Optimizing observation proposals with information theory

L. Einig, P. Palud

A. Roueff, J. Pety, Emeric Bron, F. Le Petit, P. Chainais, M. Gerin,
P.-A. Thouvenin, J. Chanusssot

and the
Orion-B Consortium - https://www.iram.fr/~pety/ORION-B/

gipsa-lab

iram
Institut de
Radioastronomie
Millimétrique

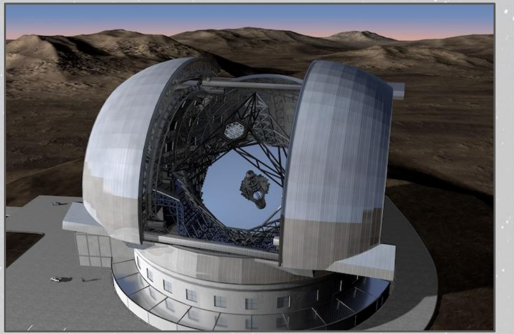
CRIStAL
Centre de Recherche en Informatique,
Signal et Automatique de Lille

l'Observatoire de Paris

ORION-B

DAOISM

The problem of line selection



*We observe the ISM to **improve our understanding**, but it is **expensive**!*
What tracers should we observe in priority and for how long, to maximize the return on investment ?

The goal

find the tracers that **best constrain** the physical conditions **in our models**

Challenges

A/ How do we quantify the potential of a line to constrain the physical conditions ?

B/ the situation is even more complex when trying to evaluate **combinations of tracers**

Previous approaches

- correlation coefficients
  (but fails for non linear relations)
- fit quality of Machine Learning regression model (Bron et al. 2021)
- inherent feature importance estimation for Random Forest models (Gratier et al. 2021)

Mutual information

Discrete entropy $H(X) = -\sum_{x \in \mathcal{X}} [\log_2 \pi(x)] \; \pi(x)$ (in bits)

differential entropy
(continuous var.)

$$h(\Theta) = -\int [\log_2 \pi(\theta)] \; \pi(\theta) \; \mathrm{d}\theta$$

uncertainty on physical conditions

conditional differential entropy

$$h\left(\Theta \mid Y^{(s)}\right) = -\int \left[\log_2 \pi\left(\theta \mid \mathbf{y}^{(s)}\right)\right] \; \pi\left(\theta, \mathbf{y}^{(s)}\right) \; \mathrm{d}\theta \; \mathrm{d}\mathbf{y}^{(s)}$$

uncertainty remaining on physical cond. after observing some tracers (averaged over noise)

Mutual information = **differential entropy** - **conditional differential entropy**

Can be estimated **from a dataset** of
(physical conditions, associated tracer intensity)
(**Monte Carlo estimation**)

Scipy's implementation: limited to one-to-one analysis.

We use our own implementation.

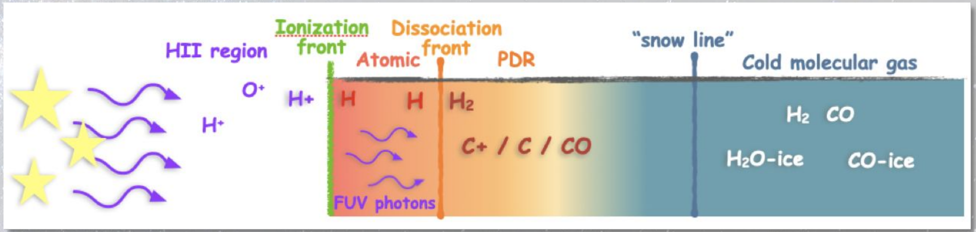
In the following, we illustrate our tools and methods on a **realistic synthetic dataset** simulating **a PDR observed with the IRAM 30m telescope**

Illustration: estimating mutual entropy from the Meudon PDR code

The Meudon PDR code: (Le Petit et al. 2006)

- Models the physics of UV irradiated neutral gas (photodissociation region)
- We restrict the physical parameters to:
  the thermal pressure P_{th} (K cm^{-3})
  the incident UV field G_0 (Habing)
  the total A_V of the slab (mag.)
- Predicts integrated intensities of thousands of emission lines
- We use the fast and accurate neural network emulator of Palud et al. 2023

