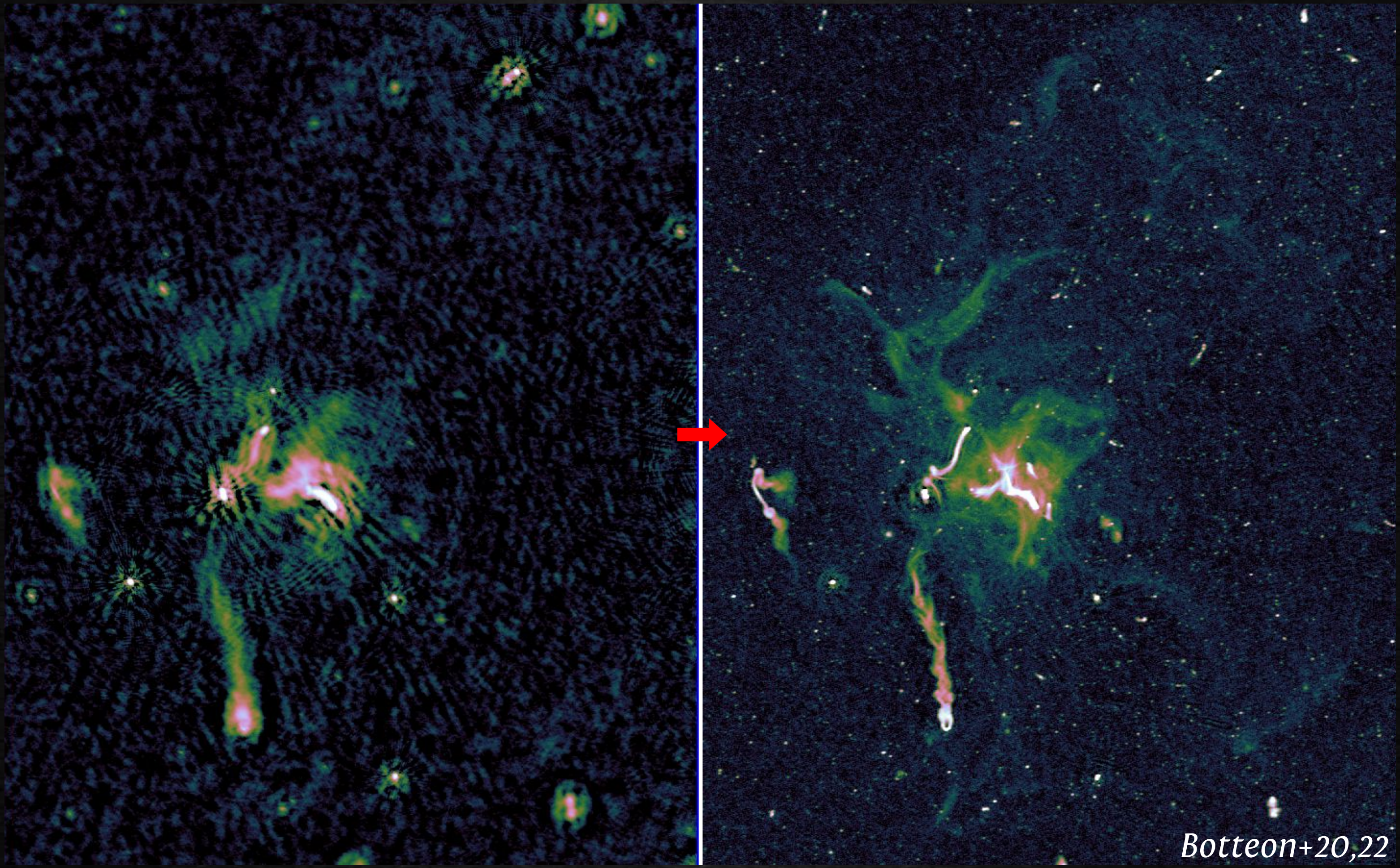
Analysis techniques to characterize extended sources

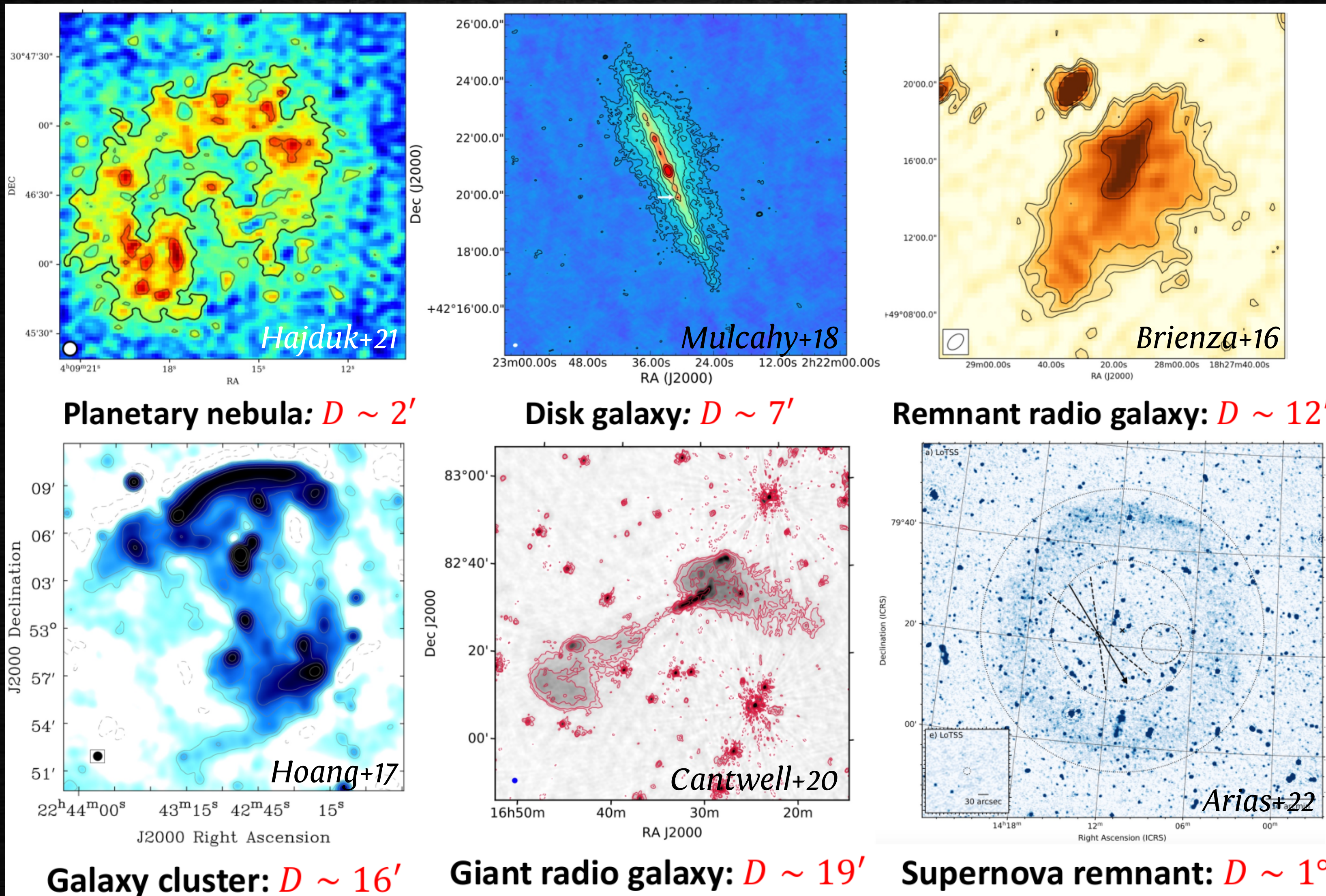


Andrea BOTTEON

INAF – IRA

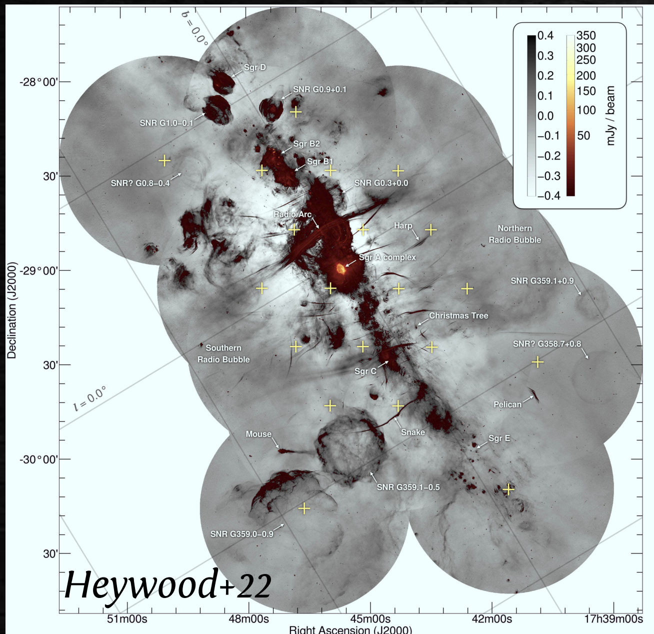
April 18, 2024 – The seventh LOFAR Data School, Dwingeloo

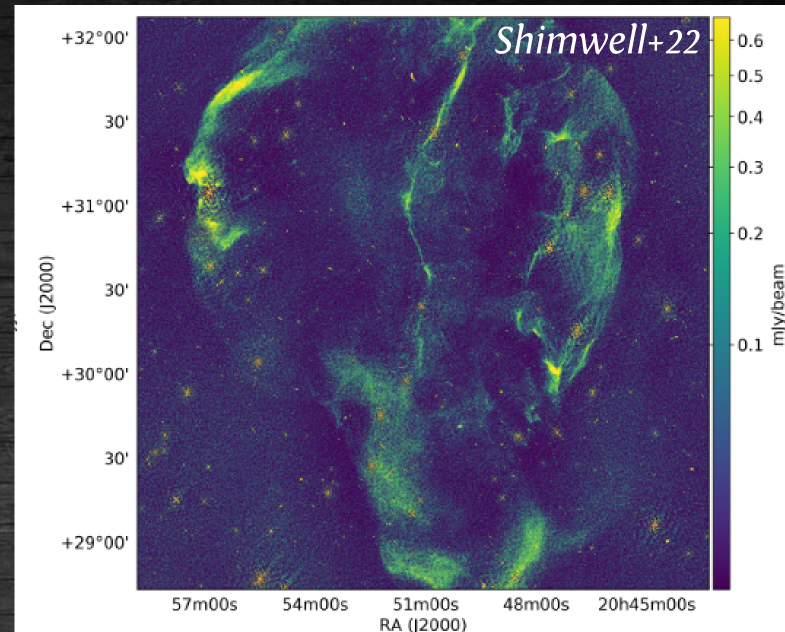
Extended radio sources...

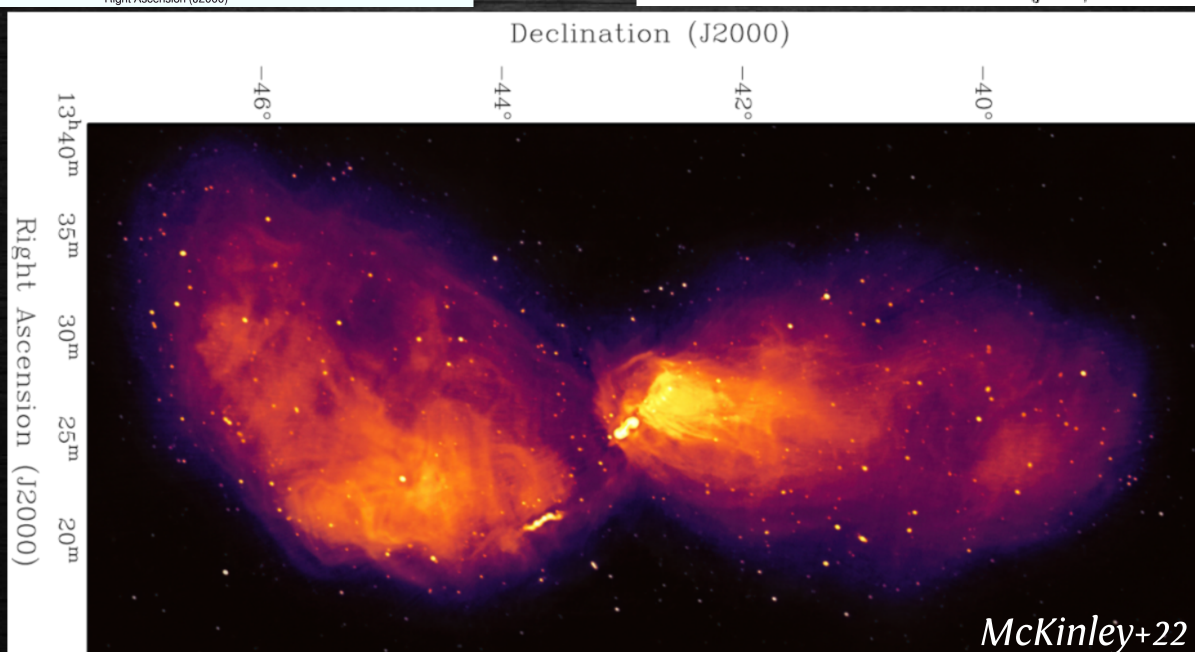


Diffuse radio sources can span a *wide* range of angular scales and can have very different origin (*galactic* and *extragalactic*)

