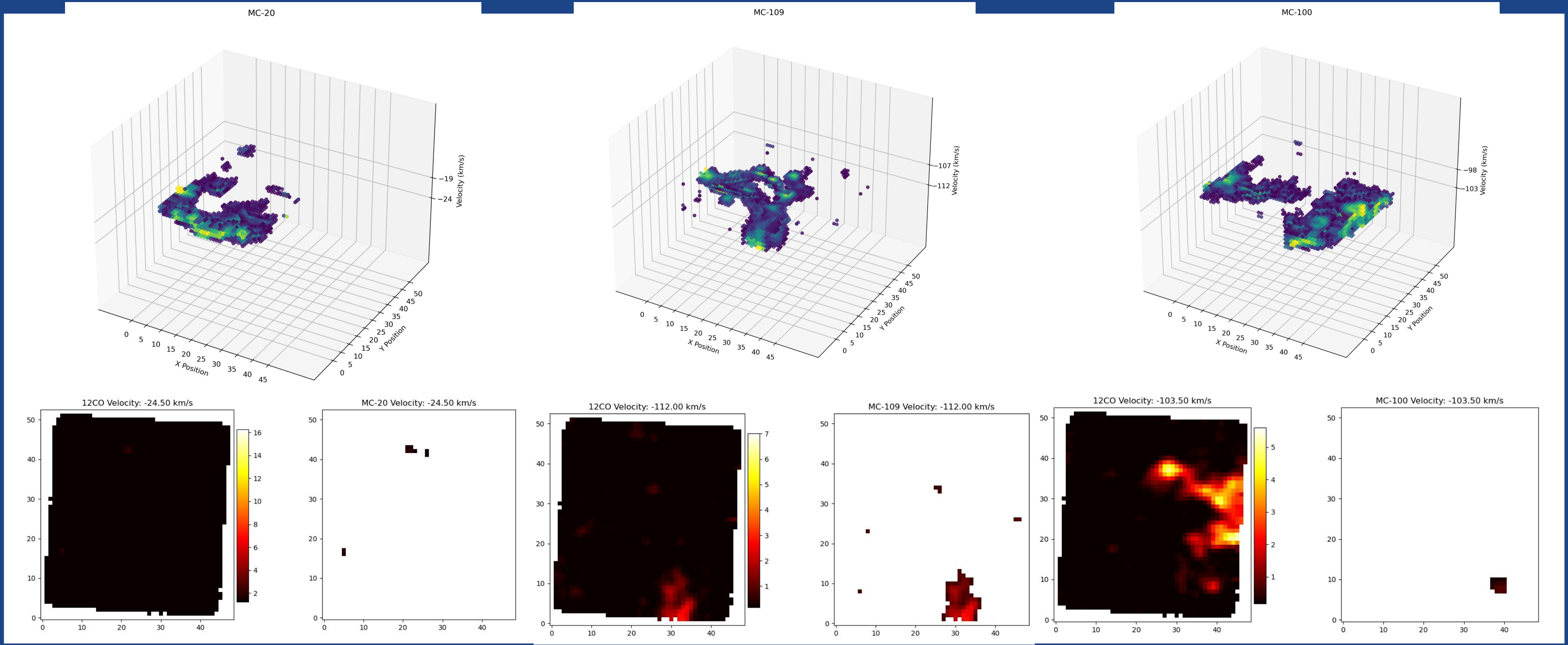


RESULTS

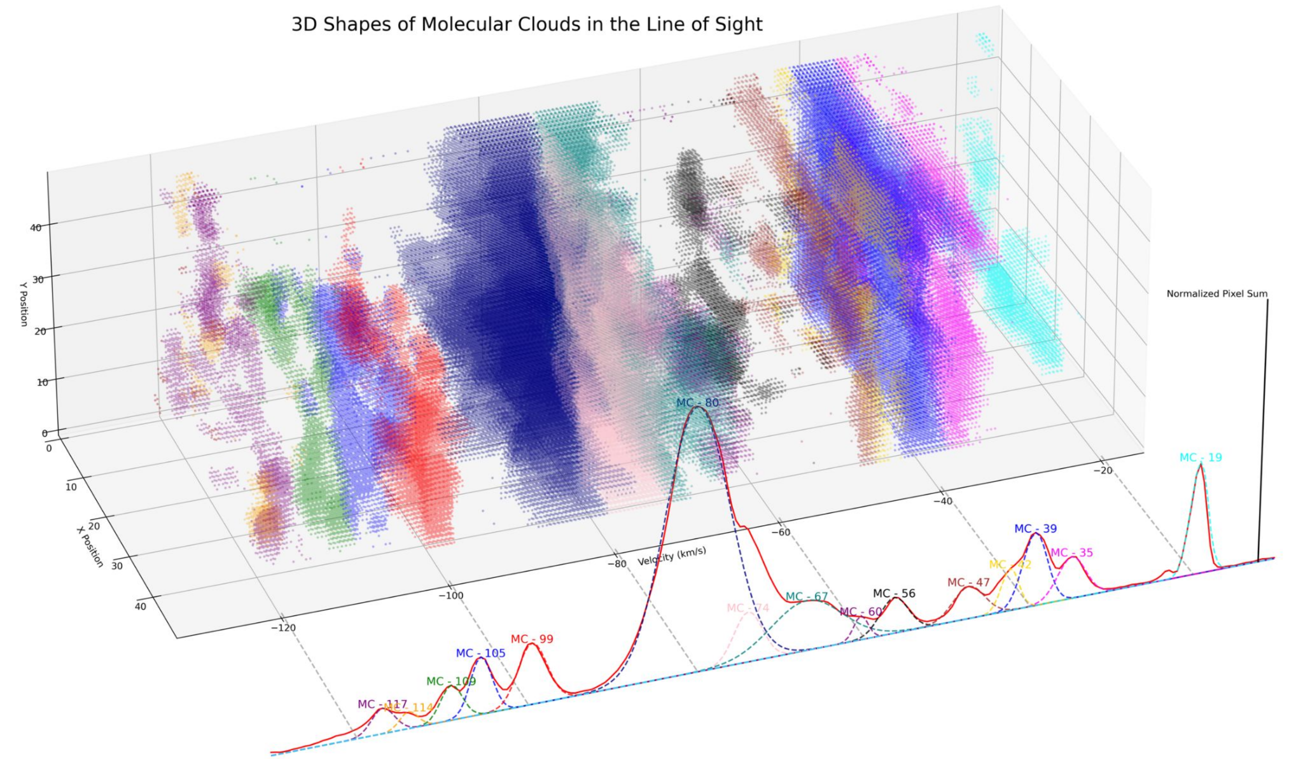


Velocity vs Pixel Sum with Gaussians and Clustered Data
Pixel Sum
Velocity
MC - 80.00
MC - 19.50
MC - 39.50
MC - 105.50
MC - 99.50
MC - 67.50
MC - 74.00
MC - 109.00
MC - 56.50
MC - 47.50
MC - 42.50
MC - 35.00
MC - 117.00
MC - 114.00
Cluster 0
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Cluster 10
Cluster 11
Cluster 12
Cluster 13
Cluster 14

RESULTS


MC-20
MC-109
MC-100
X Position
Y Position
Velocity (km/s)
12CO Velocity: -24.50 km/s
MC-20 Velocity: -24.50 km/s
12CO Velocity: -112.00 km/s
MC-109 Velocity: -112.00 km/s
12CO Velocity: -103.50 km/s
MC-100 Velocity: -103.50 km/s

RESULTS



3D Shapes of Molecular Clouds in the Line of Sight
Y Position
X Position
Velocity (km/s)
Normalized Pixel Sum
MC - 80
MC - 19
MC - 39
MC - 35
MC - 47
MC - 56
MC - 67
MC - 74
MC - 99
MC - 105
MC - 117

