PolyCLEAN

Convex Optimization for Radio Interferometry with a CLEAN-like Polyatomic Algorithm

Adrian Jarret

2023.04.19

slides: adriaj.github.io

01
Background
Specificities of radio interferometry

02
State of the Art
Current imaging methods

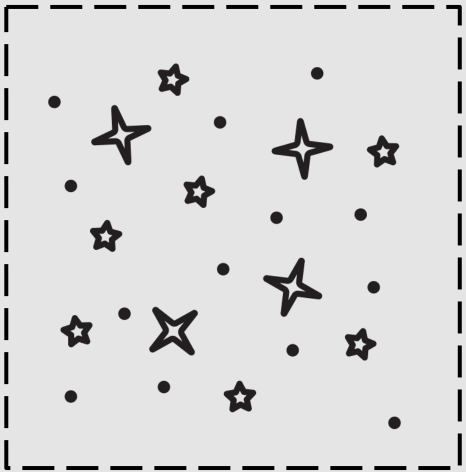
03
PolyCLEAN
Our algorithm

04
Experiments
First results in simulations

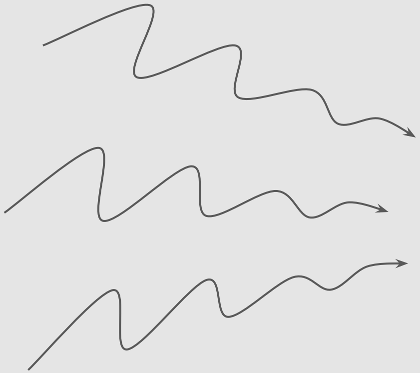
01. Background

Radio Interferometric Imaging

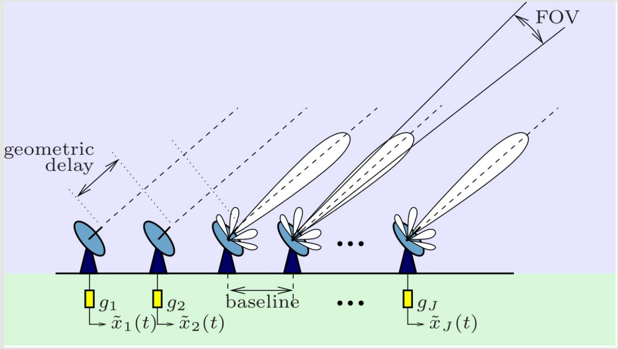
Radio Interferometry



Sky Image



EM Waves



Radio Interferometer

[Van der Veen et al.,2019]

Linear Inverse Problem

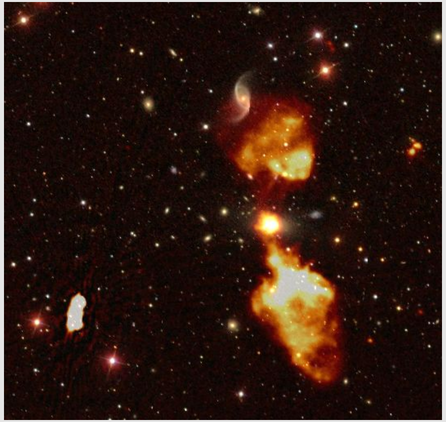
$$\boxed{\mathbf{V} = \mathbf{\Phi I}}$$

[Credits: Cyril Tasse and the LOFAR surveys team.]

Linear Inverse Problem

$$\boxed{\mathbf{V} = \mathbf{\Phi}\mathbf{I}}$$

Observed area of the sky



[Credits: Cyril Tasse and the LOFAR surveys team.]

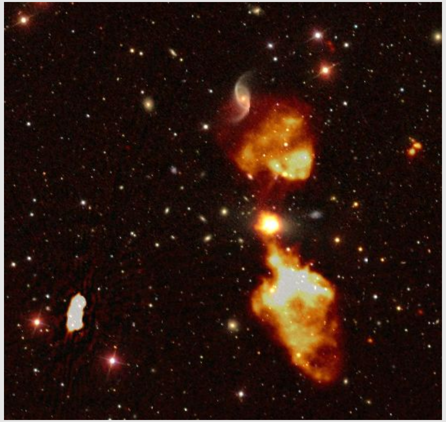
Linear Inverse Problem

Visibility measurements

Spatial frequency information, *Fourier-like* measurements, complex valued

$$\boxed{\mathbf{V} = \mathbf{\Phi I}}$$

Observed area of the sky

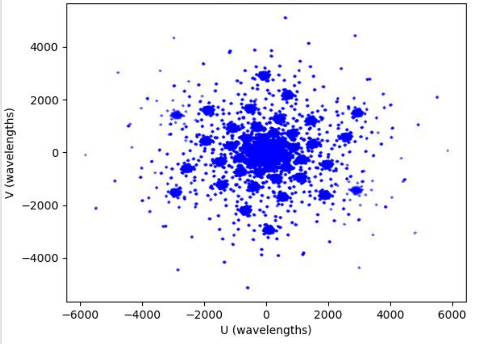


Linear Inverse Problem

$$\boxed{\mathbf{V} = \mathbf{\Phi I}}$$

Visibility measurements

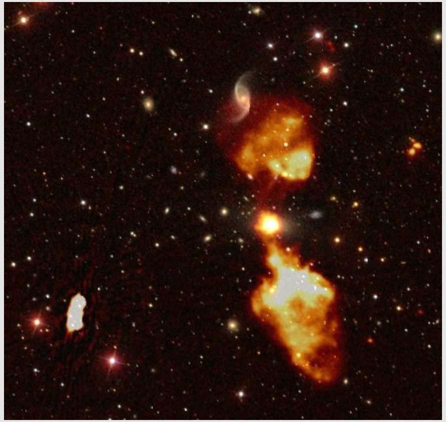
Spatial frequency information,
Fourier-like measurements,
complex valued



Interferometry operator

Depends on the location of the
antennas (*baselines)*
And the observed wavelength

Observed area of the sky



$$V_k = (\mathbf{\Phi I})_k = \sum_{i,j} \frac{e^{-j2\pi(u_k \ell_i + v_k m_j)}}{\sqrt{1-\ell_i^2 - m_j^2}} \mathcal{W}(\ell_i, m_j; w_k) I_{i,j}$$

Linear IP

$$\boxed{\mathbf{V} = \mathbf{\Phi}\mathbf{I}}$$

Challenges of RI

- Noisy measurements

$$\mathbf{V} = \mathbf{\Phi}\mathbf{I} + \varepsilon$$

Linear IP

$$\boxed{\mathbf{V} = \mathbf{\Phi}\mathbf{I}}$$

Challenges of RI

- Noisy measurements

$$\mathbf{V} = \mathbf{\Phi}\mathbf{I} + \varepsilon$$

- Ill-posed problem

$$\mathrm{Null}(\mathbf{\Phi}) \neq \{0\}$$

Challenges of RI

Linear IP

$$\boxed{\mathbf{V} = \mathbf{\Phi}\mathbf{I}}$$

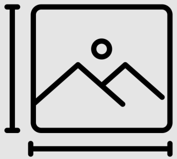
- Noisy measurements

$$\mathbf{V} = \mathbf{\Phi}\mathbf{I} + \varepsilon$$

- Ill-posed problem

$$\mathrm{Null}(\mathbf{\Phi}) \neq \{0\}$$

- Huge volumes of data



Challenges of RI

Linear IP

$$\boxed{\mathbf{V} = \mathbf{\Phi}\mathbf{I}}$$

- Noisy measurements

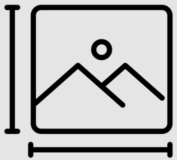
$$\mathbf{V} = \mathbf{\Phi}\mathbf{I} + \varepsilon$$

- Ill-posed problem

$$\mathrm{Null}(\mathbf{\Phi}) \neq \{0\}$$

Use of priors for reconstruction!