...and *very* extended radio sources

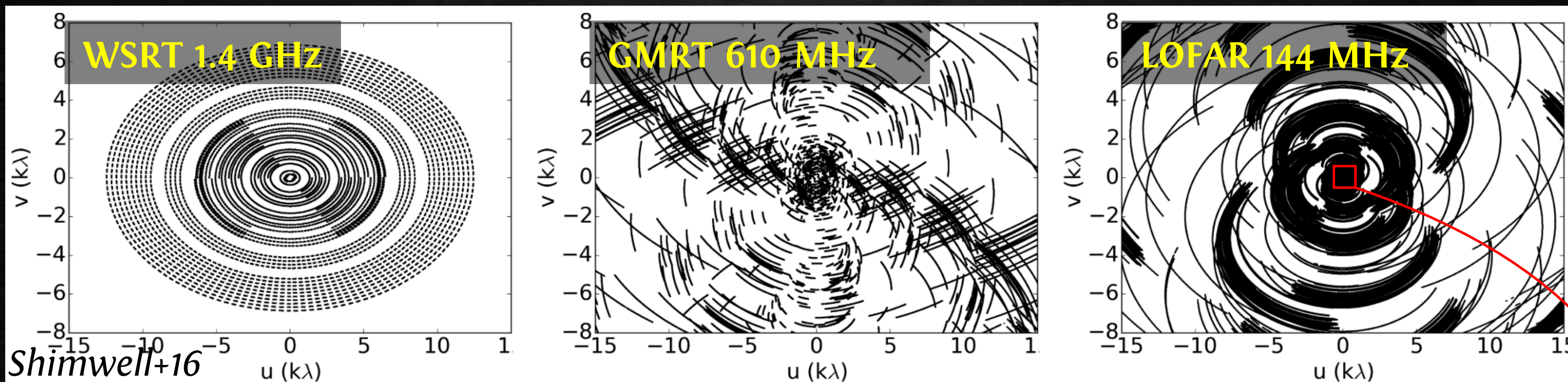






LOFAR uv sampling

Interferometers **discretely** sample the uv-plane



The *shortest baseline* determines the *maximum angular scale* that the interferometer can recover in the sky:

$$\theta_{\max} = \lambda / B_{\min} = 206265 / uv_{\min} \text{ [arcsec]}$$

(e.g. 1 deg ~ 57 lambda)

LOFAR has a *very dense inner* uv-coverage
→ critical for recovering *extended* sources

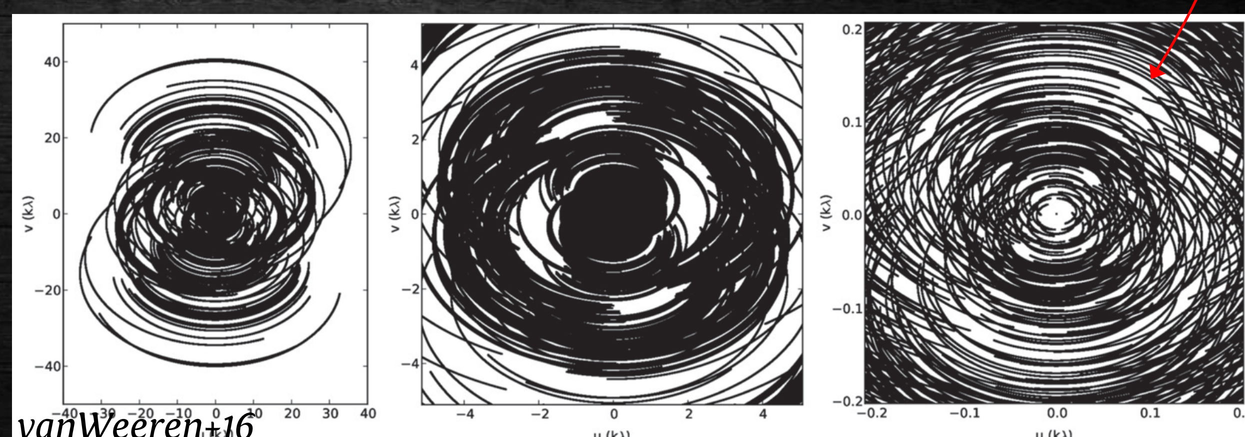
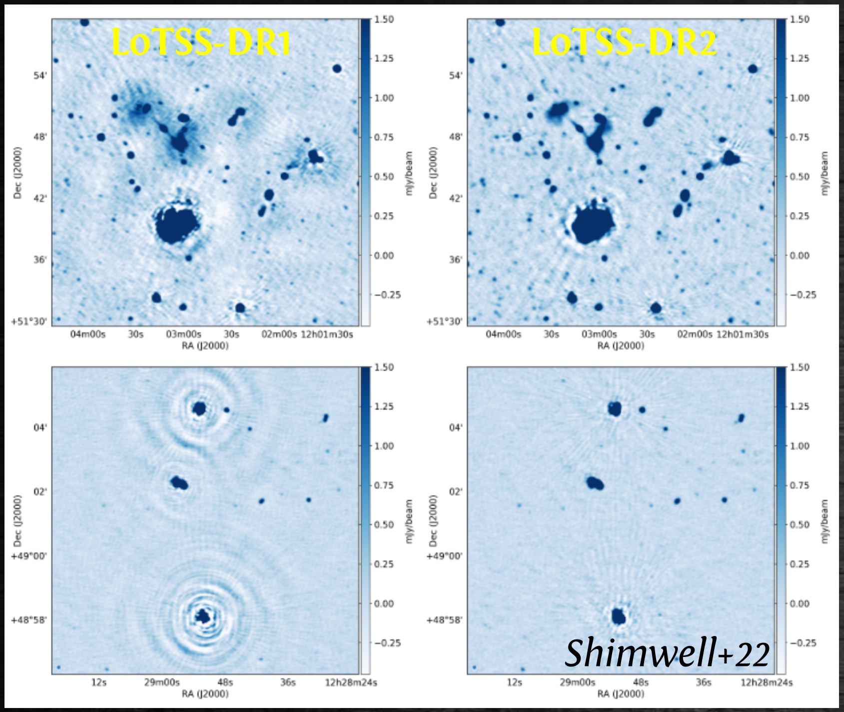
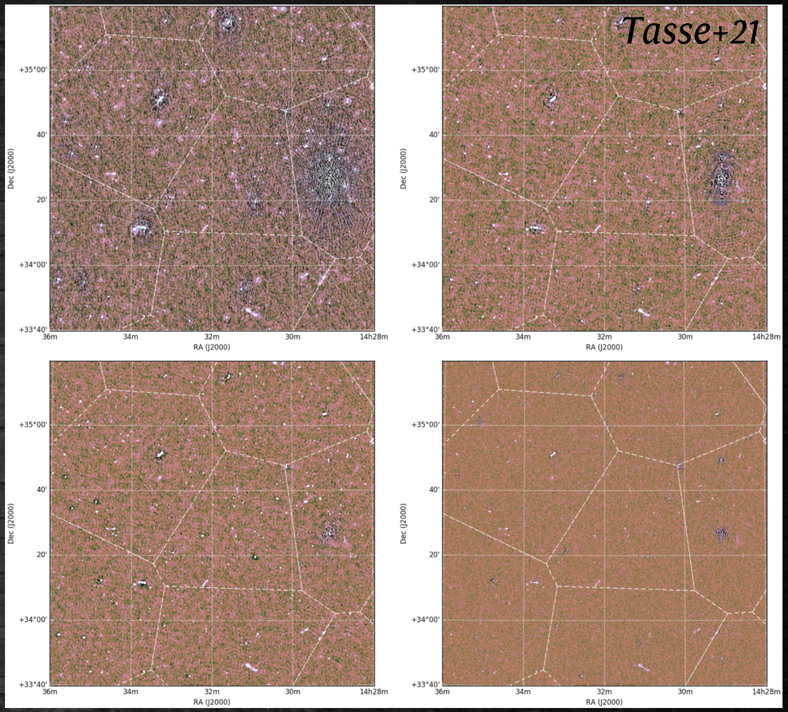


Image fidelidy

Diffuse emission has *low surface brightness*: high fidelity images are required



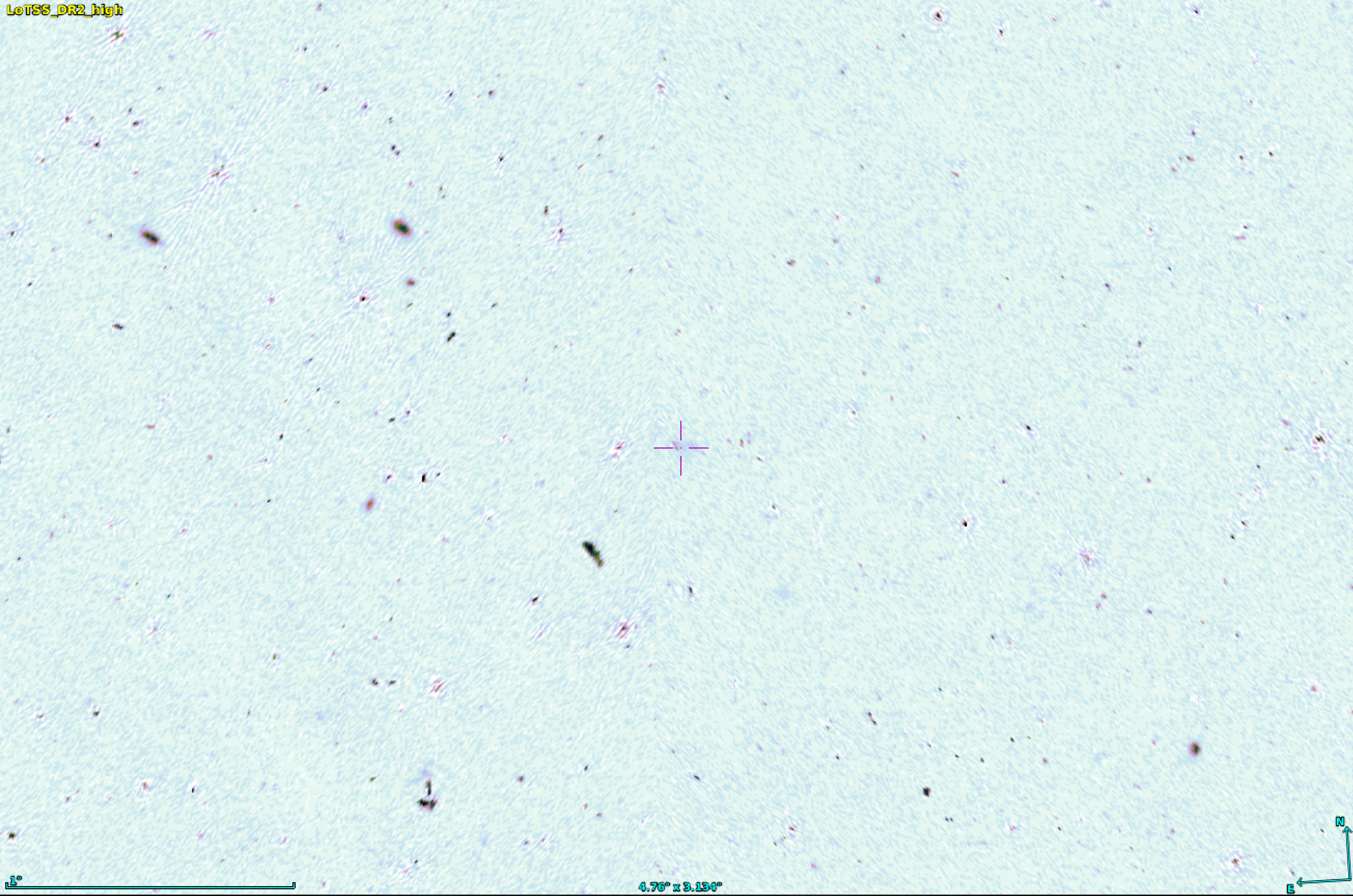


Obtaining the **best** calibration possible is *key* to produce artifact-free sensitive images with high dynamic range

