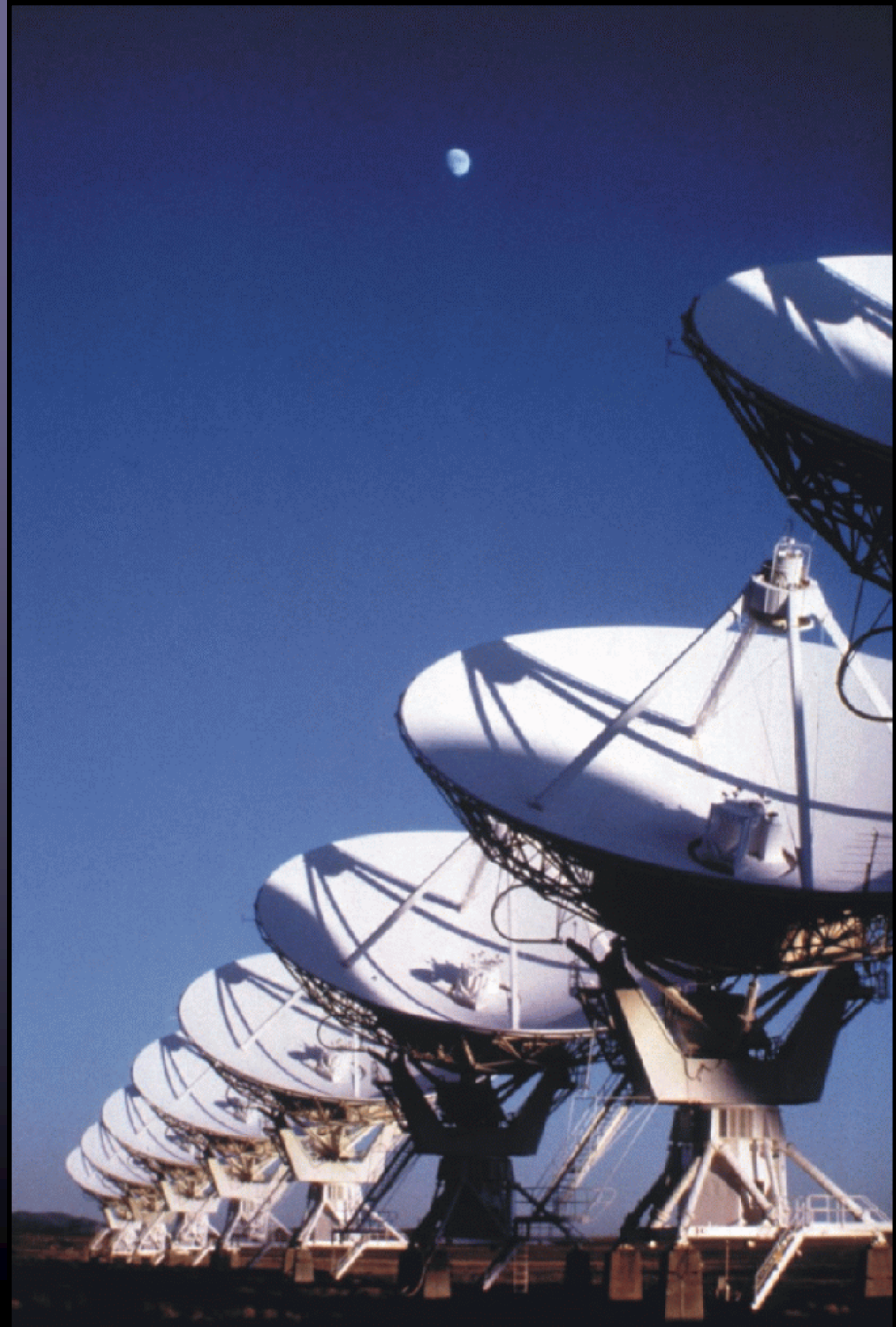
Error Recognition and Data Analysis

Greg Taylor

University of New Mexico

ASTR 423



INTRODUCTION

- Why are these two topics – ‘Error Recognition’ and ‘Data Analysis’ in the same lecture?

  -- Error recognition is used to determine defects in the (visibility) data and image during and after the ‘best’ calibration, editing, etc.

  -- Image analysis describes the many ways in which useful insight, information and parameters can be extracted from the image.

  -- non-imaging analysis describes how to extract information directly from the (u,v) data

- Perhaps these topics are related to the reaction one has when looking at an image after ‘good’ calibration, editing, self-calibration, etc.

- If the reaction is:

Fantastic Discovery or an Obvious Defect?

VLBA observations of SgrA* at 43 GHz

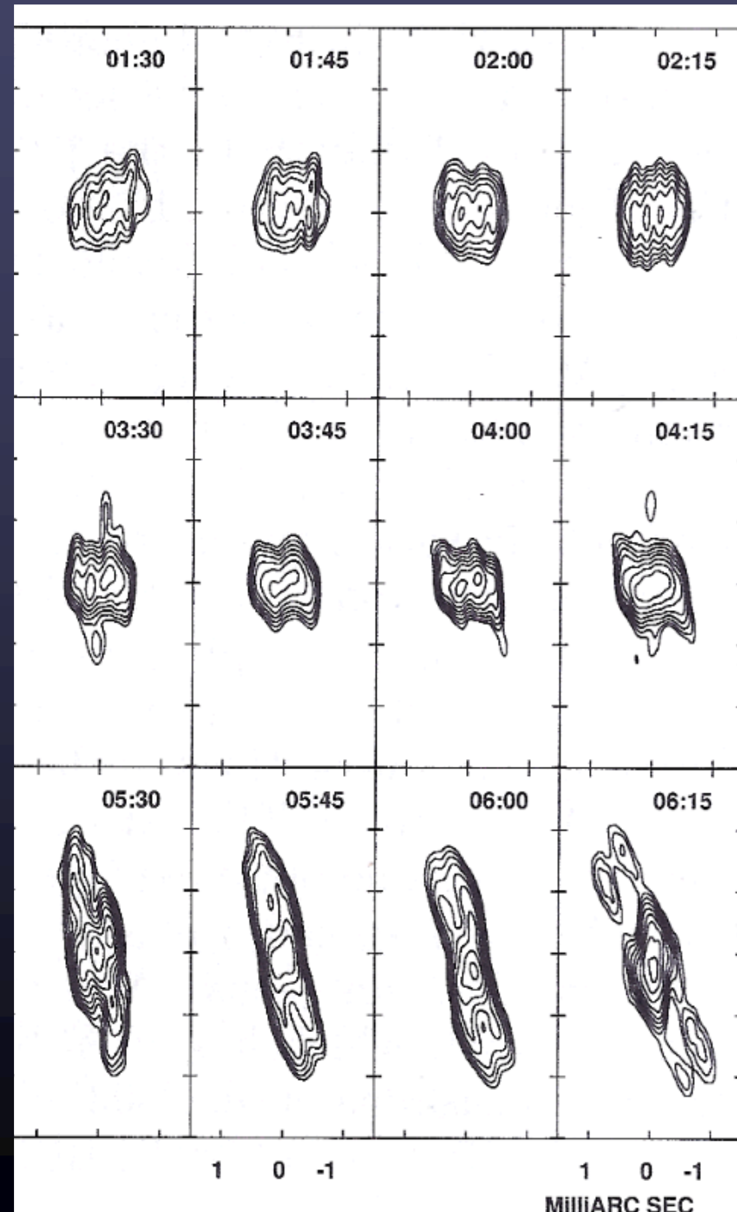
This can't be right. Either SgrA* has bidirectional jets that nobody else has ever seen or:

Clear signs of problems:

Image rms > expected rms

Unnatural features in the image

How can the problems be found and corrected?



Miyoshi et al. 2005

milliarcsec

HIGH QUALITY IMAGES

Reality

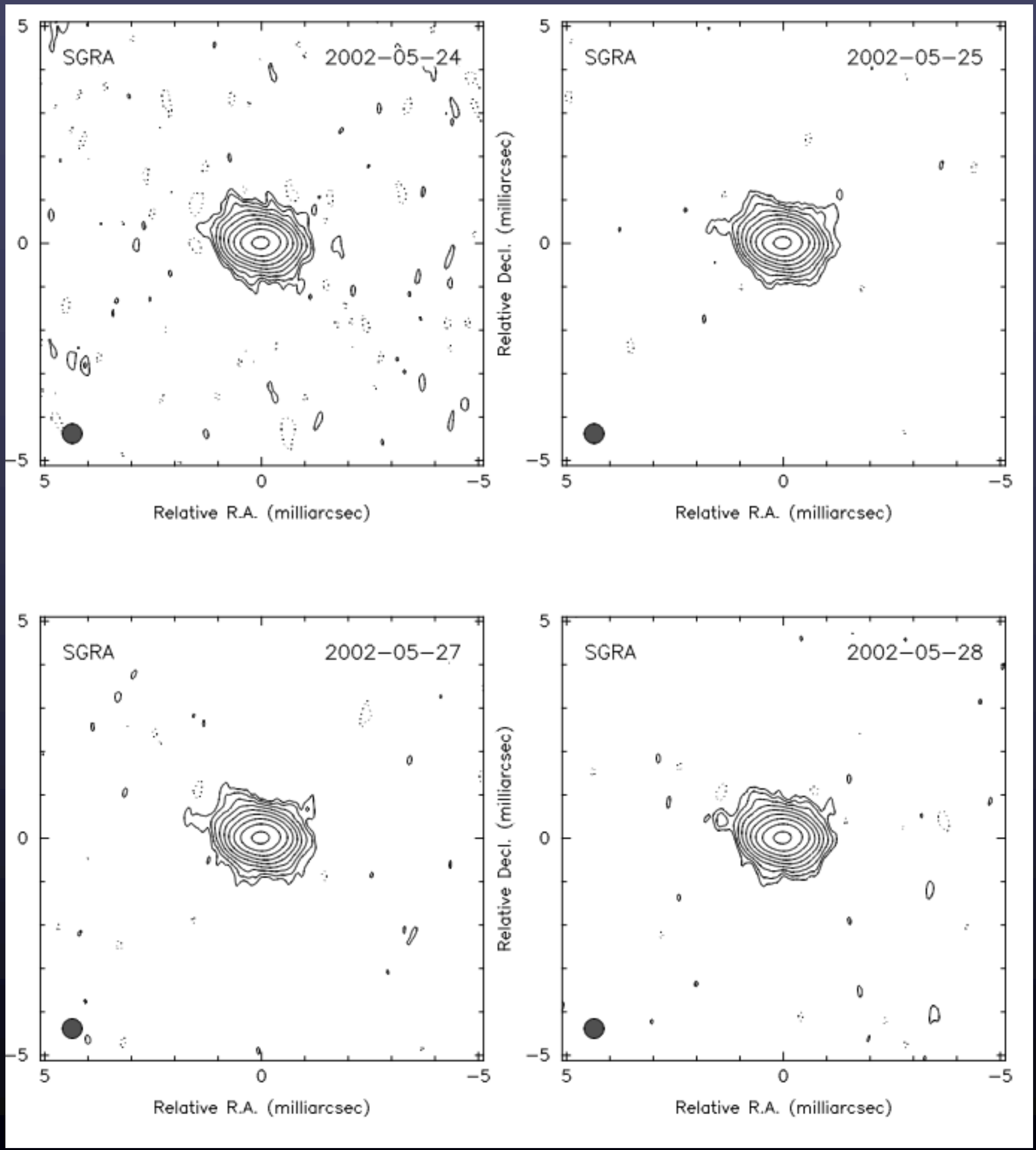
With care we can obtain good images.

What were defects?

Two antennas had ~30% calibration errors at low elevations.

This part of the lecture.

How to find the errors and remove them.



milliarcsec

GENERAL PROCEDURE

We assume that the data have been edited and calibrated reasonably successfully (earlier lectures) including self-calibration if necessary.

So, the first serious display of an image leads one—

- to inspect again and clean-up the data repeating some or all of the previous reduction steps.
  - removal of one type of problem can reveal next problem!
- once all is well, proceed to image-analysis and obtaining scientific results from the image.

But, first a digression on data and image display. First: Images

IMAGE DISPLAYS (1)

Pixel values

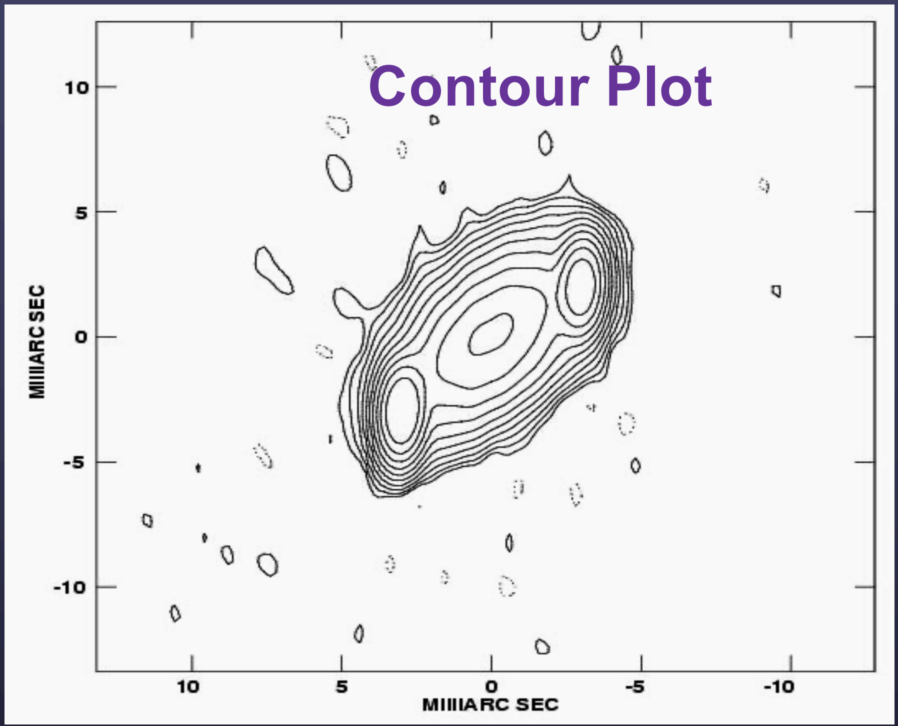
					235					245					255					265					275				
287	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
285	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
283	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0
281	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	2	3	3	3	4	3	1	0	0	0	0
279	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	3	4	4	5	5	9	12	8	3	1	0	0	0
277	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	5	7	7	8	9	9	19	32	22	6	1	0	0	0
275	0	0	0	0	0	0	0	0	0	0	0	1	2	4	6	9	13	14	15	14	16	40	72	47	12	2	0	0	0
273	0	0	0	0	0	0	0	0	1	1	1	2	4	8	12	17	22	23	24	22	27	77	136	87	19	2	0	0	0
271	0	0	0	0	0	0	0	0	1	1	2	4	8	15	21	28	35	36	37	33	43	126	217	132	28	3	0	0	0
269	0	0	0	0	0	0	0	0	1	3	4	8	15	25	34	44	54	54	53	48	61	173	288	168	34	3	0	0	0
267	0	0	0	0	0	0	0	1	2	4	8	14	25	40	52	67	79	77	74	63	75	199	316	177	34	3	0	0	0
265	0	0	0	0	0	0	0	1	3	7	14	24	40	60	77	97	109	102	93	74	79	191	289	155	29	3	0	0	0
263	0	0	0	0	0	0	1	2	5	11	22	37	59	86	108	130	137	123	105	79	73	154	220	113	20	2	0	0	0
261	0	0	0	0	0	1	1	3	8	17	33	54	81	116	139	156	156	133	107	75	61	106	140	69	12	2	0	0	0
259	0	0	0	0	0	1	2	5	12	24	45	72	105	143	162	170	161	131	99	66	47	64	75	36	6	1	0	0	0
257	0	0	0	0	0	2	4	8	19	32	58	88	124	160	171	169	152	118	86	55	36	36	36	16	3	1	0	0	0
255	0	0	0	0	1	2	7	16	27	42	70	101	135	162	164	156	134	100	71	44	27	20	16	7	1	0	0	0	0
253	0	0	0	0	1	4	15	34	43	51	77	105	133	150	146	135	112	81	56	34	19	11	7	3	1	0	0	0	0
251	0	0	0	0	1	8	34	73	70	59	79	100	120	130	122	110	88	61	41	24	12	6	3	1	0	0	0	0	0
249	0	0	0	1	2	14	69	141	112	65	73	87	100	106	96	83	64	43	27	14	7	3	1	1	0	0	0	0	0
247	0	0	0	1	3	23	121	238	167	69	62	69	77	81	70	59	42	26	16	8	3	1	0	0	0	0	0	0	0
245	0	0	0	1	3	34	180	338	217	69	48	52	56	57	47	36	25	15	8	3	1	0	0	0	0	0	0	0	0
243	0	0	0	1	4	42	222	402	242	65	36	37	38	37	29	21	14	7	4	1	0	0	0	0	0	0	0	0	0
241	0	0	0	1	4	44	229	398	228	56	26	25	25	22	16	11	7	3	1	0	0	0	0	0	0	0	0	0	0
239	0	0	0	1	3	39	196	327	179	41	18	16	15	12	8	5	3	1	1	0	0	0	0	0	0	0	0	0	0
237	0	0	0	1	3	28	139	223	118	26	11	9	8	6	4	2	1	1	0	0	0	0	0	0	0	0	0	0	0
235	0	0	0	0	2	18	82	127	64	14	6	5	4	3	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
233	0	0	0	0	1	9	40	60	29	7	3	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
231	0	0	0	0	0	4	17	23	11	3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
229	0	0	0	0	0	2	6	7	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
227	0	0	0	0	0	1	2	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
225	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
223	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

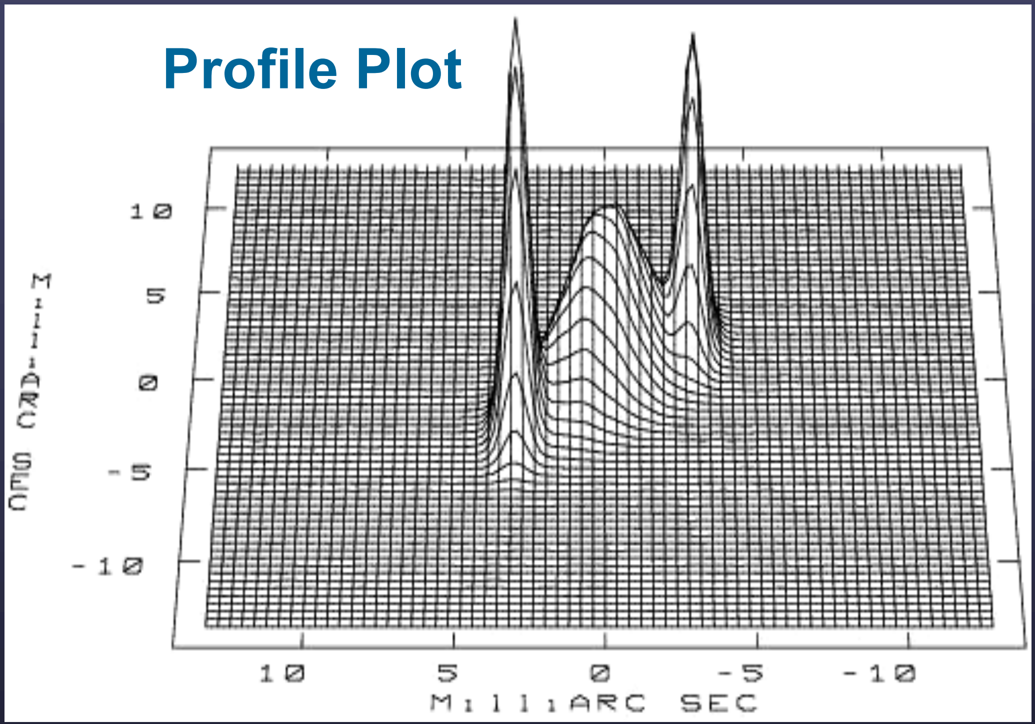
Digital image

Numbers are proportional to the intensity

Old School

IMAGE DISPLAYS (2)



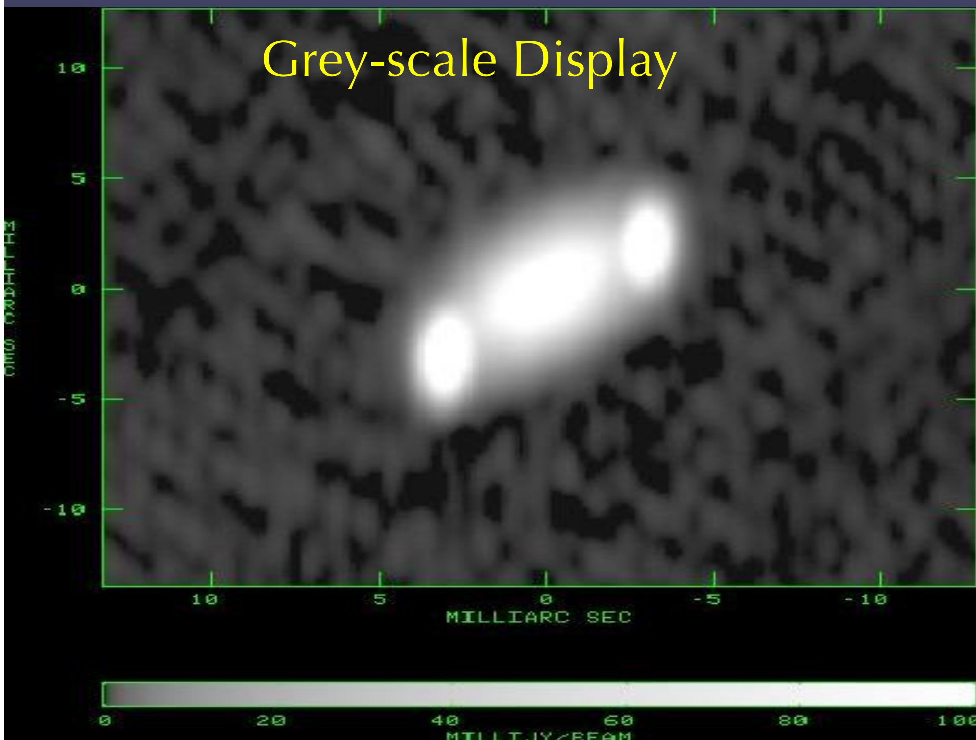


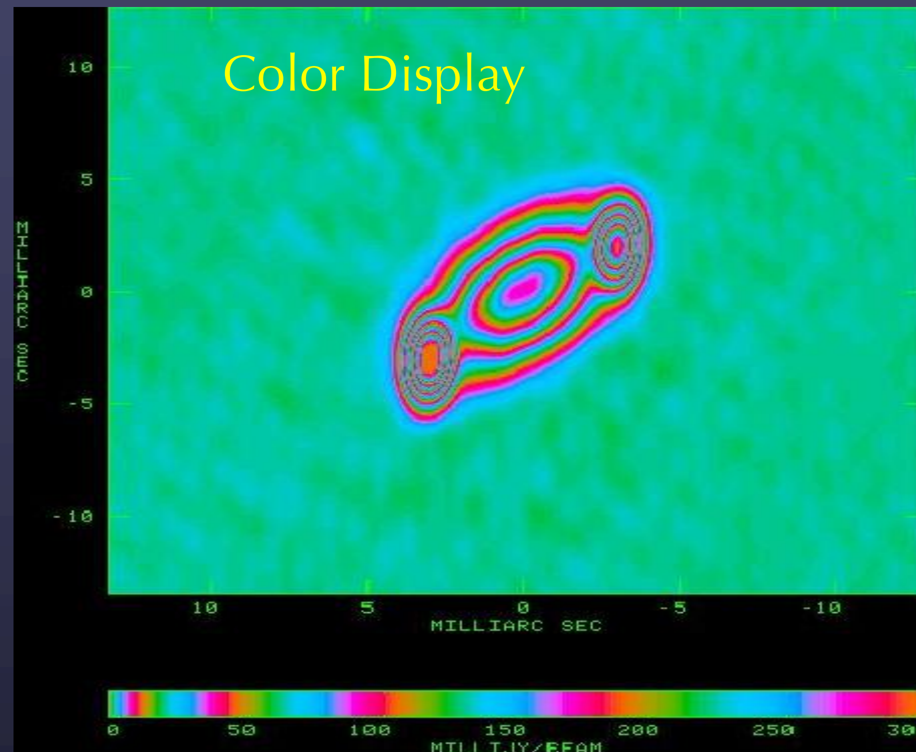
These plots are easy to reproduce and print

Contour plots give good representation of faint emission.

