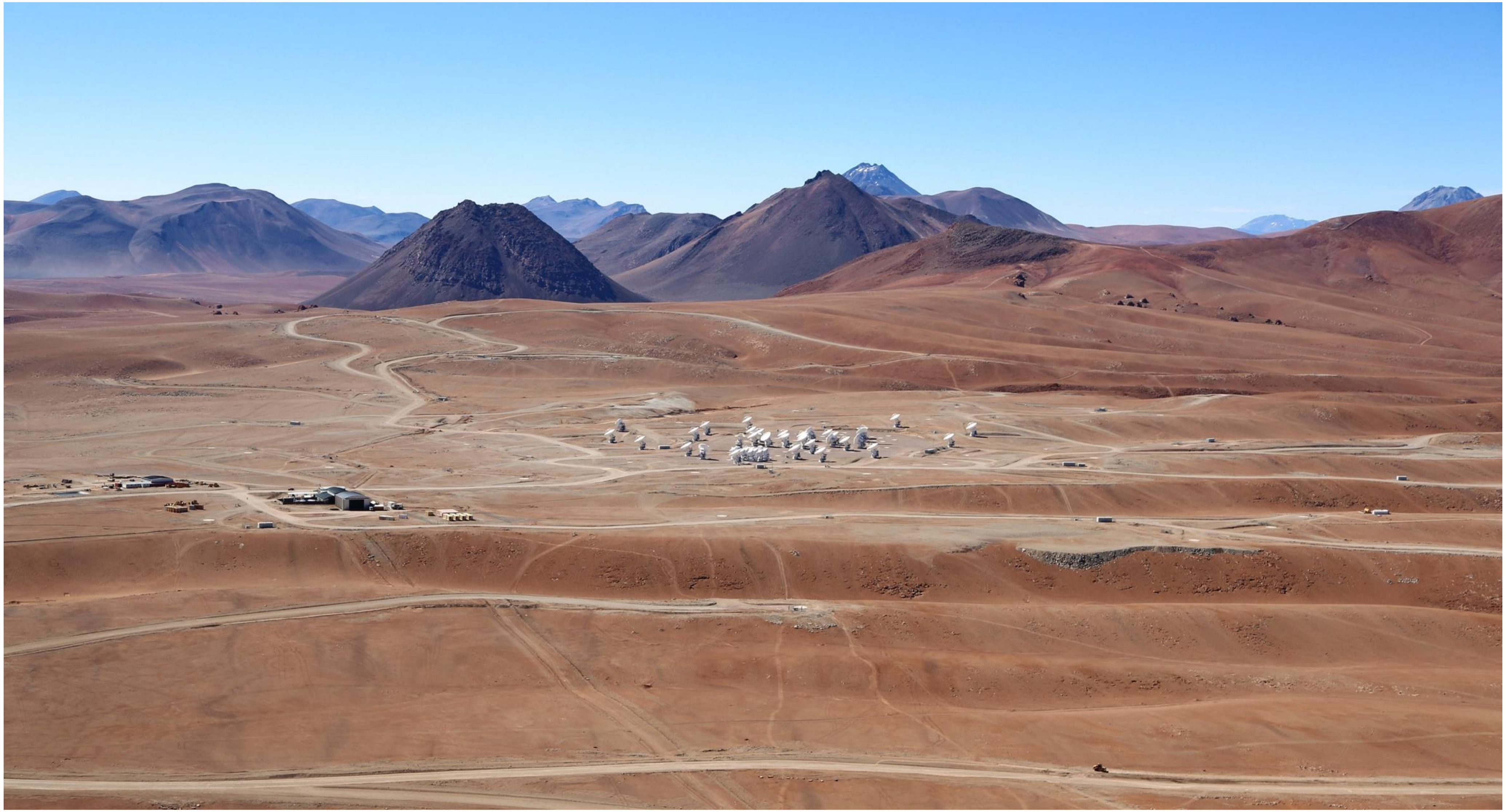
Imaging beyond CLEAN

UV fitting and AI: modern ALMA interferometric analysis

Fabrizia Guglielmetti (ESO) & Andreas Popp (ESO)