- Huge volumes of data



02. Current Methods

The landscape of prior-based imaging techniques in RA

The CLEAN-based Methods

The Convex Optimisation Methods

The CLEAN-based Methods

Implicit sparse prior,
Parametric shape of the solutions

$$\mathbf{I}^*[n] = \sum_k \alpha_k g(n - n_k)$$

The Convex Optimisation Methods

The CLEAN-based Methods

Implicit sparse prior,
Parametric shape of the solutions

$$\mathbf{I}^*[n] = \sum_k \alpha_k g(n - n_k)$$

The Convex Optimisation Methods

Penalty-based priors,
Bayes interpretation (MAP),
Representer theorems

$$\arg\min \frac{1}{2}\|\mathbf{V} - \mathbf{\Phi}\mathbf{I}\|_2^2 + \lambda \mathcal{R}(\mathbf{I})$$

The CLEAN-based Methods

Implicit sparse prior,
Parametric shape of the solutions

$$\mathbf{I}^*[n] = \sum_k \alpha_k g(n - n_k)$$

The Convex Optimisation Methods

Penalty-based priors,
Bayes interpretation (MAP),
Representer theorems

$$\arg\min \frac{1}{2}\|\mathbf{V} - \mathbf{\Phi}\mathbf{I}\|_2^2 + \lambda\mathcal{R}(\mathbf{I})$$

$$p(\mathbf{I}) \propto e^{-\mathcal{R}(\mathbf{I})}$$

The CLEAN-based Methods

Implicit sparse prior,
Parametric shape of the solutions

$$\mathbf{I}^*[n] = \sum_k \alpha_k \boxed{\delta}(n - n_k)$$

The Convex Optimisation Methods

Penalty-based priors,
Bayes interpretation (MAP),
Representer theorems

$$\arg\min \frac{1}{2}\|\mathbf{V} - \mathbf{\Phi}\mathbf{I}\|_2^2 + \lambda\boxed{\|\mathbf{I}\|_1}$$

CLEAN

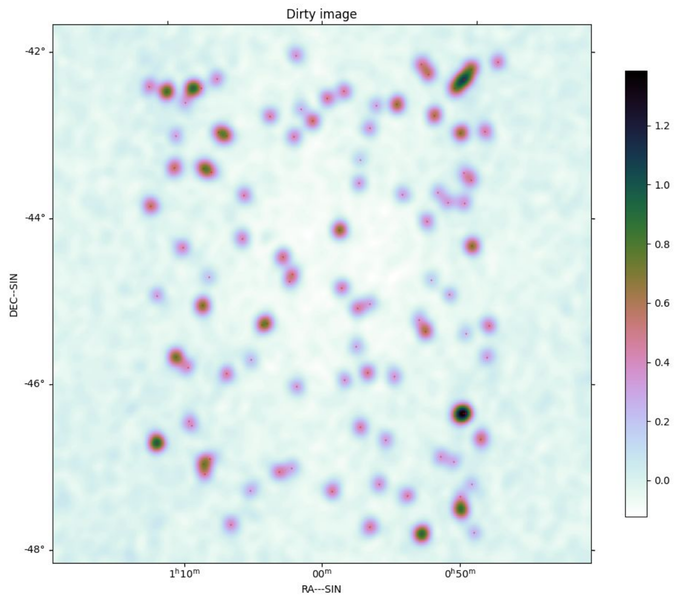
Algorithm 1 Högbom CLEAN Algorithm (Major cycles only)

Initialisation : $\mathbf{I}^{(0)} = \mathbf{0}$, $\mathbf{I}_D = \mathbf{\Phi}^* \mathbf{V}$, $\alpha > 0$

for $k = 1, 2, \cdots, k_{\max}$ **do**

1. Compute the dirty residual: $\mathbf{I}_R^{(k)} = \mathbf{I}_D - \mathbf{\Phi}^* \mathbf{\Phi} \mathbf{I}^{(k-1)}$
2. Find the next reconstructed source: $s^{(k)} = \arg\max \ \mathbf{I}_R^{(k)}$
3. Update the iterate: $\mathbf{I}^{(k)} = \mathbf{I}^{(k-1)} + \alpha \delta_{s^{(k)}}$

end for



CLEAN

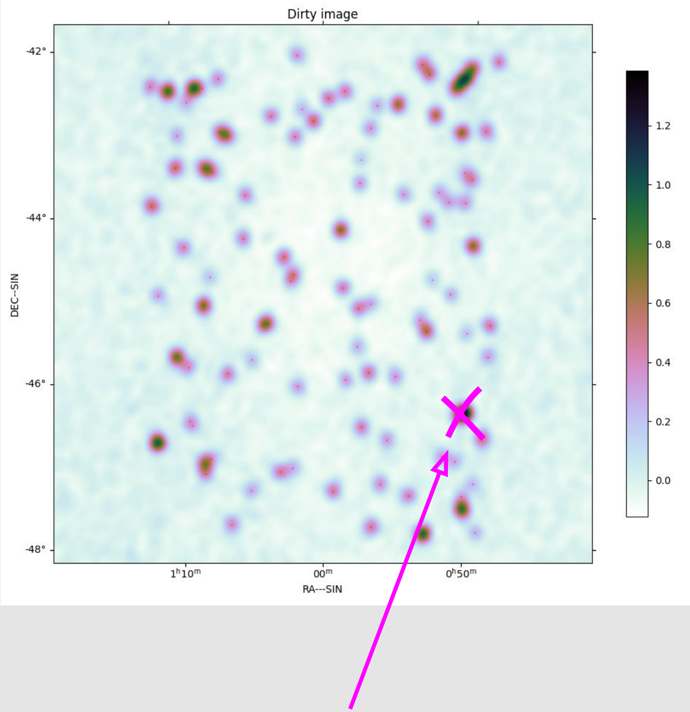
Algorithm 1 Högbom CLEAN Algorithm (Major cycles only)

Initialisation : $\mathbf{I}^{(0)} = \mathbf{0}$, $\mathbf{I}_D = \mathbf{\Phi}^* \mathbf{V}$, $\alpha > 0$

for $k = 1, 2, \cdots, k_{\max}$ **do**

1. Compute the dirty residual: $\mathbf{I}_R^{(k)} = \mathbf{I}_D - \mathbf{\Phi}^* \mathbf{\Phi} \mathbf{I}^{(k-1)}$
2. Find the next reconstructed source: $s^{(k)} = \arg\max \ \mathbf{I}_R^{(k)}$
3. Update the iterate: $\mathbf{I}^{(k)} = \mathbf{I}^{(k-1)} + \alpha \delta_{s^{(k)}}$

end for



CLEAN

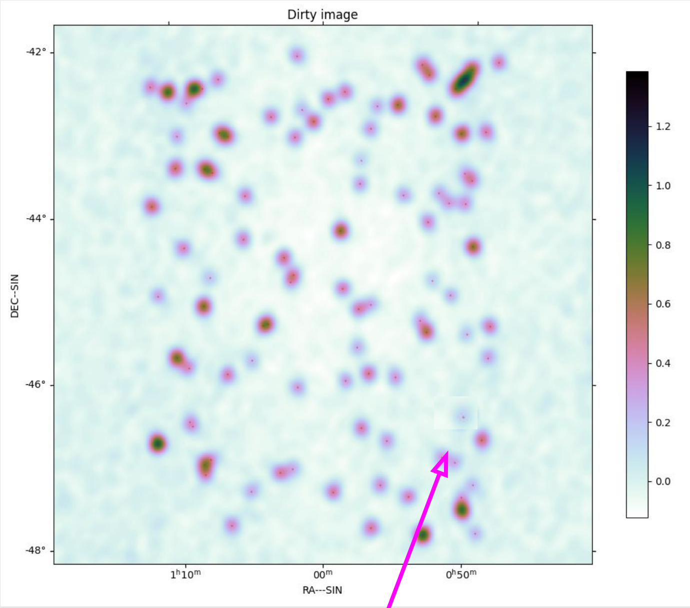
Algorithm 1 Högbom CLEAN Algorithm (Major cycles only)

