# Signal Processing of High-z Galaxies from IFU Spectral Data Cubes:
## Denoising and Data-Driven Physical Mapping Beyond Power-Law Relations

**Arnab Lahiry**

Supervised by:
**Dr. Tanio Diaz-Santos**
**Dr. Jean-Luc Starck**

13/02/2026

**Deep CosmoStat Days**, DaP, CEA 📍 **Paris-Saclay**, FR

UNIVERSITY OF CRETE
DEPARTMENT OF PHYSICS
FORTH INSTITUTE OF ASTROPHYSICS
CosmoStat
TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS
Signal Processing Laboratory

*Astronomy & Astrophysics* manuscript no. paper
February 11, 2026



# Deep and Sparse Denoising Benchmarks for Spectral Data Cubes of High-z Galaxies: From Simulations to ALMA observations

Arnab Lahiry[1,2,⋆], Tanio Díaz-Santos[1], Jean-Luc Starck[1,3], Niranjan Chandra Roy[4], Daniel Anglés-Alcázar[4], Grigorios Tsagkatakis[1,2], and Panagiotis Tsakalides[1,2]

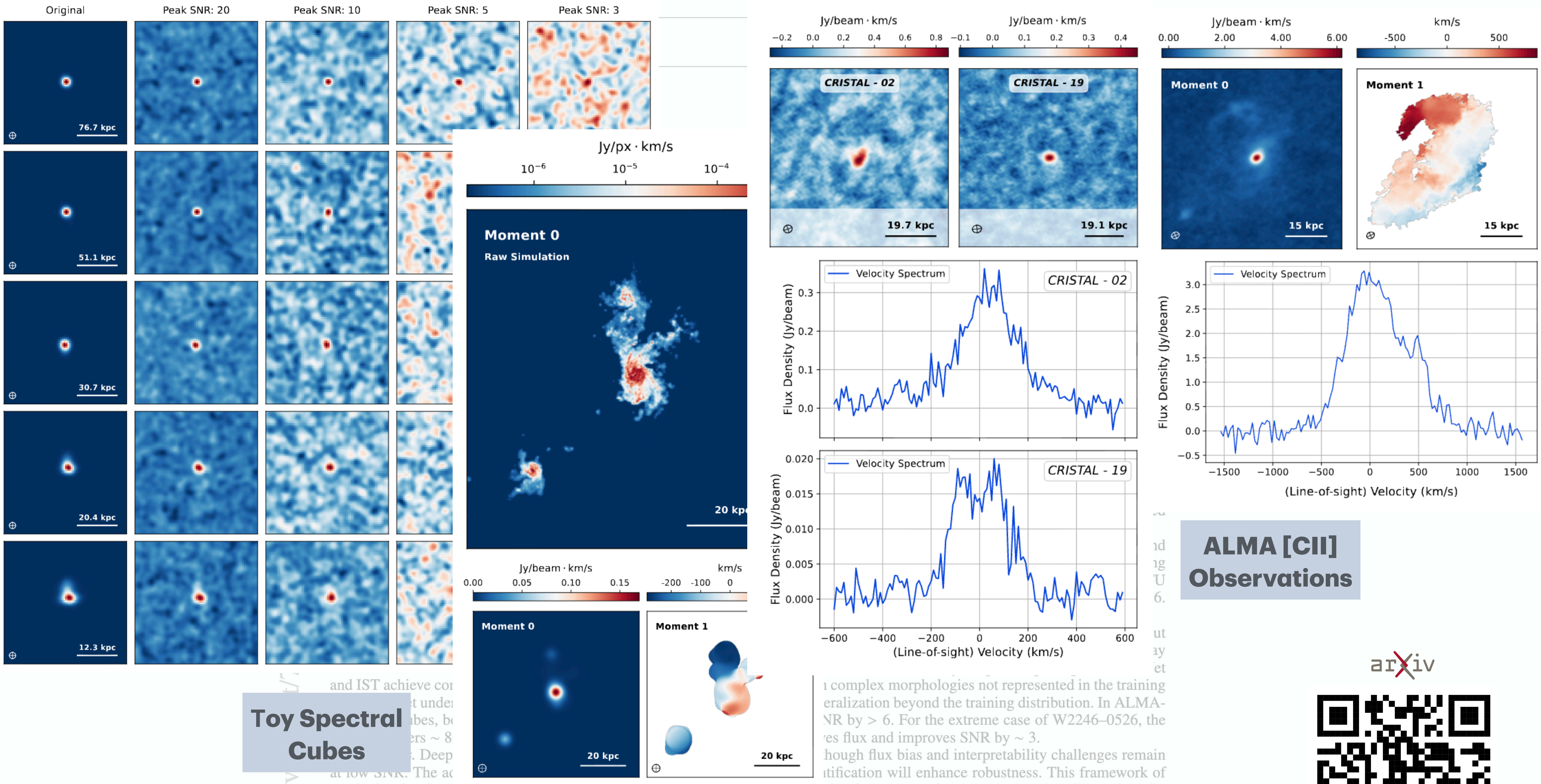
[1] Institutes of Computer Science and Astrophysics, Foundation for Research and Technology Hellas (FORTH), 100 Nikolaou Plastira str., Vassilika Vouton, Heraklion, 70013, Greece
[2] Departments of Physics and Computer Science, University of Crete, Voutes Campus, Vasilika Voutes, Heraklion, 70013, Greece
[3] Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191, Gif-sur-Yvette, France
[4] Department of Physics, University of Connecticut, 196 Auditorium Road, U-3046, Storrs, CT 06269, USA



**ABSTRACT**

*Context.* Beyond cosmic noon, galaxies usually appear as faint whispers amid overwhelming noise, yet this epoch is key to understanding massive galaxy assembly. ALMA's sensitivity to cold dust and [C II] emission allows us to probe their interstellar medium, but faint signals are still challenging, rendering robust denoising essential.
*Aims.* We evaluate denoising strategies, including classical statistical methods, sparse unsupervised representations, and supervised deep learning, to identify techniques that suppress noise while preserving flux and spectral-spatial morphology.
*Methods.* We develop a physically motivated synthetic dataset of spectral cubes simulating rotating disk galaxies for training and evaluation. We benchmark Principal Component Analysis (PCA), Independent Component Analysis (ICA), iterative soft thresholding with 2D-1D wavelets (IST), and a supervised 3D U-Net across peak SNRs of ∼2.5–8, applied to (i) toy cubes, (ii) synthetic [C II] IFU cubes from FIRE simulations, and (iii) ALMA observations of $z \sim 5$ galaxies from the CRISTAL sample and the quasar W2246−0526. Performance is assessed via RMSE, flux conservation, morphology, and SNR improvement.
*Results.* PCA and ICA provide limited noise reduction and struggle with correlated noise. IST reduces noise at moderate SNRs but can suppress emission at low SNRs. The 3D U-Net outperforms IST on synthetic cubes, particularly at low SNR, though it may overestimate flux or hallucinate faint structures in this regime. On high SNR real data with relatively simple morphologies, the U-Net and IST achieve comparable performance. However, on low SNR real data with complex morphologies not represented in the training set, the U-Net underperforms relative to IST, highlighting the challenges of generalization beyond the training distribution. In ALMA-CRISTAL cubes, both IST and U-Net conserve > 91% of flux and increase SNR by > 6. For the extreme case of W2246–0526, the U-Net recovers ∼ 80% of flux at moderate SNR, whereas IST robustly conserves flux and improves SNR by ∼ 3.
*Conclusions.* Deep learning trained on synthetic data generalizes effectively, though flux bias and interpretability challenges remain at low SNR. The addition of physically motivated priors and uncertainty quantification will enhance robustness. This framework of synthetic, simulated, and real datasets offers a pathway for transferable denoising in surveys with ALMA, VLT/MUSE, and JWST IFUs.

**Key words.** infrared: galaxies– galaxies: high-redshift – techniques: image spectroscopy – methods: data analysis – methods: statistical

