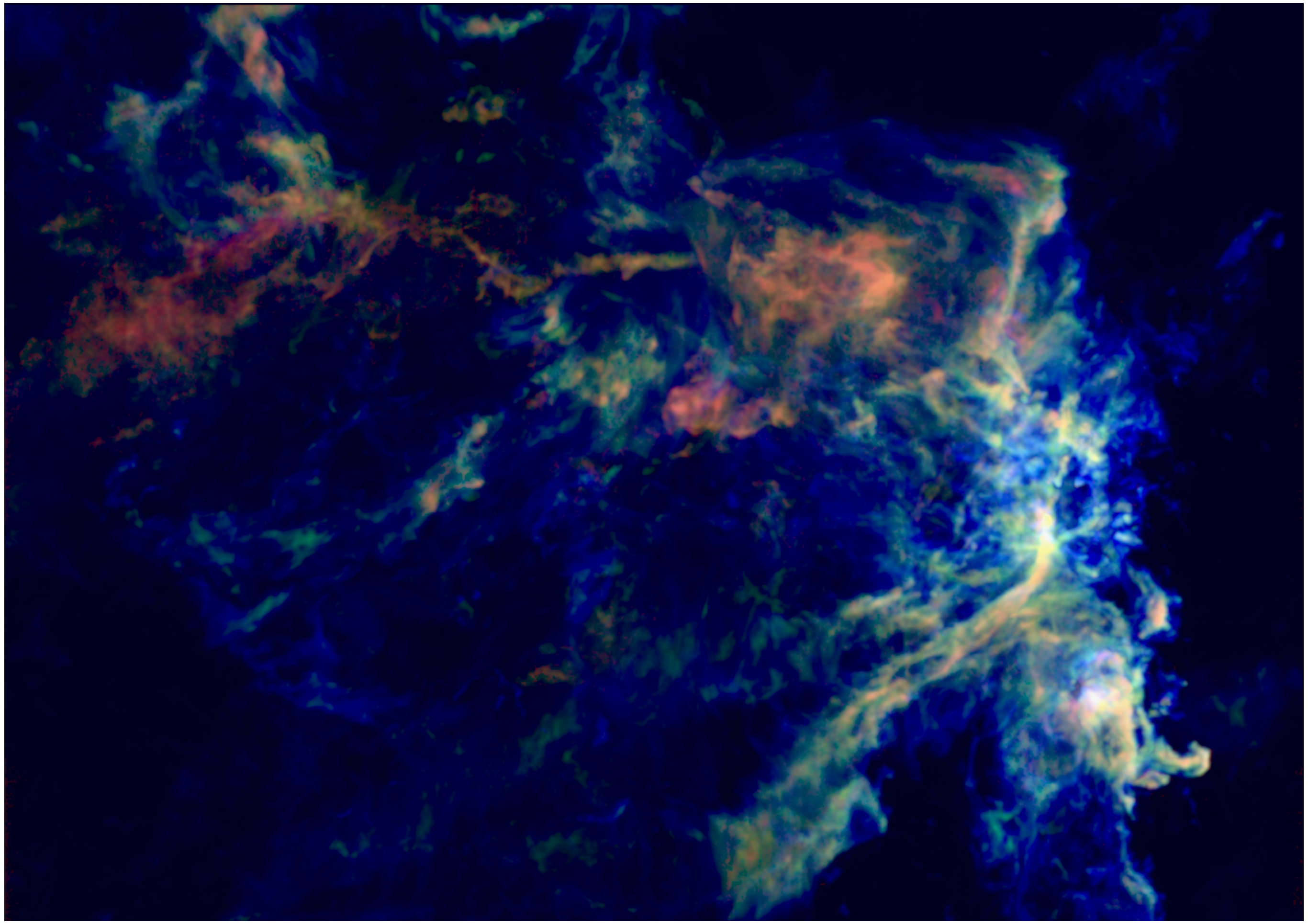
Astrophysics meet data science
for the study of Giant Molecular Clouds



$C^{18}O$ J=1–0, ^{13}CO J=1–0, ^{12}CO J=1–0

J. PETY

IRAM & Obs. de Paris/LERMA

on behalf of
I. BESLIC, S. COUDE, L. EINIG,
H. MAZUREK, J. ORKISZ, P. PALUD,
M. SANTA-MARIA, L. SEGAL,
A. ZAKARDJIAN, S. BARDEAU,
E. BRON, P. CHAINAIS,
J. CHANUSSOT, M. GAUDEL,
M. GERIN, J. R. GOICOECHEA,
P. GRATIER, V. GUZMAN,
A. HUGHES, D. LANGUIGNON,
J. LE BOURLOT, F. LE PETIT,
F. LEVRIER, D. LIS, H. LISZT,
K. OBERG, N. PERETTO,
A. ROUEFF, E. ROUEFF, A. SIEVERS,
V. de SOUZA MAGALHAES,
P. A. THOUVENIN, P. TREMBLIN,
and M. VONO

Marseille - June 6th, 2024

ORION-B (co-PIs: J. Pety & M. Gerin)

Science goals

- Elucidate link between star formation and the structure of host parent (dynamics, physics, & chemistry).
- Build an unbiased dataset, with tracers of all of the components of the host parent.

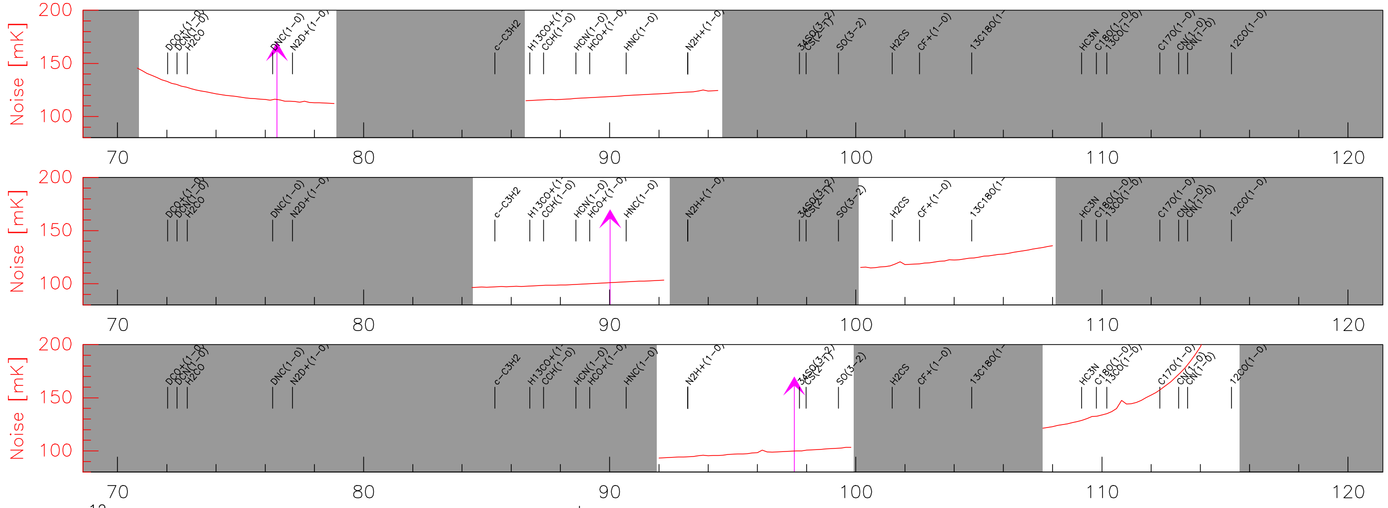
Field of view & Spatial resolution

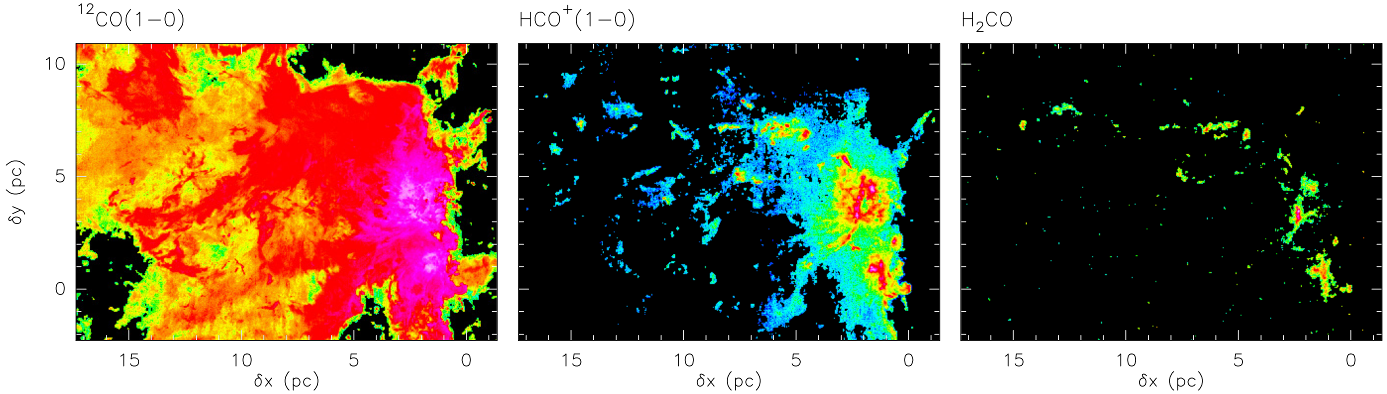
5 degree^2 or 18×13 pc
@ 26'' or ~ 50 mpc
$\Rightarrow$ One image: $\sim 10^6$ pixels.

Bandwidth & Spectral resolution

40 GHz at 200 kHz resolution
$\Rightarrow$ 200 000 images, *i.e.*, at 24 images per seconds, it makes a movie of 2h45!

A sea of noise Median noise level $0.1 - 0.5$ K $\Rightarrow$ Clear signal detected in $\sim 1\,200$ channels, or 0.5% of the data (a video of about 50 seconds).

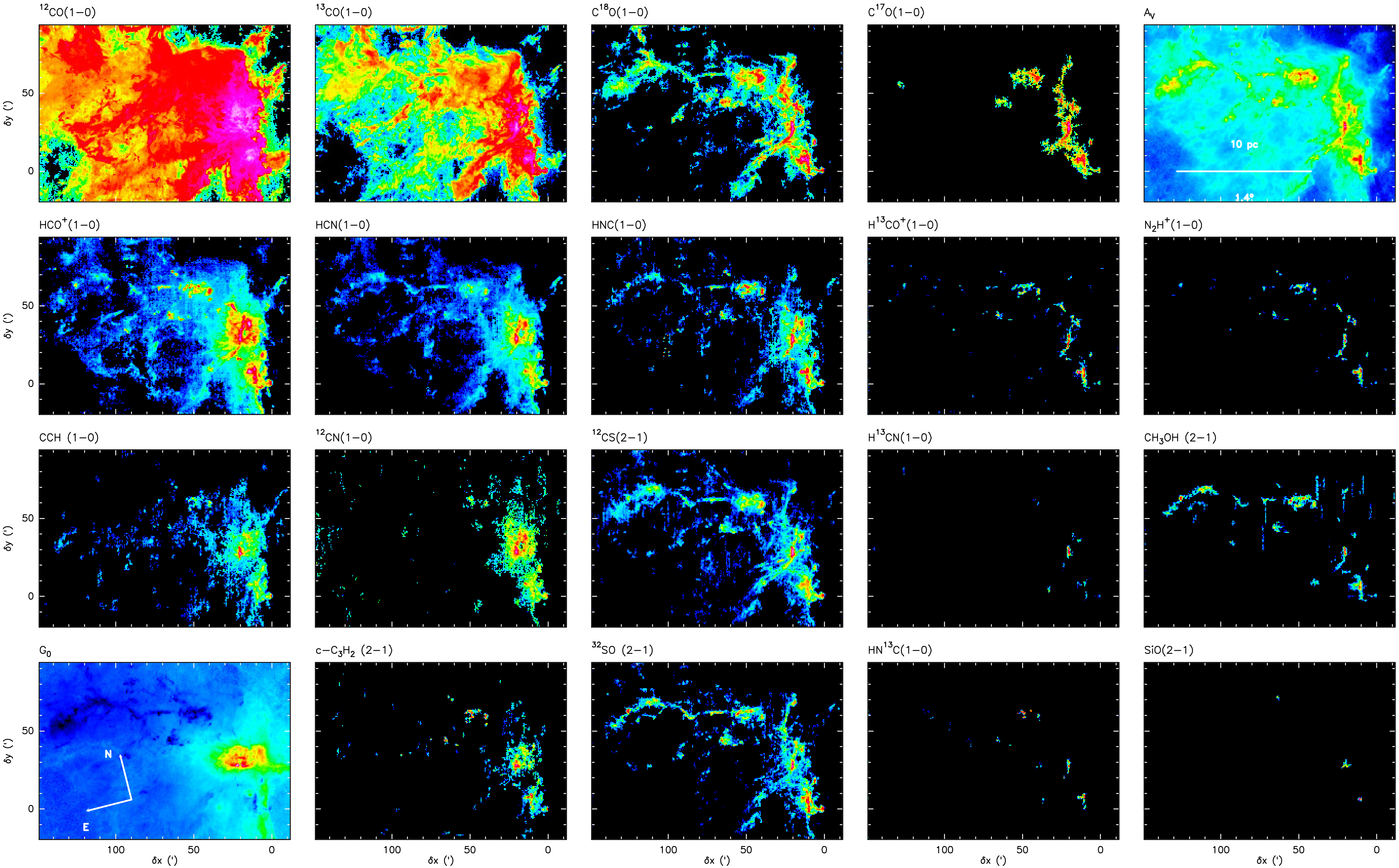




Modern statistical (Machine Learning) methods allow one

- to exploration of high dimensional data ;
- and thus to shed light on star formation that is a statistical question.