Profile plots give a good representation of the bright emission.

IMAGE DISPLAYS (3)





TV-based displays are most useful and interactive:

Grey-scale shows faint structure, but not great for high dynamic range and somewhat unbiased view of source

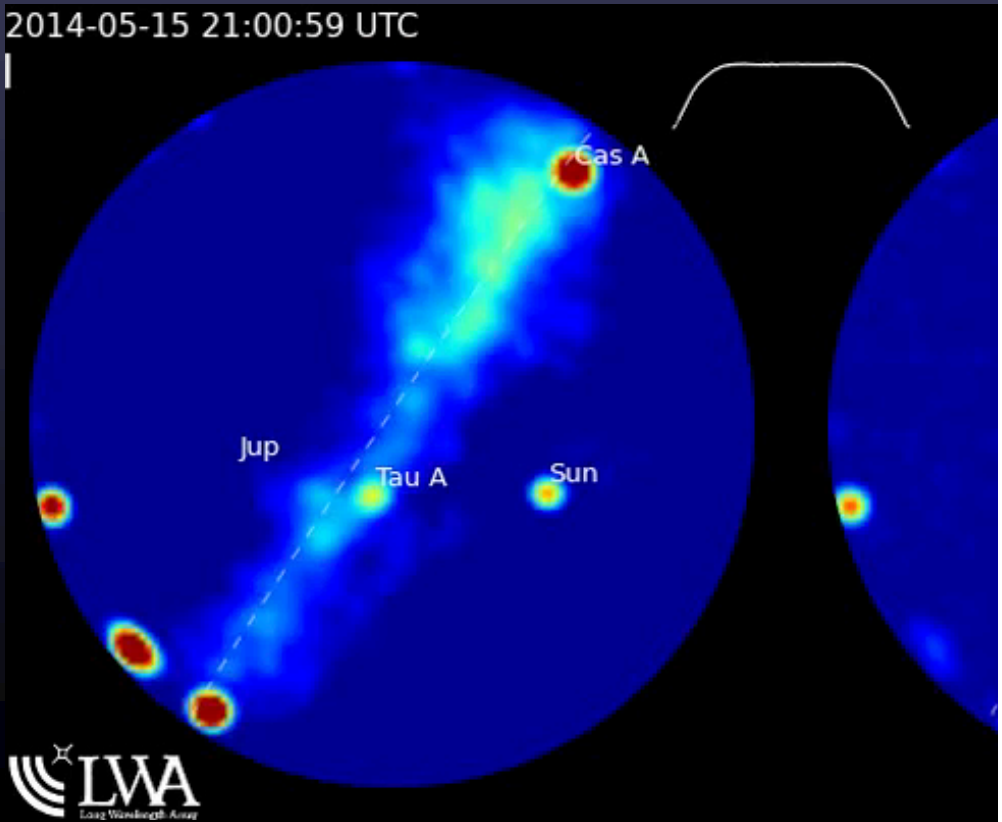
Color displays more flexible; e.g. pseudo contours

Movies and Radio Frequency Interference (RFI)

Great pressure from wireless devices (especially smartphones)
Auction in 2015 of 40 MHz raised 44 billion $$$!
Next up is 580 – 700 MHz, plus 5G networks at ~60 GHz
Dynamic allocation/Shared use of spectrum
Passive use is still useful!

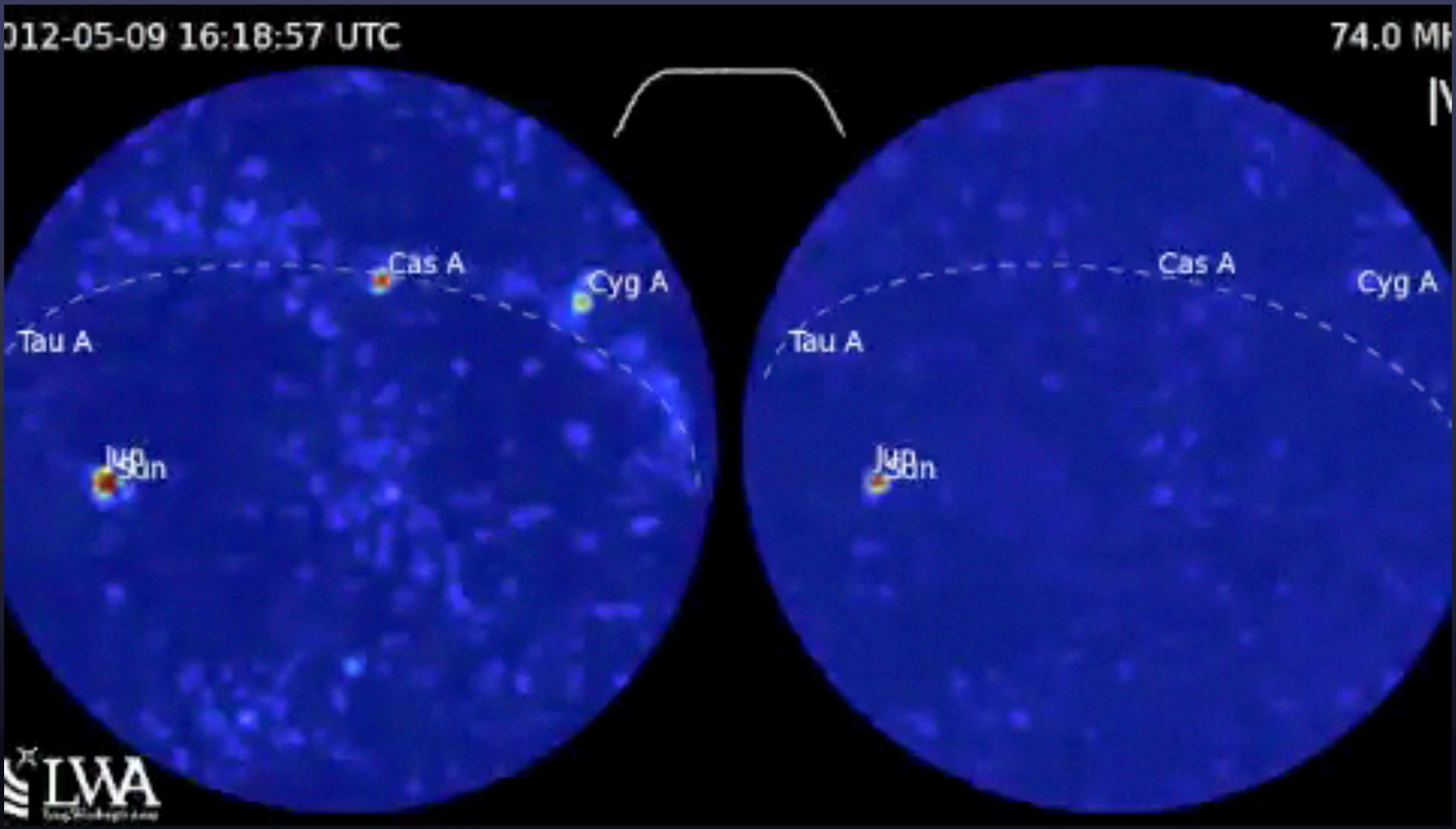
Likely changes:
608 – 614 MHz mobile comm.
1616 – 1627 MHz mobile comm.
1755 – 1850 MHz mobile comm.
2155 – 2200 MHz mobile comm
4400 – 4940 MHz AMT (planes)
5150 – 5250 MHz unlicensed devices
5725 – 5850 MHz unlicensed devices
37000 – 38600 satellite/mobile com
57240 – 59400 WiGig
71000 – 76000 automobile radar
81000 – 86000 automobile radar

LWA1 with ~250 antenna stands



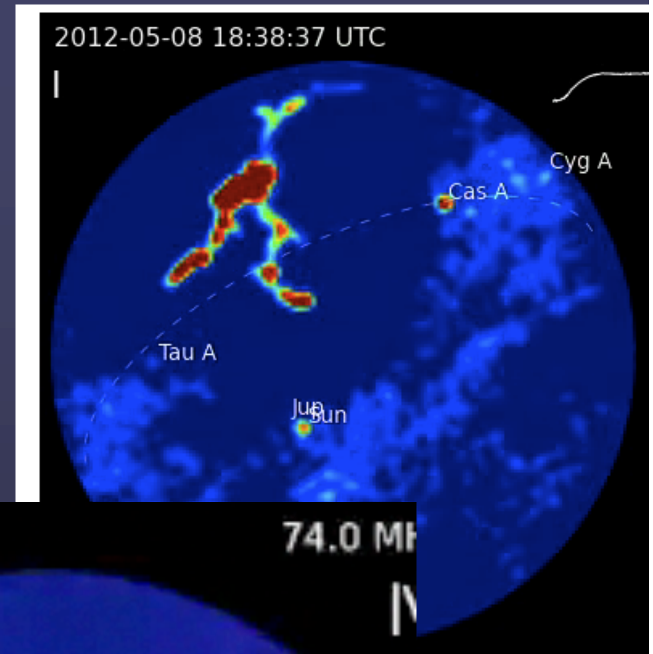
Movies and Solar Interference

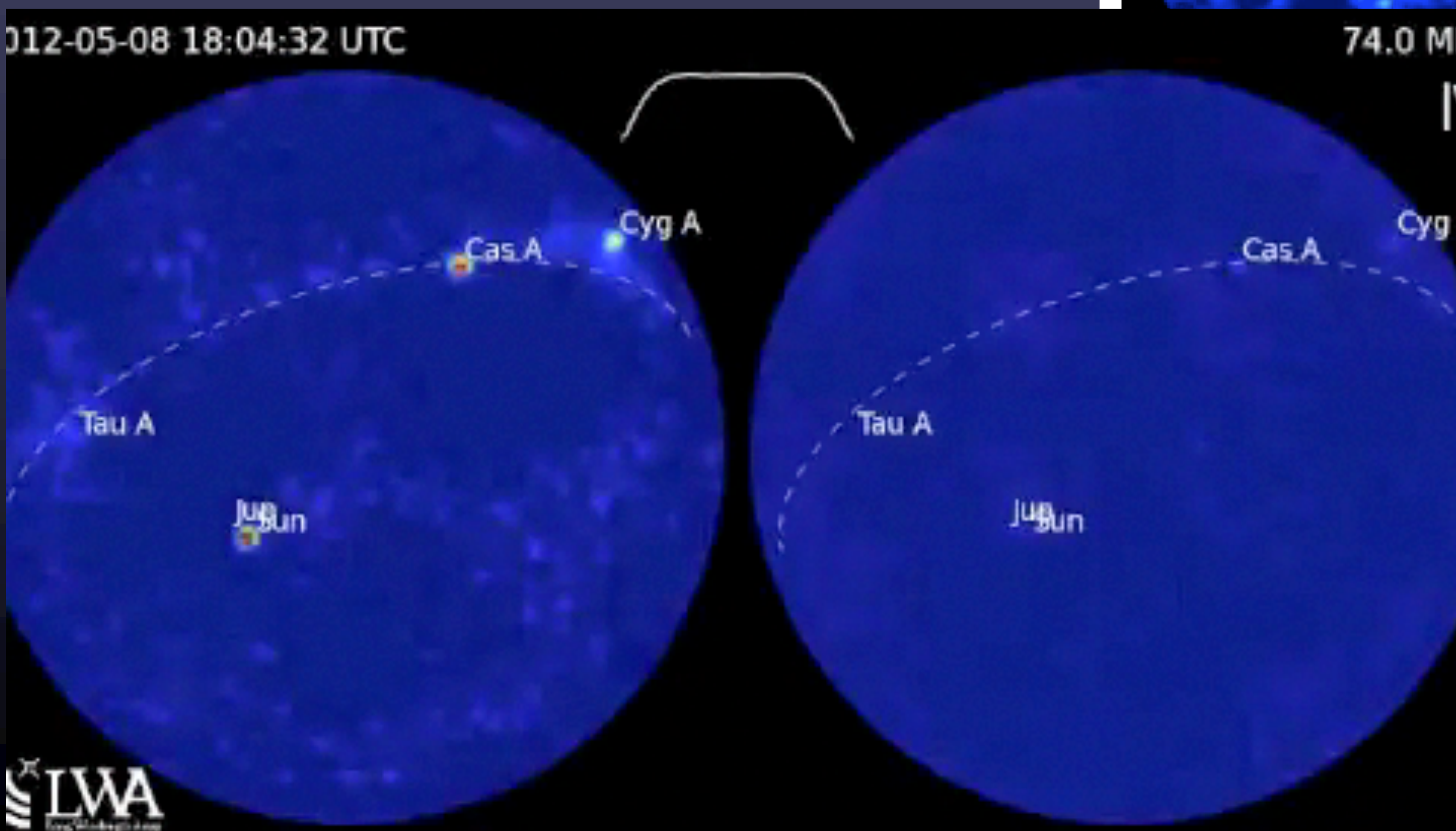
Watch out for the Active Sun



Lightning

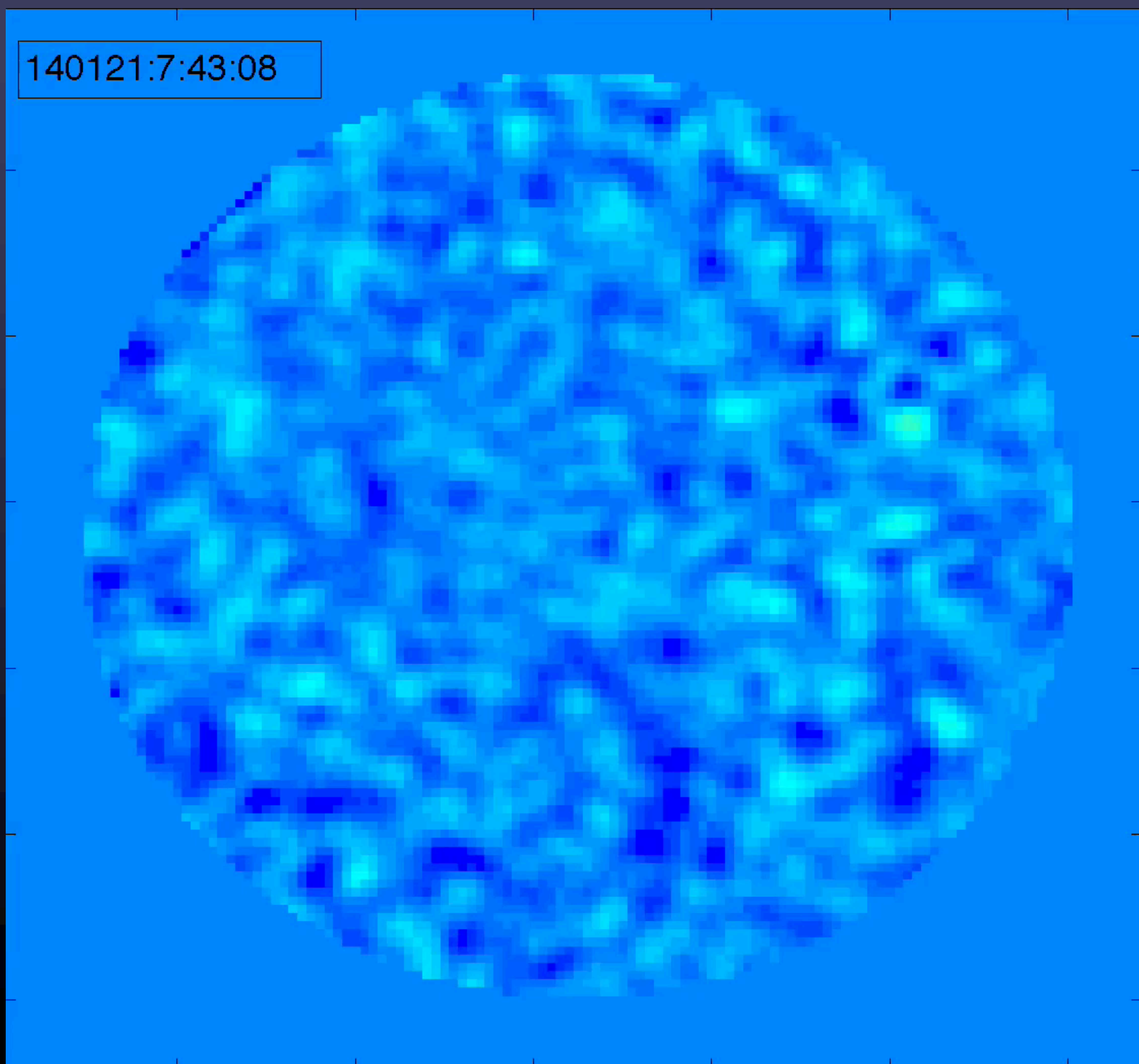
Thunderstorm season on the Plains ...