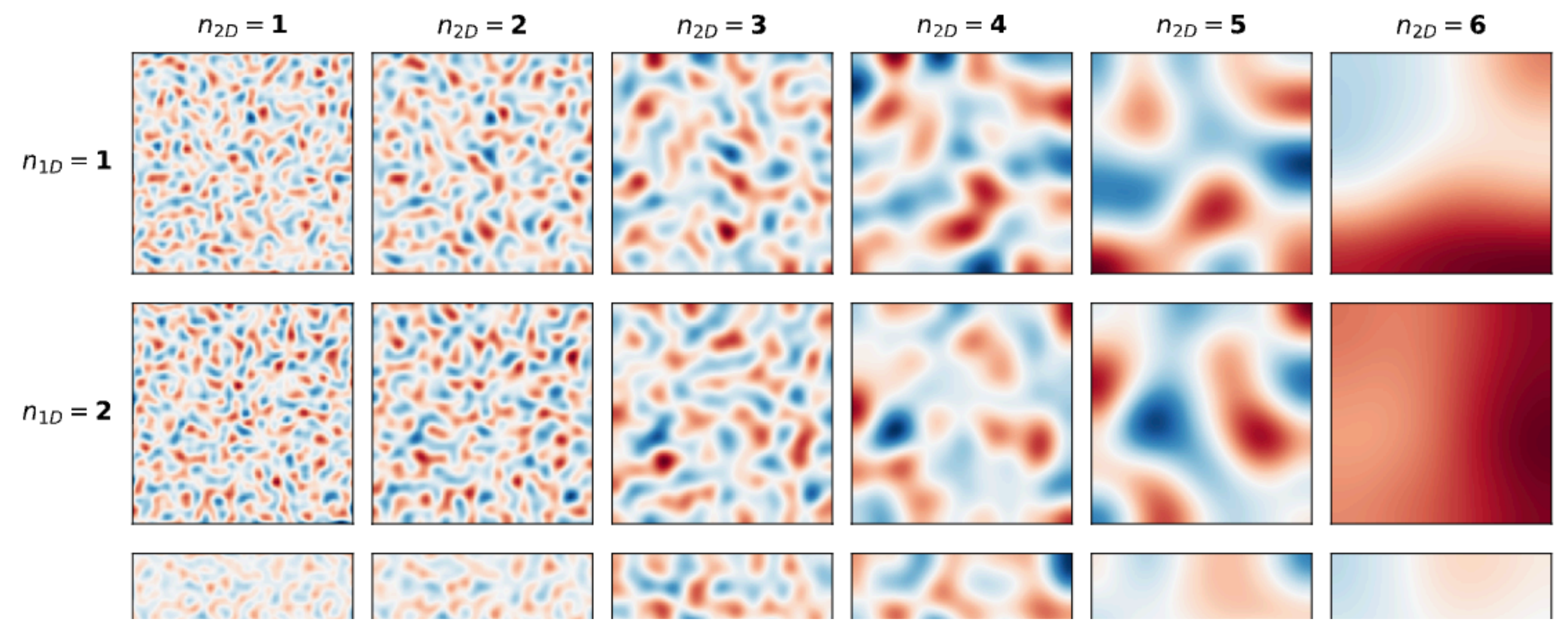


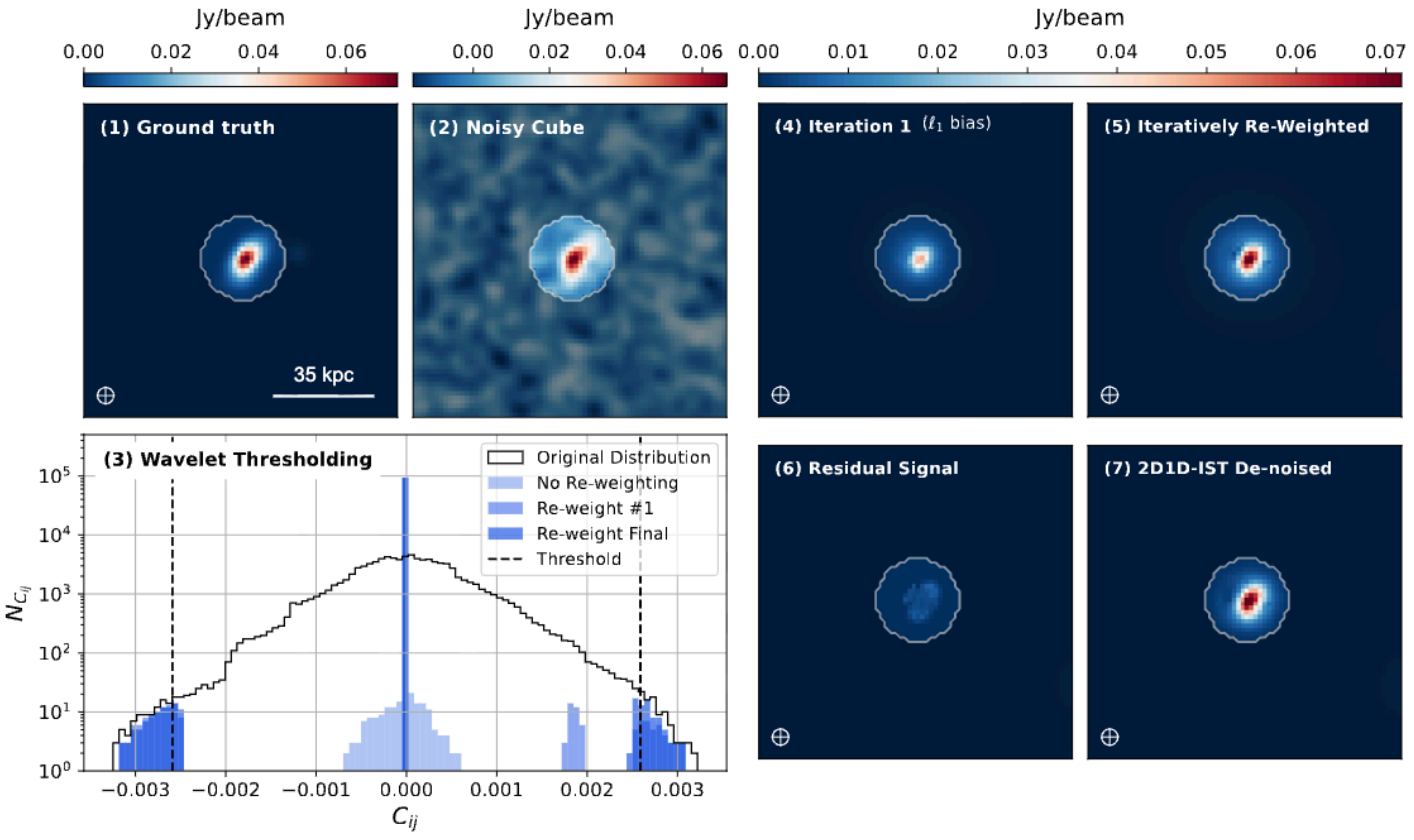

**Toy Spectral Cubes**

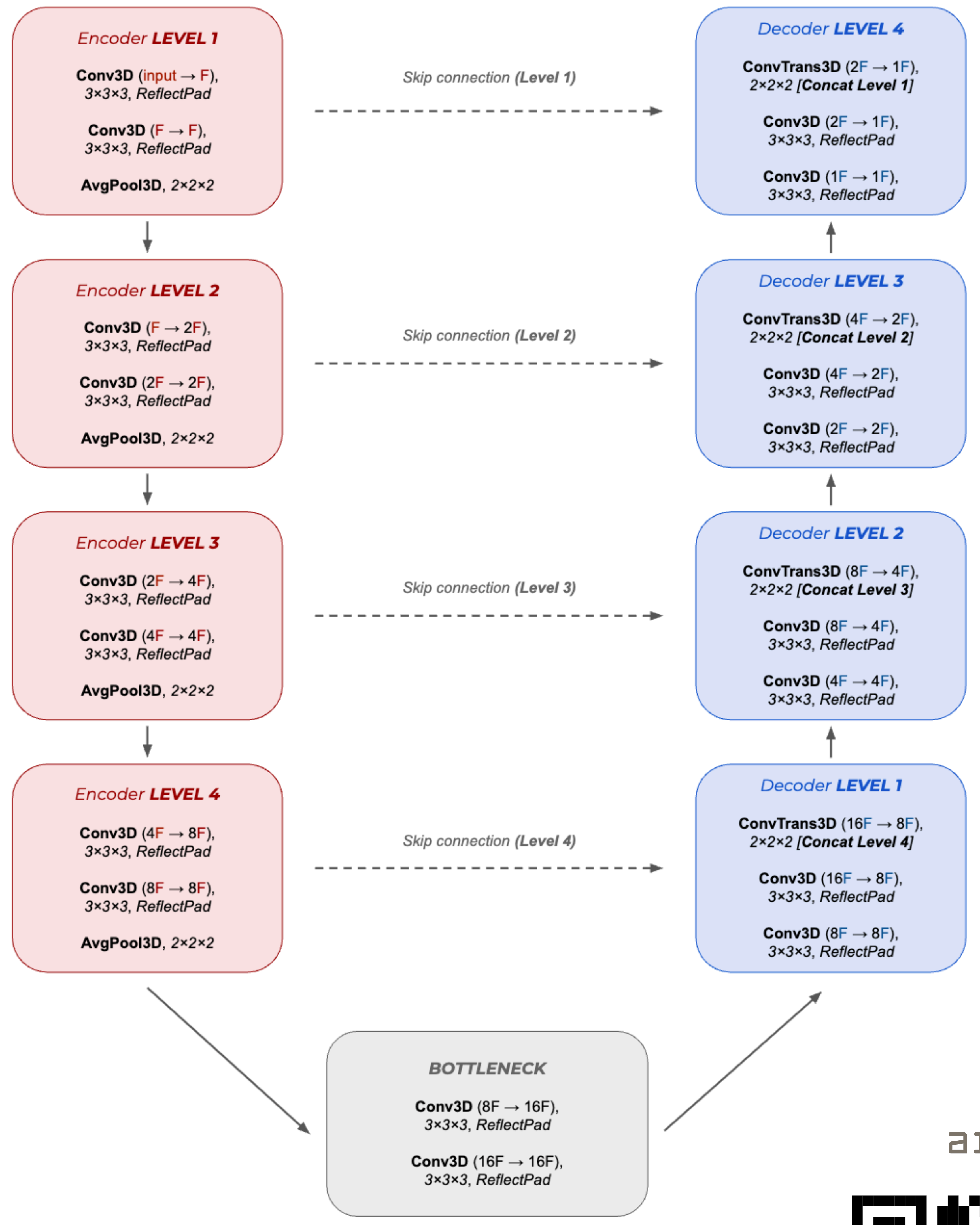
**Synthetic FIRE [CII] IFU**

**ALMA [CII] Observations**

arXiv



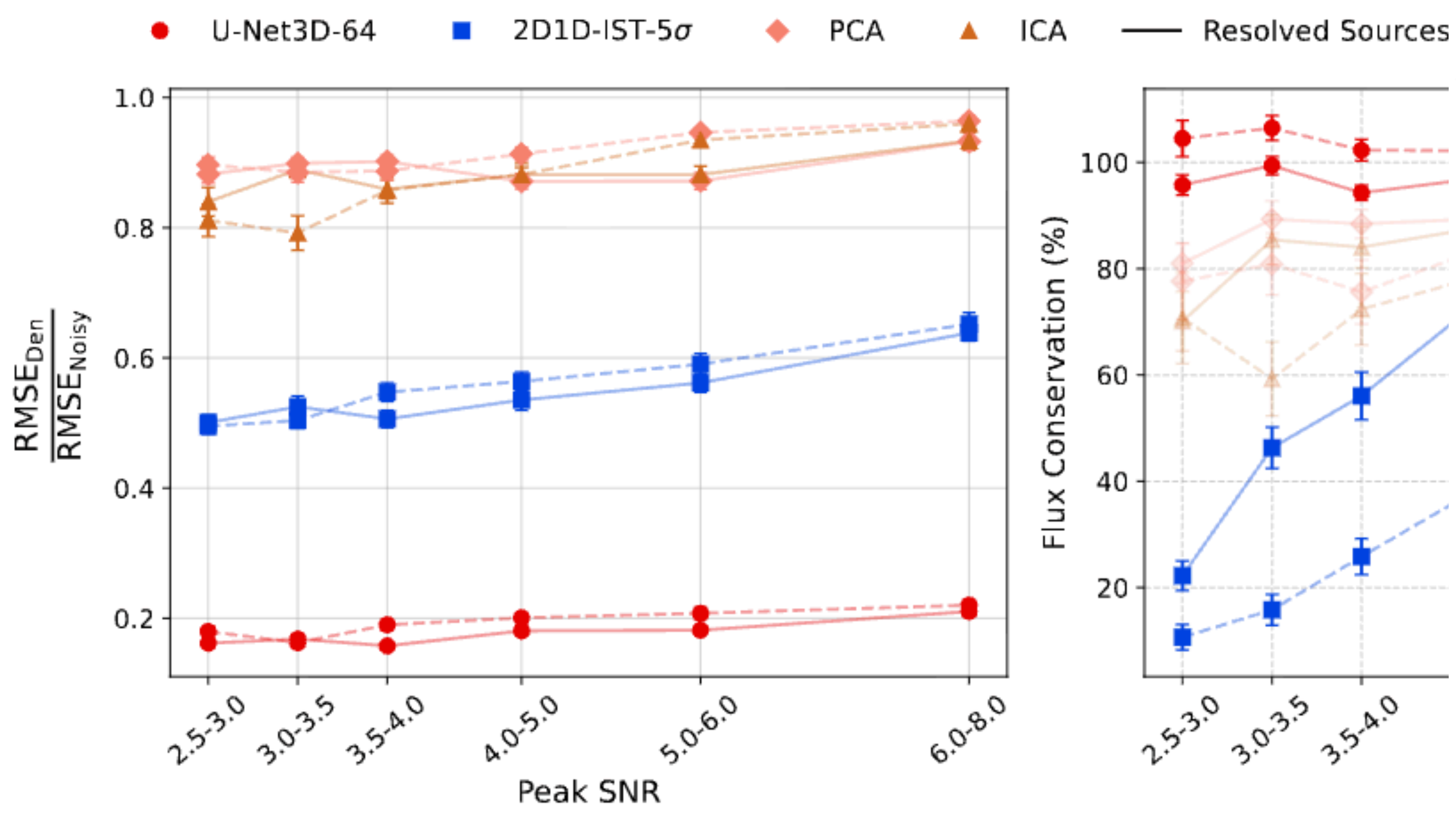

by > 6. For the extreme case of W2246–0526, the
flux and improves SNR by ∼ 3.
ugh flux bias and interpretability
ication will enhance robustness.
ig in surveys with ALMA, VLT/MUSE, and JWST

