European Organisation for Astronomical Research in the Southern Hemisphere

Atacama Large Millimeter/submillimeter Array

ESO ALMA Support Centre

Internal ALMA Development Study

ALMA 2030: Development Study on Bayesian Adaptive Interferometric Image Reconstruction Methods

Mid-term REPORT

Fabrizia Guglielmetti (ESO)

December 2022

Abridged

The Atacama Large Millimeter/submillimeter Array (ALMA) with the planned electronic upgrades will deliver an unprecedented amount of deep and high resolution observations. Sparse sampling, large variety of celestial sources' morphology and their intensities, instrumental responses, pervasive presence of noise increase complexities to the demanding task of image reconstruction. Wider fields of view are possible with the consequential cost of computing time. Alternatives to commonly used applications in image processing have to be sought and tested.
Currently ALMA is generating 1 TB of scientific data daily. Within the next decade, at least one order of magnitude of increased daily data rate is foreseen. Receivers and correlator upgrades will improve ALMA sensitivity and observing efficiency. In terms of imaging products, ALMA will produce single field and mosaic cubes of at least two orders of magnitude larger than the current cube size in the GB regime. Since the number of observed spectral lines at once will be duplicated, advanced algorithms are needed to provide shorter processing times while handling larger images. Additionally, the imaging algorithms must provide robust and reliable results to reduce human intervention.

We propose innovative research and developments exploring artificial intelligence methodologies, tackling imaging of interferometric data with the scope to refine the Common Astronomy Software Applications (CASA) package for the use of ALMA. Current challenging issues as stopping thresholding, continuum subtraction, proper detection of extended emissions, separation of point-like sources from diffuse emissions, weak signal detection, analysis of mosaics and speed up procedures are addressed. A number of prototype softwares compatible with CASA are developed for the use of ALMA data employing astro–statistics and astro–informatics techniques. In addition, a feasibility study will be performed to compare accuracy, precision, speed, data volume handling for the prototyped softwares keeping the CASA CLEAN task as a reference.

Executive Summary

Through this study we provide the initial exploration of novel imaging techniques applicable to ALMA data and in support of CASA. The novel imaging techniques are suited for the analysis of large data volume, thus enabling efficiency improvements in data processing while requiring the least amount of human intervention.

We employ two distinct software to analyse ALMA data in view of the challenges arised by the ALMA2030 development roadmap. The two techniques are different in nature, one based on astro–statistics ($RESOLVE$) the other on astro–informatics (a $Deep\ Learning\ Pipeline$ developed during this study, named $DeepFocus$). Both techniques demonstrated to be equipped with essential strengths and by required features the Big Data era is longing for.

Using real and simulated data sets, we investigate these Machine Learning techniques to tackle the synthesis imaging challenges in view of the ALMA2030 era. Specifically for real ALMA data we make use of Science Verification (SV) data and the DSHARP Large Program. SV is a process by which data quality is assured for scientific analysis and their products are ready for quality assessment. The DSHARP Large Program provides the perfect sample for testing $RESOLVE$'s potential in detecting extended emissions. For ALMA simulated data we make use of CASA, the software package commonly used to calibrate, image and simulate ALMA data. Comparisons of these techniques with tCLEAN [1] are performed. A documentation of the developed software is accessible on the GitHub and GitLab Integration Platforms.

$RESOLVE$ is a robust algorithm and founded on a principled method. It outperforms current imaging techniques for the detection of diffuse emission. Complex structures in the celestial signal and point-like sources are well detected. Super-resolution imaging is achieved. The required data for this method are the ALMA observed calibrated visibilities. The input variables, as celestial signal and power spectrum, are initialized and

estimated by the data during the optimization to the most probable image configuration. The reconstructed images provide for a reliable solution with no need of extra human intervention. $RESOLVE$ is applied to ALMA continuum images. Applications of the technique to ALMA aggregate continuum, cubes as well as mosaicking are ongoing. Although $RESOLVE$ is computationally expensive, the algorithm delivers in addition to the deconvolved image other informative products as uncertainty map, power spectrum, and its uncertainty. This software has the potentials to lay the foundations for designing a fully automated pipeline capable to learn from the data.

$DeepFocus$ demonstrated high image fidelity and high-performance computing for image reconstruction on ALMA data cubes. The technique makes use of ALMA dirty cubes. $DeepFocus$ learns the celestial sources, the noise, the instrumental point spread function from the input data. The learning process occurs both in spatial and frequency domains, accounting for the information of each channel propagated throughout the frequency space. It allows for extreme data compression by leveraging both spatial and frequency information. In the near future, $DeepFocus$ will be applied to ALMA continuum images and trained and tested on a large variety of celestial signals. $DeepFocus$ rests on a flexible and modular system, whose pipeline can be modified and refined to overcome the complexities at hand. Nevertheless, astro–informatics has the potential to revolutionise data management in Science Archives. ALMA is currently producing roughly 300-400TB worth of raw-data and reduced data products per year. Currently, 27.7% of the total data volume in the archive is occupied by images. $DeepFocus$ may allow to create images on user demand with a one-click system through a web-interface. Moreover, catalogues creation of stored data per ALMA Cycle is feasible in an automated fashion.

As a conclusive remark, based on the current investigations, $RESOLVE$ is the algorithm of choice for robust diffuse emission and faint source detection while $DeepFocus$ implemented within CLEAN will make a huge contribution to the performance problem introduced by the Wideband Sensitivity Upgrade (ALMA2030). Because of the planned upgrades, ALMA image analysis strives for algorithms capable of discovering through the data, to adapt when given new data and to get the most out of those.

Contents

Acronyms

ALMA	Atacama Large Millimeter/submillimeter Array
CASA	Common Astronomy Software Applications [2]
DeepFocus	Deep Learning Pipeline [3]
DL	Deep Learning
DSHARP	Disk Substructures at High Angular Resolution Project [4]
GRU	Gated Recurrent Unit
IFT	Information Field Theory
IoU	Intersection over Union
MLS	Machine Learning Solution
MPA	Max Planck Institute for Astrophysics, Garching
ResNets	Residual Neural Network
RESOLVE	Radio Extended SOurces Lognormal deconVolution Estimator [5]
SNR	Signal to Noise Ratio
SV	Science Verification data
SZ	Sunyaev-Zel'dovich effect
UniFi	University of Florence, Italy
UniNa	University of Naples Federico II, Italy

1 Study evolution and achievements

The starting of this study coincided with the beginning of the pandemic and a significant slow down affected the developments during 2021. Although the outbreak, crucial advancements were made to demonstrate that machine learning is suited to tackle both detection of extended emissions and to provide speed-up procedures.

A well attended kick-off meeting occurred on December 11, 2020. The great affluence to this meeting demonstrated the importance to address current open questions in image analysis from a different perspective and technology. In support of CASA, the goal of this study is to investigate advanced techniques, i.e. those algorithms capable to learn from the data and to overcome known imaging issues as detection of extended emissions and speed up convergence procedures. A core team was created with astro-informatics, astro-statistics, ALMA and CASA experts to address the goals of this study.
Following the requests during the initial meeting, seminars were provided by Kumar Golap, Tak Tsutsumi, Ben Bean (NRAO) and Dirk Petry (ESO) to address the tCLEAN task specifications, mosaicking issues and the ALMA data structure.

A weekly work cadence is established to keep sense of achievements and to quantify the progress. A group in Teams was created to keep track of the advancements, including minutes of the meetings, presentations, publications and sharable files.

The milestones of this study can be summarized as follows:

1. December 2020: **Kick-off** meeting
2. December 2020-February 2021: coordinated **lectures** on ALMA data structure and analysis
3. September 2021+: **ALMA SV data** applications
4. March 2022: Proposal submission to **C2PAP**, titled “Enabling Big Data Science in ALMA 2030 with machine learning” (Delli Veneri M., Guglielmetti F., Testi L.)
   - 3 computational nodes, 256 GB of RAM and 3 TB of storage
5. April 2022: application to **Leibniz Data Center** through C2PAP support
   - Cloud services: CPU : 1 computational node, 256 GB of RAM, 3 TB storage, GPU: 2 GPUs with at least 12 GB of memory, 480 hours for Training and Inference, between 720 and 2160 hours of fine tuning depending on the cluster availability
6. May-June 2022: Michele Delli Veneri is an ESO **visitor** for one month
   - Contributed talk at **SciOps2022**, titled “Data Cleaning, Detection and Characterization of Sources in ALMA Data through Deep Learning” (Delli Veneri M., Tychoniec Ł., Guglielmetti F., Villard E., Longo G.)
   - **ESO Azure MLS** accounts (Delli Veneri M., Tychoniec Ł., Guglielmetti F.) CL-PROD-001 MLS workspace. This is a Microsoft Data Science Solution for Machine Learning supported by ESO. MLS environment tested with our software, but on limited practice due to unsafe ownership on the software.
   - Contributed talk at **AI Forum**, titled “3D Source Detection and Characterization of Sources in ALMA Data through Deep Learning” (Delli Veneri M.)
   - **ALMA data cubes simulation development and acceleration**. Available on GitHub with 100 ALMA data cubes (256x256x128) created in 20 seconds.

7. July 2022:
   - **MaxEnt2022 conference**: Invited talk "Bayesian and Machine Learning Methods in the Big Data era for astronomical imaging" (Guglielmetti F., Arras P., Delli Veneri M., Enßlin T., Longo G., Tychoniec Ł., Villard E.) [6]
   - **MaxEnt2022 conference**: Contributed talk "Bayesian statistics approach to imaging of aperture synthesis data: RESOLVE meets ALMA " (Tychoniec Ł., Guglielmetti F., Arras P., Enßlin T., Villard E.) [7]
8. August 2022: article submission to **MNRAS** "3D Detection and Characterisation of ALMA Sources through Deep Learning" (Delli Veneri M., Tychoniec Ł., Guglielmetti F., Longo G., Villard E.) [3]
9. September 2022: **NFMCP (ECML-PKDD 2022)** workshop: Contributed talk "3D Detection of ALMA Sources through Deep Learning" (Delli Veneri M. et al.)
10. October 2022:
    - **ARC All–Hands Meeting**: Contributed talk "Update on the ESO internal ALMA development study: ALMA2030 Bayesian Reconstruction through Adaptive Image Notion" (Guglielmetti F.)
    - **RESOLVE Workshop 2022** at MPI for Radioastronomy. Contributed talk "Bayesian and Machine Learning Methods in the Big Data era for astronomical imaging" (Guglielmetti F. et al.)
    - **Comparison** of DeepFocus with tCLEAN on a set of 1000 ALMA data cubes simulations.
    - First applications of RESOLVE to aggregate continuum and mosaicking in the search of SZ effect.
11. November 2022:
    - Contribution to **ORIGINS Annual Science Meeting** for GPU usage employing the C2PAP infrastructure.
    - Taking advantage of the *DRM F2F meeting* at NRAO (Charlottesville, VA): Presentation of the study to **Amanda Kepley (NRAO), Ryan Loomis (NRAO), Theodoros Nakos (JAO)** with a talk titled:"ESO internal ALMA development study: ALMA2030 Bayesian Reconstruction through Adaptive Image Notion" (Guglielmetti F.). Proposal from Ryan Loomis to make use of the new software in an hybrid system in CASA.
12. January 2023: **Contribution to VLTI and ALMA imaging** workshop, since RESOLVE and DeepFocus have the potential to overcome the challenges also for VLTI synthesis imaging.
13. **RESOLVE is on GitLab** at MPCDF: All improvements obtained during the development study are publicly available.
14. The **DeepFocus pipeline** will be publicly available on **GitHub** as soon as the PhD Thesis of Delli Veneri M. is accepted. The software is going to be available at this link.

2 The Core Team

Fabrizia Guglielmetti is ARC Scientist at ESO and P.I. of this study. Her interest is on the development of advanced imaging techniques. She earned her PhD in astrophysics on image analysis employing Bayesian Probability Theory for a joint estimation of background and celestial signals (LMU, 2010).

Philipp Arras, formerly PostDoc at MPA, is expert on Bayesian statistics applied to imaging algorithms for radio telescopes. He developed theoretical models for inference algorithms for the unification of calibration and imaging, multi-spectral imaging, polarization imaging, data fusion with single dish data and Very Long Baseline Interferometry (VLBI). Expert on the RESOLVE technique.

Michele Delli Veneri (UniNa) has a master in Astrophysics and currently PhD student at the Electrical Engineering and Information Technology Department of the University of Naples Federico II. He is the developer and designer of DeepFocus. He gained expertise on ALMA and Square Kilometer Array (SKA) imaging.

Torsten Enßlin (MPA), Associate Professor at the Ludwig-Maximilians-University and Information Field Theory Group Lead. He is expert in image analysis for multi-wavelength observations, with special emphasis on radio synthesis observations.

Giuseppe Longo (UniNa), Professor in astrophysics, chair of the Data Science Initiative at the University Federico II: Data Science program, current president of the International AstroInformatics Association. His research interests are in development and application of novel machine learning methods to a variety of problems.

Łukasz Tychoniec is Research Fellow at ESO, where he conducts his own scientific research using ALMA telescopes, other ESO observatories and the James Webb Space telescope. He is expert in star and planet formation, protostellar jets, protoplanetary disks, early planet formation, submillimeter interferometry. Part of his functional work is dedicated to this study.

Eric Villard, ESO/ALMA Advanced Data Product Scientist (ESO) involved with ALMA observatory in operations since 2010: Commissioning Scientist, System Astronomer, Deputy Head of Data Management Group, Head of the ALMA Array Performance Group.

Since November 2022, Philipp Arras (MPA) moved to new ventures. **Jakob Roth** and **Carmen Blanco** are contributing to the study. They are both working at the IFT group (MPA), lead by T. Enßlin. **Jakob Roth**' PhD topic is Computational Structure Formation in Physical Cosmology. **Carmen Blanco**' Master thesis is focused on detection of SZ effect on ALMA data.
Aida Ahmadi (Allegro ARC node in Leiden) is going to contribute to the study with applications of RESOLVE on own science topic (planet and star formation). Ivano Baronchelli (INAF) is new contributor to the study with his machine learning and ALMA data analysis expertise.

This Study is also supported by Paola Andreani (ESO), Massimo Brescia (UniNa), Stefano Cavuoti (UniNa), Ed Fomalont (NRAO), Alessandro Marconi (UniFi), Federico Montesino Pouzols (ESO), Urvashi Rau (NRAO), Andy Strong (MPE), Udo von Toussaint (IPP), Crystal Brogan (NRAO), Sanjay Bhatnagar (NRAO), Allen Caldwell (MPP), Sandra Castro (ESO), John Hibbard (NRAO), Elizabeth Humphreys (JAO), Mark Lacy (NRAO), Dirk Muders (MPI for Radioastronomy), Martin Zwaan (ESO).

3 Software selection

Applications of two algorithms (RESOLVE and DeepFocus) to ALMA data are shown. RESOLVE and DeepFocus are thoroughly and distinctly discussed in two scientific papers [5] and [3], respectively. Both RESOLVE and DeepFocus bring significant advantages in terms of the quality of data products, and they have the potential to largely improve the user data-processing experience both if used as a standalone tool and in combination with CASA. These algorithms were included in the original proposal. On the one hand, RESOLVE has been successfully applied to VLA data [8], along with the ability to perform calibration and imaging with a joint uncertainty quantification [9]. RESOLVE showed to be a promising candidate also for ALMA data applications and for investigating possible advantages over traditional algorithms. This application exploits the participation of the Information Field Theory group at the Max-Planck institute for Astrophysics. On the other hand, Autoencoders, i.e. neural network architectures, have been explored and set the base of the DeepFocus algorithm. The Deep Learning models employed within DeepFocus took advantage of the team experience of the data science group at the University of Naples on, e.g., SKA data cubes but also on other fields of science (e.g. autonomous driving, medicine). The selection of the Deep Learning models were dictated on optimised tasks most suited to ALMA data. Convolutional Autoencoders, Variational Autoencoders, U-NETs (convolutional neural networks), and Region Proposal Networks were tested with extensive fine-tuning of each model architecture as well as hyper-parameters. DeepGRU was tested against Transformers and Long short-term memory units (LSTMs), while the ResNets were compared against EfficientNets, DenseNets, CNNs and several pre-trained models available within the literature in a transfer learning approach. Due to a recent success on the application of 3D Deep Learning models for the resolution of the imaging problem for SKA, BlobsFinder and DeepGru are planned to be ameliorated, including the introduction of a 3D Classification ResNet.

Two additional algorithms are accounted in the original study's proposal. Firstly, the Bayesian Mixture Model technique [10] is an unsupervised learning method. The mixture model technique [10] is equipped with a defined model to explore the data and extract the required information from the data for a robust background-source separation. Based on Bayesian probability theory, the technique is capable to jointly estimate source signal and background, providing a multi-resolution analysis for the detection of faint sources. This technique for the use of ALMA needs further development and exploration, mainly driven by technology advancements, e.g. [11]. Please note that there is another technique worth of dedicating time and effort, that is described in [12]. However, additional workforce is needed to achieve these developments. Secondly, the maximum entropy deconvolution (MEM) algorithm [13] is not further supported. MEM is a nonlinear deconvolution algorithm and useful to correct for the limited sampling of the u-v plane. However, it assumes a sky brightness of the sources as a prior knowledge. This algorithm showed the prior knowledge not adapting to the data and hindering faint source detection.

