CASA VLBI

WORKSHOP 2020

2-6 NOVEMBER 2020

LECTURE 6: TYPICAL DATA ISSUES

I. MARTI-VIDAL
(University of Valencia, SPAIN)

CASA
Common Astronomy Software Applications

JIVE
Joint Institute for VLBI ERIC

RadioNet

JUMPING JIVE
Joint Institute for VLBI ERIC

Typical Data Issues with VLBI

Identifying and Correcting Gain-related and Noise-related Problems

Ivan Martí-Vidal

Observatori Astronòmic & Dpt. Astronomia i Astrofísica
Universitat de València
(GIDEGENT Research Program, GVA)

CASA-VLBI Workshop 2020 — JIVE (Netherlands)

Typical Data Issues

- **Bad Gains.**
  - Effects on visibilities and images.
  - The "effective" PSF from bad gains. Dynamic-range limitations.
  - Amplitude vs. phase gain issues.
  - Locating problematic visibilities during deconvolution.
- **Bad Data.**
  - Sensitivity limitation.
  - Noise-dominated visibility distributions.
  - A "crazy" self-calibration.
- **Bad Sources.**
  - Bandwidth effects.

BAD GAINS

Back to the Origins

The *true* image brightness distribution, $\mathcal{I}^{\rm true}(x,y)$ is related to the visibility function, $\mathcal{V}(u,v)$ via the equation:

$$\mathcal{I}^{\rm true}(x,y) = \int_{(u,v)} \mathcal{V}(u,v)\, e^{2\pi j(ux+vy)} du\, dv$$

Back to the Origins

The *true* image brightness distribution, $\mathcal{I}^{\mathrm{true}}(x,y)$ is related to the visibility function, $\mathcal{V}(u,v)$ via the equation:

$$\mathcal{I}^{\mathrm{true}}(x,y) = \int_{(u,v)} \mathcal{V}(u,v)\, e^{2\pi j(ux+vy)} du\, dv$$

However, we only measure $\mathcal{V}(u,v)$ in a finite set of (u,v) points, a.k.a. the uv-coverage, $\mathcal{C}(u,v)$. We thus reconstruct a dirty image, $\mathcal{I}^{\mathrm{dirty}}(x,y)$:

$$\mathcal{I}^{\mathrm{dirty}}(x,y) = \int_{(u,v)} \mathcal{V}(u,v)\, \mathcal{C}(u,v)\, e^{2\pi j(ux+vy)} du\, dv$$

Back to the Origins

The *true* image brightness distribution, $\mathcal{I}^{\rm true}(x,y)$ is related to the visibility function, $\mathcal{V}(u,v)$ via the equation:

$$\mathcal{I}^{\rm true}(x,y) = \int_{(u,v)} \mathcal{V}(u,v)\, e^{2\pi j(ux+vy)} du\, dv$$

However, we only measure $\mathcal{V}(u,v)$ in a finite set of (u,v) points, a.k.a. the uv-coverage, $\mathcal{C}(u,v)$. We thus reconstruct a dirty image, $\mathcal{I}^{\rm dirty}(x,y)$:

$$\mathcal{I}^{\rm dirty}(x,y) = \int_{(u,v)} \mathcal{V}(u,v)\, \mathcal{C}(u,v)\, e^{2\pi j(ux+vy)} du\, dv$$

Using the well-known *Convolution Theorem*: $\mathcal{I}^{\rm dirty}(x,y) = \mathcal{I}^{\rm true}(x,y) * B(x,y)$

$$\text{where } B(x,y) = \int_{(u,v)} \mathcal{C}(u,v)\, e^{2\pi j(ux+vy)} du\, dv$$

The Origins Revisited

We actually measure a somewhat *corrupted* version of $\mathcal{V}^{\rm obs}(u,v)$:

$$\mathcal{V}^{\rm obs}(u,v) = \mathcal{V}(u,v) \times (1 + \mathcal{G}(u,v))$$

Hence, the recovered image is:

$$\mathcal{I}^{\rm obs}(x,y) = \int_{(u,v)} \mathcal{V}(u,v)\,\mathcal{C}(u,v)(1 + \mathcal{G}(u,v))\, e^{2\pi j(ux+vy)} du\, dv$$

The Origins Revisited

We actually measure a somewhat *corrupted* version of $\mathcal{V}^{\mathrm{obs}}(u,v)$:

$$\mathcal{V}^{\mathrm{obs}}(u,v) = \mathcal{V}(u,v) \times (1 + \mathcal{G}(u,v))$$