→ Provides the relationship between physical parameters and line intensities



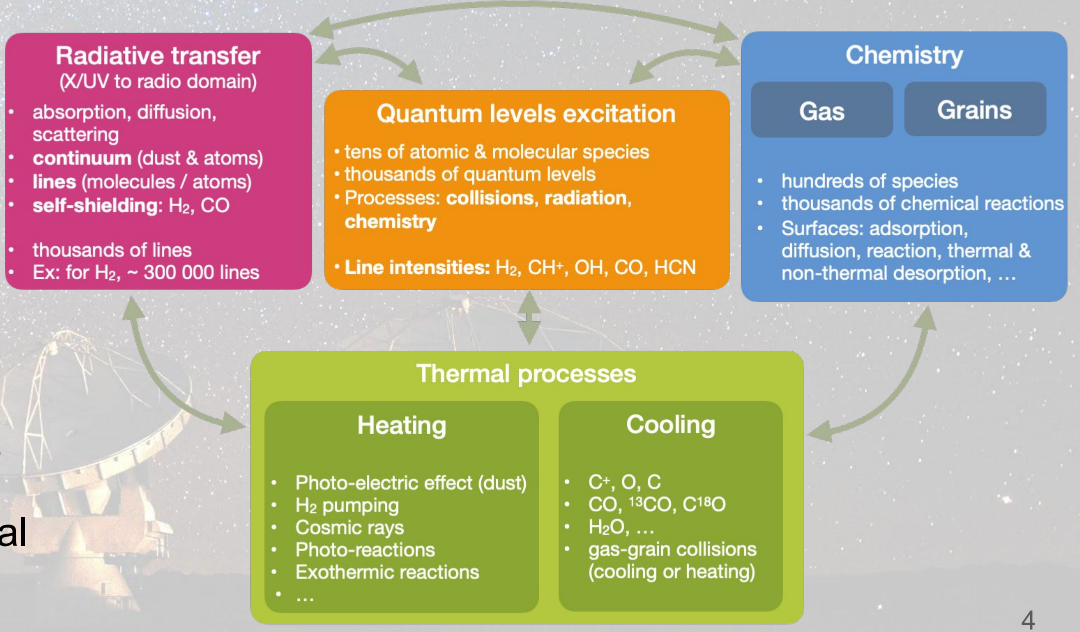


Illustration: estimating mutual entropy from the Meudon PDR code

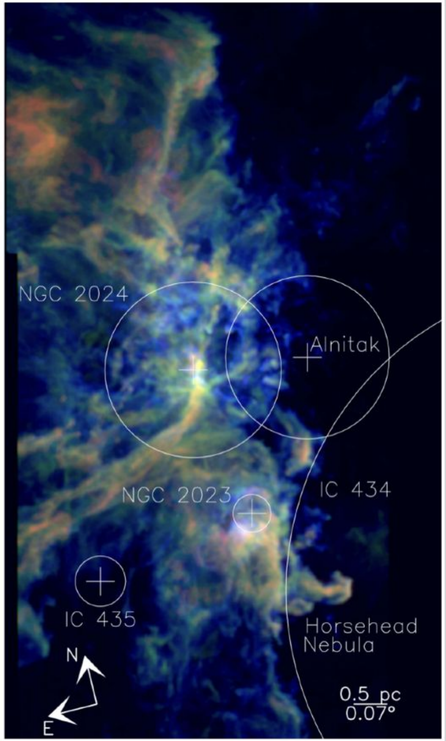
The observation model:

- We follow the properties of the ORION-B Large Program
- Accounts for both
  - additive gaussian noise (thermal noise) with noise level taken from ORION-B program (Einig et al. 2023)
  - multiplicative lognormal calibration uncertainty, taken from EMIR properties (5 to 10% depending on band)
- Keep lines observable with IRAM-30m → 33 lines

The ORION-B Large Program:

- IRAM-30m Large Program
- PIs: J. Pety & M. Gerin
- 850 h
- 5 square degrees
- 26” resolution
- 71 - 116 GHz
- ~200 kHz spectral resolution
- median noise level 0.1 - 0.5 K
- unbiased dataset of a full GMC

(Pety et al. 2017)



Our Goal: *to **evaluate the full constraining power** of a (combination of) tracer(s) with respect to physical conditions*

Selected quantitative criterion: **mutual information (MI) from information theory**

accounts for **non-linear** relationships

value **independent from** an algo class (neural network, random forest, etc.)