Initialisation : $\mathbf{I}^{(0)} = \mathbf{0}$, $\mathbf{I}_D = \mathbf{\Phi}^*\mathbf{V}$, $\alpha > 0$

for $k = 1, 2, \cdots, k_{\max}$ **do**

1. Compute the dirty residual: $\mathbf{I}_R^{(k)} = \mathbf{I}_D - \mathbf{\Phi}^*\mathbf{\Phi}\mathbf{I}^{(k-1)}$
2. Find the next reconstructed source: $s^{(k)} = \arg\max\ \mathbf{I}_R^{(k)}$
3. Update the iterate: $\mathbf{I}^{(k)} = \mathbf{I}^{(k-1)} + \alpha\delta_{s^{(k)}}$

end for



CLEAN

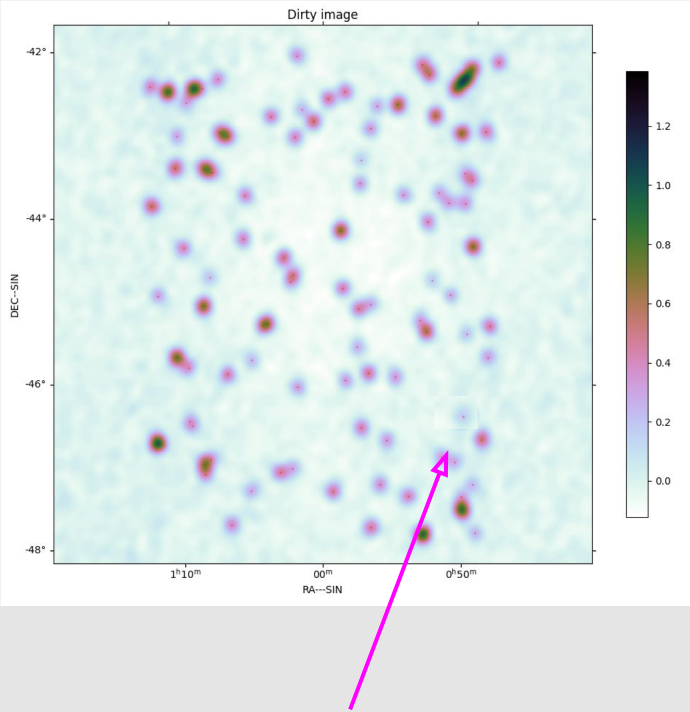
Algorithm 1 Högbom CLEAN Algorithm (Major cycles only)

Initialisation : $\mathbf{I}^{(0)} = \mathbf{0}$, $\mathbf{I}_D = \mathbf{\Phi}^*\mathbf{V}$, $\alpha > 0$

for $k = 1, 2, \cdots, k_{\max}$ **do**

1. Compute the dirty residual: $\mathbf{I}_R^{(k)} = \mathbf{I}_D - \mathbf{\Phi}^*\mathbf{\Phi}\mathbf{I}^{(k-1)}$
2. Find the next reconstructed source: $s^{(k)} = \arg\max\ \mathbf{I}_R^{(k)}$
3. Update the iterate: $\mathbf{I}^{(k)} = \mathbf{I}^{(k-1)} + \alpha\delta_{s^{(k)}}$

end for



\+ Optional post-processing (convolution, residual)

CLEAN-Like methods (continued)

- ✓ Long date expertise
- ✓ A lot of hacks and tips to make them very fast
- ✓ Atomic method (scalable)
- ✓ Calibration-compliant

CLEAN-Like methods (continued)

- ✓ Long date expertise
- ✓ A lot of hacks and tips to make them very fast
- ✓ Atomic method (scalable)
- ✓ Calibration-compliant

- ✗ Only denoising = enforcing the prior model
- ✗ Very sensitive to stop
- ✗ Objective function unclear

Optimization methods

- History of Compressed Sensing : next generation
- Proximal methods: **Fast algorithms + Convergence guarantees**
  - FISTA = APGD (LOFAR sparse image reconstruction[1])

$$\mathbf{I}^{(k+1)} \leftarrow \mathrm{prox}_{\lambda\tau\mathcal{R}}\left(\mathbf{I}^{(k)} - \tau\nabla f(\mathbf{I}^{(k)})\right)$$

  - PDS (SARA algorithms[2, 3, 4])

$$\mathbf{x}_{k+1} \leftarrow \mathrm{prox}_{\tau\mathcal{R}_1}\left(\mathbf{x}_k - \tau\nabla f(\mathbf{x}_k) - \tau\mathbf{K}^*\mathbf{z}_k\right) \qquad \mathbf{z}_{k+1} \leftarrow \mathrm{prox}_{\sigma\mathcal{R}_2^*}\left(\mathbf{z}_k + \sigma\mathbf{K}[2\mathbf{x}_{k+1} - \mathbf{x}_k]\right)$$

[1] H. Garsden *et al.*, "LOFAR sparse image reconstruction," *Astronomy & Astrophysics*, Mar. 2015

[2] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, *Monthly Notices of the Royal Astronomical Society*, Oct. 2012

[3] A. Abdulaziz, A. Dabbech, and Y. Wiaux, *Monthly Notices of the Royal Astronomical Society*, Oct. 2019

[4] P.-A. Thouvenin, A. Abdulaziz, A. Dabbech, A. Repetti, and Y. Wiaux, *Monthly Notices of the Royal Astronomical Society*, 2021.

Optimization methods (continued)

- ✓ Denoising (with only one parameter!)
- ✓ Can handle very complex priors
- ✓ Fast principled algorithms
- ✓ Bayesian interpretation

Optimization methods (continued)

- ✓ Denoising (with only one parameter!)
- ✓ Can handle very complex priors
- ✓ Fast principled algorithms
- ✓ Bayesian interpretation

- ✗ Completely different implementation paradigm
- ✗ Scalability issues
- ✗ Non calibration-compliant
- ✗ Shrinkage of the reconstructed intensity

Didn’t reach the production stage

The CLEAN-based Methods

Astronomers

The Convex Optimisation Methods

Signal processing community

Two worlds with different goals and different constraints

03.

PolyCLEAN

The place and contribution
of our method

1. Optimization method

Penalty-based prior (atomic norm)

2. Atomic behavior

CLEAN-like algorithmic structure
and minor cycles

3. Focus on scalability

Sparsity-informed computations with
Pycsou and HVOX (nufft)

1. Optimization method

Penalty-based prior (atomic norm)

$$\boxed{\lambda\|\mathbf{I}\|_1\,,\ \mathbf{I} \geqslant 0}$$

2. Atomic behavior

CLEAN-like algorithmic structure
and minor cycles

3. Focus on Scalability

Sparsity-informed computations with
Pycsou and HVOX (nufft)

1. Optimization method

Penalty-based prior (atomic norm)

$$\boxed{\lambda \|\mathbf{I}\|_1 \, , \, \mathbf{I} \geqslant 0}$$

2. Atomic behavior

CLEAN-like algorithmic structure
and minor cycles

$$\boxed{\mathbf{I} = \sum \alpha_k \boldsymbol{\delta}_{i_k}}$$

3. Focus on scalability

Sparsity-informed computations with
Pycsou and HVOX (nufft)

1. Optimization method

Penalty-based prior (atomic norm)

$$\boxed{\lambda\|\mathbf{I}\|_1 \,, \ \mathbf{I} \geqslant 0}$$

2. Atomic behavior

CLEAN-like algorithmic structure
and minor cycles

$$\boxed{\mathbf{I} = \sum \alpha_k \boldsymbol{\delta}_{i_k}}$$

3. Focus on scalability

$$\boxed{\mathbf{\Phi}\left(\mathbf{I}^{(k)}\right)}$$

Sparsity-informed computations with
Pycsou and HVOX (nufft)

The Landscape of Methods

CLEAN

Convex Optimization

- L1 penalty and positivity
- Interpretability
- Support estimation

- Atomic algorithm
- Scalability

PolyCLEAN

The Algorithm

Algorithm 1 PolyCLEAN

Initialisation : $\mathbf{I}^{(0)} = \mathbf{0}$, $\mathcal{S}^{(0)} = \mathrm{Supp}(\mathbf{I}^{(0)}) = \emptyset$, $\mathbf{I}_D = \mathbf{\Phi}^* \mathbf{V}$

while `stopping_criterion(`$\mathbf{I}^{(k)}$`)` not reached **do**

1. Compute the dirty residual: $\mathbf{I}_R^{(k)} = \mathbf{I}_D - \mathbf{\Phi}^* \mathbf{\Phi} \mathbf{I}^{(k-1)}$

end while

The Algorithm

Algorithm 1 PolyCLEAN

Initialisation : $\mathbf{I}^{(0)} = \mathbf{0}$, $\mathcal{S}^{(0)} = \mathrm{Supp}(\mathbf{I}^{(0)}) = \emptyset$, $\mathbf{I}_D = \mathbf{\Phi}^* \mathbf{V}$

while `stopping_criterion`$(\mathbf{I}^{(k)})$ not reached **do**

1. Compute the dirty residual: $\mathbf{I}_R^{(k)} = \mathbf{I}_D - \mathbf{\Phi}^* \mathbf{\Phi} \mathbf{I}^{(k-1)}$

2. Place many candidate sources: $s_1^{(k)}, s_2^{(k)}, \cdots =$ `highest_level_set` $(\mathbf{I}_R^{(k)})$
   Update active set : $\mathcal{S}^{(k)} \leftarrow \mathcal{S}^{(k-1)} \cup \{s_1^{(k)}, s_2^{(k)}, \ldots\}$

end while

The Algorithm

Algorithm 1 PolyCLEAN

Initialisation : $\mathbf{I}^{(0)} = \mathbf{0}$, $\mathcal{S}^{(0)} = \mathrm{Supp}(\mathbf{I}^{(0)}) = \emptyset$, $\mathbf{I}_D = \mathbf{\Phi}^*\mathbf{V}$

while `stopping_criterion`$(\mathbf{I}^{(k)})$ not reached **do**

1. Compute the dirty residual: $\mathbf{I}_R^{(k)} = \mathbf{I}_D - \mathbf{\Phi}^*\mathbf{\Phi}\mathbf{I}^{(k-1)}$

2. Place many candidate sources: $s_1^{(k)}, s_2^{(k)}, \cdots =$ `highest_level_set` $(\mathbf{I}_R^{(k)})$
   Update active set : $\mathcal{S}^{(k)} \leftarrow \mathcal{S}^{(k-1)} \cup \{s_1^{(k)}, s_2^{(k)}, \dots\}$