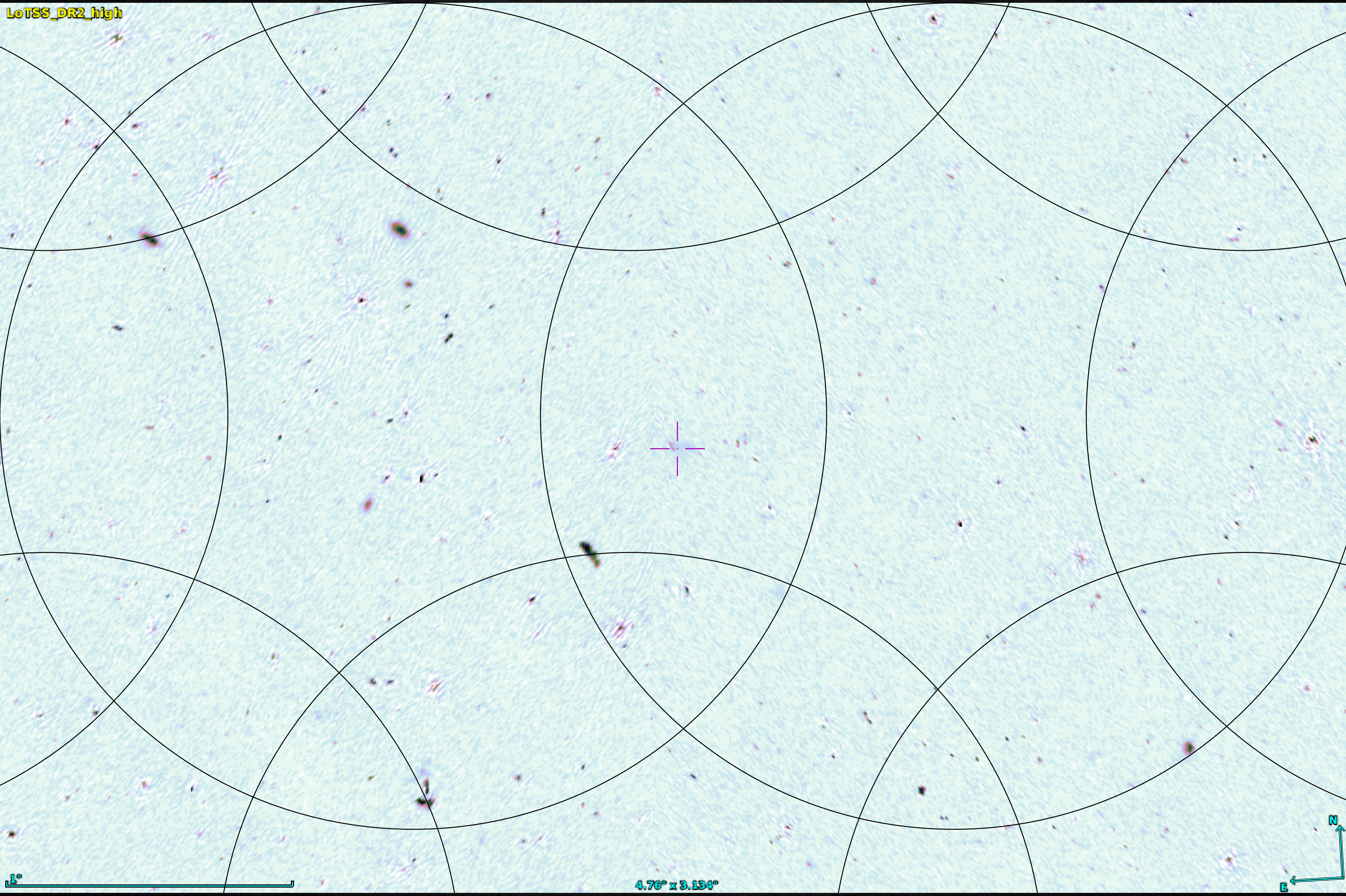
The *pipelines* developed to process **LOFAR surveys** (LoTSS & LoTSS) do a very good job, but *improvements are possible* towards specific (extended) targets

→ extract+selfcal method (https://github.com/rvweeren/lofar_facet_selfcal)

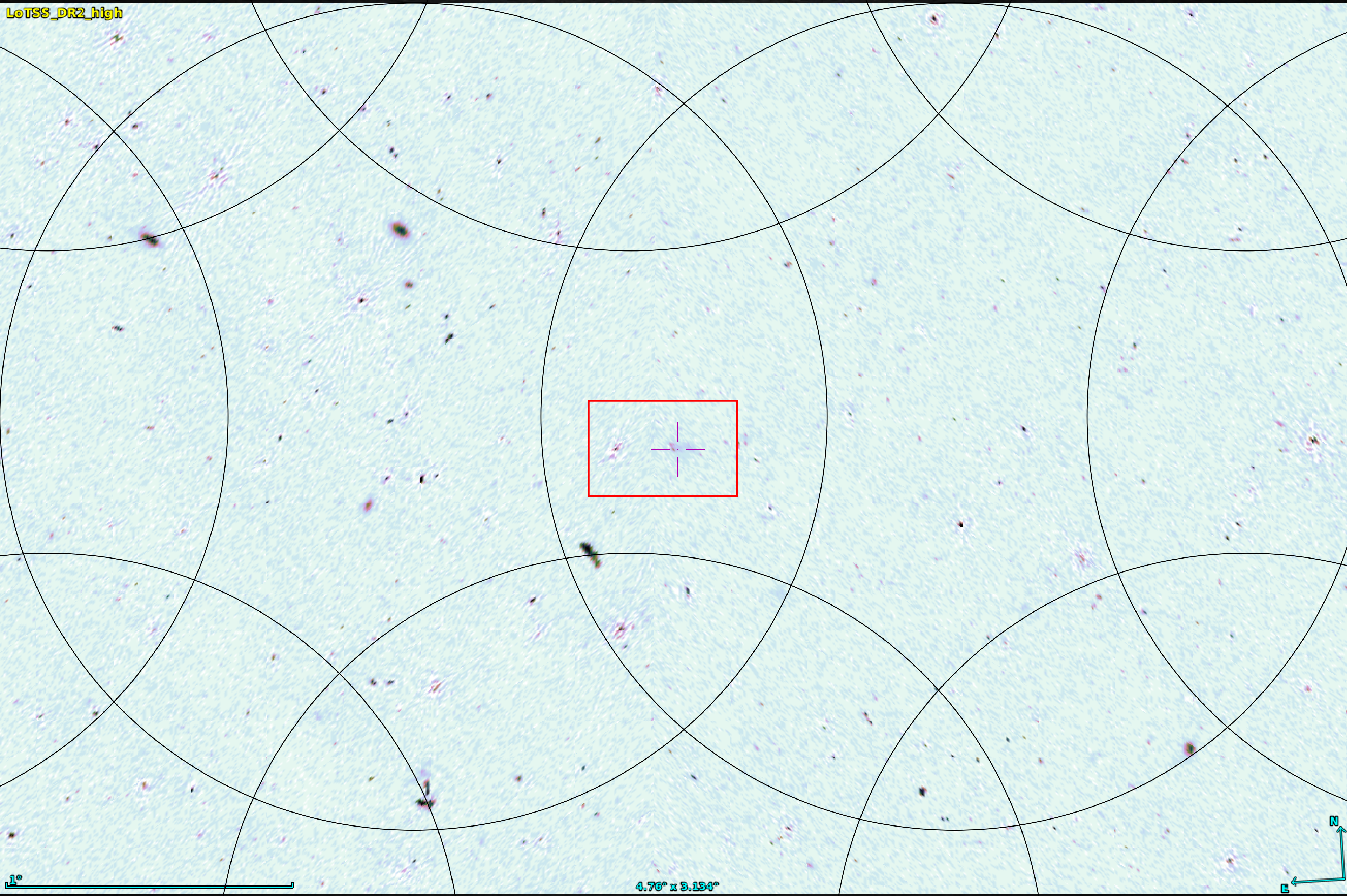
Extraction+selfcal



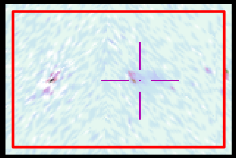
Extraction+selfcal



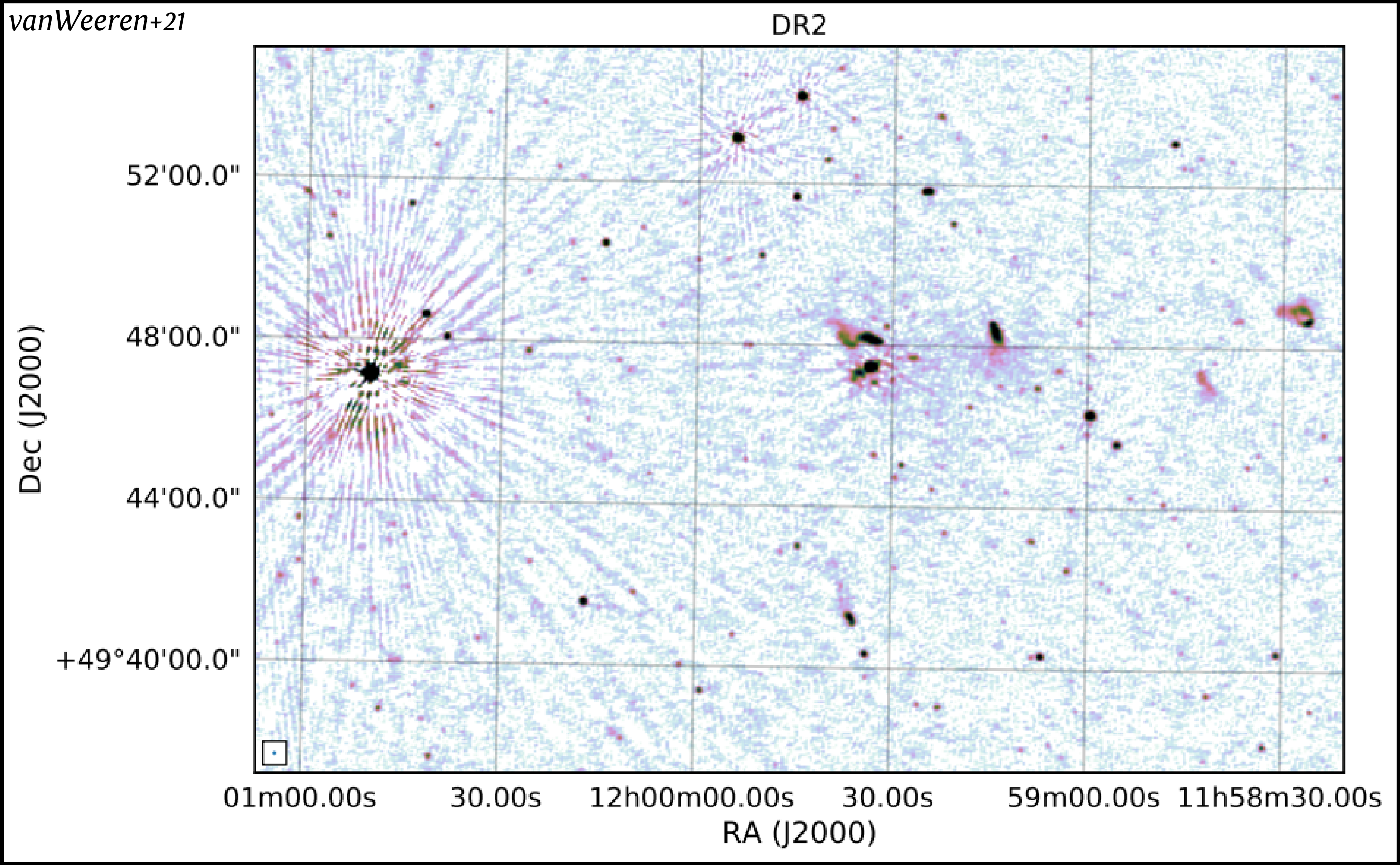
Extraction+selfcal



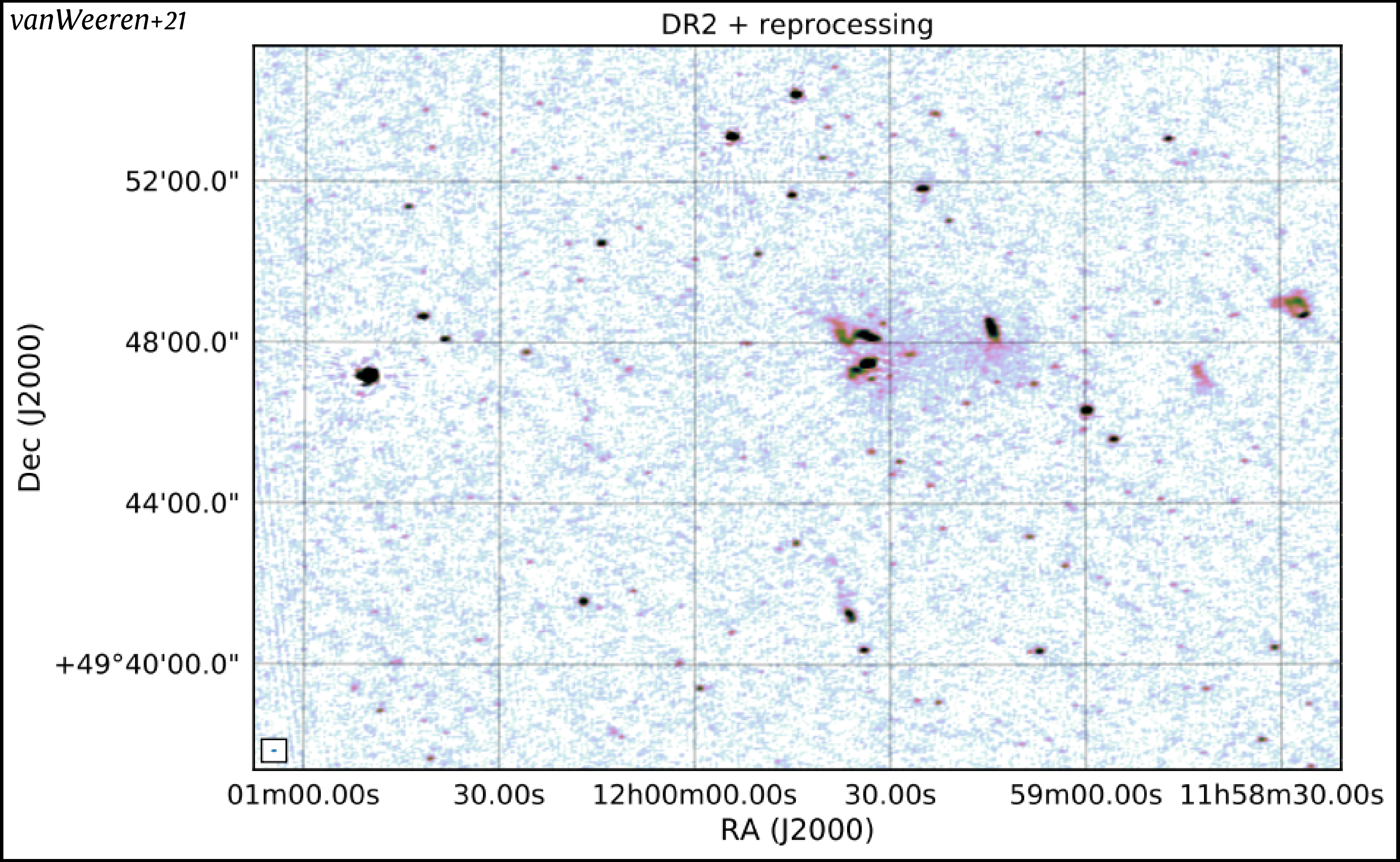
Extraction+selfcal



Extraction+selfcal



Extraction+selfcal



Exploit LoTSS-DR2 by yourselves

The pipeline used to process **LoTSS** observations is available at
https://github.com/mhardcastle/ddf-pipeline

The document `docs/extract.md` contains information on how to extract and calibrate targets from LoTSS-DR2

Example:

```
extraction.py NGC507
```

Look up the object NGC507 to determine a position and do a default-size extraction of a region of 0.5 degrees square around the catalogued position.

```
extraction.py myfield 0.4 286.1918961 59.8494461
```

Extract a region of 0.4 degrees square around the specified RA and DEC, naming the working directory myfield.