allows to analyze **many-to-many** relations (many physical conditions, many tracers)

includes the **noise model** of the considered instrument

New tools produced

- Understanding relation between ISM tracers and physical conditions
- Help optimizing observation proposals

Mutual information **maps**

Evolution of MI with **integration time**

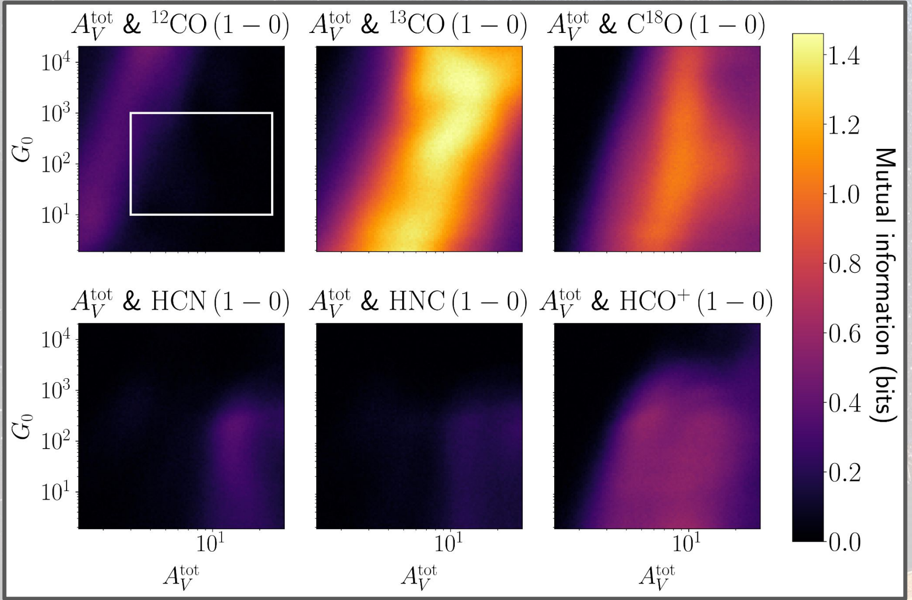
selection of **most informative single line** at given S/N

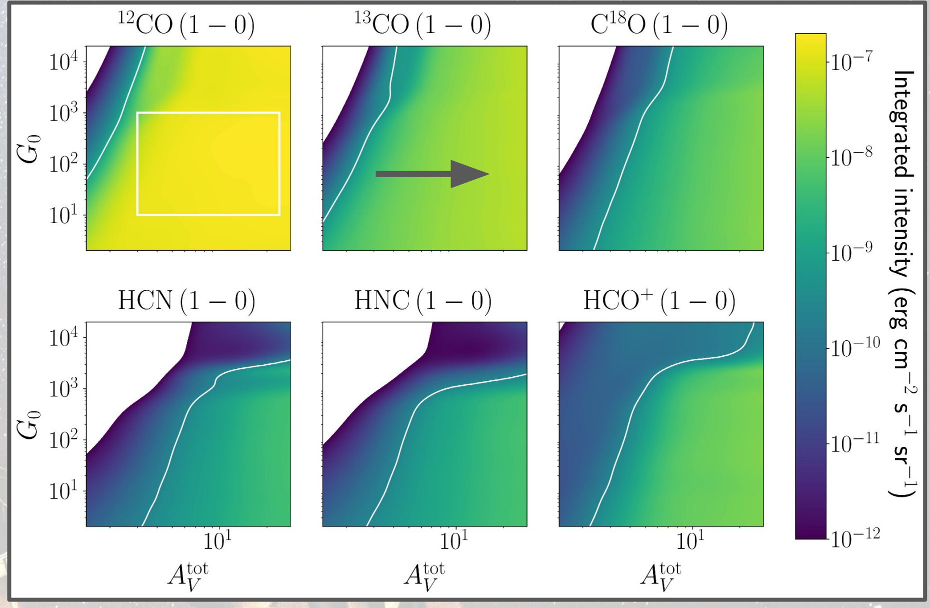
selection of **most informative combination of K lines** at given S/N

Tool 1: mutual information maps (on individual tracers)

For individual tracers, **maps of MI** can be compared with **maps of integrated intensity**

When can we trace the total column density (total A_V) ?