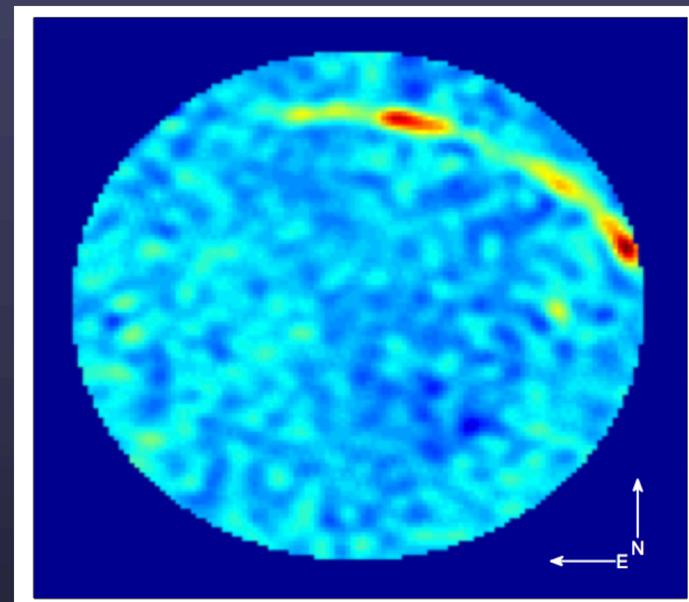




Fireballs

Discovery of broad-band emission from meteors





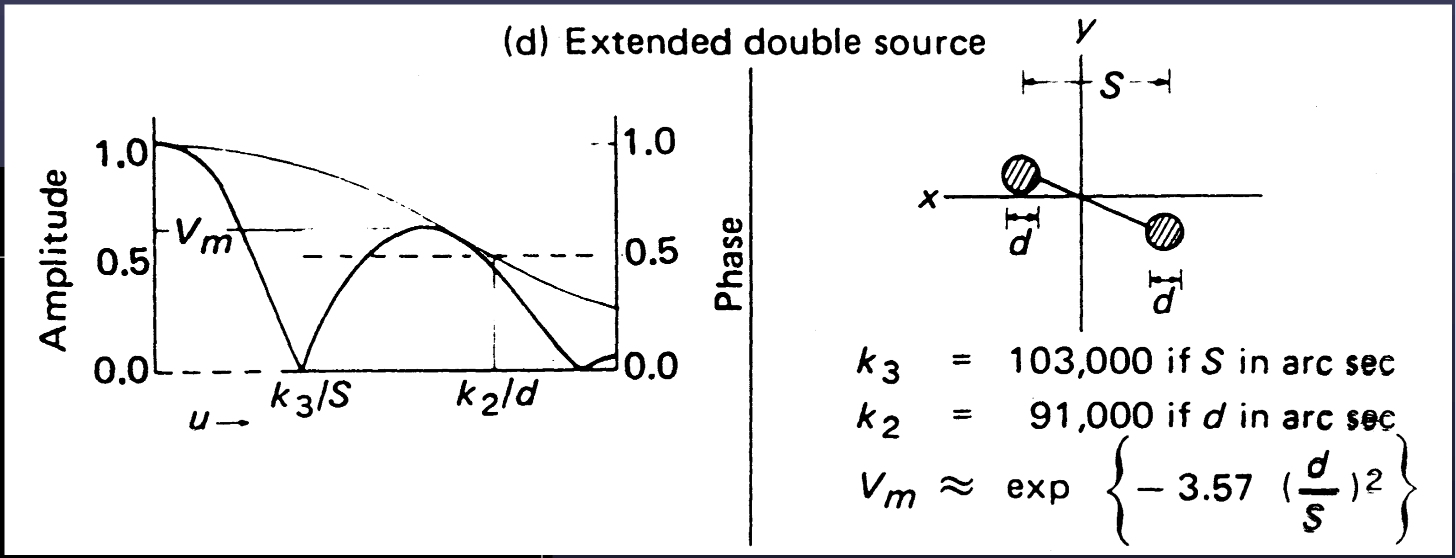
These happen about 1/week
Peak is ~3000 Jy at 40 MHz

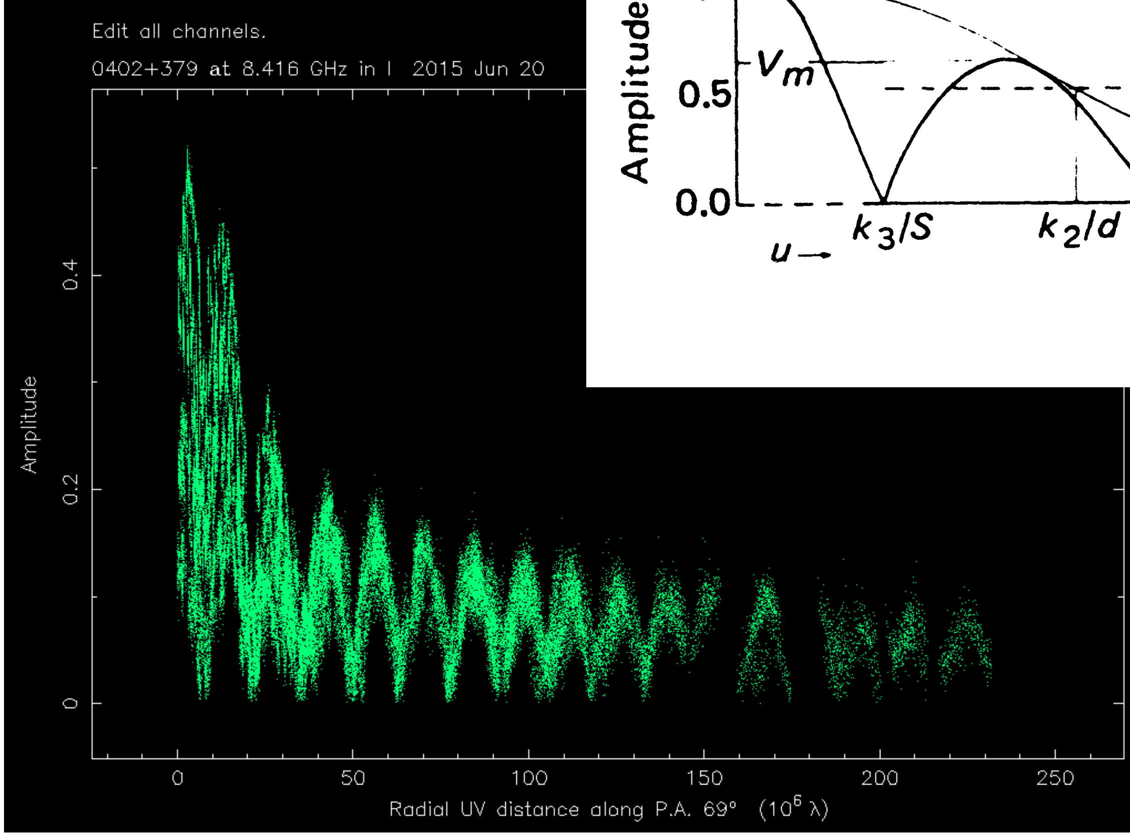
Obenberger et al. 2014

Trial model

By inspection, we can derive a simple model:

Two roughly equal components, each 0.1 Jy, separated by about 15 milliarcsec in p.a. 69º, each about 0.5 milliarcsec in diameter (gaussian FWHM)





To be refined later…

DATA DISPLAYS(1)

List of (u,v) Data

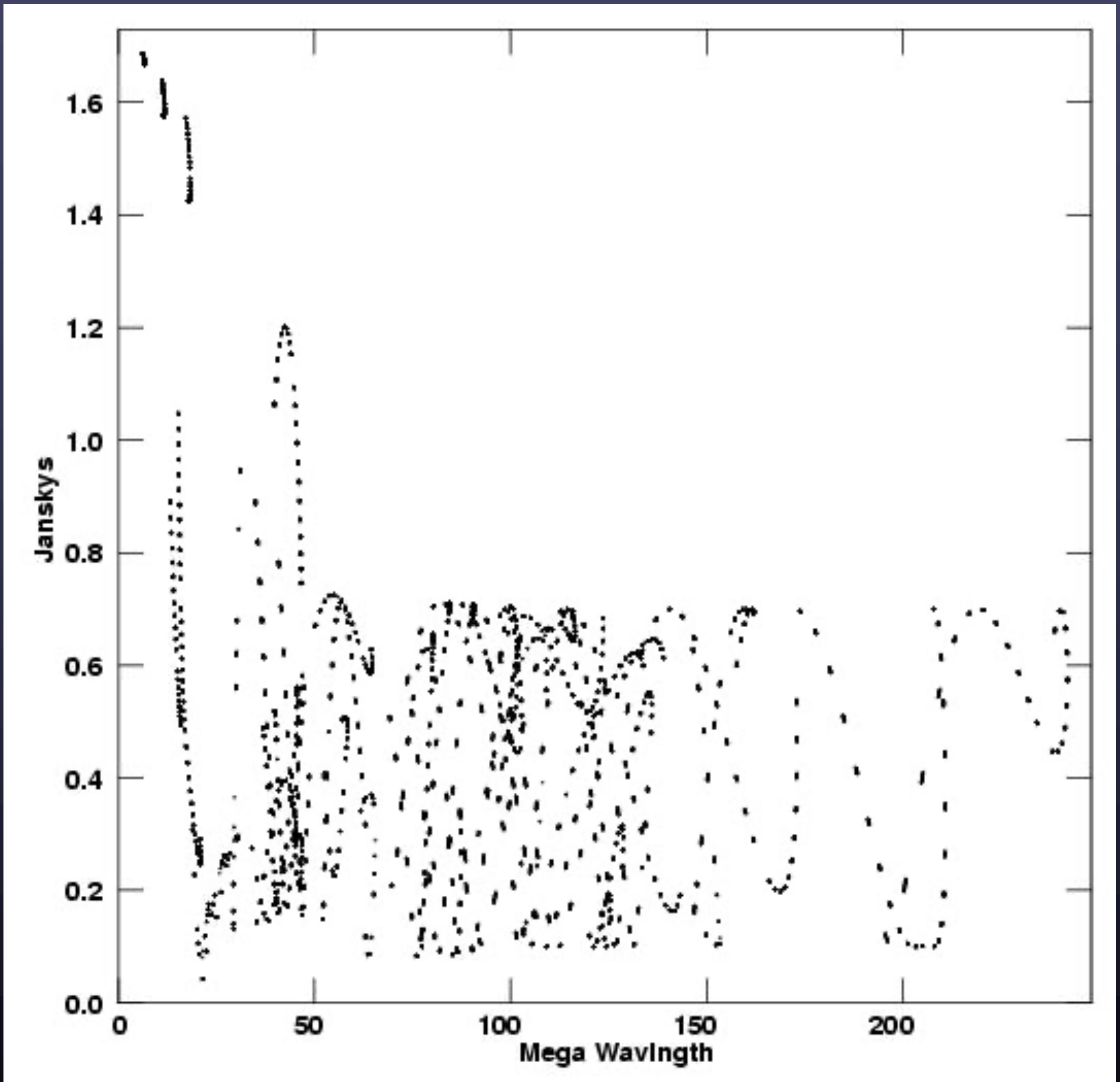
Source= J0121+11 Freq= 8.434858511 Sort= TB 1 RR

Vis #	IAT	Ant	Su	Fq	U(klam)	V(klam)	W(klam)	Amp	Phas	Wt
2191	0/22:35:08.22	5- 6	1	0	94220	23776	100371	0.614	-16	1.0000
3971	0/22:43:43.34	5- 6	1	0	97659	24517	96844	0.508	-13	1.0000
6431	0/23:07:05.15	5- 6	1	0	106307	26661	86632	0.154	17	1.0000
6611	0/23:07:14.98	5- 6	1	0	106364	26677	86557	0.152	17	1.0000
6791	0/23:07:24.81	5- 6	1	0	106421	26692	86483	0.150	18	1.0000
6971	0/23:07:34.64	5- 6	1	0	106477	26708	86408	0.148	19	1.0000
7151	0/23:07:44.47	5- 6	1	0	106534	26724	86333	0.146	19	1.0000
7331	0/23:07:54.30	5- 6	1	0	106591	26739	86259	0.144	20	1.0000
7511	0/23:15:06.84	5- 6	1	0	109027	27438	82930	0.101	74	1.0000
7691	0/23:15:16.67	5- 6	1	0	109081	27454	82854	0.101	75	1.0000
7871	0/23:15:26.50	5- 6	1	0	109135	27470	82777	0.102	77	1.0000
8051	0/23:15:36.33	5- 6	1	0	109189	27486	82701	0.102	78	1.0000
8231	0/23:15:46.16	5- 6	1	0	109243	27502	82624	0.103	79	1.0000
8411	0/23:15:55.99	5- 6	1	0	109297	27518	82547	0.104	81	1.0000
9701	0/23:31:02.36	5- 6	1	0	114020	29035	75322	0.260	134	1.0000
9791	0/23:31:06.29	5- 6	1	0	114040	29042	75290	0.261	134	1.0000
10301	0/23:31:29.88	5- 6	1	0	114156	29082	75098	0.266	134	1.0000
10861	0/23:39:02.08	5- 6	1	0	116320	29863	71379	0.348	139	1.0000
10951	0/23:39:06.01	5- 6	1	0	116339	29870	71346	0.348	139	1.0000
11171	0/23:39:15.84	5- 6	1	0	116384	29887	71264	0.350	139	1.0000

Old School, but sometimes worth-while: e.g. , can search on e.g. Amp > 1.0, or large weight. Often need precise times in order to flag the data appropriately.

DATA DISPLAYS(2)

Jy



Mega Wavelength

Visibility Amplitude versus Projected (u,v) spacing

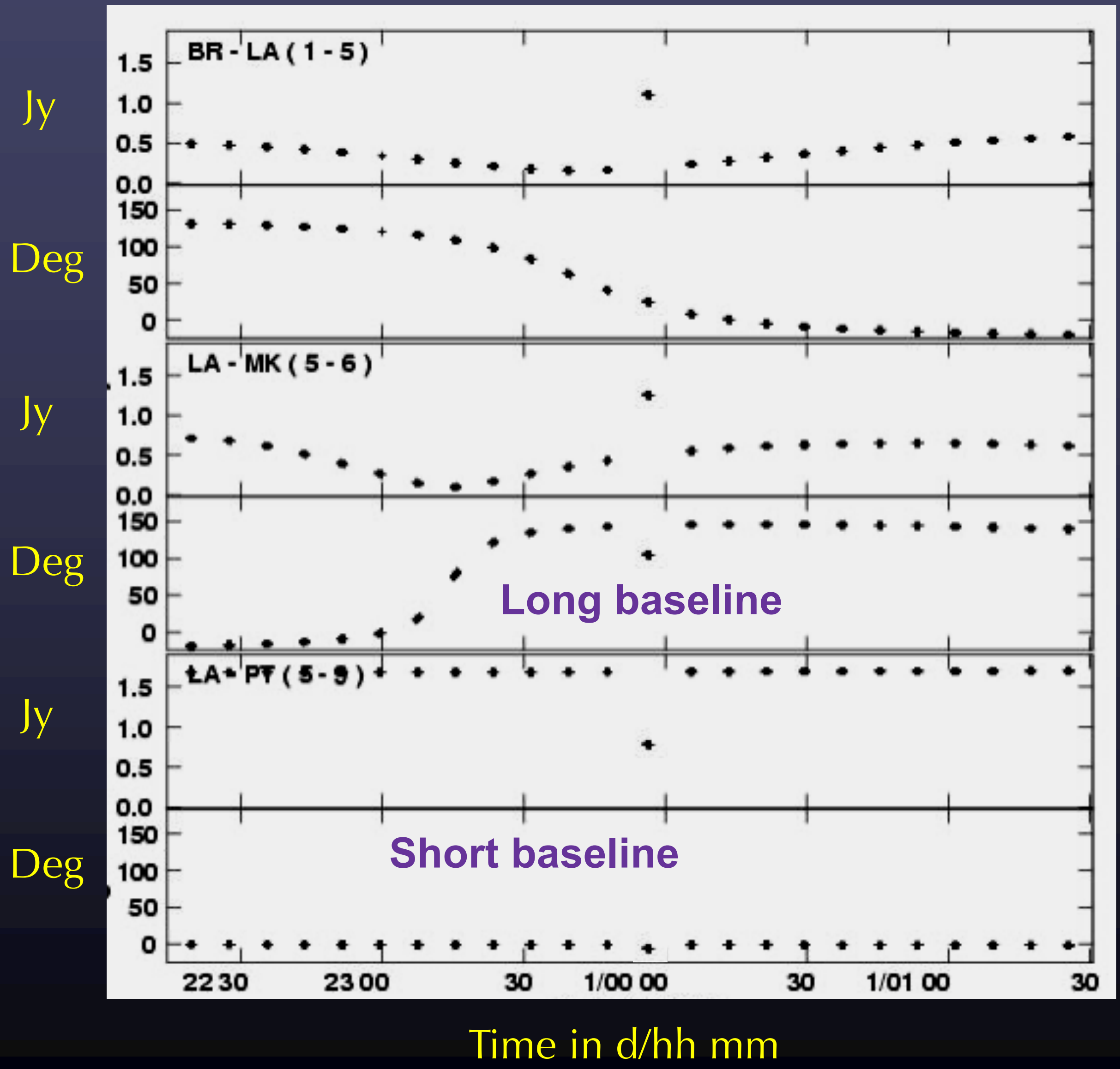
General trend of data.

Useful for relatively strong sources.

Triple source model. Large component cause rise at short spacings.

Oscillations at longer spacings suggest close double.

DATA DISPLAYS(3)



Visibility amplitude and phase versus time for various baselines

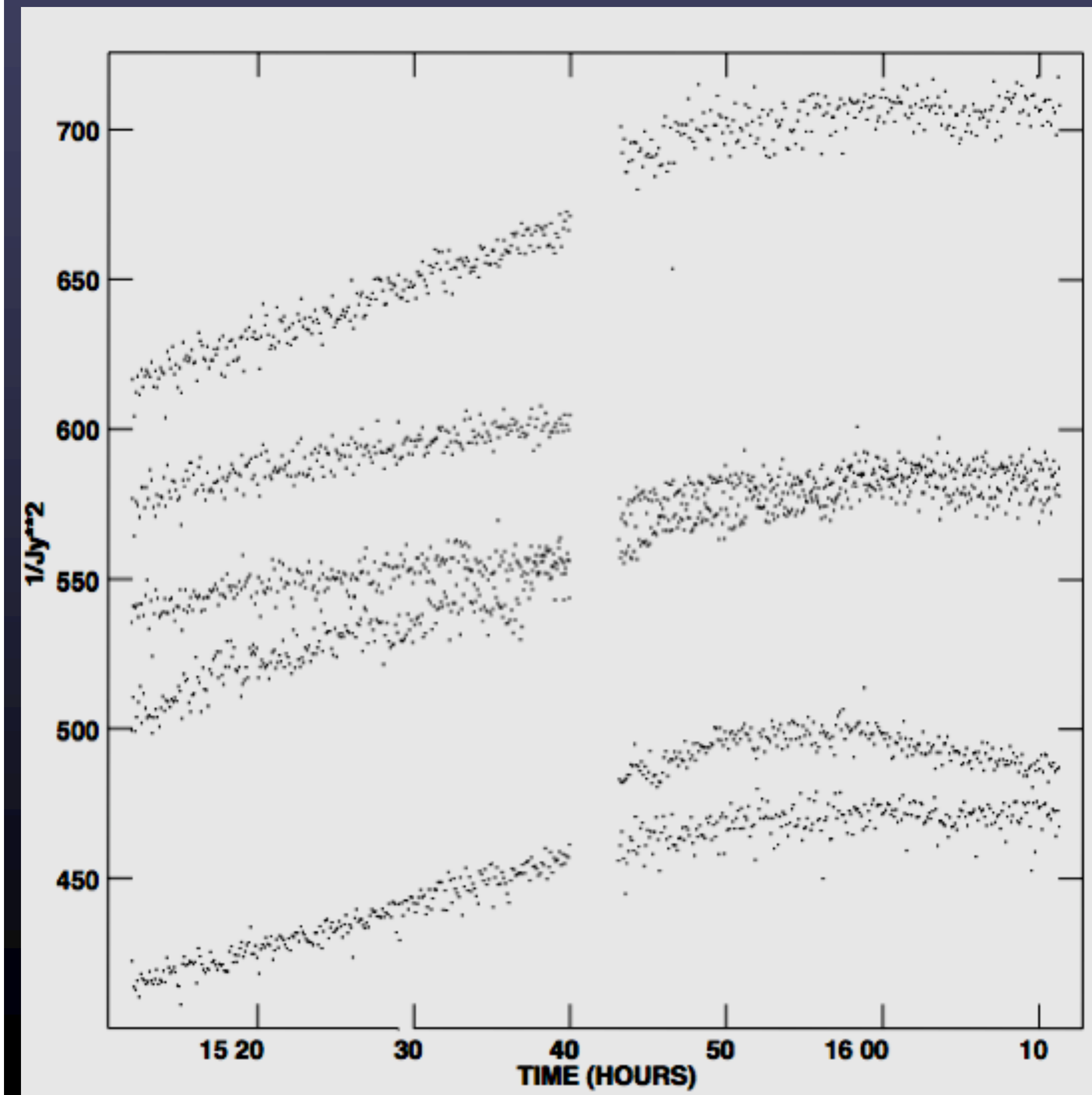
Good for determining the continuity of the data

Should be relatively smooth with time

Outliers are obvious.

DATA DISPLAYS(4)

Weights of antennas 4 with 5,6,7,8,9



All (u,v) data points have a weight.

The weight depends on the antenna sensitivity, measured during the observations

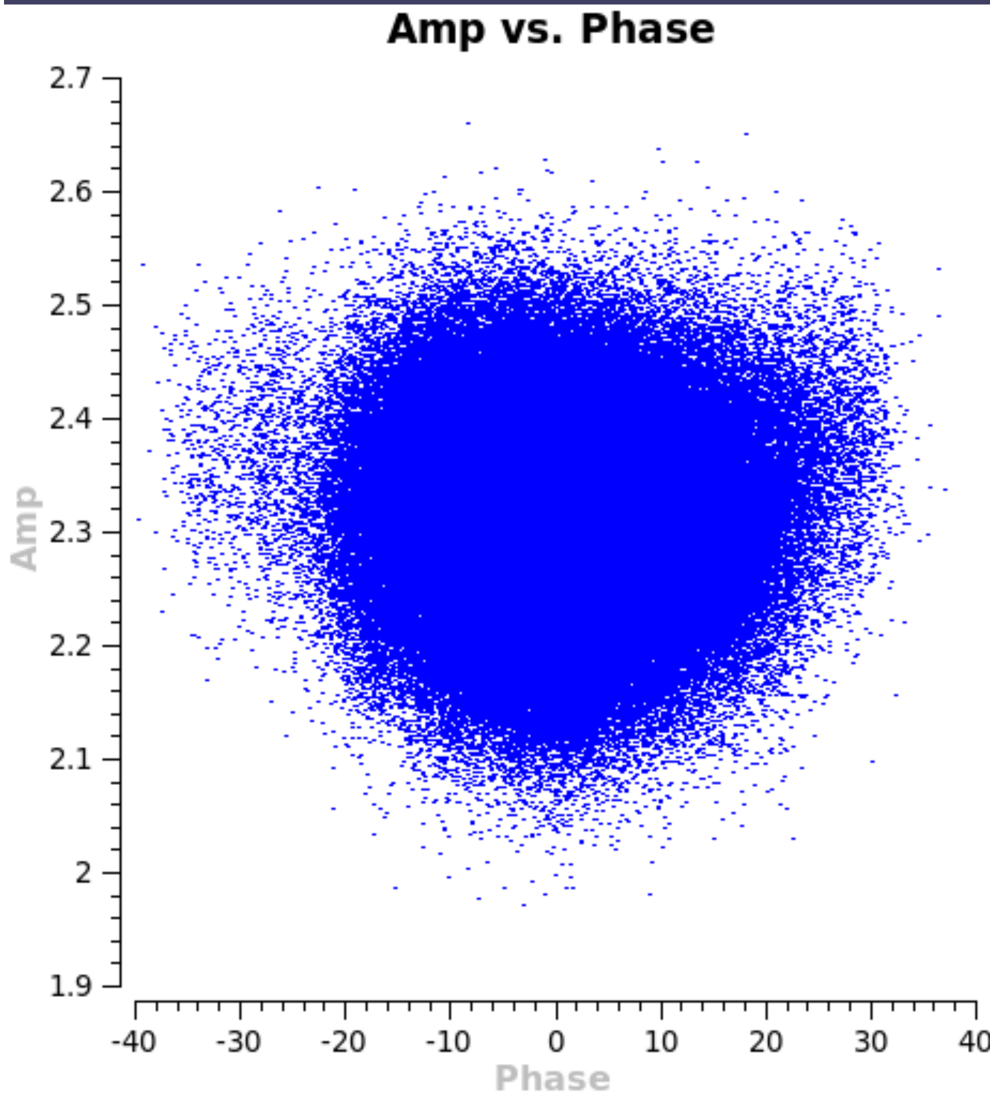
The amplitude calibration values also modify the weights.