*Extraction will create a directory with the downloaded pipeline output and concatenated broad-band measurement sets, one per observation (which may mean more than one per field). These measurements sets will be in per-observation directories with names *.dysco.sub.shift.avg.weights.ms.archive?.*

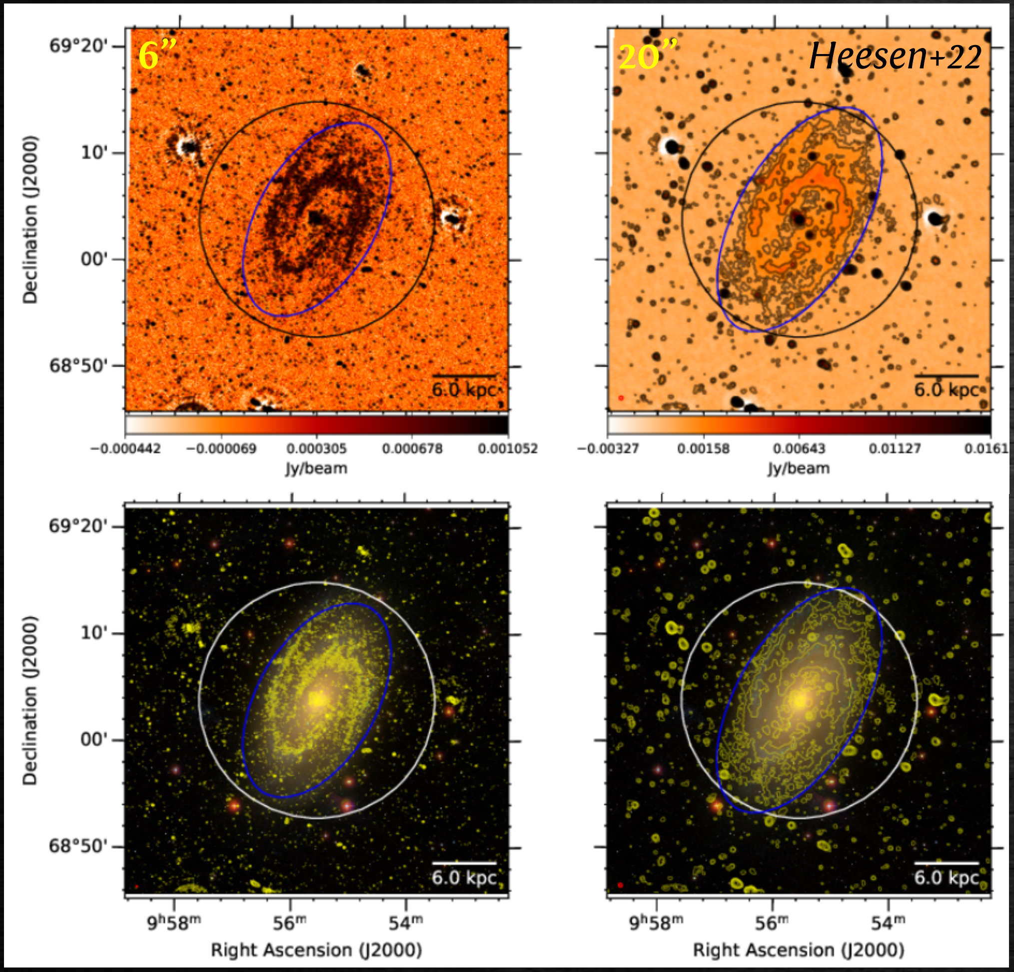
Inside the working directory for the target, created by the extraction, do

```
facetselfcal.py --auto --remove-flagged-from-startend \
 --helperscriptspath PATH -b NAME.ds9.reg -i NAME \
 *.dysco.sub.shift.avg.weights.ms.archive?
```

where the NAME is the name of your target and PATH the location of your directory with facetselfcal.py

What to do once data are calibrated?

Even if **LOFAR** can recover extended emission at relatively *high resolution*, *low resolution* imaging is desired to *increase the S/N of the detection* and better characterize the properties of the diffuse emission



In both cases, contours start at **3** sigma

Low resolution images can be obtained:

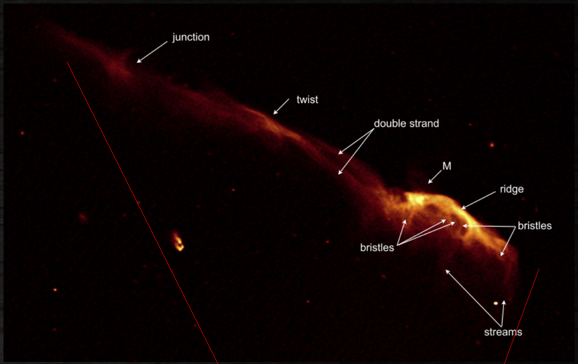
1) *a posteriori*, convolving high-resolution images by a Gaussian

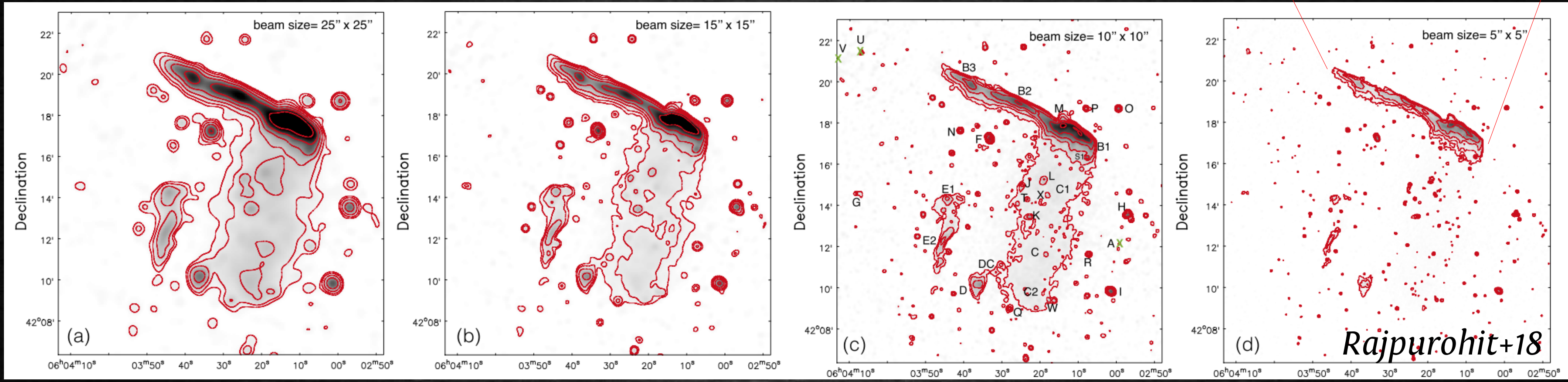
2) *a priori*, using a taper to give more weight to short baselines during imaging

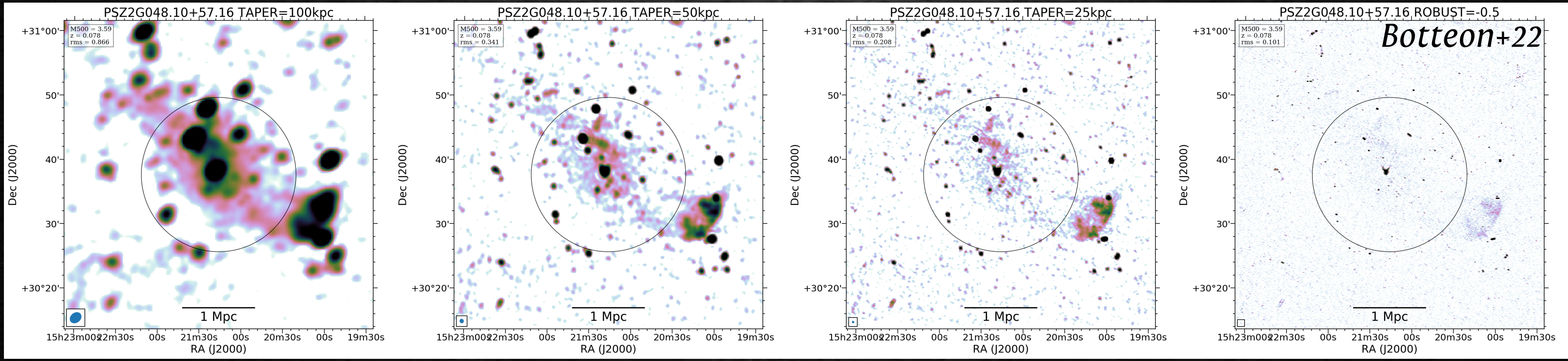
The deconvolution of extended emission **is not a simple task**, that's why approach **#1** is *risky*. Also, remember that **multi-scale cleaning** is *fundamental* for diffuse sources

What does low resolution mean?

It depends on the target/scientific goal. Generally, images at *multiple resolutions* are produced because they provide **complementary** information





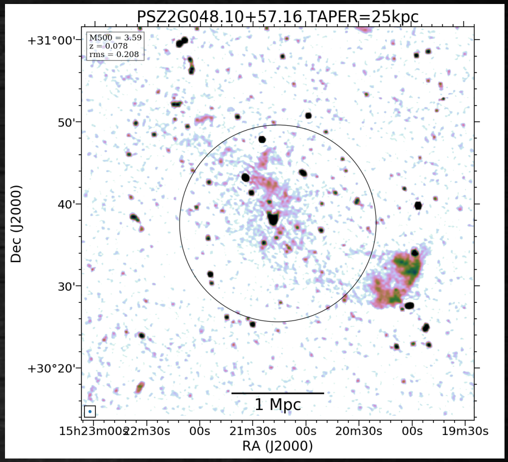


Do you see the problem of low resolution imaging?

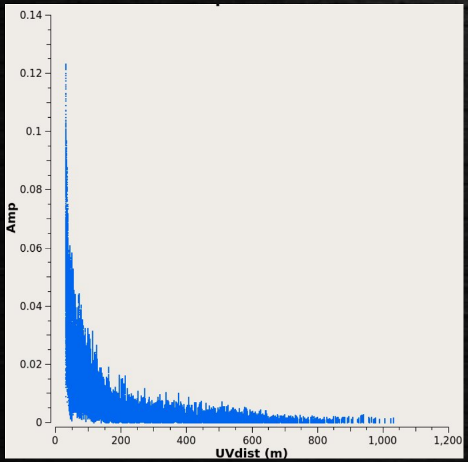
Subtraction of discrete sources

Discrete sources (mainly AGN) **contaminate** the extended emission

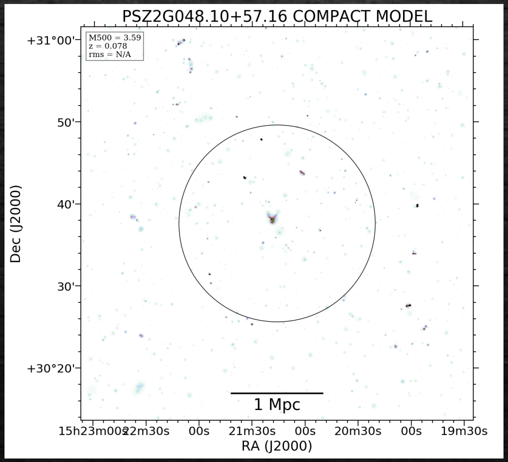
Image-plane subtraction: the contribution of the contaminating sources is subtracted a posteriori, in the *final image*. Fast but can be problematic (e.g. when your image has a low resolution)



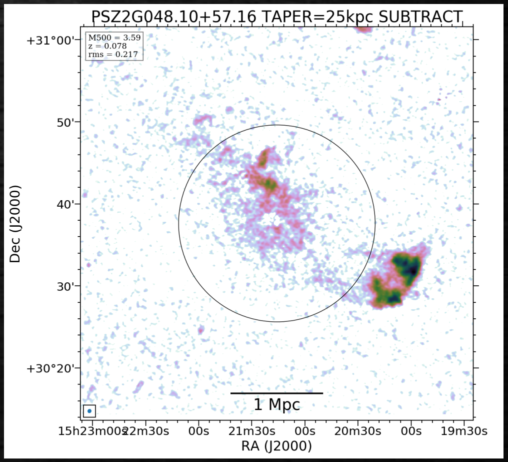
Uv-plane subtraction: clean components of discrete sources are directly subtracted from the *visibility data*. Powerful but more time consuming