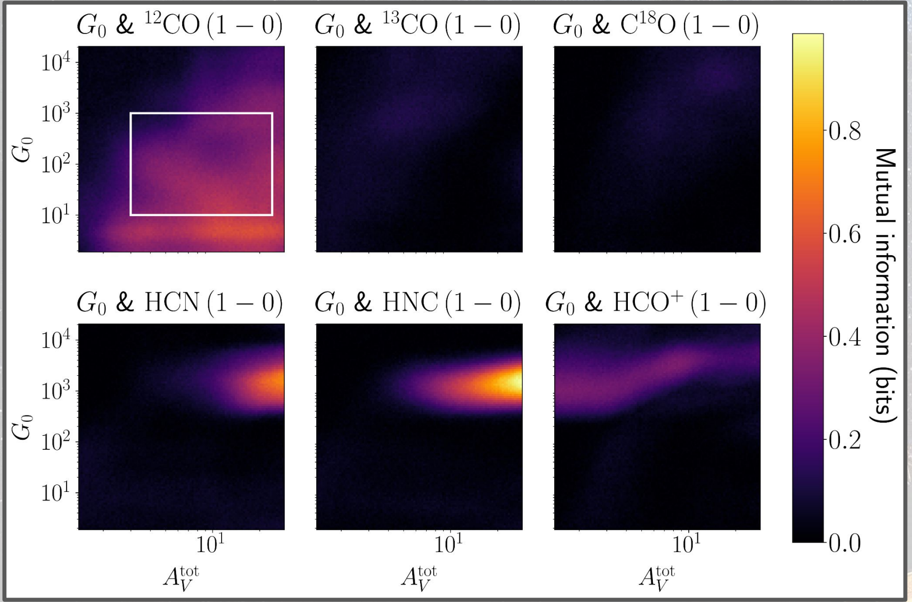
To have a high MI: large S/N (>10) & large gradient

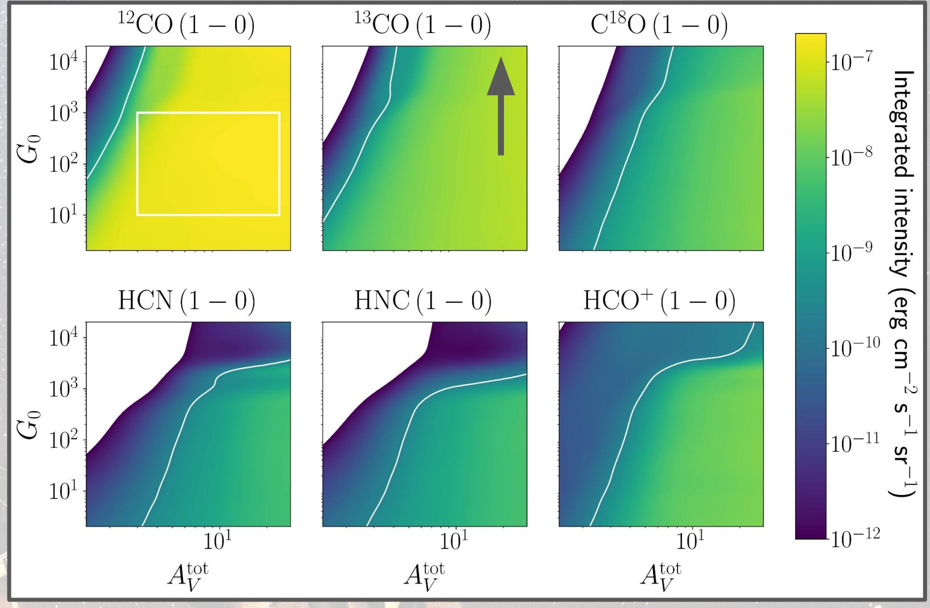
Einig et al. 2024

Tool 1: mutual information maps (on individual tracers)

For individual tracers, **maps of MI** can be compared with **maps of integrated intensity**

When can we trace the UV irradiation (G_0) ?