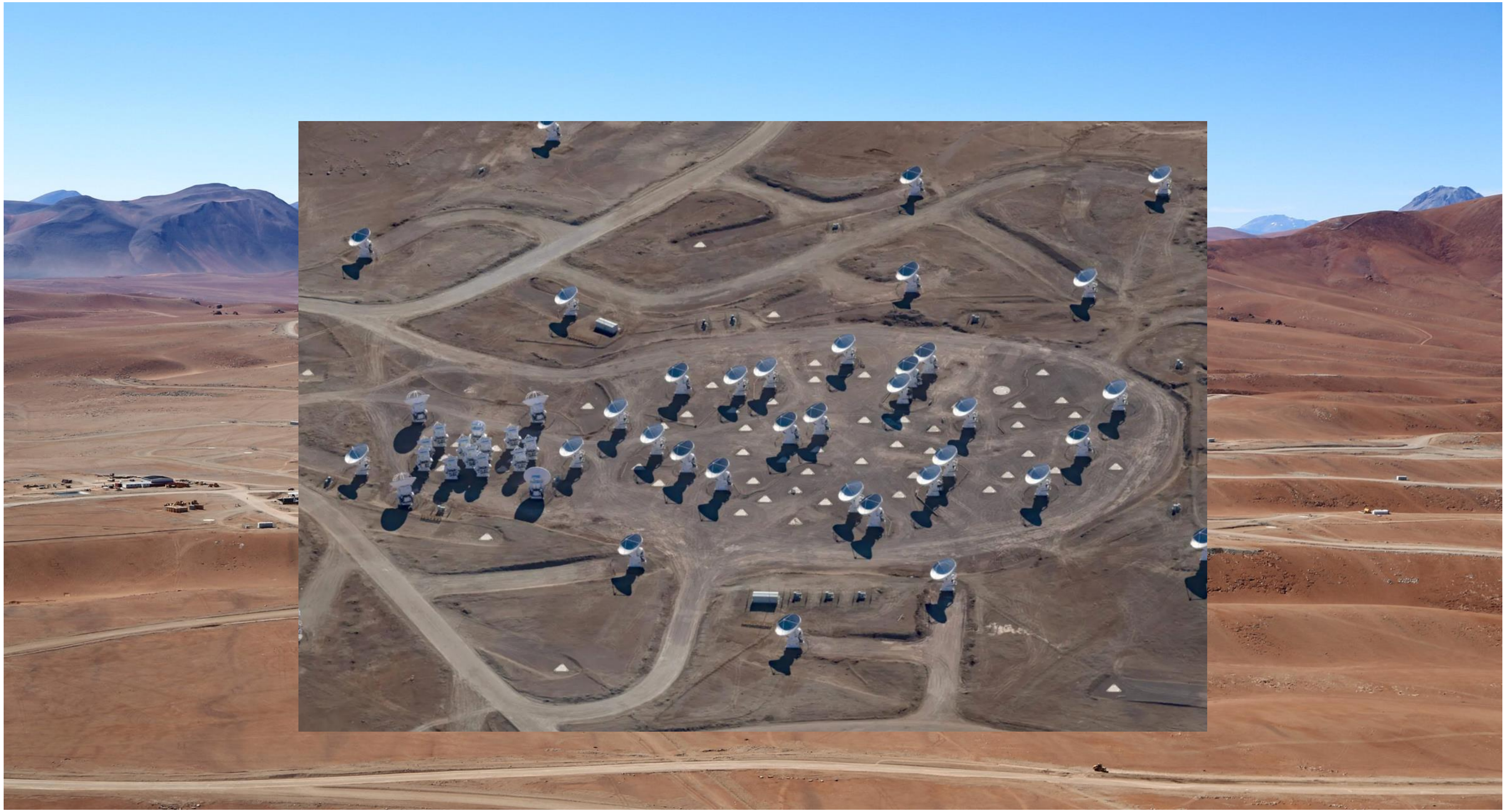
The Atacama Large mm/sub-mm Array

35-950 GHz



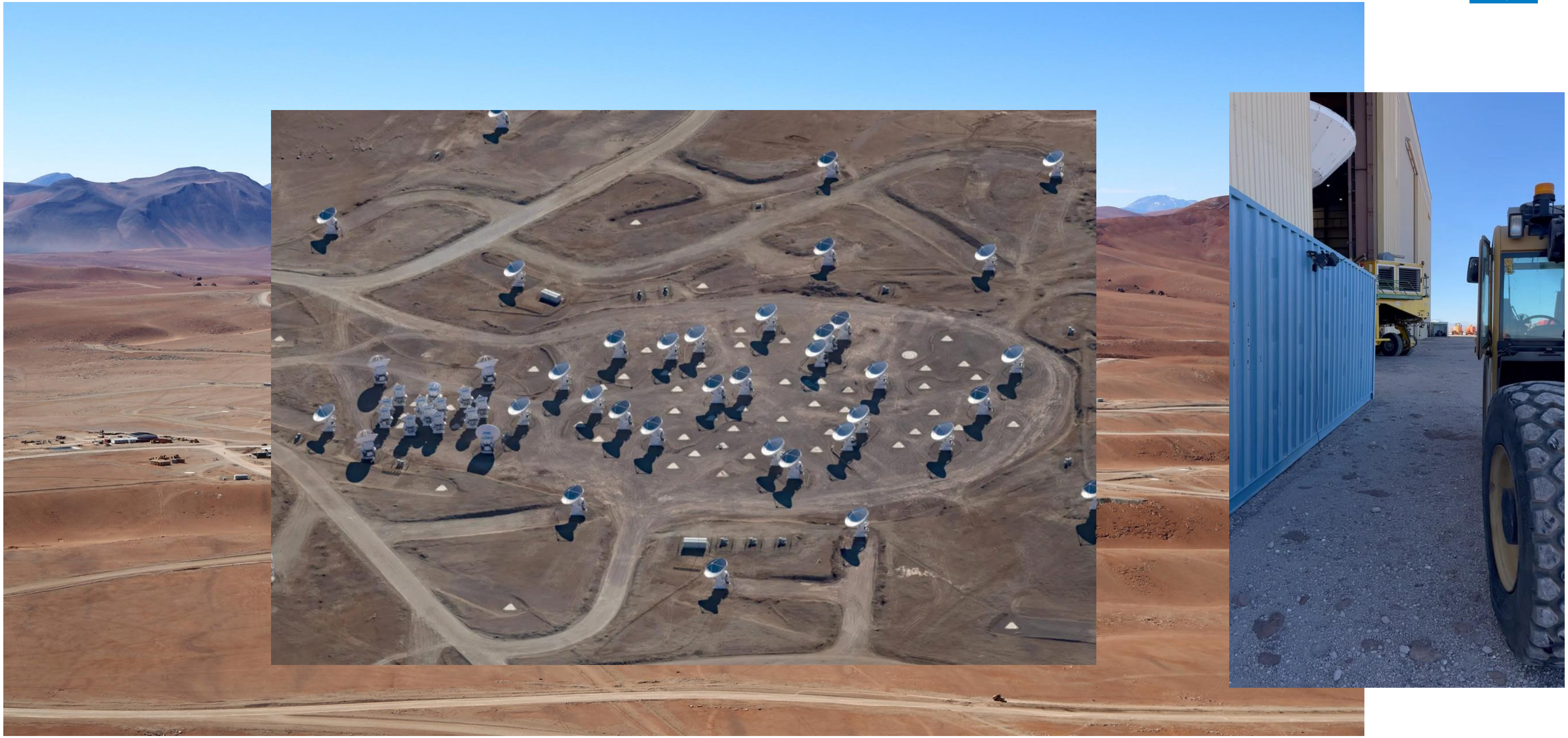
The Atacama Large mm/sub-mm Array

35-950 GHz



The Atacama Large mm/sub-mm Array

35-950 GHz

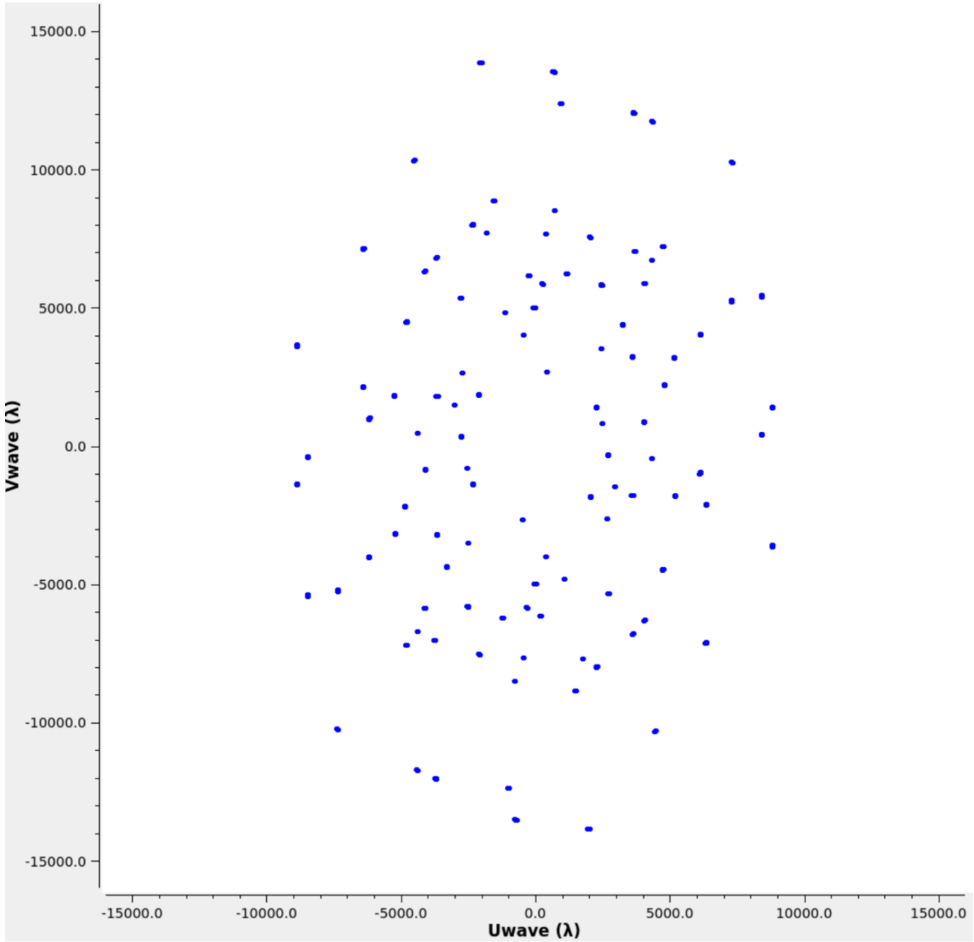


What does ALMA (IF) measure

No direct sky brightness distribution

- ALMA does **not** measure images
  - It measures **VISIBILITIES**
- Sampling is **incomplete** (uv-coverage)
- **Noise, calibration-errors, missing short spacing**
- Imaging is an **ill-posed inverse problem**

Units in the uv-plane are λ because interferometry is fundamentally wave-based.

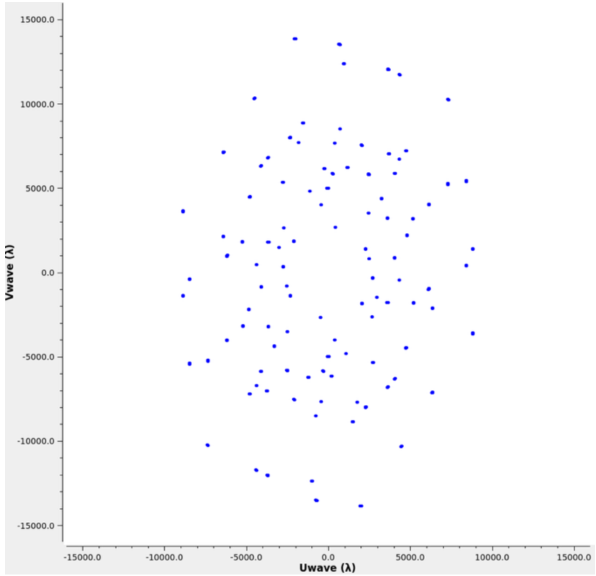


uv-space

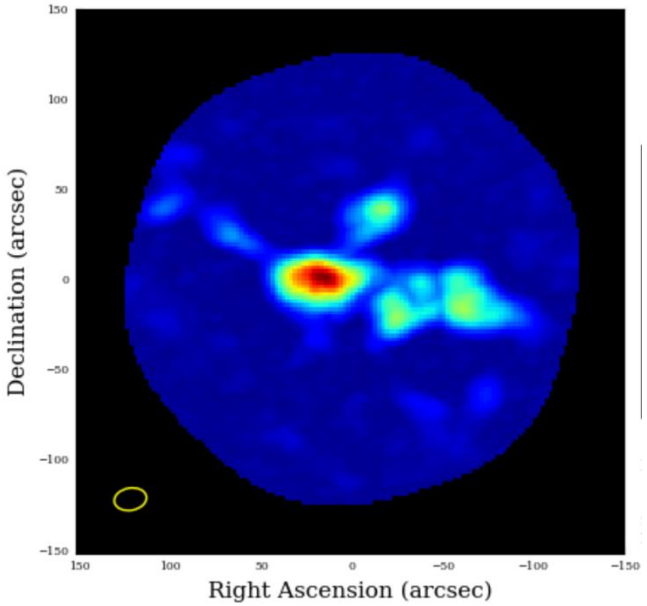
From Visibilities to Images

No direct sky brightness distribution

incomplete sampled data of the Fourier Transform of the sky x Sampling Function



$$\vec{V} = [A]\vec{I}$$

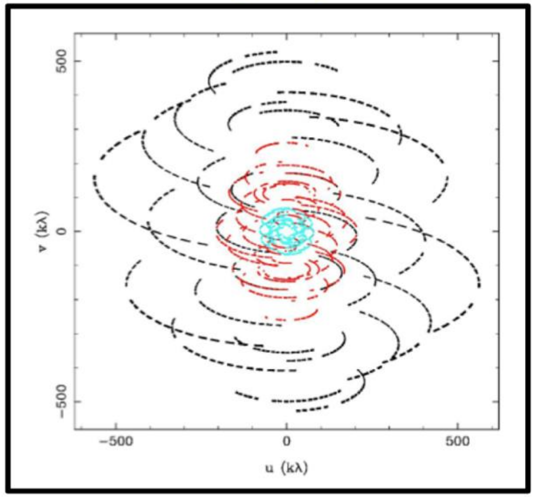


- **V= Calibrated Visibility Data**
- [A] = Measurement operator
- I = Model of the sky brightness distribution

uv-space

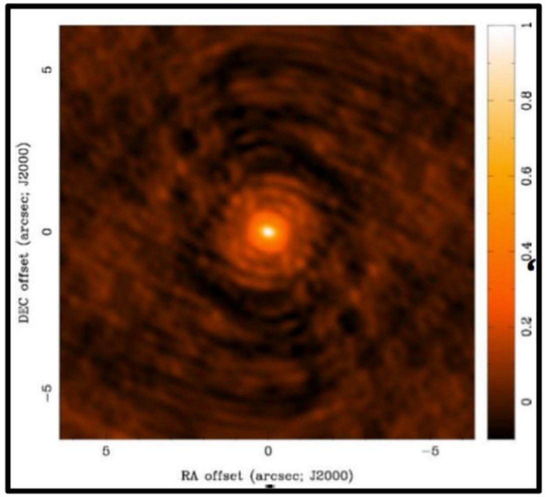
Image-space

S(u,v)



FT

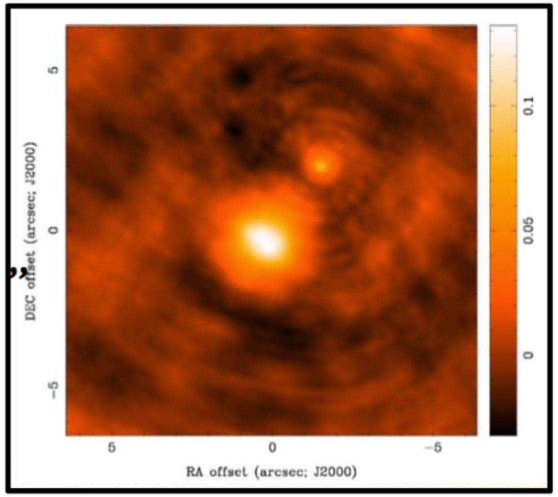
s(l,m)



Dirty Beam

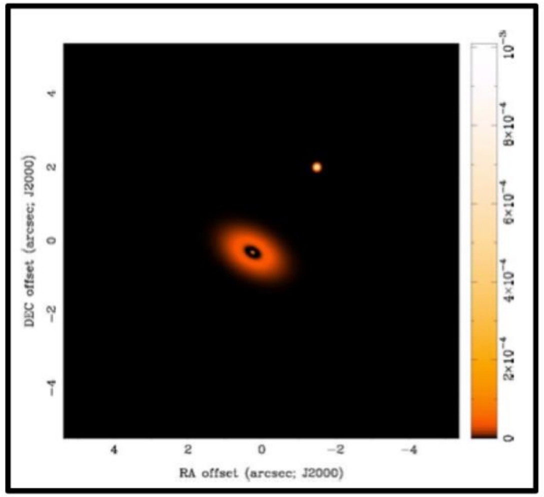
* (Convolution)

Dirty Image



Deconvolution

D(l,m)

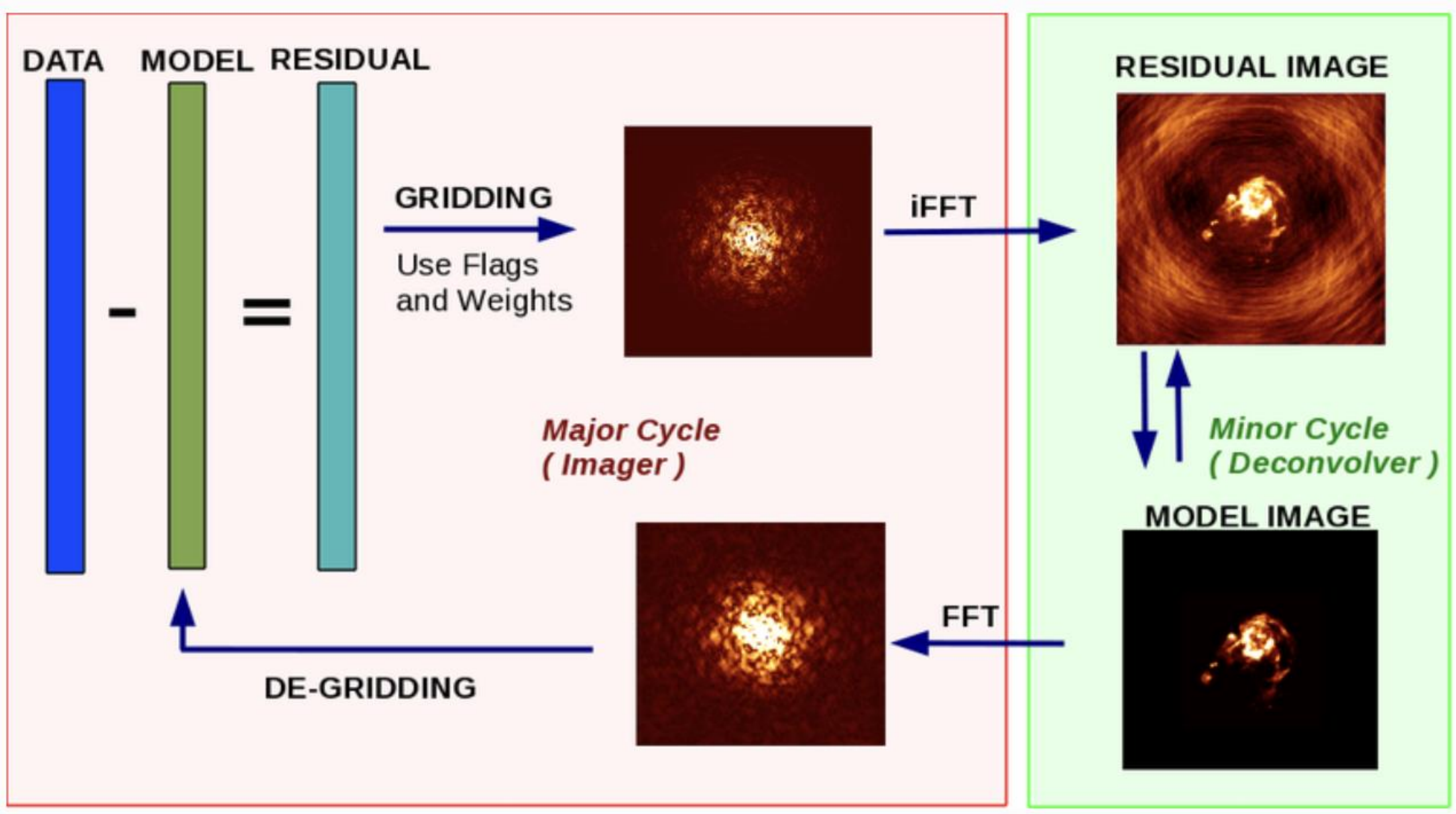


T(l,m)

Synthesis Imaging in a classical way

Iterative Image Reconstruction

Casa Doc on Synthesis imaging



ALMA image analysis beyond CLEAN algorithm

CLEAN (Högbom,J.A., A&A Suppl., 15, 417, 1974) was developed in an era when imaging was primarily limited by **computing power**

- Sky ≈ sum of point sources
- Iterative deconvolution

In modern interferometry, the dominant limitations have shifted toward

- **model fidelity and statistical rigor**
- **computational constraints remain significant** due to the scale of contemporary datasets

ALMA image analysis beyond CLEAN algorithm

CLEAN strengths

- Fast on small data sets, but not efficient on large datasets
- Operationally simple
- Operationally mature

CLEAN Limitations

- Heuristic
- Weak uncertainty estimates
- Super-resoution not statistically defined

Pipelines rely on CLEAN because it is reliable.