4 RESOLVE

The RESOLVE algorithm has been refined for the use of ALMA data. The input data are the ALMA calibrated measurement set (MS). Taking the information from the calibrated visibilities in the u-v plane, the Response operator R acquires the information of the dirty beam. R is an integral ingredient for the reconstruction of the detectable celestial signal. The input data (or calibrated MS) d are modelled as a combination of celestial signal s corrupted by the dirty beam and by the noise n (systematic and random errors): $d = Re^s + n$. Synthesis imaging reconstruction occurs through the posterior estimation of the possible true sky signal configurations arised by the Hamiltonian sampling: $P(s|d) = \frac{e^{-H(s,d)}}{Z(d)}$. More information on the foundations and description of the algorithm can be found at [14, 15, 5, 16, 17].

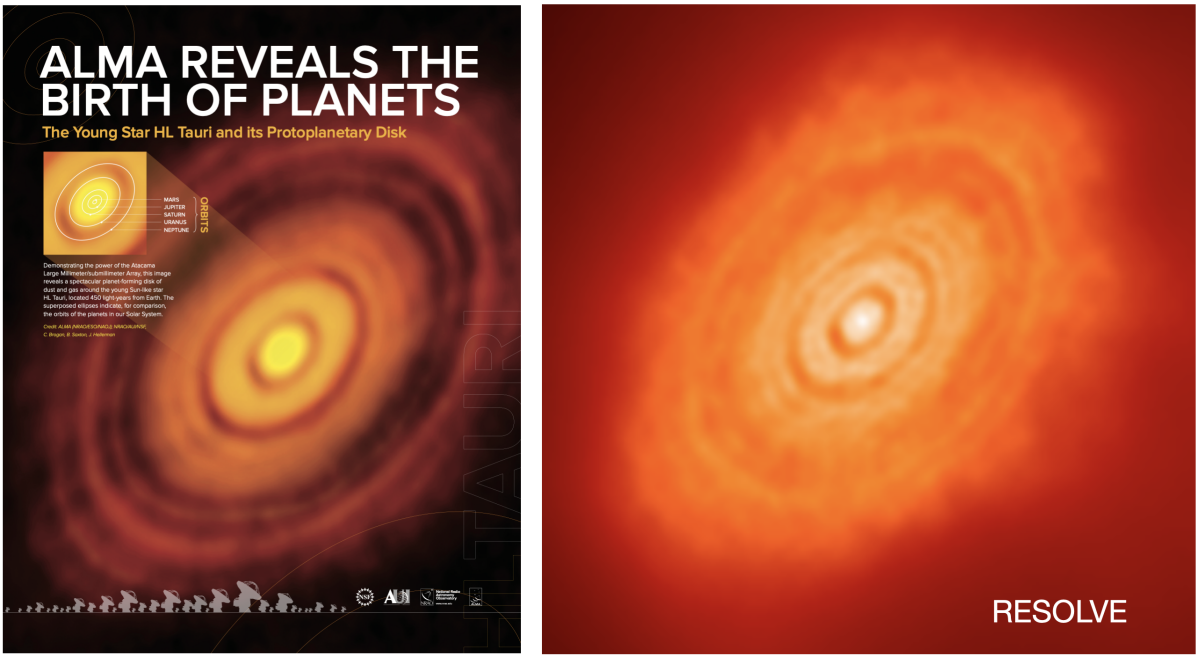

Figure 1: On the left, HL-Tau press release combining bands 6 and 7 [18]. On the right, HL-Tau imaging with RESOLVE on band 6 and one spectral window (one quarter of the available data for band 6).

4.1 Proof of concepts

Working on the u-v plane, the long lasting weighting issue of ALMA visibilities was addressed and solved [6, 19]. The successful application to HL-Tau data (SV data) with RESOLVE and the comparison with CASA provided the proof of concepts. In Fig. 1 (image on the right), the continuum image at 1.3 mm (233 GHz) applied to one spectral window at 1.827 GHz (128 channels) is obtained with RESOLVE. The image on the left shows the reconstructed IF image by the CASA team using the full band 6 and band 7 data [18].
The image, shown on the right of Fig. 1, is derived by the algorithm as representation of the posterior mean of the Hamiltonian sampling of the reconstructed celestial signal. A full set of posterior density function estimations of the celestial signal are provided and shown in Fig. 2. For each $P(s|d)$ a corresponding uncertainty map is estimated. At the converged sample, RESOLVE's answer to the ill-posed problem of image reconstruction provides a posterior mean of the sampled estimated signals and a mean uncertainty map: See Fig. 3, on the upper/lower left are shown the posterior mean of the signal detection and the corresponding uncertainty map. Still in Fig. 3, the power spectrum $P_s(\vec{k})$ of the process that generated the signal s as a function of spatial frequency k and its uncertainty estimation are shown. A representation of the distribution of data weights versus baselines is also reported (right).

4.1.1 The power spectrum

The signal's power spectrum $P_s(\vec{k})$ describes how the signal's variance is distributed over the different frequencies of the signal [14]. Low- and high-frequency modes correspond to large- and small-scale features of the signal, respectively. An example of power spectrum reconstructed from the RESOLVE procedure for image reconstruction is shown in Fig. 3.

The power spectrum contains interesting information about the statistical properties of the physical signal. If the power spectrum falls with higher frequencies, it corresponds to higher variations on large-scale features

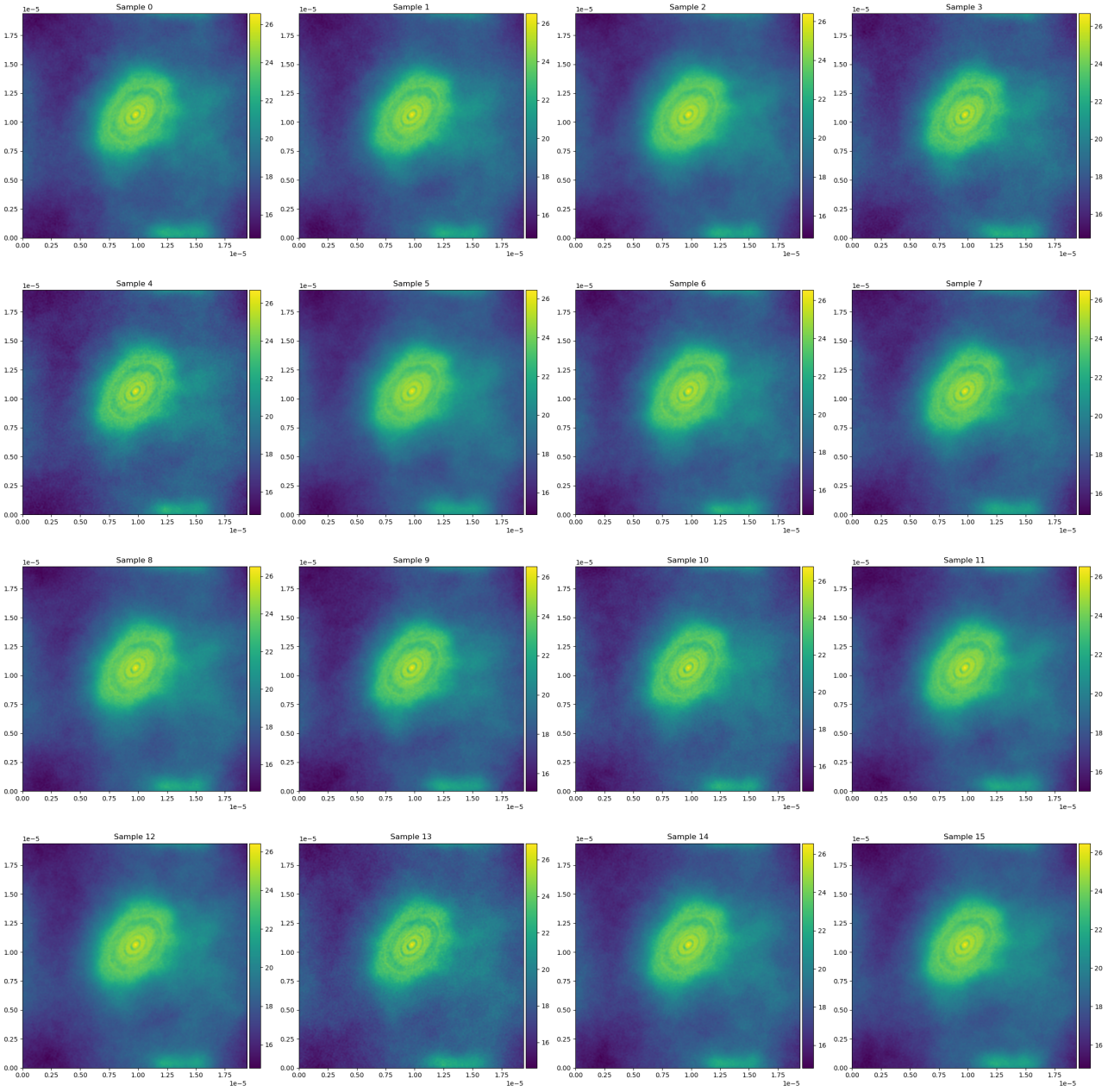

Figure 2: RESOLVE: Set of Hamiltonian samples showing the estimated posterior probability density function of the estimated celestial signal for the HL-Tau data shown in Fig. 1, right. Each sample is similar with negligible deviations in reconstructed intensity and shape. This is because the last iteration of the converged optimization procedure is shown.

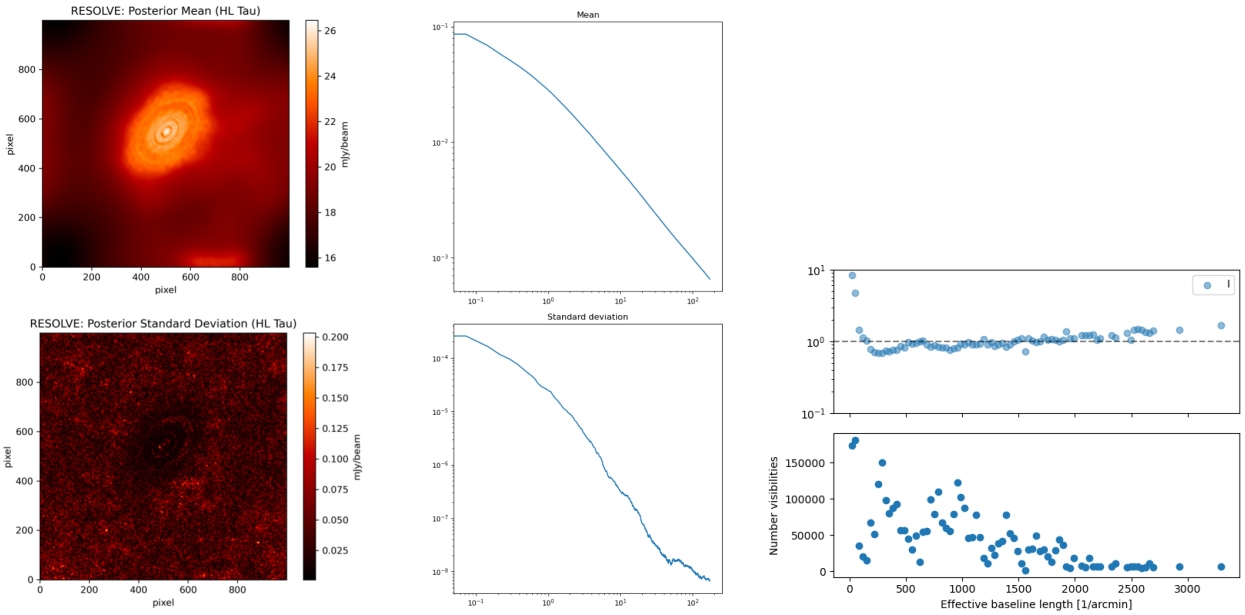

Figure 3: RESOLVE: Left, the mean posterior probability density function (upper) and its uncertainty (lower) of the estimated celestial signal for the HL-Tau data shown in Fig. 1, right. Center, power spectrum estimation corresponding to the estimated posterior mean (upper) and its uncertainty (lower). Right, representation of RESOLVE's distribution of data weights versus baselines (upper) and number of visibility versus effective baseline length (lower) at the convergence iteration.

compared to small-scale features. This is a typical behavior for many physical signals. When analyzed with sufficient resolution, such signals are smooth functions. The slope of the power spectrum determines the degree of smoothness. Steep spectra correspond to smooth signals, while flatter spectra correspond to signals with more small-scale variations. Peaks and bumps in the power spectrum indicate oscillations with the respective frequency. Flat power spectra correspond to white noise.

4.2 Simulations

The major advantage in using simulated datasets for quality assessment of the imaging procedures is the selection of the input parameters, allowing to control the shape and brightness of the sources, as well as configuration and behaviour of the telescope array. The downside of this approach is the difficulty to realistically model the noise acquired during an observation and the simplistic assumptions about the sky brightness. Nonetheless, this test allows to understand how RESOLVE works and to compare source detection and flux estimates with the tCLEAN task [1] in CASA [20, 2]. In the following, two applications are shown that are tailored at testing the algorithm capabilities to overcome source confusion and weak signal detection in single pointing and continuum images. The RESOLVE algorithm is shown to be successfully applied to ALMA simulated data.

The simulation in Fig. 4 aims at challenging the algorithm to detect weak and extended as well as bright point–like sources spread over an ALMA synthetic observation. The simulated data set is processed with RESOLVE and the tCLEAN imaging task in CASA. Taking the simulated data set as benchmark, RESOLVE provides both weak and extended source detection and more realistic source characteristics with respect to tCLEAN. RESOLVE powerfully provides detection of diffuse emission, weak signal, point-sources embedded

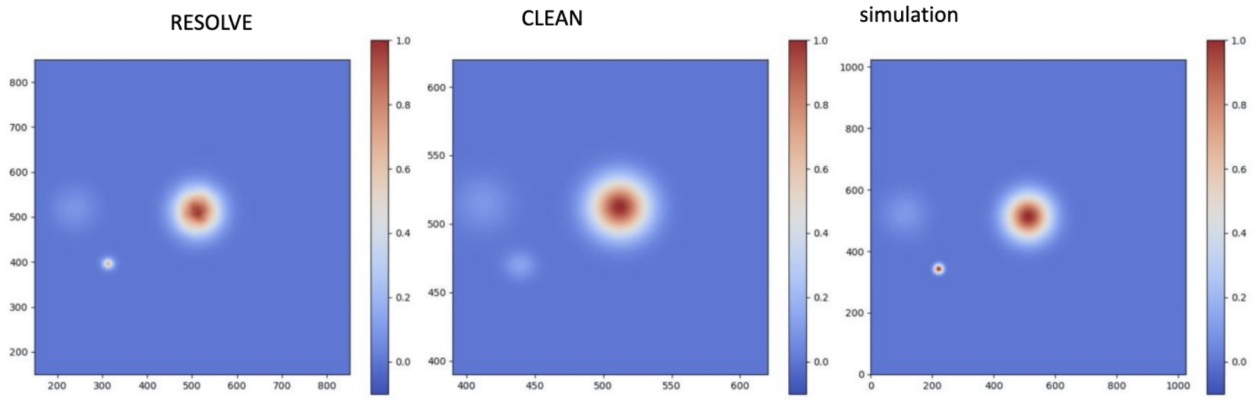

Figure 4: RESOLVE (left) and tCLEAN (center) application to an ALMA continuum simulated data set (right). The colorbar shows a normalized flux intensity to allow for a visual comparison on recovered source characteristics.

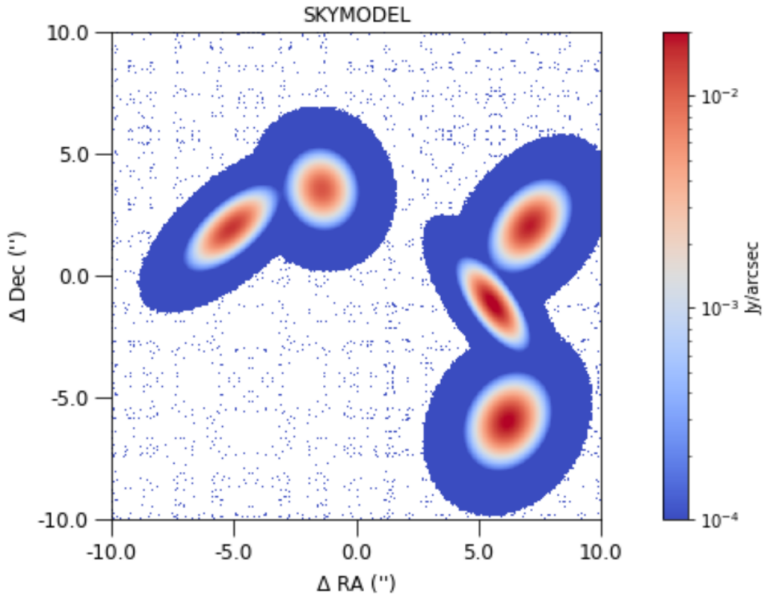

Figure 5: Sky model which serves as an input for creating simulated ALMA Measurement Set: Single pointing and continuum image.