To have a high MI: large S/N (>10) & large gradient

Einig et al. 2024

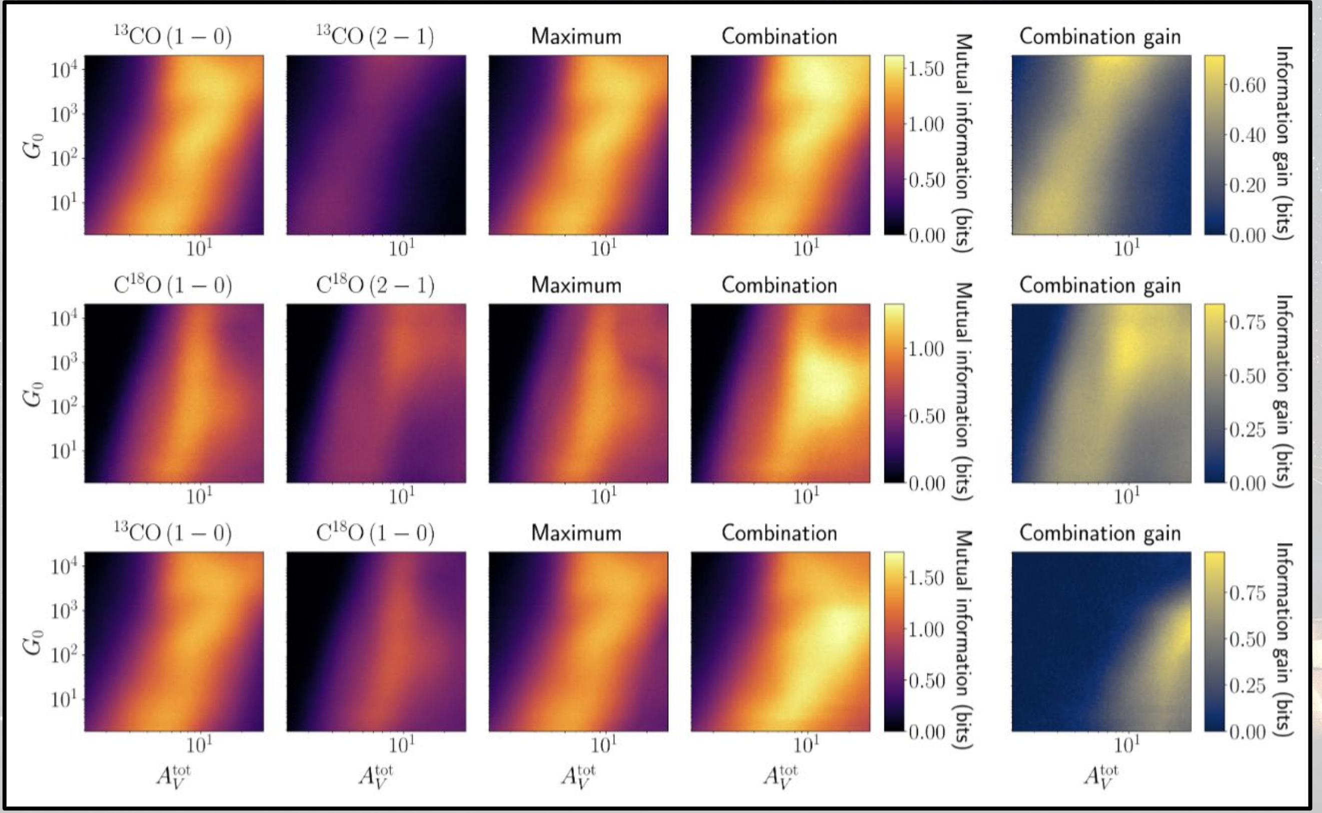
Tool 1: mutual information maps (on combination of tracers)

Constraint on A_V

Best individual MI1

Combining both lines MI2

gain = MI2 - MI1



combination of ^{13}CO first 2 lines: informative where **both** individual lines are informative

combination of $C^{18}O$ first 2 lines: informative even where **none** of the individual lines is informative (high Av, small G0)

^{13}CO and $C^{18}O$ first line: Best combination of the three, **covers well** the parameter space with a **high MI value**