troscopy – methods: data analysis – methods: statis-

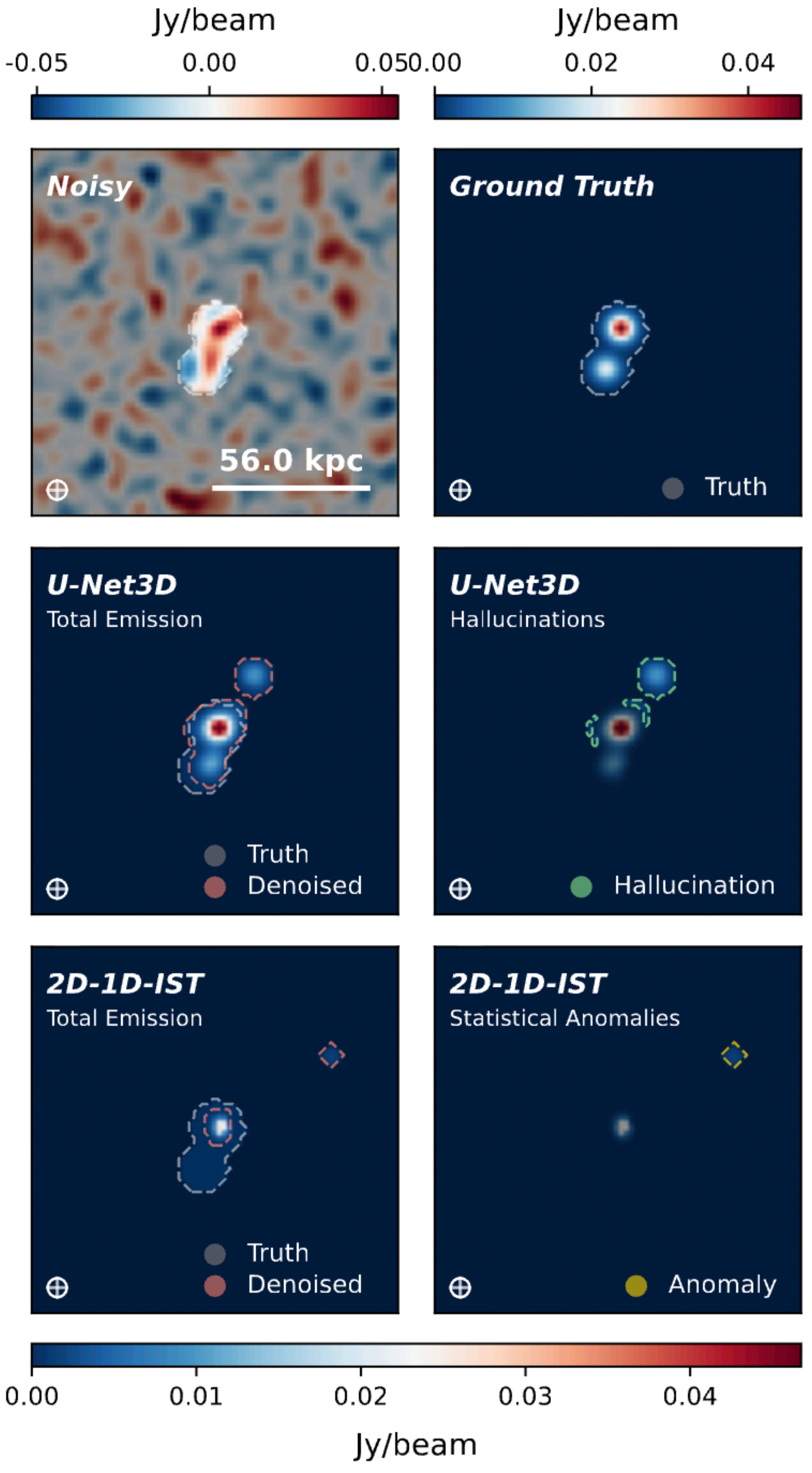
**U-Net3D**

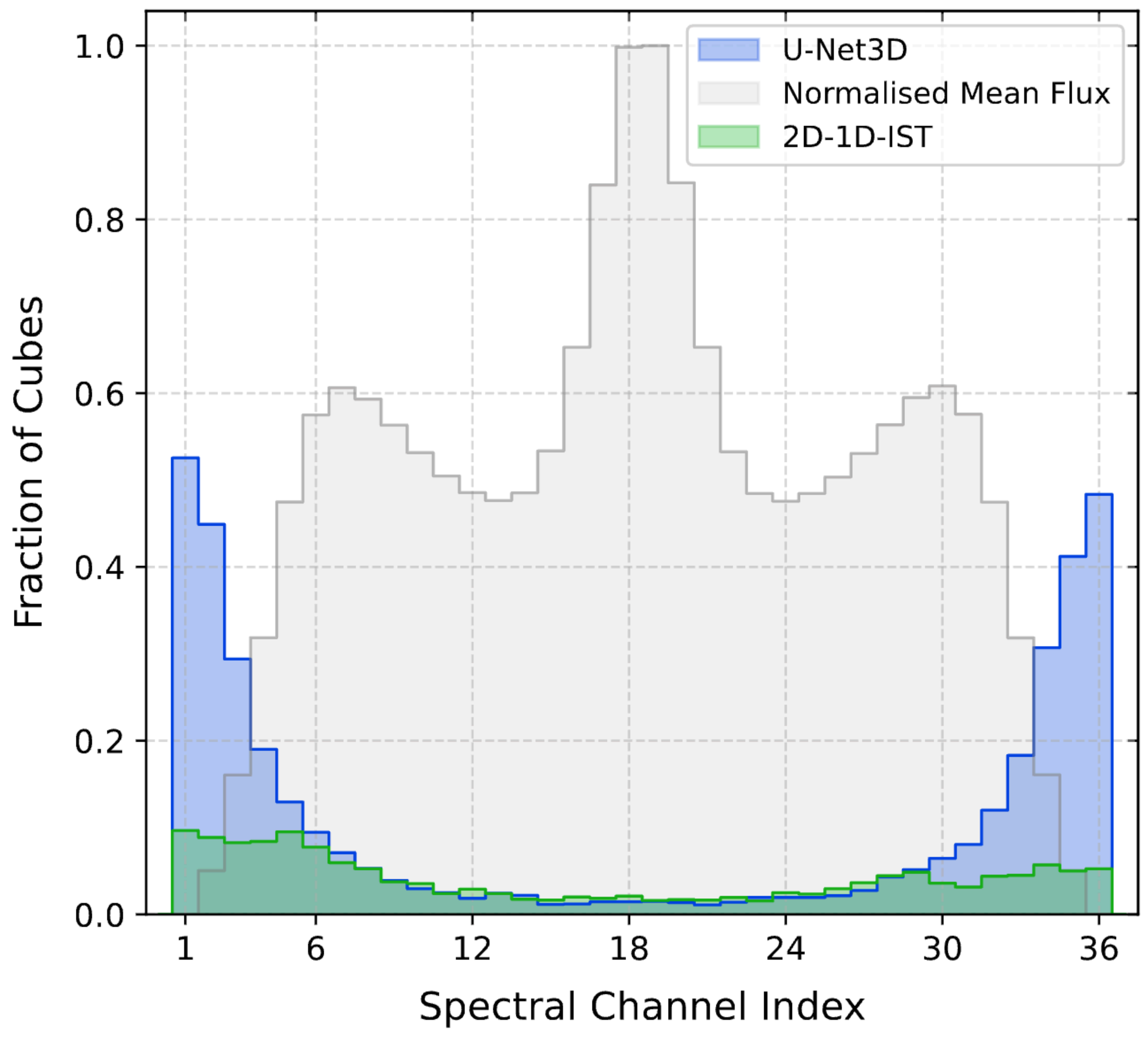
**Reweighted Iterative Soft Thresholding (IST) Denoising**