How to best combine these line emissions to improve our knowledge of GMCs and star formation?



Tracing the amount of gas along the line of sight

Direct observation of cold H_2 is impossible $\Rightarrow$ Use detectable emission from other tracers.

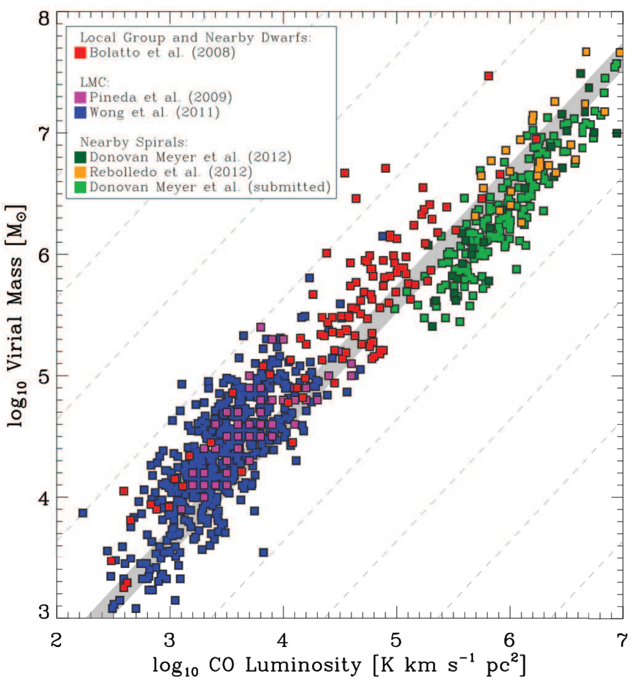
Tracer #1: FIR thermal emission of dust grains $\Rightarrow$ Requires space telescope (e.g., Herschel).

Tracer #2: Molecular tracers like CO and a conversion factor (X_{CO}).

- **Pro** Observable from the ground $\Rightarrow$ high angular observation.
- **Con** Relationship between intensity and the amount of gas saturates.

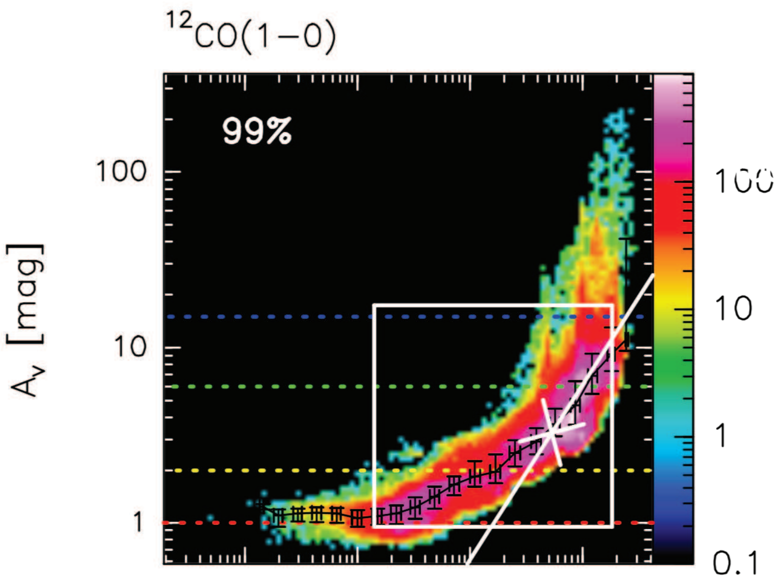
$\Rightarrow$ We can do better with multi-line observations and machine learning (eg, random forests.)

Unresolved GMCs



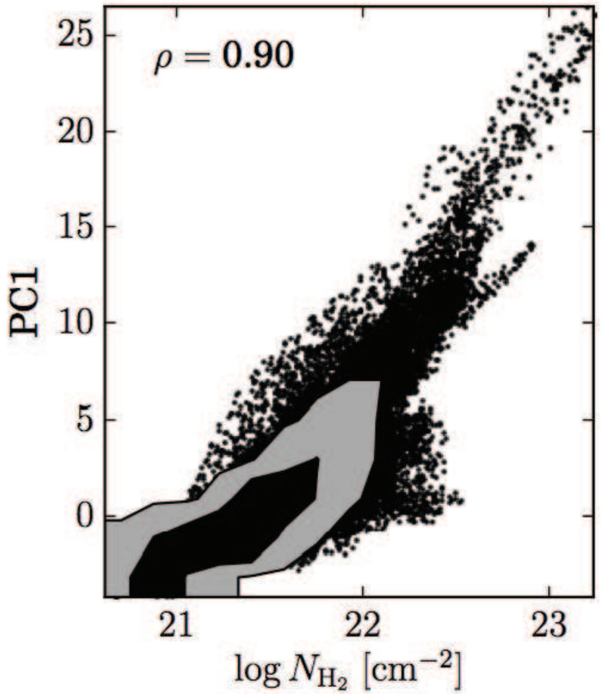
Bolatto et al. 2013

Orion (resolved)



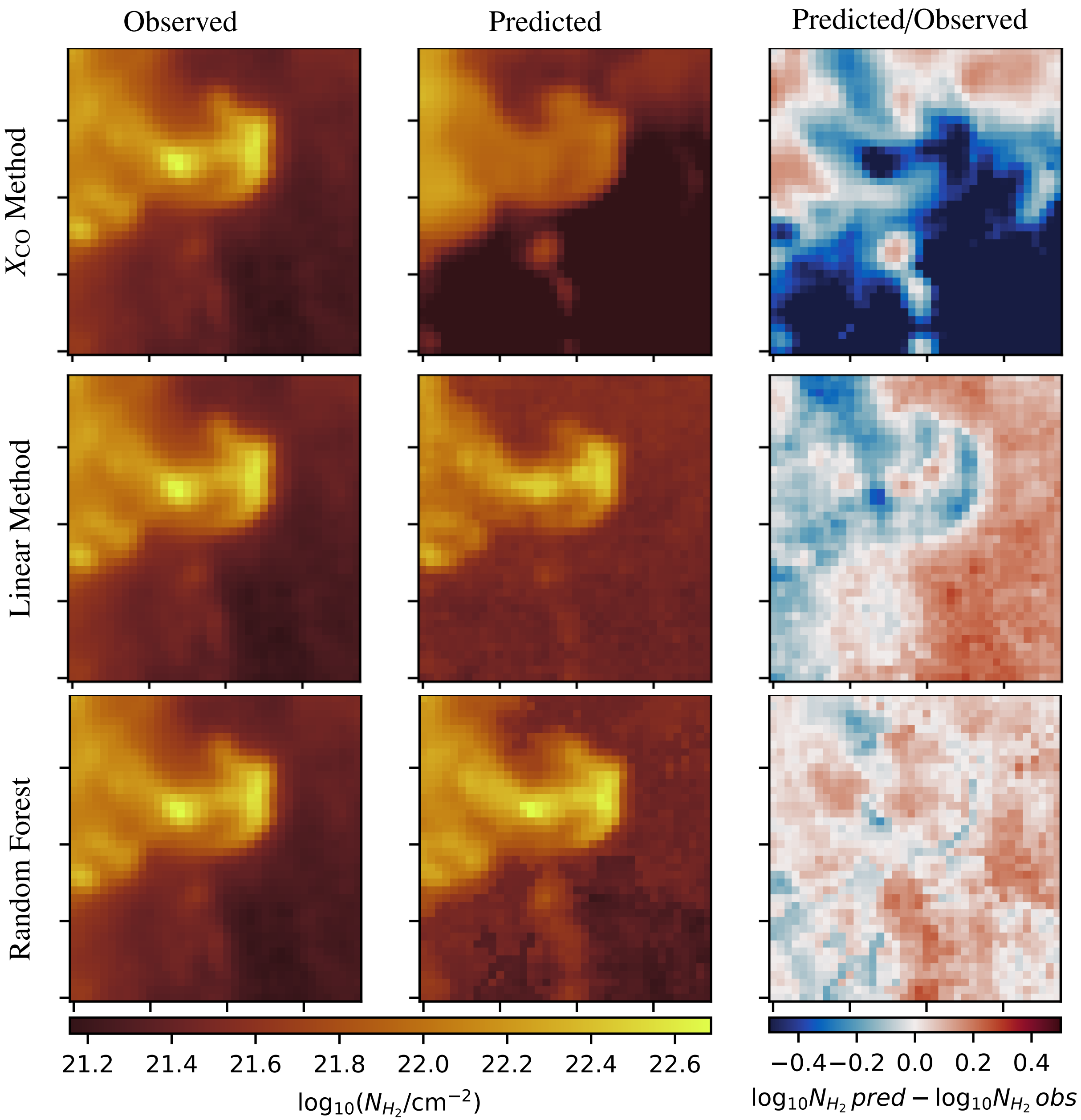
Pety et al. 2017

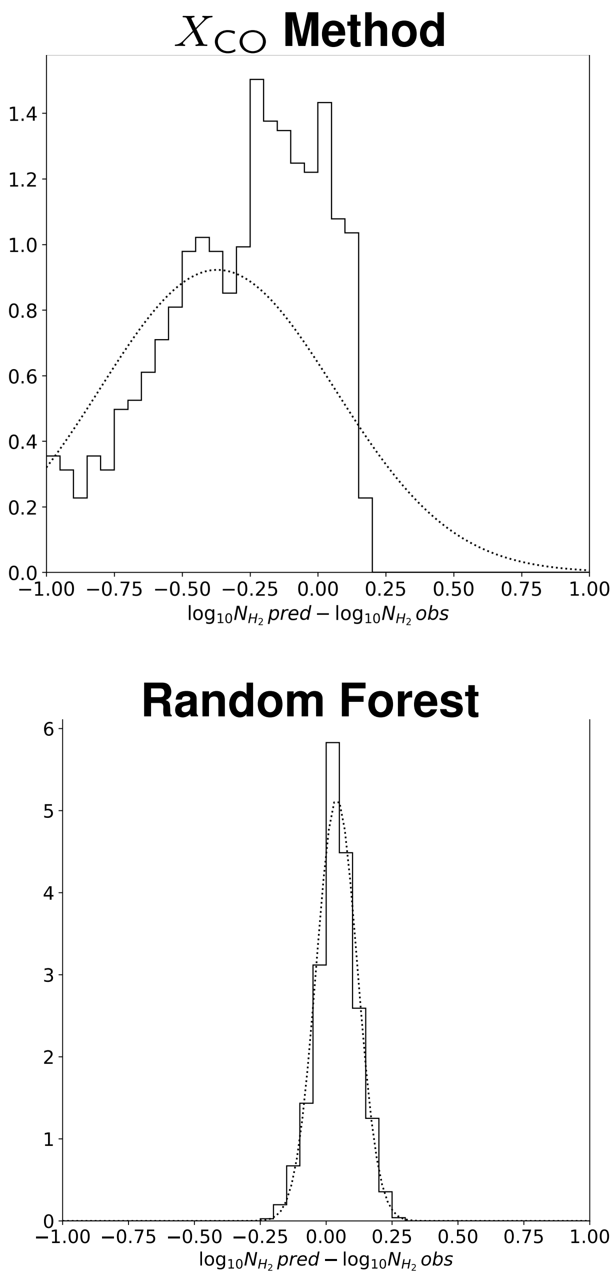
PCA



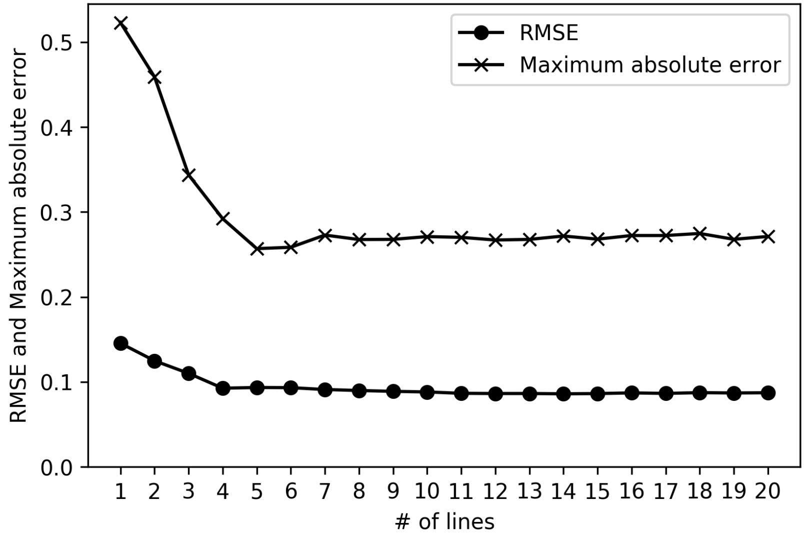
Gratier et al. 2017

Quality of the $N(\mathrm{H_2})$ inference as a function of the method (Gratier et al. 2021)