But science increasingly needs uncertainties, not just images.

ALMA image analysis beyond CLEAN algorithm

ALMA upgrades (e.g., *Gonzalez, A. et al., 13102, SPIE, 2024)* are driving a paradigm shift in IF data analysis:

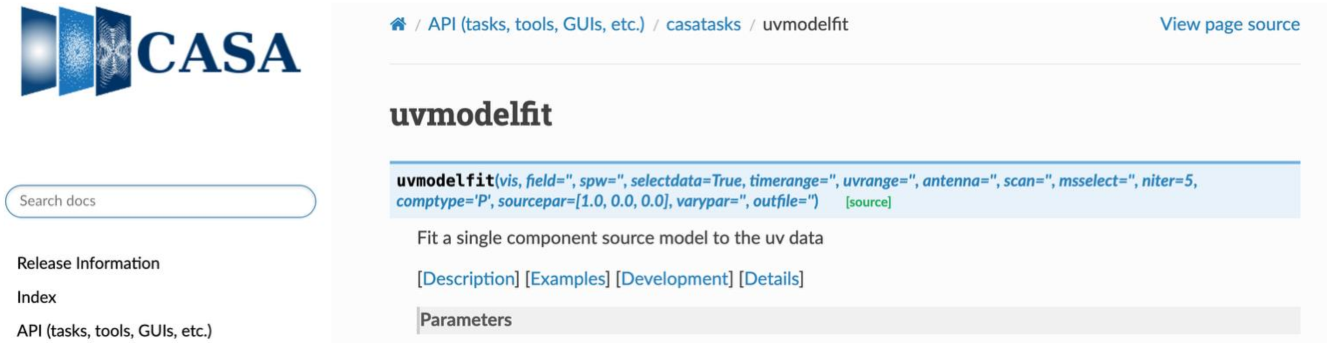
- Model-based analysis in the UV plane
- Super-resolution and inference-driven imaging methods
- Machine Learning approaches to handle scale, complexity, and discovery

Beyond CLEAN

UV Fitting : model-driven analysis

ALMA image analysis beyond CLEAN algorithm

- Fit models **directly to visibilities**
- Parametric source descriptions
- Avoids imaging artifacts
- Naturally respects what ALMA actually measures



Note

INFO: The Nordic ALMA Regional Center Node has developed tools for fitting multiple components of any shape to the visibilities. Their versatile uvmultifit [1] package is provided on their software tools page.

casadocs uvmodelfit

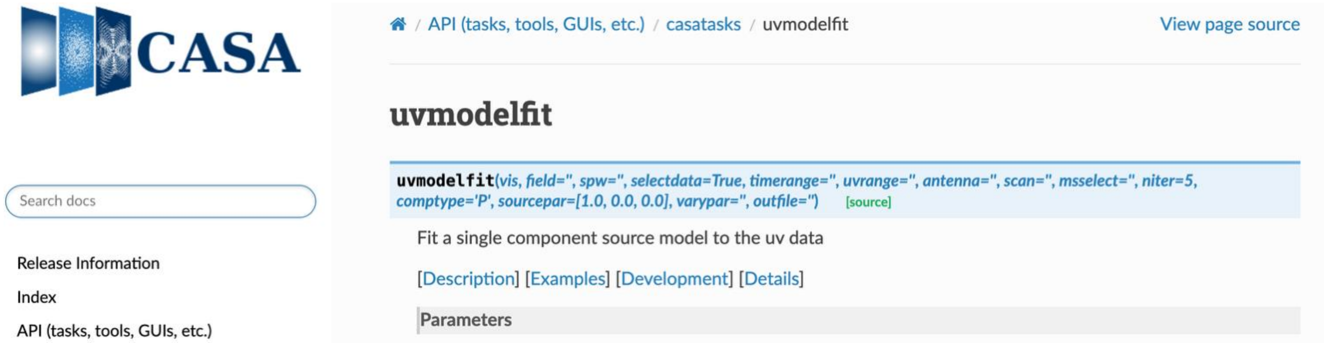
casadocs uv_manipulation

- ✓ Skip imaging using directly the data
- ✓ Noise is well-defined in uv-plane
- ✓ Works best when you know what you are looking for

UV Fitting : model-driven analysis

ALMA image analysis beyond CLEAN algorithm

- Fit models **directly to visibilities**
- Parametric source descriptions
- Avoids imaging artifacts




```
# Note: It's best to channel average the data if there are many channels before running a modelfit
split('ngc5921.ms','1445+099_avg.ms', datacolumn='corrected',field='1445*',width='63')

# Initial guess is that it's close to the phase center and has a flux of 2.0 (a priori we know it's 2.4
uvmodelfit('1445+099_avg.ms', # use averaged data
           niter=5, # Do 5 iterations
           comptype='P', # P=Point source, G=Gaussian, D=Disk
           sourcepar=[2.0,.1,.1], # Source parameters for a point source
           spw='0',
           outfile='gcal.cl') # Output component list file
```

UV Fitti

ALMA image

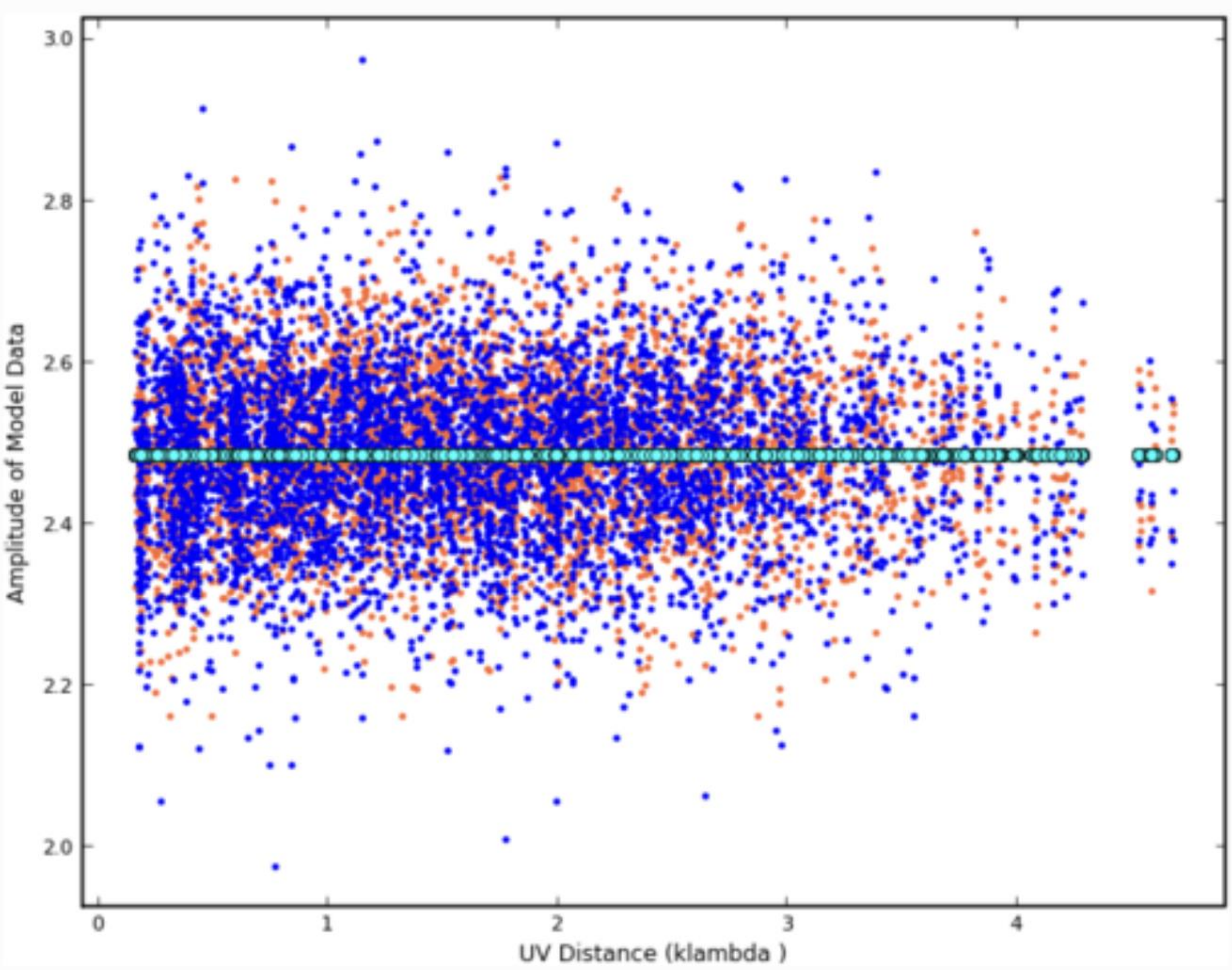
- Fit models
- Parametric
- Avoids ima

```
# Note: It's b                                   a modelfit
split('ngc5921

# Initial gues                                   we know it's 2.4
uvmodelfit('14
          nit
          com
          sou
          spw
          out
```

View page source

ntenna='', scan='', msselect='', niter=5,



corrected data (red and blue points) and uv model fit (green circles)

UV Fitting : model-drive

ALMA image analysis beyond CLEAN algo

A&A 563, A136 (2014)
DOI: 10.1051/0004-6361/201322633

Astronomy & Astrophysics

UVMULTIFIT: A versatile tool for fitting astronomical radio interferometric data

I. Martí-Vidal, W. H. T. Vlemmings, S. Muller, and S. Casey

European ALMA Regional Center (Nordic Node), Onsala Space Observatory (Chalmers University of Technology), Observatorievägen 90, 43992 Onsala, Sweden
e-mail: mivan@chalmers.se

Received 9 September 2013 / Accepted 22 January 2014

ABSTRACT

Context. The analysis of astronomical interferometric data is often performed on the images obtained after deconvolving the interferometer's point spread function. This strategy can be understood (especially for cases of sparse arrays) as fitting models to models, since the deconvolved images are already non-unique model representations of the actual data (i.e., the visibilities). Indeed, the interferometric images may be affected by visibility gridding, weighting schemes (e.g., natural vs. uniform), and the particulars of the (non-linear) deconvolution algorithms. Fitting models to the direct interferometric observables (i.e., the visibilities) is preferable in the cases of simple (analytical) sky intensity distributions.
Aims. We present UVMULTIFIT, a versatile library for fitting visibility data, implemented in a Python-based framework. Our software is currently based on the CASA package, but can be easily adapted to other analysis packages, provided they have a Python API.
Methods. The user can simultaneously fit an indefinite number of source components to the data, each of which depend on any algebraic combination of fitting parameters. Fits to individual spectral-line channels or simultaneous fits to all frequency channels are allowed.
Results. We have tested the software with synthetic data and with real observations. In some cases (e.g., sources with sizes smaller than the diffraction limit of the interferometer), the results from the fit to the visibilities (e.g., spectra of close by sources) are far superior to the output obtained from the mere analysis of the deconvolved images.
Conclusions. UVMULTIFIT is a powerful improvement of existing tasks to extract the maximum amount of information from visibility data, especially in cases close to the sensitivity/resolution limits of interferometric observations.