arXiv

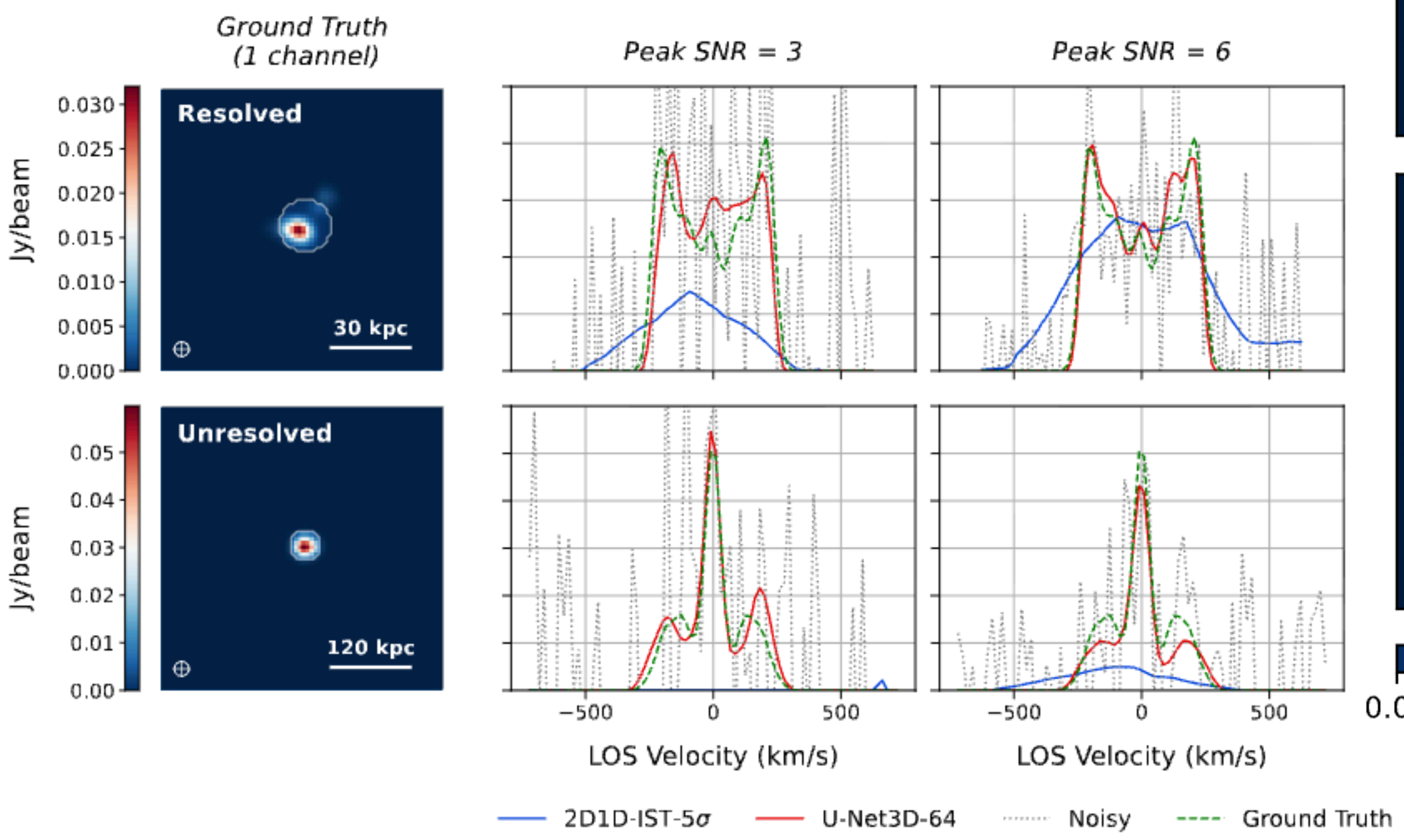



## Benchmarking toy cubes

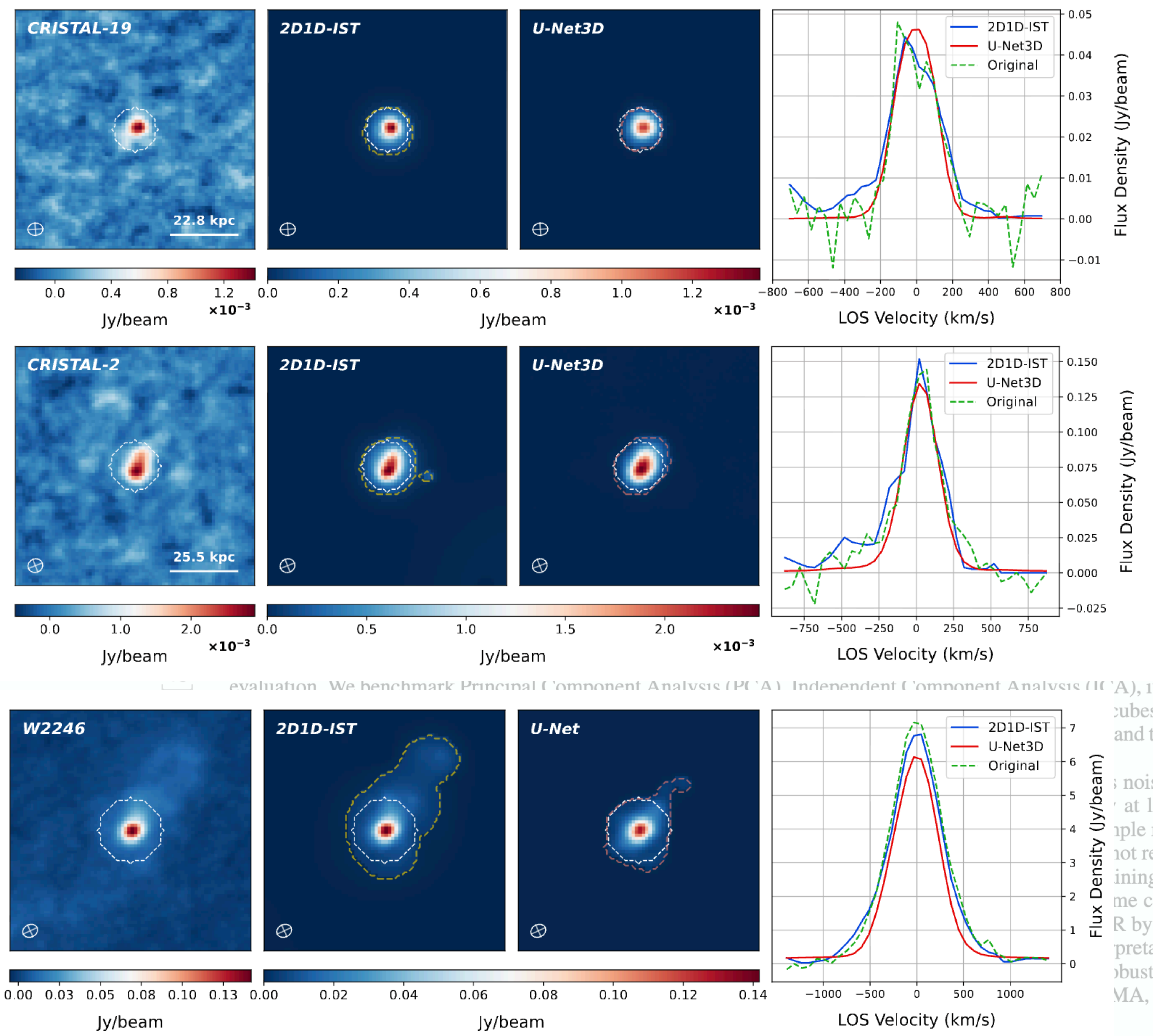



(Aperture-based) **Hallucination Quantification**




## Spectral shape recovery




l Data Cubes
serv
iel An
I), 100 N
aklion, 7

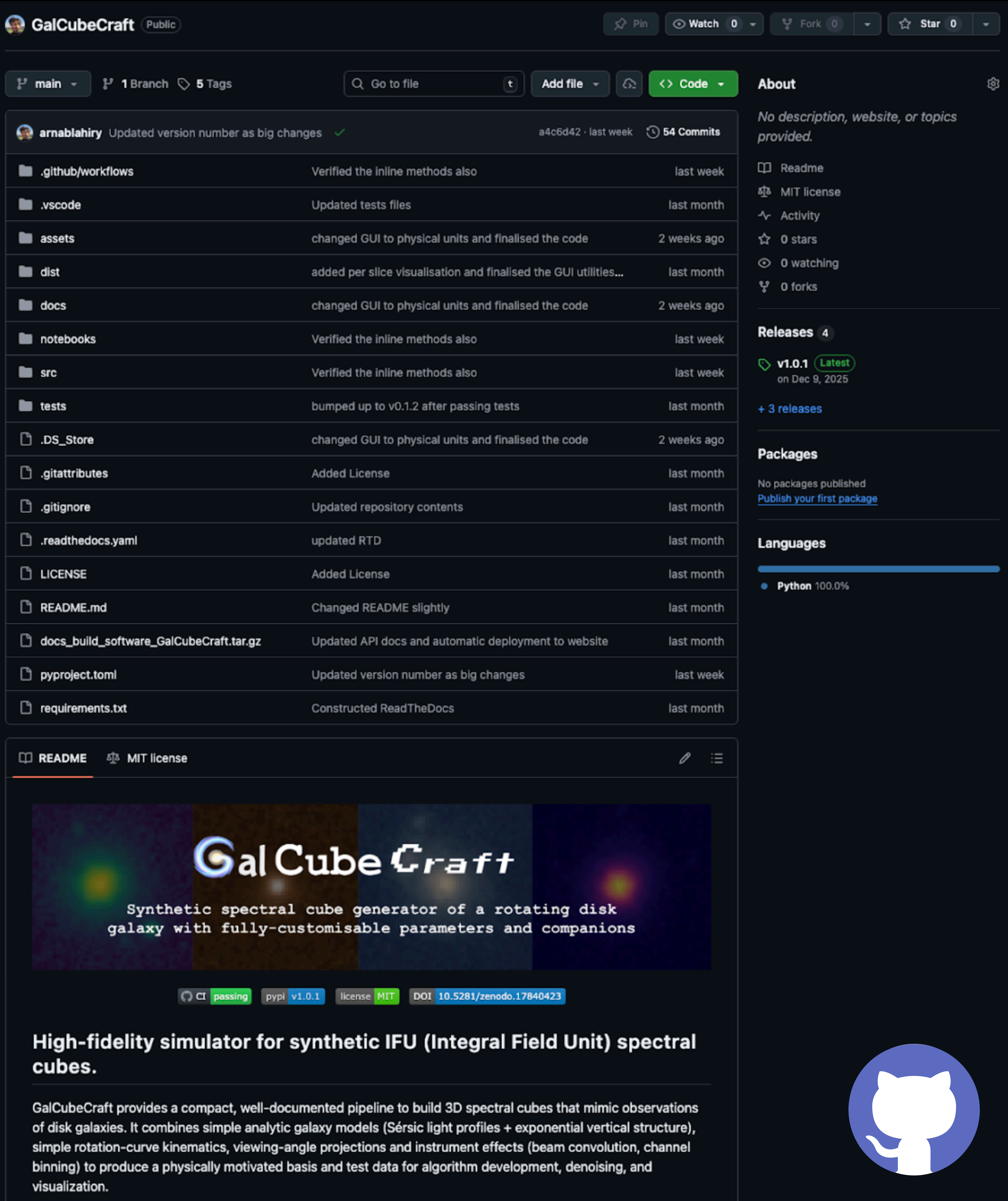
## Application to [CII] 158 µm observations from ALMA:

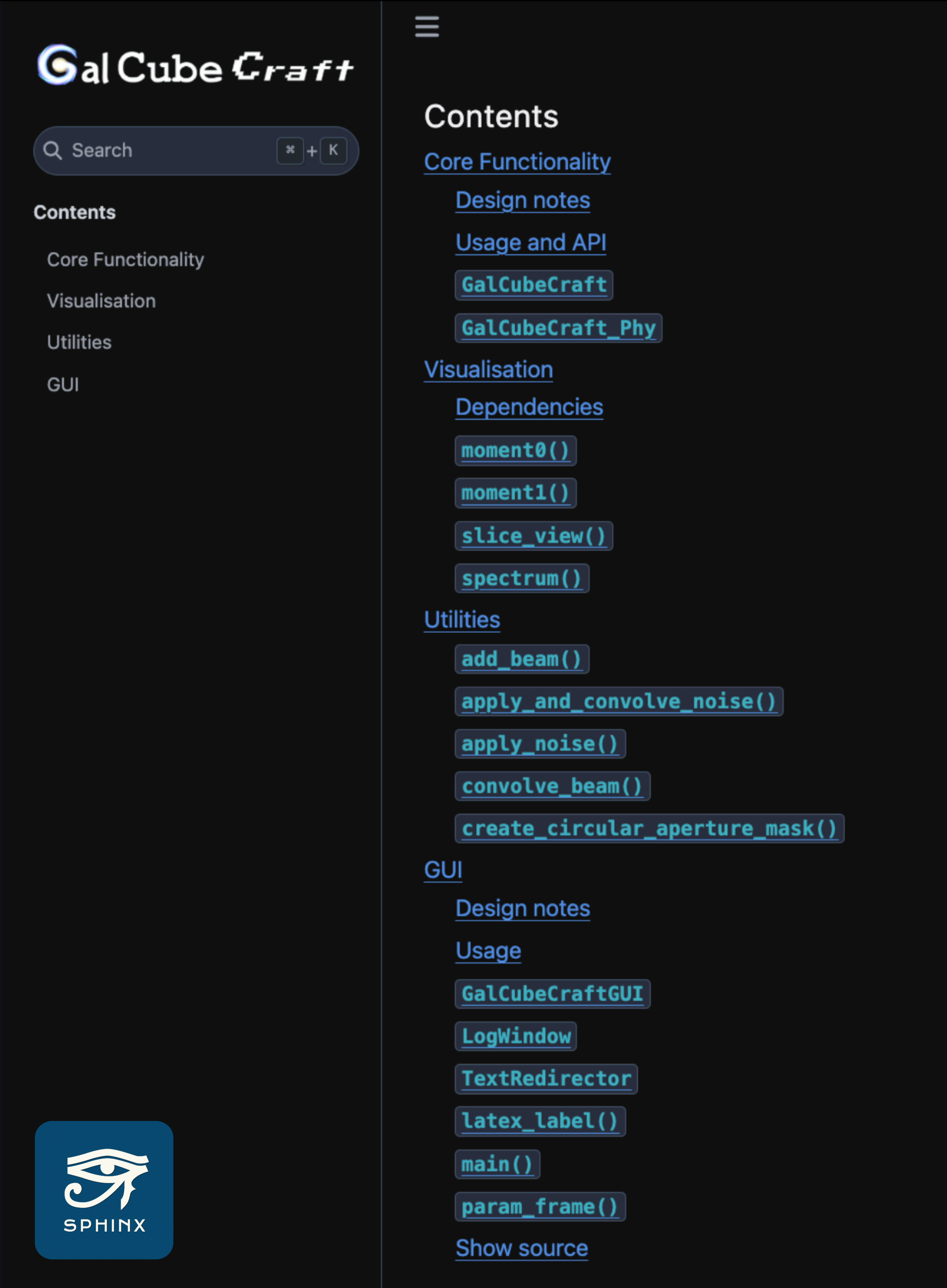
1. Resolved & Unresolved CRISTAL samples
2. Quasar: W2246−0526

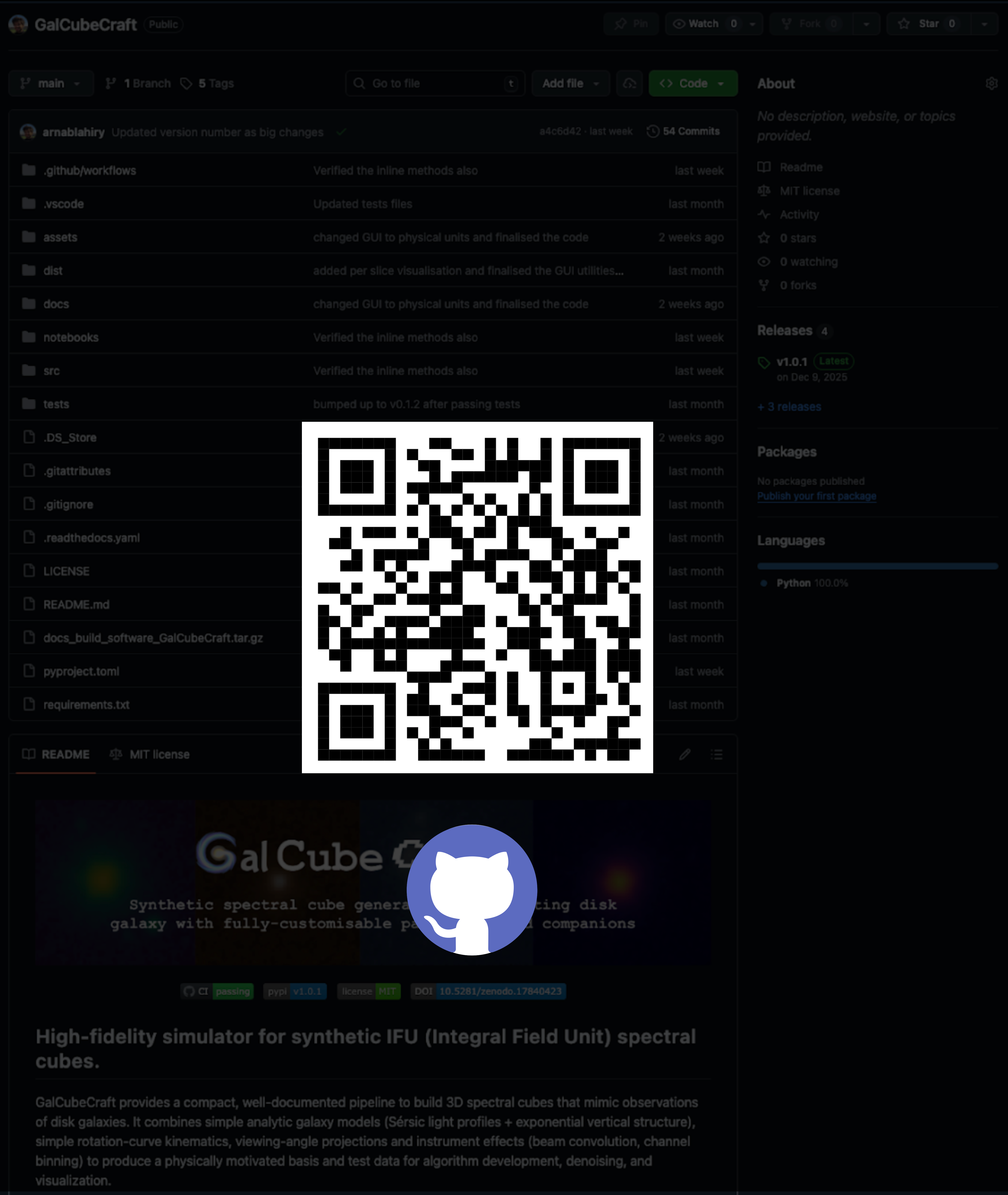
s epoch is key to under-
heir interstellar medium,
ntations, and supervised
gy.
galaxies for training and
evaluation. We benchmark Principal Component Analysis (PCA), Independent Component Analysis (ICA), iterative soft thresholding
cubes, (ii) synthetic [C II] IFU
and the quasar W2246−0526.
noise at moderate SNRs but
at low SNR, though it may
iple morphologies, the U-Net
not represented in the training
ining distribution. In ALMA-
me case of W2246–0526, the
R by ∼ 3.
pretability challenges remain
bustness. This framework of
MA, VLT/MUSE, and JWST