3. Update the iterate:

$$\mathbf{I}^{(k)} = \underset{\substack{\mathrm{Supp}(\mathbf{I}) \subset \mathcal{S}^{(k)} \\ \mathbf{I} \geqslant 0}}{\arg\min} \frac{1}{2} \|\mathbf{V} - \mathbf{\Phi}\mathbf{I}\|_2^2 + \lambda \|\mathbf{I}\|_1 \tag{R}$$

end while

The Algorithm

Algorithm 1 PolyCLEAN

Initialisation : $\mathbf{I}^{(0)} = \mathbf{0}$, $\mathcal{S}^{(0)} = \mathrm{Supp}(\mathbf{I}^{(0)}) = \emptyset$, $\mathbf{I}_D = \mathbf{\Phi}^*\mathbf{V}$

while `stopping_criterion`$(\mathbf{I}^{(k)})$ not reached **do**

1. Compute the dirty residual: $\mathbf{I}_R^{(k)} = \mathbf{I}_D - \mathbf{\Phi}^*\mathbf{\Phi}\mathbf{I}^{(k-1)}$
2. Place many candidate sources: $s_1^{(k)}, s_2^{(k)}, \cdots =$ `highest_level_set` $(\mathbf{I}_R^{(k)})$
   Update active set : $\mathcal{S}^{(k)} \leftarrow \mathcal{S}^{(k-1)} \cup \{s_1^{(k)}, s_2^{(k)}, \ldots\}$
3. Update the iterate: (approximately)

$$\mathbf{I}^{(k)} = \underset{\substack{\mathrm{Supp}(\mathbf{I}) \subset \mathcal{S}^{(k)} \\ \mathbf{I} \geqslant 0}}{\arg\min} \frac{1}{2} \left\| \mathbf{V} - \mathbf{\Phi}\mathbf{I} \right\|_2^2 + \lambda \left\| \mathbf{I} \right\|_1 \tag{R}$$

end while

Support Identification

LASSO dual certificate:

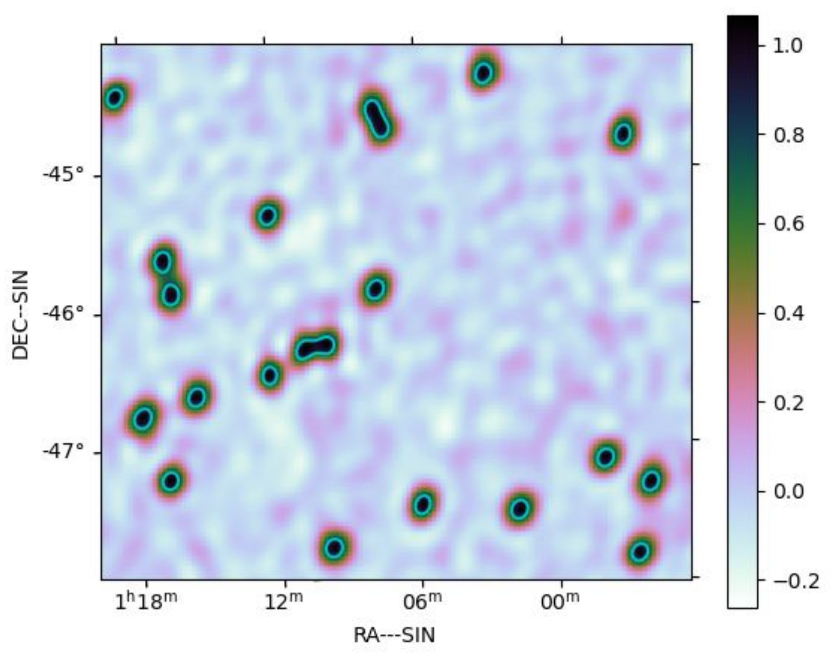
$$\begin{cases} \boldsymbol{\eta} \in [-1, 1]^N \\ \langle \boldsymbol{\eta}, \mathbf{I}^* \rangle = \|\mathbf{I}^*\|_1 \end{cases}$$

Support Identification

LASSO dual certificate:

$$\begin{cases} \boldsymbol{\eta} \in [-1, 1]^N \\ \langle \boldsymbol{\eta}, \mathbf{I}^* \rangle = \|\mathbf{I}^*\|_1 \end{cases}$$

$$\boxed{\boldsymbol{\eta} = \frac{1}{\lambda}\boldsymbol{\Phi}^* \left(\mathbf{V} - \boldsymbol{\Phi}\mathbf{I}^*\right)}$$



Extensions

- Extended Sources
- PolyCLEAN +

Extensions

- Extended Sources

Parametric expression of the sky image (dictionary, wavelets):

$$\mathbf{I} = \mathbf{\Psi\Theta}$$

$$\boxed{\arg\min \frac{1}{2}\|\mathbf{y} - \mathbf{\Phi\Psi\Theta}\|_2^2 + \lambda\|\mathbf{\Theta}\|_1}$$

- PolyCLEAN +

Extensions

- Extended Sources

Parametric expression of the sky image (dictionary, wavelets):

$$\mathbf{I} = \mathbf{\Psi\Theta}$$

$$\boxed{\arg\min \frac{1}{2}\|\mathbf{y} - \mathbf{\Phi\Psi\Theta}\|_2^2 + \lambda\|\mathbf{\Theta}\|_1}$$

- PolyCLEAN +

Post-processing: Account for the shrinking of the LASSO with least squares reweighting

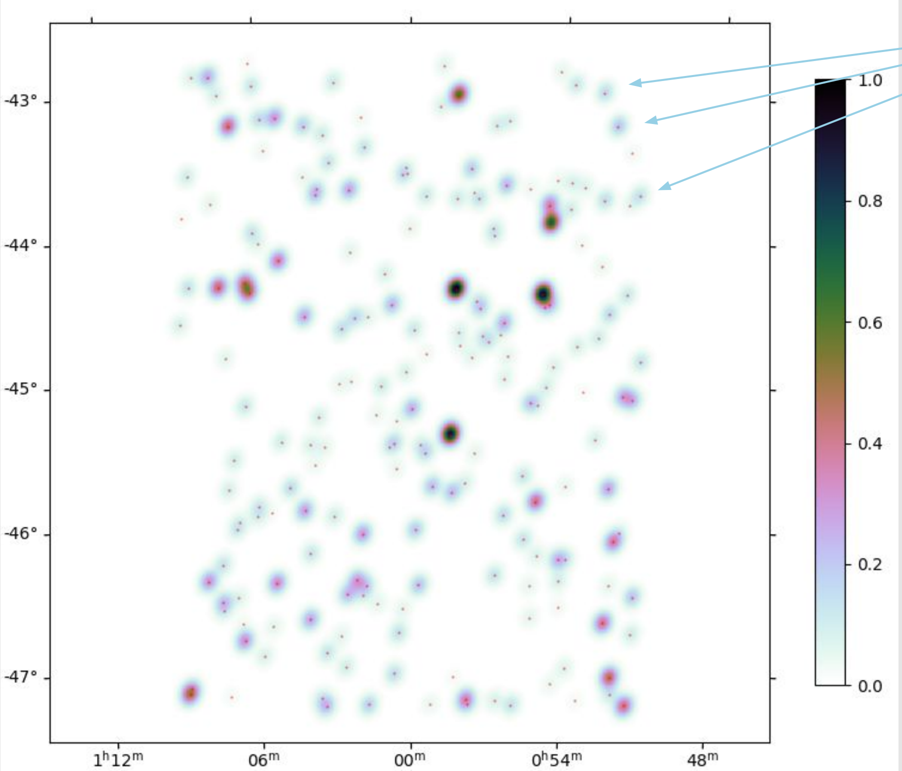
$$\boxed{\underset{\substack{\mathbf{I}\geqslant 0 \\ \mathrm{Supp}(\mathbf{I})\subset\mathrm{Supp}(\mathbf{I}^*)}}{\arg\min} \|\mathbf{V} - \mathbf{\Phi I}\|_2^2}$$