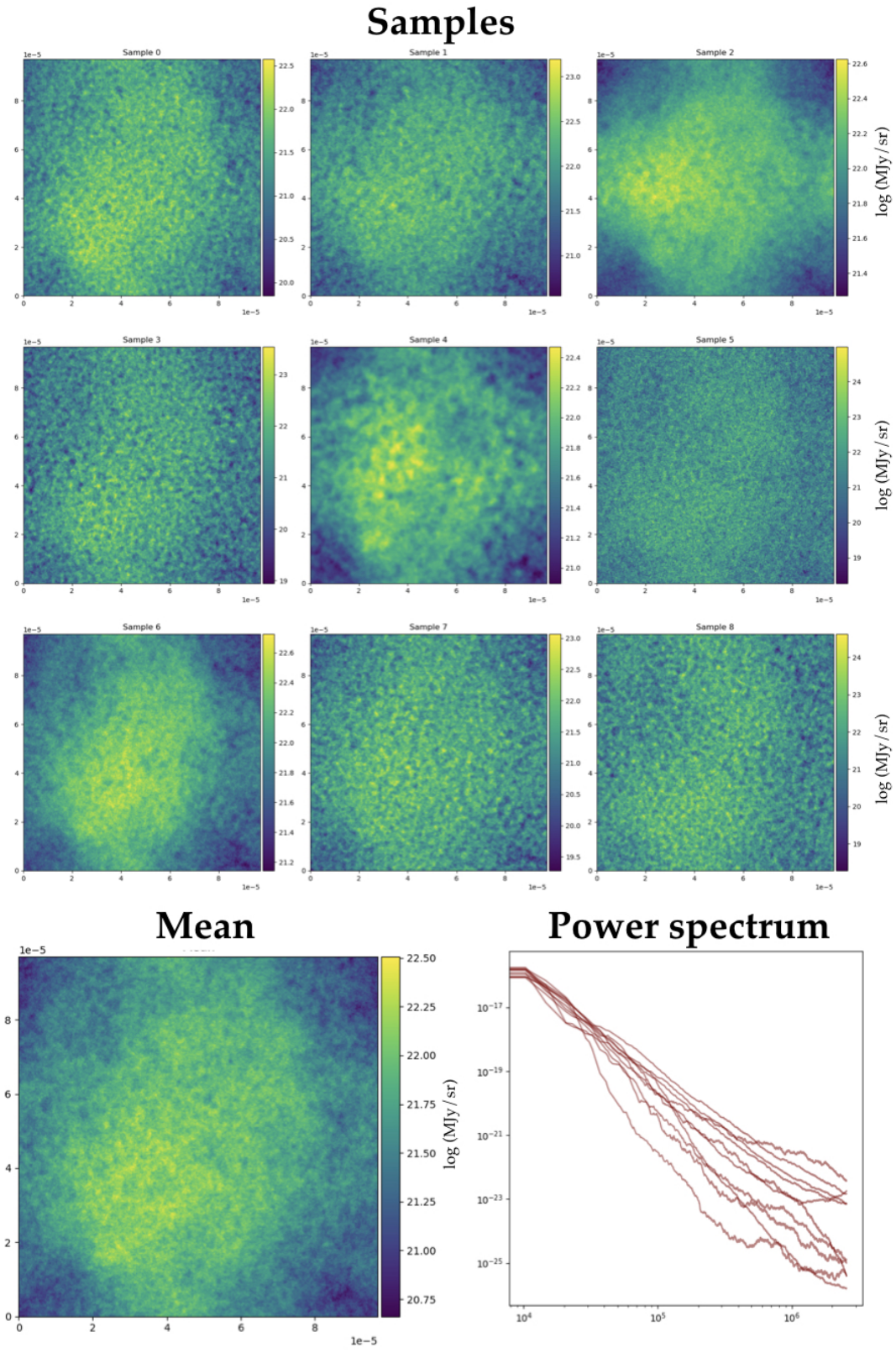

Figure 6: ***Iteration 0****: First set of Hamiltonian sampling from the starting iteration of ALMA single pointing and continuum simulated data set (Fig. 5). The initial uninformative guess of* $RESOLVE$ *is shown.*

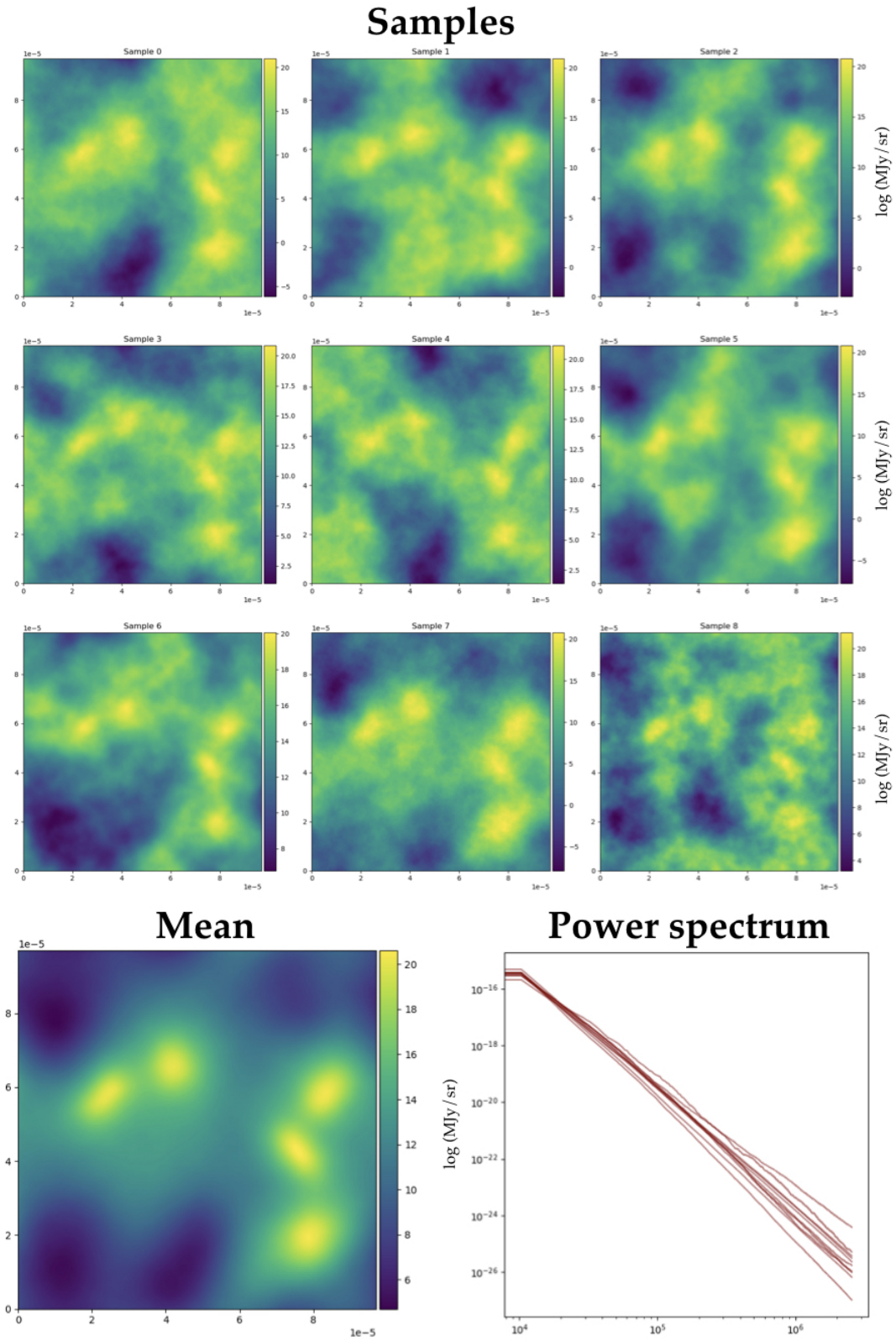

Figure 7: ***Iteration 5****: After 5 iterations, the set of Hamiltonian sampling of the ALMA single pointing and continuum simulated data set (Fig. 5) shows that the algorithm learns about the signal* s.

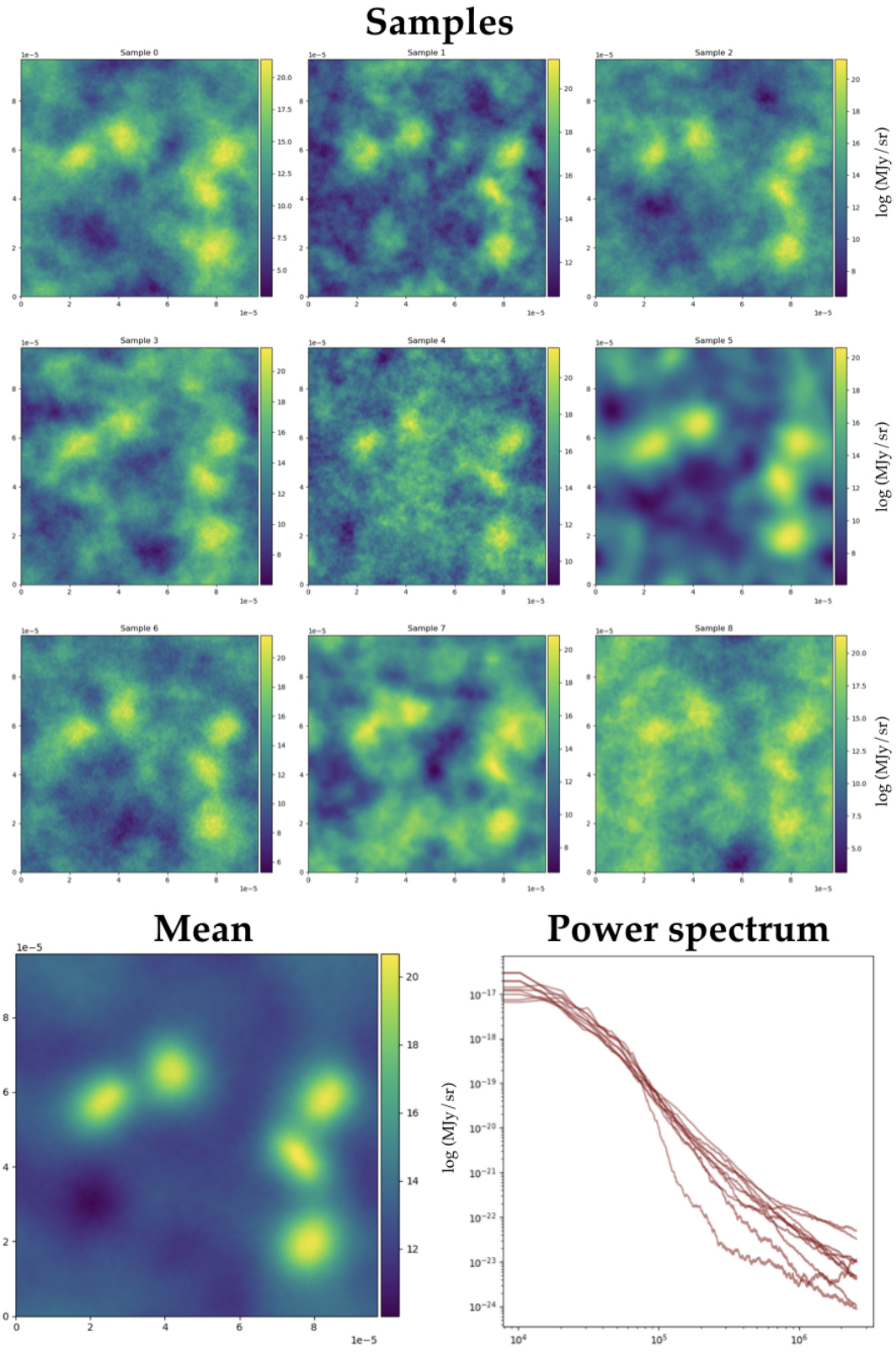

*Figure 8: RESOLVE learns the signal s (Fig. 5) as shown at (almost) convergence stage (**iteration twenty-four**). The mean of the posterior probability density functions of the detected signal s given the data d is shown in the lower left image. The image on the lower right shows the reconstructed power spectrum samples, starting to deviate from the initial uninformative guess.*

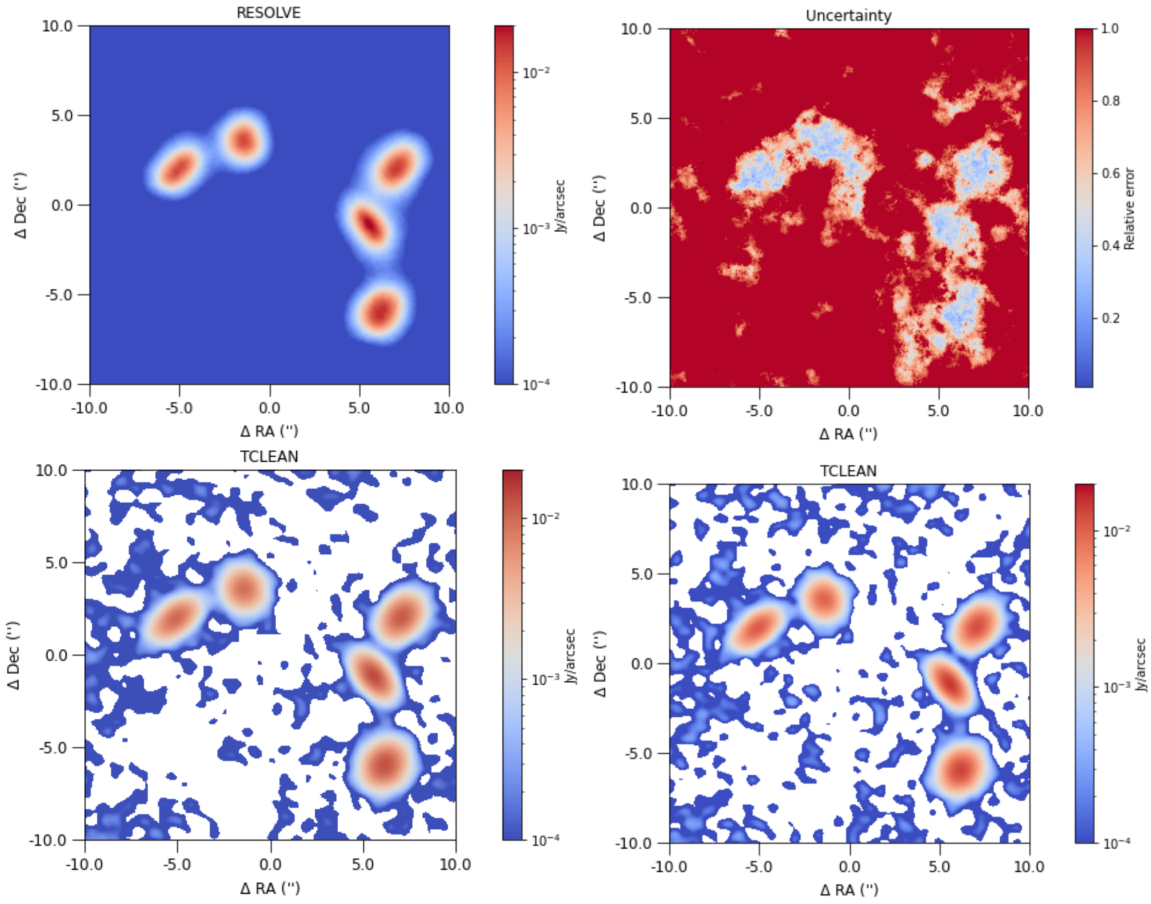

Figure 9: Maps of the brightness distribution in the sky plane of the simulated data (Fig. 5). Upper row: RESOLVE's reconstructed image (left) and its uncertainty map (right). Lower row: tCLEAN solution using natural weighting (left) and robust = 0.5 following the Briggs weighting scheme (right).

Parameter	Mean
Offset	26
Zero mode	1±0.1
Fluctuations	5±1
Power spectrum slope	-2±0.5
Flexibility	1.2±0.4
Asperity	0.2±0.2

Table 1: Example of input parameters for the RESOLVE application to simulated data and its estimated values.

in diffuse emissions.