DSH SIMULATIONS

(this part is done by Dr. Emrah Kalemci & Ahmed Abdullah Abbasi)

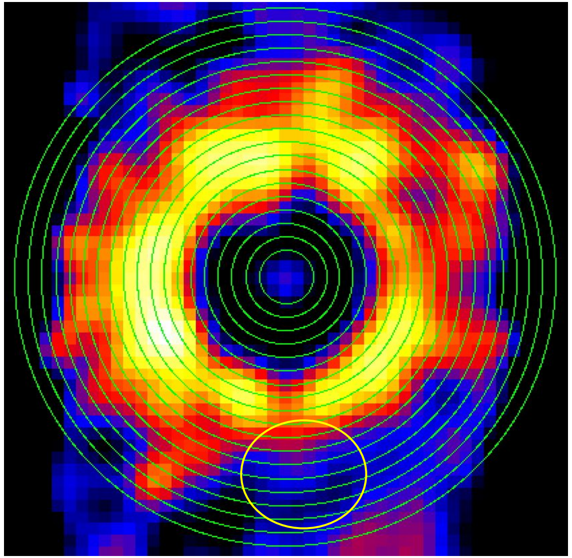
- ➔ Free parameters: distance to the source, whether the clouds are at the near or far distance and an overall normalization.

- ➔ Distance varied 0.1 kpc between 4 - 7 kpc and 10-15 kpc
- ➔ 15 clouds are placed in both near and far distances

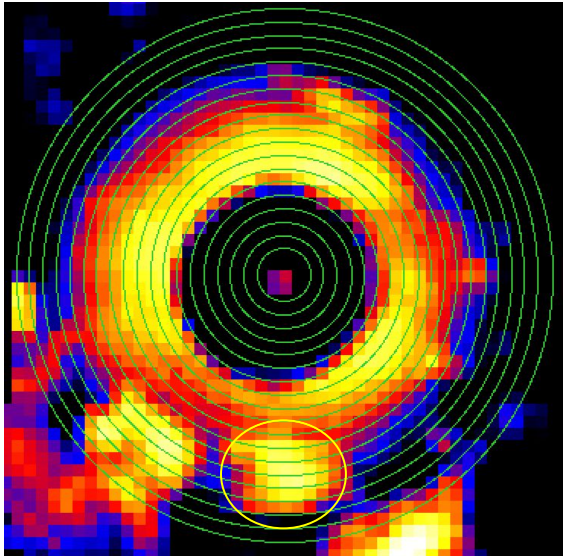
~2.6 million simulated images

- ➔ Chandra image and generated images are binned radially and axially. A $\chi 2$ minimization is applied to the histogram of amplitude of signals in the same radial and axially divided areas to find the best combination of cloud and source distances.

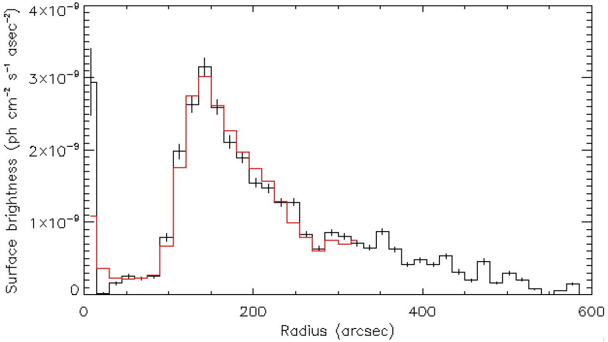
DSH Image fitting, radial (this part is done by Dr. Emrah Kalemci & Ahmed Abdullah Abbasi)