-



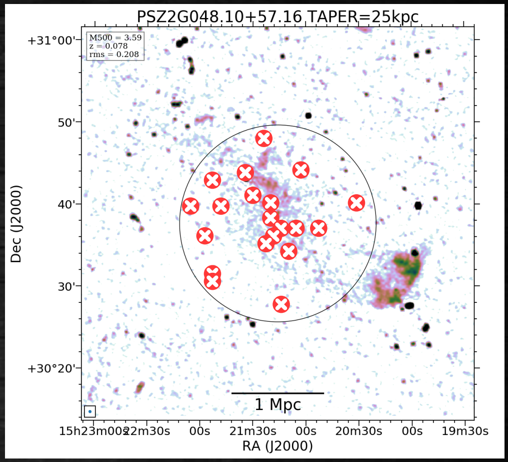
=



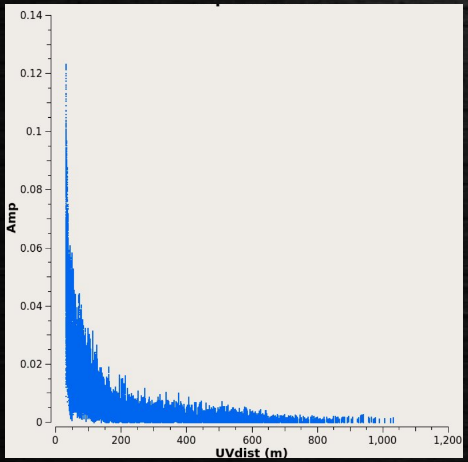
Subtraction of discrete sources

Discrete sources (mainly AGN) **contaminate** the extended emission

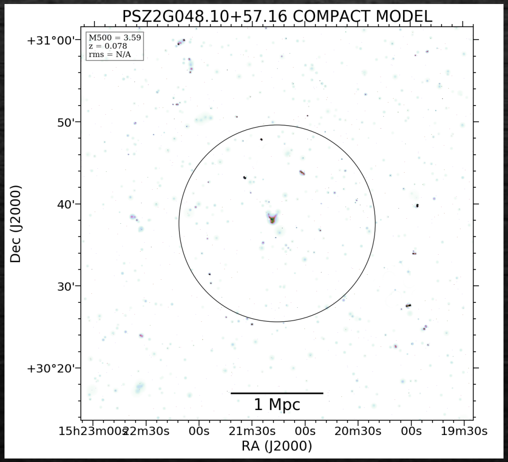
Image-plane subtraction: the contribution of the contaminating sources is subtracted a posteriori, in the *final image*. Fast but can be problematic (e.g. when your image has a low resolution)



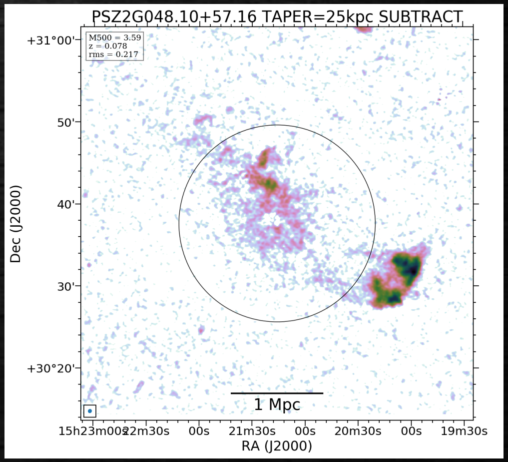
Uv-plane subtraction: clean components of discrete sources are directly subtracted from the *visibility data*. Powerful but more time consuming



-



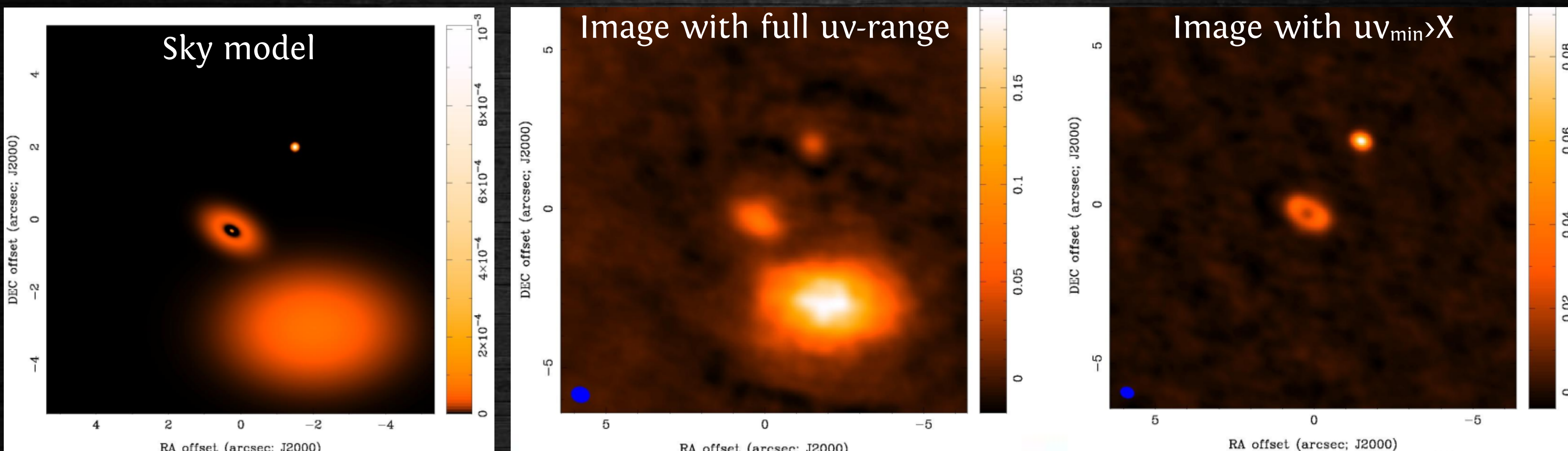
=



Uv-plane subtraction

How to make the model for the subtraction?

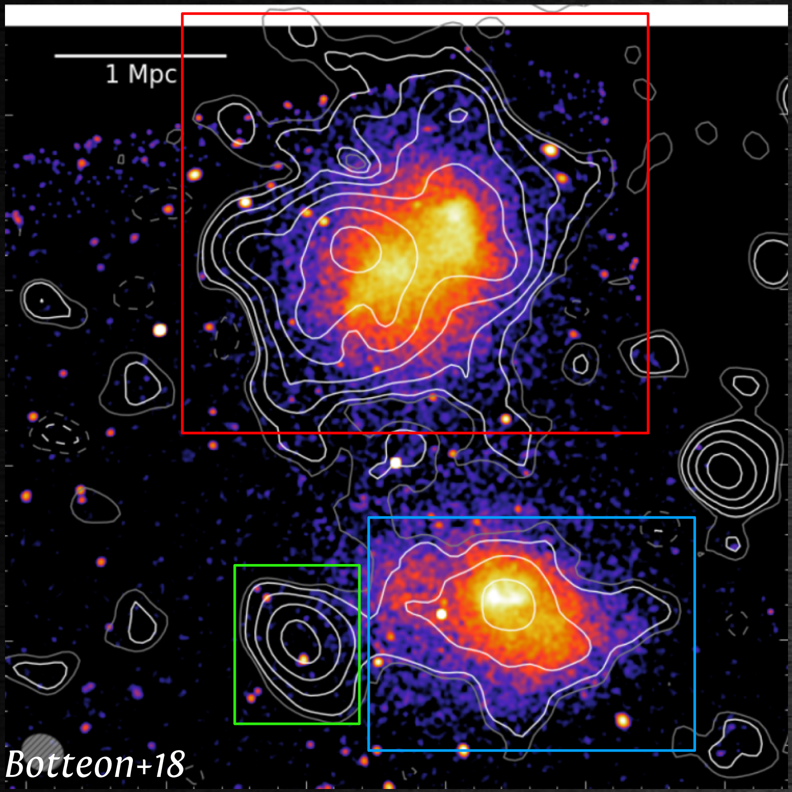
1) High resolution + high uv_{min} imaging $\quad \theta_{\rm max} = \lambda / B_{\rm min} = 206265 / uv_{\rm min}$ [arcsec]
2) Subtract from the visibility data the model obtained
3) Reimage the residual visibilities

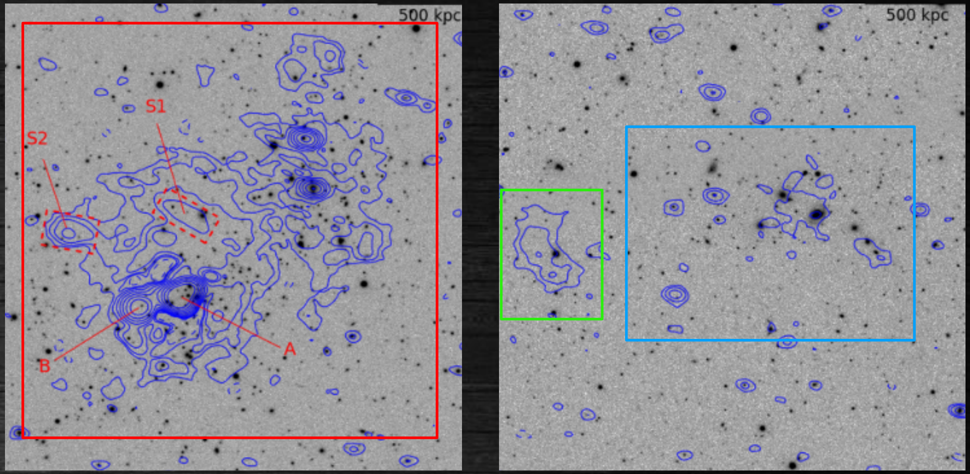


The extended emission disappeared because it can be recovered only by baselines *shorter than X*

→ we can use this to our advantage to *filter out* the extended emission and image *only* the discrete sources that we want to subtract

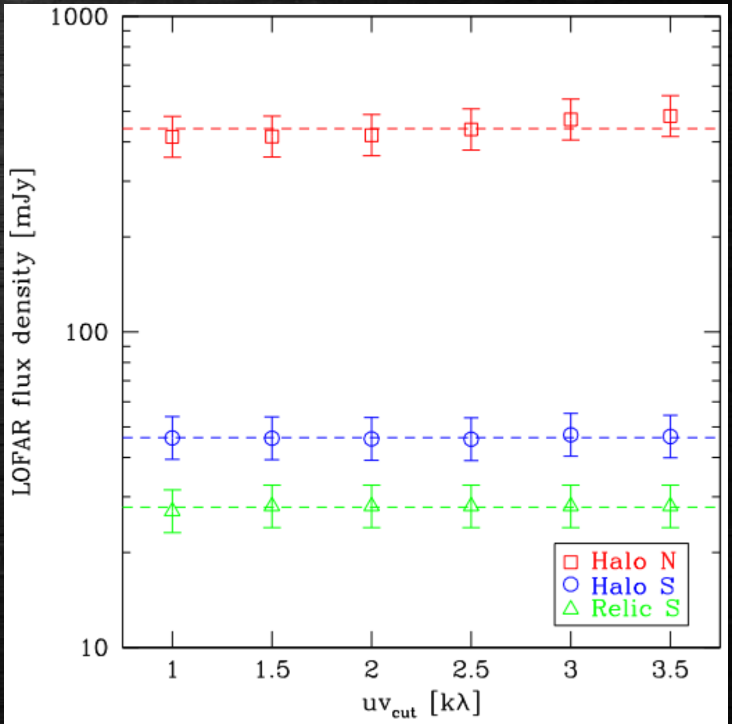
Choosing the uv_{min}





The flux density of Halo N depends on the uv_{min} adopted for the subtraction:

Low uv_{min}: better modeling of the extended emission from A and B
High uv_{min}: better filtering of the halo diffuse emission from the model

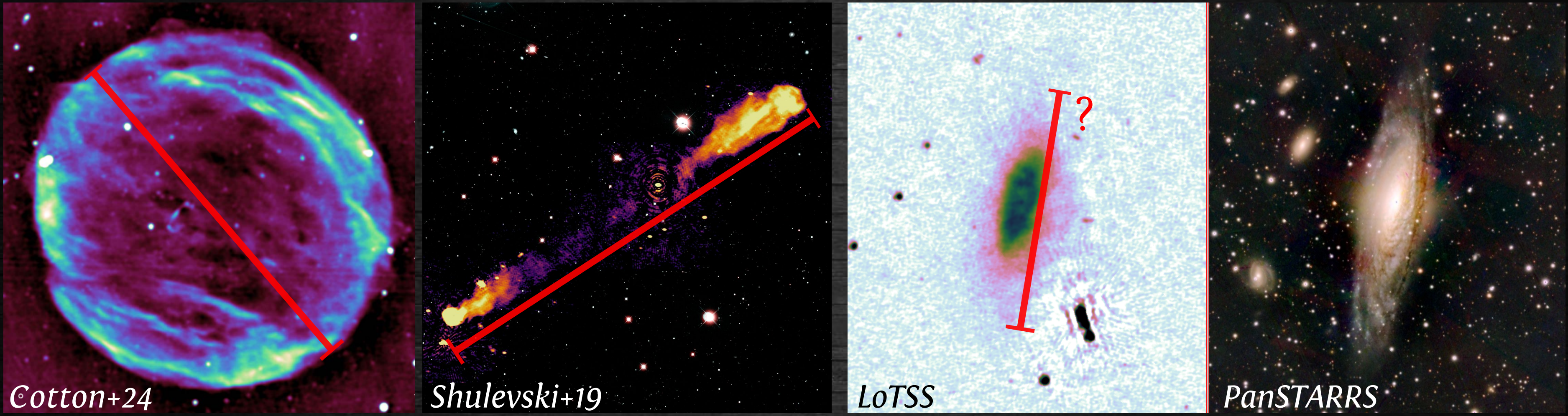


Properties of extended sources

Now that you have images at multiple resolutions, with/without discrete sources removed, you are ready to measure the **source properties**

- **Largest-angular size** (LAS) and **largest-linear size** (LLS):

→ probably the *easiest* property that you can derive…but pay attention that often diffuse sources do not have sharp boundaries!