Fig. 5 shows the sky model generated with simalma task in CASA. A simple 2D array consisting of five Gaussian components of different brightness, size and position angle are generated. The sky model is generated, imposing physical properties to the sky, such as spherical coordinates, pixel dimensions, field-of-view and brightness in physical units. Afterwards, the task is simulating observations of the given sky model with the ALMA observatory. For this specific case we simulate observations with ALMA configuration C-3 at 230 GHz (ALMA band 6), which results in an effective resolution of 0.7 arcsec, due to the longest available baseline being ~ 500 m. The simalma task returns a calibrated MS, that consists of complex visibilities. Those visibilities are a Fourier transform of the sky brightness, therefore inverse Fourier transformation provides a dirty image of the observed sky.

In Fig. 9, lower row, the reconstructed images with tCLEAN are shown. The task tCLEAN is run without any constraint on where to look for point sources in the image (i.e. without any masking as input parameter), for 20000 iterations, or until threshold of 0.3 mJy/beam is reached. The pixel size of the reconstructed image is set to 0.1 arcsec and the image size is set to 512×512 pixels. The noise threshold level of 0.3 mJy/beam is selected so that the tCLEAN algorithm does not attempt to find sources from the residual image consisting purely of noise. The RMS of the dirty image is 0.1 mJy/ beam and the SNR is 3σ. In the default settings, the weighting of the baselines is set to $natural$ (image on the left), which means it associates the baseline with weight proportional to the sampling density (i.e. the most covered baselines have the highest weight). Since at larger scales the sampling is much denser, this puts more weight on lower resolution, which results in achieving lower resolution than the sky model image. Natural weighting will give the largest beam and the best surface brightness sensitivity. Therefore we also attempt imaging with Briggs weighting [21] with robust parameter 0.5 (image on the right), which moves the balance of weighting toward longer baselines increasing the resolution but decreasing signal-to-noise ratio.

For RESOLVE image of the simulated data, 30 iterations are needed to reach convergence. In Figs. 6, 7, 8, the first, sixth, twenty-fourth iterations of RESOLVE are shown. The $P(s|d)$ samples are displayed, including the posterior mean $< P(s|d) >$ and the power spectrum samples $P_s(\vec{k})$. Given an initial state of no information, the algorithm searches for the optimal sky configuration from the given data (visibilities and dirty beam). In Fig. 9 (upper row), the converged posterior mean probability density function of the detected simulated signal $< P(s|d) >$ and its uncertainty map are shown. The uncertainty map displays the relative error as a measurement of precision, to determine the magnitude of the absolute error in terms of the actual size of the measurement process. No assumption was made on the presence of point sources in the data. Input parameters of the RESOLVE run are summarized in Table 1: See [8] for a detailed explanation of the parameters.

RESOLVE's reconstructed image exhibit a smooth background in agreement with the simulation (Fig.5). Bayesian probability theory allows us to estimate the uncertainty of the hypothesis of detecting a source signal. What is not a celestial source and potentially noise is captured by the description of uncertainty on the full image. Therefore, the uncertainty map does not only address detected sources. Structures in the image away from the detected sources provide evidence that the uncertainty quantification can not be a smooth

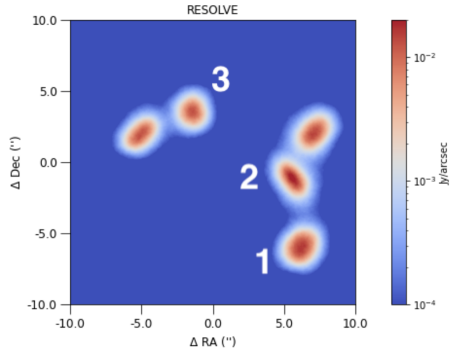

Figure 10: Example of reconstructed fluxes (sources $1-3$) from the application of RESOLVE *and* $tCLEAN$ *to Fig. 5*

Comp	model	tCLEAN natural	tCLEAN Briggs	RESOLVE
1	34.11	24.67	24.96	26.68±5.76
2	23.31	21.68	22.19	23.59±6.72
3	16.69	16.31	16.57	17.82±3.86

Table 2: Peak flux (in Jy $arcsec^{-2}$*) of different components on the simulated image.*

Parameter	Mean
Offset	20
Zero mode	1±0.2
Fluctuations	3±1
Power spectrum slope	-4±2
Flexibility	4±0.8
Asperity	2±0.8

Table 3: Example of input parameters for the $RESOLVE$ *application to DSHARP data's Sz114 and its estimated values.*

function. The tCLEAN reconstructed images are characterized by negative values in several areas. Negative values arise by the recurrent application of subtraction during the major-minor cycles and residual estimations. Although the non-physical representation of the background image with tCLEAN, it is known that those variations are only noise. Moreover, a comparison of the reconstructed fluxes $[Jy \cdot arcsec^{-2}]$ with the two algorithms is shown in Table 2 where the corresponding components numbered 1-2-3 are indicated in Fig. 10. The model column provides the simulated sky model's flux values. Both algorithms provide flux reconstruction close to the ideal values. tCLEAN with $natural$ weigthing provides the worst estimates, being those simulated sources mainly point-like sources. RESOLVE gives reconstructed flux values (and their uncertainties) 7% closer to the simulated values with respect to tCLEAN.

4.3 DSHARP Large Program

The application of RESOLVE to real data is performed on a well-studied data set, whose survey was designed to optimize the spatial resolution and contrast sensitivity to continuum emission substructures [4]. This is the ALMA Large Project 2016.1.00484.L, known as Disk Substructures at High Angular Resolution Project (DSHARP). The sample is shown in Fig. 11. The data were taken at long baselines allowing us to test RESOLVE on high resolution data while detecting extended emission. In fact, a secondary goal of this project was to identify corresponding gas structures and infer other relevant bulk disk properties (e.g. geometry). Expertise on this science topic is brought into the study by Ł. Tychoniec (ESO). For more details on this application please refer to [7]. The protoplanetary disk Sz114 is chosen within the sample, because of its relatively smooth structure [7]. Sz114 is shown as first image on the left, second row, in Fig. 11.

Both tCLEAN and RESOLVE are applied to a single spectral window to directly compare the two techniques. Currently, RESOLVE does not have a capability to make images of multiple spectral windows. From the publicly available calibrated MS, we extracted spectral window 9 with split task in CASA and binned it into

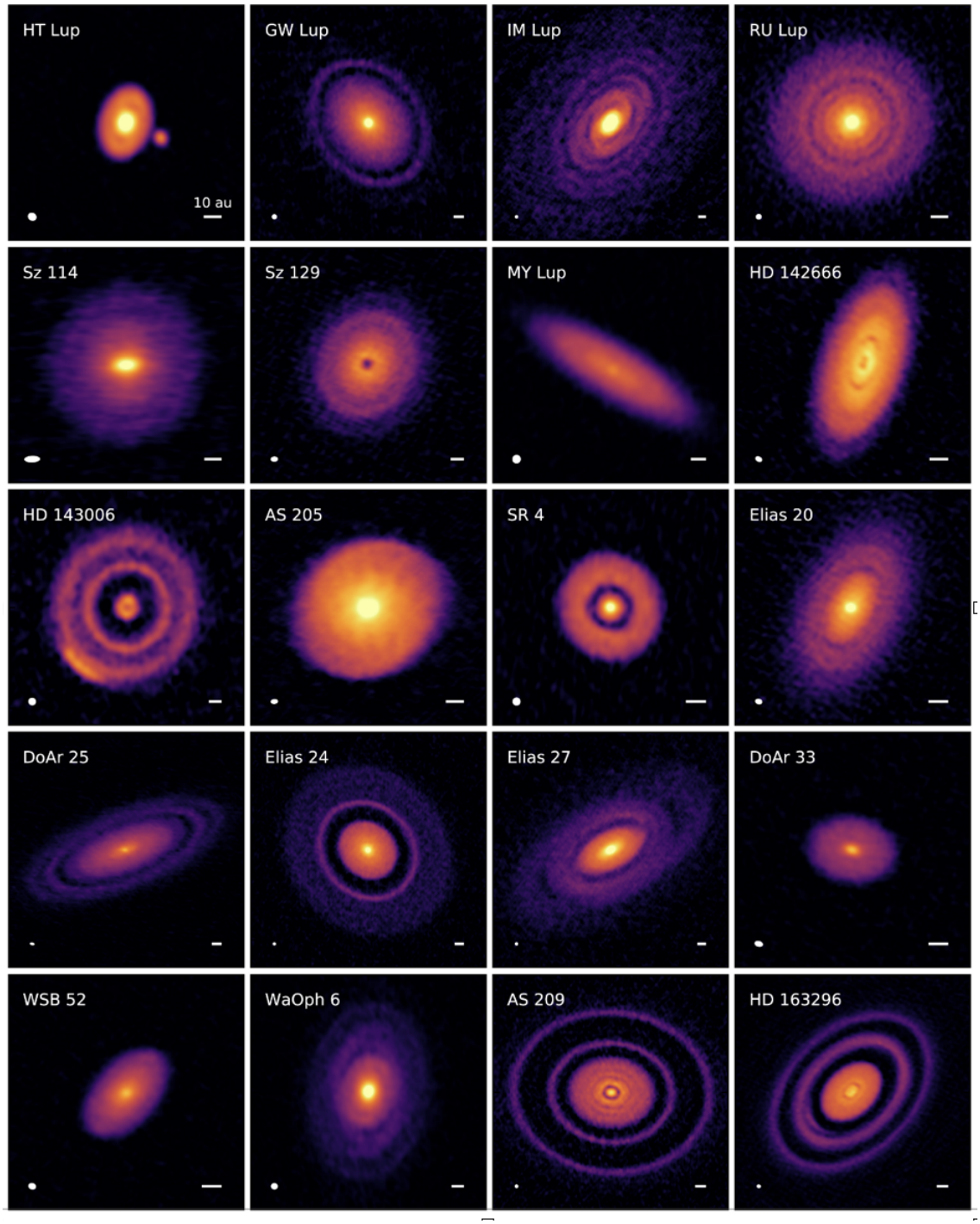

Figure 11: This image is taken from [4]: Gallery of 240 GHz (1.25 mm) continuum emission images for the disks in the DSHARP sample. Beam sizes and 10 au scalebars are shown in the lower left and right corners of each panel, respectively.

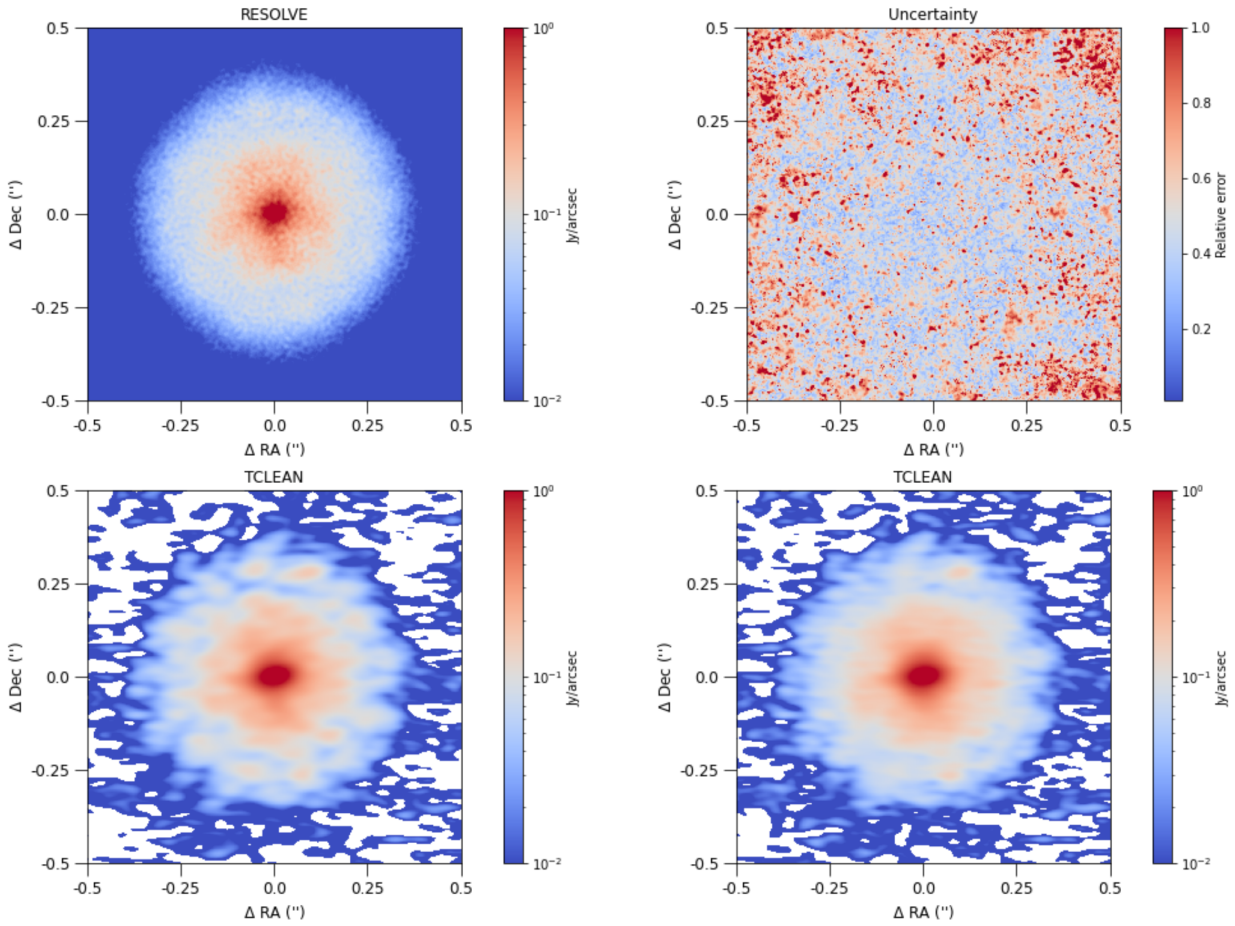

Figure 12: Image reconstruction with $RESOLVE$ and $tCLEAN$ of ALMA observation of Sz114 (Fig. 11). $Upper\ row$: reconstructed image by $RESOLVE$ (left) and the relative error map (right). $Lower\ row$: $tCLEAN$ reconstructed images with standard cleaning algorithm (right) and with multi-scale algorithm (left).

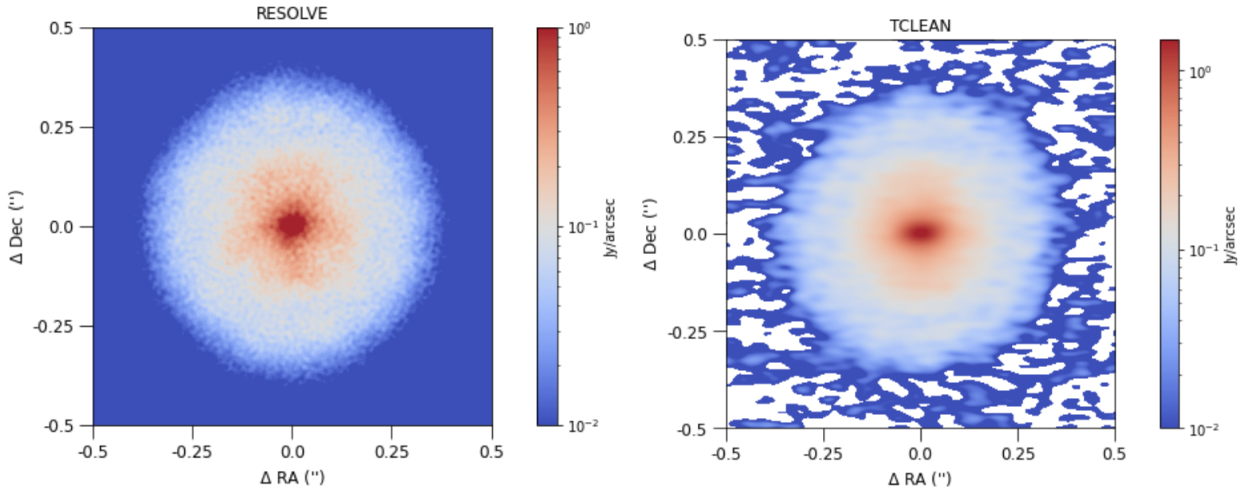

Figure 13: On the left, the reconstructed image by RESOLVE of Sz114 as shown in Fig. 12. On the right, the reconstructed image of Sz114 with tCLEAN as delivered by the Large Program team [4] (In table 4 indicated with "best tclean").

Radius arcsec	best tclean	tclean hogbom	tclean multiscale	resolve
peak	1.57	1.55	1.49	5.8±3.82
0.06	8.63	8.85	8.95	9.48±1.07
0.15	22.79	22.82	22.53	23.21±0.68
0.35	47.45	47.07	47.24	47.24±2.79

Table 4: Peak (in Jy·arcsec^2) and integrated flux at given radii of the Sz114 disk.

a single channel.

The estimated input parameters of the RESOLVE application to Sz114 is shown in Table 3. Please note the power spectrum slope with respect to the result in Table 1. The achieved steeper power spectrum slope on the real application is reasonable since less structures are present at small scales with respect to the simulated case. This result is in agreement with the reconstructed image of Sz114 by the RESOLVE algorithm (Fig. 12), upper left, and the uncertainty map, upper right. The uncertainty map shows that no aggregate structures are detected and that the quantified uncertainty increases at increasing signal detection (absolute uncertainty).