CHANDRA



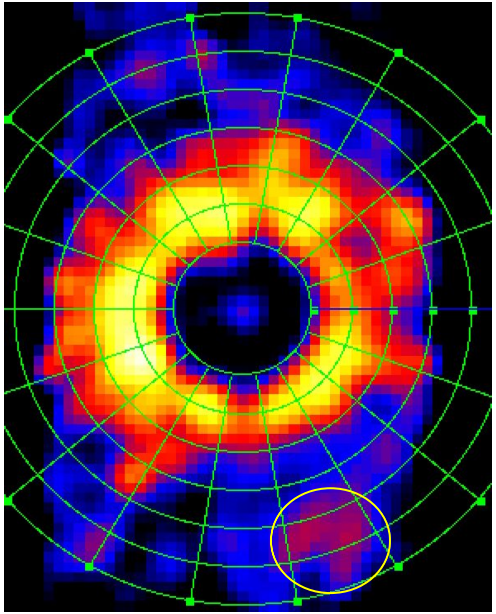
Generated best fit image



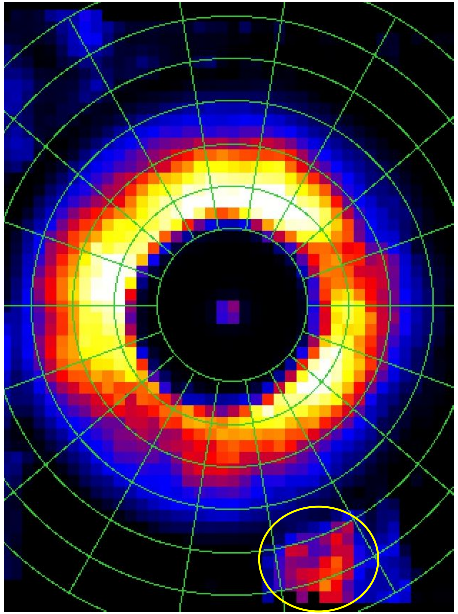
Radial binning best fit:
Distance 13.6 kpc

However some features are not present in Chandra image.

DSH Image fitting, axial
(this part is done by Dr. Emrah Kalemci & Ahmed Abdullah Abbasi)



CHANDRA



Generated best fit image

The image is divided into ~90 wedges with similar signal to noise ratio. Fitting the amplitude distribution again yield 13.6 kpc.

Generating images allow us to compare features and determine near/far distances in an alternative manner. Axial fitting produced better results in terms of picking the features.

FUTURE WORK

- ➔ Complete generating energy dependent images taking into account energy dependent absorption in clouds.
- ➔ Once method is confirmed, we will write new proposals for APEX and other mm wave observatories together with imaging X-ray telescopes (Chandra, XRISM, SWIFT, IXPE) to apply the method for other sources.
- ➔ We will look for relic halos in eROSITA.
- ➔ We will prepare for future observatories, conduct simulations with ATHENA using REFLEX+SIXTE (Work already started)
- ➔ We will investigate if we could use Artificial Neural Networks to do the image fitting to obtain cloud parameters and source distance. (Discussions underway on methodology)

MACHINE LEARNING FOR ASTROPHYSICS
2ND EDITION CATANIA, 8-12 JULY, 2024

Sabancı Üniversitesi

Thank you for your attention.

ANY QUESTIONS?

Efe Öztaban
Sabancı University

efeoztaban@sabanciuniv.edu

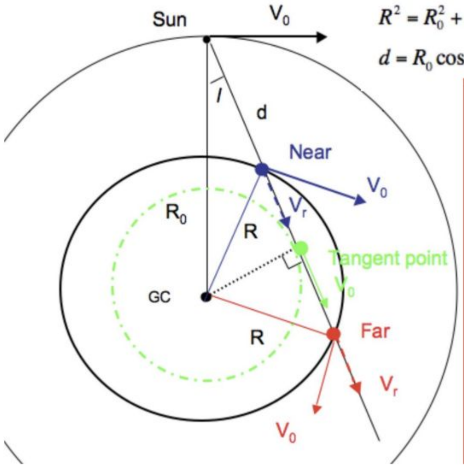
Galactic Rotation Curve and Distance Ambiguity