04. Experiments

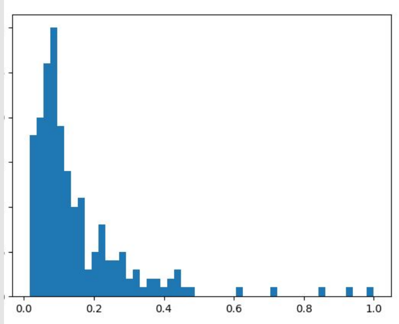
Simulations

Point sources with
sharp smoothing kernel

1024 pixels
=
5° FOV



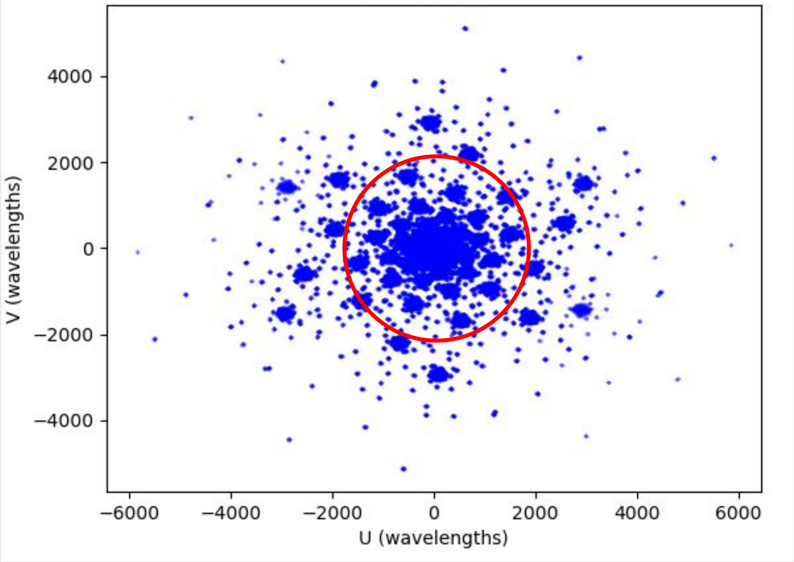
200 sources



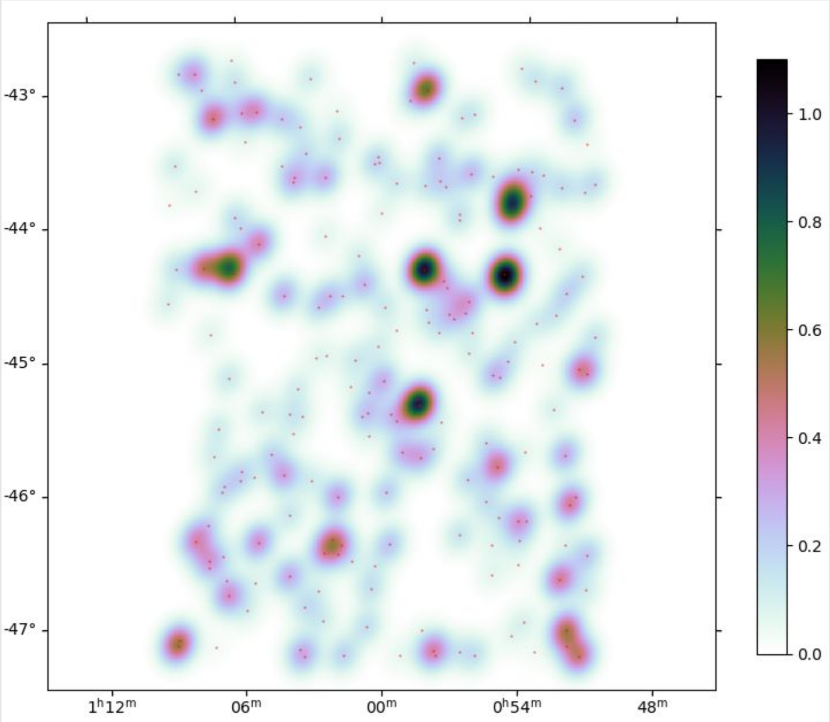
Simulations

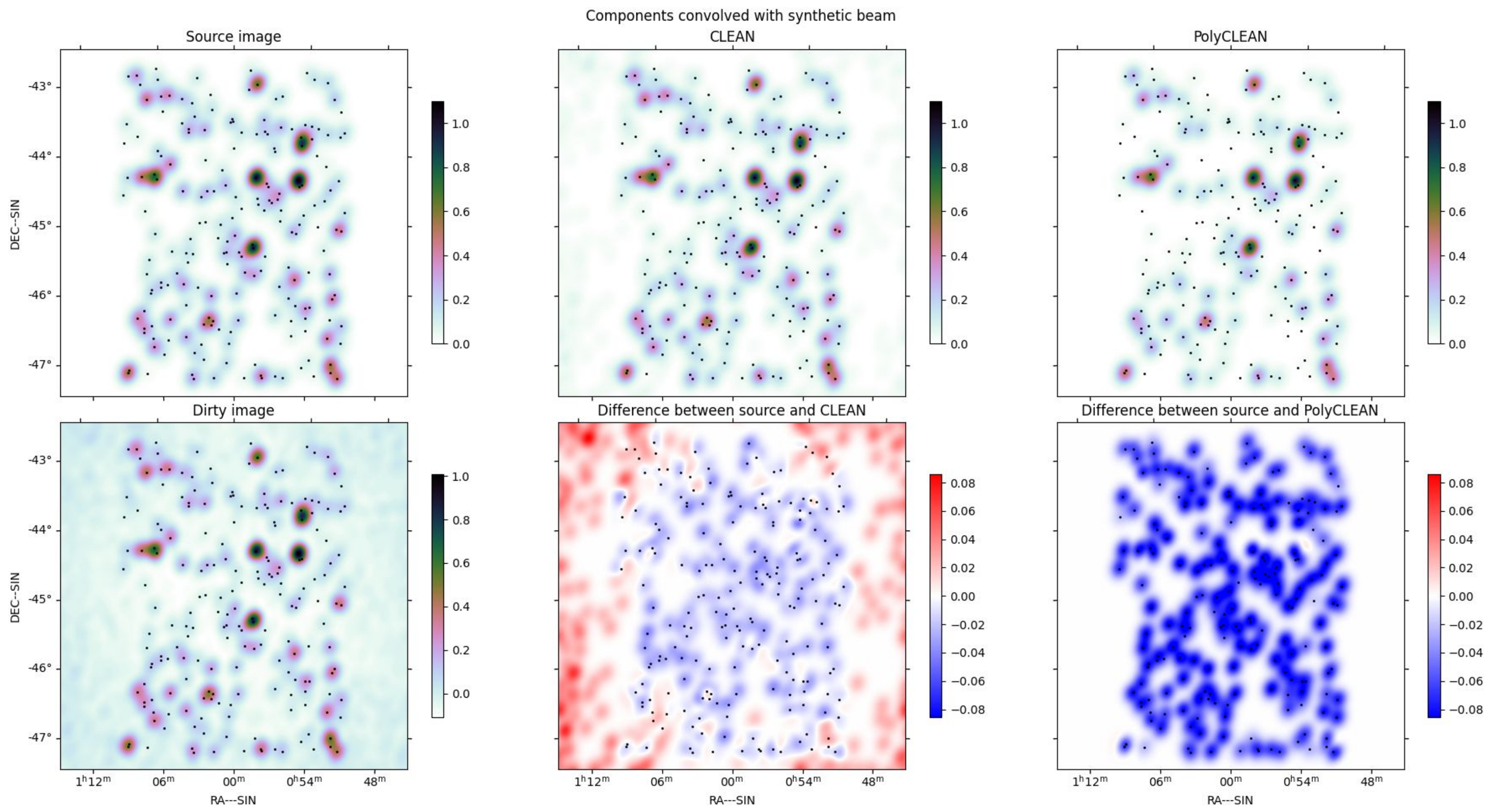
Number of baselines
(SKA Low configuration):