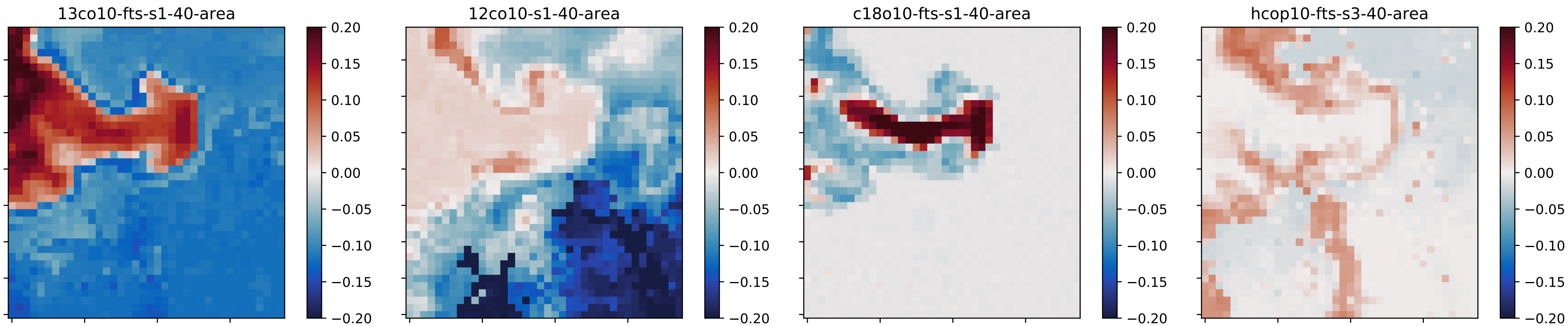




Which tracers have the largest contribution? (Gratier et al. 2021)

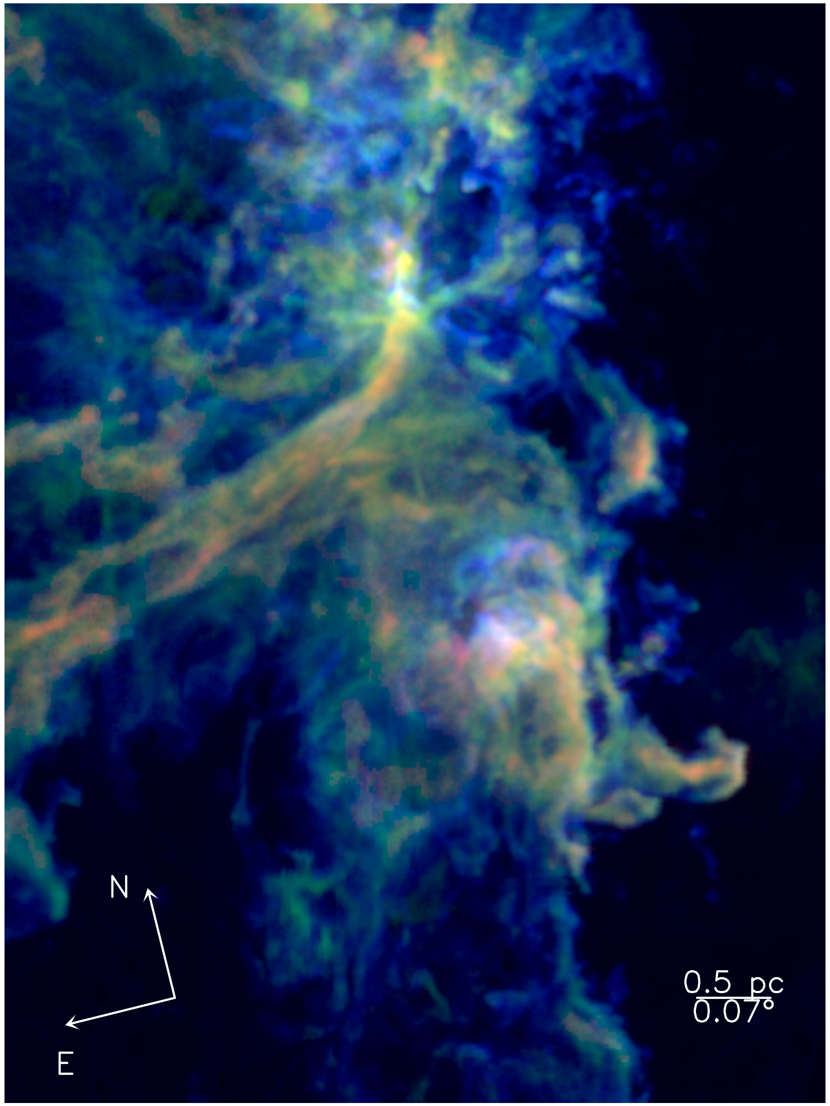


Four key species ^{12}CO, ^{13}CO, C^{18}O, HCO^{+} (1-0) lines.
Decision path for each pixel ⇒ Contribution of each observable to the column density.
⇒ Insight into the physics and chemistry.



^{13}CO (1-0) ^{12}CO (1-0) C^{18}O (1-0) HCO^{+} (1-0)

Why is it possible to constrain the H_2 column density mostly based on CO isotopologues? (Roueff et al. 2021)



$C^{18}O$ J=1–0, ^{13}CO J=1–0, ^{12}CO J=1–0

Simple reasoning CO isotopologues have

1. similar chemistries;
2. similar radiative transfer properties.

$\Rightarrow$ Their line ratios should be simple probes of the elemental abundance of carbon and oxygen to monitor the progress of the stellar nucleosynthesis in nearby galaxies.

The image suggests this reasonning is incorrect Why?

$\Rightarrow$ Need to invert the observations.

Difficulties

- Limited signal-to-noise ratio.
- Huge number of pixels.

$\Rightarrow$ Cramer-Rao Bound: a robust statistical tool to tell us which amount of information can actually be extracted.

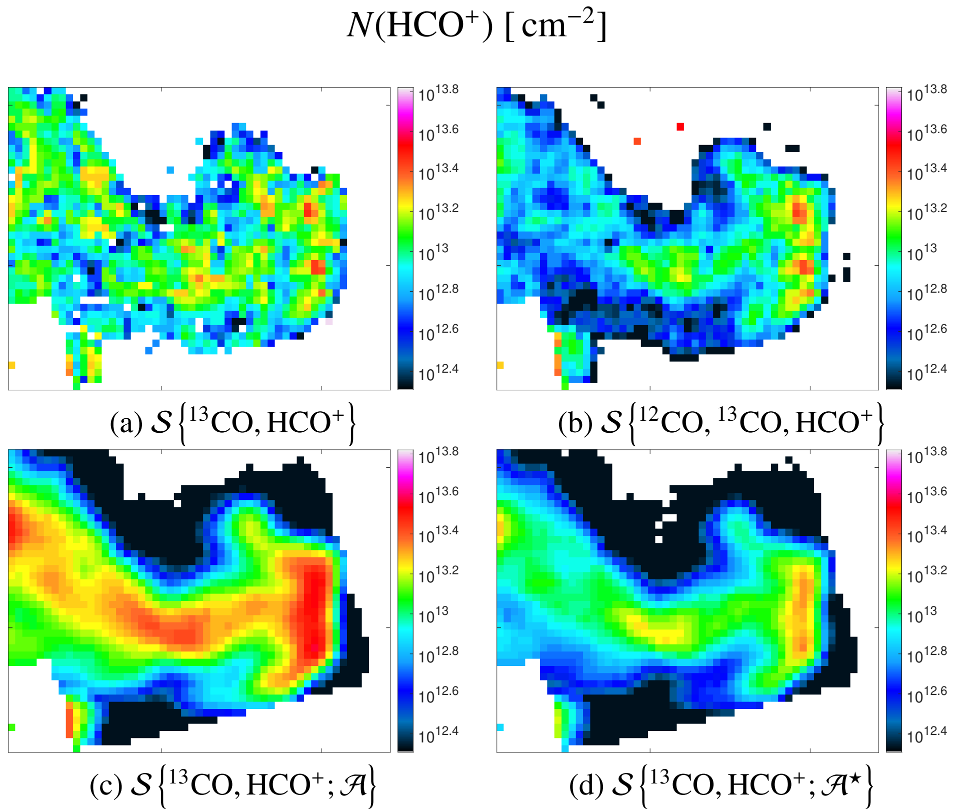
$\Rightarrow$ Automated quality assessment.

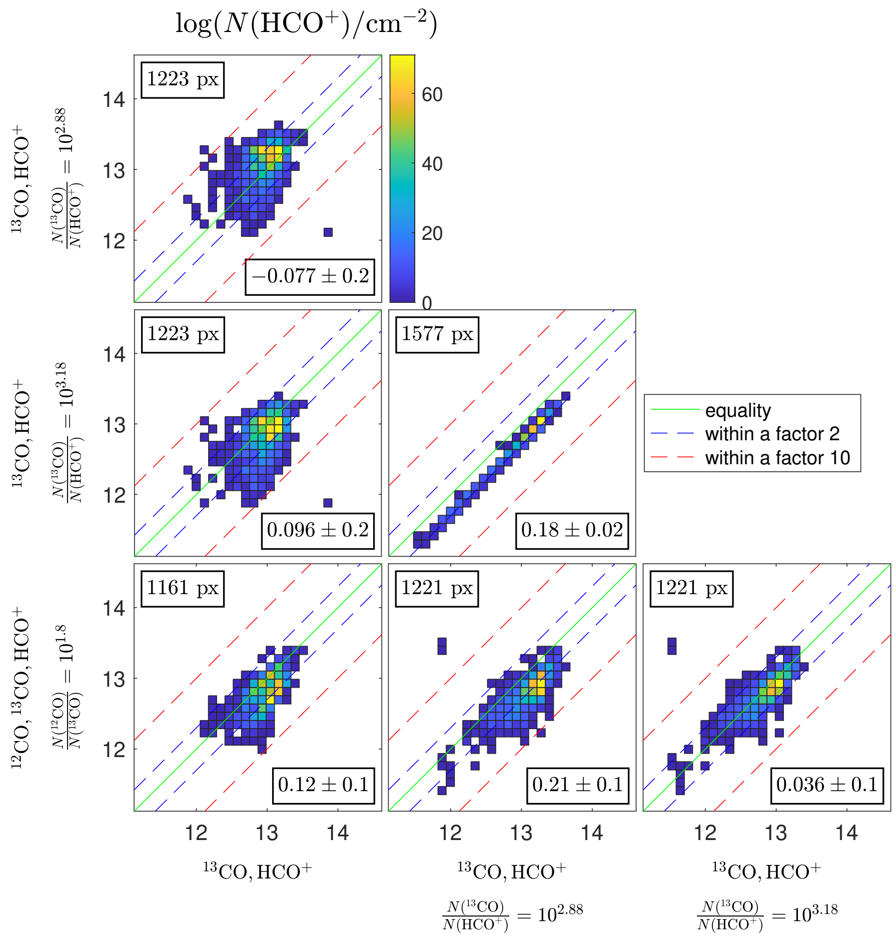
Results Molecular chemistry is a key player!

- Fractionation of C^+: $^{12}CO + {}^{13}C^+ \rightarrow {}^{13}CO + {}^{12}C^+$
- Selective photo-dissociation destroys more $C^{18}O$ than ^{13}CO and ^{12}CO.