Occasionally the weight of the points become very large, often caused by subtle software bugs.

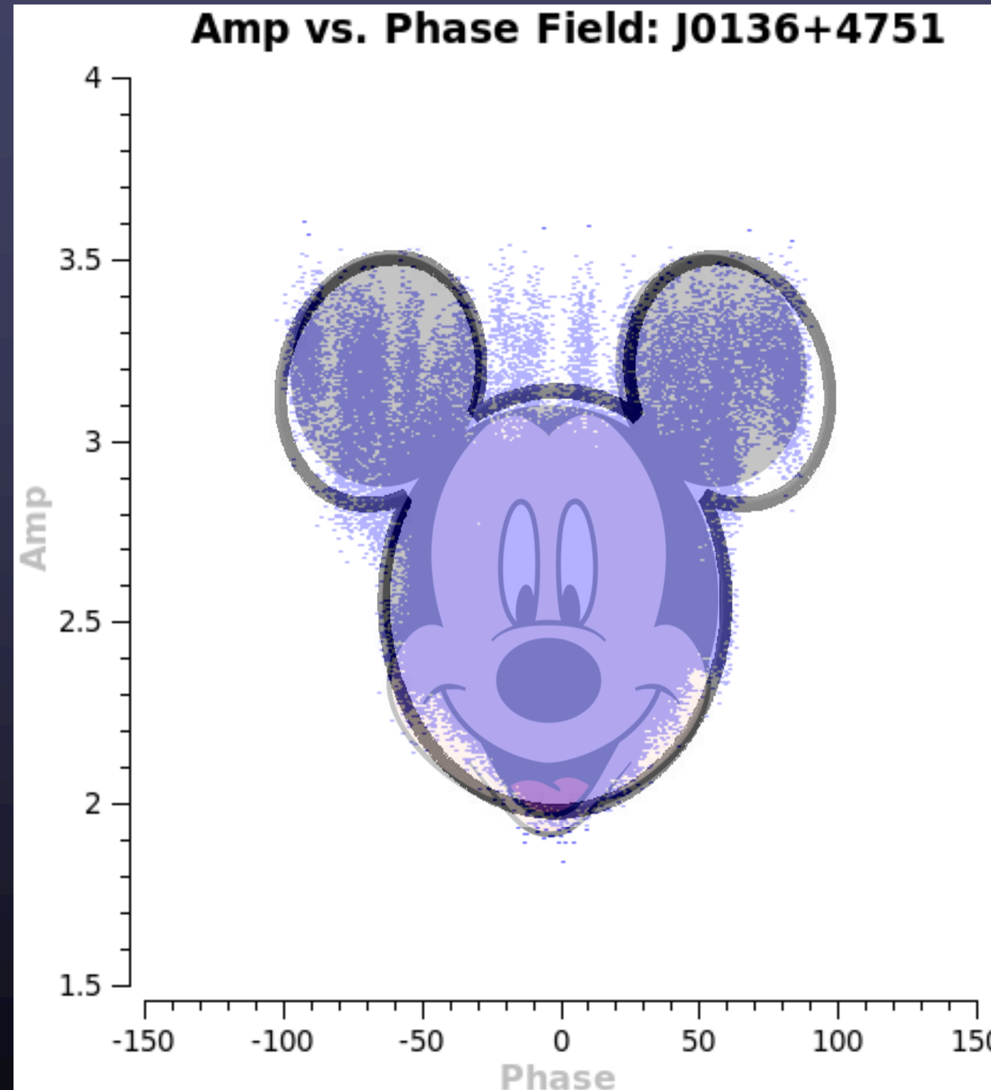
A large discrepant weight causes the same image artifacts as a large discrepant visibility value.

Please check weights to make sure they are reasonable.

DATA DISPLAYS(5) – Amplitude vs Phase



Good



Bad

IMAGE PLANE OR DATA (U,V) PLANE INSPECTION?

Errors obey Fourier transform relationship

- Narrow feature in (u,v) plane <-> wide feature in image plane
- Wide feature in (u,v) plane <-> narrow feature in image plane
  - Note: often easier to spot narrow features
- Data (u,v) amplitude errors <->symmetric image features
- Data (u,v) phase errors <-> asymmetric image features
- An obvious defect may be hardly visible in the transformed plane
- A small, almost invisible defect may become very obvious in the transformed plane
- Noise bumps can have sidelobes!

FINDING ERRORS

---Obvious outlier data (u,v) points:

100 bad points in 100,000 data points gives an 0.1% image error

(unless the bad data points are 1 million Jy)

LOOK at DATA to find gross problem (you'd be hard pressed to find it in the image plane other than a slight increase in noise)

---Persistent small data errors:

e.g. a 5% antenna gain calibration error is difficult to see in (u,v) data (not an obvious outlier), but will produce a 1% effect in image with specific characteristics (more later).

USE IMAGE to discover problem

---Non-Data Problems:

Data ok, but algorithms chosen aren't up to the task.

ERROR RECOGNITION IN THE (u,v) PLANE

Editing obvious errors in the (u,v) plane

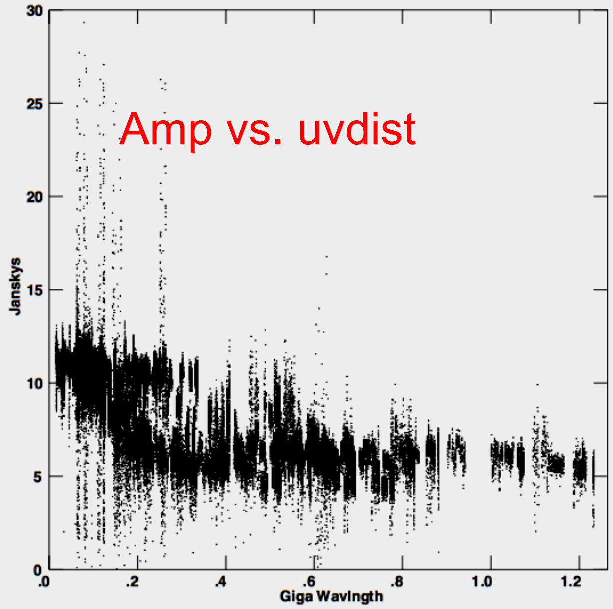
---Mostly consistency checks assume that the visibility cannot change much over a small change in (u,v) spacing

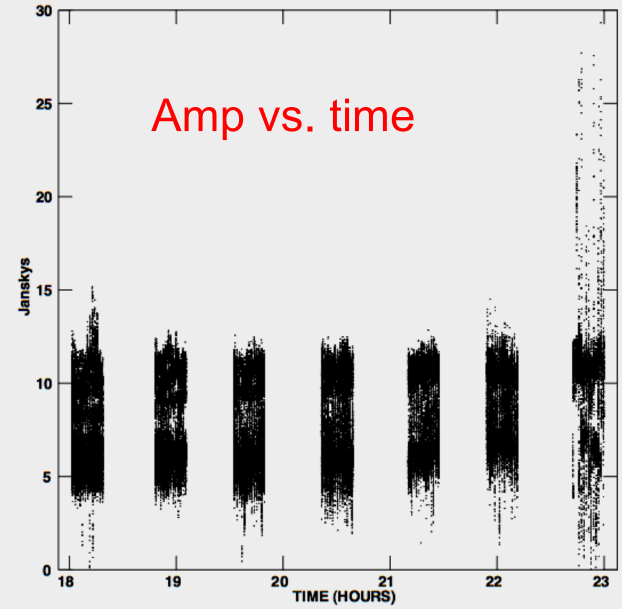
---Also, look at gains and phases from calibration processes. These values should be relatively stable.

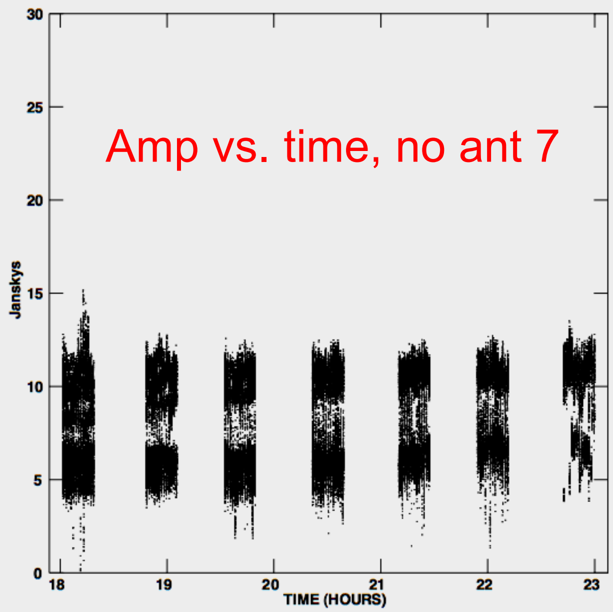
See Summer school lecture notes in 2002 by Myers

See ASP Vol 180, Ekers, Lecture 15, p321

VISIBILITY AMPLITUDE PLOTS







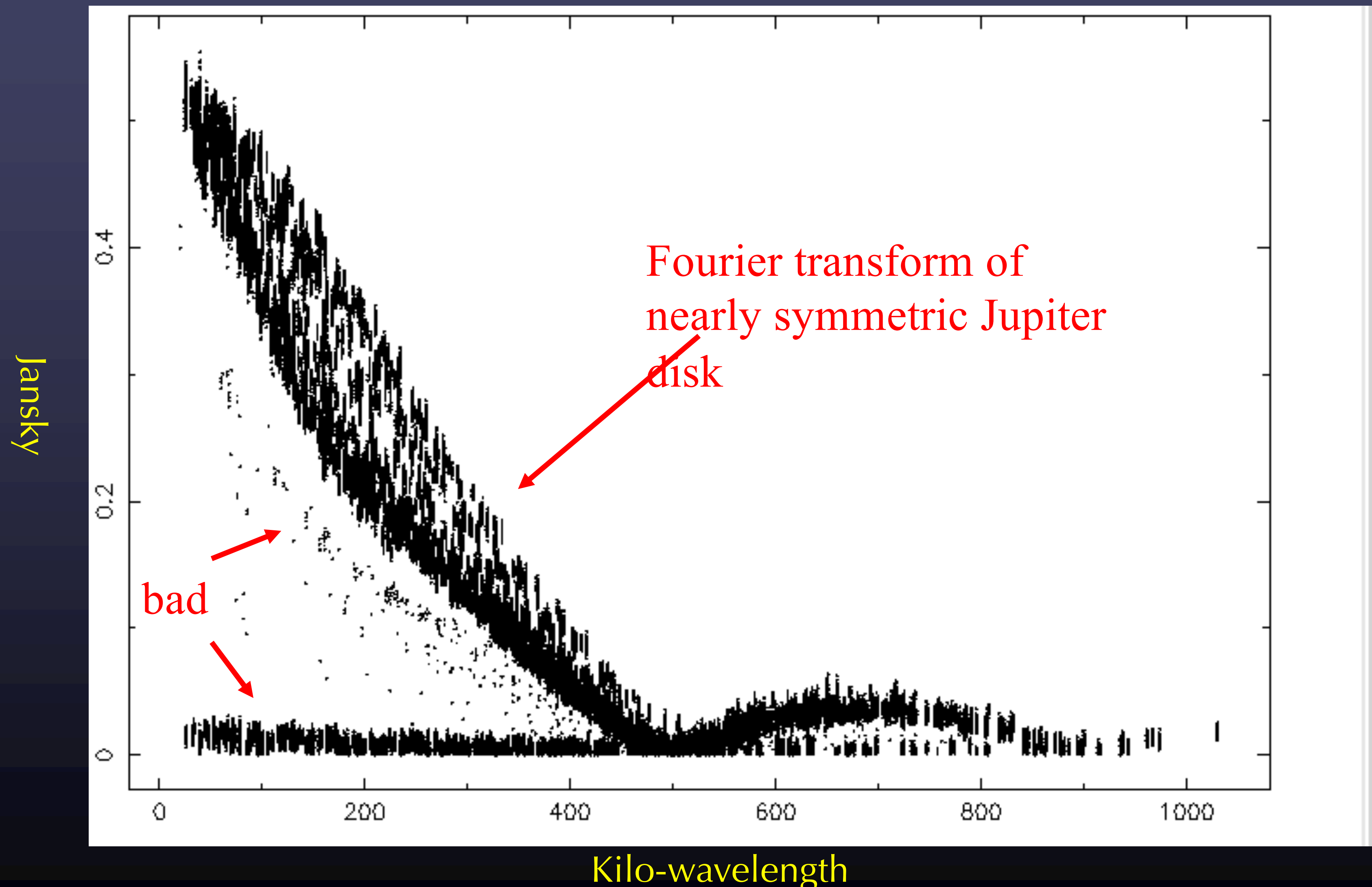
Amp vs. uvdist shows outliers

Amp vs. time shows outliers in last scan

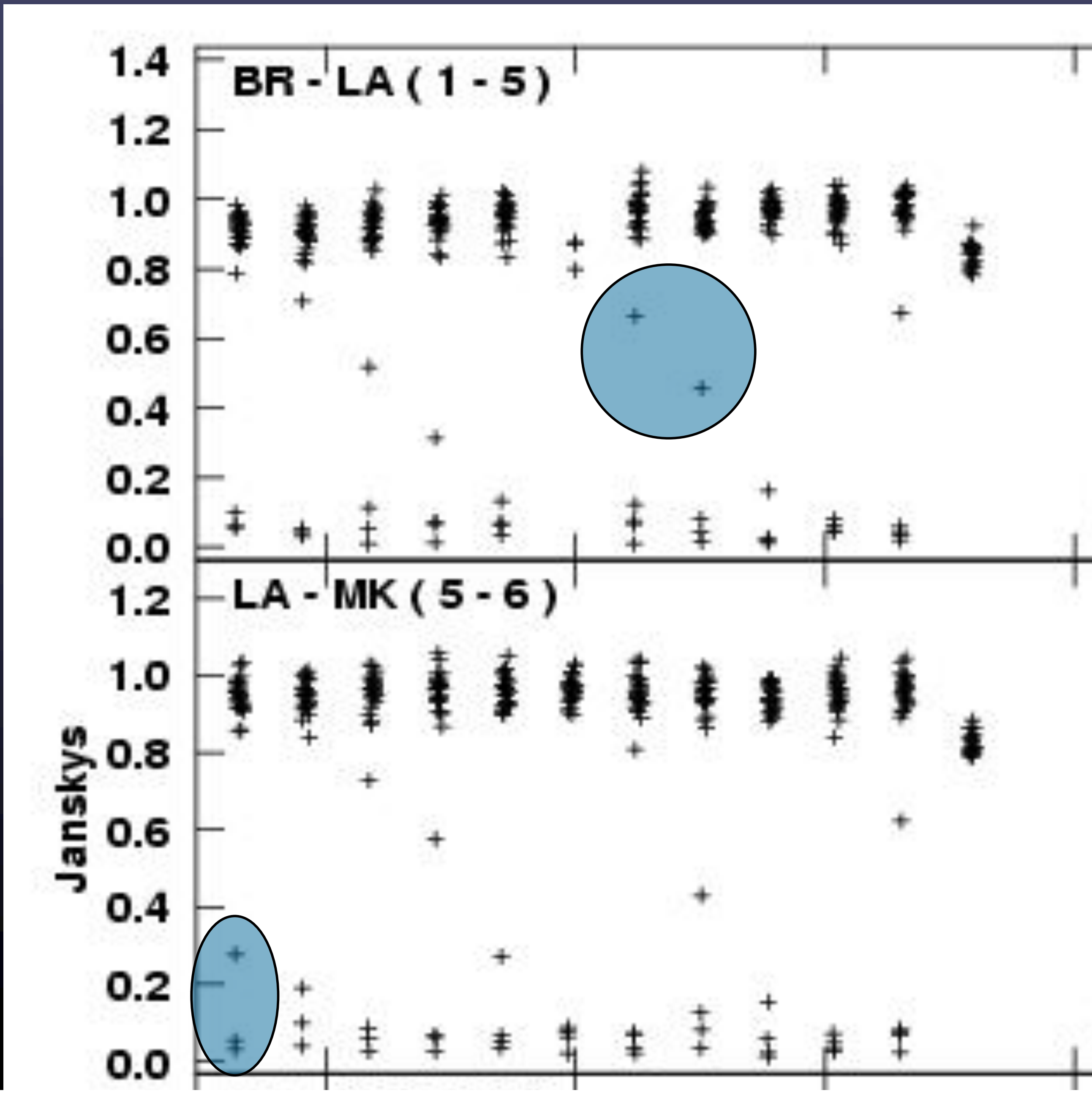
Amp vs. time without ant 7 shows good data

(3C279 VLBA data at 43 GHz)

Example Edit – plotms



Drop-outs at Scan Beginnings



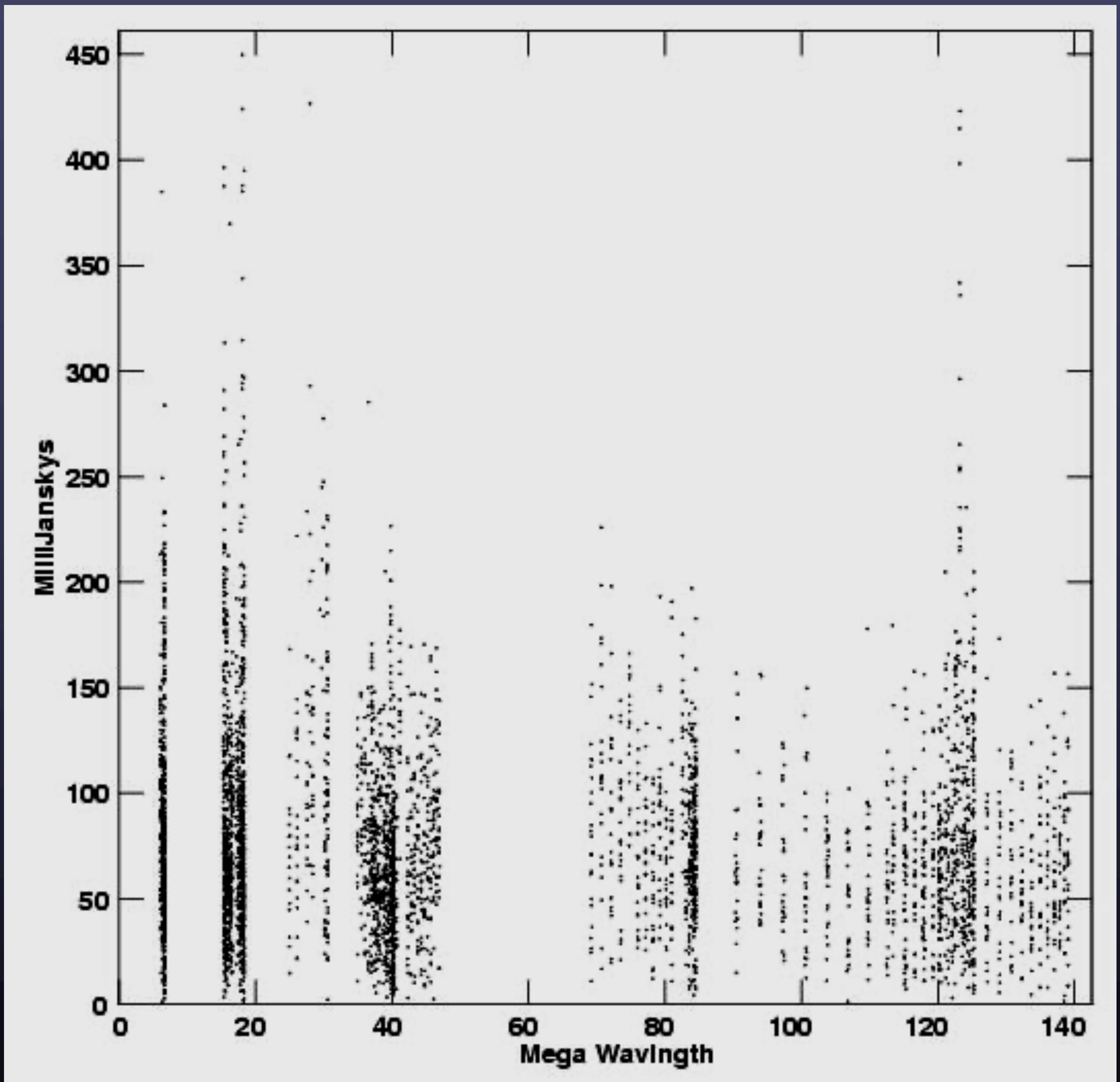
Often the first few points of a scan are low. E.g. antenna not on source.