Key words. techniques: interferometric – methods: data analysis

https://www.aanda.org/articles/aa/pdf/2014/03/aa22633-13.pdf

https://github.com/onsala-space-observatory/UVMultiFit

GPU enabled algorithms

Algorithms commonly benchmarked for ALMA data

Several well-established algorithms are widely used and benchmarked for ALMA IF analysis

- **GALARIO** https://academic.oup.com/mnras/article/476/4/4527/4867987?login=false

  Tazzari , Beaujean, Testi +2018

  - Computes synthetic visibilities directly in the **uv plane**
    - Inputs:
    - - A **model image** *or* a **parametric radial brightness profile**
    - - Observed visibilities
    - Key features:
    - - Enables **forward modelling** of IF data
    - - Well suited for **multi-wavelength** and **multi-configuration** analyses

GPU enabled algorithms

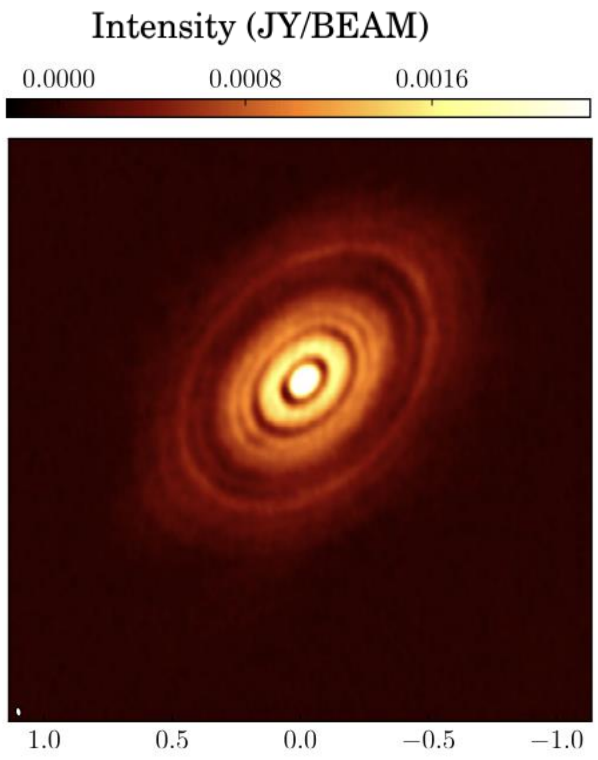
Algorithms commonly benchmarked for ALMA data

Several well-established algorithms are widely used and benchmarked for ALMA IF analysis

- **GPU-MEM** https://www.sciencedirect.com/science/article/pii/S2213133717300094

  Cárcamo, Román, Casassus, Moral, Rannou +2017

  - **Image-domain deconvolution** algorithm
  - Based on **maximum entropy regularization**
    - Key features:
    - - Produces smooth, physically plausible images
    - - Particularly effective for **extended and diffuse emission**



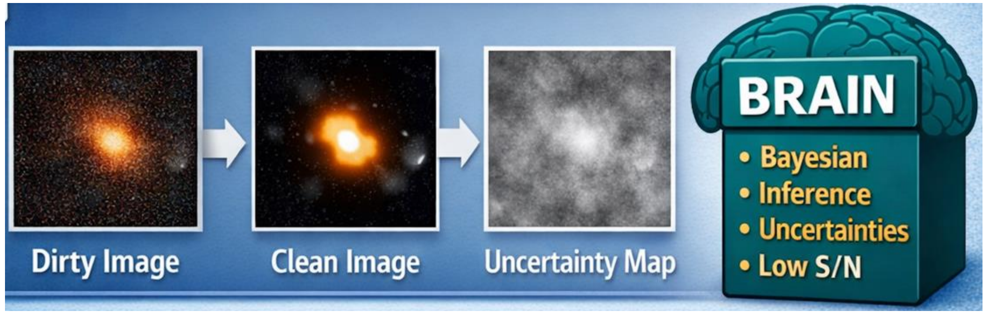
Data driven approaches

Bayesian Reconstruction through Adaptive Image Notion (BRAIN) Study

ALMA Development Studies: Motivation and Context

Why ALMA Development studies?

- Enable community-led innovation beyond standard operations
- Prepare for future scientific and technical challenges
- Test high-risk/high-gain ideas before operational adoption



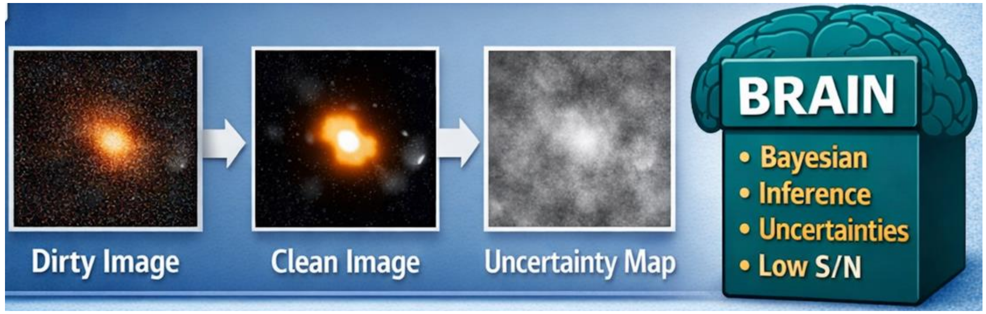
ALMA Development Studies: Motivation and Context

Why ALMA Development studies?

- Enable community-led innovation beyond standard operations
- Prepare for future scientific and technical challenges
- Test high-risk/high-gain ideas before operational adoption

Challenges motivating new approaches

- Increasing sensitivity exposes faint, extended emission
- Complex source morphologies are difficult to image robustly
- Standard techniques may rely on manual tuning
- Limited propagation of uncertainties to science products



ALMA Development Studies: Motivation and Context

Why ALMA Development studies?

- Enable community-led innovation beyond standard operations
- Prepare for future scientific and technical challenges
- Test high-risk/high-gain ideas before operational adoption

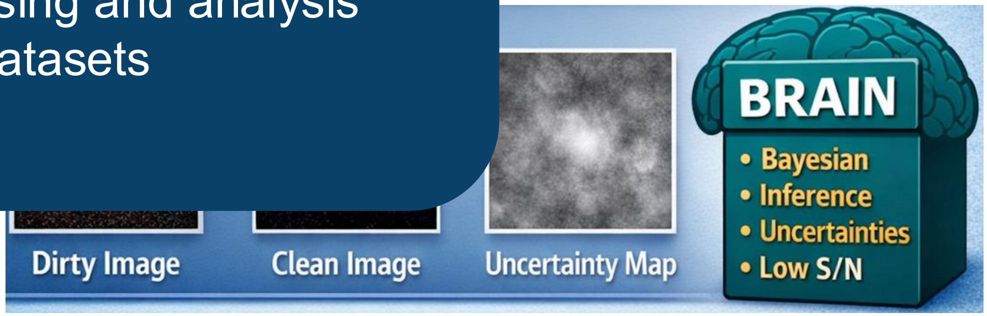
Challenges motivating new approaches

- Increasing sensitivity exposes faint, extended emission
- Complex source morphologies are difficult to image robustly
- Standard techniques may rely on manual tuning
- Limited propagation of uncertainties to science products

The BRAIN Study

- Enhance image analysis using data-driven algorithms
  - ➢ Prioritize objective insights over subjective interpretations
  - ➢ Achieve more precise and efficient image processing and analysis
- Improve computational efficiency for handling large datasets
- Automate data processing to elevate image quality

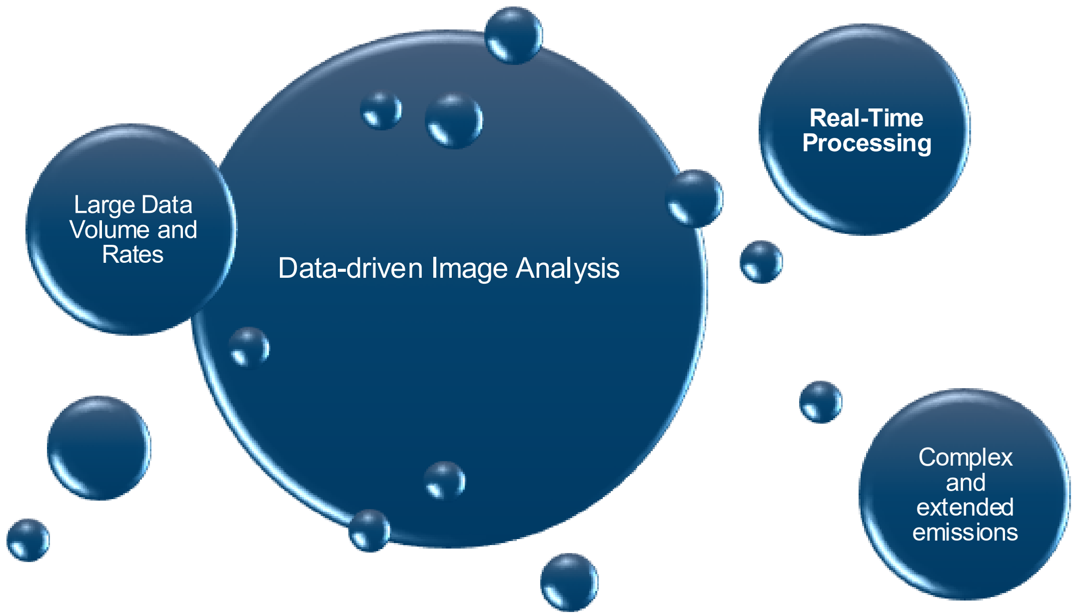
Guglielmetti+2019, Guglielmetti+2022, Guglielmetti+2023



Enhancing Data Processing in Astrophysics: The Brain Study

ALMA2030 Operational Challenges

(Gonzalez, A. et al., 13102, SPIE, 2024; Carpenter, J. et al., ALMA memo 621, 2022)



Solution Pathways:

Cutting-edge Resources and AI Innovation:

- **Expertise in AI and Astronomy:** Engage teams with dual expertise to create advanced algorithms tailored for astronomical data.
- **Enhanced Source Detection Accuracy**: AI-based approaches can refine source detection, leveraging knowledge embedded in Point Spread Functions (PSF), noise patterns, source distribution…

ALMA Simulation Tools:

- **Purpose**: Develop an ALMA simulator designed to generate datasets for AI training, testing and validation.
- **Applications**: This simulator is supported by empirical noise studies and used to validate AI models, ensuring alignment with real-world conditions in the ALMA2030 environment.