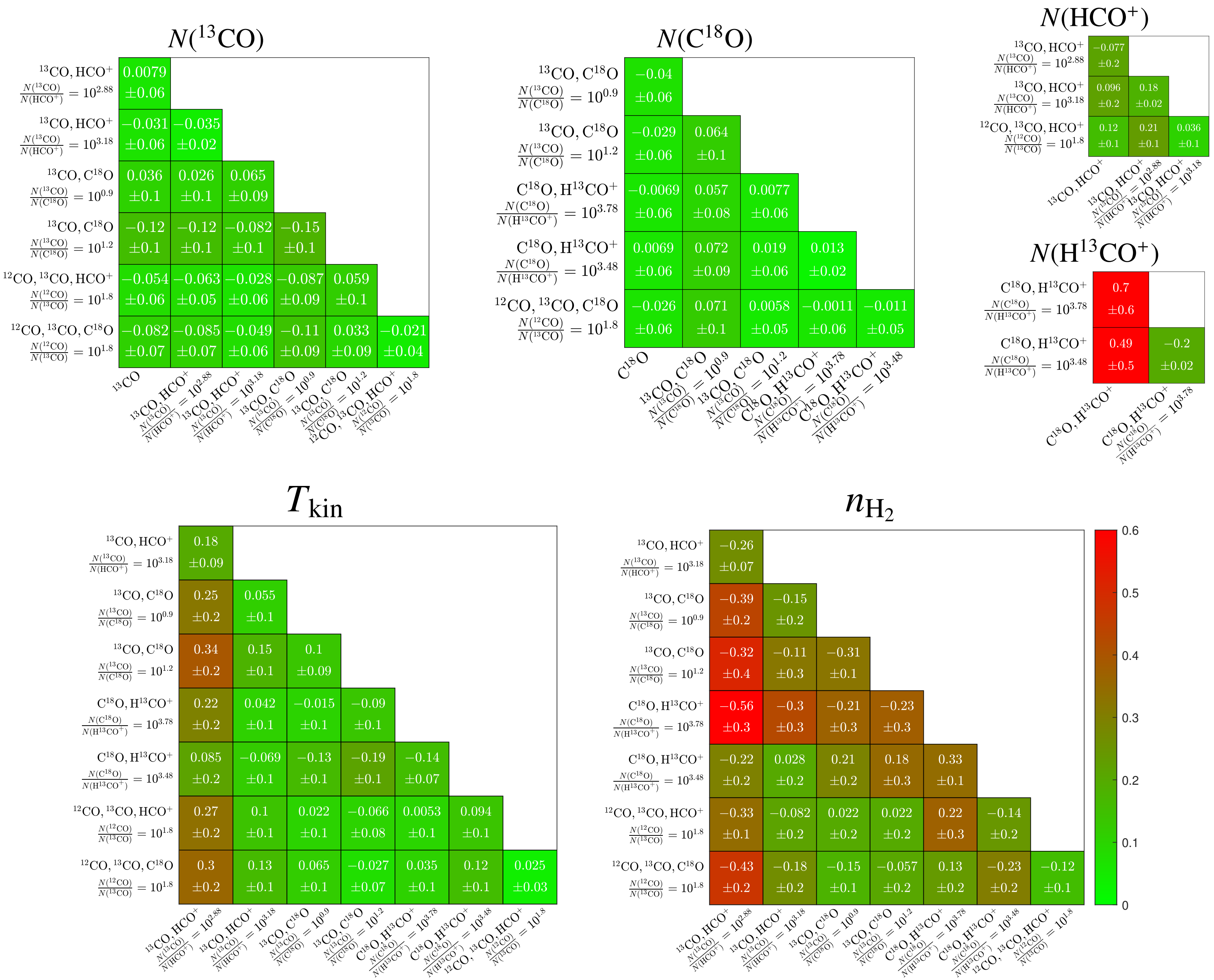
Bias vs variance when fitting multi-species molecular lines with a non-LTE radiative transfer model (Roueff et al. 2024)

Same data J=1–0 lines of CO isotopologues and HCO^+.
Same model RADEX.
Different assumptions Combination of lines or relative abundances.
$\Rightarrow$ **Varied results.**



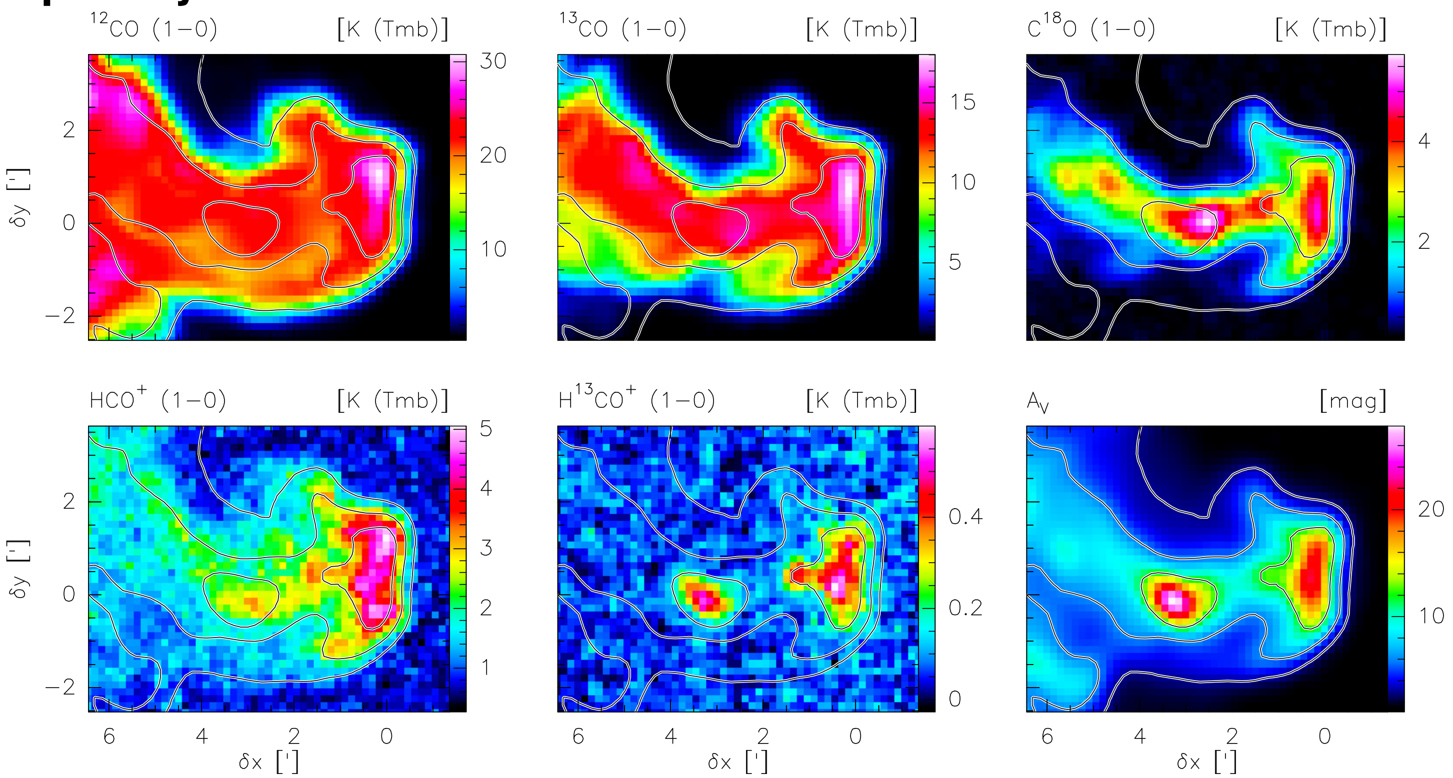


Accuracy(Column Density) > Accuracy(Kinetic Temperature) > Accuracy(Volume Density)

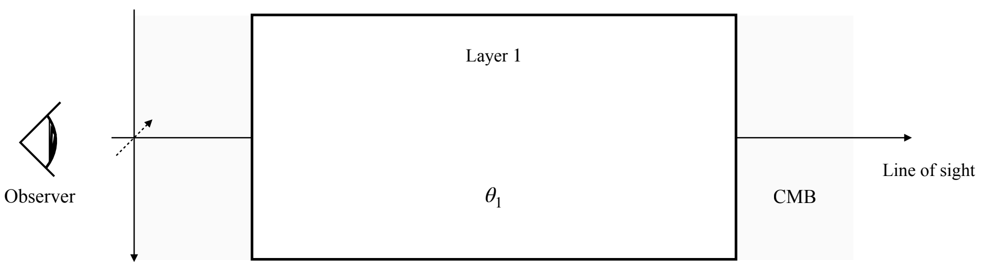


On the need of a sandwich model (Ségal et al., in prep.)

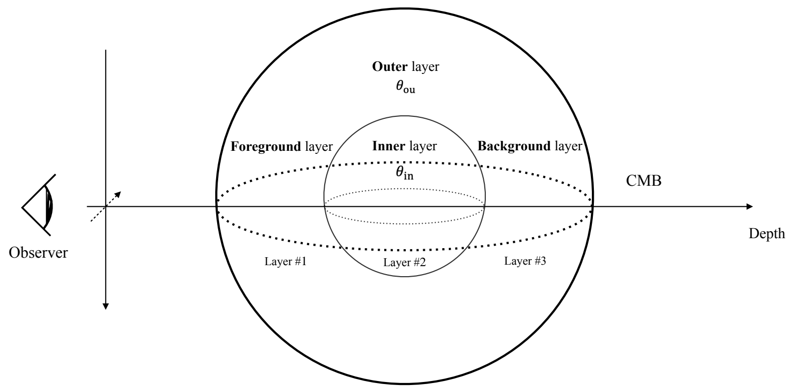
Optically thick lines are more sensitive to the outer translucent gas.
Optically thin lines are easier to detect in the inner dense cores.



From a mono-layer model



To a sandwich model



Inverting PDR models to derive physical parameters
Challenges and solutions (Palud et al., 2023a)

Sophisticated PDR models

- Many physical and chemical processes must be modeled.
- Micro-physical parameters (eg, chemical rates) may still be uncertain.

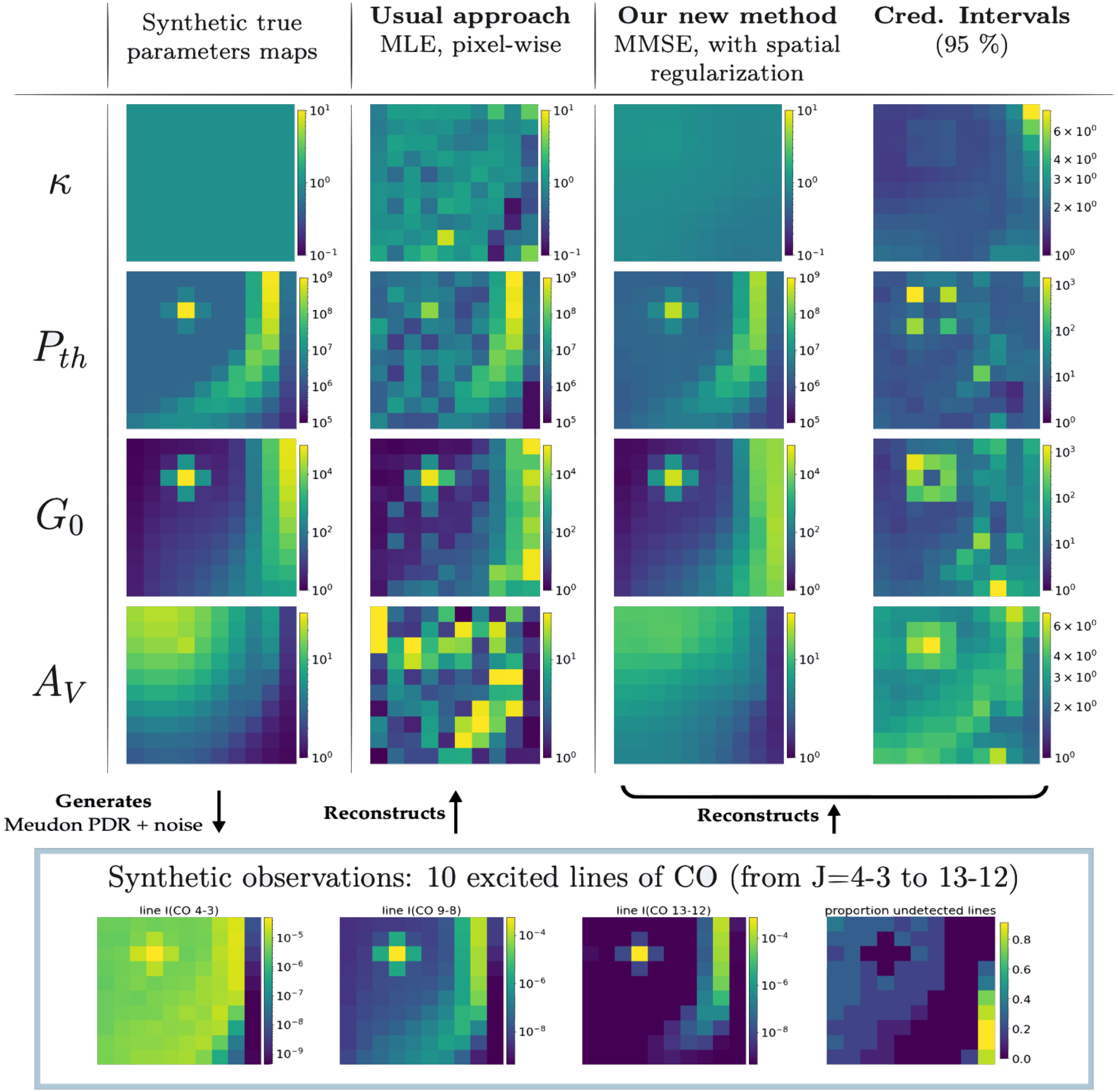
Large dynamical range

- Calibration uncertainty (often neglected multiplicative “noise”) is important at high Signal-to-Noise Ratio.
- Censored information at low Signal-to-Noise ratio.