Software can remove these points (aips,casa 'quack')

Flag extension:

Should flag all sources in the same manner even though you cannot see dropout for weak sources

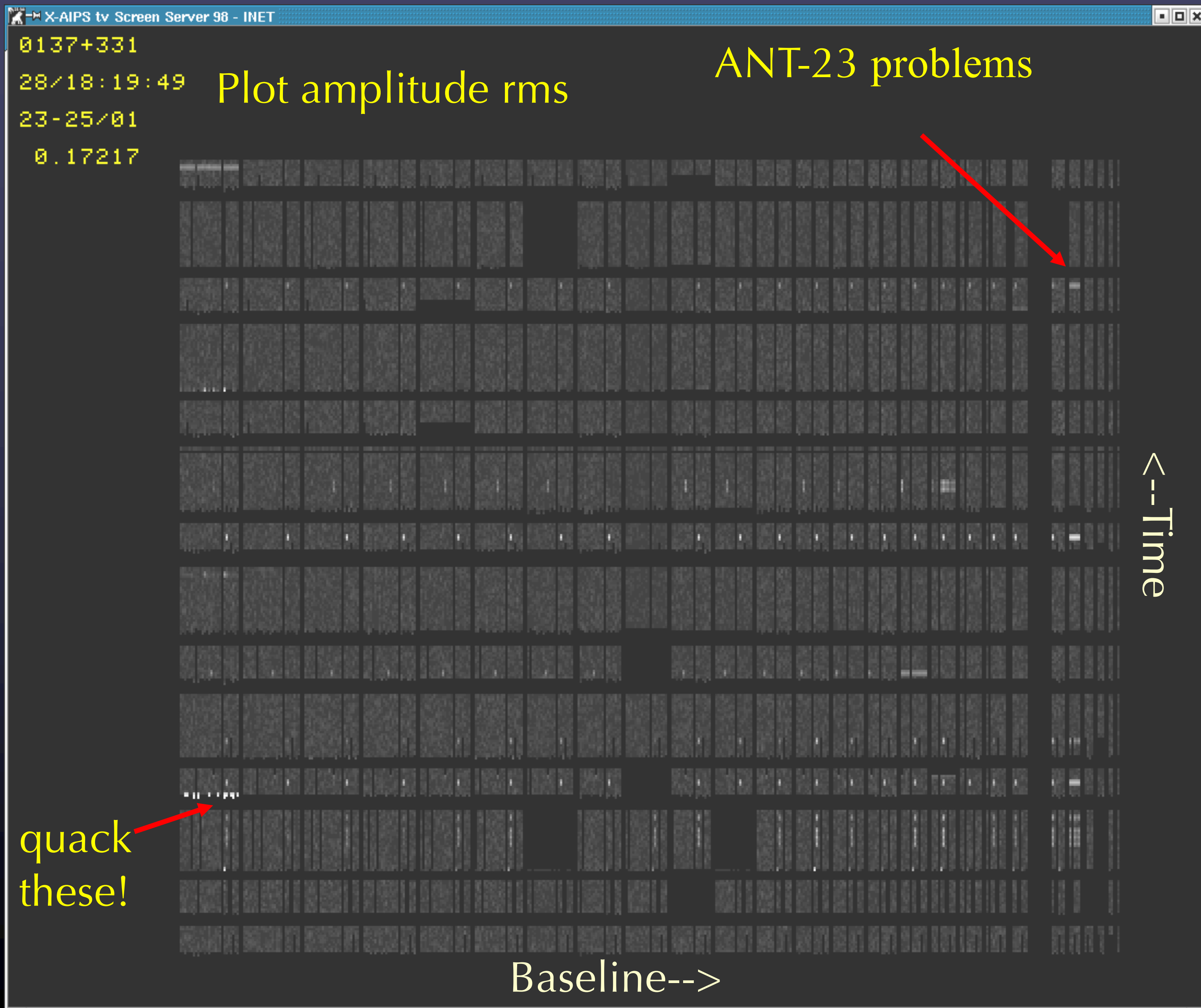
Editing Noise-dominated Sources



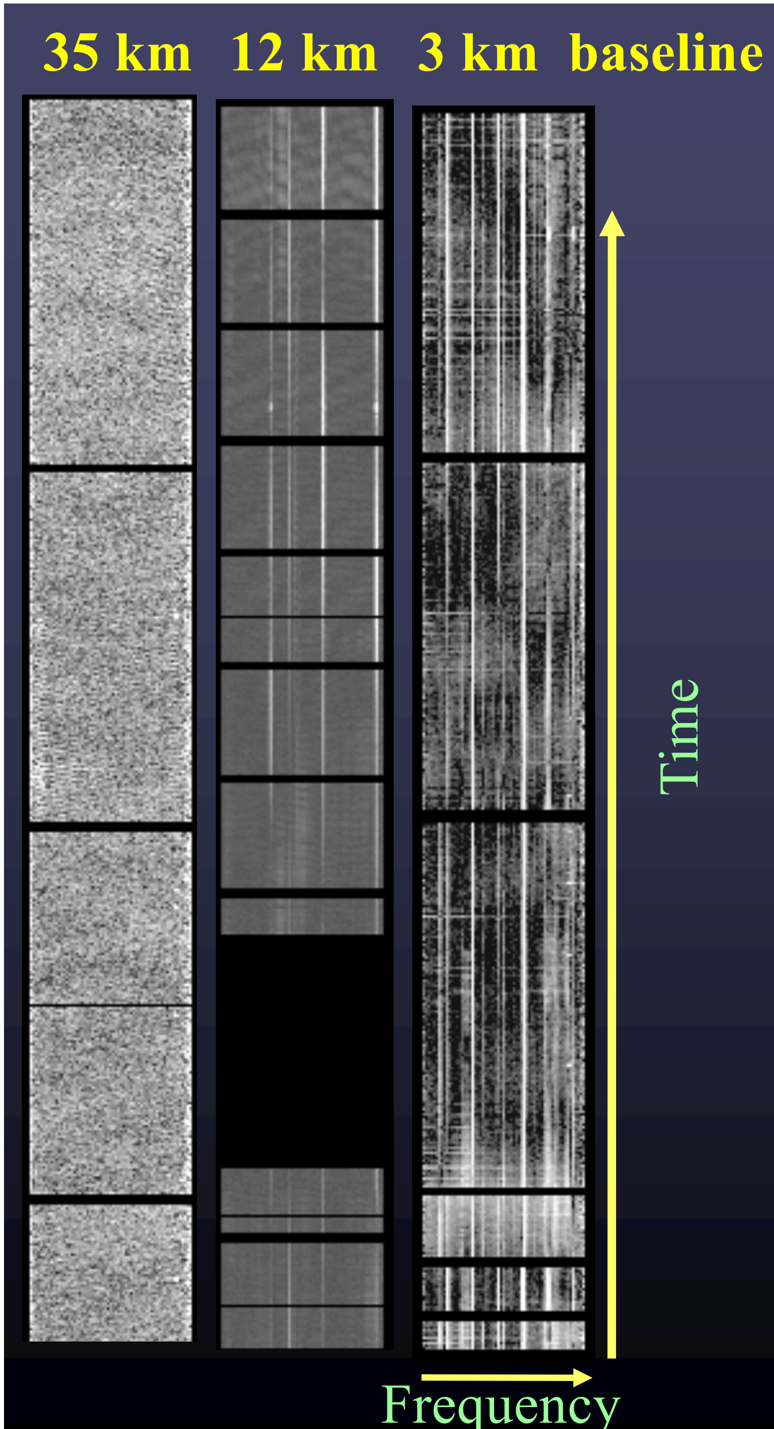
No source structure information is detected. Noise dominated.

All you can do is quack and remove outlier points above ~3sigma (0.3 Jy). Precise level not important as long as large outliers are removed.

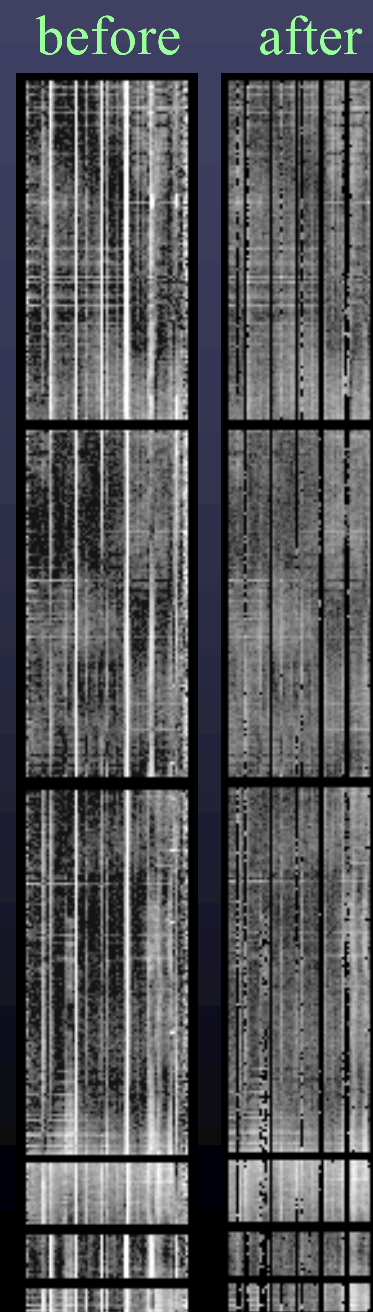
USING TVFLG (VIEWER) DISPLAY on a source



RFI Excision



AIPS: SPFLG



RFI environment worse on short baselines

Several 'types': narrow band, wandering, wideband, ...

Wideband interference hard for automated routines

AIPS tasks FLGIT, RFLAG, SPFLG and CASA flagdata, mode=‘rfi’

AOFlagger by Offringa

Automation is crucial for WIDAR (wide band, lots of data)

From Tracy Clarke (NRL)

ERROR RECOGNITION IN THE IMAGE PLANE

Some Questions to ask:

Noise properties of image:

- Is the rms noise about that expected from integration time?
- Is the rms noise much larger near bright sources?
- Are there non-random noise components (faint waves and ripples)?

Funny looking Structure:

- Non-physical features; stripes, rings, symmetric or anti-symmetric
- Negative features well-below 4xrms noise
- Does the image have characteristics that look like the dirty beam?

Image-making parameters:

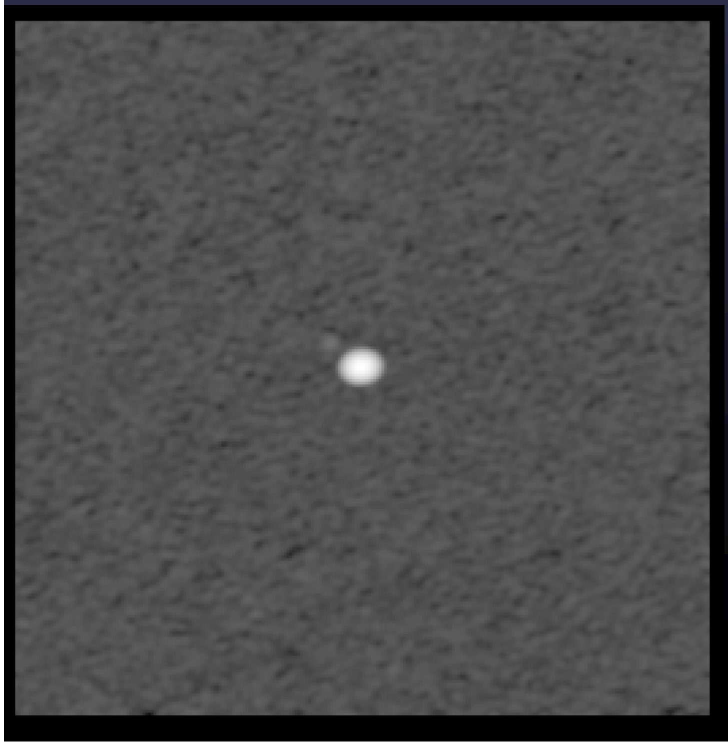
- Is the image big enough to cover all significant emission?
- Is cell size too large or too small? ~4 points per beam okay
- Is the resolution too high to detect most of the emission?

EXAMPLE 1

Data bad over a short period of time

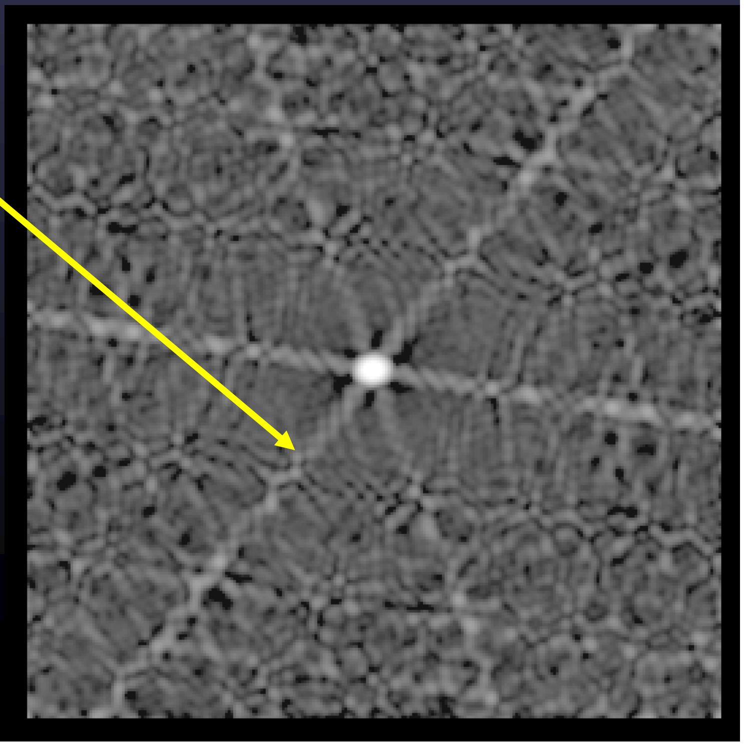
Results for a point source using VLA.   13 x 5min observation over 10 hr.
Images shown after editing, calibration and deconvolution.

no errors:
max 3.24 Jy
rms 0.11 mJy



10% amp error for all antennas for 1 time period
rms 2.0 mJy

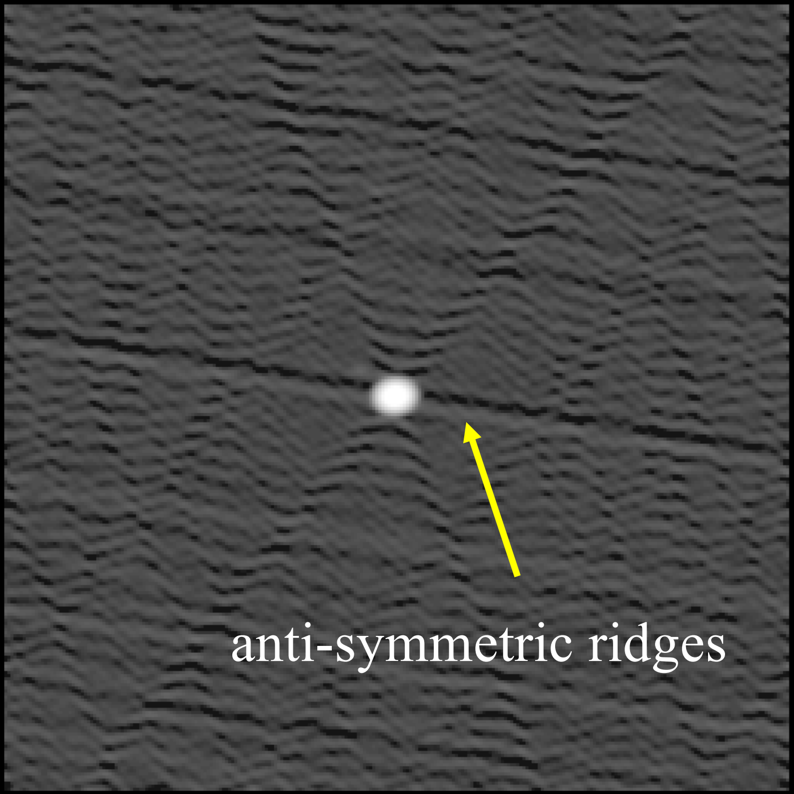
6-fold symmetric pattern due to VLA “Y”.
Image has properties of dirty beam.



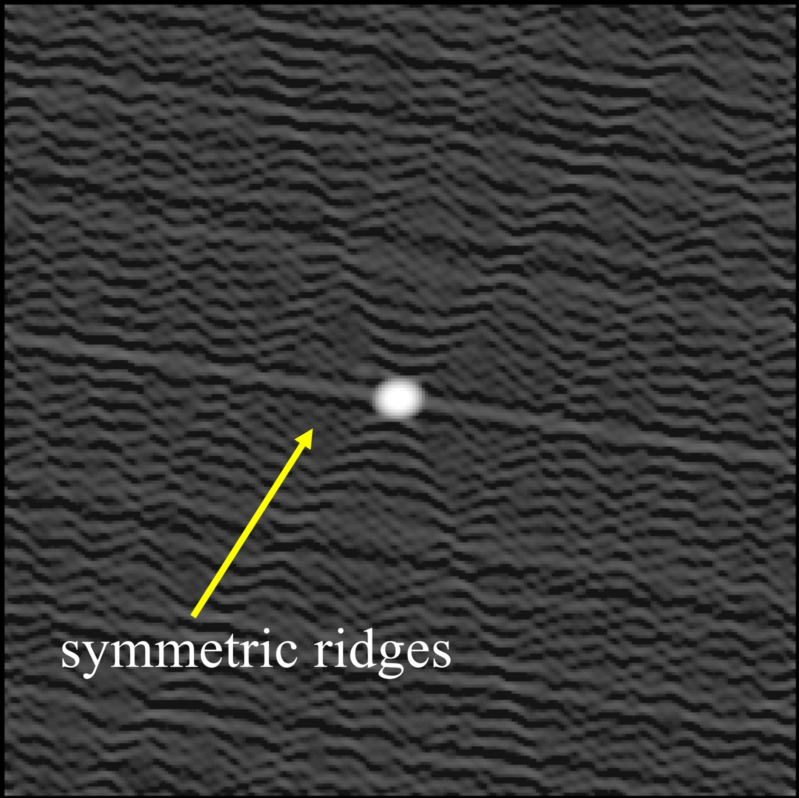
EXAMPLE 2
Short burst of bad data

Typical effect from one bad antenna

10 deg phase error for
one antenna at one time
rms 0.49 mJy



20% amplitude error for
one antenna at one time
rms 0.56 mJy

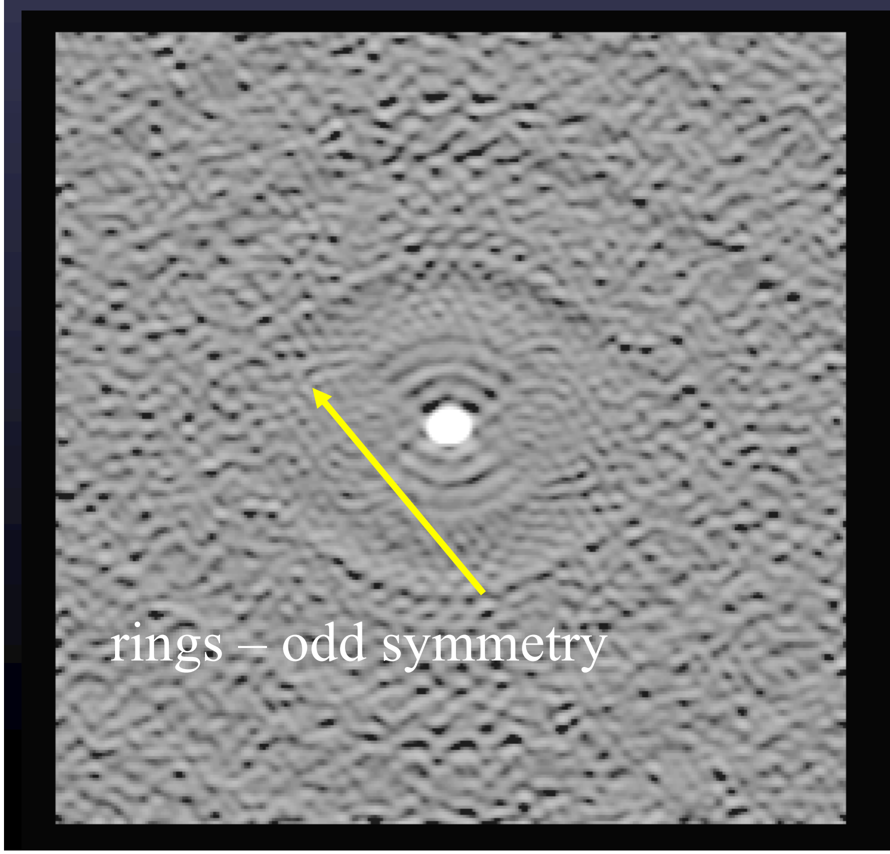


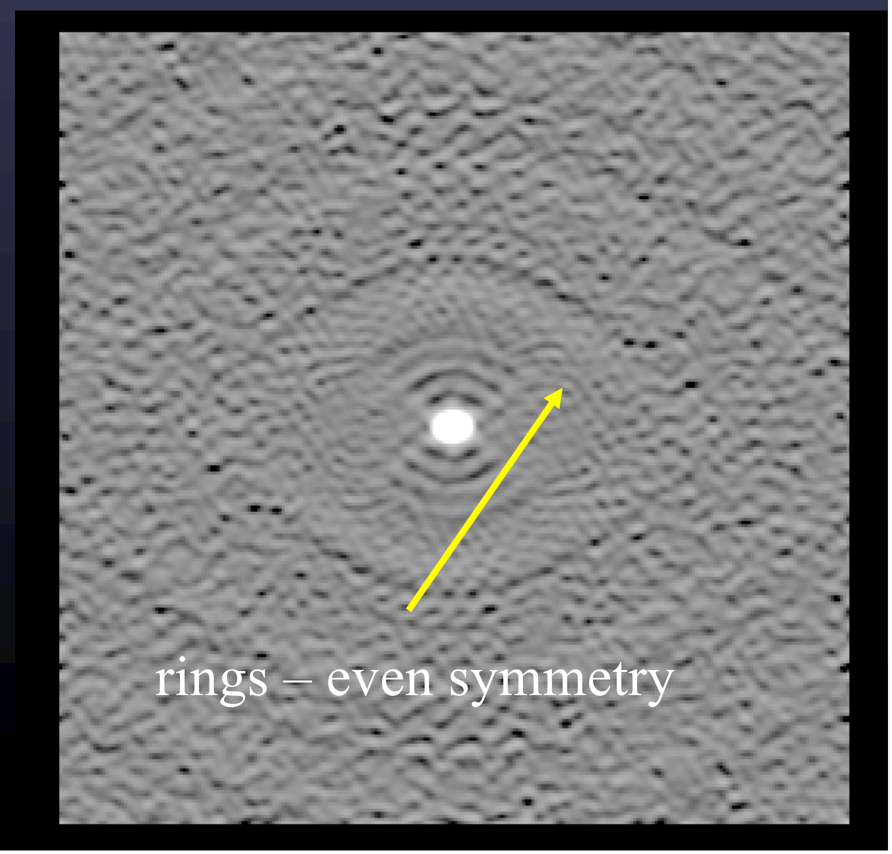
EXAMPLE 3
Persistent errors over most of observations