RESOLVE:

- **Methodology:**
  - ✓ Bayesian approach within IFT
  - ✓ Synthesizing radio apertures
  - ✓ Enabling precise image reconstruction
- **Key Capabilities**:
  - ✓ Low signal-to-noise diffuse emissions detection
  - ✓ Data combination
  - ✓ Self-calibration, Full Polarization imaging
- **Results**:
  - ✓ High accuracy in identifying and preserving weak celestial signals, even in noisy observational environments

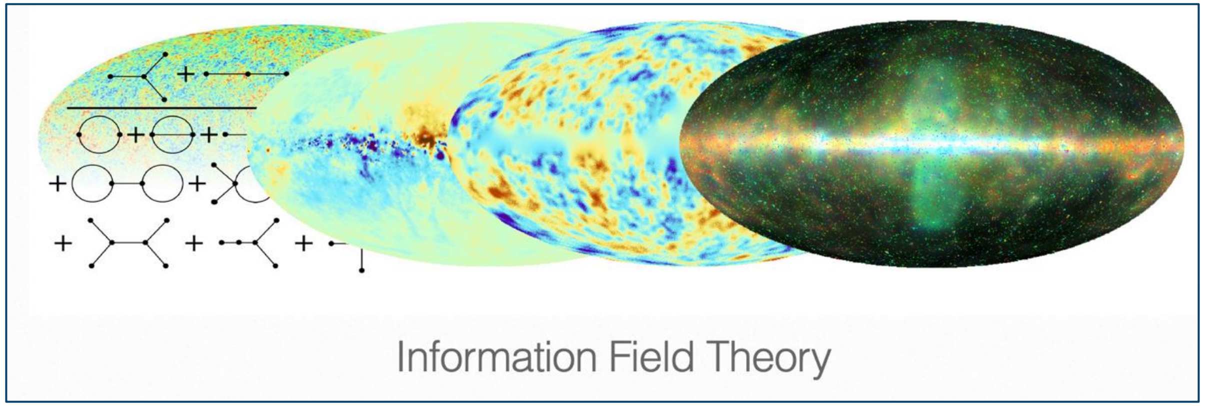
https://github.com/information-field-theory/nifty
https://ift.pages.mpcdf.de/nifty/index.html

https://github.com/NIFTy-PPL/J-UBIK
https://ift.pages.mpcdf.de/j-ubik/

https://gitlab.mpcdf.mpg.de/ift/resolve
https://ift.pages.mpcdf.de/resolve/

Resolve user guide

This guide is an overview and explains the main idea behind resolve. More details are found in the API reference.



Research, Applications, Software and Publications of the IFT group

RESOLVE:

https://github.com/information-field-theory/nifty
https://ift.pages.mpcdf.de/nifty/index.html

https://github.com/NIFTy-PPL/J-UBIK
https://ift.pages.mpcdf.de/j-ubik/

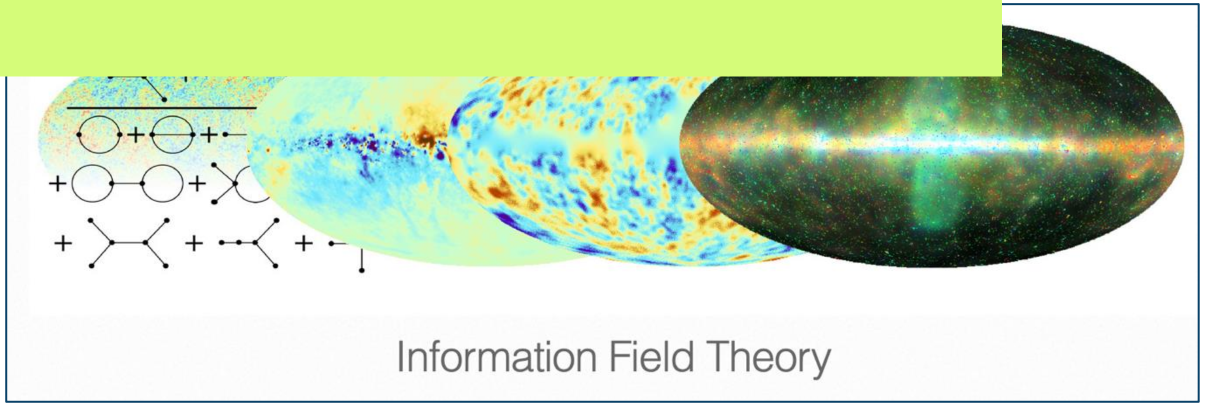
- **Methodology:**
  - ✓ Bayesian approach within IFT
  - ✓ Synthesizing radio apertures

https://gitlab.mpcdf.mpg.de/ift/resolve

- **Ke**

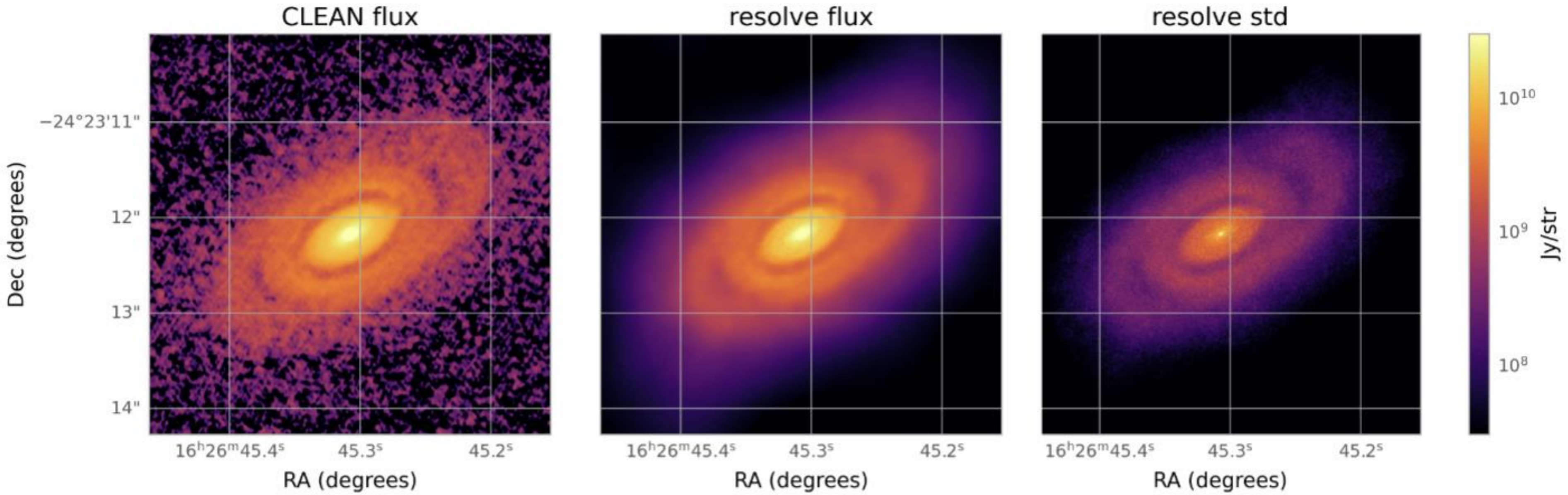
Enßlin+2009, Junklewitz+2016, Tychoniec+2022, Roth+2024, Johnson+2026, Popp+2026, Rüstig+2026

- **Results**:
  - ✓ High accuracy in identifying and preserving weak celestial signals, even in noisy observational environments



Research, Applications, Software and Publications of the IFT group

Elias 27 at 240 GHz (1.25 mm) from the DSHARP Large Program : Protoplanetary Disks



Fiducial CLEAN Image

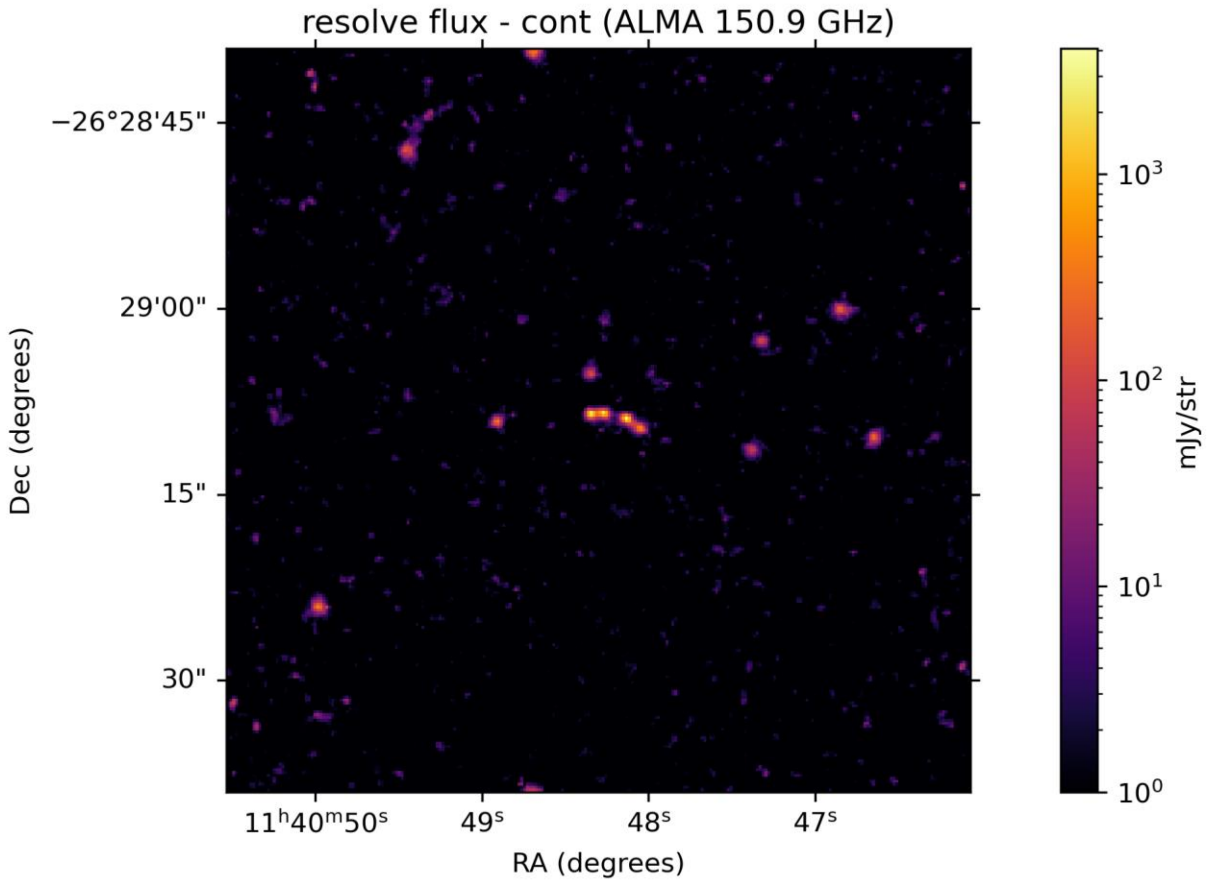
resolve posterior mean

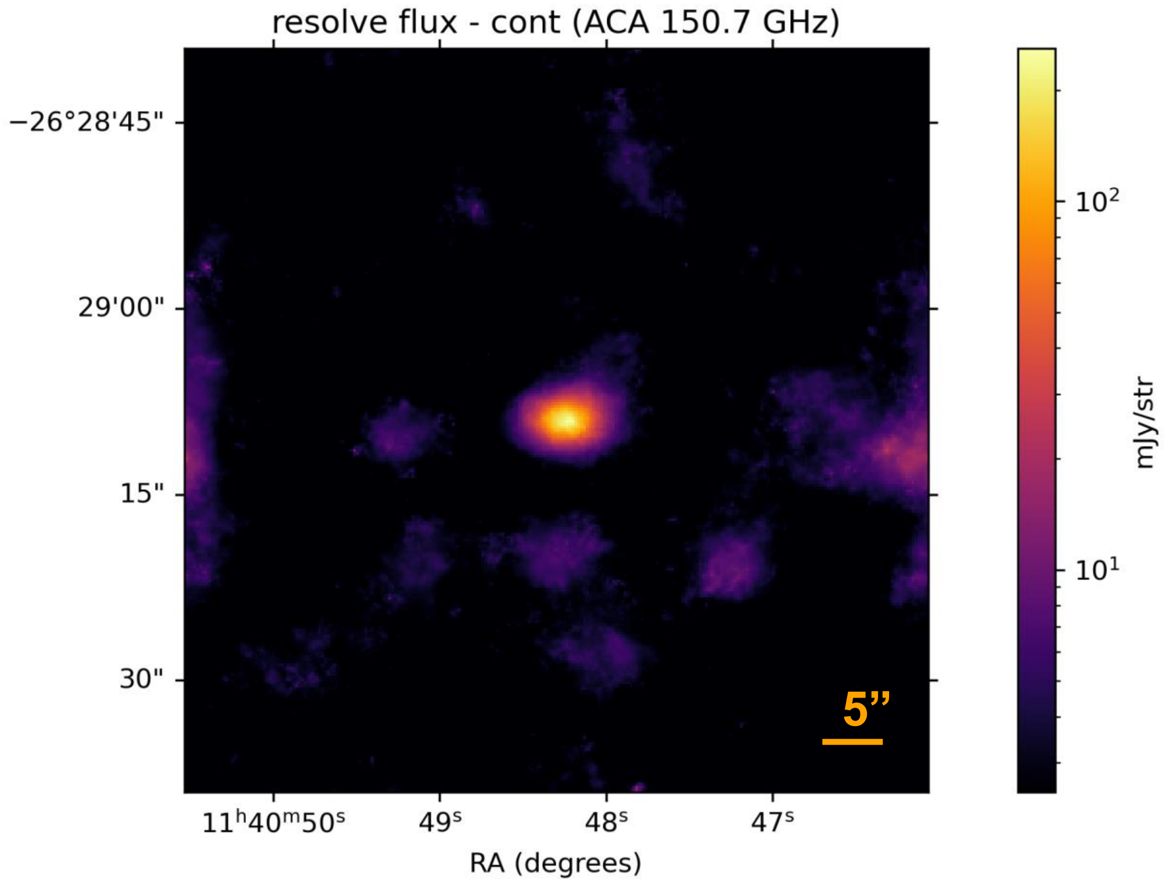
resolve standard deviation for uncertainty quantification (UQ)

Andrews et al. 2018; Huang et al. 2018; Huang et al. 2018

Johnson et al., in prep.