Hidden variables

- Geometry.
- Observations of only minor tracers of the gas.

Solution New, fast Bayesian framework that takes into account additional and multiplicative noise, censored information, and spatial regularization $\Rightarrow$ Derivation of parameters and associated credibility intervals.



Fast, accurate, and robust, neural network-based, emulation of complex physical models (Palud, Einig, et al. 2023b)

State-of-the-art ISM models

- Many physical and chemical processes.
- Too slow for Bayesian inference.

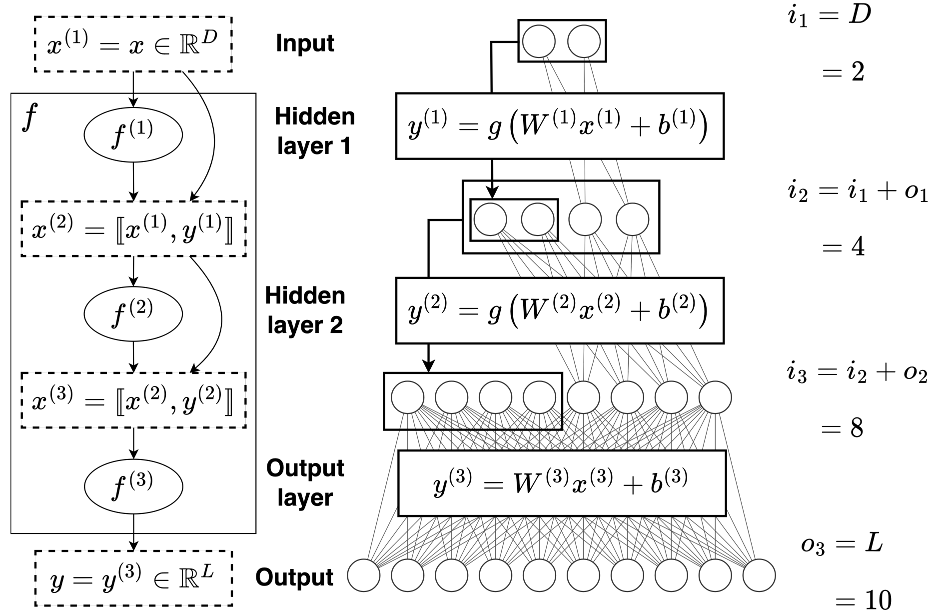
Solution Emulate the code with a neural network trained on a grid of pre-computed models.

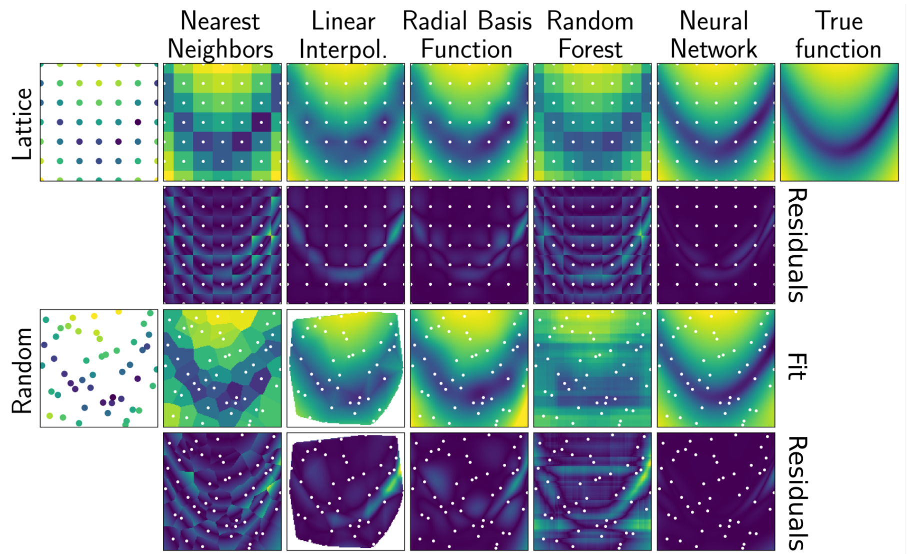
Specific strategies to be accurate

1. Outlier removal procedure to avoid reproducing numerical instabilities.
2. Clustering method to divide the prediction job.
3. Dimension reduction technique to size the neural network.
4. Augmented inputs to ease the learning of non-linearities.
5. Dense architecture to ease the learning of simple relations.

Results on Meudon PDR code:

Average error 4.5%.
Speed 100 to 1000 $\times$ faster.
Memory imprint 10 $\times$ smaller.





Deep learning denoising by dimension reduction (Einig, et al. 2023)

State-of-the-art observations

- Large intensity dynamic.
- Interesting science often happens at low/medium S/N.

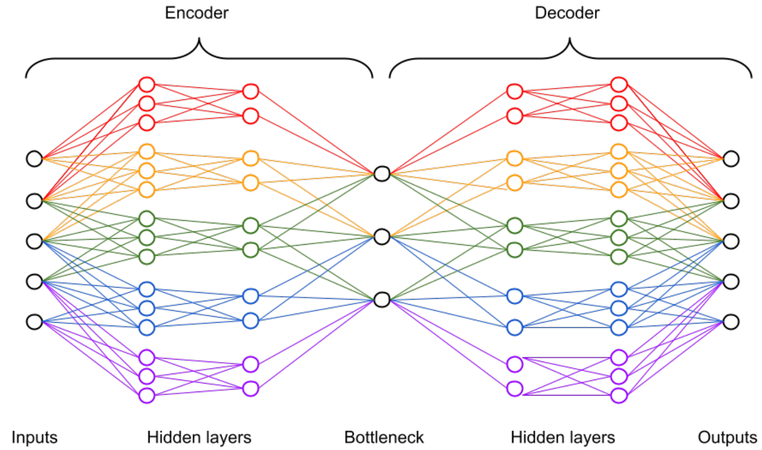
Solution Filter noise by compressing/decompressing useful information $\Rightarrow$ Neural Network Autoencoder.

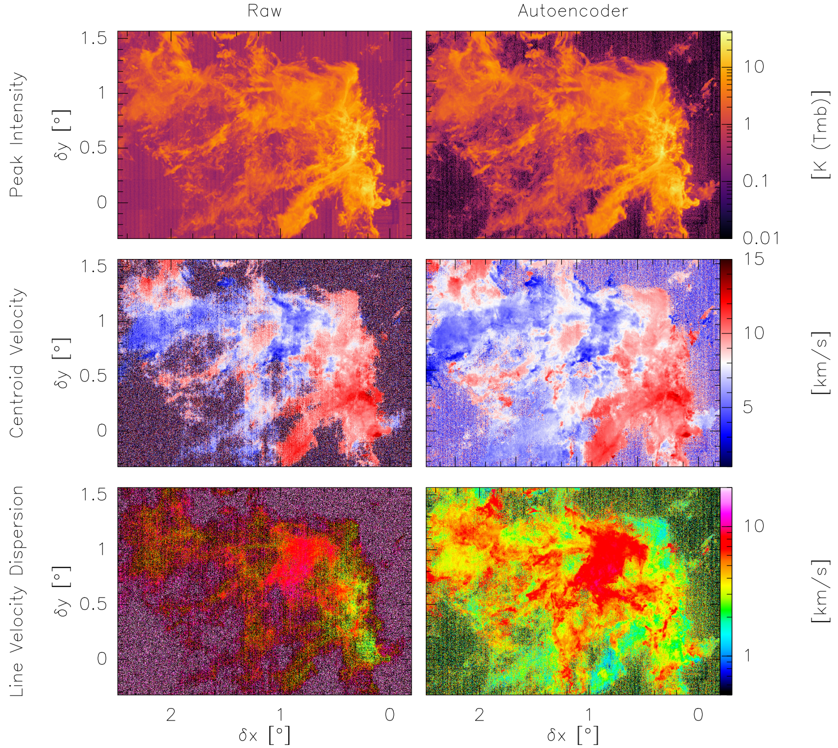
Specific strategies to be avoid deformation

1. Estimate the amount of redundant information in the dataset to fix the size of the autoencoder.
2. Use a locally connected autoencoder to only explore correlated channels.
3. Take into account the fact that signal voxels are embedded in a sea of noise with a specific pay-attention loss function.

Result

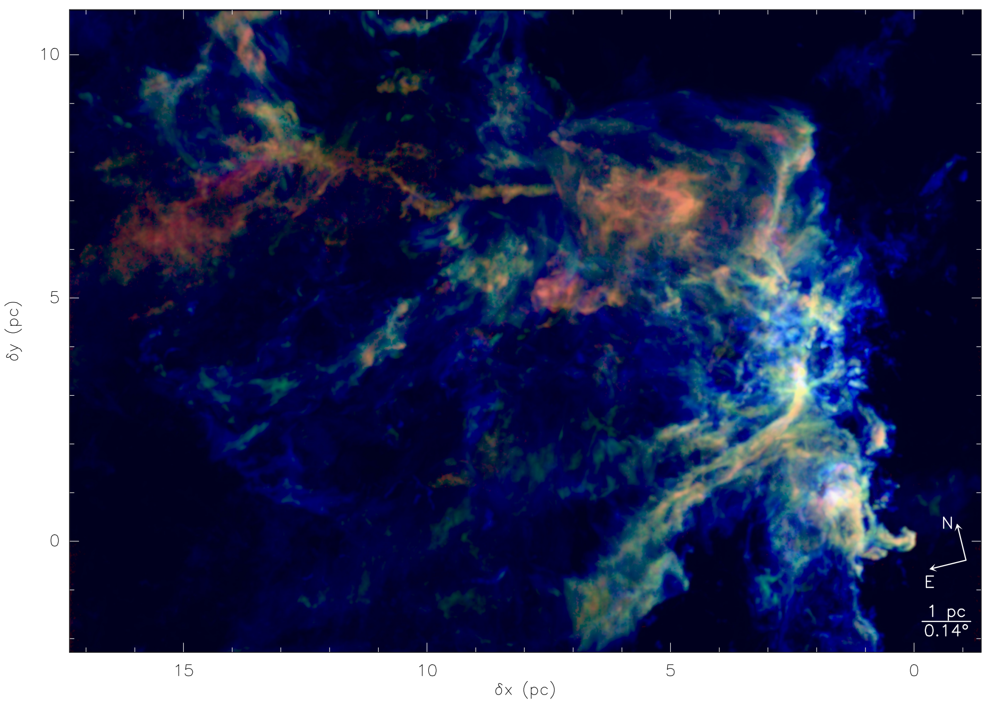
- Reduced (not suppressed) noise $\Rightarrow$ much better determination of the moments of the line.
- No deformation of the signal at high S/N.
- No creation of false signal at low S/N.



