ALMA source extractions tools

1. Tools of source extraction

1.1 Sextractor

refs:

Survey	Area ($'^2$)	Depth (μJy)	N	before/ after PB	Refs
ALMACAL (1.2mm)	149	—	228	before	Oteo et al. 2016
ASPECS (1.2mm/3mm)	4.2	9.3	35	before	Araveba et al. 2016
ALMACAL (650μm)	5.5	100	21	before & reason	Klitsch et al. 2020
Reason: We perform the source detection using SEXTRACTOR on the clean maps before correcting for the primary beam attenuation, to ensure uniform noise properties.					
Archive Data (1.2mm)	120p	20	133	before	Fujimoto et al. 2016
AS2UDS (870μm)	716p	300	708	- -	Stach et al. 2018; Stach et al. 2019
AS2COSMOS (870μm)	182p	190	260	before	Simpson et al. 2020

pros:

- Speed. Sextractor is made to go through data quickly.
- Does decent photometry.
- Robust, it'll run with idiotic input.
- Controllable, most steps can be influenced by user.
- The possibility to accept user specified flag images or weight images.
- The possibility to detect sources in one image and do the photometry in another.
- There is follow-up software (GIM2D or GALFIT) to decomposition of galaxy profiles.

cons:

- Only as good as its settings. SE is dependent on some of it's setting and these are crucial for the detection and photometry. It will run on just about any set of input parameters but give back output that may be total bogus.
- Limited accuracy.
- Classification of objects is of very limited use.
- Breaks down in crowded fields eventually.
- Corrections of photometry for the 'wings' of object profiles is very rudimentary.

1.2 SoFiA

refs:

Survey	Area ($'^2$)	Depth (μJy)	N	before/ after PB	Refs
COALAS	21	—	46	dirty image	Jin et al. 2021
Tips: Most of the rest papers are conducted with radio survey thus not shown here.					

pros:

- SoFiA2 is wrote in C language, thus it runs with high speed and uses less memory.
- It has 100% reliability and 90% completeness at an integrated SNR>5.
- The accuracy of source parameters is high, such as spatial/spectral centroids, intergrated fluxes.

cons:

- Special settings are required for SoFiA to work on 2D images, such as frequency-related control parameters.
- Sources with low SNR might be mis-resolved and partially detected.
- Peak flux density is more strongly affected by the noise at low SNR.
- Full-scale bias assessment is required if one need higher parameterization accuracy.

1.3 PyBDSF

refs:

Survey	Area ($'^2$)	Depth (μJy)	N	before/ after PB	Refs
GOODS-ALMA 2.0 (1.1mm)	72	68.4	88	before	Gómez-Guijarro et al. 2022
REBELES SURVEY	—	—	12	before	Inami et al. 2022
Tips: Most of the rest papers are conducted with radio survey thus not shown here.					

pros:

- Specially designed for the interferometric instrument.
- Capability of decomposing an image into a set of Gaussians, shapeless or wavelets.
- Measurement of the PSF variation across an image.
- Capture of components at various scales.
- Performing of other filtering operations post-extraction for science purposes.

cons:

- Not specialized on 3D cubes. The sources are identified in the collapsed 2D image if a 3D cube is given.
- The typical size of the artifacts around bright sources (e.g., rms_box) are needed to take care to specify.

1.4 AEGEAN

refs:

Survey	Area ($'^2$)	Depth (μJy)	N	before/ after PB	Refs
SXDF	2	55	23	before	Hatsukade et al. 2016
ASAGAO	26	34	45	before	Hatsukade et al. 2018
ADF22	20	73	35	before	Umehata et al. 2018
Tips: Most of the rest papers are conducted with radio survey thus not shown here.					

pros:

- Creating background and noise images in a lower false detection rate, especially in the case where the background or noise properties are changing quickly within an image.

- The prioritised fitting with a choice of the number of degrees of freedom, and includes a regrouping algorithm that ensures that overlapping sources and components are fit jointly.
- Capabilities of describing the variation of PSF in the image with wide field of view.

cons:

- Possibility of wrongly describing the noise properties of the image, due to the sensitivity limitation from a combination of side-lobe and classical confusion.

1.5 BLOBCAT

refs:

Survey	Area ($'^2$)	Depth (μJy)	N	before/ after PB	Refs
GOODS-ALMA 1.0	69	180	20	- -	Franco et al. 2018
Tips: Most of the rest papers are conducted with radio survey thus not shown here.					

pros:

- Unresolved and resolved Gaussian blobs are automatically and accurately processed by BLOBCAT
- Precedure for estimating the background RMS noise are robust and accurate, also suitable for other tools.

cons:

- Non-Gaussian blobs need to be addressed manually.
- Modification of BLOBCAT's algorithms may be required to account for wavelength- and instrument-specific descriptions of point spread functions and pixellation errors.
- Not suitable for optical data.