Standard and multiscale tCLEAN imaging are used because the observed disk presents large variety of spatial scales. Images are characterized by 0.005 arcsec (pixel size), 1024×1024 pixels and 20000 iterations or until the noise threshold of 0.05 mJy is reached. In case of multiscale clean, three scales (0, 7, 28 pixels) are specified to find three types of sources in the data: point source (scale=0), extended (Gaussian) components with FWHM of 7 and 28 pixels. The resulting tCLEAN images are displayed in Fig. 12, lower row.

In Fig.13, a visual comparison between the reconstructed images of Sz114 from RESOLVE and from tCLEAN. This tCLEAN image is produced employing the aggregate continuum as delivered by the Large Program team [4]. We define this tCLEAN solution on the aggregate continuum as the best tCLEAN image. RESOLVE is capable to achieve improved emission detection. Table 4 provides a comparison of integrated flux estimates measured over different areas of the disk (radius). The three tCLEAN solutions are very similar. The fluxes provided by RESOLVE are close to the benchmark values. The only deviation found is at the peak. tCLEAN approximate the dirty beam to a Gaussian, while RESOLVE encodes the dirty beam from the uv-plane into the Response operator. The distribution at the peak collects more flux than a Gaussian. Therefore,

it is reasonable that at the peak the measured flux is larger than the value provided by tCLEAN.

RESOLVE is providing an high-fidelity image of the protoplanetary disk Sz114 and robust estimation on the fluxes and their uncertainties.

4.4 Ongoing developments and outlook

The technique RESOLVE, applied to ALMA data, provides for image reconstruction from sparse samples and proper uncertainty quantification. Bayesian probability theory allows one to extend from point estimates to a distribution of solutions whose critical mass under the multi-dimensional integral contains the information to quantify optimal values and their uncertainties. Information Field theory allows us to tackle imaging on a pixel–free approach. Input information to the imaging technique is the observed and calibrated visibilities. The dirty beam is derived from the uv coordinates and incorporated into the Response operator in the deconvolution process. The input data are processed within a probabilistic approach allowing one to model the celestial signal and noise in the data in one unique algorithm. Products of the technique are reconstructed ALMA deconvolved image and its uncertainty map, the power-spectrum and its uncertainty, estimates of input parameters and their uncertainties. Through simulated and real data, we demonstrate that RESOLVE is applicable to ALMA data. Detections of extended emissions and structures in protoplanetary disks are promising. Considering the application to ALMA data with RESOLVE occurred on one spectral window for continuum detection only, when comparing with tCLEAN' solution on the aggregate continuum, RESOLVE is capable to provide an improved detection of Sz114 while employing one quarter of the available data.

The capabilities of RESOLVE are currently under elaboration in order to

- apply to aggregate bandwidth for continuum imaging,
- detect the Sunyaev-Zel'dovich (SZ) effect (absorption detection),
- analyse mosaics of images.

RESOLVE showed outstanding detection of celestial signal in emission. We want to challenge the algorithm to detect absorption features. Considering that a substantial cosmological radio beackground exists, the SZ effect arises due to Comptonization of the background leading to a signal decrement at the ALMA high frequencies. Currently an ALMA band 7 data set is under scrutiny. This application is planned to be completed by February 2023.

When the application to the aggregate continuum will succeed, the next step is to tackle cube imaging. Ideally, the Information Field Theory (IFT) framework shall allow for a joint continuum and line detection. This exploration may require few months of intense work and may go beyond the available time frame of end 2023.

RESOLVE has the functionality to be executed in parallel. Therefore we are going to test on our ARC cluster at ESO the execution speed of RESOLVE in order to define its computational cost. Another exploration worth of dedication and effort is the exploitation of distributed data objects to further accelerate the technique [22]. While the latter application is intensive for providing conclusive results in a short time range, during the beginning of 2023 we are going to have an answer for the former one. We consider the utilization of distributed data objects of low priority at the moment and hardly it will be tackled in 2023.

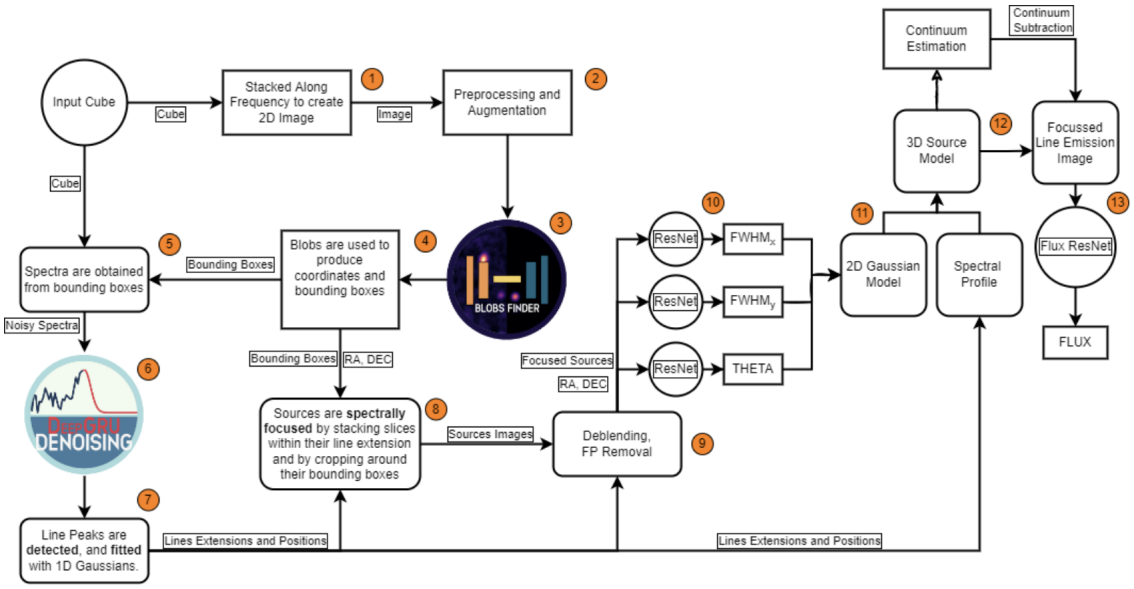

Figure 14: Deep Learning Pipeline schema. Numbers indicate the logical flow of the data within the pipeline.

5 DeepFocus

The Deep Learning Pipeline, named DeepFocus, is fully described in the refereed article accepted for publication on MNRAS: click on this link to be redirected to the article [3].
DeepFocus is designed as a modular system allowing to expand the pipeline according to the problem at hand. In operations we always need to cope with new modes or instrumental improvements (e.g. electronic upgrades) [23]. New units can be connected to the pipeline as well as replacements with others during software refinements. The schema of the composition of the Pipeline can be seen in Fig. 14. Currently DeepFocus is composed by six Deep Learning models: a Convolutional Autoencoder for source detection within the spatial domain of the integrated data cubes (Blobs Finder, stage 3), a Recurrent Neural Network for denoising and peak detection within the frequency domain (DeepGRU, stage 6), and four Residual Neural Networks for source characterization (ResNets, stage 10).
The algorithm solves for $I^D = I_{db} * I + n$, where I^D, I_{db}, I and n are the ALMA dirty cube, the dirty beam, the target sky image and any additional noise in the image, respectively. The information contained in each channel in the ALMA cube is accounted in the whole frequency domain, correlating information between pixels along the frequency axis of the cube. It has been demonstrated that the combination of spatial and frequency information has the capability to improve completeness while decreasing spurious signal detection. Comparison with tCLEAN perfomed analysing a set of 1000 simulated ALMA data cubes indicated that DeepFocus improves in speed up procedures by a factor of at least 140.

5.1 Proof of concepts

Supervised Machine Learning techniques need training, testing and validation to become that powerful tool to outperform any current algorithm to date in processing time. Most of ALMA archived data are galaxies, with compact and almost point-like shape. The potential capabilities of this algorithm have been tested on ALMA SV interferometric data: BR1202-0725 [24]. Please refer to the article [6] (reachable clicking this link) to get more details on this application (Subsection 2.2 and Figure 4).

This trial shows that both galaxies in the ALMA cube are detected with DeepFocus, while the standard approach in CASA missed the serendipitous quasar (southernmost object). In [24], self-calibration was performed to detect this serendipitous source. The source fluxes derived by DeepFocus agree with the ones reported in [24]. The computing time for the image restoration of one ALMA cube for BR1202-0725 with DeepFocus occurred in $\sim 35\ \mu s$.

5.2 Simulations

Simulated ALMA cubes are used for training/testing/validation as well as reliability and quality assessment of the developed Deep Learning pipeline. Section 3 of the journal article "3D Detection and Characterisation of ALMA Sources through Deep Learning" [3] provides a detailed description of the developed artificial data sets. The CASA simulator capabilities [2],[25] were extended for the creation of ALMA cubes. 2D Gaussian Components in the spatial plane with 1D Gaussian component (emission lines) in frequency space are used to create the artificial emission lines in addition to the continuum signal. ALMA Cycle 9 C-3 configuration with 43 antennas was chosen within the simalma task in CASA. ALMA interferometric measurement sets are created employing the CASA' simobserve task. Dirty cubes are produced employing tCLEAN. Corrupted by white noise, the dirty cubes' RMS is adjusted to a wished SNR.

A set of 5000 ALMA cubes (360x360 pixels, 128 channels and total bandwidth of 1.28 GHz) are created with a source at the center and other randomly distributed in the image with random extension. The brightest source is located at the center and characterized by a SNR > 10. The minimum and maximum flux densities generated are respectively 0.97 and 407.4 mJy/beam. Uniformity on the distribution of simulated source parameters is achieved (Fig. 3 in [3]). See upper and middle rows of each set in Figs. 21 and 22 for some examples of produced ALMA dirty cubes and model images.

At the following link in GitHub, the python package developed to create the simulations of ALMA cubes for observations of galaxies and/or point sources can be downloaded.

5.3 Novelty of the method

The pipeline architecture can be roughly divided into three phases, based on the assumptions made on the data. First, sources are assumed to be present within the image and the algorithm is trained to detect those sources. Second, sources to be deemed true must show emission lines in the frequency domain. The algorithm searches for those emission lines by removing noise in order to boost SNR and recognise spectral peaks. If spectral peaks are found, the algorithm preserves the sources for characterization, otherwise those initial identifications are discarded as false detection. The latter operation is performed through SNR and geometrical criteria involving the reference integrated dirty image and all images produced by integration along the detected emission ranges. Third, sources passing the selection criteria are fed to an array of ResNets to regress the morphological source parameters.

5.3.1 Blobs Finder

Blobs Finder solves for the deconvolution problem in the image domain, i.e. to recover the normalized denoised integrated sky images from the integrated dirty images (stage 3 in Fig. 14). The output probability maps are censored and all pixels with probability higher than a given threshold are connected into potential sources through a friend of friend algorithm. Bounding boxes are extracted around the islands of connected pixels to define source spatial boundaries. Bounding boxes around source candidates are used to extract dirty spectra from the input dirty cube by summing, for each frequency slice (channel), all pixels within the bounding boxes. Figures 15, 16, and 17 show, respectively, an example of an input integrated dirty cube containing 6 simulated

sources (green boxes) with two spatially blended sources, the target sky model image (green and red highlight the target and the predicted bounding boxes), and the 2D prediction map (red bounding boxes).

5.3.2 DeepGRU

The obtained spectra are standardized and fed to DeepGRU (step 6 in Fig. 14) which is tasked to solve a 1D denoising problem and outputs 1D probabilistic maps of source emission lines which are then analyzed in search for peaks. Each peak is fitted with a 1D Gaussian function. Position z and extension $\Delta_z = 2 * FWHM_z$ (where $FWHM_z$ is the FWHM of the Gaussian peak) are recorded. Fig. 18 shows the dirty spectrum extracted from the two blended sources shown in Fig. 17, and the DeepGRU's predicted emission probability map. Blue and red vertical bars limit the true and predicted, respectively, emission ranges of the two sources within the spectrum. In order to detect possible false positives produced by Blobs Finder, all potential candidates showing no meaningful peak in their spectra are removed. If more than one peak is found alongside the spectrum, three possibilities may arise: detected peaks may indicate distinct celestial sources which are spatially blended, detected peaks may belong to the same source, or one or more peaks are false detection(s). None of these possibilities can be excluded a priori. On each peak we perform *spectral focusing*, i.e. we crop a 64×64 pixel image around the source center in the spatial plane and integrate within the peak extension in frequency. In order to estimate the SNR of a source, two SNR measurements are accounted:

Global SNR:

$$SNR \stackrel{\text{def}}{=} \frac{median(x_s(r))}{var(x_n(R-r))} \tag{1}$$

where $x_s(r)$ are the pixel values of the source within a radius r inscribing the bounding box, and $x_n(R-r)$ are the pixel values within an annulus of internal radius r and external radius R which has the same area of the inscribed circumference;

Pixel SNR:

$$snr \stackrel{\text{def}}{=} \frac{x_i}{var(X)} \tag{2}$$

where x_i is the value of the given pixel, and $var(X)$ is the variance computed on the full image.

The two signal-to-noise ratio measurements are used to distinguish falsely detected from true sources and to deblend overlapping sources within a blob. Fig. 19 summarises the false positive detection pipeline. In case of not blended detections, the process works as follow:

- if $SNR \geq 6$ eq. (1) (empirical bright source SNR threshold) in the integrated dirty cube and the source is not flagged for deblending, the detected source is focused;
- if $SNR < 6$ eq. (1) in the integrated dirty cube, the detected source goes through a check for focus:
  - if SNR eq. (1) increases, the source is kept;
  - otherwise the source is discarded as false positive.

This is the condition marked with 1 in Fig. 19. In case of blended sources, the procedure for the identification of true sources proceed as follows:

1. focusing on the highest peak (primary peak) by integrating within its extension allows for the SNR eq. (1) calculation. The same logic described above is followed.
2. the snr measurement eq. (2) is used to identify the pixel with strongest intensity in the image $p(x, y)$. This reference pixel is used in the next phase of the deblending process.

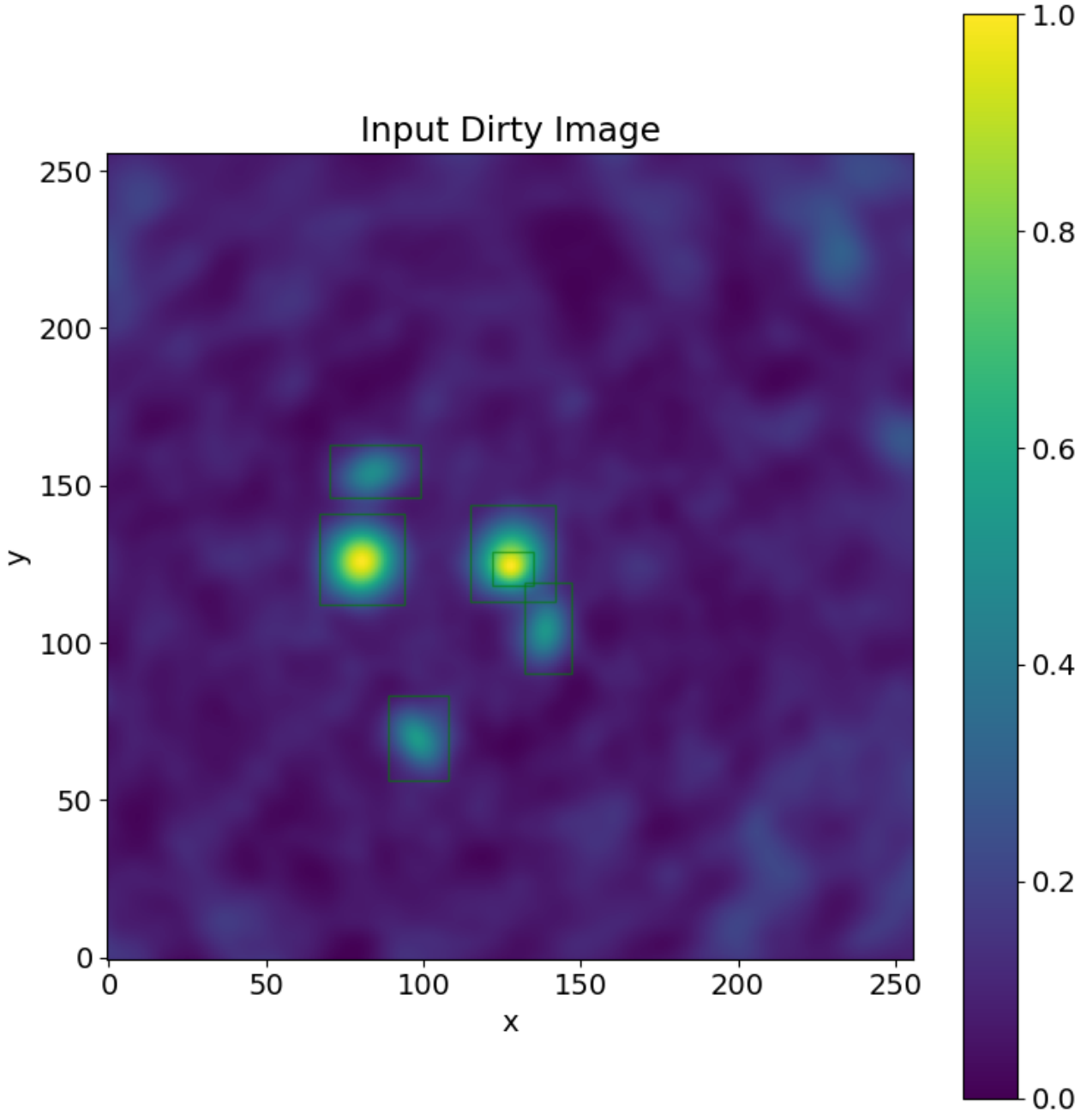

Figure 15: An example of Blobs Finder's input 2D integrated dirty cube produced by integrating an input dirty cube over the entire frequency range. Superimposed in green, are the target bounding boxes outlining the emissions of the 6 sources present in the cube. The image contains an example of two spatially blended sources located around the centre of the image, one is a bright point-like source, the other a fainter and diffuse source laying behind.