NOTE: 10 deg phase error to 20% amplitude error cause similar sized artifacts

10 deg phase error for one antenna all times
rms 2.0 mJy

20% amp error for one antenna all times
rms 2.3 mJy





EXAMPLE 4
Spurious Correlator Offset Signals

Occasionally correlators produce ghost signals or cross talk signals
Occurred during change-over from VLA to EVLA system

Symptom: Garbage near phase center, dribbling out into image

Image with correlator offsets

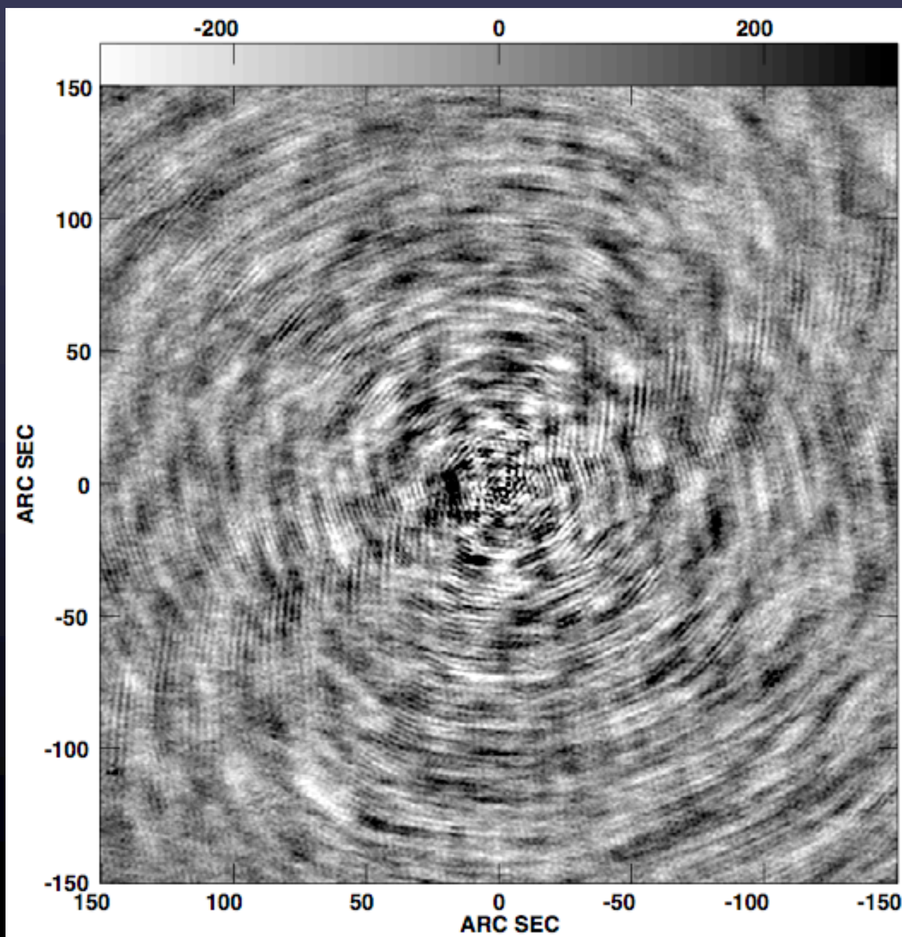
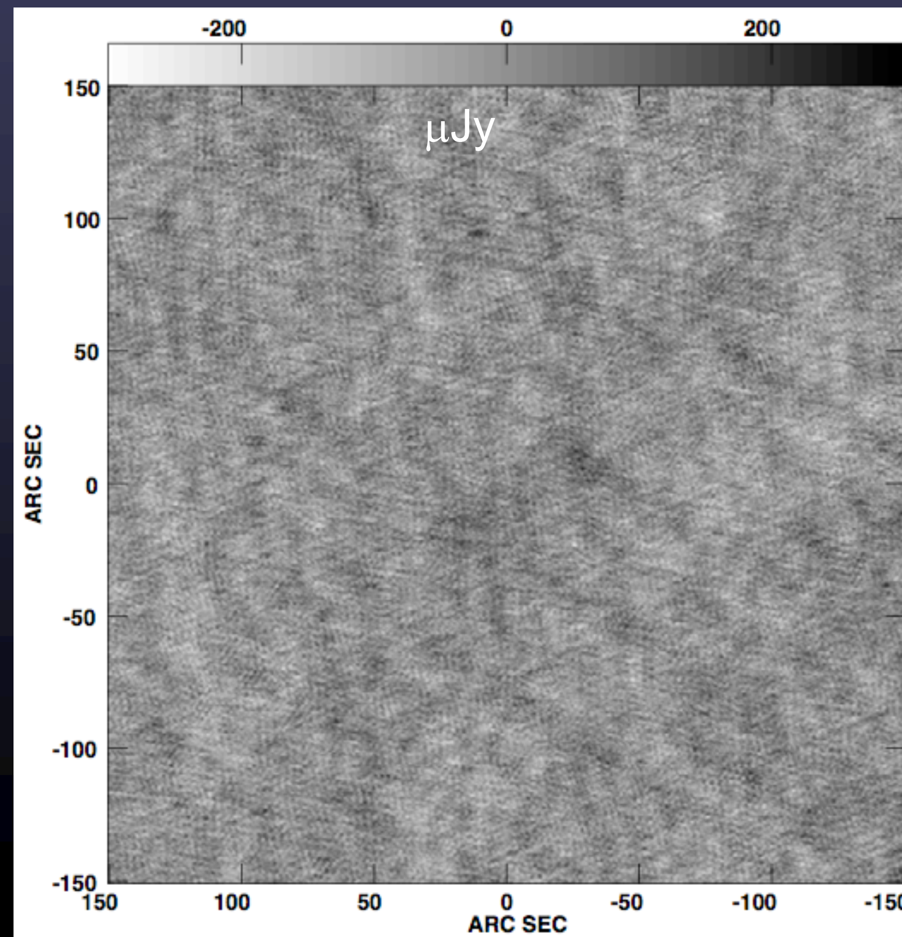
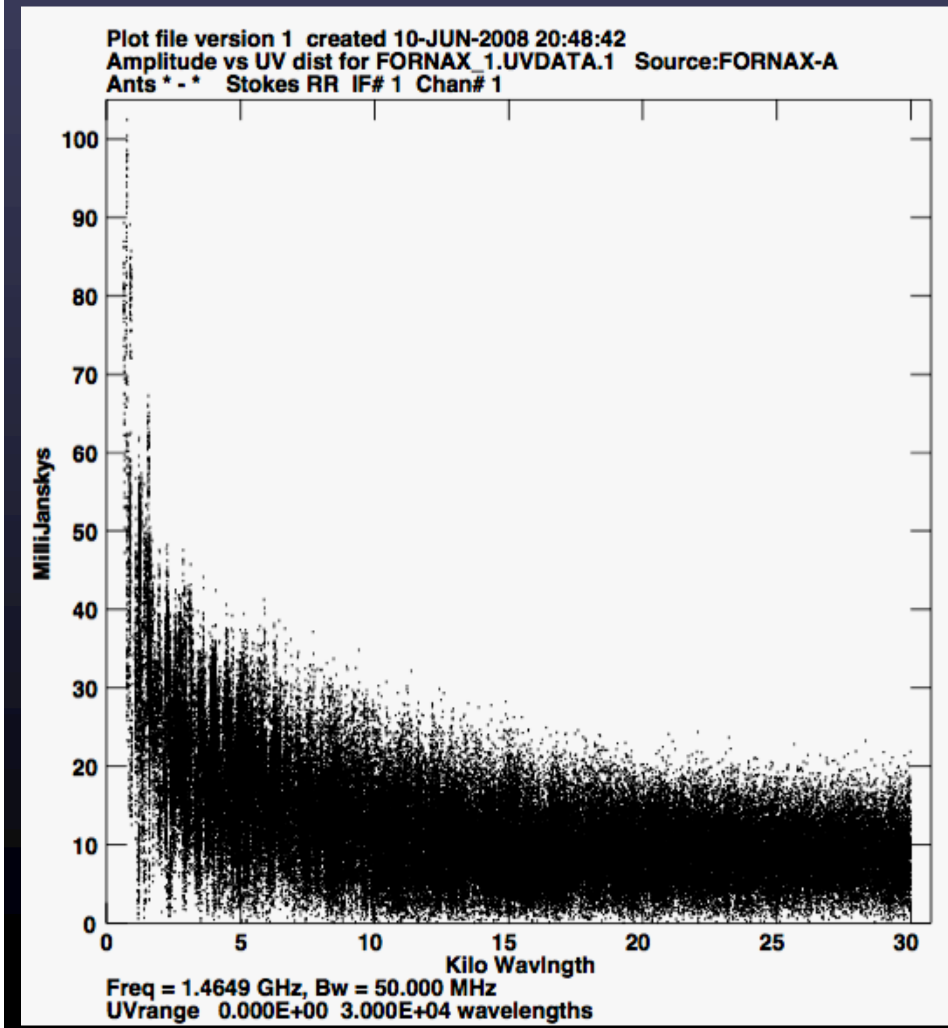


Image after correction of offsets



DECONVOLUTION ERRORS

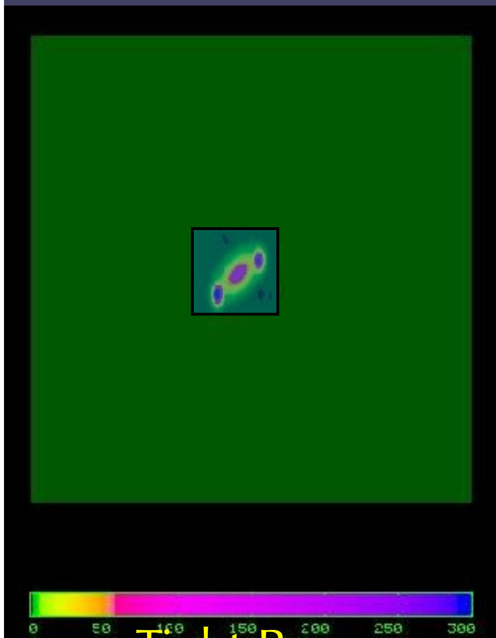
Even if the data are perfect, image errors and uncertainties can occur if the (u,v) coverage is not adequate to map the source structure.



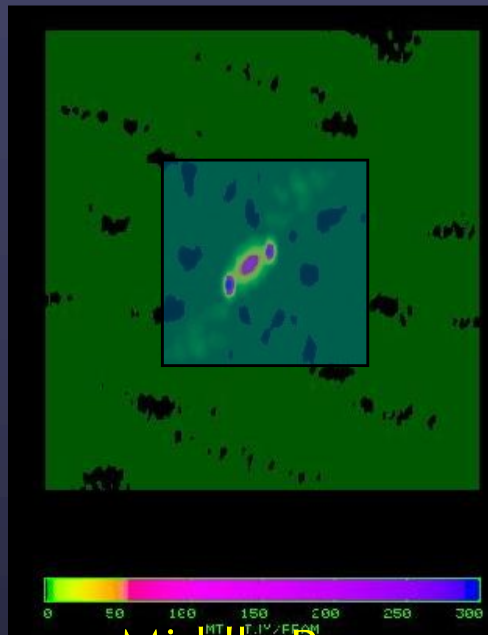
The extreme rise of visibility at the short spacings makes it impossible to image the extended structure. You are better off imaging the source with a cutoff below about 2 kilo-wavelengths

Get shorter spacing or single-dish data

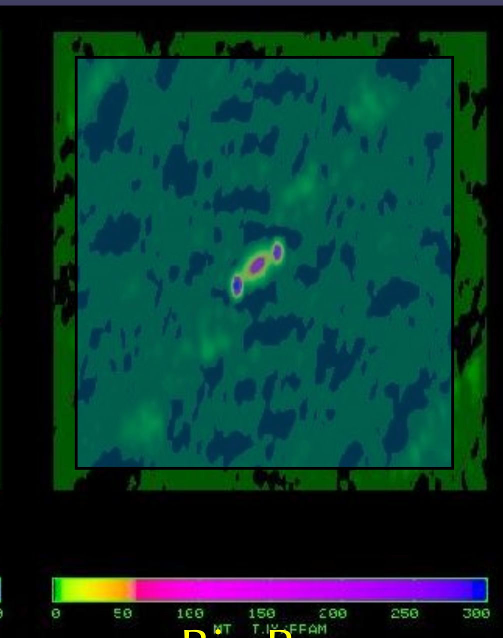
CLEANING WINDOW SENSITIVITY



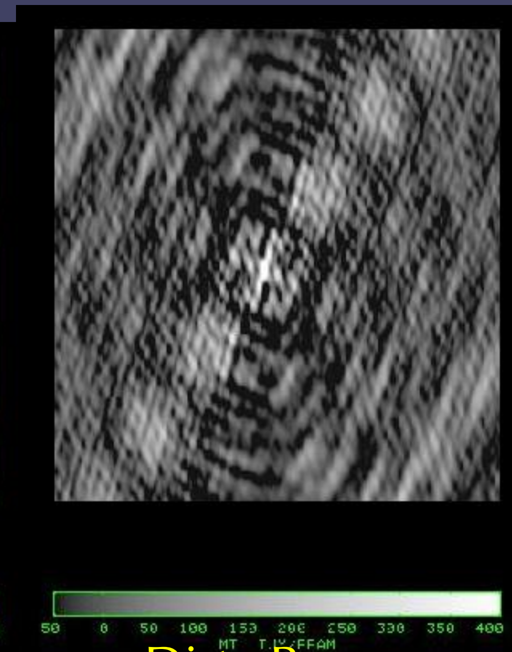
Tight Box



Middle Box



Big Box



Dirty Beam

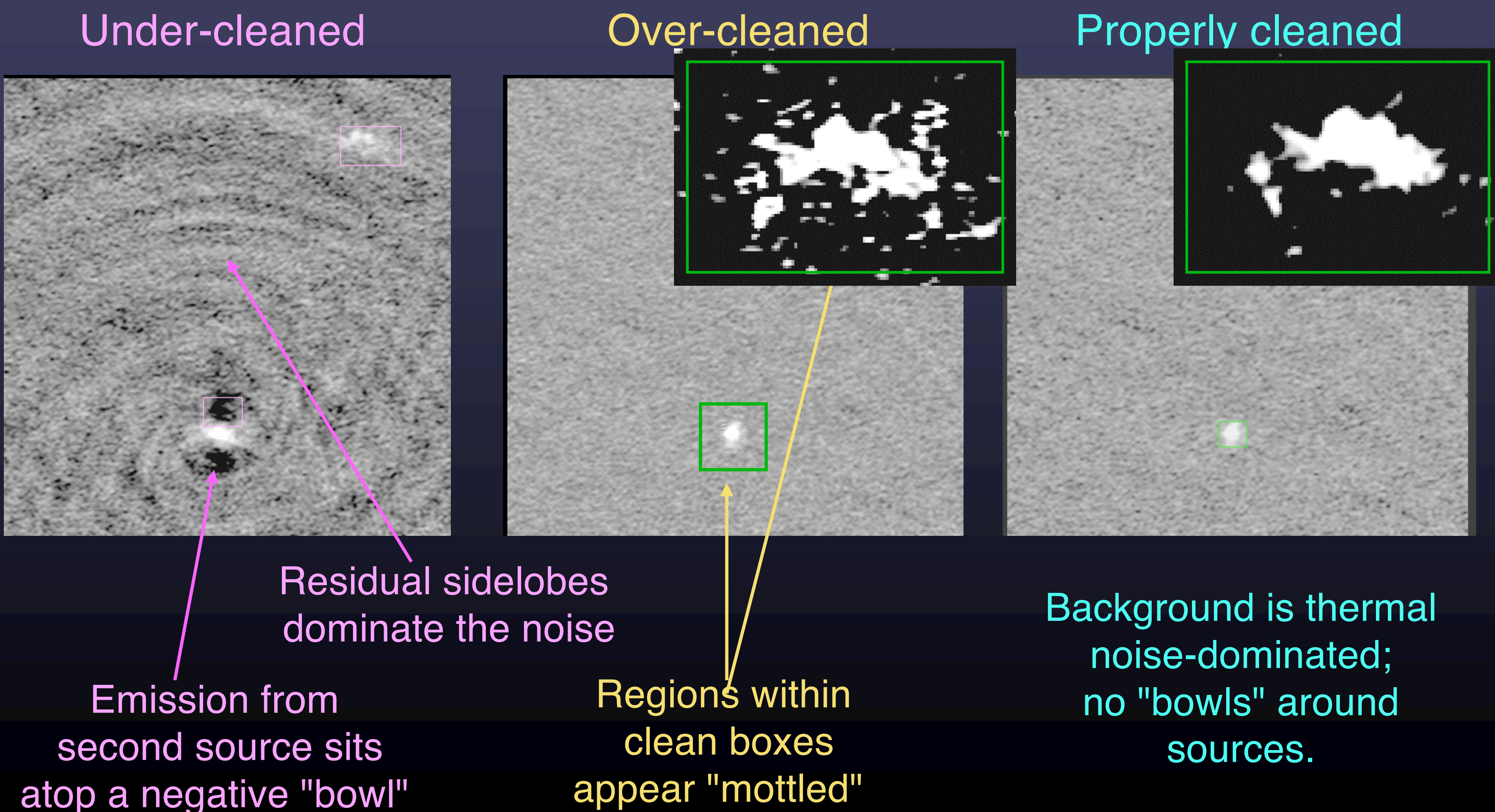
One small clean box

One clean box around all emission

Clean entire inner map quarter

Make box as small as possible to avoid cleaning noise interacting with sidelobes

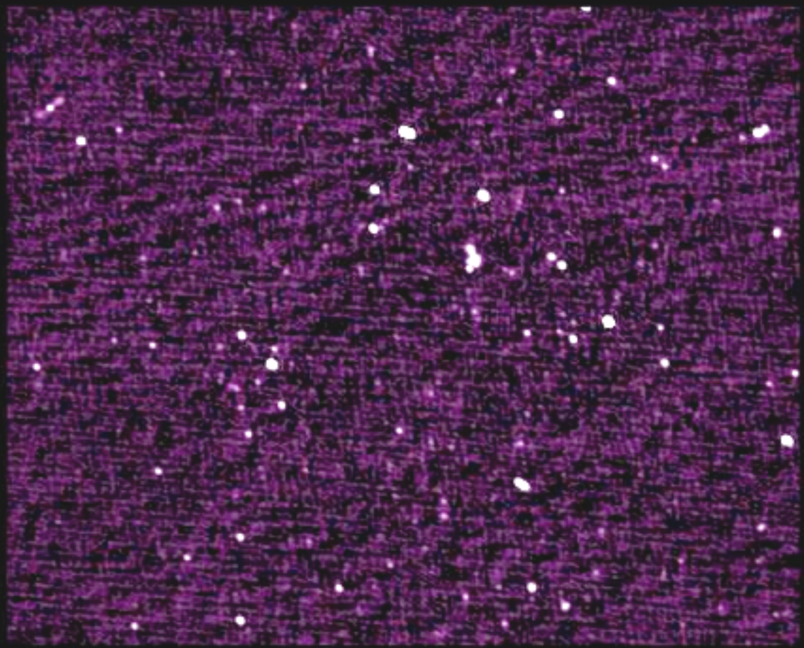
How Deep to Clean?