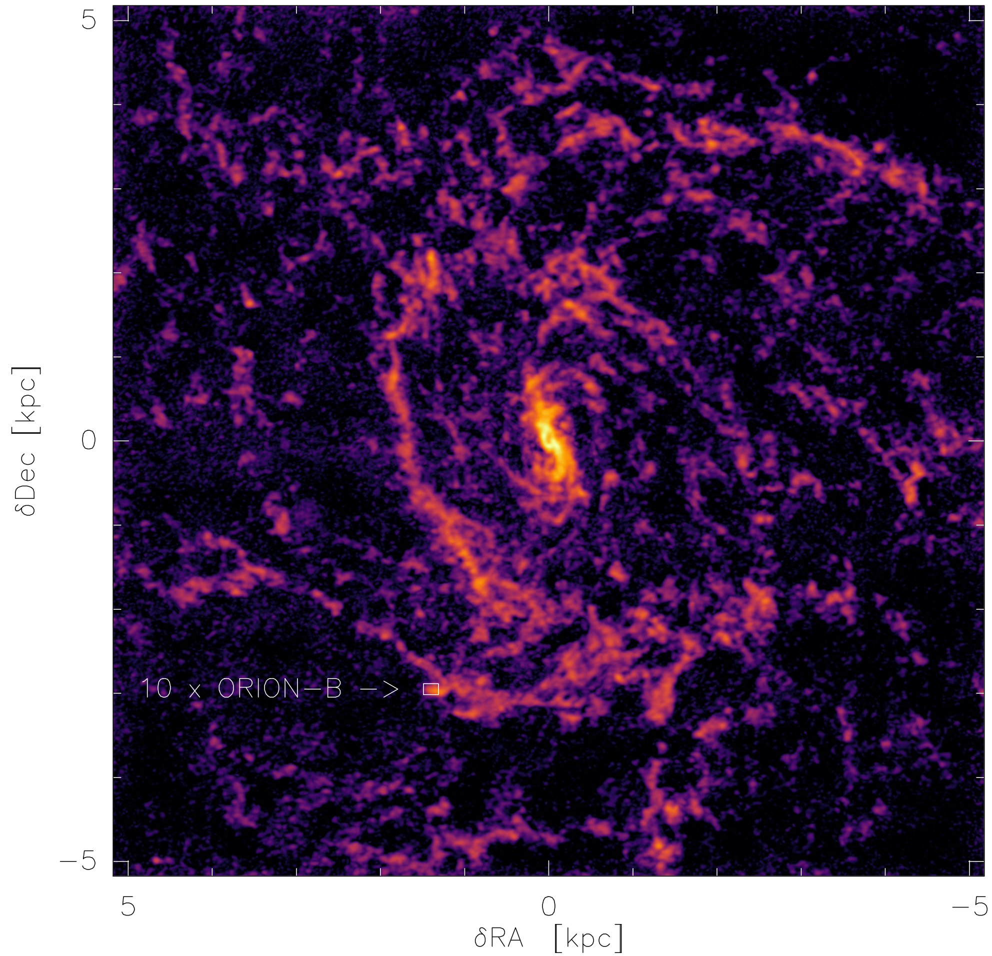


How to bridge the gap of spatial scales between Milky Way and nearby galaxy studies?

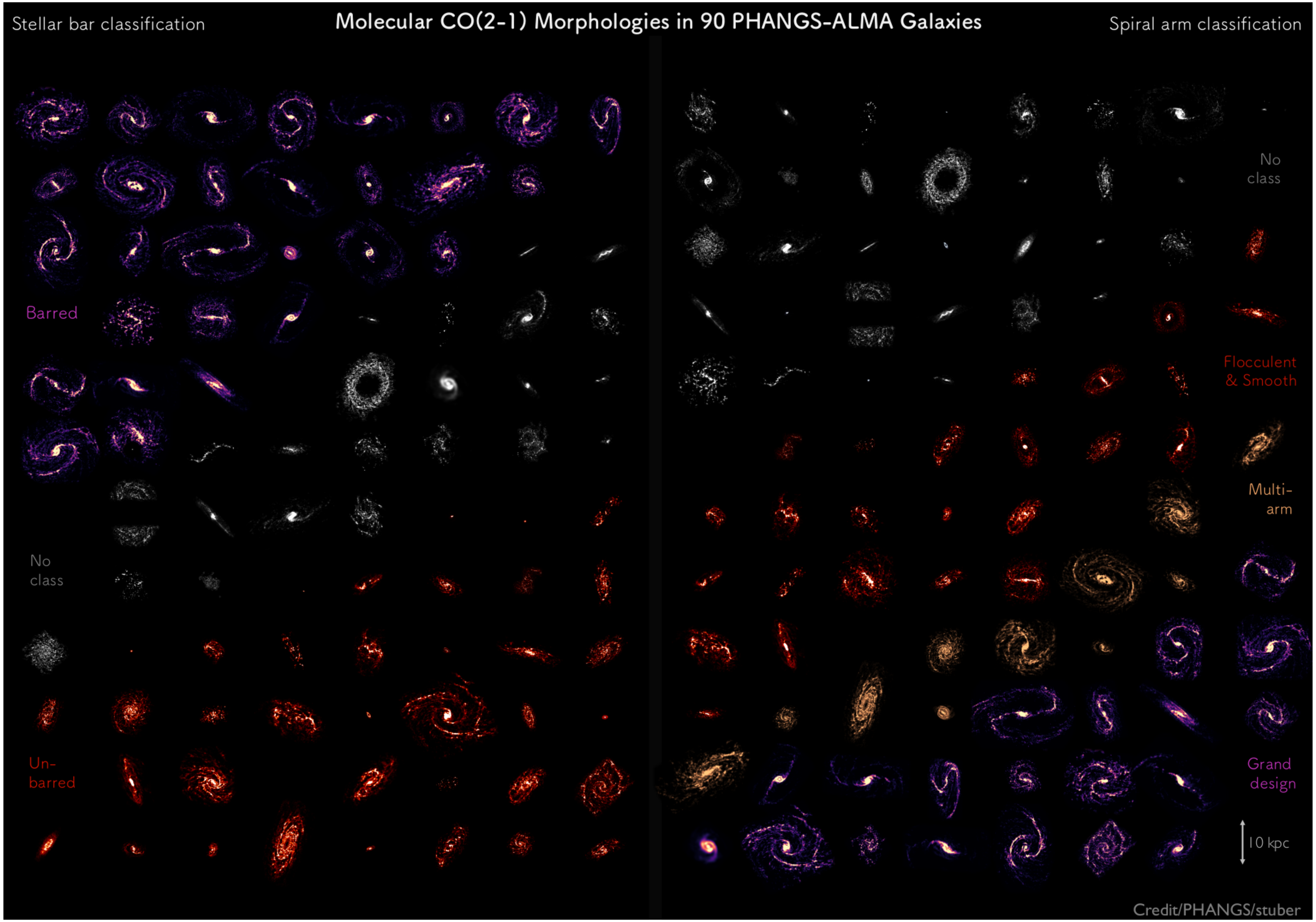
^{12}CO, ^{13}CO, and C^{18}O J=1–0 in Orion B



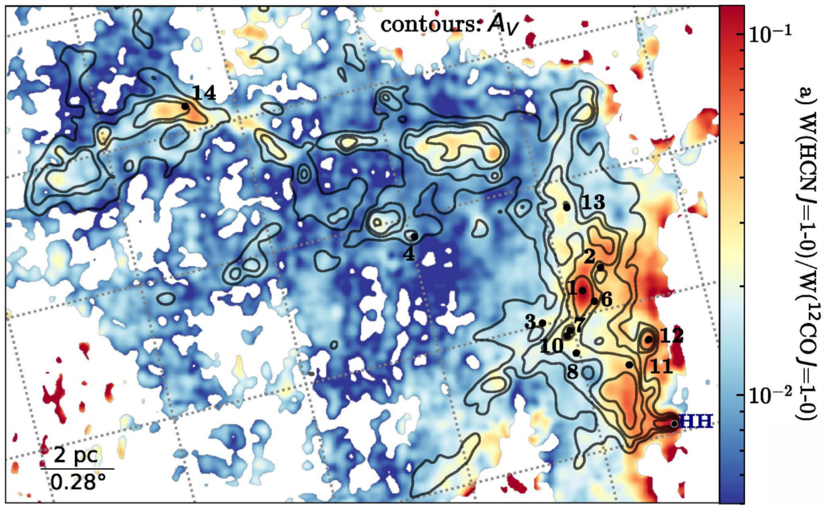
^{12}CO J=1–0 in IC 342 (Querejeta et al. 2023)



To make the best use of, e.g., the PHANGS survey?



Is $W(\mathrm{HCN\ J{=}1{-}0})/W(^{12}\mathrm{CO\ J{=}1{-}0})$ a dense gas tracer? (Santa-Maria et al. 2024)

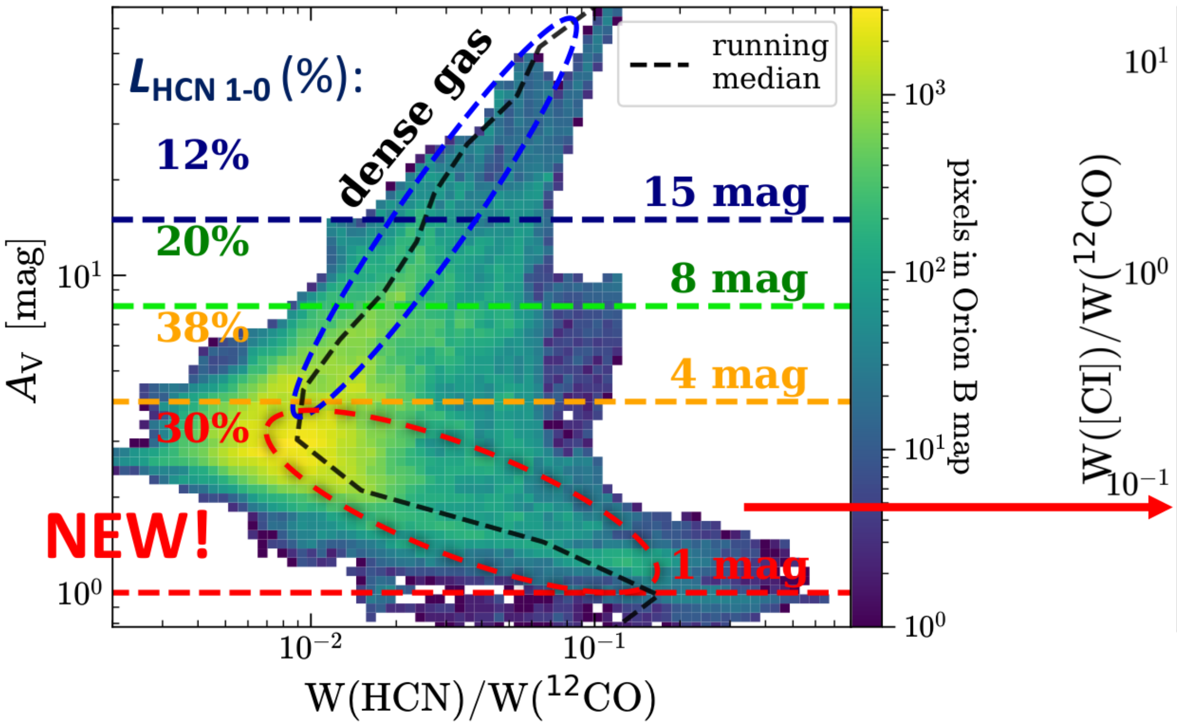


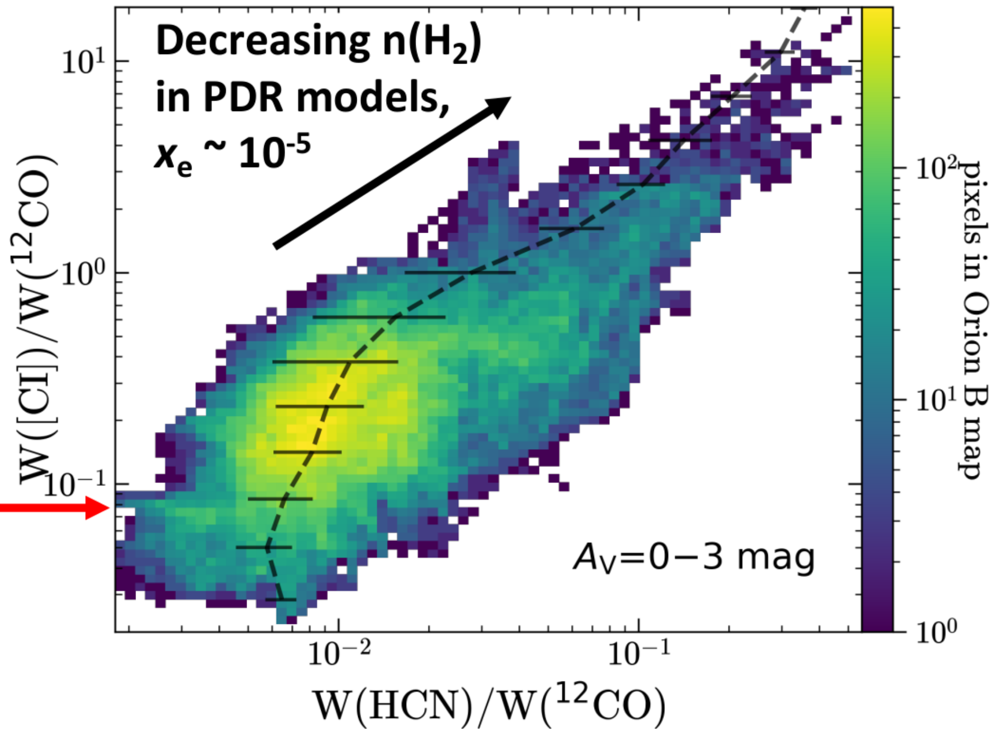
In part

- High values in dense filaments and cores.
- Low values in diffuse gas.

Also sensitive to far UV field Higher values in Photo-dissociation Regions at the edges of HII regions.

The net outcome depends on the proportion of both kind of gas in the beam.





Understanding spatially unresolved measurements of molecular line emission (Zakardjian et al. subm.)