- **Volume** (V):

→ a real problem in astronomy due to the missing 3rd dimension: usually *spherical/ellipsoidal/cylindrical* geometries are assumed

- **Flux density** (S_ν):

→ probably *the most relevant* property that you can derive: with S_ν you can derive the *power* (P_ν), with P_ν + V you can derive the *emissivity* (ε_ν)

Measuring the flux density

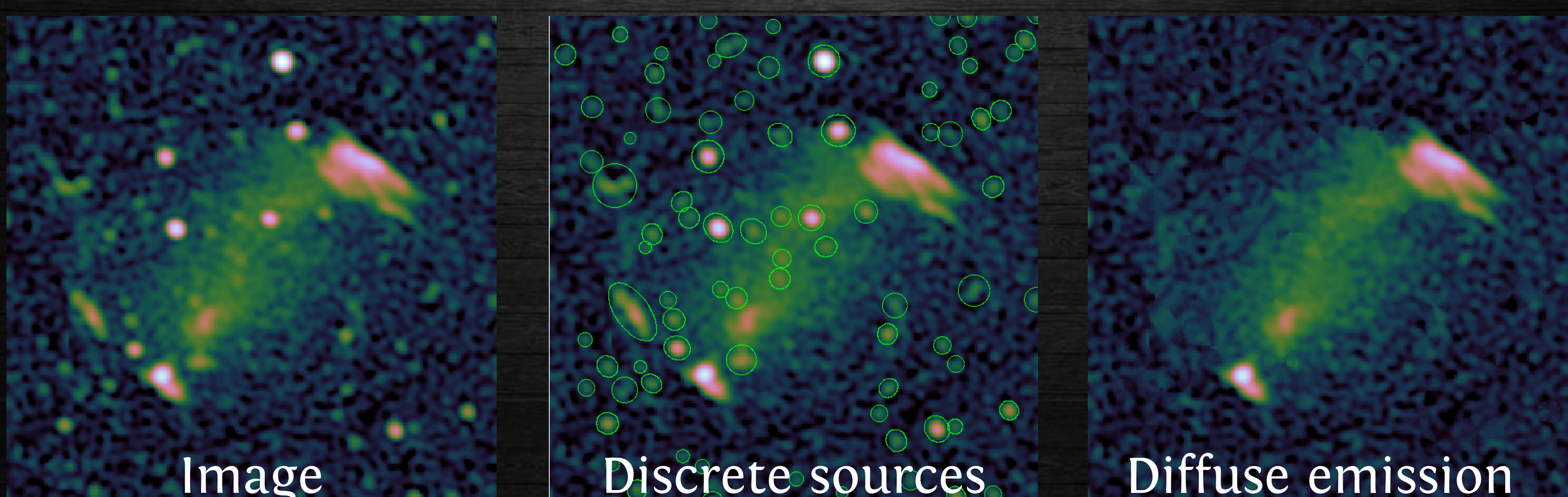
Different approaches:

1) Follow the 2 (or 3) sigma contour (polygon)
2) Adopt a circle/ellipse/polygon that follows the emission
3) Model the surface brightness profile of the source

Methods **#1** and **#2** use *regions files* and one of the following:

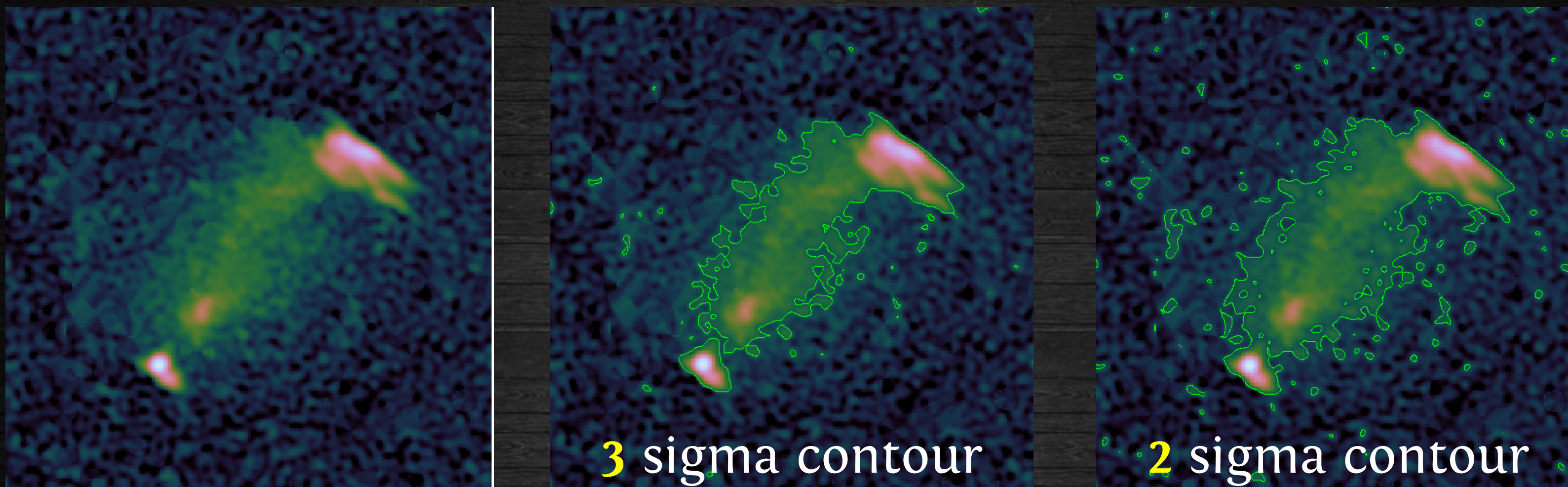
- ds9 + radioflux (https://github.com/mhardcastle/radioflux)
- casaviewer
- CARTA
- python/pyregion

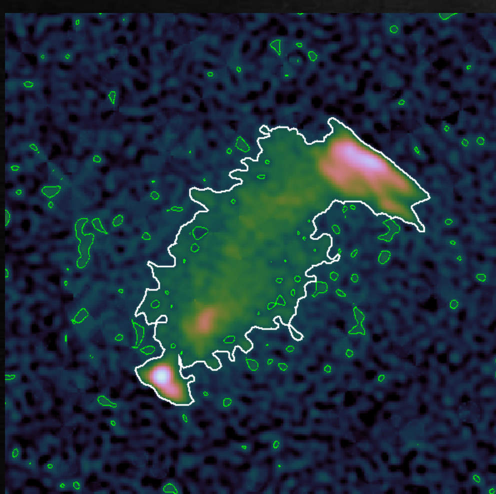
Method **#3** requires some *assumptions* on the source and probably some *coding* to perform the surface brightness profile *fitting*



Measuring the flux density

1) Follow the 2 (or 3) sigma contour (polygon)
2) Adopt a circle/ellipse/polygon that follows the emission
3) Model the surface brightness profile of the source





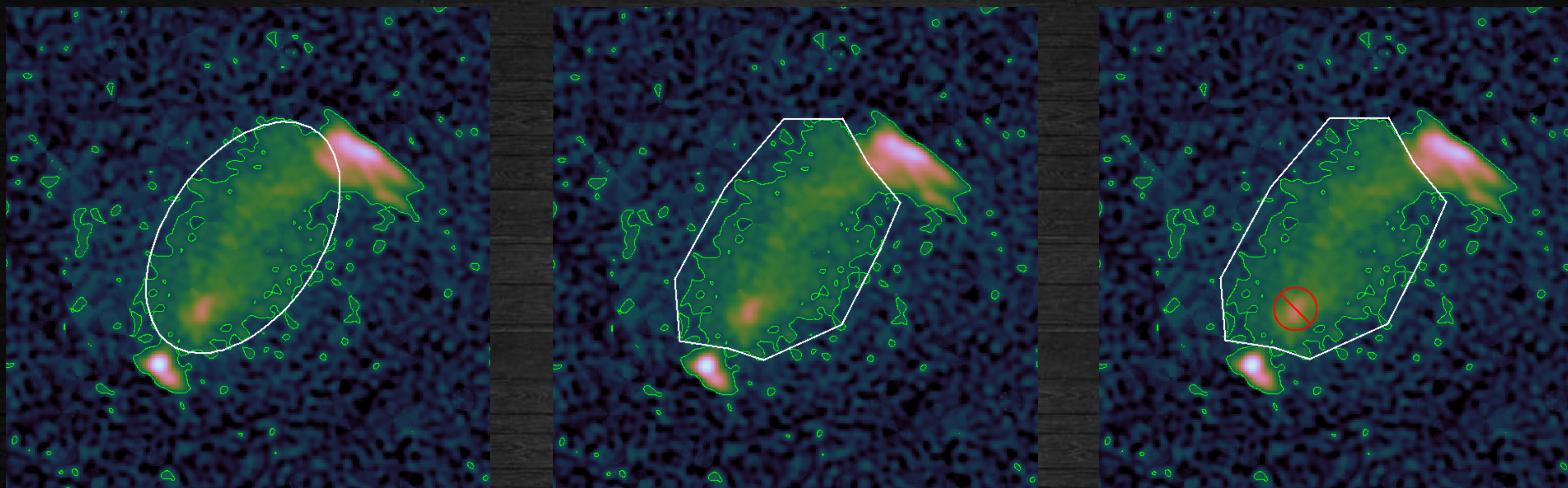
Strictly following a given contour is **not correct**:

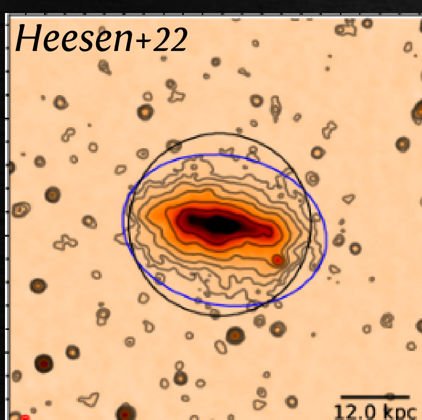
1. if you perform a deeper observation of the same target, sigma will decrease: a property of a source *should not depend on a parameter of the observation*!
2. do these fragmented borders make sense? No!

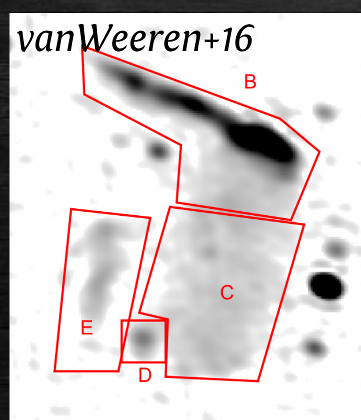
Remember that *often diffuse sources do not have sharp boundaries*

Measuring the flux density

1) Follow the 2 (or 3) sigma contour (polygon)
2) Adopt a circle/ellipse/polygon that follows the emission
3) Model the surface brightness profile of the source







It's a good practice to report the regions used

In the case of masking:

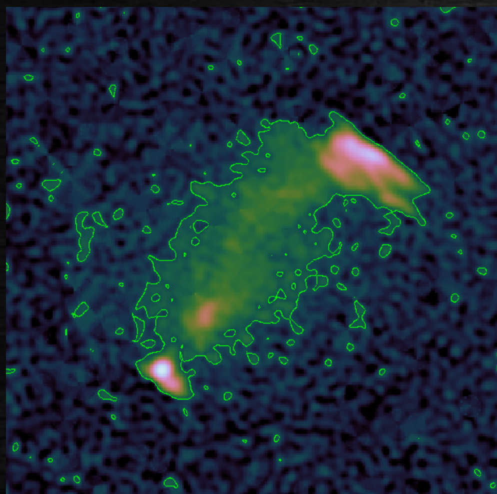
$$S_{\rm tot} = S_{\rm m} \times \frac{A_{\rm tot}}{A_{\rm m}}$$

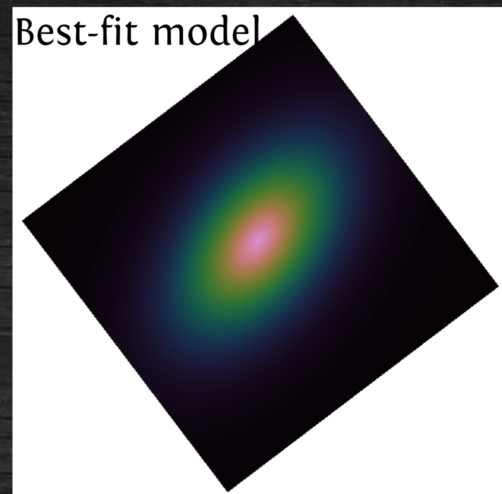
S_m is the flux density measured in the area A_m, where discrete sources were masked