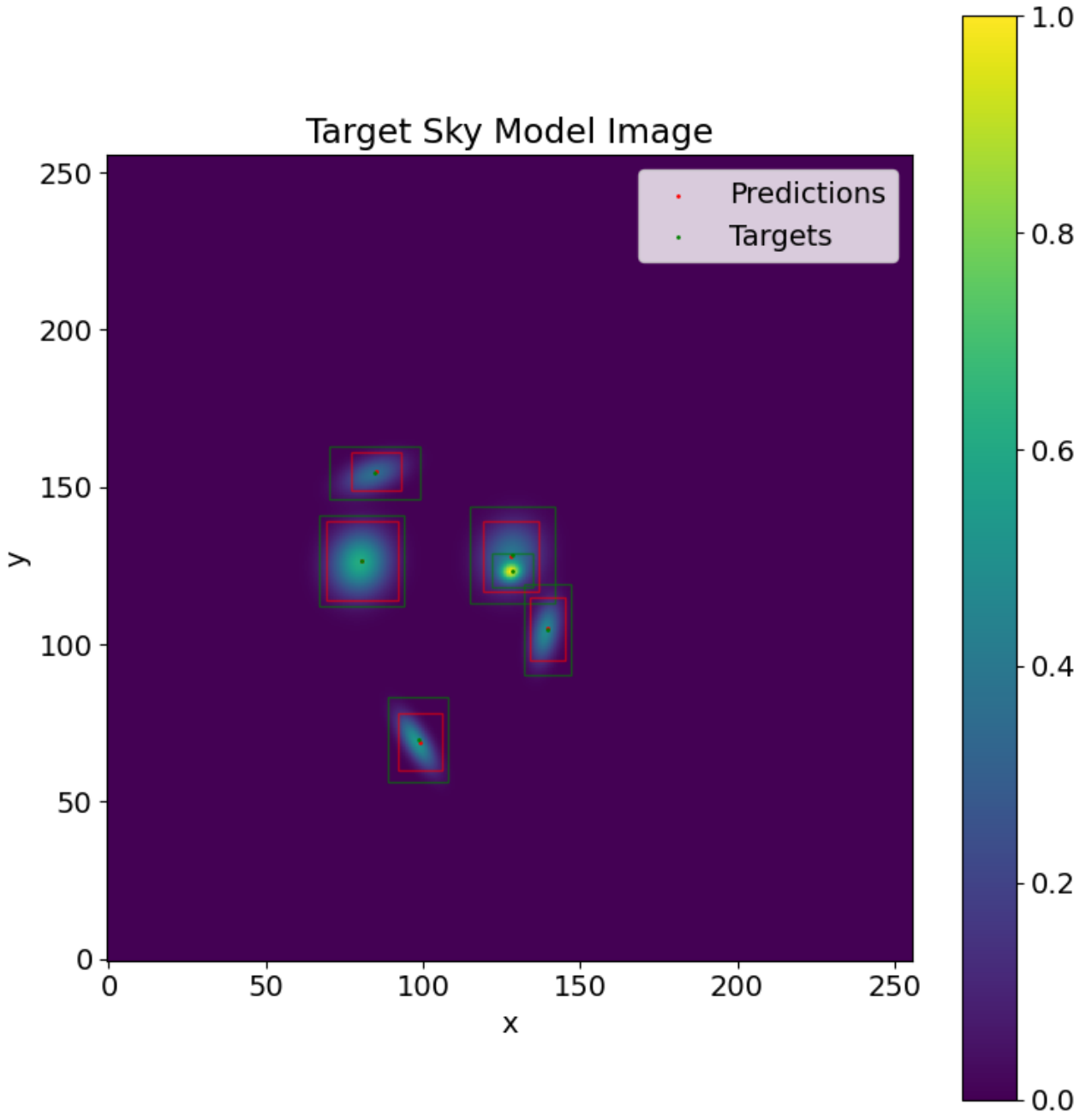

Figure 16: An example of Blobs Finder's target 2D Sky Model image with the target bounding boxes highlighted in green and the predicted bounding boxes extracted through the tresholding operation on Blobs Finder's probabilistic output, highlighted in red. Predicted and true bounding box centers are also plotted as, respectively, red and green dots.

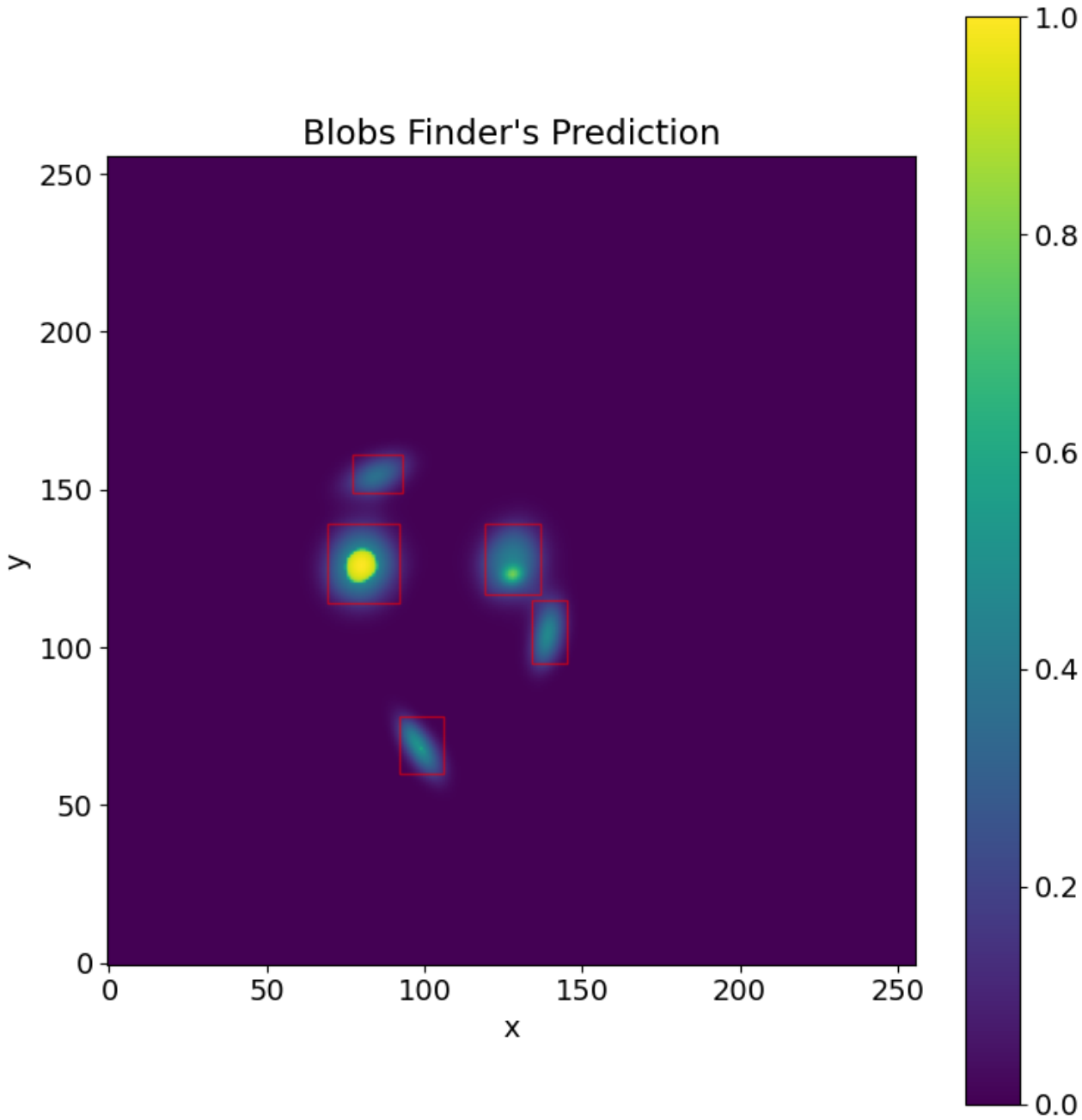

Figure 17: An example of Blobs Finder's output 2D probabilistic source detection map with the predicted bounding boxes extracted through thresholding, highlighted in red.

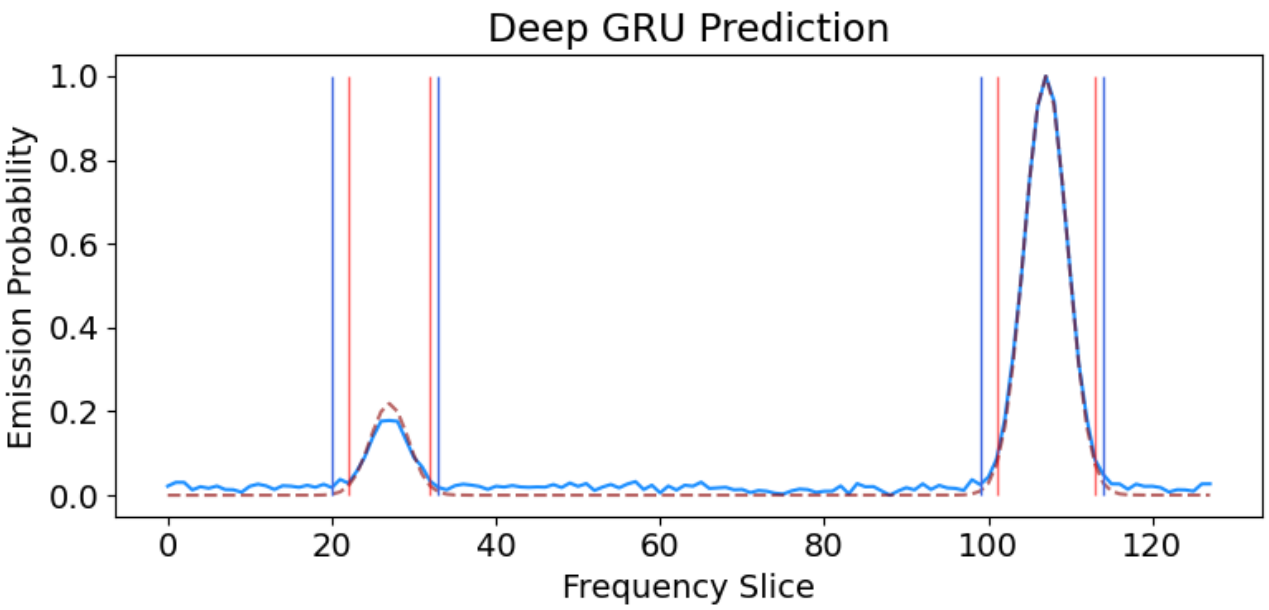

Figure 18: In blue the dirty spectrum extracted from the central source bounding box predicted by Blobs Finder (Fig. 16), in dotted-red the DeepGRU's prediction. Vertical blue bars delimit the true emission ranges, while red bars the predicted emission ranges. A secondary fainter source emission peak is detected by DeepGRU and thus the source is flagged for deblending.

3. secondary peaks not overlapping in frequency with the primary peak are analysed:
   - the snr measurements eq. (2) find the reference pixel in the image $s(x, y)$.
   - friend of friends algorithm is used to link pixels around the new reference pixel in $s(x, y)$ until a saturation level is reached when calculating the SNR eq. (1) iteratively.
   - a bounding box is created to encompass all the selected pixels, and a $[64, 64]$ pixel image is cropped around the bounding box.
4. secondary peaks overlapping in frequency with the primary peak are inspected:
   - if the primary and secondary peaks coincide spatially ($p(x, y) = s(x, y)$), then the secondary peak is discarded as a false detection (condition marked as 2 in Fig. 19).
     - e.g. DeepGRU may predict a single peak as two separate peaks or Blobs Finder predicts a single true source as two very close blobs.
   - if $p(x, y) = s(x, y)$ but SNR eq. (1) increases : the secondary peak is deemed as part of the primary source and the source emission range is extended accordingly.
     - it may happen if DeepGRU overpredicts the true emission range.

Finally all spectrally focused sources with SNR lower than 1 eq. (1) are flagged and removed.

Fig. 20 gives an example of Spectral Focusing applied to the potential sources detected by Blobs Finder and DeepGRU in the test cube already displayed in Fig. 15. By focusing on the two peaks detected by the DeepGRU Fig. 18, the two blended sources produce two different images (Focused Source 0 and 1) which can be analysed independently.

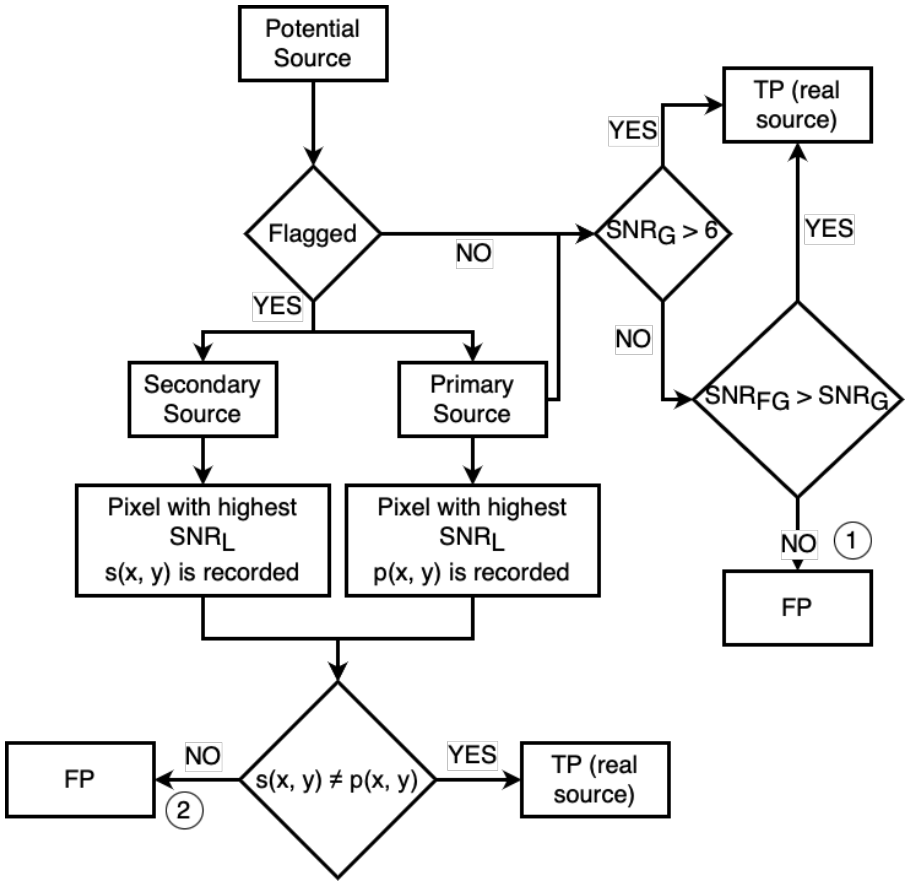

Figure 19: Schema of the False Positives detection and source deblending pipeline which constitutes step 9 in Fig. 14. Numbers 1 and 2 show two possible conditions: a potential source being defined as false positive and discarded from further analysis. The subscript FG (focused global) indicates that the Global SNR is measured on the focused source, while L implies a (local) Pixel SNR measurement. Flagged expresses that multiple peaks are detected within the potential source's spectrum and thus the source is flagged for deblending.