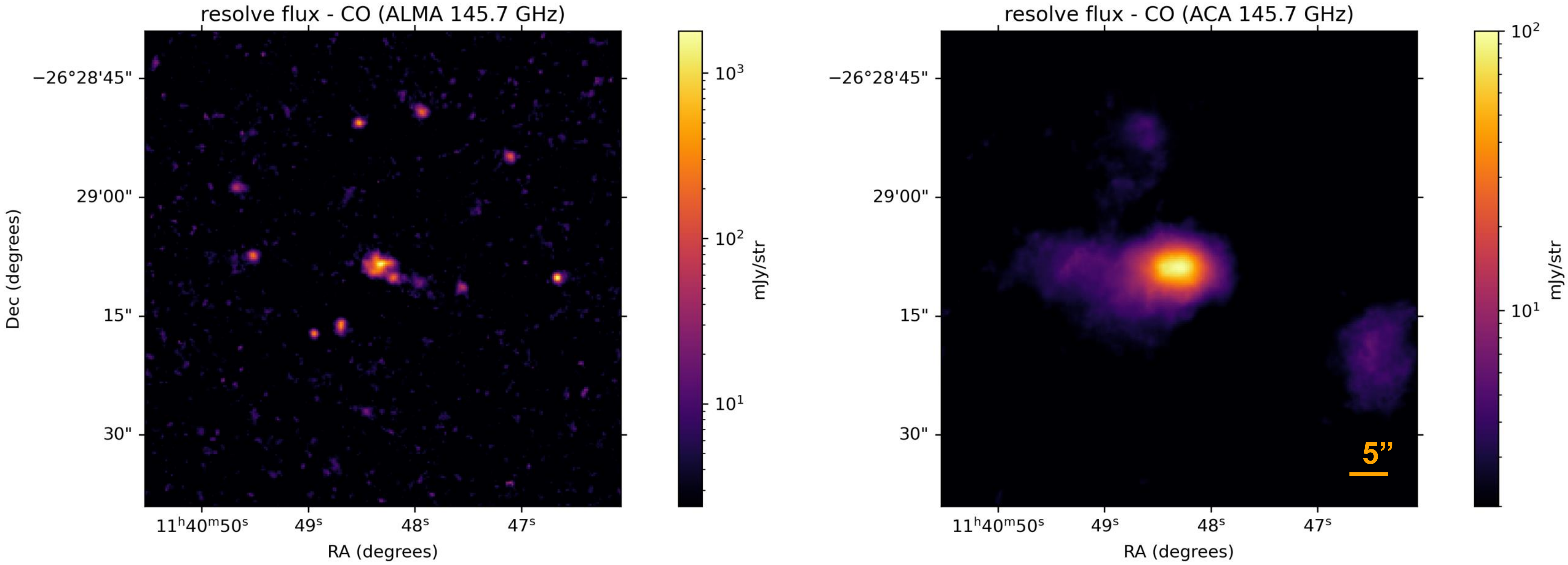
Targeting the CGM in ALMA Archival Data : continuum detection - **MRC 1138–262**





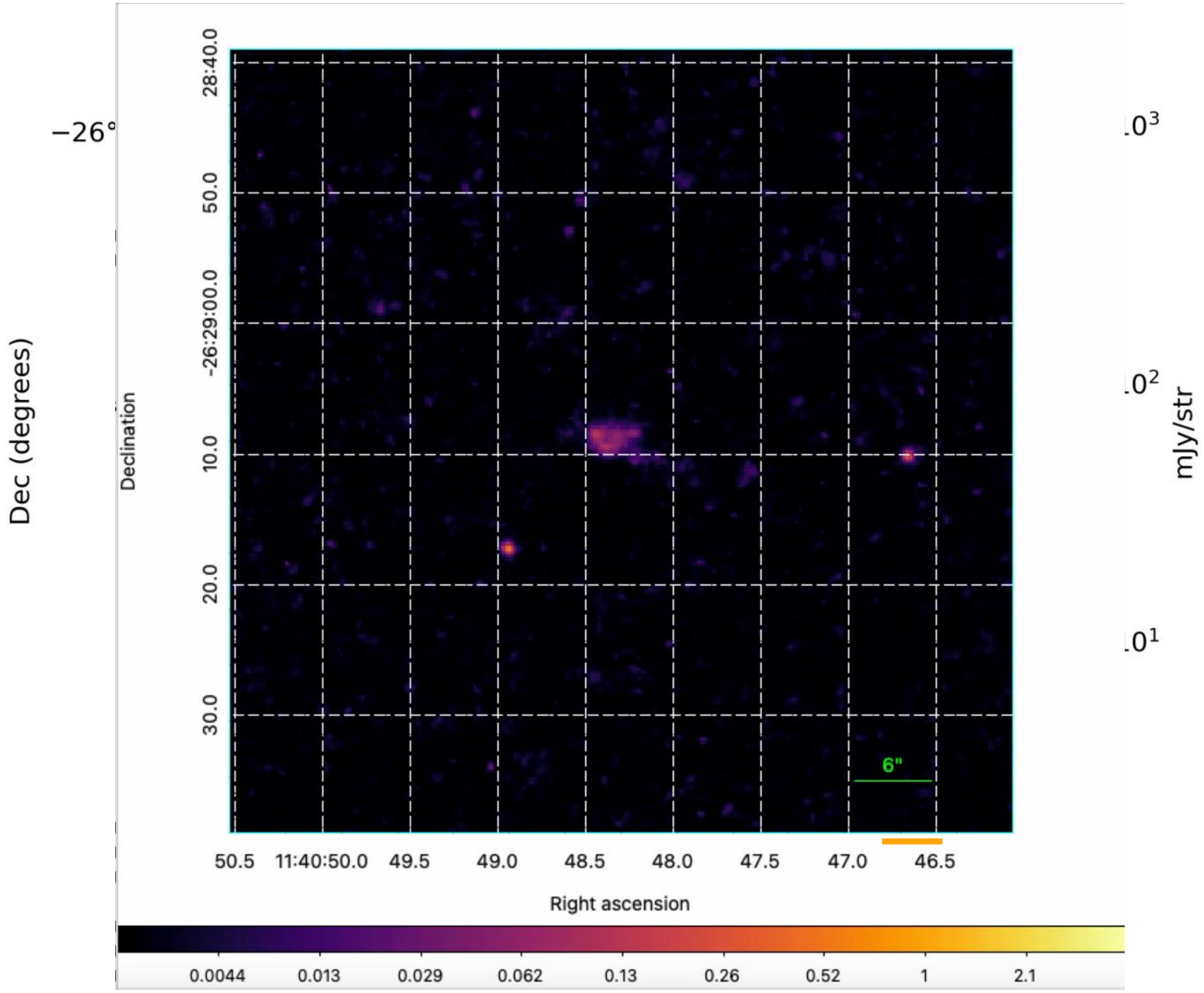
Popp et al., in prep.

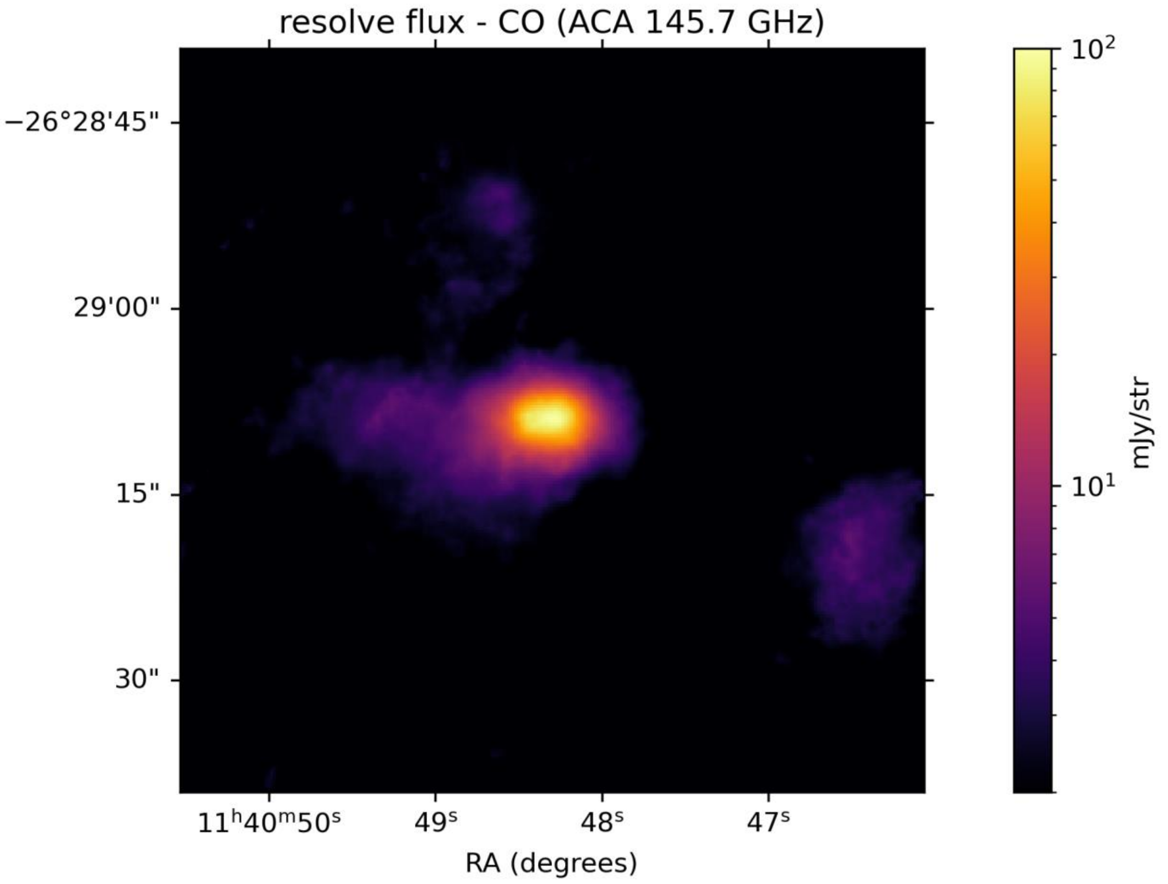
Targeting the CGM in ALMA Archival Data : Carbon Monoxide - MRC 1138–262



Popp et al., in prep.