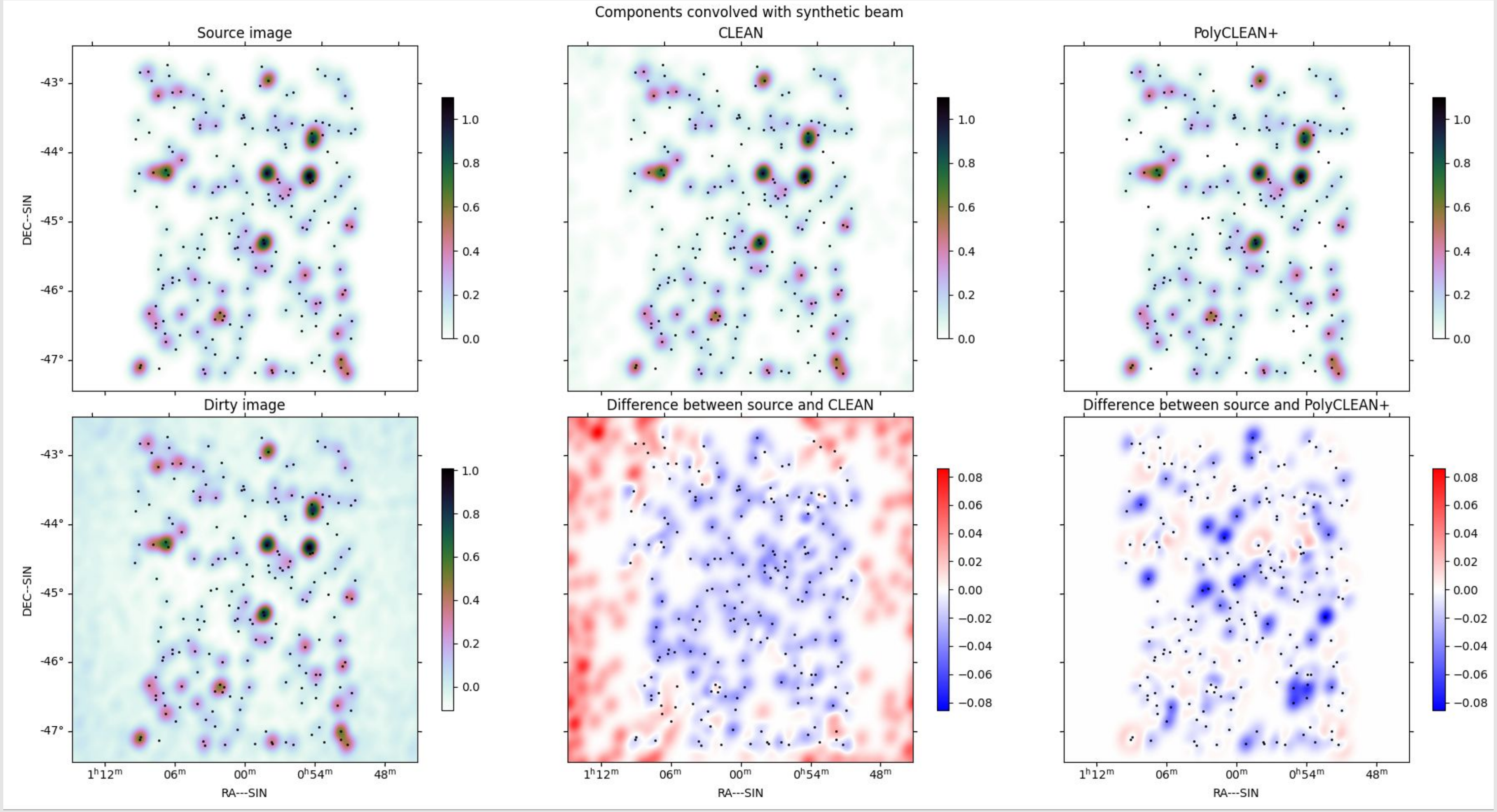
500m -> 18500



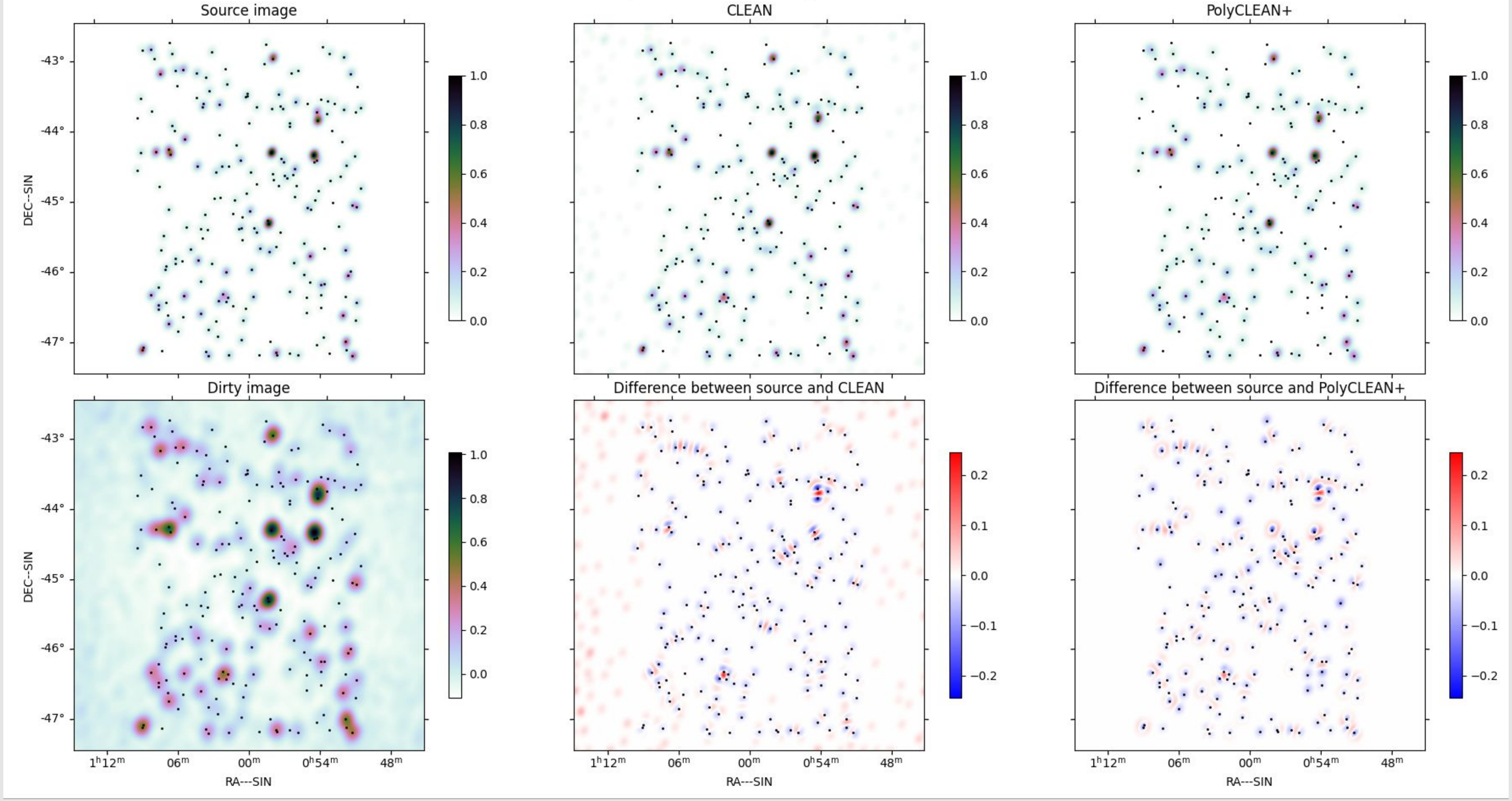
Sky image convolved with synthetic beam



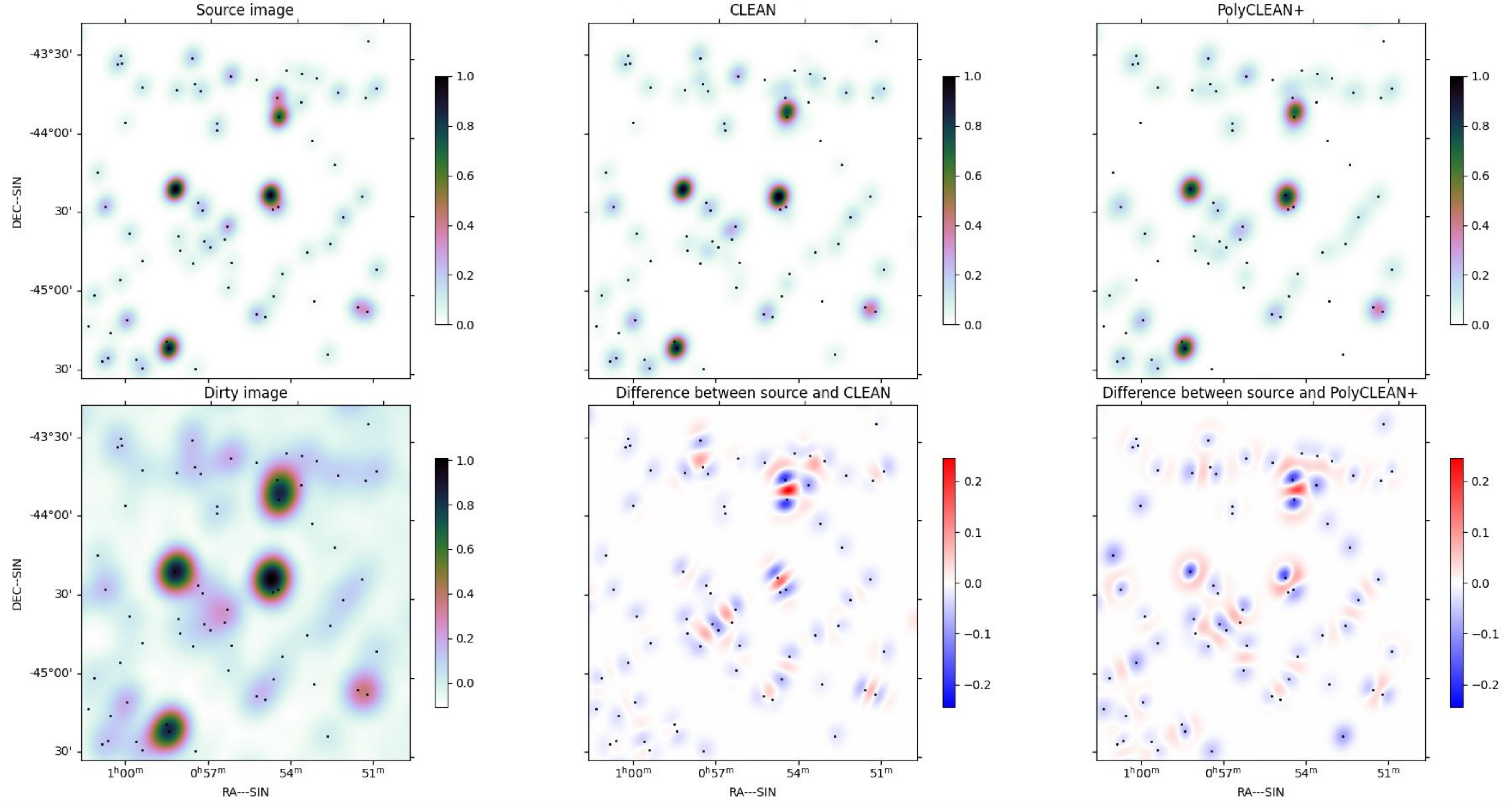

Components convolved with synthetic beam
Source image
CLEAN
PolyCLEAN
Dirty image
Difference between source and CLEAN
Difference between source and PolyCLEAN
DEC--SIN
RA---SIN
-43°
-44°
-45°
-46°
-47°
1h12m
06m
00m
0h54m
48m
0.0
0.2
0.4
0.6
0.8
1.0
0.08
0.06
0.04
0.02
0.00
−0.02
−0.04
−0.06
−0.08


Components convolved with synthetic beam
Source image
CLEAN
PolyCLEAN+
Dirty image
Difference between source and CLEAN
Difference between source and PolyCLEAN+
DEC--SIN
RA---SIN
-43°
-44°
-45°
-46°
-47°
1h12m
06m
00m
0h54m
48m
1.0
0.8
0.6
0.4
0.2
0.0
0.08
0.06
0.04
0.02
0.00
−0.02
−0.04
−0.06
−0.08

Components convolved with very sharp beam



Components convolved with very sharp beam (zoom)



Results

Reconstruction metrics

	rmax = 500 (~ 18500 baselines)				