**Key words.** infrared: galaxies– galaxies: high-redshift – techniques: image spectroscopy – methods: data analysis – methods: statistical


GalCubeCraft GUI
GalCubeCraft
Synthetic spectral cube generator of a rotat
galaxy with fully-customisable parameters and
Number of galaxies
1 2 3 4 5 6
Spatial Resolution (
Field of View [kpc]
FOVX ; FOVY 275
Beam Information [k
Bmin 11.0 Bmaj 13.0
Sérsic index (n) [−]
1.000
Scale height (hz) [kp
Central effective flux density (Se) [Jy]
0.1
Satellite offset from
Spectral Resolution (Δvz) [km s−1]
20
Velocity dispersion (
Inclination angle (θX) [deg]
45
Azimuthal angle (φY)
Generate
Slice
Moments
Spectrum
Save
Reset
arnablahiry — python -m GalCubeCraft.gui — 80×
IFU viewer
Channel 26 : v = −90.0 km s−1
94.
0.00
0.01
0.02
0.03
0.04
Jy beam−1
Channel: 26 : v = -90.0 km/s
Moment 1
km s−1
-200
-100
0
100
200
Moment 1
94.8 kpc
Logs
Creating disk #1...
Calculating the flux density values at each spatial location
Calculating and assigning velocity vectors...
Rotating 45.00 degrees about X axis and 30.00 degrees about Y axis:
1. Rotating/transforming the whole system...
2: Rotating the individual velocity vectors...
Disk #1 generated!
Creating disk #2...
Calculating the flux density values at each spatial location
Calculating and assigning velocity vectors...
Rotating 67.55 degrees about X axis and 107.73 degrees about Y axis:
1. Rotating/transforming the whole system...
2: Rotating the individual velocity vectors...
Disk #2 generated!
Creating disk #3...
Calculating the flux density values at each spatial location
Calculating and assigning velocity vectors...
Rotating 146.75 degrees about X axis and -53.31 degrees about Y axis:
1. Rotating/transforming the whole system...
2: Rotating the individual velocity vectors...


GalCubeCraft Public
Pin
Watch 0
Fork 0
Star 0
main
1 Branch
5 Tags
Go to file
Add file
Code
arnablahiry Updated version number as big changes
a4c6d42 · last week
54 Commits
.github/workflows Verified the inline methods also last week
.vscode Updated tests files last month
assets changed GUI to physical units and finalised the code 2 weeks ago
dist added per slice visualisation and finalised the GUI utilities... last month
docs changed GUI to physical units and finalised the code 2 weeks ago
notebooks Verified the inline methods also last week
src Verified the inline methods also last week
tests bumped up to v0.1.2 after passing tests last month
.DS_Store changed GUI to physical units and finalised the code 2 weeks ago
.gitattributes Added License last month
.gitignore Updated repository contents last month
.readthedocs.yaml updated RTD last month
LICENSE Added License last month
README.md Changed README slightly last month
docs_build_software_GalCubeCraft.tar.gz Updated API docs and automatic deployment to website last month
pyproject.toml Updated version number as big changes last week
requirements.txt Constructed ReadTheDocs last month
About
No description, website, or topics provided.
Readme
MIT license
Activity
0 stars
0 watching
0 forks
Releases 4
v1.0.1 Latest
on Dec 9, 2025
+ 3 releases
Packages
No packages published
Publish your first package
Languages
Python 100.0%
README
MIT license
GalCubeCraft
Synthetic spectral cube generator of a rotating disk galaxy with fully-customisable parameters and companions
CI passing
pypi v1.0.1
license MIT
DOI 10.5281/zenodo.17840423
High-fidelity simulator for synthetic IFU (Integral Field Unit) spectral cubes.
GalCubeCraft provides a compact, well-documented pipeline to build 3D spectral cubes that mimic observations of disk galaxies. It combines simple analytic galaxy models (Sérsic light profiles + exponential vertical structure), simple rotation-curve kinematics, viewing-angle projections and instrument effects (beam convolution, channel binning) to produce a physically motivated basis and test data for algorithm development, denoising, and visualization.


GalCubeCraft
Search
Contents
Core Functionality
Visualisation
Utilities
GUI
Contents
Core Functionality
Design notes
Usage and API
GalCubeCraft
GalCubeCraft_Phy
Visualisation
Dependencies
moment0()
moment1()
slice_view()
spectrum()
Utilities
add_beam()
apply_and_convolve_noise()
apply_noise()
convolve_beam()
create_circular_aperture_mask()
GUI
Design notes
Usage
GalCubeCraftGUI
LogWindow
TextRedirector
latex_label()
main()
param_frame()
Show source


GalCubeCraft Public
Pin
Watch 0
Fork 0
Star 0
main
1 Branch
5 Tags
Go to file
Add file
Code
arnablahiry Updated version number as big changes
a4c6d42 · last week
54 Commits
.github/workflows Verified the inline methods also last week
.vscode Updated tests files last month
assets changed GUI to physical units and finalised the code 2 weeks ago
dist added per slice visualisation and finalised the GUI utilities... last month
docs changed GUI to physical units and finalised the code 2 weeks ago
notebooks Verified the inline methods also last week
src Verified the inline methods also last week
tests bumped up to v0.1.2 after passing tests last month
.DS_Store 2 weeks ago
.gitattributes last month
.gitignore last month
.readthedocs.yaml last month
LICENSE last month
README.md last month
docs_build_software_GalCubeCraft.tar.gz last month
pyproject.toml last week
requirements.txt last month
About
No description, website, or topics provided.
Readme
MIT license
Activity
0 stars
0 watching
0 forks
Releases 4
v1.0.1 Latest
on Dec 9, 2025
+ 3 releases
Packages
No packages published
Publish your first package
Languages
Python 100.0%
README
MIT license
Synthetic spectral cube genera... ...ting disk
galaxy with fully-customisable pa... ...d companions
CI passing
pypi v1.0.1
license MIT
DOI 10.5281/zenodo.17840423
High-fidelity simulator for synthetic IFU (Integral Field Unit) spectral cubes.
GalCubeCraft provides a compact, well-documented pipeline to build 3D spectral cubes that mimic observations of disk galaxies. It combines simple analytic galaxy models (Sérsic light profiles + exponential vertical structure), simple rotation-curve kinematics, viewing-angle projections and instrument effects (beam convolution, channel binning) to produce a physically motivated basis and test data for algorithm development, denoising, and visualization.


Gal Cube Craft
Search
Contents
Core Functionality
Visualisation
Utilities
GUI
Contents
Core Functionality
Design notes
Usage and API
GalCubeCraft
GalCubeCraft_Phy
Visualisation
Dependencies
GUI
Design
Usage
LogWindow
TextRedirector
latex_label()
main()
param_frame()
Show source

# *Beyond Moment-0: Galaxy property inference from High-z Spectral Data Cubes*

**Arnab Lahiry**, Tanio Díaz-Santos, Jean-Luc Starck
Niranjan Chandra Roy, Daniel Anglés-Alcázar,

# M O T I V A T I O N

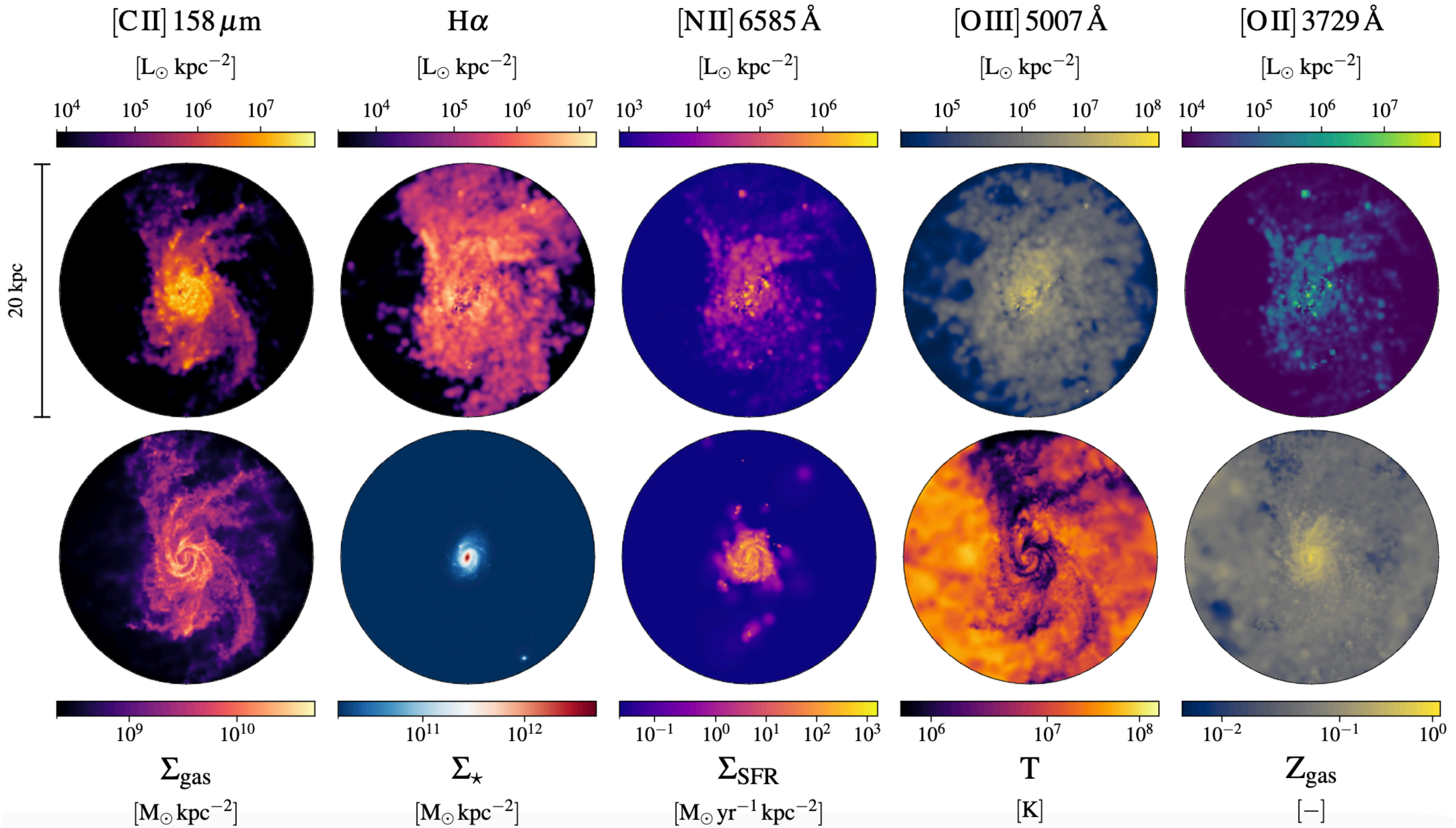
Current norms for resolved sources involve computing empirical scaling relations between the moment 0 map/luminosity (surface density) and the star formation rate (surface density).

- Use synthetic simulated IFUs and use machine learning to go 'beyond moment-0' and constrain a physically motivated and realistic relationship between emission lines and physical properties.