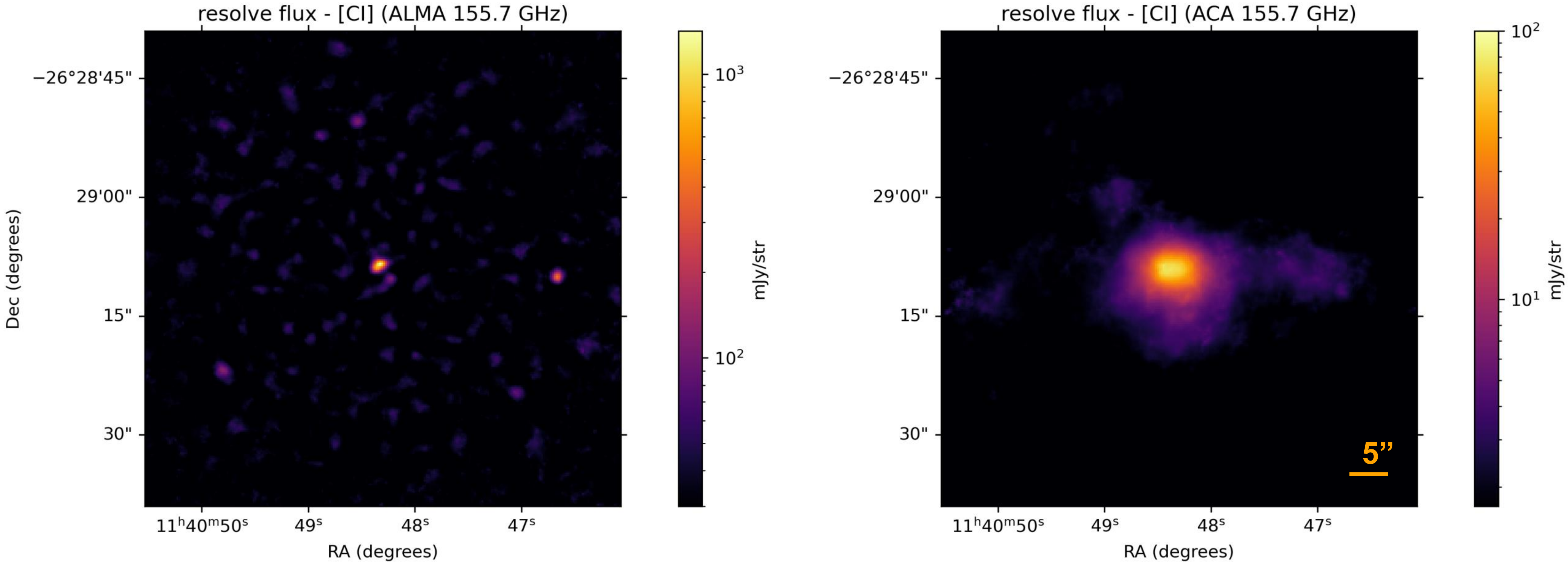
Targeting the CGM in ALMA Archival Data : Carbon Monoxide - MRC 1138–262





Popp et al., in prep.

Targeting the CGM in ALMA Archival Data : Neutral Atomic Carbon - MRC 1138–262



Popp et al., in prep.

DeepFocus:

- **Overview**:
  - ✓ **Deep Learning** architecture
  - ✓ **Meta-learner:** Taxonomy with Bayesian Surrogates
  - ✓ Designed to analyze and process **large** ALMA data cubes
- **Enhanced by Simulation**:
  - ✓ Utilizes **ALMASIM** for realistic data generation and **NOISEMPIRE** for accurate noise representation, ensuring fidelity to real ALMA observations
- **Outcomes**:
  - ✓ Real-time image analysis
  - ✓ Catalogues of extracted sources

GitHub

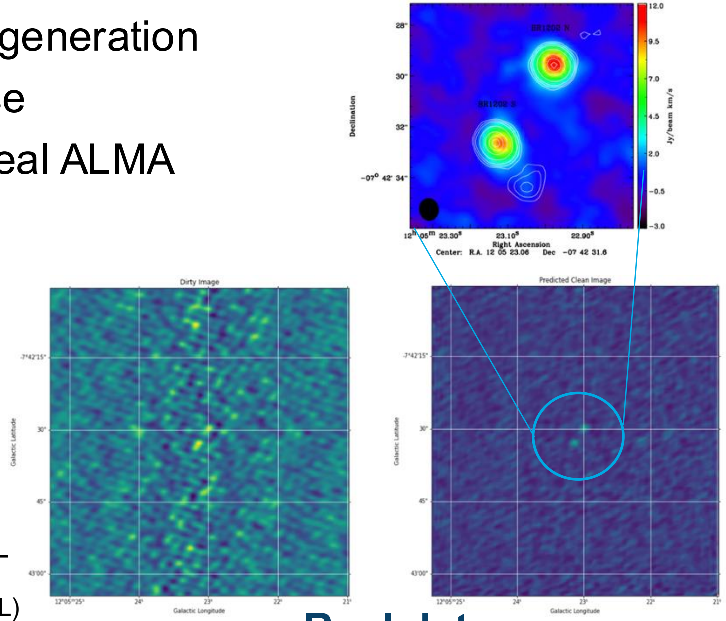
https://github.com/MicheleDelliVeneri/DeepFocus

Delli Veneri +2023

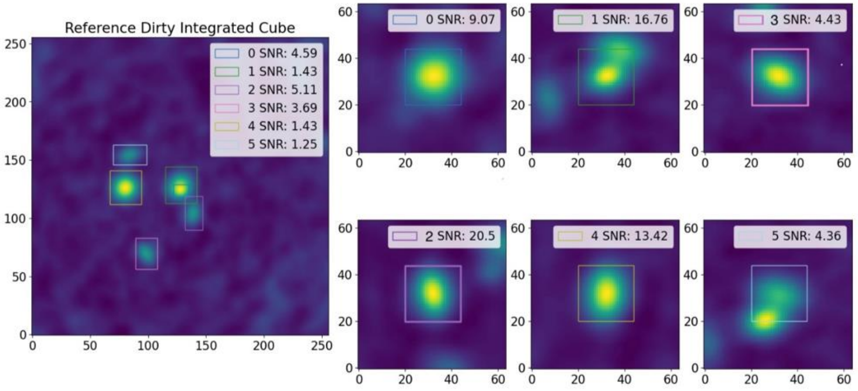
https://github.com/MicheleDelliVeneri/ALMASim

https://github.com/lbaronch/NOISEMPIRE. **Baronchelli, I.**

Carniani et al.+2013



Real data



Simulated data

INPUT DIRTY CUBE

↓

DEEP FOCUS

↓

TAXONOMY

↓

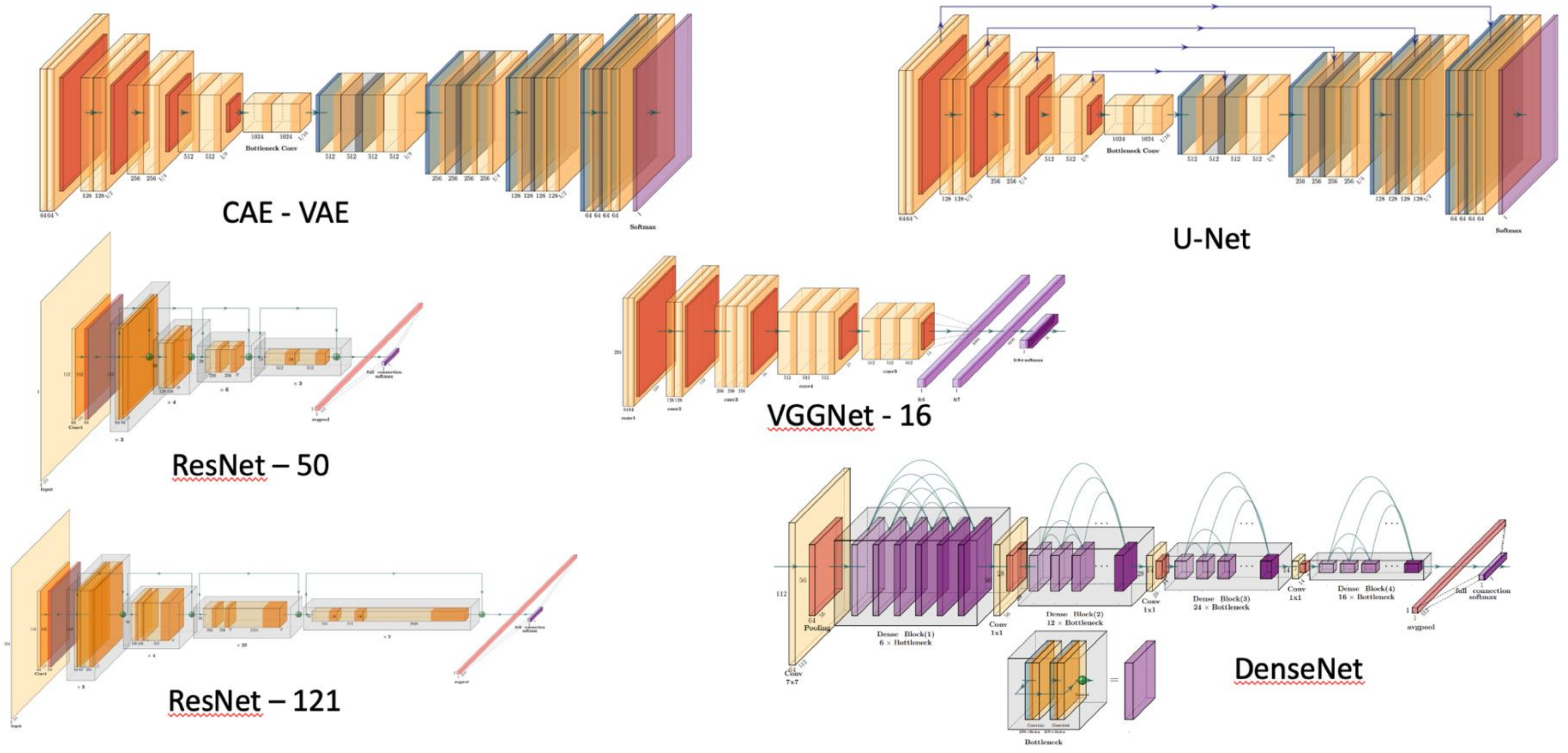
DECONVOLUTION AND DENOISING

↓

FOCUSING

↓

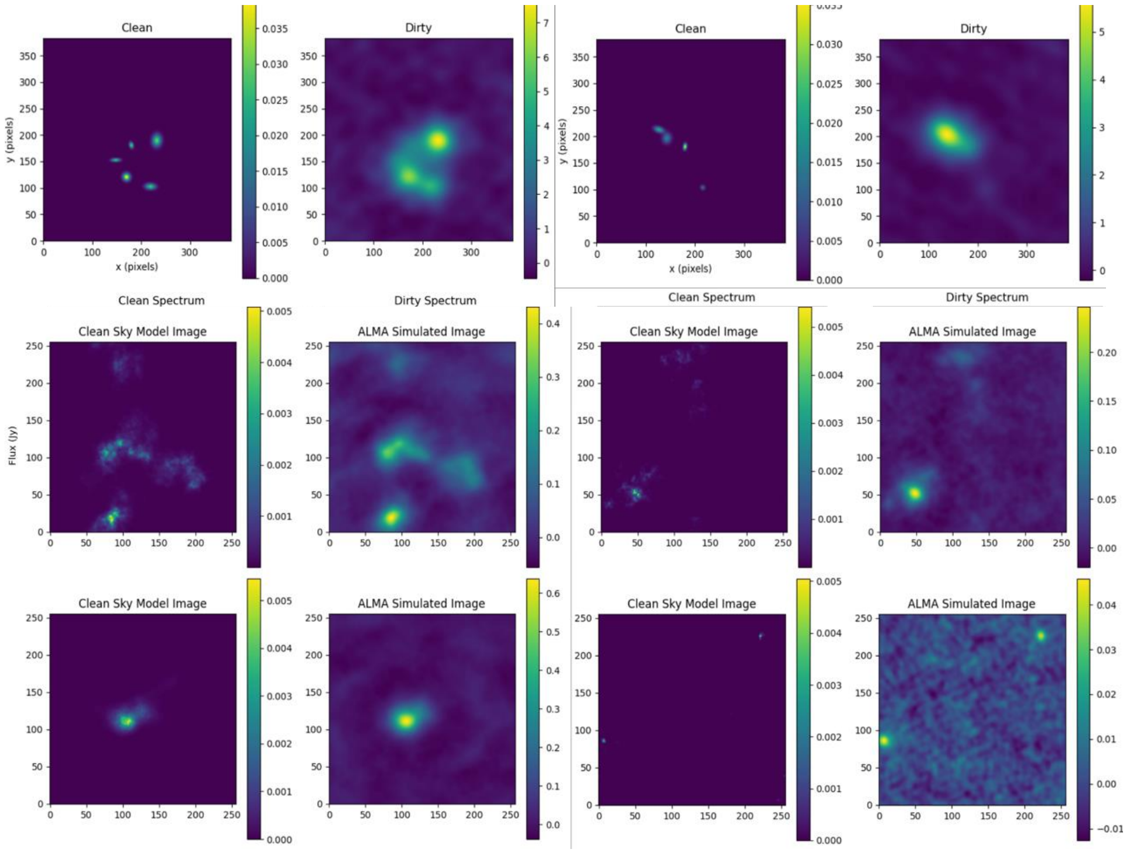
REGRESSION AND CLASSIFICATION



https://github.com/lbaronch/NOISEMPIRE
https://github.com/MicheleDelliVeneri/ALMASim.git
https://github.com/MicheleDelliVeneri/DeepFocus