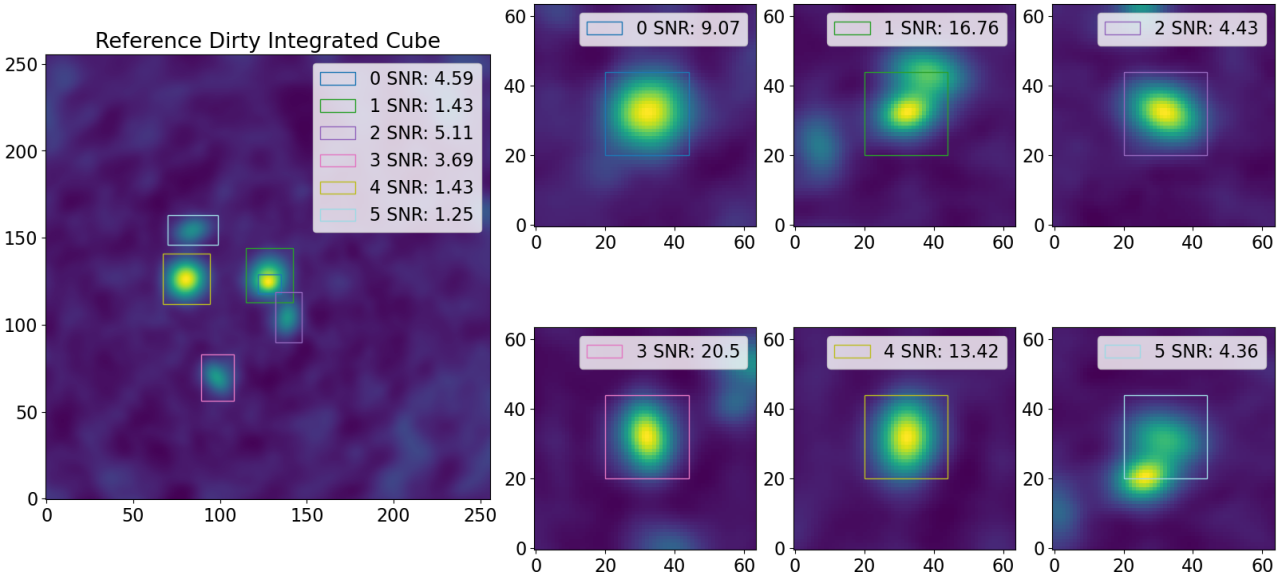

Figure 20: An example of source spectral focusing of sources within a test set image. On the Left, as reference, we plot the dirty integrated cube with the predicted 2D bounding boxes obtained by Blobs Finder highlighted in different colours. The legend matches the source number to the bounding box colour in the image and the measured Global SNR eq. (1). On the right, there are the 6 Spectrally Focused images obtained by integrating over the predicted line extensions found by DeepGRU and cropping a $[64, 64]$ pixel image around Blobs Finder's predicted bounding boxes centres. In each focused image it is also showcased the measured Global SNR. A substantial increase in SNR occurs when sources are focused around their actual emission ranges.

An additional advantage of focusing is the improved dynamic range of the detected signal with respect to the registered one in the reference dirty integrated image.

Generalization to more complex spectral profiles: DeepGRU makes the assumption that emission lines are mostly consisting of a single Gaussian. This simple approximation is valid for a large fraction of ALMA targets, but there are other sources with complex velocity structures such as lopsided gas distributions in galaxies, combinations of Giant Molecular Clouds, etc. In order to generalize the treatment to cover more complex spectral profiles, improved simulations are needed to train the model. In fact, DL models do not generalize well outside the boundaries of their training set. The improvement of the simulation is planned as follow: 1. Employment of multiple ALMA array configurations and observational parameters to address QSO simulations and, e.g. through the collaboration with the SKA Source Finding Focus group, complex morphologies of elliptical and spiral galaxies are generated through physics based modelling. This will be further extended to simulate other objects with complex structures (for instance gas distributions, molecular clouds) and multiple spectral lines but at a later stage; 2. Recover a targeted and comprehensive selection of simulated ALMA observations of interesting objects. Real data and simulated data will be labelled as True and False and a Generative Adversarial Network will be tasked to modify the simulated images until they cannot be discerned anymore from the real data. This will allow us to generate truly realistic moc data on which to re-train our pipeline.

5.3.3 ResNets

Prediction of morphological parameters of the detected sources (FWHM, coordinates x and y, projection angle pa) is performed by ResNets (stage 10 in Fig. 14). Celestial coordinates are computed as photometric baricenters (pixel-weighted centers) of the Blobs Finder predicted bounding box. Source fitting in spatial and frequency domains are combined to create a 3D Gaussian profile.
A 3D segmentation map is created. The segmentation map is dilated by a factor of 1.5 to account for the convolution process spreading the continuum and the line emission signals in the image. This is performed to make sure that all the source signal is contained within the 3D segmentation map. A dilated 3D segmentation mask is used to create the model-masked cube by multiplying it with the dirty cube. The inverse mask is instead used to capture the continuum cube. The continuum image is created by averaging the continuum cube in frequency. The line emission cube is created through the following formula:

$$L_z[x,y] = M_z[x,y] - f(z) * C[x,y] \quad \text{with } z \in \Delta_z \tag{3}$$

where $L_z[x,y]$ is the 2D line emission image at slice z, $M_z[x,y]$ is the model masked 2D image at slice z, $C[x,y]$ is the continuum image and $f(z)$ is the 1D continuum model. The line emission cube is integrated along the frequency to create the line emission image which is fed to a specialized ResNet predicting the source flux density in mJy/beam.

In summary, ResNets allow us to estimate morphological parameters of the detected sources, measure their continuum and the line emission in addition to create the reconstructed cubes. Please note that in case of overlapping/blended sources, the fitting of the morphological parameters is executed simultaneously. Each detected source (primary or discovered seconday source) is fed to the ResNet to regress the source morphological parameters. The characterization is thus performed simultaneously and no information about the two sources, or any previous source seen in inference, is employed to predict the source parameters.

5.4 Train, Test, Validation

The 5000 simulated ALMA Dirty cubes (described in Subsec. 4.2) are grouped in sets and a split ratio of 60-20-20 to train, test and validate is employed.

- The training set is used to train the DL models within the pipeline.
- The test set is used to measure the pipeline performances in detecting sources and in regressing their parameters.
- The validation set is used to measure the training progress and assess generalization capabilities.

5.4.1 Training and Validation

The training allows the model to learn initially a median representation of the data (which should contain information about the dirty beam and the noise patterns) up to the nuances in the data, such as source positions and morphological properties (the shape and sizes of the galaxies). Details on of the training strategies are in Subsec. 4.5 of [3].
At first training occurs in parallel on the whole pipeline starting from pairs of dirty input images (dirty cubes integrated along the frequency) and target sky model images (target sky model cubes integrated along the frequency) with the Blobs Finder model. DeepGRU is trained on pairs of dirty spectra (extracted from the dirty cubes) and clean spectra (extracted from the sky model cubes). The DeepGRU predictions are used in combination with Blobs Finder's predictions to extract the spectrally focused galaxy images. Targets for training are the simulated source parameters. The three ResNets for morphological parameters estimation are trained simultaneously. In the first training iterations, care is taken in DL model to prevent overfitting. Successively, each model is trained independently on the un-augmented training set predictions of the previous model. In this way, each model corrects for biases introduced by the previous one.
Validation is used to fine tune the training stage. Sec. 4 of [3] provides the description of the validation loss implemented in the algorithm.

5.4.2 Testing: Accuracy evaluation of detected sources after training

The test set is used to evaluate whether the algorithm can generalize well to an unseen dataset (i.e. the 1000 ALMA simulated cubes). Blobs Finder and DeepGRU detection capabilities are performed quantifying the overlap of the Ground Truth and the Prediction region, given by the Intersection over Union (IoU). For Blobs Finder, the 2D IoU between the true 2D bounding box and the predicted one is measured. For DeepGRU, the 1D IoU is measured between the true emission ranges and the detected ones. At least 60% of the 3D emission range of a source has to be captured and a true positive (TP) is detected. The 0.6 IoU threshold is chosen to guarantee that 90% of the true emission range is captured within it, given that the line emission image is created through the dilated segmentation mask.

Blobs Finder succeeds with an 89$\%$ efficiency and a 0.1$\%$ contamination. Spectra from the detected dirty cubes are extracted and fed to DeepGRU, that provides a 99$\%$ efficiency and a 0.02$\%$ contamination. Sources are "spectrally focused" within the predicted frequency emission ranges Δ_z, and SNR checks are made. It allows to further investigate false (FP) and true positives. The full logic of the FP removal process is shown in Fig. 19. Blobs Finder's false detection is eliminated by DeepGRU, as described in conditions 1 and 2. Figs. 21 and 22 show some examples of Blobs Finder predictions on the test set. For each block, the upper, middle and bottom rows show the input integrated dirty cubes, the target sky models, and Blobs Finder predictions, respectively.

The clean peaks found by DeepGRU are characterized by the ResNets. Fig. 18 of [3] shows the scatter plots of the true parameters versus the predicted ones and the corresponding residuals histograms. The vast majority of residuals lies within $\pm 1\sigma$, indicating that the process is perfectly under control.

Last, $DeepFocus$ is shown to obtain better results than other machine learning techniques when applied on the same test data set. Please refer to Sec. 5 of [3] for more details.

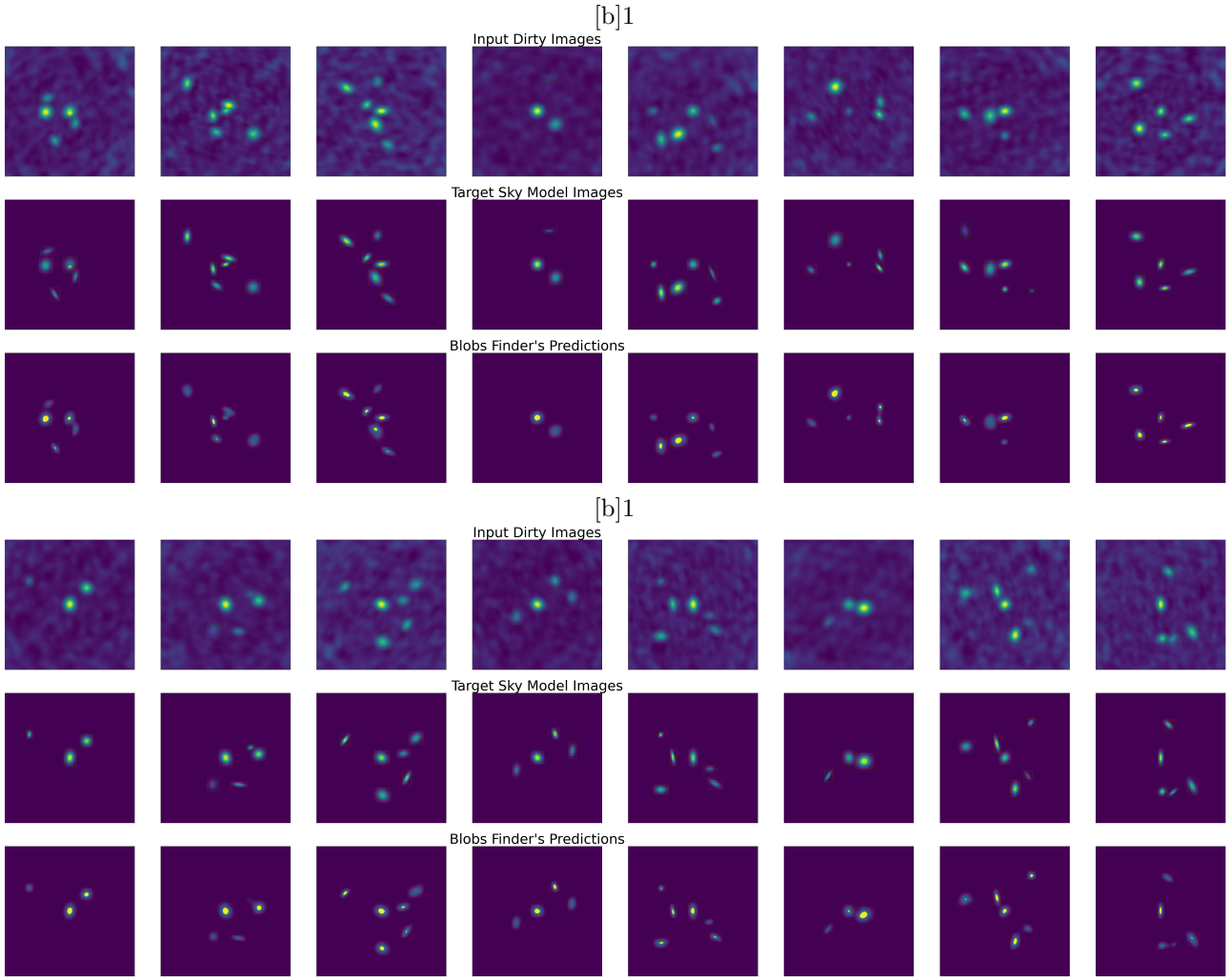

Figure 21: Examples of Blobs Finder predictions on the test Set. The first row shows input integrated dirty cubes, the middle row the target sky models, and the bottom row, Blobs Finder predicted 2D Source Probability maps. In green are outlined (in the dirty and sky models images) the true bounding boxes, while in red the predicted bounding boxes extracted by thresholding the probability maps.

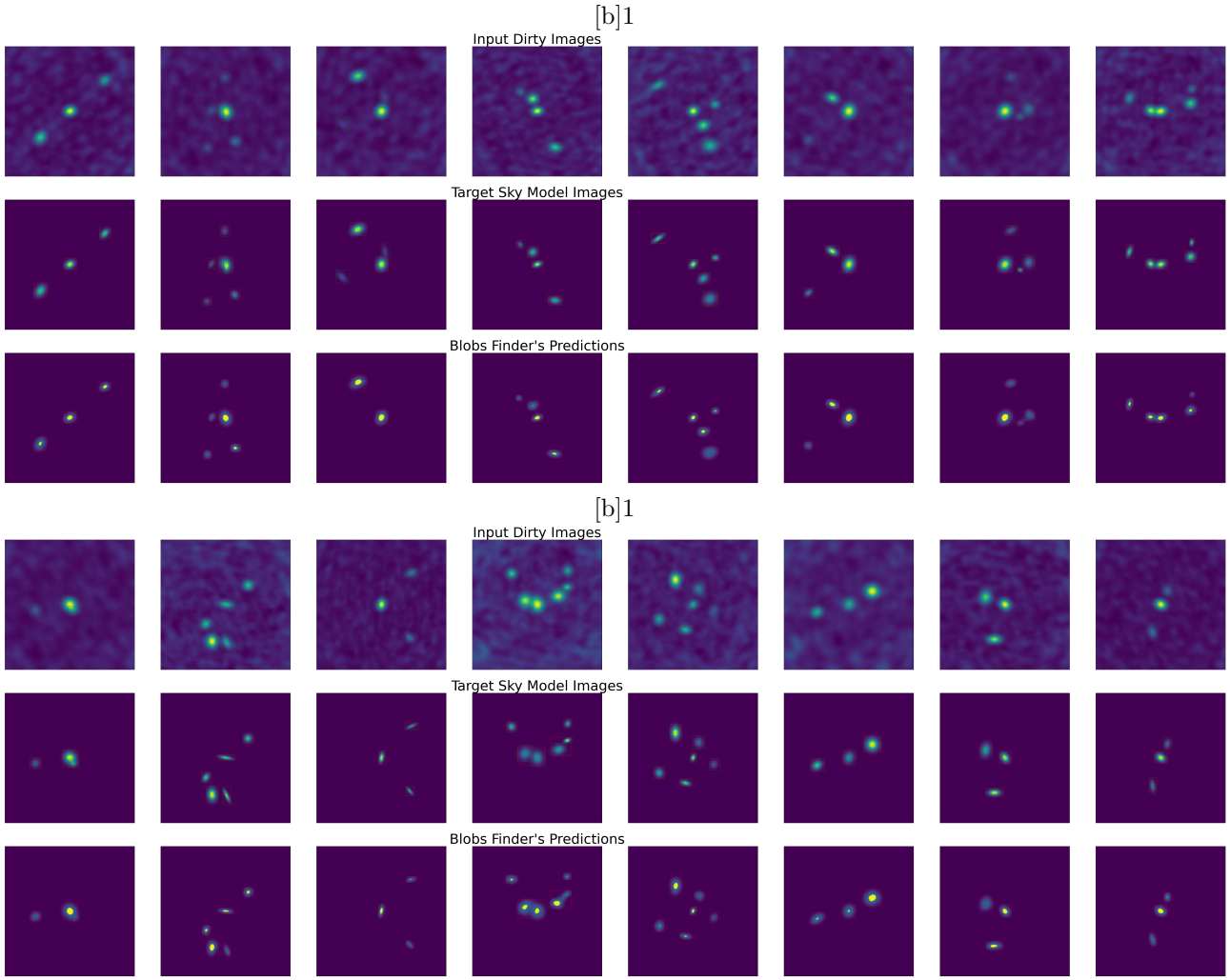

Figure 22: Examples of Blobs Finder predictions on the Test Set. The first row shows input integrated dirty cubes, the middle row the target sky models, and the bottom row, Blobs Finder predicted 2D Source Probability maps. In green are outlined (in the dirty and sky models images) the true bounding boxes, while in red the predicted bounding boxes extracted by thresholding the probability maps.