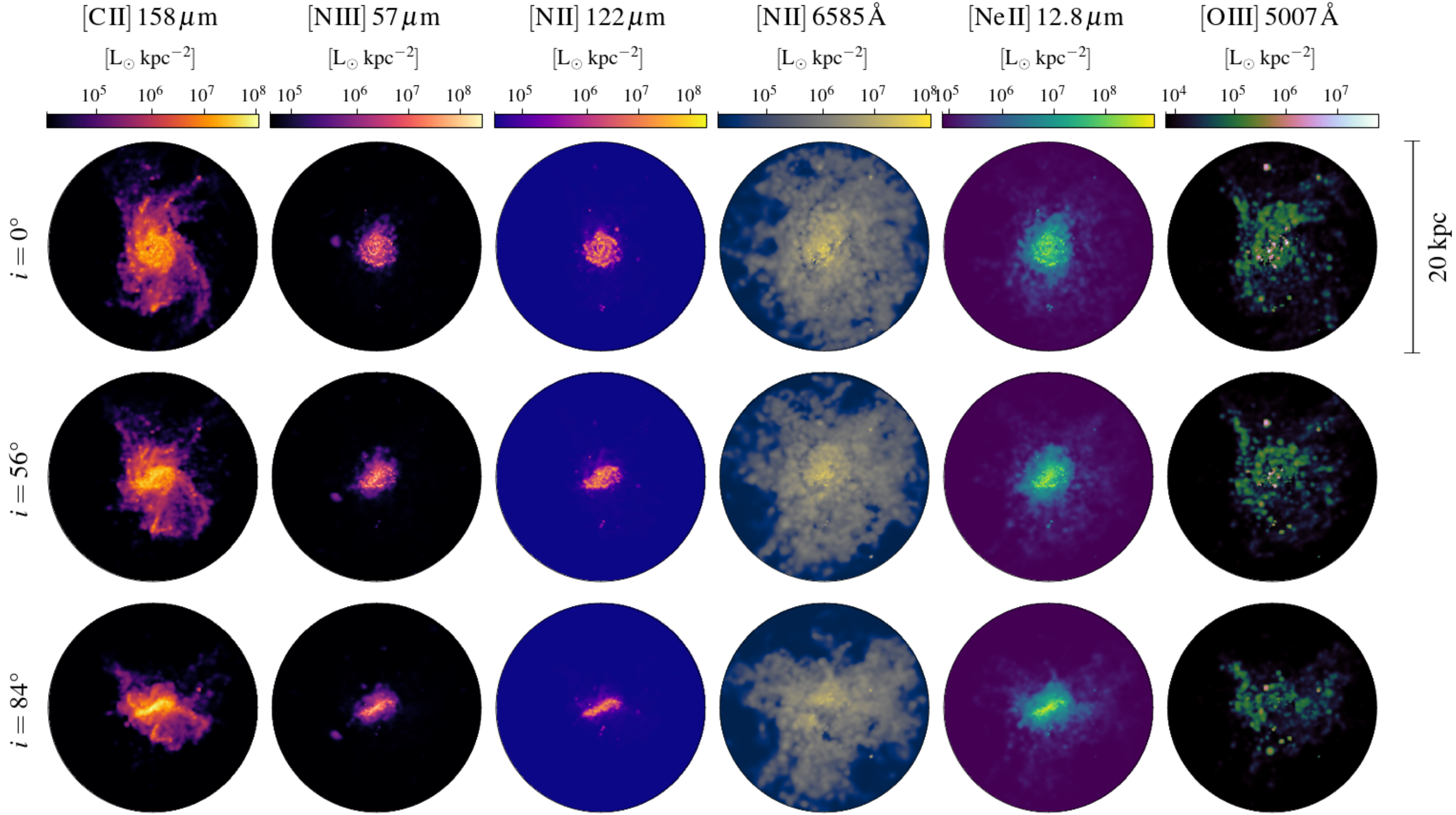
  - Analyse whether the information from the spectral axis contributes significantly to physical property inference, and compute significance scores

- Aim to improve SFR constraints for observational sources

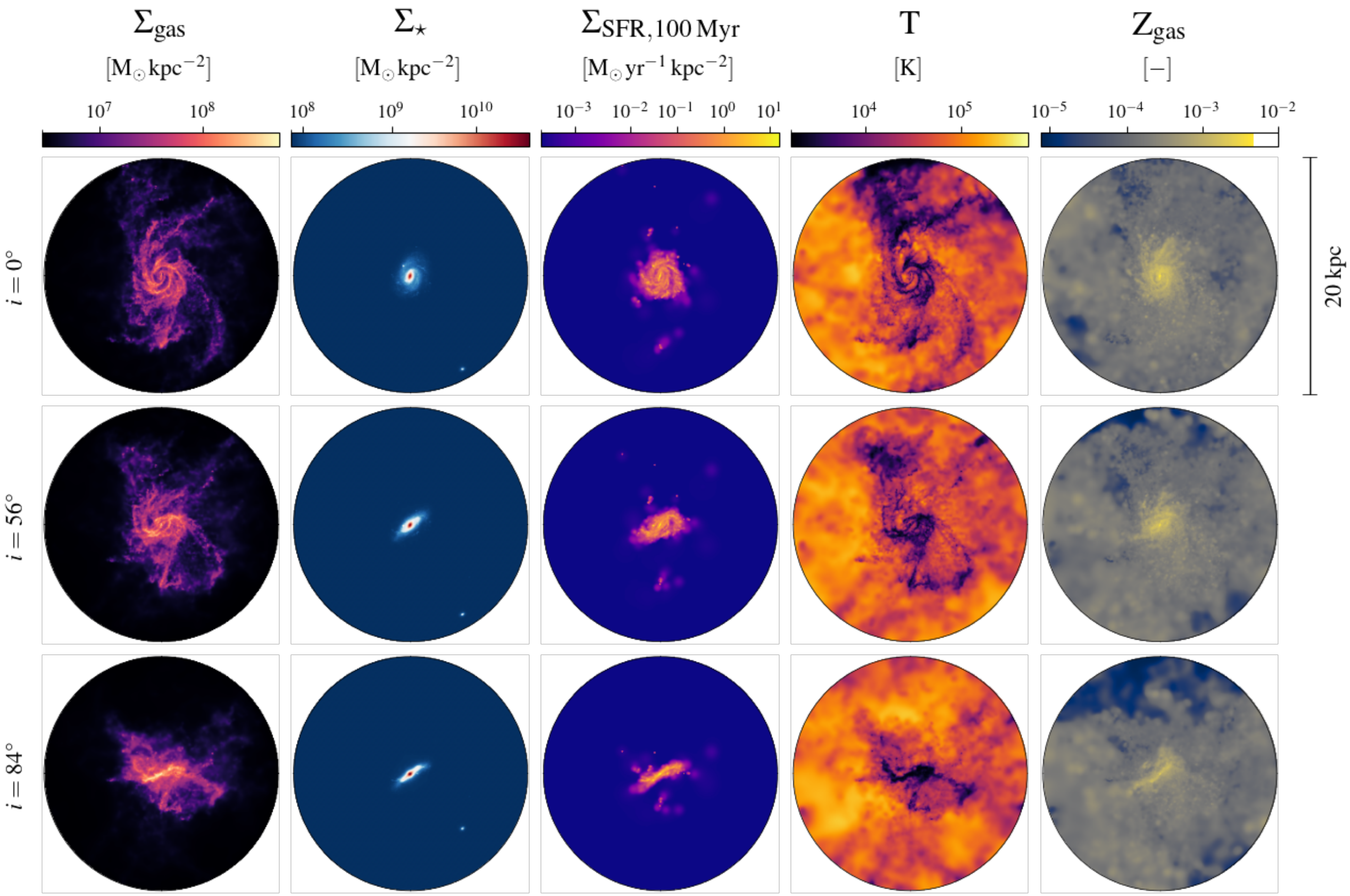
# **Synthetic IFU Spectral Cubes & Physical Properties** - FIRE Simulations + CHIMES + RADMC-3D



# **Multiple Inclinations** (Individual Runs)



# Corresponding Physical Properties



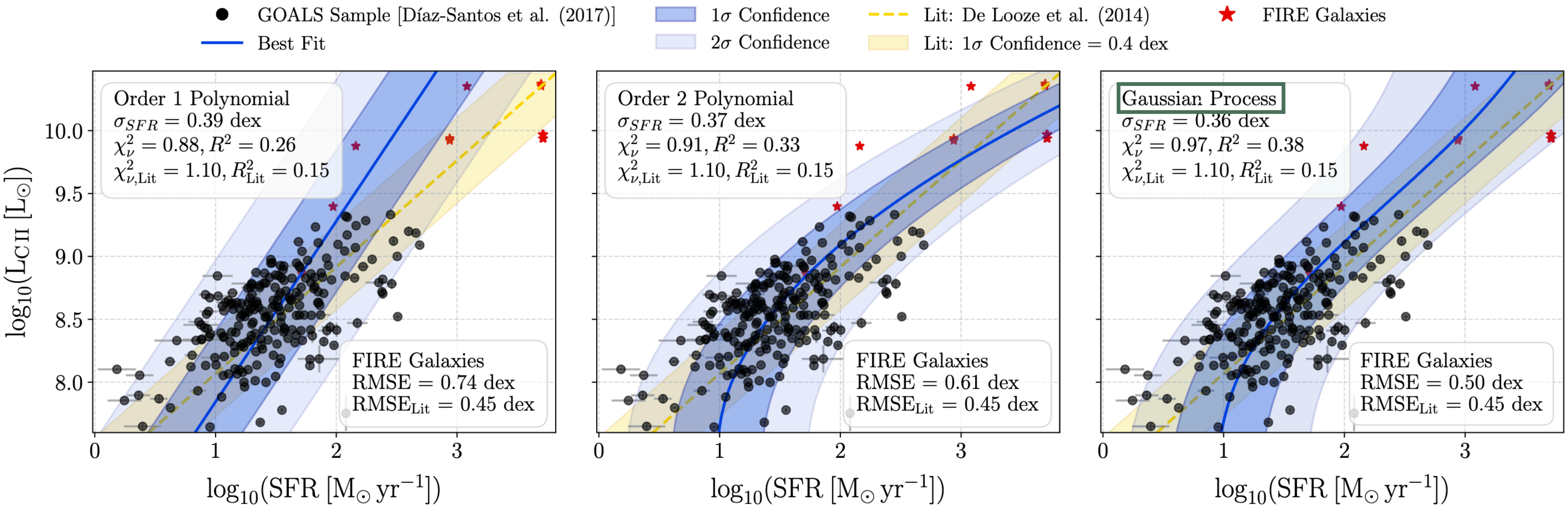
# Global scaling relation between [C II] luminosities and SFR

THE ASTROPHYSICAL JOURNAL, 846:32 (22pp), 2017 September 1 https://doi.org/10.3847/1538-4357/aa81d7


**GOALS**
sample

## A *Herschel*/PACS Far-infrared Line Emission Survey of Local Luminous Infrared Galaxies

T. Díaz-Santos[1], L. Armus[2], V. Charmandaris[3,4], N. Lu[5,6], S. Stierwalt[7,8], G. Stacey[9], S. Malhotra[10], P. P. van der Werf[11], J. H. Howell[2], G. C. Privon[12,13], J. M. Mazzarella[14], P. F. Goldsmith[15], E. J. Murphy[8], L. Barcos-Muñoz[8,16], S. T. Linden[7,8], H. Inami[17], K. L. Larson[2], A. S. Evans[7,8], P. Appleton[14,18], K. Iwasawa[19,20], S. Lord[21], D. B. Sanders[22], and J. A. Surace[2]



**Data** (FIRE)

**Input**: $N_{samples}$ X $N_{lines}$ X $N_{features}$

- **[C II]** 158 μm
- **[N II]** 122 μm
- **[N II]** 6585 Å
- **[O III]** 5007 Å
- ...

- **Integrated Luminosity** ($L_\odot$ kpc$^{-2}$)
- Mom1 - **Velocity Centroid** (km s$^{-1}$)
- Mom2 - **Velocity Dispersion** (km s$^{-1}$)
- Mom3 - **Skewness** (—)
- Mom4 - **Kurtosis** (—)

**Target**: *$N_{samples}$ X Property$_i$*

- Stellar Mass - $\Sigma_\star$ ($M_\odot$ kpc$^{-2}$)
- Gas Mass - $\Sigma_{gas}$ ($M_\odot$ kpc$^{-2}$)
- **SFR$_{100Myr}$ ($M_\odot$ yr$^{-1}$ kpc$^{-2}$)**
- Temperature - T (K)
- Metallicity - Z (—)

SFR **Thresholding** + **Log$_{10}$(..)** scaling **+** Z-Score **Normalisation**

→ Selection of **M** << $N_{samples}$ virtual **Inducing Points** (Dimensionality & Complexity reduction)

**Model**

## Variational Gaussian Process

***Deterministic Mean***

`1. LinearMean`

[Captures baseline linear scaling trend]

***Stochastic Covariance***

`2. RBF + Matern + RQ`

[Additive kernels capture multi-scale deviations from trend]

***Variational* Inference**

`3. VariationalStrategy`

[Interpolating data, inducing points & target]

`4. CholeskyVariationalDistribution`

[Gaussian distribution to represent initial uncertainty at inducing points]