ALMASim (Simulator)

- **Metadata retrieval** : query the ALMA Archive
- **Source reasoning** :
  - Extract fitting lines for the observation window
  - Calculate continuum and line fluxes
- **Sky model generation** :
  - Serendipitous sources
  - Storing source properties and observations on disk
- **Mock ALMA Observations** :
  - Accounts for atmospheric noise and gain errors
  - Save data products to disk



ALMASim: the ALMA Simulator for AI Developments

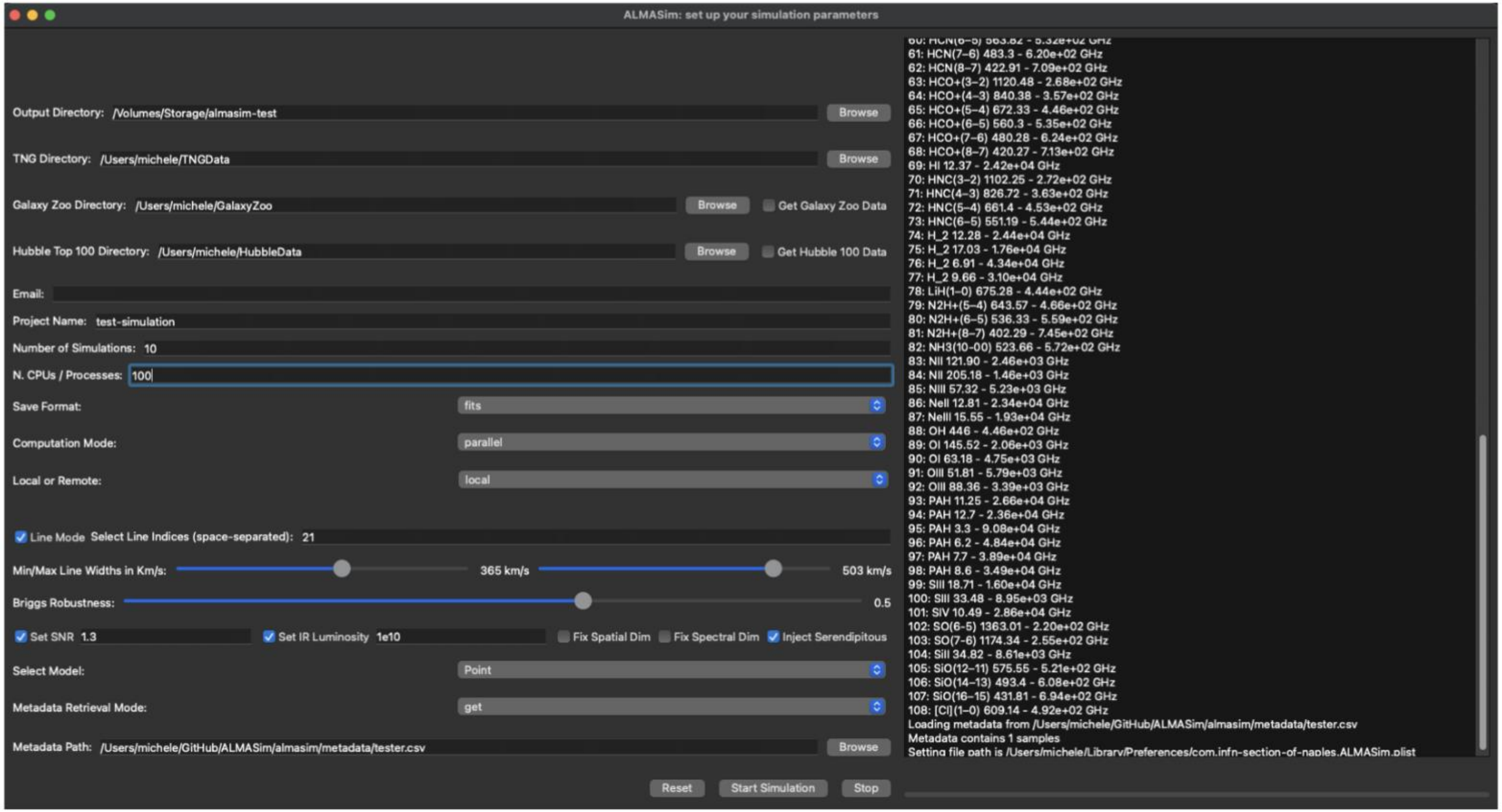
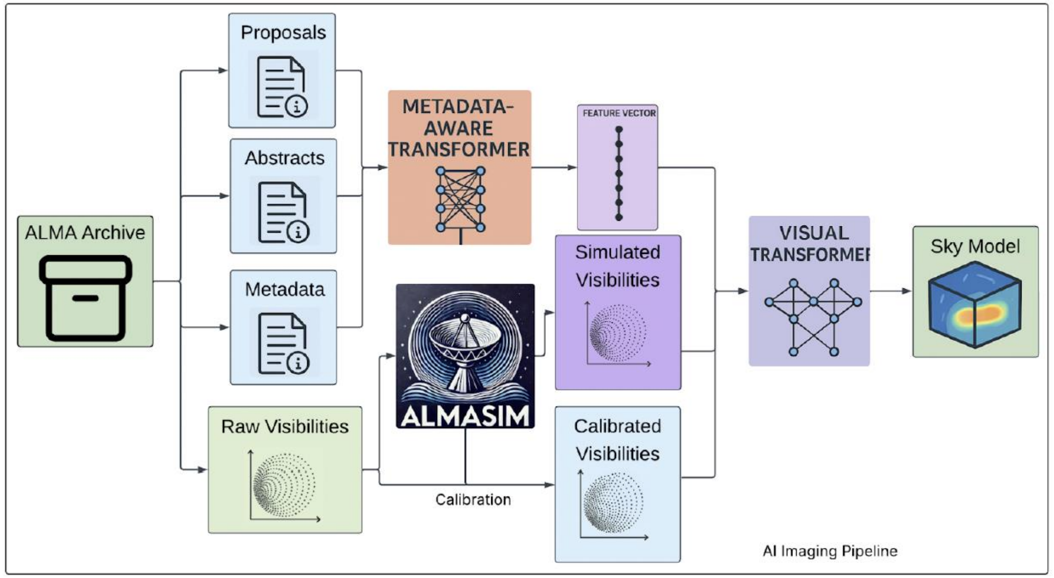


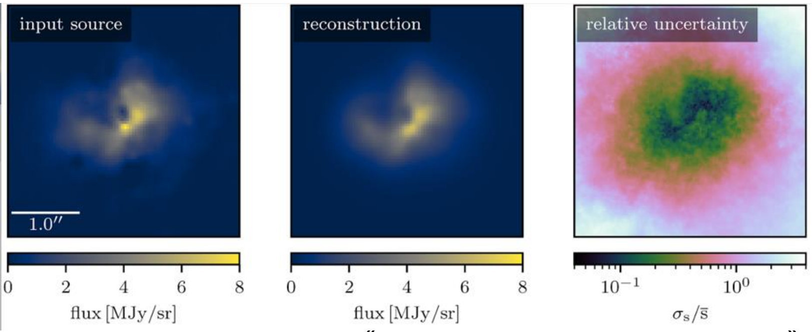
Figure 33: The ALMASim Graphical User Interface.

Deep Learning Models: Future extensions

Metadata Aware transformer for Direct Imaging from Visibilities

- *Visual Transformer architecture as deconvolver/imager operating on visibility domain*
- *Metadata-Aware Transformer as prior knowledge on signal distribution*
- *ALMASim will learn from ALMA Archival data*
- *Generative models will be used to create simulated data*
- **Applications to ALMA Archival data**



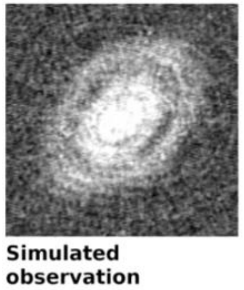


Rüstig+2024

RESOLVE

IFT + UQ
Improves
accuracy in
ALMA imaging

Baronchelli+2024



NOISEMPIRE 1.0.3 Noce

Empirical noise simulator

Simulates pure noise 2D images given a real ALMA cube or image in input. A real sky image (same size as reference image) can also be added.

NOISEMPIRE

Separates noise
levels and
patterns.
Real-time ALARM
Systems for
Observatory's
health?

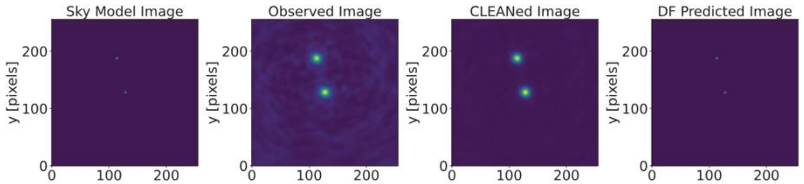


Fig. 1. – From left to right: Sky Model Image, Observed Image, CLEAN image, DF Prediction of the Sky Model.

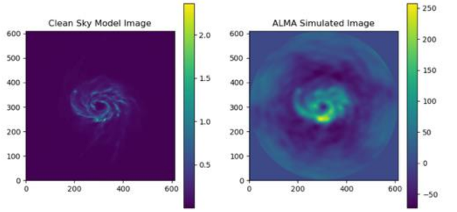
Delli Veneri+2024

DeepFocus

Meta Learner
Enables
Real-time
ALMA imaging
+ cataloguing

ALMASim

Encodes +100
Molecular
transitions at
several z, HPC
enabled



Delli Veneri, ALMASim documentation

ESO/ ARC


F. Guglielmetti (PI)


E. Villard

ESO FELLOWS
L. Tychoniec
G. Cosentino

Felix Stoehr
(ALMA Archive SSS)

MPA


T. Enßlin


P. Arras


J. Roth

Julian Rüstig


A. Popp

UniNa


M. Delli Veneri


M. Brescia


G. Longo

Alvi Rownock

Luca Sannino

Jonseo Kim


V. Johnson

INFN IRA/ ARC


I. Baronchelli


M. Bonato

IFT GROUP & DZA Head of Data Science

https://wwwmpa.mpa-garching.mpg.de/~ensslin

https://www.deutscheszentrumastrophysik.de

https://wwwmpa.mpa-garching.mpg.de/~ensslin/Bayes_Forum/bayes_forum.html

Machine Learning

http://datascience.unina.it/

The core team

Conclusion on Imaging beyond CLEAN

AI = Modern shift in Data Analysis	✓Objectivity above subjectivity ✓Model fidelity and statistical rigor ✓Data-driven approaches to overcome common limitations
RESOLVE	✓Improved the detection of faint signals, expanding the frontiers of astronomical discovery ✓Enhanced uncertainty quantification for more reliable results
DeepFocus	✓Real-time image analysis for immediate insights ✓Automated identification of interesting features, cataloguing
ALMASim	✓Enabled supervised ML approaches for more accurate modeling ✓Generated more realistic simulations for training AI models
Future	❑Improve AI techniques to image analysis, popularize to community, intensify data mining

Thank you!

the BRAIN Study Team

fgugliel@eso.org

https://gitlab.mpcdf.mpg.de/ift/resolve

https://github.com/NIFTy-PPL

https://github.com/MicheleDelliVeneri/DeepFocus

https://pypi.org/project/ALMASim

https://github.com/lbaronch/NOISEMPIRE

@ESOAstronomy

@esoastronomy

@ESO

european-southern-observatory

@ESOobservatory