Einig et al. 2024

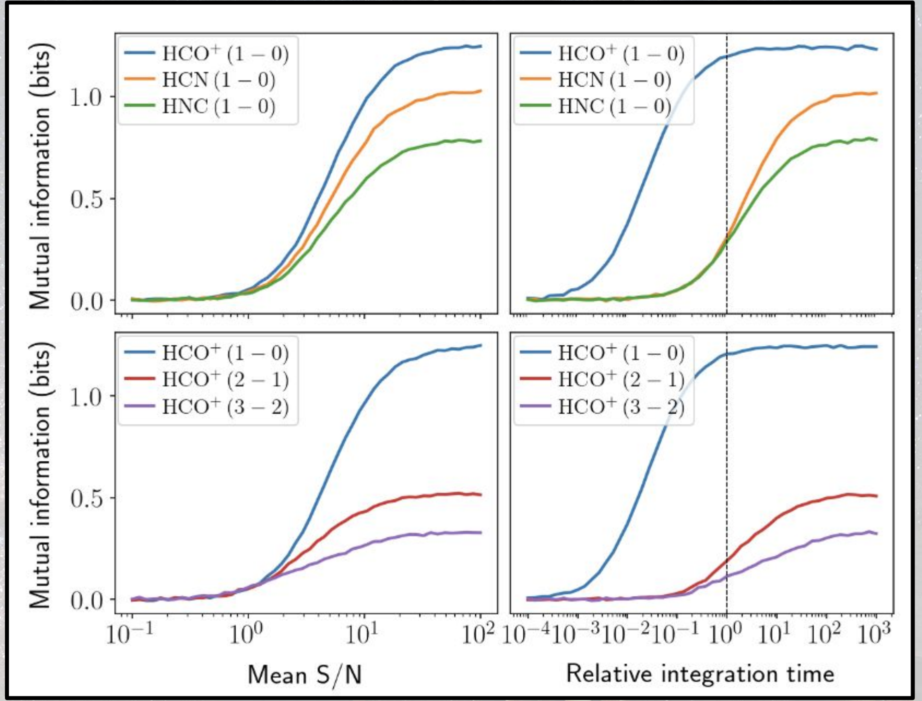
Tool 2: how long should we observe? To target which S/N?

Constraint on A_V

Comparing…

ground state transition of multiple molecules

first three lines of HCO^+



Same curves, juste horizontally shifted
(different noise standard deviation for each line)

Key properties

relation S/N and MI: not linear
all the curves have an S-shape

at low S/N, MI = 0

MI saturates at high S/N

different saturation value

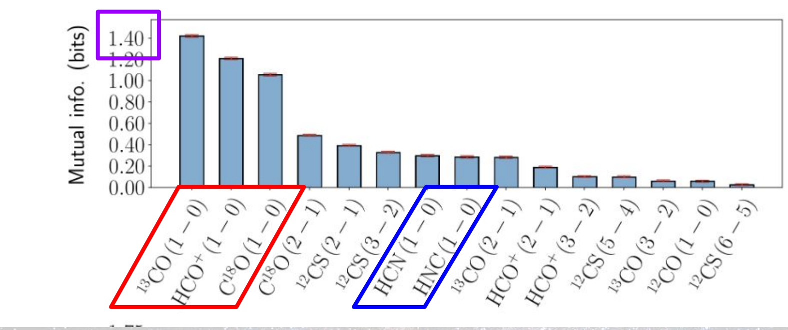
Increasing the S/N
does not necessarily mean
improving estimation accuracy on physical conditions

possibility: adjust integration time to **aim for the saturation S/N**

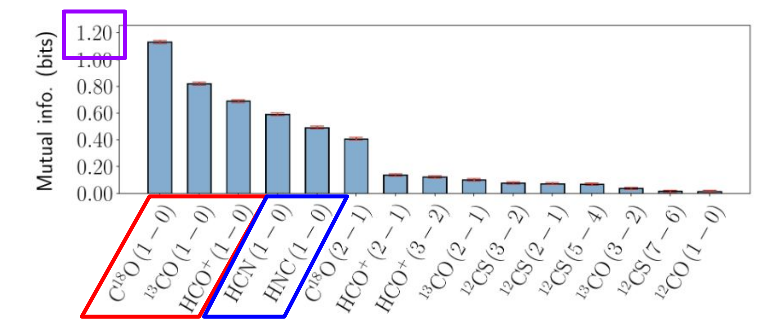
Einig et al. 2024

Tool 3: at given integration time, which individual line to observe?

Constraint on A_V



filamentary gas
6 < Av < 12 mag



dense cores
12 < Av < 24 mag

With MI, one can rank individual tracers to select the most informative one

The ranking depends on the expected physical regime:

^{13}CO (1-0) best line for filamentary gas
C^{18}O (1-0) best line for dense cores