**Application**

## Model Training

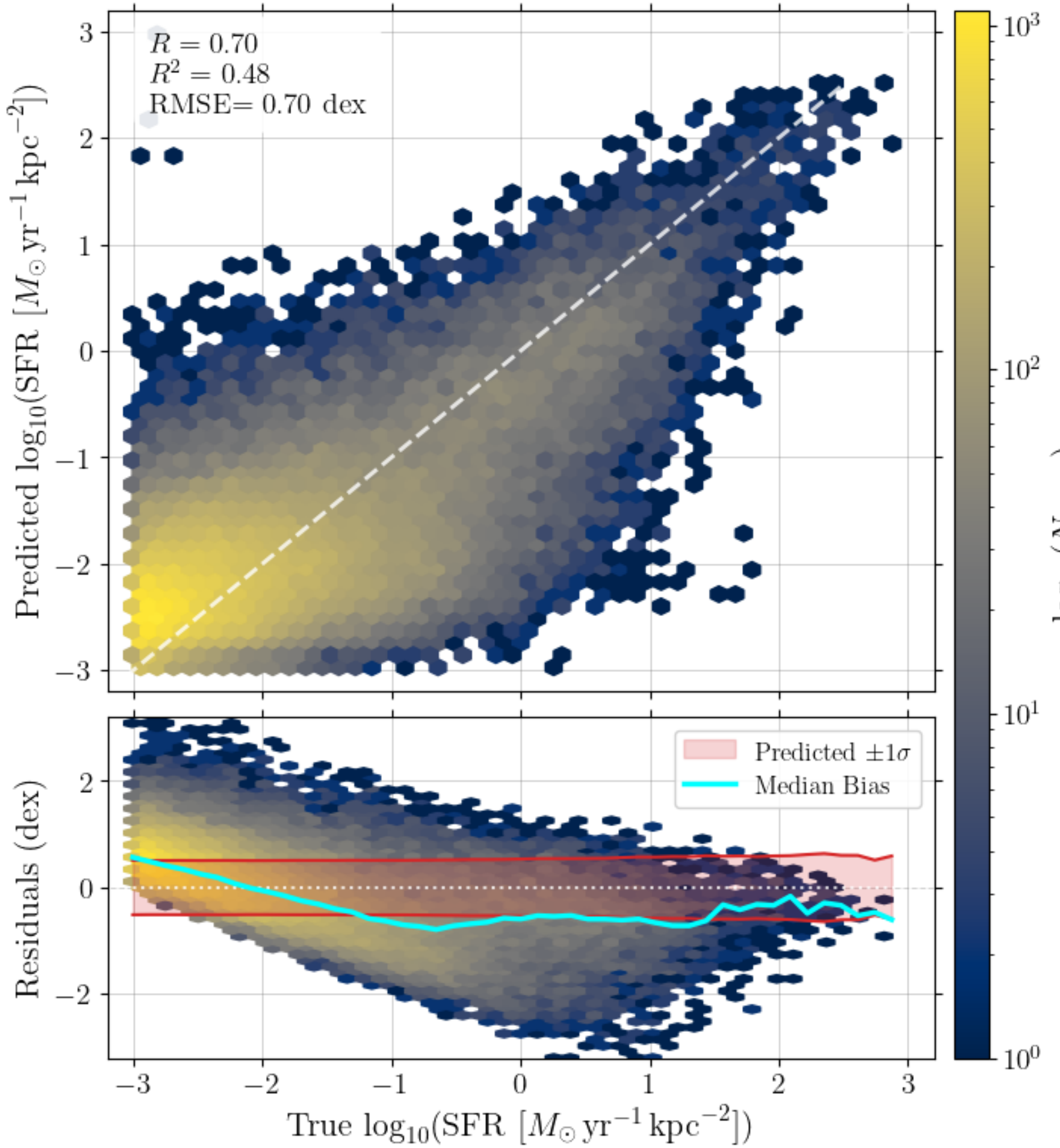
1. **ELBO** Objective
2. **Adam** Optimiser
3. Leave-One-Galaxy-Out **Cross-Validation**

[Rotating folds: Trained on N-i galaxies, tested on i$^{th}$ galaxy]

→ **Results**

## Probabilistic Output

**`MultivariateNormal`** distribution, providing both a mean prediction and a standard deviation for every spatial pixel.



**Alternate approach:**
Can we directly leverage spatial correlations to constrain SFR?

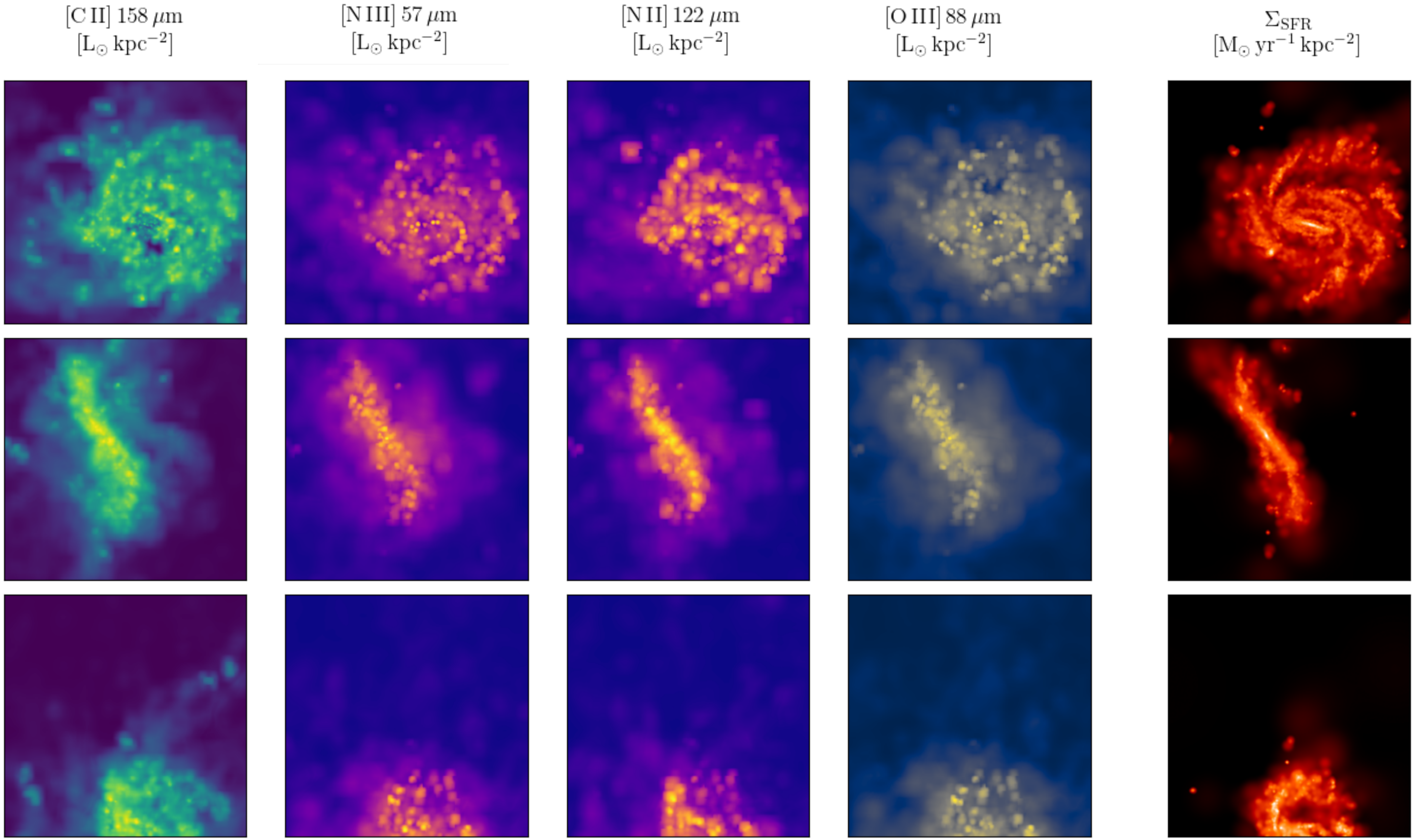
Trying to mitigate the lack of a large dataset

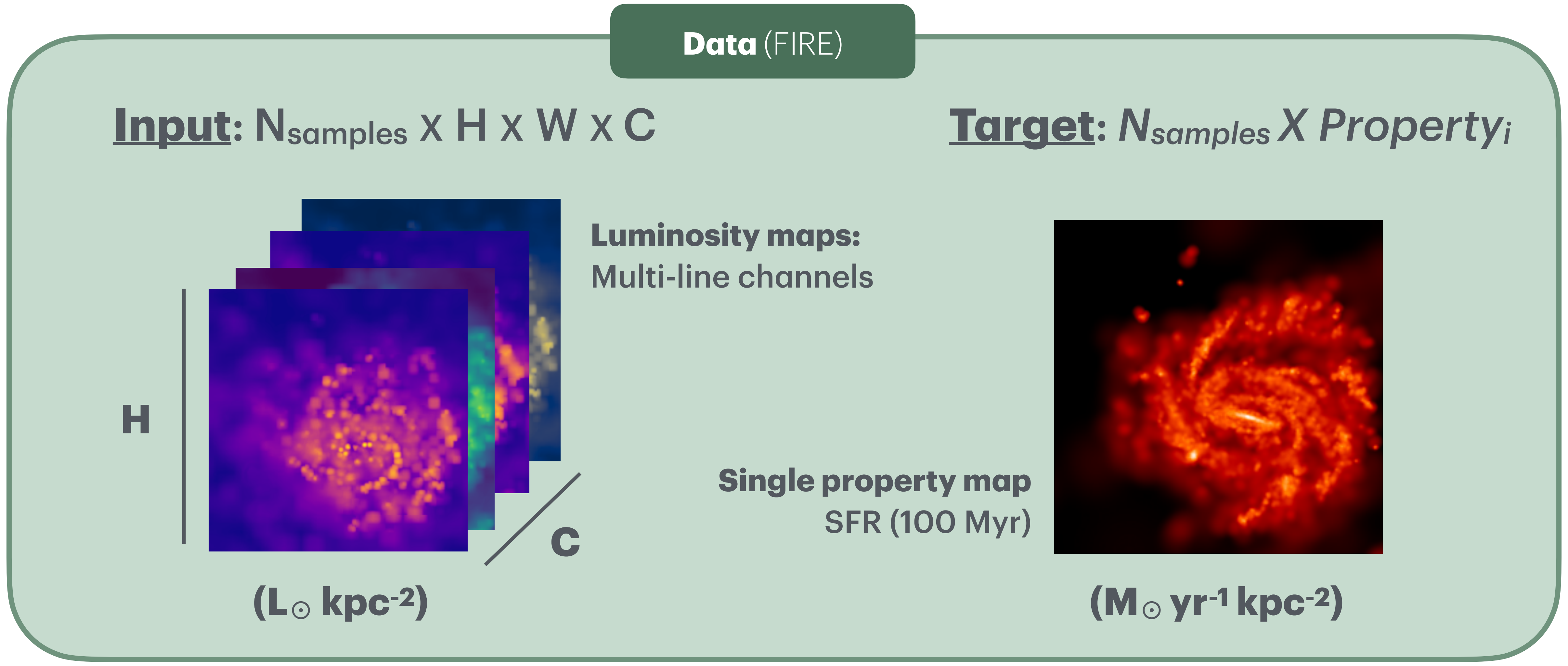
**Patching spatial area**

(1024 X 1024 px$^2$) ——> 16 X (256 X 256 px$^2$)