5.5 Comparison with tCLEAN and speed-up estimation

Preliminary results on comparing the capabilities of DeepFocus with respect to tCLEAN on a sample of **1000 simulated ALMA cubes** (256x256x128) are provided. In Fig. 23 part of the sample is shown with the first and second columns representing a sample of the integrated ALMA cubes and sky models. The third column shows the solution provided by DeepFocus. tCLEAN solution run with 200 cleaning iterations is given in the last column. Both algorithms are capable to find all sources in this sample. Residuals are measured to compare the true sky model reconstruction performance of the two algorithms: Blobs Finder's residuals are within $\pm 1\sigma$, while for tCLEAN the residuals are deviating above the $\pm 5\sigma$ threshold.

The measured residuals will certainly decrease while increasing the number of cleaning iterations within tCLEAN, but to the detriment of the computational cost. Employing the sample of 1000 ALMA cubes, we can compare the computational cost of the two algorithms when using 200 cleaning iterations for tCLEAN:

- **Blobs Finder made its predictions on the entire Test set in 23 seconds employing a single NVIDIA Tesla K20.**
- **tCLEAN took 4.3 minutes per cube utilising 8 Intel Xeon E5-2680 CPUs. Given the 400 CPUs at our disposal, we run it on 50 cubes at a time in parallel obtaining a total computational time of 1.5 hours**.

Employing BlobsFinder for the reconstruction task on the entire Test set results in a speed-up factor of 200 on our system with respect to tCLEAN. If we consider the possibility to accommodate DeepFocus within tCLEAN, if we account for the major cycle procedure execution time, this workout will provide an improvement in **speeding up the CASA procedures of at least a factor of 150**.

We are aware that the promising speed up in the procedures when employing DeepFocus within tCLEAN will not solve unfortunately the performance problem of the Wideband Sensitivity Upgrade [23] alone. However, it will make a huge contribution. If this is not enough for ALMA2030 era, we may explore the generalization of the algorithm to perform a whole deconvolution process independently to tCLEAN.

Note on the speed calculation and sustainability: We are planning to measure execution speed of DeepFocus and tCLEAN in a setting suited to ALMA operations. Estimates of costs will also be accounted and compared.
So far, the provided comparison in speed is made with prediction time only. The total pipeline training time lasted around 5 hours, while $tCLEAN$ 1.5 hours (as $tCLEAN$ does not need training time). Nevertheless this is a one time investment when DeepFocus is used within the CASA framework. DeepFocus is shareable as a python script plus a series of weight files (like pre-trained networks are shared by Google, Pytorch and so on), and the community would thus benefit of an active learning paradigma (the model is periodically trained on new and improved data) in order to provide the best possible performances.
DeepFocus algorithm is run on a graphics rather than central processing units, with a potential lower cost than several CPUs on a long run operational perspective.

5.6 Ongoing developments and outlook

So far DeepFocus is applied to simulated data sets. An application of DeepFocus to real data is ongoing. An archived ALMA QSO sample is under development to be fed to the algorithm in the search of serendipitous galaxies. This work will stand on previous work of a well-known sample. Although the goal is to detect faint serendipitous objects, the robustness of the algorithm is going to be strengthened. In fact, while Deep GRU and Blobs Finder may produce false detections, previous work on other Radio Interferometric data (such as SKA Data Challenge 2, Westerbork Synthesis Radio Telescope data) seems to suggest that the problem of false detection can be reliably solved by running classifiers after the detection process. Those classifiers are capable to distinguish true from false detections. ResNets Classificators and Random Forests have been

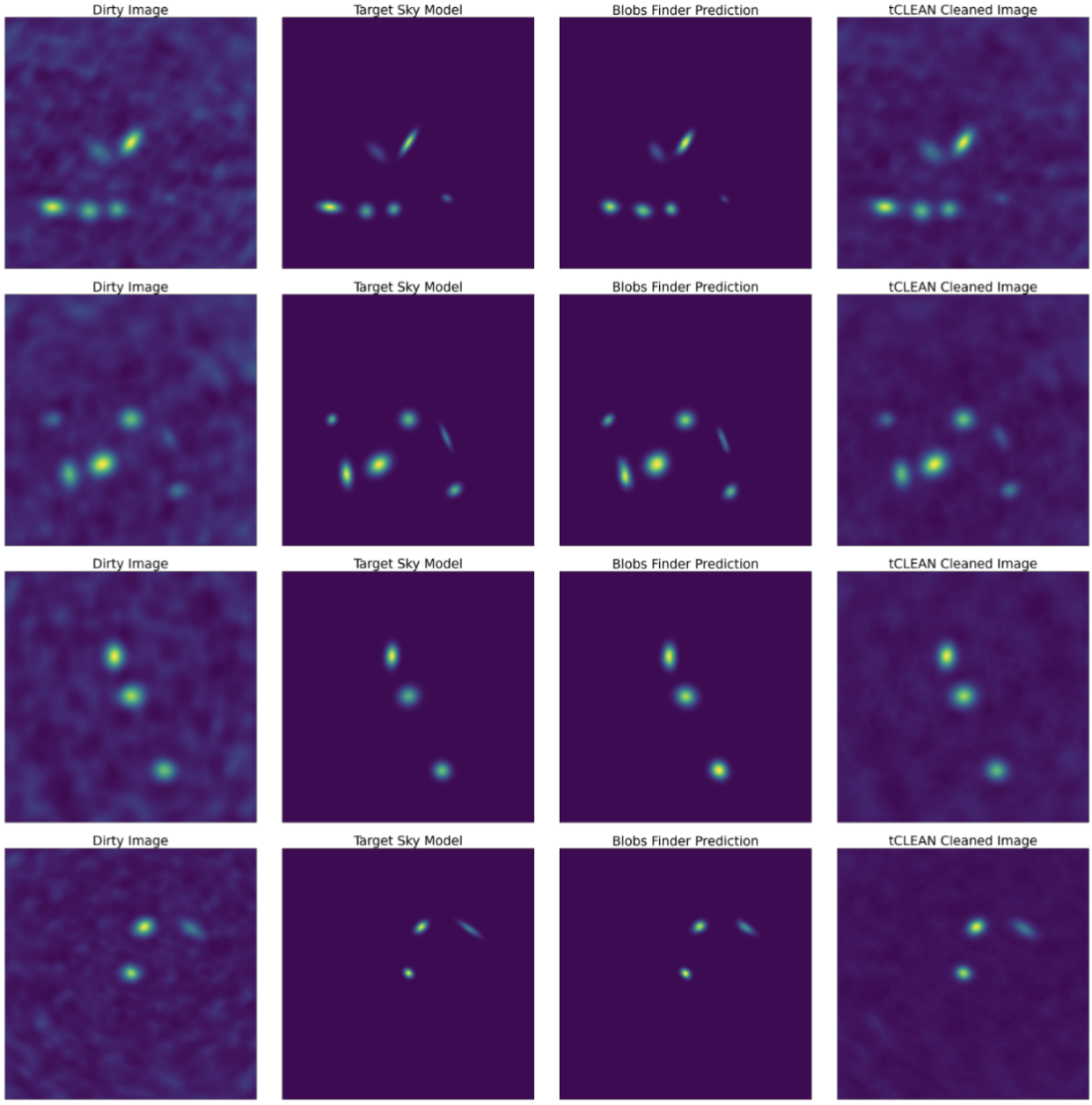

Figure 23: *Comparison of $DeepFocus$ and $tCLEAN$ on a set of 1000 simulated ALMA cubes. In the images, the first, second, third and fourth columns shows a sample of Dirty Images, target Sky Models, Blobs Finder's reconstruction and tCLEAN reconstructions with niter = 200.*

already successfully tested for this task.
Simulations for the creation of the ALMA cubes have been discussed for further improvements. On the short term, tests shall be performed to cross-check if sources that are off axis from one channel to another are well detected by the algorithm: introducing extreme cases with respect to what already implemented. Larger data cubes shall be created and tested to meet the challenges introduced by the approved upgrades [23]. The technique shall also be tested and trained to detect continuum only in continuum imaging and noise only in cubes and continuum imaging.
On the long term, we want to refine the simulations employing physical properties to include more complex source morphologies, use physically-based models for the galaxy kinematics, and employ spectral catalogues to generate several spectral profiles for different class of sources with the primary goal to improve the quality of our simulations, and the additional goal of having a publicly available, and easy to use simulation code that the community may use to generate common data sets on which compare different architectures. The pipeline is foreseen to be modified to account for such more complex spectral profiles.
Refinements on the simulation code and detection algorithm are also needed to detect other complex and reach environments as when observing the galactic center or the SZ effect for absorption lines.
Because of the several and complex celestial radio sources detectable with ALMA, the algorithm shall be expanded to allow for transfer learning. For instance, domain adaptation will allow the algorithm's ability to learn from both artificial and real data. The basis of this development are important in operations in view of ALMA2030 [23].
Last, we want to start the endevour to support CASA in an attempt to develop an hybrid approach, as suggested by Ryan Loomis (NRAO), to have both full tCLEAN and ML capabilities implemented. This approach has the big advantage to allow DeepFocus to learn during operation and refine the ML algorithm while still relying on an established and understood technique as tCLEAN.

6 Final remarks and future perspectives

We demonstrated the capabilities of two machine learning techniques applicable to ALMA data and capable to overcome current issues in synthesis imaging. RESOLVE is capable to detect diffuse emissions as well as point sources. DeepFocus has exceptional capabilities in speeding up procedures. Both techniques need further testing. The team is motivated to move forward with further developments. A number of young and enthusiastic researchers would like to join our enterprise. We appreciate your feedback and support for continuing this study.

7 Appendix

- Bayesian and Machine Learning Methods in the Big Data era for astronomical imaging (Guglielmetti, F. et al.)
- Bayesian statistics approach to imaging of aperture synthesis data: RESOLVE meets ALMA (Tychoniec, L. et al.)
- 3D Detection and Characterisation of ALMA Sources through Deep Learning (Delli Veneri, M. et al.)

References

[1] J. A. Hogbom. Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines. *Astron.Astrph.Suppl.*, 15:417, June 1974.

[2] CASA Team, Ben Bean, Sanjay Bhatnagar, Sandra Castro, Jennifer Donovan Meyer, Bjorn Emonts, Enrique Garcia, Robert Garwood, Kumar Golap, Justo Gonzalez Villalba, Pamela Harris, Yohei Hayashi,

Josh Hoskins, Mingyu Hsieh, Preshanth Jagannathan, Wataru Kawasaki, Aard Keimpema, Mark Kettenis, Jorge Lopez, Joshua Marvil, Joseph Masters, Andrew McNichols, David Mehringer, Renaud Miel, George Moellenbrock, Federico Montesino, Takeshi Nakazato, Juergen Ott, Dirk Petry, Martin Pokorny, Ryan Raba, Urvashi Rau, Darrell Schiebel, Neal Schweighart, Srikrishna Sekhar, Kazuhiko Shimada, Des Small, Jan-Willem Steeb, Kanako Sugimoto, Ville Suoranta, Takahiro Tsutsumi, Ilse M. van Bemmel, Marjolein Verkouter, Akeem Wells, Wei Xiong, Arpad Szomoru, Morgan Griffith, Brian Glendenning, and Jeff Kern. CASA, the Common Astronomy Software Applications for Radio Astronomy. , 134(1041):114501, November 2022.

[3] Michele Delli Veneri, Łukasz Tychoniec, Fabrizia Guglielmetti, Giuseppe Longo, and Eric Villard. 3D detection and characterization of ALMA sources through deep learning. , 518(3):3407–3427, January 2023.

[4] Jane Huang, Sean M. Andrews, Cornelis P. Dullemond, Andrea Isella, Laura M. Pérez, Viviana V. Guzmán, Karin I. Öberg, Zhaohuan Zhu, Shangjia Zhang, Xue-Ning Bai, Myriam Benisty, Tilman Birnstiel, John M. Carpenter, A. Meredith Hughes, Luca Ricci, Erik Weaver, and David J. Wilner. The disk substructures at high angular resolution project (dsharp). ii. characteristics of annular substructures. *The Astrophysical Journal Letters*, 869(2):L42, dec 2018.

[5] H. Junklewitz, M. R. Bell, M. Selig, and T. A. Enßlin. RESOLVE: A new algorithm for aperture synthesis imaging of extended emission in radio astronomy. *Astronomy & Astrophysics*, 586, Feb. 2016.

[6] Fabrizia Guglielmetti, Philipp Arras, Michele Delli Veneri, Torsten Enßlin, Giuseppe Longo, Łukasz Tychoniec, and Eric Villard. Bayesian and Machine Learning Methods in the Big Data era for astronomical imaging. *arXiv e-prints*, page arXiv:2210.01444, October 2022.

[7] L. Tychoniec. Bayesian statistics approach to imaging of aperture synthesis data: Resolve meets alma. In *International Conference on Bayesian and Maximum Entropy methods in Science and Engineering*, number 67, 18 - 22 July 2022.

[8] Philipp Arras, Hertzog L. Bester, Richard A. Perley, Reimar Leike, Oleg Smirnov, Rüdiger Westermann, and Torsten A. Enßlin. Comparison of classical and bayesian imaging in radio interferometry. *Astronomy & Astrophysics*, 646:A84, Feb. 2021.

[9] Philipp Arras, Philipp Frank, Reimar Leike, Rüdiger Westermann, and Torsten A. Enßlin. Unified radio interferometric calibration and imaging with joint uncertainty quantification. *A&A*, 627:A134, Jul 2019.

[10] F. Guglielmetti, R. Fischer, and V. Dose. Background-source separation in astronomical images with bayesian probability theory - i. the method. *MNRAS*, 396:165–190, 2009.

[11] K-E Harabi, T. Hirtzlin, T. Clément, E. Vianello, R. Laurent, J. Droulez, P. Bessière, J-M Portal, M. Bocquet, and D. Querlioz. A memristor-based Bayesian machine. *Nature Electronics*, 6:52–63, December 2022.

[12] F. Guglielmetti, E. Villard, and Ed Fomalont. Bayesian reconstruction through adaptive image notion. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering—MaxEnt 2019*, volume 33, 8, pages 95–101, 30 June - 5 July 2019.

[13] T. J. Cornwell and K. F. Evans. A simple maximum entropy deconvolution algorithm. *A&A*, 143:77–83, February 1985.

[14] Torsten A. Enßlin, Mona Frommert, and Francisco S. Kitaura. Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis. *Physical Review D*, 80(10):105005, Nov. 2009.

[15] Torsten Enßlin. Information field theory. In Udo von Toussaint, editor, *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 32nd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume 1553 of *American Institute of Physics Conference Series*, pages 184–191, Aug. 2013.

[16] Maksim Greiner, Valentina Vacca, Henrik Junklewitz, and Torsten A. Enßlin. fastRESOLVE: fast Bayesian imaging for aperture synthesis in radio astronomy. *arXiv e-prints*, page arXiv:1605.04317, May 2016.

[17] Philipp Arras, Jakob Knollmüller, Henrik Junklewitz, and Torsten A. Enßlin. Radio Imaging With Information Field Theory. *arXiv e-prints*, page arXiv:1803.02174, March 2018.

[18] C. L. Brogan and et al. THE 2014 ALMA LONG BASELINE CAMPAIGN: FIRST RESULTS FROM HIGH ANGULAR RESOLUTION OBSERVATIONS TOWARD THE HL TAU REGION. *The Astrophysical Journal*, 808(1):L3, July 2015.

[19] Łukasz Tychoniec, Fabrizia Guglielmetti, Philipp Arras, Torsten Enßlin, and Eric Villard. Bayesian statistics approach to imaging of aperture synthesis data: RESOLVE meets ALMA. *arXiv e-prints*, page arXiv:2210.02408, October 2022.

[20] CASA documentation. `https://casa.nrao.edu/casadocs-devel`. Accessed: 2019-02-07.

[21] D. S. Briggs. High Fidelity Interferometric Imaging: Robust Weighting and NNLS Deconvolution. In *American Astronomical Society Meeting Abstracts*, volume 187, page 112.02, Dec 1995.

[22] Theo Steininger, Maksim Greiner, Frederik Beaujean, and Torsten A. Ensslin. D2O: Distributed Data Object. Astrophysics Source Code Library, record ascl:2203.010, March 2022.

[23] John Carpenter, Daisuke Iono, Francisca Kemper, and Al Wootten. The alma development program: Roadmap to 2030. 2020.

[24] S. Carniani and et al. Strongly star-forming rotating disks in a complex merging system at z = 4.7 as revealed by ALMA. *Astronomy & Astrophysics*, 559:A29, Nov. 2013.

[25] J. P. McMullin, B. Waters, D. Schiebel, W. Young, and K. Golap. CASA Architecture and Applications. In R. A. Shaw, F. Hill, and D. J. Bell, editors, *Astronomical Data Analysis Software and Systems XVI*, volume 376 of *Astronomical Society of the Pacific Conference Series*, page 127, Oct. 2007.