HCN (1-0) and HNC (1-0) are relatively better tracers in dense cores

in dense cores, Av is harder to reconstruct accurately

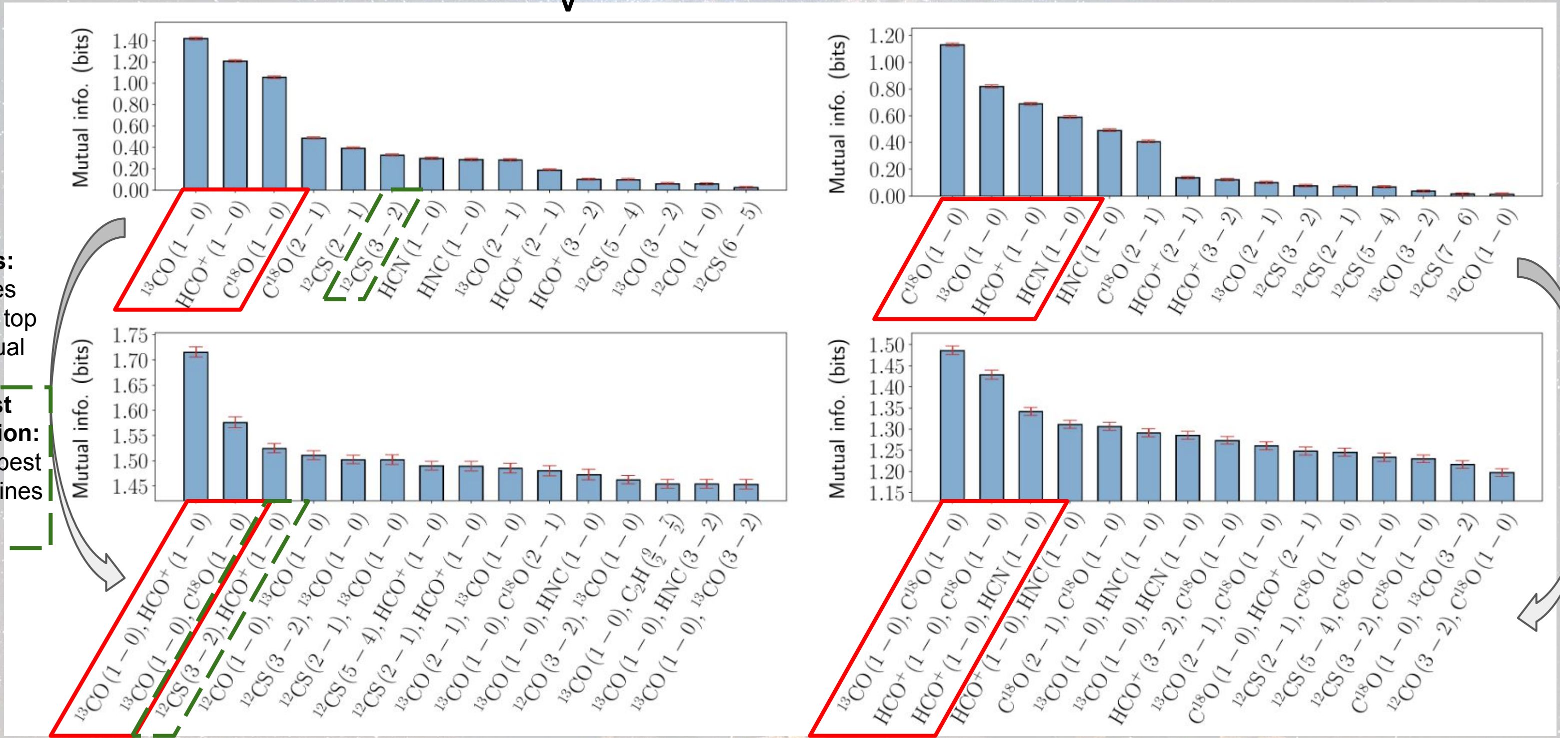
Einig et al. 2024

Tool 4: at given integration time, which line couple to observe?

Constraint on A_V

Top 2 couples: combines lines from top 3 individual

3rd best combination: 2nd + 6th best individual lines



Top 3 couples: combines lines from top 4 individual

filamentary gas
6 < AV < 12 mag

dense cores
12 < AV < 24 mag

Einig et al. 2024

The best combination of K lines might contain the K most informative lines, **but not necessarily**

Conclusions and perspectives

- A tool for tracer selection to help in proposal preparation
- The method is **very general:** not limited to Meudon PDR code, nor to IRAM-30m mm emission lines
- **No limiting assumption:** not limited to linear relations, nor normal distributions
- **Not limited to one-to-one** relations: can quantify constraining power for sets of multiple tracers, and also to constraining multiple parameters at once

Einig et al. 2024

What's next ?