DISTANCES TO MOLECULAR CLOUDS

$$V_{radial} = V_0 \sin l \, (\frac{R_0}{R} - 1)$$

$$R^2 = R_0^2 + d^2 - 2Rd\cos l$$

$$d = R_0 \cos l \pm \sqrt{R^2 - R_0^2 \sin^2 l} = R_0 \cos l \pm \sqrt{R^2 - R_{\min}^2}$$



- Based on a rotation curve model
- Large errors where non-circular motions (spiral arms….)
- Unique solution for the galactocentric radius
- Two solutions for the distance (with respect to the tangent point)
  - "Near"
  - "Far"

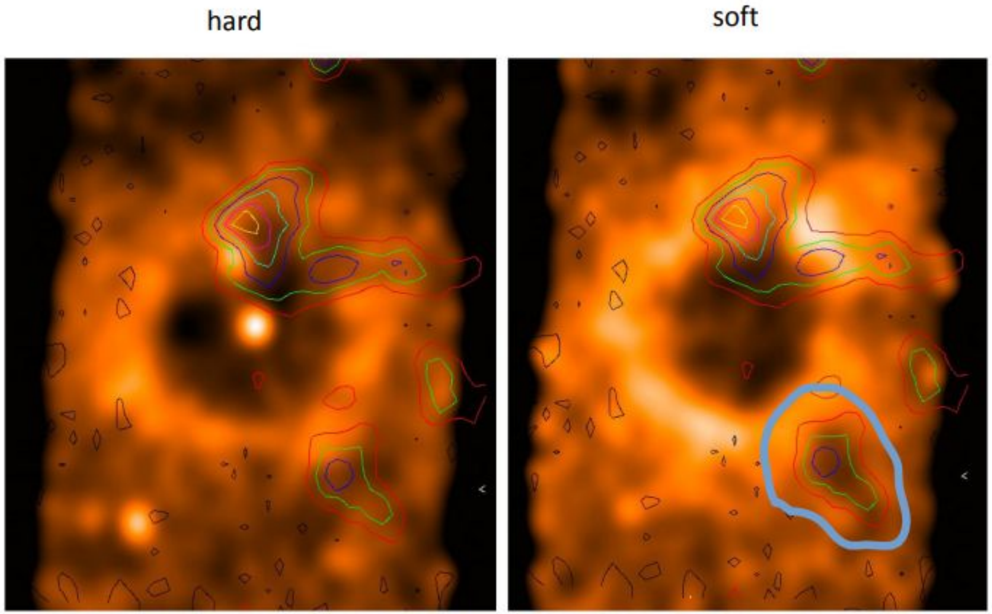
⇒ Kinematic Distance Ambiguity (KDA)

distance ambiguity

Name	V_{lsr} [a] (km/s)	ΔV(FWHM) (km/s)	Near Dist. (kpc)	Far Dist. (kpc)
MC −111	-111.0	9.0	5.75	9.70
MC −105	-104.9	3.6	5.53	9.92
MC −99	-98.6	4.7	5.32	10.15
MC −79	-79.4	9.6	4.63	10.86
MC −72	-72.3	4.1	4.37	11.14
MC −67	-66.6	10.1	4.14	11.38
MC −56	-56.1	3.9	3.71	11.84
MC −47	-47.0	4.0	3.30	12.27
MC −39	-39.3	6.4	2.92	12.67
MC −33	-33.5	3.2	2.62	13.01
MC −19	-19.1	2.8	1.76	13.94

Resolving Distance Ambiguity with Multi-Energy Analysis

(this part is done by Dr. Thomas Stanke and Enno Nussbaum)



- ➔ “black spot” in soft X-ray image
- ➔ not visible in hard X-ray image
- ➔ absorbing cloud at -105 km/s (d= 6.2/8.8 kpc)
- ➔ scattering cloud must be at far distance
- ➔ This resolves the near/far distance problem in ApJ paper.