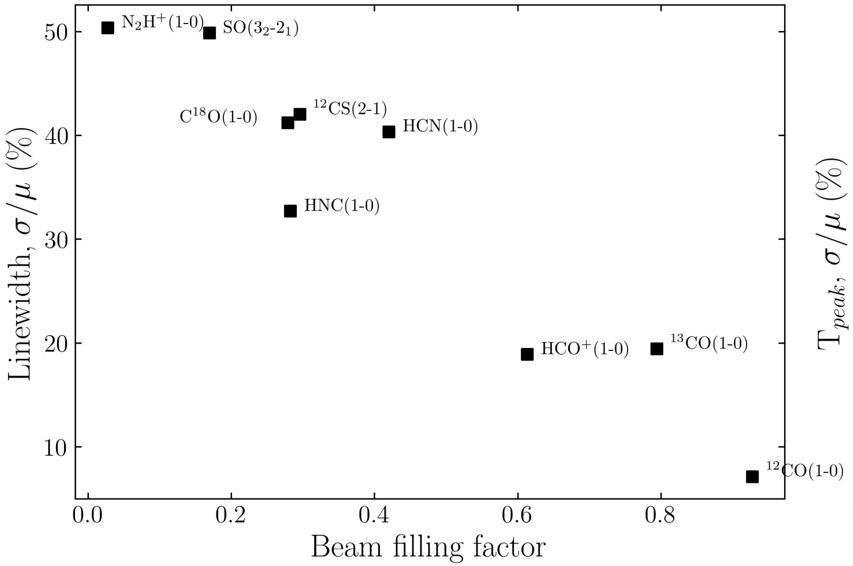
Generating 1000 realistic unresolved observations of a molecular cloud

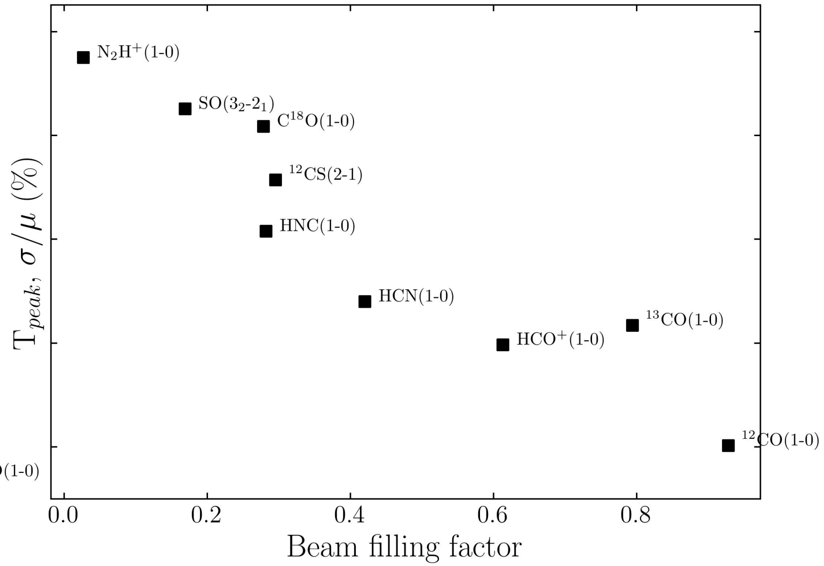
$\Rightarrow$ a single unresolved spectrum per line and mocked cloud.

- Draw random fields of the column density and centroid velocity using fractional Brownian motion.
- Shuffle the spectra of the ORION-B cubes to follow these random fields.
- Average the resolved spectra.

Effect on linewidths and peak temperatures

- Unresolved line profile parameters vary significantly purely because of the sub-beam distribution of the emission.
- Variability of up to a factor of 2 for N_2H^+ J=1–0.
- Variability of less than 5% for CO J=1–0.





Investigating the magnetic field strength across the Flame Nebula (Beslic et al., 2024, PI: D. Lis)

Status Lack of knowledge of the magnetic field in the evironment of new-born stars at scales of 0.1 pc or 50″.

Our goal Correlate changes in the magnetic field geometry with corresponding changes in the kinematic properties of the gas.

Results

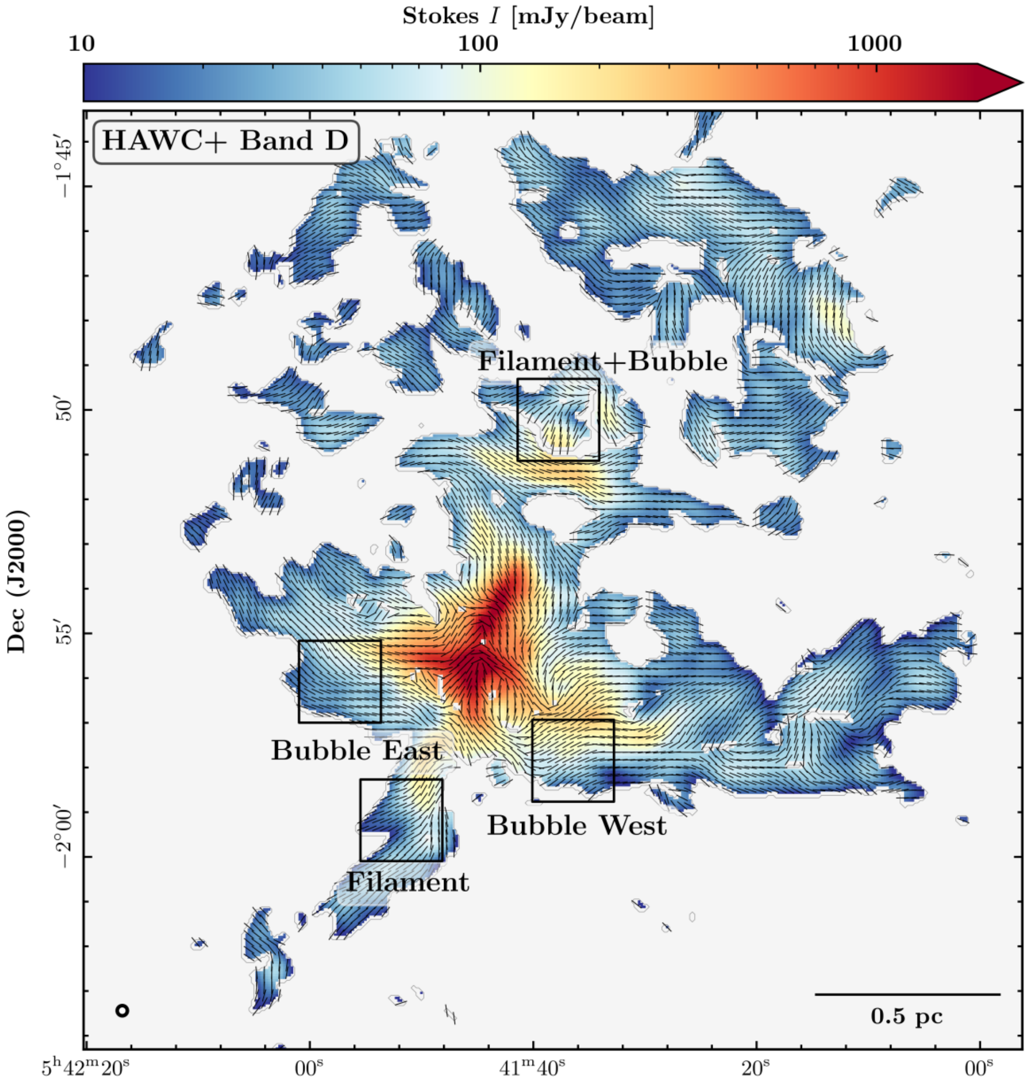
B-field morphology

- Parallel to edges of the expanding HII region.

B-field strenght using

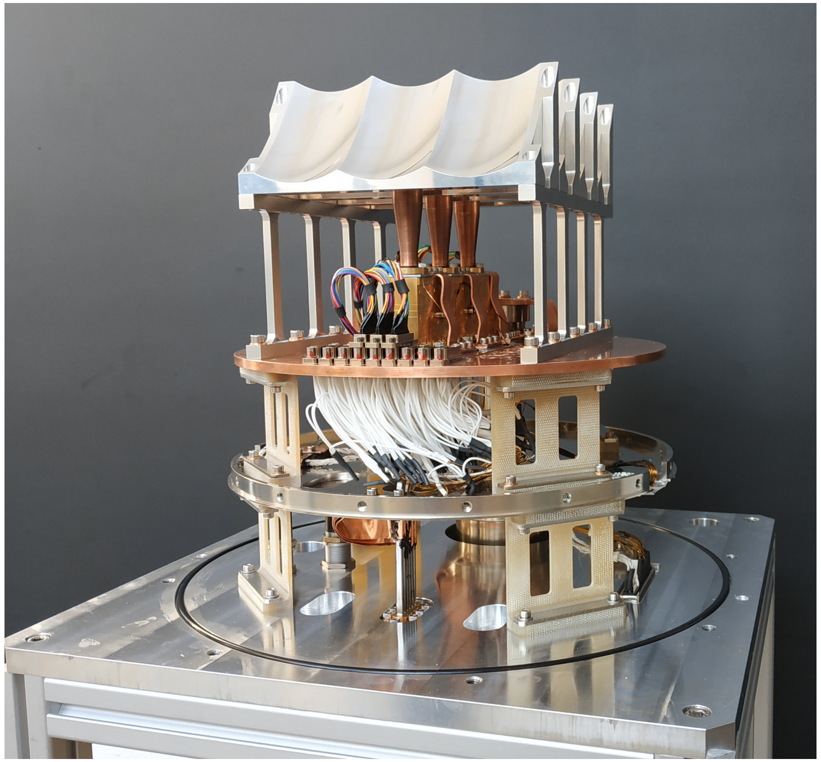
$$B_{\rm pos} \sim 1.8 \times \sqrt{n_{\rm H_2}} \frac{\Delta v}{\sqrt{\sigma_c(\phi)}} [\mu\text{G}].$$

Bubble West	$30 - 90\,\mu$G
Bubble East	$100 - 200\,\mu$G
Filament	$30 - 100\,\mu$G
Filament + Bubble	$80 - 150\,\mu$G

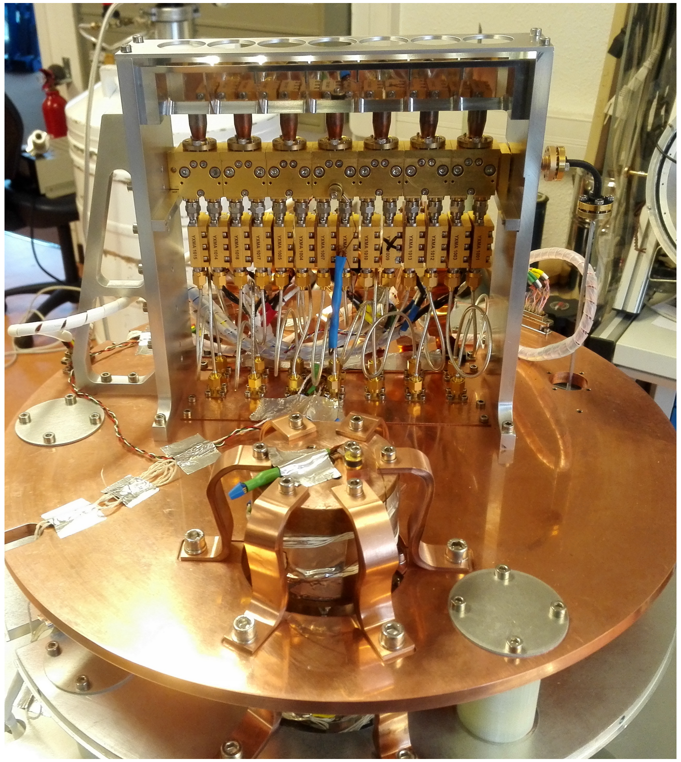


What could be done in ∼1000 hr with the IRAM-30m ALHAMBRA

Instrument	# 3 mm ind. Beams	FoV degree^2	Sensitivity mK	Data TB	Comment
EMIR at 30m	1	5	100	0.5	ORION-B like
NG Multi-Beam at 30m	25	5	15	0.5	Deep ORION-B
NG Multi-Beam at 30m	25	125	100	17.0	% of galactic plane



3 mm HEMT 3-beam prototype.
P. Serres, O. Garnier, & the front-end group



1 mm SIS 7-beam prototype.
D. Maier et Q. Moutote & the front-end group

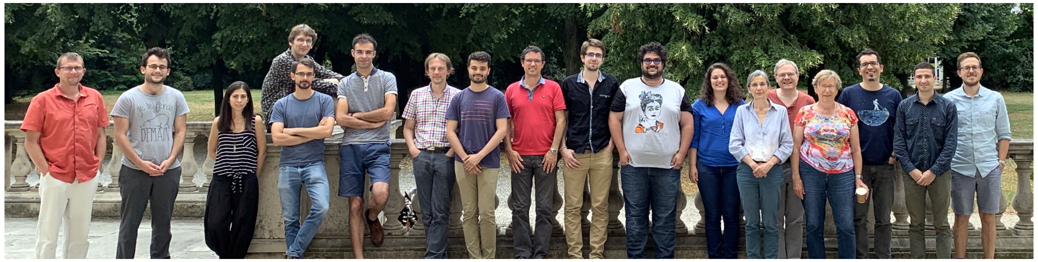
An actual collaboration between data scientists and astronomers

Collaboration started in may 2018

- 3 co-directed PhD students:
  P. Palud, L. Einig, L. Segal.