S_{tot} is the total flux density measured in the area A_{tot}

Measuring the flux density

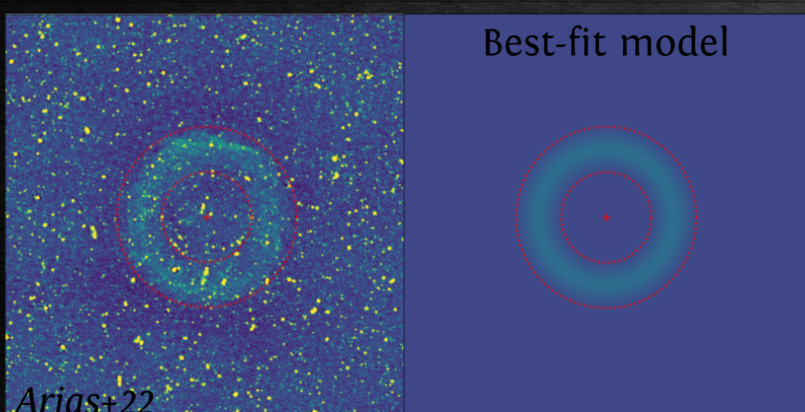
1) Follow the 2 (or 3) sigma contour (polygon)
2) Adopt a circle/ellipse/polygon that follows the emission
3) Model the surface brightness profile of the source



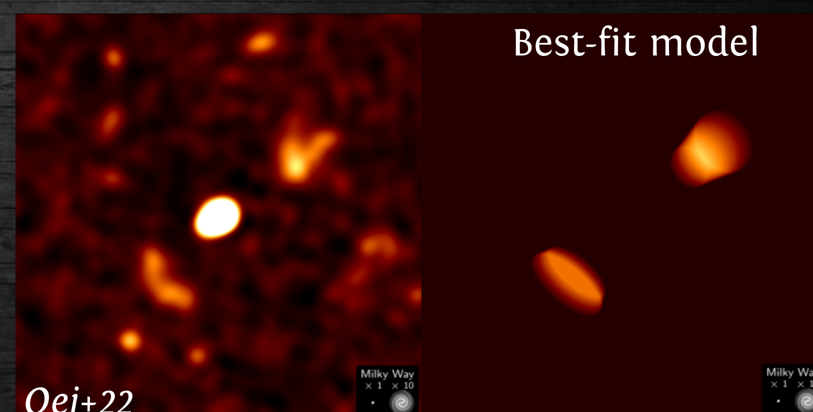


Typical models used are *exponentials* and *Gaussians* (**easy** to integrate!)

Useful to obtain *characteristic scales* and for *population comparison*



A supernova remnant as a...
Gaussian ring



The lobes of a radio galaxy as...
doubly truncated cones

Summary on S_ν

Always quote the method used to measure the flux density!

Often, measurements with *different methods*, as well as with *different approaches* used to subtract discrete sources, are reported

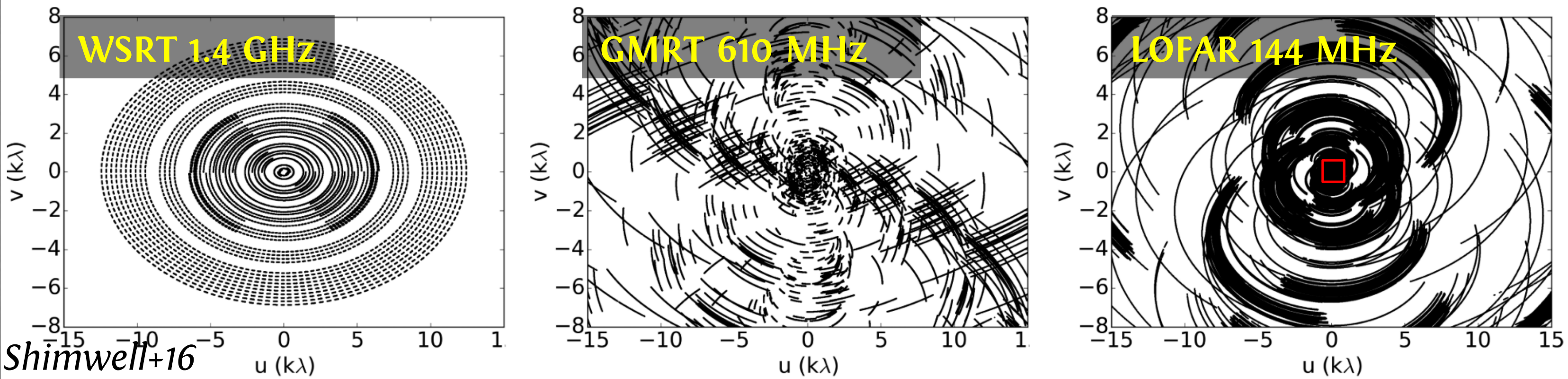
The *error on the flux density* of a diffuse source is due to different factors:

$$\Delta S_{\rm diffuse} = \sqrt{\sigma^2_{\rm stat} + \sigma^2_{\rm sys} + \sigma^2_{\rm sub}}$$
$$= \sqrt{\left(\sigma_{\rm rms} \times \sqrt{\frac{A_{\rm s}}{A_{\rm b}}}\right)^2 + (\delta_{\rm cal} \times S_{\rm diffuse})^2 + (\xi_{\rm res} \times S_{\rm discrete})^2}$$

σ_{rms} is the image noise
A_s is the area of the source
A_b is the area of the beam
δ_{cal} is the systematic uncertainty on the flux scale (10% for LOFAR)
ξ_{res} is the residual error on the source subtraction (see e.g. Botteon+22)

Combine LOFAR with other data

Spectral analysis of diffuse sources involving different interferometers needs caution because the **different uv-coverages** of the instruments

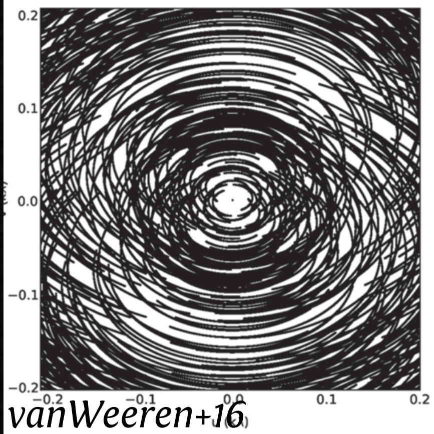


To make a **fair** comparison between different interferometers:

- *use a common uv_{min}*, where uv_{min} is the *shortest well-sampled baseline* by all instruments
- *compensate the different uv-sampling* by matching the uv-planes as closely as possible or by using a uniform weighting

Most of the times, this implies to **reduce** the capabilities of **LOFAR** to recover extended emission. If you do not do that, you risk to **bias** the spectral analysis, obtaining *steeper spectra* than what you should

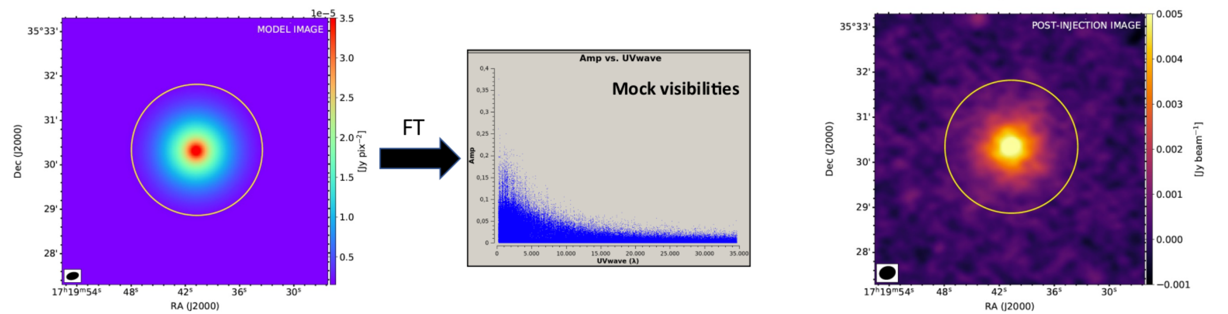
Recovering diffuse emission: a note

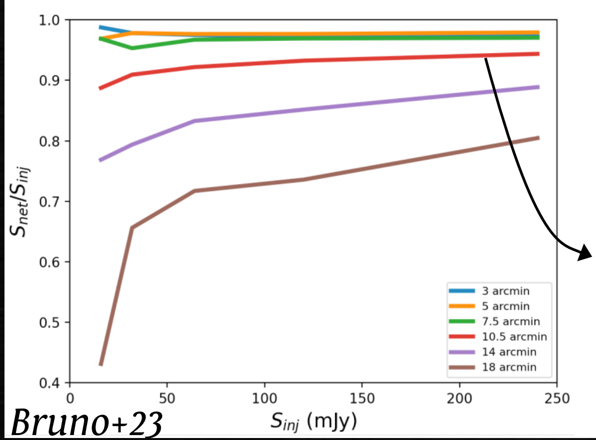


The uv-plane is *sparsely sampled* and has a *central "hole"* (missing short spacings)

→ you can inject *fake sources* in the visibilities data, with different sizes **D** and flux densities **S_{inj}**, to understand the **ability** of an interferometer to recover the *diffuse emission*

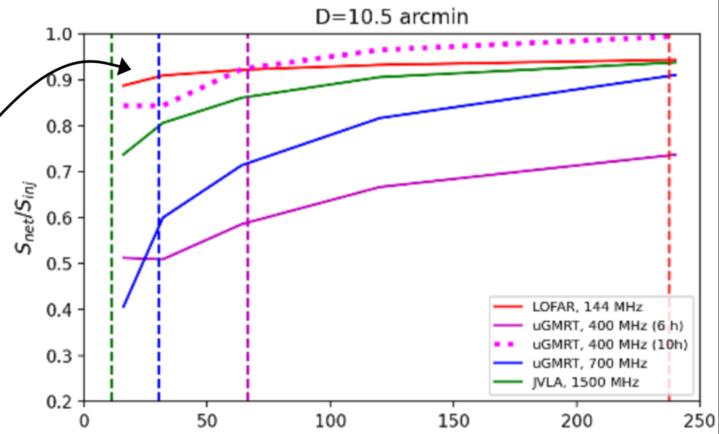
https://github.com/lucabruno2501/MUVIT/





Recovered flux density
(single 8-hr pointing)

Diameter	Recovered fraction
3'	> 95%
5'	> 95%
7.5'	> 95%
10.5'	~ 90 – 95%
14'	~ 80 – 90%
18'	~ 40 – 80%



LOFAR: *negligible* losses for sources of **D < 15'**!

LOFAR vs **uGMRT** vs **JVLA**

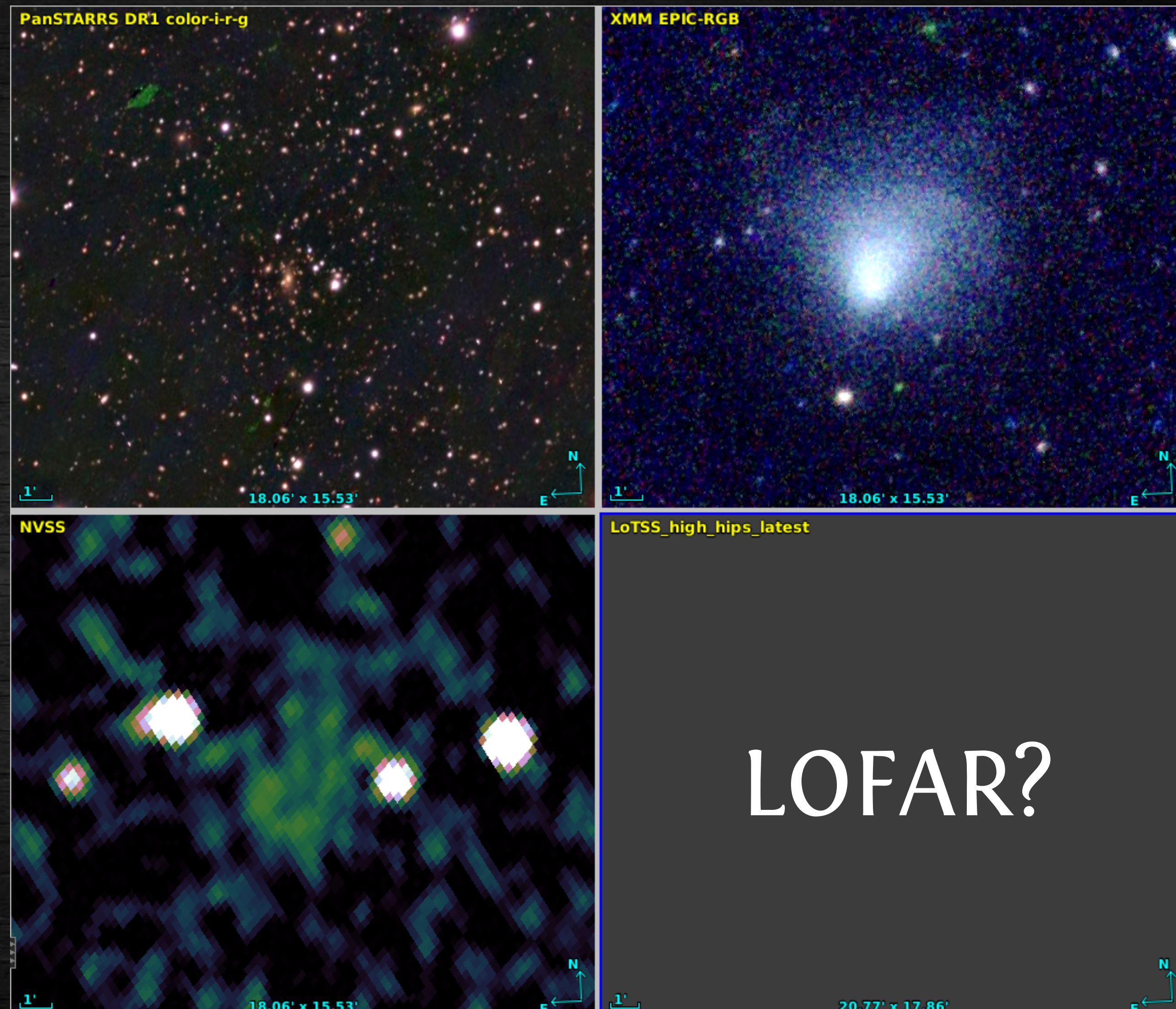
Tutorial

Our target: Abell 665

Abell 665 is a *galaxy cluster* at z=0.181 (**1” = 3.062 kpc**)

It hosts an extended source, a *radio halo*, which is detected also at 1.4 GHz by the **NVSS**

Abell 665 lays in the **LoTSS-DR2** footprint: how does **LOFAR** recover the cluster diffuse radio emission?