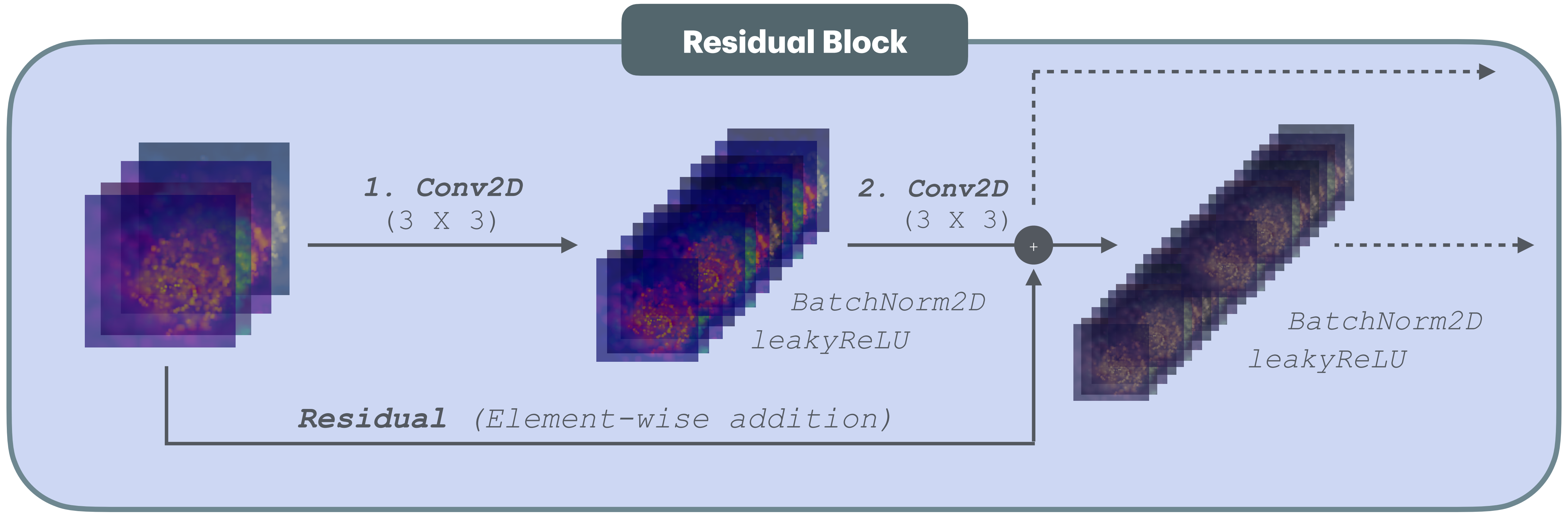
**+ Rotations & Flips**

Increasing dataset size with augmentations



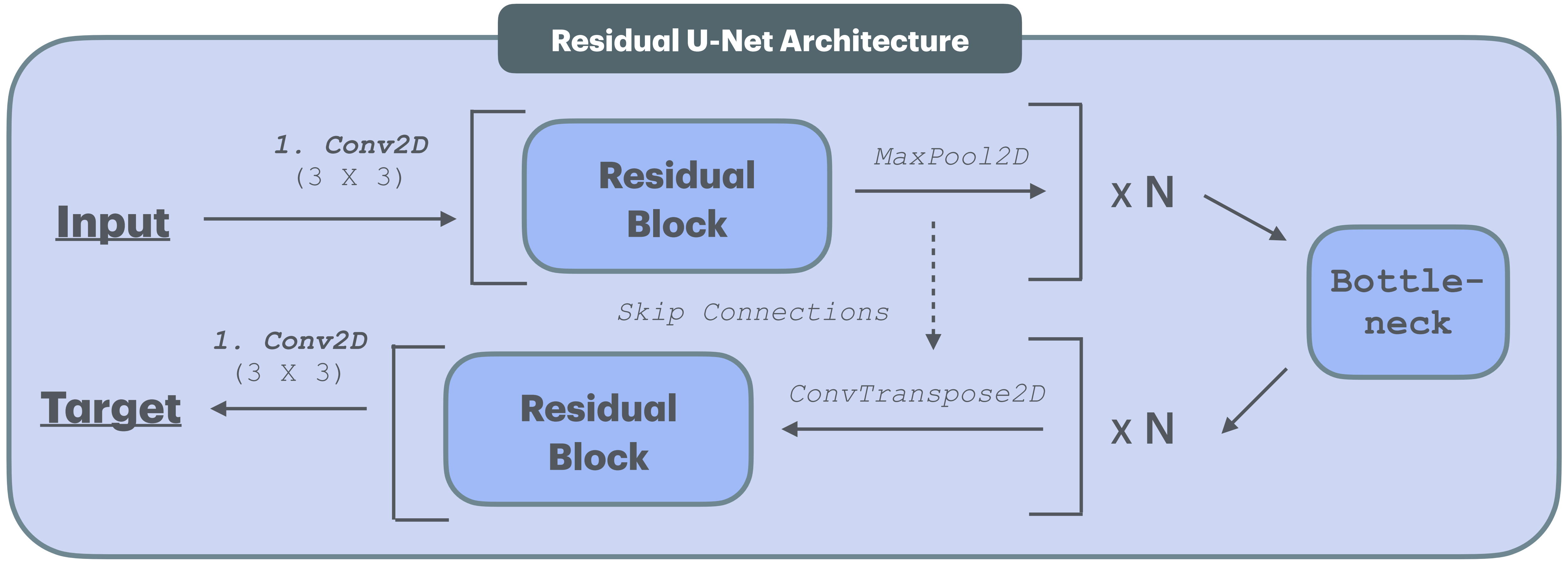



# Alternate approach:
Can we directly leverage spatial correlations to constrain SFR?




## Evaluation

**Evaluation** was conducted with training on 6 galaxies, validating on one and testing on one completely unseen galaxy (all inclinations, ensuring no data leakage)



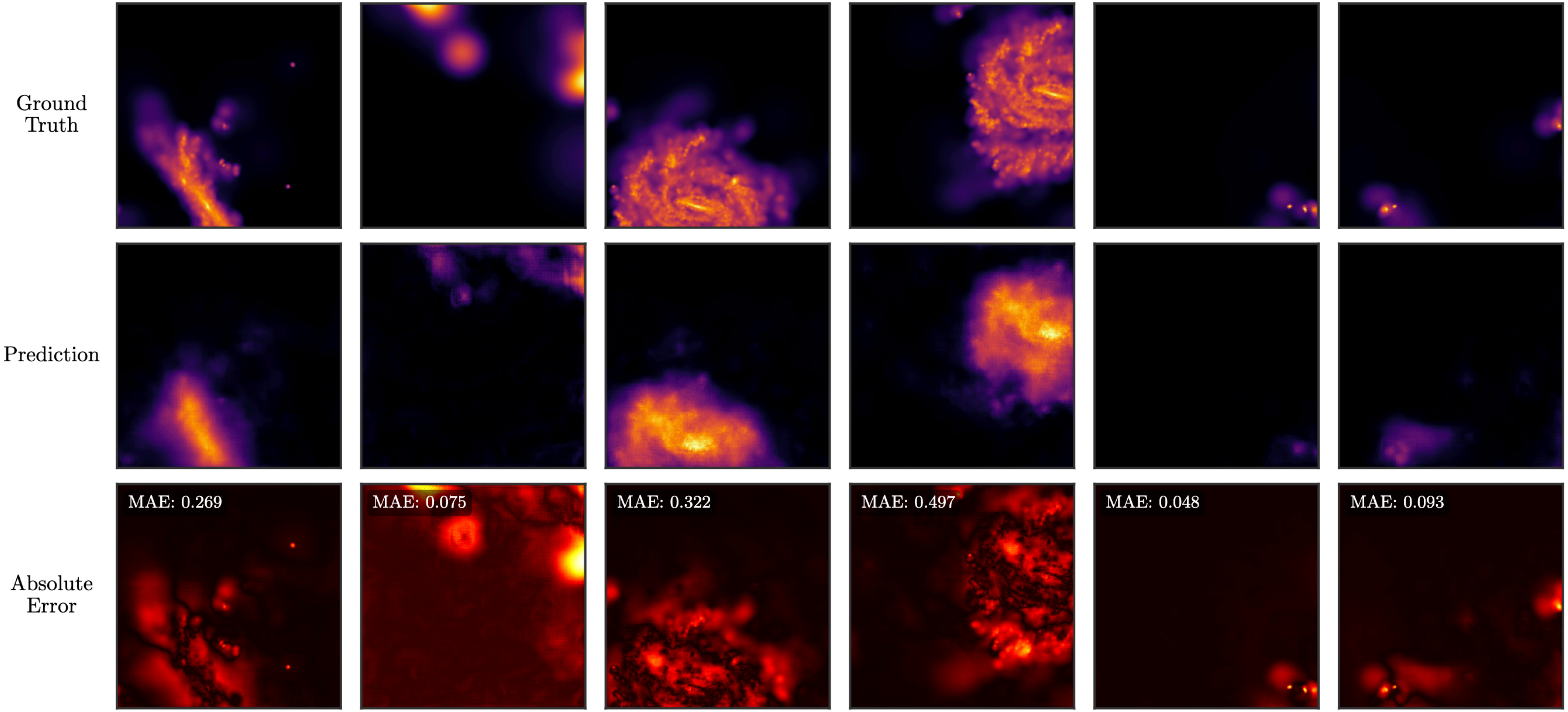

## Training Model

**Optimization:** `AdamW` with `weight_decay` = $10^{-4}$ and a `ReduceLROnPlateau` scheduler.

**Loss Function:** `MSE Loss`

# ResNet Predictions

Yes, it is bad, BUT-




Overfitting and lack of generalisation - probably direct consequence of **only 8 unique galaxies**

Network unable to learn the general physical connections YET

# Need more unique galaxies!!!


**Table 1.** FIRE-2 simulations included in Data Release 2

| suite name | physics variation | final redshift | # of snaps. | halo mass [$M_\odot$] | # of sims. | simulation name | directory |
|---|---|---|---|---|---|---|---|
| Core | Base | 0 | 601 | $10^9$ | 1 | m09 | core/ |
| | | | | $10^{10}$ | 2 | m10(q, v) | |
| | | | | $10^{11}$ | 6 | m11(b, d, e, h, i, q) | |
| | | | | $10^{12}$ | 8 | m12(b, c, f, i, m, r, w, z) | |
| | | | | $10^{12}(\times 2)$ | 3 | m12(ThelmaLouise, RomeoJuliet, RomulusRemus) | |
| | Dark Matter Only | 0 | 61 | $10^{10}$ | 2 | m10(q, v) | core/ dm_only/ |
| | | | | $10^{11}$ | 5 | m11(d, e, h, i, q) | |
| | | | | $10^{12}$ | 9 | m12(b, c, f, i, m, q, r, w, z) | |
| | | | | $10^{12}(\times 2)$ | 3 | m12(ThelmaLouise, RomeoJuliet, RomulusRemus) | |
| | Later Reionization | 0 | 601 | $10^9$ | 1 | m09 | core/ reionize_later/ |
| | | | | $10^{12}$ | 3 | m12(f, i, m) | |
| | MHD+ | 0 | 61 | $10^9$ | 1 | m09 | core/ mhd/ |
| | | | 61 | $10^{10}$ | 2 | m10(q, v) | |
| | | | 61 | $10^{11}$ | 9 | m11(a, b, c, d, e, h, i, q, v) | |
| | | | $61-309$ | $10^{12}$ | 4 | m12(f, i, m, z) | |
| | Cosmic Ray | 0 | 61 | $10^9$ | 1 | m09 | core/ cosmic_ray/ |
| | | | 61 | $10^{10}$ | 1 | m10v | |
| | | | 61 | $10^{11}$ | 8 | m11(a, b, c, d, e, h, i, v) | |
| | | | $61-601$ | $10^{12}$ | 4 | m12(f, i, m, z) | |
| Massive Halo | Base | 1 | 278 | $0.3-5\times10^{13}$ | 8 | A1, A2, A4, A8, B1, B2, C1, C2 | massive_halo/ |
| High Redshift | Base | 5 | 68 | $10^9$ | 2 | z5m09(a, b) | high_redshift/ |
| | | | | $10^{10}$ | 6 | z5m10(a, b, c, d, e, f) | |
| | | | | $10^{11}$ | 9 | z5m11(a, b, c, d, e, f, g, h, i) | |
| | | | | $10^{12}$ | 5 | z5m12(a, b, c, d, e) | |
| | | 7 | 42 | $10^{11}$ | 3 | z7m11(a, b, c) | |
| | | | | $10^{12}$ | 3 | z7m12(a, b, c) | |
| | | 9 | 27 | $10^{11}$ | 3 | z9m11(a, b ,c) | |
| | | | | $10^{12}$ | 3 | z9m12(a, b, c) | |
| Boxes | DM Only | 0 | 11 | - | 4 | L86, L108, L136, L172 | boxes/ |

**What I have now** and have run analysis with

**Next set of data** I am in the process of obtaining

Niranjan Chandra Roy

Dr. Daniel Anglés-Alcázar

UCONN
UNIVERSITY OF CONNECTICUT

**Ongoing** run

# Summary

- **Deep and Sparse Denoising Benchmarks**
  - First review addressed, resubmitted to A&A *(2026)*
  - Developed `GalCubeCraft` - Lightweight mock spectral cube simulator

arXiv

- **Beyond Moment0 Galaxy Property Inference**
  - Developed spaxel-based approach - Variational Gaussian Process Regression and Computer Vision-based approach - Residual U-Net to constrain star formation rate.
  - Currently limited by lack of unique galaxies, BUT (near) **future work:**
    - Increasing data set size by at least 12 unique high-z massive galaxies for generalisation
    - Constraining multiple physical properties (gas mass, stellar mass) using GPR
    - Leveraging spectral features of the cubes in form of a Residual UNet**3D**
    - Feature importance analysis of emission lines and emission morphology
    - Hyperparameter optimisation of ML methods for tighter constraints.