	Dirty Image	CLEAN	Poly CLEAN	Poly CLEAN+	APGD
Duration (s)	-	**11.2**	17.3	19.3	38.7
MSE ($\times 10^{-4}$)					
wide	252	7.14	38.2	**2.66**	35.3
synthetic	52.4	2.38	12.0	**1.13**	11.2
sharp	25.6	1.19	4.86	**0.81**	4.48
very sharp	42.3	1.27	3.02	**1.11**	2.76
components	92.2	0.13	0.09	**0.09**	0.09
Sparsity	-	**486**	~ 45000	~ 43500	~ 37000

Perspectives

1. Real world datasets

- Many parameters
- A lot of flexibility for CLEAN as well as PolyCLEAN
- Difficult to simulate noise

➢ Define an experimental setup

2. Extended sources

- Many possibilities:
  - Dictionary
  - Wavelets
- Few code required
  - Generic framework
  - Mostly done

3. Framework byproducts

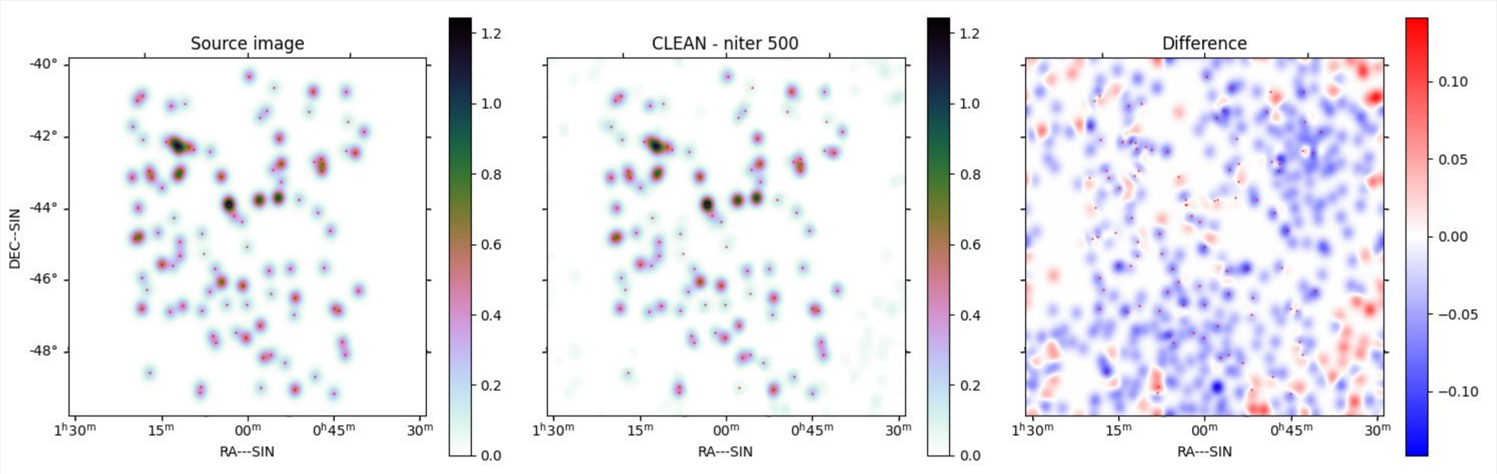
- Analysis of the dual certificate
- Bayesian tests

Thanks!

	CLEAN	MAP Estimation	PolyCLEAN
Sparse iterates	✅	❌	✅
Flexible priors	❌	✅	~
Fast solvers	✅	~	✅
Calibration compliant	✅	❌	✅
Interpretable obj. function	❌	✅	✅