- Can be applied in a **model-independent** way on observational dataset
  → Application to ORION-B dataset and physical conditions derived from Herschel dust observations.
  (Einig et al. in prep)
- Applications to come to other astrophysical models and other telescopes

These tools are now available online!

Our Mutual Information tool

https://pypi.org/project/**infovar**/

infovar 0.2.0

```
pip install infovar
```

Latest version

Released: Oct 2, 2024

To reproduce the figures of this presentation

https://github.com/einigl/
iram-30m-emir-obs-info

IRAM 30-meter EMIR observations informativity

docs passing | coverage 64%

This package implements tools to quantitatively estimate the usefulness of spectral line observations for estimating physical conditions. It provides a tool for simply reproducing observations made at IRAM 30-meter millimeter-wave telescope coupled with the EMIR receiver. Other instruments can also be simulated.

Line intensity predictions are made using a neural network emulation of the Meudon PDR code. This emulator enables a thousand predictions to be made in around 10 ms on a laptop, with an average error of less than 5%.

https://arxiv.org/abs/
2408.08114

arXiv > astro-ph > arXiv:2408.08114

Search...

Help | A

Astrophysics > Astrophysics of Galaxies

[Submitted on 15 Aug 2024 (v1), last revised 21 Sep 2024 (this version, v2)]

Quantifying the informativity of emission lines to infer physical conditions in giant molecular clouds. I. Application to model predictions

Lucas Einig, Pierre Palud, Antoine Roueff, Jérôme Pety, Emeric Bron, Franck Le Petit, Maryvonne Gerin, Jocelyn Chanussot, Pierre Chainais, Pierre-Antoine Thouvenin, David Languignon, Ivana Bešlić, Simon Coudé, Helena Mazurek, Jan H. Orkisz, Miriam G. Santa-Maria, Léontine Ségal, Antoine Zakardjian, Sébastien Bardeau, Karine Demyk, Victor de Souza Magalhães, Javier R. Goicoechea, Pierre Gratier, Viviana V. Guzmán, Annie Hughes, François Levrier, Jacques Le Bourlot, Dariusz C. Lis, Harvey S. Liszt, Nicolas Peretto, Evelyne Roueff, Albrecht Sievers

Observations of ionic, atomic, or molecular lines are performed to improve our understanding of the interstellar medium (ISM). However, the potential of a line to constrain the physical conditions of the ISM is difficult to assess quantitatively, because of the complexity of the ISM physics. The situation is even more complex when trying to assess which combinations of lines are the most useful. Therefore, observation campaigns usually try to observe as many lines as possible for as much time as possible. We search for a quantitative statistical criterion to evaluate the constraining power of a (or combination of) tracer(s) with respect to physical conditions in order to improve our understanding of the statistical relationships between ISM tracers and physical conditions and helps observers to motivate their observation proposals. The best tracers are obtained by comparing the mutual information between a physical parameter and different sets of lines. We apply this method to simulations of radio molecular lines emitted by a photodissociation region similar to the Horsehead Nebula that would be observed at the IRAM 30m telescope. We search for the best lines to constrain the visual extinction A_V^{tot} or the far UV illumination G_0. The most informative lines change with the physical regime (e.g., cloud extinction). Short integration time of the CO isotopologue $J = 1 - 0$ lines already yields much information on the total column density most regimes. The best set of lines to constrain the visual extinction does not necessarily combine the most informative individual lines. Precise constraints on G_0 are more difficult to achieve with molecular lines. They require spectral lines emitted at the cloud surface (e.g., [CII] and [CI] lines). This approach allows one to better explore the knowledge provided by ISM codes, and to guide future observation campaigns.

Our Goal: *to* ***evaluate the full constraining power*** *of a (combination of) tracer(s) with respect to physical conditions*

Selected quantitative criterion: **mutual information (MI) from information theory**

- accounts for **non-linear** relationships
- value **independent from** an algo class (neural network, random forest, etc.)
- allows to analyze **many-to-many** relations (many physical conditions, many tracers)
- includes the **noise model** of the considered instrument

Produced new tools

Understand relation between ISM tracers and physical conditions

Tool 1
Evolution of MI with **integration time**

Tool 2
MI **maps**

Help optimizing observation proposals

Tool 3
selection of **most informative single line** at given S/N

Tool 4
selection of **most informative combination of K lines** at given S/N