Setup the enviroment

```
tar xzvf A665.avg.tar.gz

apptainer shell flocs_v5.0.0_sandybridge_sandybridge_mkl_cuda.sif \
 --noprofile --norc
```

INFO: the FoV of the image is 20'x20'

```
This color will highlight the running time on my machine (12xCPU, 30GB RAM)
```

Basic clean

```
wsclean -no-update-model-required \
 -minuv-l 80.0 -size 1000 1000 -scale 1.2arcsec \
 -weight briggs -0.5 \
 -mgain 0.8 -data-column DATA \
 -join-channels -channels-out 6 \
 -fit-spectral-pol 3 \
 -auto-mask 2.5 -auto-threshold 0.5 \
 -niter 60000 \
 -baseline-averaging 10 \
 -name A665_ROBUST-0.5 \
 A665.ms.avg

real     0m53.005s
```

Check image and residual.

```
ds9 A665_ROBUST-0.5-MFS-image.fits A665_ROBUST-0.5-MFS-residual.fits -scale log -scale limits -1e-4 5e-2 -cmap cubehelix0 -lock frame wcs -lock scale yes -lock colorbar yes
```

Do you notice the issue?

Add multiscale

```
wsclean -no-update-model-required \
 -minuv-l 80.0 -size 1000 1000 -scale 1.2arcsec \
 -weight briggs -0.5 \
 -mgain 0.8 -data-column DATA \
 -join-channels -channels-out 6 \
 -fit-spectral-pol 3 \
 -auto-mask 2.5 -auto-threshold 0.5 \
 -niter 60000 \
 -baseline-averaging 10 \
 -name A665_ROBUST-0.5ms \
 -multiscale \
 A665.ms.avg

real     1m21.486s
```

Check image and residual

```
ds9 A665_ROBUST-0.5ms-MFS-image.fits A665_ROBUST-0.5ms-MFS-residual.fits -scale log -scale limits -1e-4 5e-2 -cmap cubehelix0 -lock frame wcs -lock scale yes -lock colorbar yes
```

Compare previous model with present model

```
ds9 A665_ROBUST-0.5-MFS-model.fits A665_ROBUST-0.5ms-MFS-model.fits -scale linear -scale limits 0 5e-5 -cmap cubehelix0 -lock frame wcs -lock scale yes -lock colorbar yes
```

Taper!

```
wsclean -no-update-model-required \
 -minuv-l 80.0 -size 300 300 -scale 4.0arcsec \
 -weight briggs -0.5 \
 -mgain 0.8 -data-column DATA \
 -join-channels -channels-out 6 \
 -fit-spectral-pol 3 \
 -auto-mask 2.5 -auto-threshold 0.5 \
 -niter 60000 \
 -baseline-averaging 10 \
 -name A665_ROBUST-0.5msTAPER15 \
 -multiscale \
 -taper-gaussian 15.0 \
 A665.ms.avg

real	0m34.552s
```

Check image and residual

```
ds9 A665_ROBUST-0.5msTAPER15-MFS-image.fits A665_ROBUST-0.5msTAPER15-MFS-residual.fits -scale log -scale limits -1e-4 5e-2 -cmap cubehelix0 -lock frame wcs -lock scale yes -lock colorbar yes
```

Compare previous image with present image, what are the rms?

```
ds9 A665_ROBUST-0.5ms-MFS-image.fits A665_ROBUST-0.5msTAPER15-MFS-image.fits -scale log -scale limits -1e-4 5e-2 -cmap cubehelix0 -lock frame wcs -lock scale yes -lock colorbar yes
```

Source subtraction

Discrete sources contaminate the diffuse emission: we want to subtract them

Let's make an image containing only discrete sources, then we subtract the corresponding model from the visibilities.

To make a image containing only discrete sources we need:
1) high resolution
2) filter out the diffuse emission

Make discrete sources model

```
wsclean -no-update-model-required \
 -minuv-l 4000.0 -size 1200 1200 -scale 1.0arcsec \
 -weight briggs -1.0 \
 -mgain 0.8 -data-column DATA \
 -join-channels -channels-out 6 \
 -fit-spectral-pol 3 \
 -auto-mask 2.5 -auto-threshold 0.5 \
 -niter 60000 \
 -baseline-averaging 10 \
 -name A665_ROBUST-1.0uvmin4kl \
 A665.ms.avg

real     1m21.486s
```

Check image and residual

```
ds9 A665_ROBUST-1.0uvmin4kl-MFS-image.fits A665_ROBUST-1.0uvmin4kl-MFS-residual.fits -scale log -scale limits -1e-4 5e-2 -cmap cubehelix0 -lock frame wcs -lock scale yes -lock colorbar yes
```

What does the model tell us?

```
ds9 A665_ROBUST-1.0uvmin4kl-MFS-model.fits -scale linear -scale limits 0 5e-5 -cmap cubehelix0
```

Guide the clean with a mask

```
MakeMask.py --RestoredIm=A665_ROBUST-1.0uvmin4kl-MFS-image.fits \
 --Th=3.0 \
 --Box=100,2
```

Check the generated mask

```
ds9 A665_ROBUST-1.0uvmin4kl-MFS-image.fits A665_ROBUST-1.0uvmin4kl-MFS-
image.fits.mask.fits -scale log -scale limits -1e-4 5e-2 -cmap cubehelix0 -lock
frame wcs -lock scale yes -lock colorbar yes
```

Clean with mask

```
wsclean -no-update-model-required \
 -minuv-l 4000.0 -size 1200 1200 -scale 1.0arcsec \
 -weight briggs -1.0 \
 -mgain 0.8 -data-column DATA \
 -join-channels -channels-out 6 \
 -fit-spectral-pol 3 \
 -auto-mask 2.5 -auto-threshold 0.5 \
 -niter 60000 \
 -baseline-averaging 10 \
 -name A665_maskROBUST-1.0uvmin4kl \
 -fits-mask A665_ROBUST-1.0uvmin4kl-MFS-image.fits.mask.fits \
 A665.ms.avg

real	0m4.592s
```

Model prediction

Compare previous model with present model

```
ds9 A665_ROBUST-1.0uvmin4kl-MFS-model.fits A665_maskROBUST-1.0uvmin4kl-MFS-
model.fits -scale linear -scale limits 0 5e-5 -cmap cubehelix0 -lock frame wcs
-lock scale yes -lock colorbar yes
```

Satisfied? How could it be improved further?

Fill the MODEL_DATA column

```
wsclean -predict \
 -channels-out 6 \
 -name A665_maskROBUST-1.0uvmin4kl \
 A665.ms.avg

real     0m21.160s
```

Now we will make a new column "DIFFUSE_SUB", by subtracting the MODEL_DATA column to the DATA column

Subtraction

In a ipython shell:

```
import casacore.tables as pt

ms = 'A665.ms.avg'
outcolumn = 'DIFFUSE_SUB'

ts  = pt.table(ms, readonly=False)
colnames = ts.colnames()

desc = ts.getcoldesc('DATA')
desc['name']=outcolumn
ts.addcols(desc)

data = ts.getcol('DATA')
model = ts.getcol('MODEL_DATA')
ts.putcol(outcolumn,data-model)
ts.close()

real    0m24.526s
```

Imaging only the diffuse emission 1/2

```
wsclean -no-update-model-required \
 -minuv-l 80.0 -size 1000 1000 -scale 1.2arcsec \
 -weight briggs -0.5 \
 -mgain 0.8 -data-column DIFFUSE_SUB \
 -join-channels -channels-out 6 \
 -fit-spectral-pol 3 \
 -auto-mask 2.5 -auto-threshold 0.5 \
 -niter 60000 \
 -baseline-averaging 10 \
 -name A665_subROBUST-0.5ms \
 -multiscale \
 A665.ms.avg

real     1m4.461s
```

Compare with non-subtracted image

```
ds9 A665_ROBUST-0.5ms-MFS-image.fits A665_subROBUST-0.5ms-MFS-image.fits -scale log -scale limits -1e-4 5e-2 -cmap cubehelix0 -lock frame wcs -lock scale yes -lock colorbar yes
```

Imaging only the diffuse emission 2/2

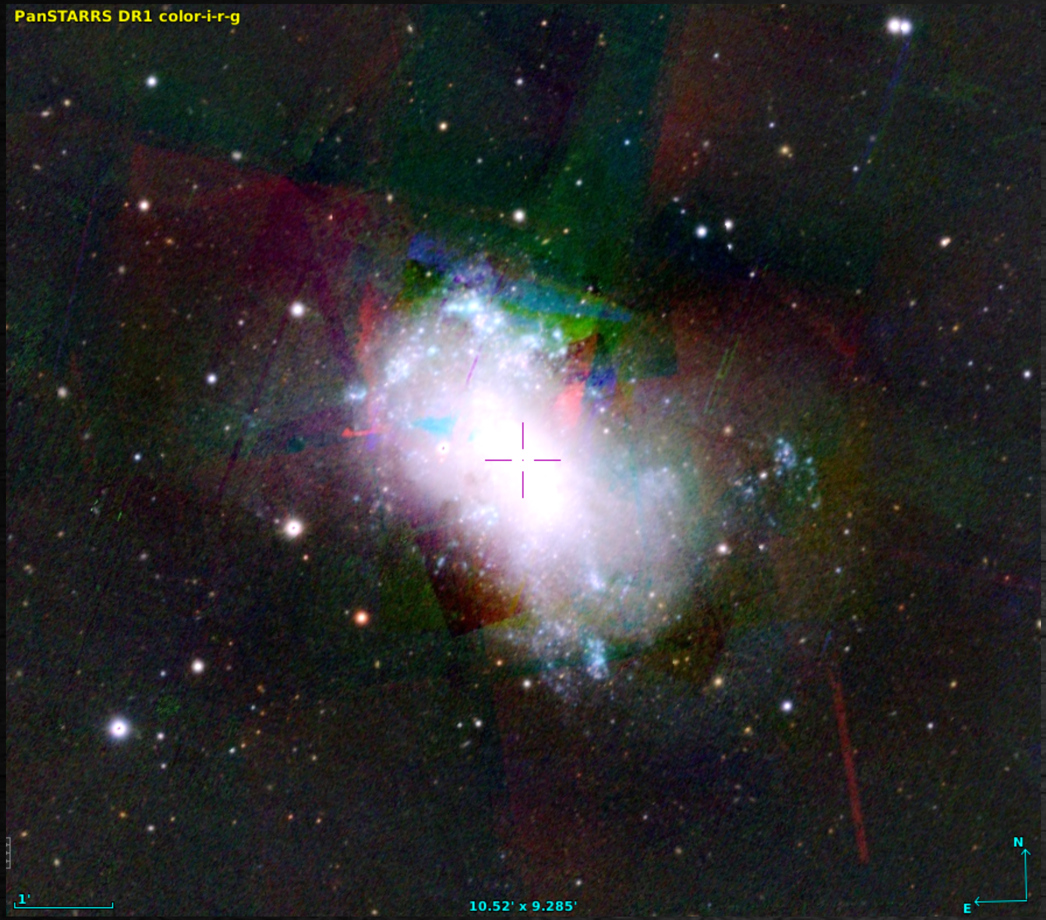
```
wsclean -no-update-model-required \
 -minuv-l 80.0 -size 300 300 -scale 4.0arcsec \
 -weight briggs -0.5 \
 -mgain 0.8 -data-column DIFFUSE_SUB \
 -join-channels -channels-out 6 \
 -fit-spectral-pol 3 \
 -auto-mask 2.5 -auto-threshold 0.5 \
 -niter 60000 \
 -baseline-averaging 10 \
 -name A665_subROBUST-0.5msTAPER15 \
 -multiscale \
 -taper-gaussian 15.0 \
 A665.ms.avg

real	1m4.461s
```

Compare with non-subtracted image

```
ds9 A665_ROBUST-0.5msTAPER15-MFS-image.fits A665_subROBUST-0.5msTAPER15-MFS-image.fits -scale log -scale limits -1e-4 5e-2 -cmap cubehelix0 -lock frame wcs -lock scale yes -lock colorbar yes
```

Now you are ready to measure LAS, LLS, S_ν, P_ν, ε_ν, etc.

Science contest: NGC 4449



An irregular galaxy at D_L = 4.02 Mpc

1. Produce images at different resolutions (FoV of the image is 20'x20')
2. Subtract emission of discrete sources
3. Measure properties of the diffuse emission (angular/linear sizes, flux density, power and, if you are brave, the emissivity

For point 2: explore different -minuv-l and the usage of the multiscale in the creation of the model of the discrete sources to subtract. How does this impact the final measurement?
For point 3: use the different approaches discussed during the lecture to measure the flux density. What are the differences of the various methods (if any)?