Articles

- 5 in common in 2023.
- 3 of them led by data scientists.



Conclusions

New inter-disciplinary collaboration It takes time and will but it pays off.

The black box approach most often fails Adapting the machine learning approach to the astrophysical problem leads to success.

Only the beginning of the story We need multi-beams to increase the sampling of conditions at high S/N.

Chemistry challenge A more quantitative understanding of the detailed of the photo-chemistry of the CO isotopologues is required.

Web site http://www.iram.fr/~pety/ORION-B

References

- Antoine Roueff, et al., 2024, "Bias versus variance when fitting multi-species molecular lines with a non-LTE radiative transfer model", https://doi.org/10.1051/0004-6361/202449148
- Ivana Bešlić, et al., 2024, "The magnetic field in the Flame nebula", https://doi.org/10.1051/0004-6361/202348376
- Miriam Santa-Maria, et al., 2024, "HCN emission from translucent gas and UV-illuminated cloud edges revealed by wide-field IRAM 30 m maps of the Orion B GMC. Revisiting its role as a tracer of the dense gas reservoir for star formation", https://doi.org/10.1051/0004-6361/202346598
- Pierre Palud, Lucas Einig, et al., 2023, "Neural network-based emulation of interstellar medium models", https://doi.org/10.1051/0004-6361/202347074
- Lucas Einig, et al., 2023, "Deep learning denoising by dimension reduction: Application to the ORION-B line cubes", https://doi.org/10.1051/0004-6361/202346064
- Pierre Palud, et al., 2023, "Efficient Sampling of Non Log-Concave Posterior Distributions With Mixture of Noises", https://arxiv.org/pdf/2211.12915
- Mathilde Gaudel, et al., 2022, "Gas kinematics around filamentary structures in the Orion B cloud", https://doi.org/10.1051/0004-6361/202142109
- Pierre Gratier, et al., 2021, "Quantitative inference of the H_2 column densities from 3 mm molecular emission: A case study towards Orion B", https://doi.org/10.1051/0004-6361/202037871
- Antoine Roueff, et al., 2021, "C18O, 13CO, and 12CO abundances and excitation temperatures in the Orion B molecular cloud: An analysis of the precision achievable when modeling

spectral line within the Local Thermodynamic Equilibrium approximation", https://doi.org/10.1051/0004-6361/202037776

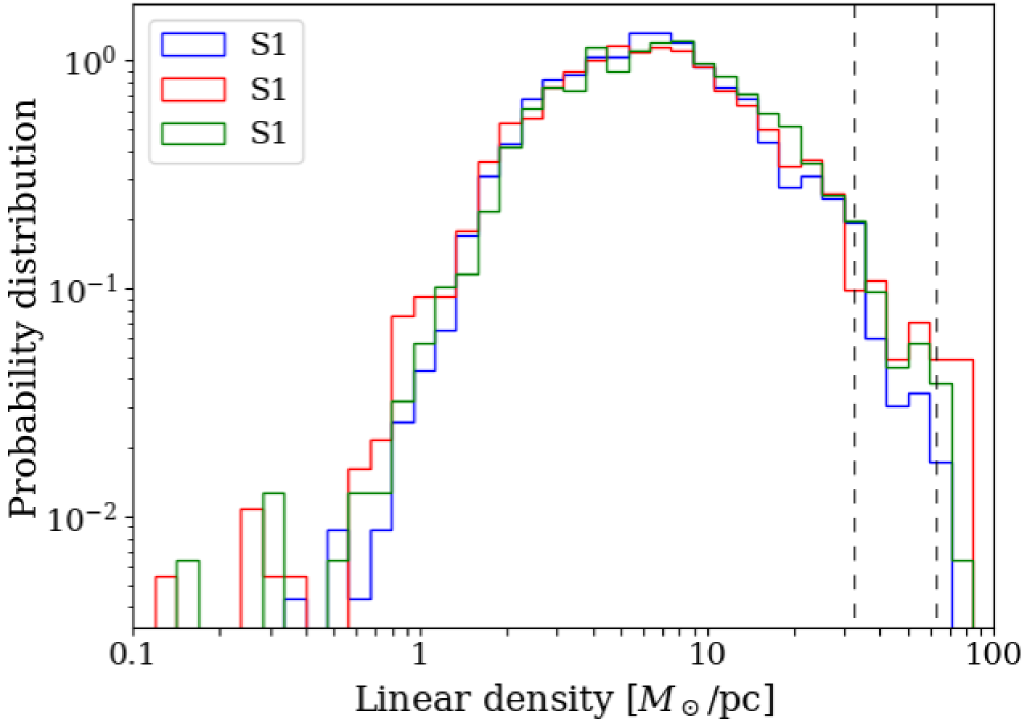
- Emeric Bron, et al., 2021, "Which tracers for the ionization fraction in the translucent and dense phases of molecular clouds? Automated exploitation of massive astrochemical model grids", https://doi.org/10.1051/0004-6361/202038040
- Jan H. Orkisz, et al., 2019, "A dynamically young, gravitationally stable network of filaments in Orion B", https://doi.org/10.1051/0004-6361/201833410
- Emeric Bron, et al., 2018, "Clustering the Orion B giant molecular cloud based on its molecular emission", https://doi.org/10.1051/0004-6361/201731833
- Jan H. Orkisz, et al., 2017, "Turbulence and star formation efficiency in molecular clouds: solenoidal versus compressive motions in Orion B", https://doi.org/10.1051/0004-6361/201629220
- Pierre Gratier, et al., 2017, "Dissecting the molecular structure of the Orion B cloud: Insight from Principal Component Analysis", https://doi.org/10.1051/0004-6361/201629847
- Jérôme Pety, et al., 2017, "The anatomy of the Orion B Giant Molecular Cloud: A local template for studies of nearby galaxies", https://doi.org/10.1051/0004-6361/201629862

Additional material

Filamentary network in Orion B (Orkisz et al. 2019)

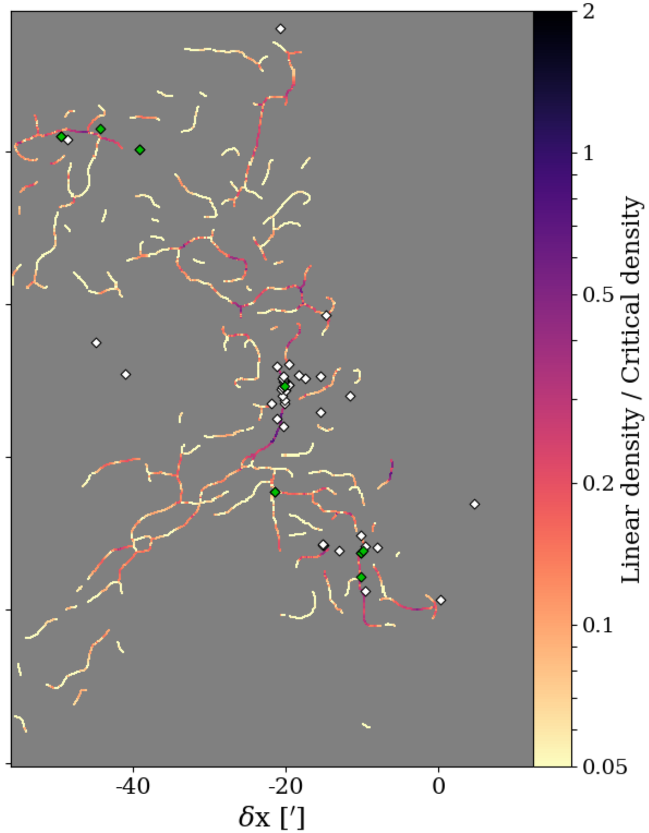
Wide range of densities

- Linear ($1 - 100\,\mathrm{M_{sun}\,pc^{-1}}$) and volume ($2 \times 10^3 - 2 \times 10^5\,\mathrm{cm^{-3}}$) densities.
- Upper end is similar to other studies $\Rightarrow$ Many filaments have low intensity contrast compared to the background.



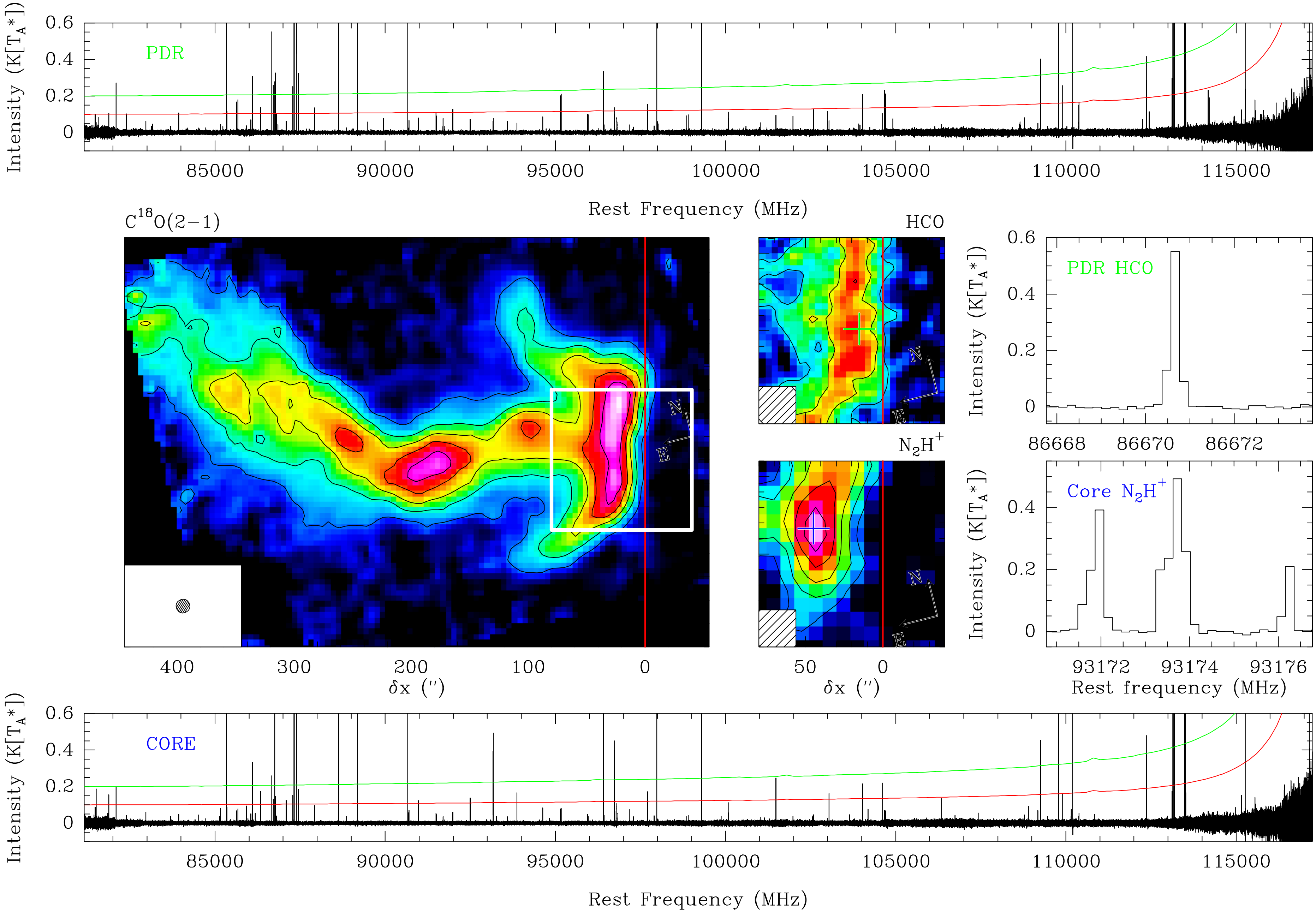
Dominated by low-density, thermally subcritical filaments

- Most of the filaments are not collapsing to form stars $\Rightarrow$ Correlated with low star formation efficiency (Lada et al. 2010, Megeath et al. 2012).
- Only 1% of the mass in super-critical filaments inside the star forming regions (NGC 2023 and NGC 2024).



Horsehead WHISPER ⇒ A forest of faint lines

Red curve: 5σ & Green curve: 10σ with $\sigma \sim 5\,\mathrm{mK}$



Astrophysics meet data science for the study of Giant Molecular Clouds

J. Pety and the DAOISM consortium 2024