1.6 Peak-finding

refs:

Survey	Area ($'^2$)	Depth (μJy)	N	before/after PB	Refs
HUDF	4.5	35	16	- -	Dunlop et al. 2017
MORA (2mm)	184	—	13	after	Zavala et al. 2021

Survey	Area ($'^2$)	Depth (μJy)	N	before/after PB	Refs
3mm Archive	130p	>10	16	before & reason	Zavala et al. 2018
Tips: This method is not a specific software, wrote by authors themselves.					
Reason: Source extraction was performed using the uncorrected primary beam continuum maps (which have the benefit of a constant noise)					

pros:

- It is simple and easily to write by someone self.
- Work on single point data well and fast.

cons:

- Too simple to detect sources in images with large field of view.

1.7 LineSeeker

refs:

Survey	Area ($'^2$)	Depth (μJy)	N	before/after PB	Refs
ASPECS	4.2	9.3	35	before	González-López et al. 2019/2020
Tips: No more documents are found thus no pros and cons here.					

1.8 Maxfinder

refs:

Survey	Area ($'^2$)	Depth (μJy)	N	Refs
COALAS	21	—	46	Jin et al. 2021
Tips: No more documents are found thus no pros and cons here.				

1.9 Comments

- The software designed for 2D images would be better at current stage, not those specifically designed for searching spectral lines in 3D cubes.
- Therefore the tools of Sextractor, PyBDSF, AEGEAN and BLOBCAT are recommended, not those of SoFiA, Maxfinder, LineSeeker.
- The software could be considered If we indeed need to search for the potential spectral lines in the 3D cubes.

1.10 Primary beam correction

About the reason of using the un PB corrected image to do source extraction:

- I went through the alma archive school docs, there is no information about the source extraction and PB correction. I check all the reference you sent me and I found, almost all the papers used the un PB corrected images to do source extraction. There are two papers stated why they use the un PB corrected iamges and I copied here
- "We perform the source detection using SEXTRACTOR on the clean maps before correcting for the primary beam attenuation, to ensure uniform noise properties." --Klitsch+2020
- "Source extraction was performed using the uncorrected primary beam continuum maps (which have the benefit of a constant noise)" -- Zavala+2018

I also discussed with Jianhang last week. We come out the following thought:

- Since the source extraction procedure need an RMS map, one have to produce an RMS map using the flux map (the latter is only product of the interferometric data). After PB correction, the flux will increase from center to outer regions, especially in marginal areas. So the RMS map will heavily affect by the primary beam response and the method of RMS estimation. While in the un PB corrected images, the flux in the whole regions are flat. The RMS map is easily calculated and uniform (or even constant). So one can extract source in the un PB corrected images just like in optical bands, getting more reliable result (especially in marginal areas) without the influence of primary beam response and the RMS estimating methods.

- Note that the source extraction is only used to determine the positions of the reliable detections, the “wrong” flux value will not impact the further flux estimate. Meanwhile, the SNR of the detections remain unchanged before and after the PB correction, since the Flux and Rms are both increase after the PB correction.

Considerations for our situation:

- There are several parameters to affect the estimation of the RMS maps, I didn't find the “reasonable criteria” to determine them by now. I think we could have deeper understanding after check and compare the result from both the before and after PB corrected images.

2. Background RMS estimate

2.1 Stardard deviation (std) of a window

- Calculating the std of the surrounding 100*100 pixels; Umehata et al. 2018
- Calculating the std in the window of size 10 * √(1.331*a*b); Dunlop et al. 2017

2.2 Build in the softwar

- Autocalculated by PyBDSF; Gómez-Guijarro et al. 2022

3. Flux estimate

3.1 Peak flux (useful for unresolved sources)

- Gómez-Guijarro et al. 2022
- Aravena et al. 2016

3.2 Aperture photometry

- Elliptical apertures same as synthesis beam; Chen et al. 2022
- Apertures with 1.6 arcsec diameter; Gómez-Guijarro et al. 2022

3.3 2D Gaussian fiiting

- 2D functional fitting with imfit in CASA; Gómez-Guijarro et al. 2022
- 2D functional fitting with imfit in CASA; Fujimoto et al. 2016
- 2D Gaussian fitting with astropy.modeling; Chen et al. 2022
- 2D Gaussian fitting with Sextractor; Aravena et al. 2016

3.4 Other methods

- Galfit; Peng et al. 2010
- PyBDSM; Mohan&Rafferty et al. 2015
- BLOBCAT; Hales et al. 2012