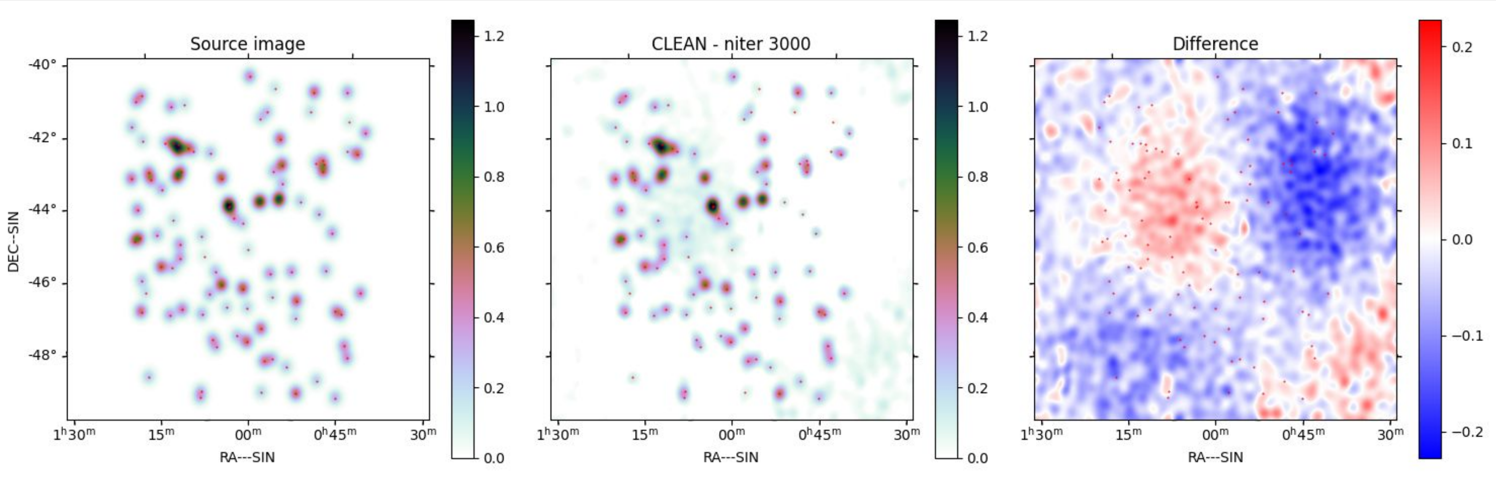
Sensibility of CLEAN w.r.t Stop

Iterations	500	1000	2000	3000
Run time (s)	12.6	44.8	81.0	107.2
MSE (e-3)	0.6	1.1	2.7	3.7



Sensibility of CLEAN w.r.t stop

Iterations	500	1000	2000	3000
Run time (s)	12.6	44.8	81.0	107.2
MSE (e-3)	0.6	1.1	2.7	3.7



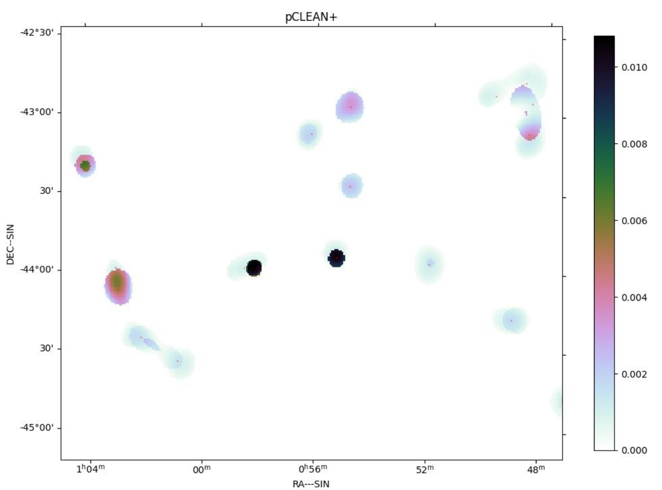
Sparsity issue

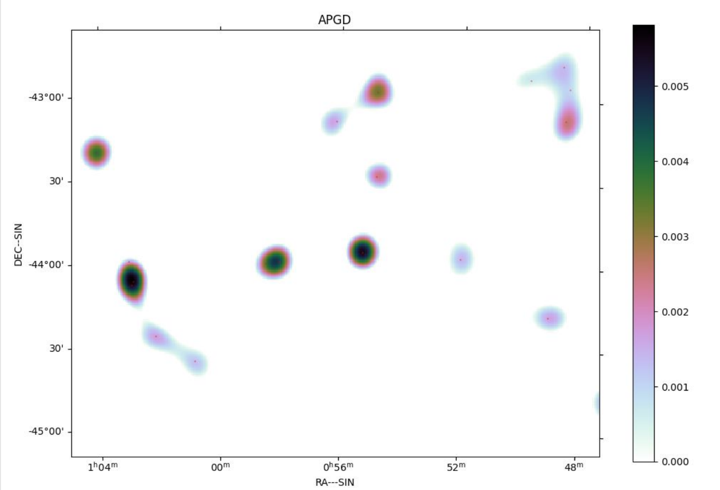
With the sizes of the data involved, LASSO solutions are not as sparse as CLEAN:

32k vs 1k PS

Workarounds:

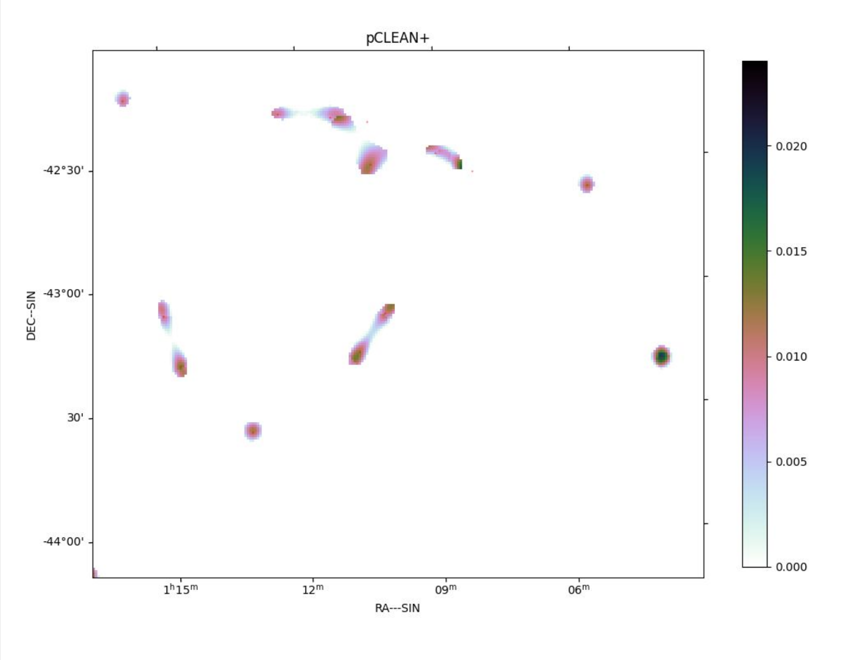
- Higher lambda
- More precise stopping crit
- Change the multi spikes strategy
- Gaussian dictionaries ?

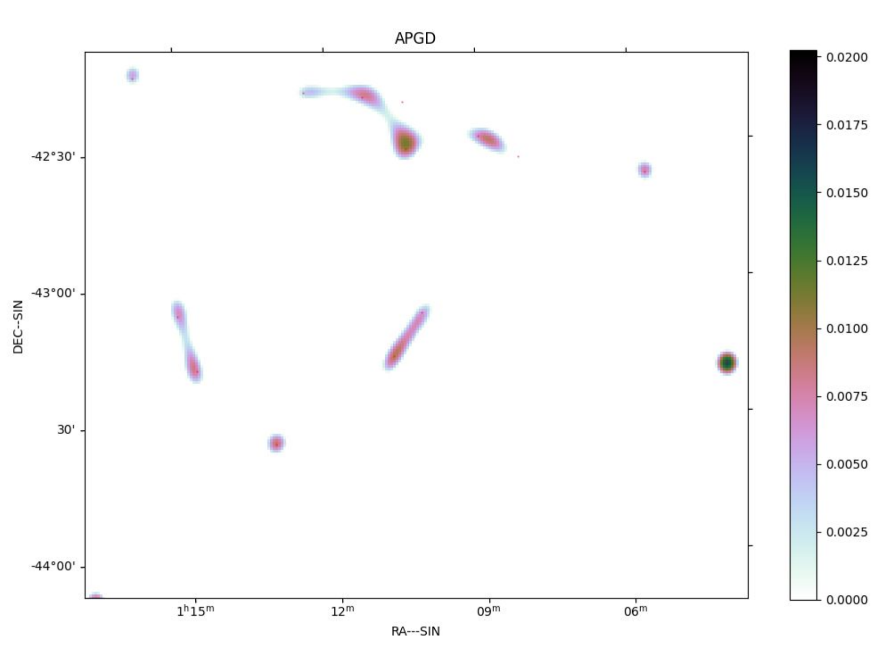




Sparsity issue

- More accuracy (so longer to run) *[PolyCLEAN+: 40s - APGD 179s - CLEAN: 24s]*





The CLEAN-based Methods

- Astronomers
- Efficient methods
- Produce science content: End goal
- Small brick in a long pipeline

The Convex Optimisation Methods

- Signal processing community
- Principled and satisfying methods, but difficulties to reach the astronomers
    - Algorithms,
    - Complex methods,
    - Scalability,
    - Lack of trust on the images.