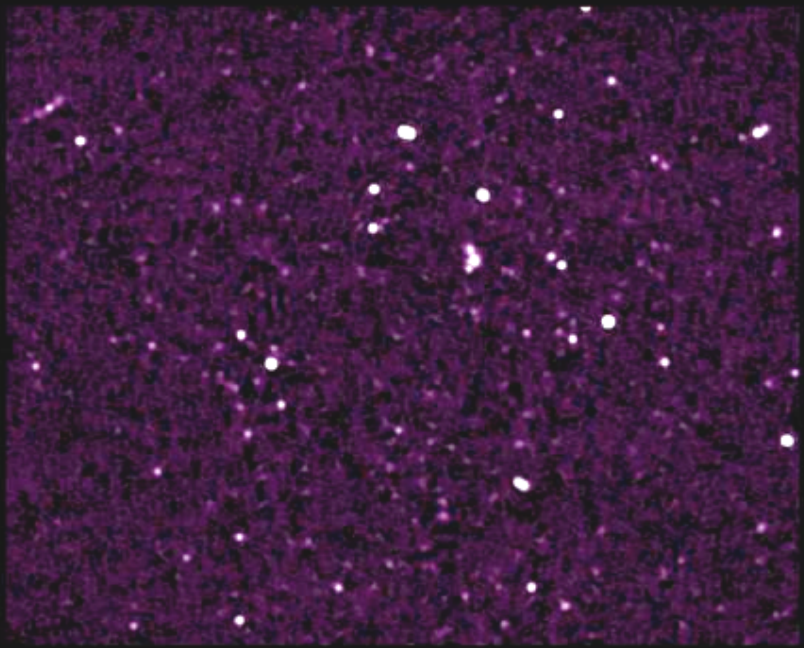
Improvement of Image

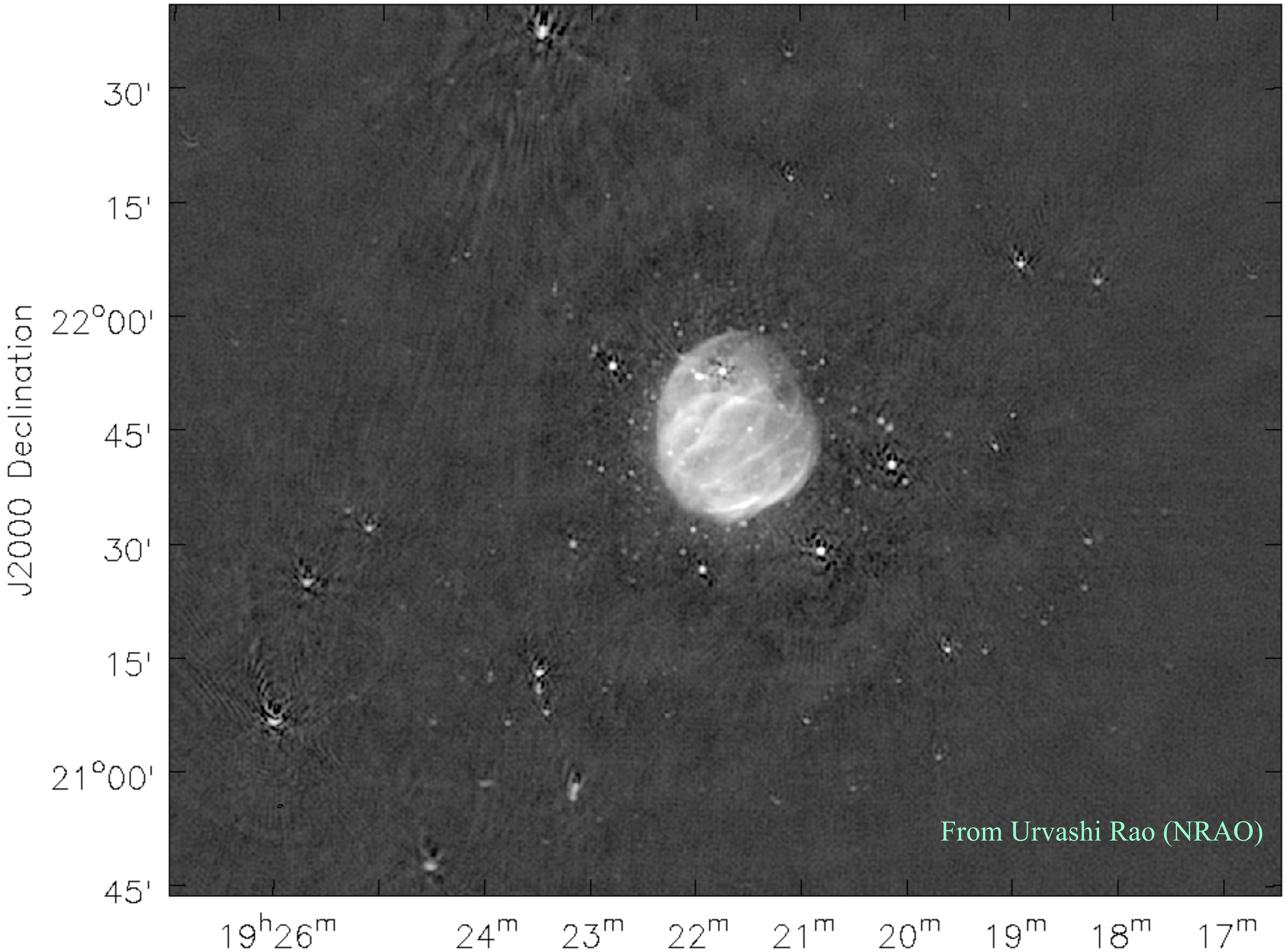
Removal of low level ripple improves detectability of faint sources

Before editing

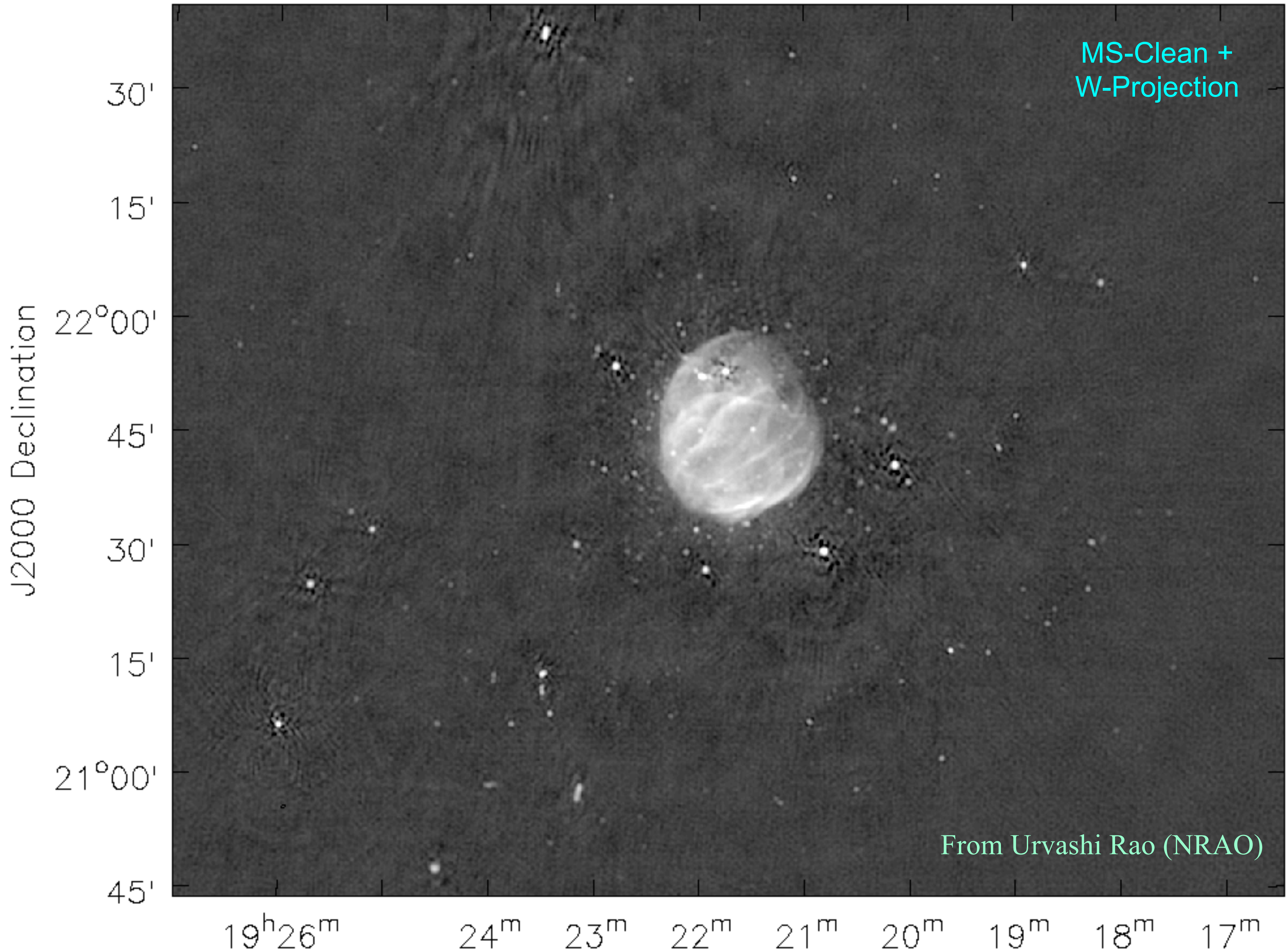


After editing

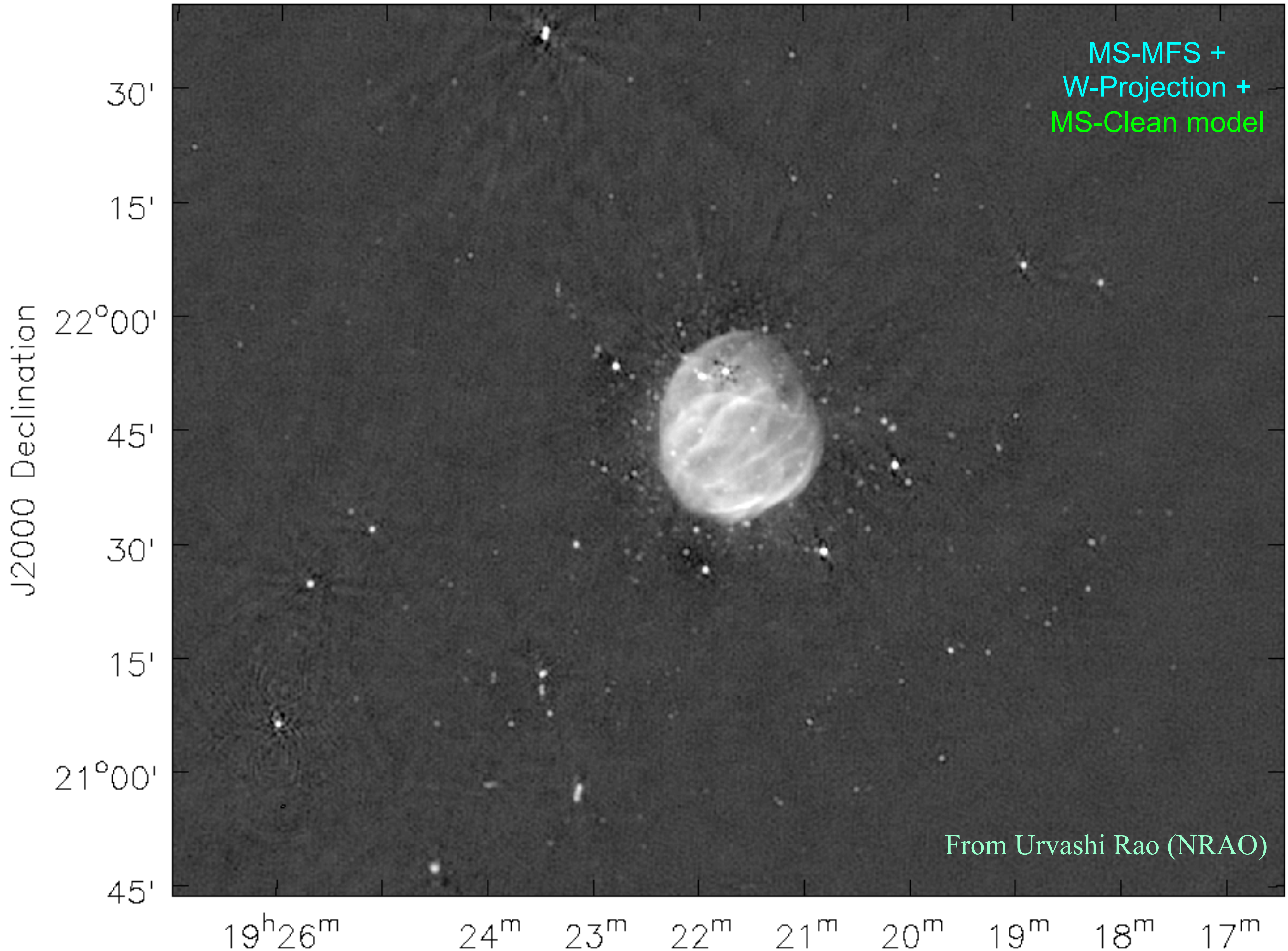




J2000 Declination
30'
15'
22°00'
45'
30'
15'
21°00'
45'
19h26m
24m
23m
22m
21m
20m
19m
18m
17m
From Urvashi Rao (NRAO)



MS-Clean +
W-Projection
From Urvashi Rao (NRAO)
J2000 Declination
30'
15'
22°00'
45'
30'
15'
21°00'
45'
19h26m
24m
23m
22m
21m
20m
19m
18m
17m


MS-MFS +
W-Projection +
MS-Clean model
From Urvashi Rao (NRAO)
J2000 Declination
30'
15'
22°00'
45'
30'
15'
21°00'
45'
19h26m
24m
23m
22m
21m
20m
19m
18m
17m

SUMMARY OF ERROR RECOGNITION

Source structure should be 'reasonable', the rms image noise as expected, and the background featureless. If not,

(u,v) data

Look for outliers in (u,v) data using several plotting methods.

Check calibration gains and phases for instabilities.

Look at residual data (u,v data - clean components)

IMAGE plane

Do defects resemble the dirty beam?

Are defect properties related to possible data errors?

Are defects related to possible deconvolution problems?

Are other corrections/calibrations needed?

Does the field-of-view encompass all emission?

IMAGE ANALYSIS

- Input: Well-calibrated data-base producing a
  high quality image
- Output: Parameterization and interpretation
  of image or a set of images

This is very open-ended

Depends on source emission complexity

Depends on the scientific goals

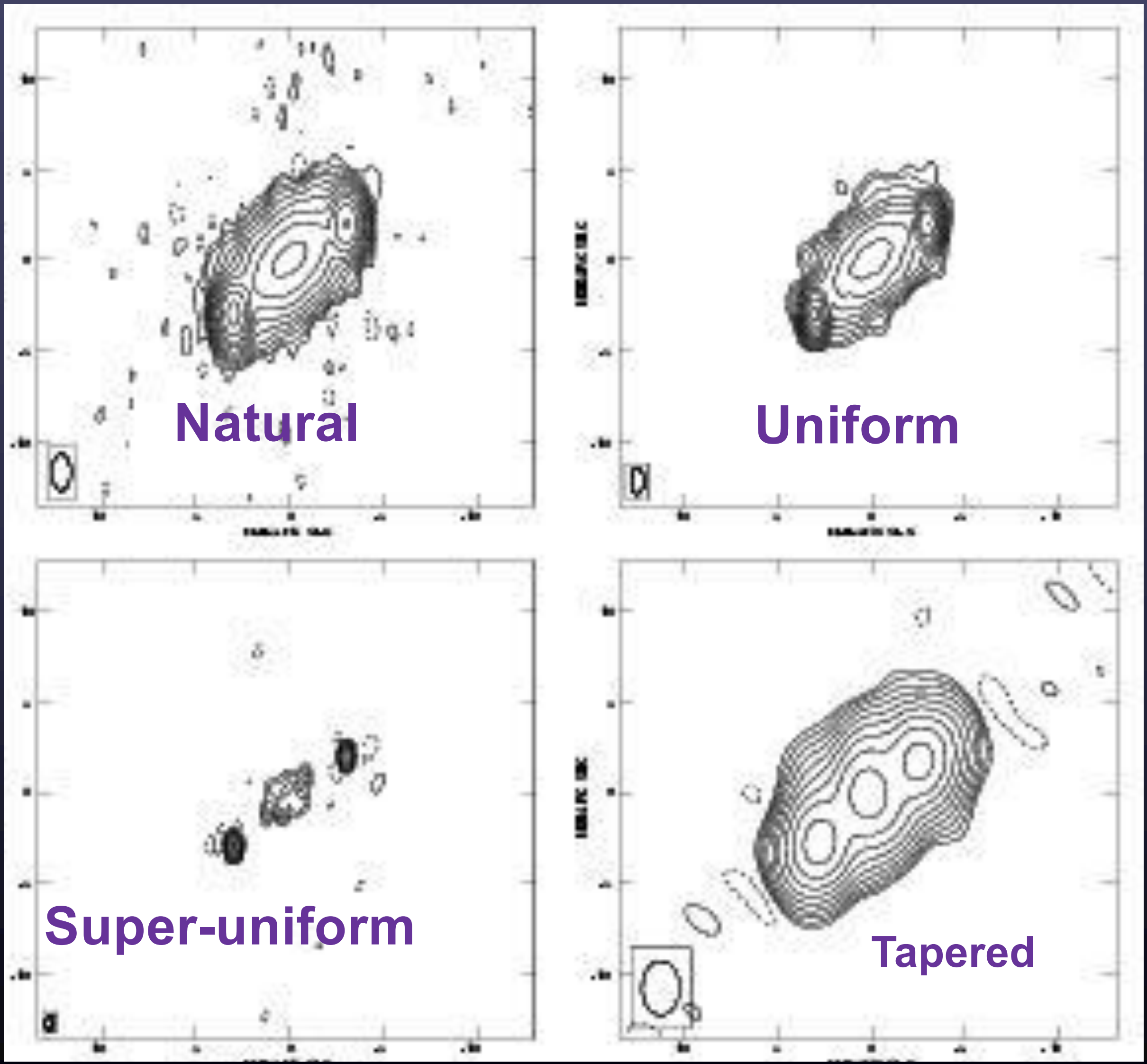
Examples and ideas are given.

Many software packages, besides AIPS
and Casa (e.g.. IDL, DS-9) are available.

IMAGE ANALYSIS OUTLINE

- Multi-Resolution of radio source.
- Parameter Estimation of Discrete Components
- Image Comparisons
- Positional Registration

IMAGE AT SEVERAL RESOLUTIONS



Milli-arcsec

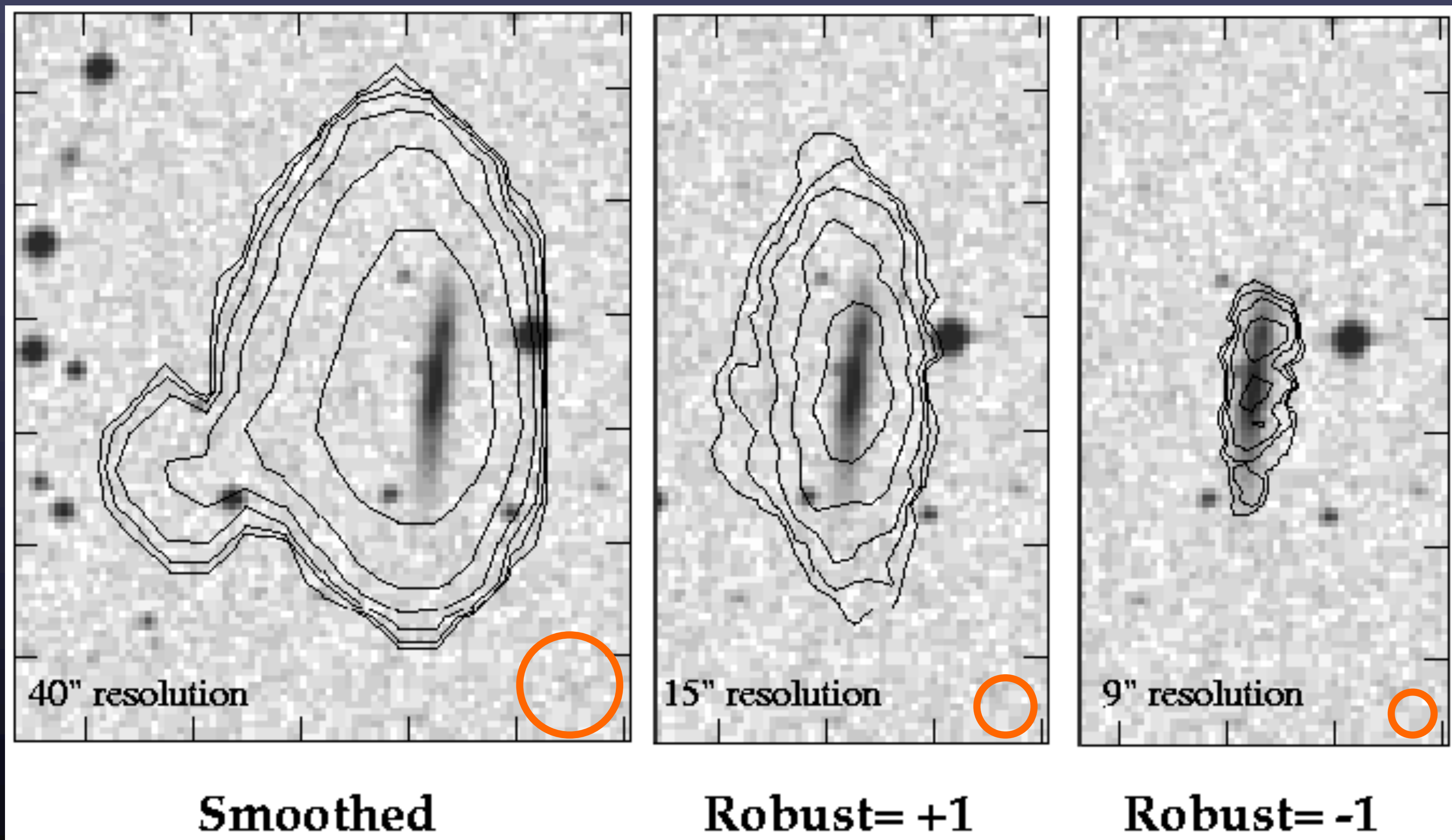
Different aspect of source structure can be see at various resolutions, shown by the ellipse in the lower left corner of each box.

SAME DATA USED FOR ALL IMAGES

For example,
Outer components are small from SU resolution
There is no extended emission from low resolution

Imaging and Deconvolution of Spectral Line Data:

Type of weighting in imaging



HI contours overlaid on optical images of an edge-on galaxy

PARAMETER ESTIMATION

Parameters associated with discrete components

- Fitting in the image
  - Assume source components are Gaussian-shaped
  - Deep cleaning restores image intensity with Gaussian-beam
  - True size * Beam size = Image size, if Gaussian-shaped. Hence, estimate of true size is relatively simple.
- Fitting in (u,v) plane (aka model-fitting)
  - Better estimates of parameters for simple sources
  - May be possible even when imaging is not
  - Can fit to more source models (e.g. Gaussian, ring, disk)
- Error estimates of parameters
  - Simple ad-hoc error estimates
  - Estimates from fitting programs
  - Monte Carlo simulations if model-fitting

IMAGE FITTING

```
Component   2-Gaussian
  Peak intensity     = 0.104 +/- 0.005 JY/BEAM
  Integral intensity= 0.998 +/- 9,47   JANSKYS
  X-position         =     255.986 +/-    0.0029 pixels
  Y-position         =     257.033 +/-    0.0032 pixels
    Major ax               19.99     +/-    0.02   pixels
    Minor ax                9.98     +/-    0.03   pixels
    Pos ang               135.3      +/-    0.1    deg
```


```
Component   1-Gaussian
  Peak intensity     = 0.300 +/- 0.005 JY/BEAM
  Integral intensity= 0.302 +/- 0.008 JANSKYS
  X-position         =     270.991 +/-    0.001 pixels
  Y-position         =     267.018 +/-    0.001 pixels
    Major ax                0.53     +/-    0.01  pixels
    Minor ax                0.00     +/-    0.05  pixels
    Pos ang                21.6      +/-    1.1   deg
```

```
Component   3-Gaussian
  Peak intensity     = 0.393 +/- 0.004 JY/BEAM
  Integral intensity= 0.403 +/- 0.008 JANSKYS
  X-position         =     241.007 +/-    0.001 pixels
  Y-position         =     241.988 +/-    0.001 pixels
    Major ax                1.54     +/-    0.01  pixels
    Minor ax                0.21     +/-    0.01  pixels
    Pos ang                 3.6      +/-    0.2   deg
```

AIPS task: JMFIT

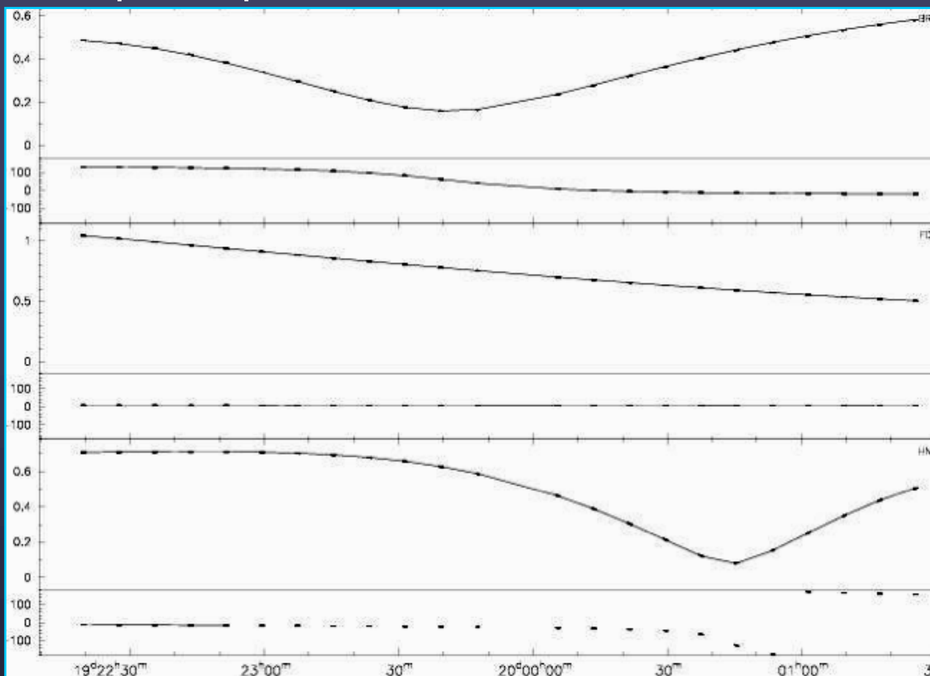
Casa tool

imfit

(u,v) DATA FITTING

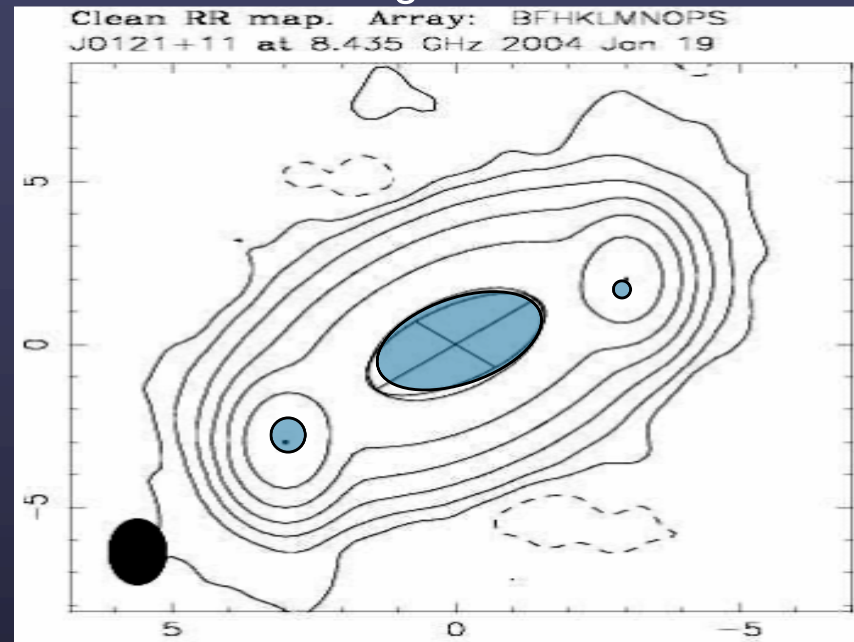
Amp and phase vs. time for three baselines

Jy
Deg
Jy
Deg
Jy
Deg



Time

Contour image with model fits



milliarcsec

milliarcsec

DIFMAP has good (u,v) fitting algorithm

Fit model directly to (u,v) data

Compare model to data

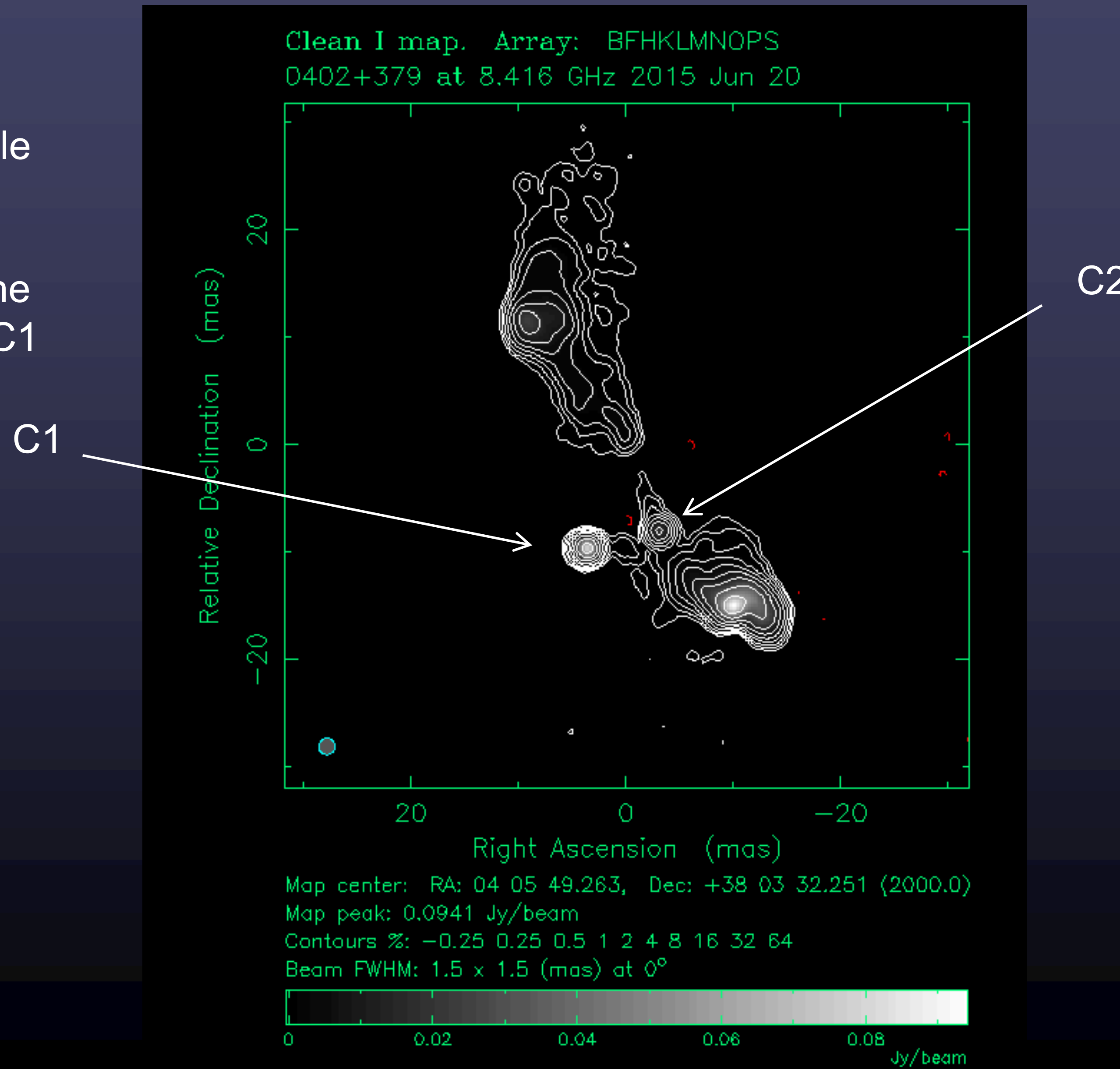
Contour display of image

Ellipses show true component size. (SNR dependent resolution)

Demo!

Supermassive Binary Black Hole Candidate 0402+379

Goal is to measure motion of the jets and the core components C1 and C2



COMPONENT ERROR ESTIMATES

P = Component Peak Flux Density

σ = Image rms noise $\qquad$ P/σ = signal/noise = ***S***

B = Synthesized beam size

θ_i = Component image size

ΔP = Peak error = σ

ΔX = Position error = $B / 2S$

$\Delta\theta_i$ = Component image size error = $B / 2S$

θ_t = True component size = $(\theta_i^2 - B^2)^{1/2}$

$\Delta\theta_t$ = Minimum component size = $B / S^{1/2}$

eg. S=100 means can determine size of B/10

Comparison and Combination of Images of Many Types

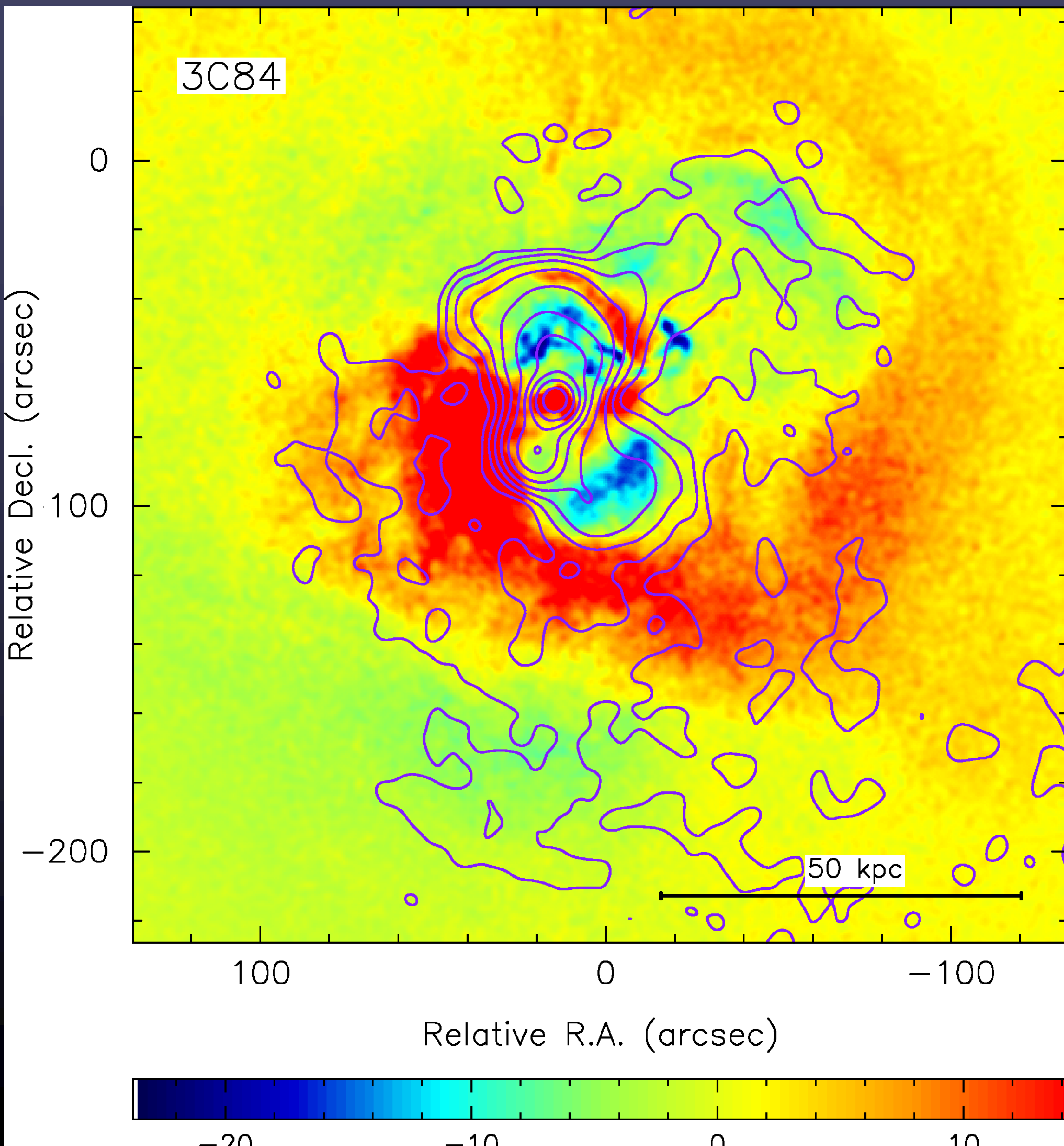
FORNAX-A Radio/Optical field

Radio is red
Faint radio core in center of NGC1316

Optical in blue-white

Frame size is 60′ x 40′

COMPARISON OF RADIO/X-RAY IMAGES



Contours of radio intensity at 1.4 GHz

Color intensity represents X-ray intensity smoothed to radio resolution

IMAGE REGISTRATION AND ACCURACY

- Separation Accuracy of Components on One Image due to residual phase errors, regardless of signal/noise:

  Limited to 1% of resolution

  Position errors of 1:10000 for wide fields, i.e. 0.1″ over 1.4 GHz PB

- Images at Different Frequencies:

  Multi-frequency. Use same calibrator for all frequencies. Watch out at frequencies < 2 GHz when ionosphere can produce displacement. Minimize calibrator-target separation

- Images at Different Times (different configuration):

  Use same calibrator for all observations. Daily troposphere changes can produce position changes up to 25% of the resolution.

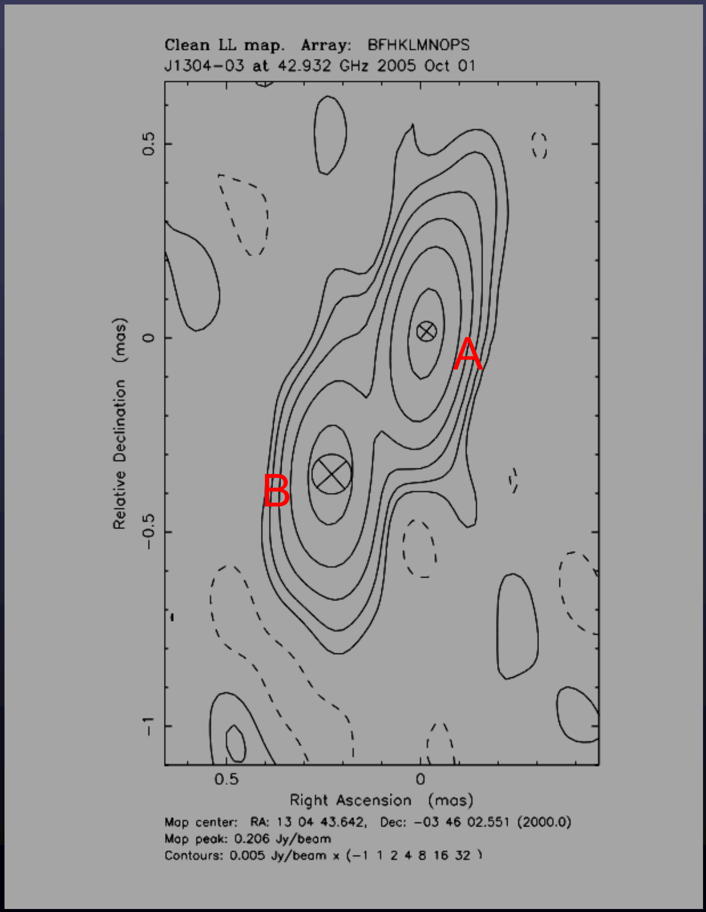
- Radio versus non-Radio Images:

  Header-information of non-radio images often much less accurate than for radio. For accuracy $<1''$, often have to align using coincident objects.

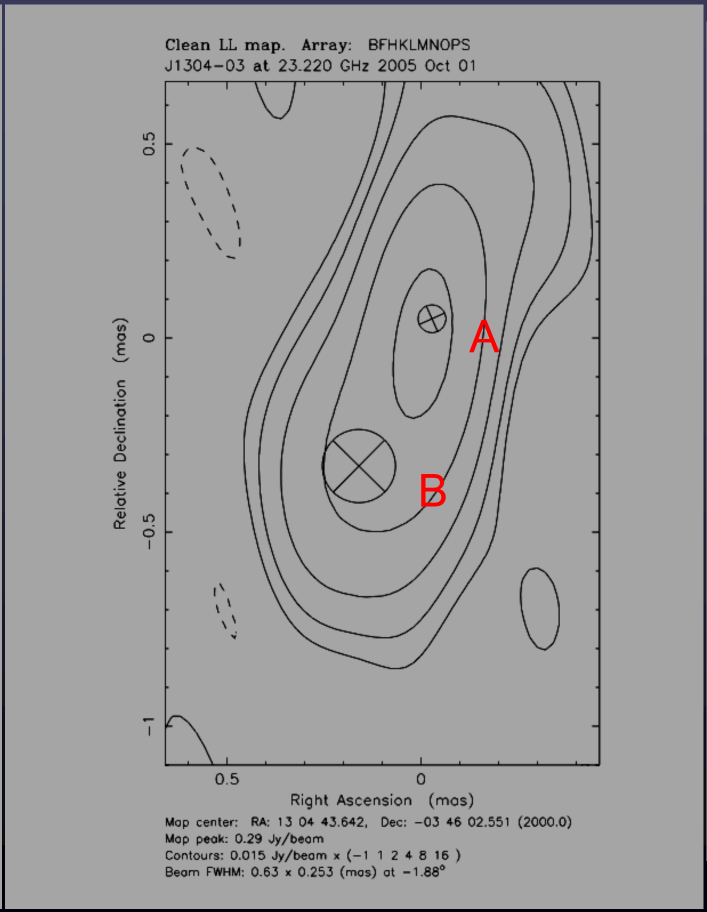
Radio Source Alignment at Different Frequencies

Self-calibration at each frequency aligns maximum at (0,0) point
Frequency-dependent structure causes relative position of maximum to change
Fitting of image with components can often lead to proper registration

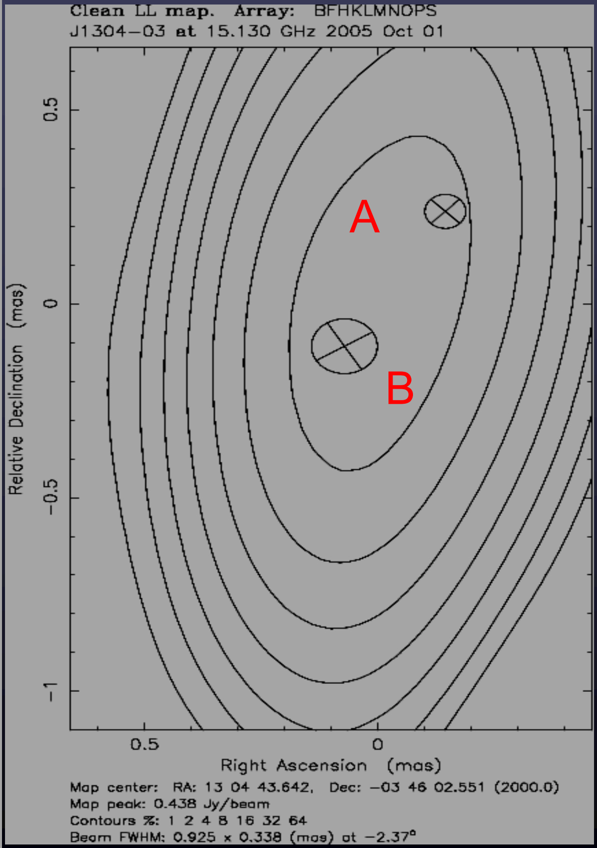
43 GHz: res = 0.3 mas



23 GHz: res = 0.6 mas



15 GHz: res = 0.8 mas



ANALYSIS: SUMMARY

- Analyze and display data in several ways
    - Adjust resolution to illuminate desired interpretation, analysis
- Parameter fitting useful, but be careful of error estimates
    - Fitting in (u,v) plane and/or image plane
- Registration of a field at different frequencies or wave-bands can be subtle.
    - Whenever possible use the same calibrator
    - May be able to align using 'known' counterparts
- Check spectral index image for artifacts

Further Reading

- http://www.nrao.edu/whatisra/
- www.nrao.edu

- 2010 Lecture on Non-Imaging Analysis
- Synthesis Imaging in Radio Astronomy
- ASP Vol 180, eds Taylor, Carilli & Perley
- Numerical Recipes, Press et al. 1992