Hence, the recovered image is:

$$\mathcal{I}^{\mathrm{obs}}(x,y) = \int_{(u,v)} \mathcal{V}(u,v)\,\mathcal{C}(u,v)(1 + \mathcal{G}(u,v))\, e^{2\pi j(ux+vy)} du\, dv$$

From the *Convolution Theorem*, we get: $\mathcal{I}^{\mathrm{obs}}(x,y) = \mathcal{I}^{\mathrm{true}}(x,y) * (B(x,y) + G(x,y))$

$$\text{where } G(x,y) = \int_{(u,v)} \mathcal{C}(u,v)\,\mathcal{G}(u,v)\, e^{2\pi j(ux+vy)} du\, dv$$

The Gain Problem

The images $\mathcal{I}^{\rm obs}(x,y)$ and $\mathcal{I}^{\rm dirty}(x,y)$ are convolved with different PSFs.
However, we use the PSF of $\mathcal{I}^{\rm dirty}(x,y)$ to deconvolve $\mathcal{I}^{\rm obs}(x,y)$.
In other words, **we use the wrong point-spread function** for the deconvolution!:

$$\mathcal{I}^{\rm obs}(x,y) = \mathcal{I}^{\rm true}(x,y) * \left(B(x,y) + G(x,y)\right) \quad ; \quad \mathcal{I}^{\rm dirty}(x,y) = \mathcal{I}^{\rm true}(x,y) * B(x,y)$$

The Gain Problem

The images $\mathcal{I}^{\mathrm{obs}}(x,y)$ and $\mathcal{I}^{\mathrm{dirty}}(x,y)$ are convolved with different PSFs.
However, we use the PSF of $\mathcal{I}^{\mathrm{dirty}}(x,y)$ to deconvolve $\mathcal{I}^{\mathrm{obs}}(x,y)$.
In other words, **we use the wrong point-spread function** for the deconvolution!:

$$\mathcal{I}^{\mathrm{obs}}(x,y) = \mathcal{I}^{\mathrm{true}}(x,y) * (B(x,y) + G(x,y)) \quad ; \quad \mathcal{I}^{\mathrm{dirty}}(x,y) = \mathcal{I}^{\mathrm{true}}(x,y) * B(x,y)$$

Wrong antenna gains have a convolution-like effect in the image plane. The true convolving PSF and the one computed from the uv-coverage are different.

The Gain Problem

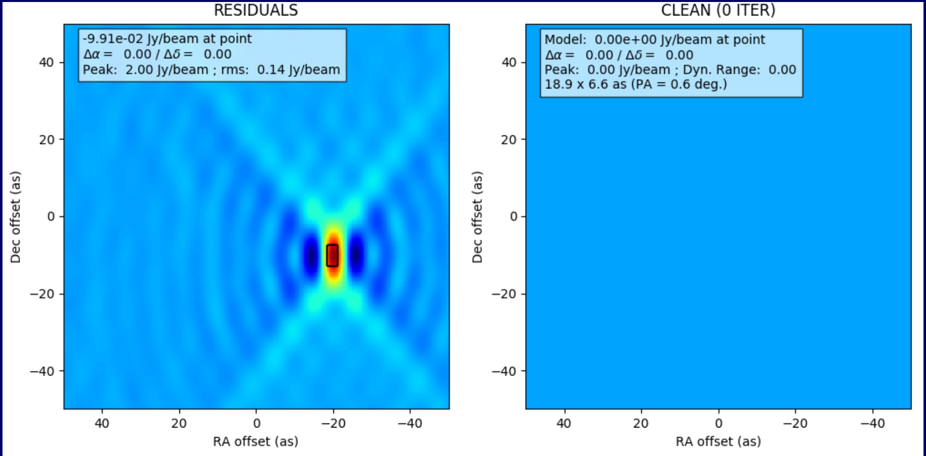
The images $\mathcal{I}^{\rm obs}(x,y)$ and $\mathcal{I}^{\rm dirty}(x,y)$ are convolved with different PSFs.
However, we use the PSF of $\mathcal{I}^{\rm dirty}(x,y)$ to deconvolve $\mathcal{I}^{\rm obs}(x,y)$.
In other words, **we use the wrong point-spread function** for the deconvolution!:

$$\mathcal{I}^{\rm obs}(x,y) = \mathcal{I}^{\rm true}(x,y) * (B(x,y) + G(x,y)) \quad ; \quad \mathcal{I}^{\rm dirty}(x,y) = \mathcal{I}^{\rm true}(x,y) * B(x,y)$$

Wrong antenna gains have a convolution-like effect in the image plane. The true convolving PSF and the one computed from the uv-coverage are different.

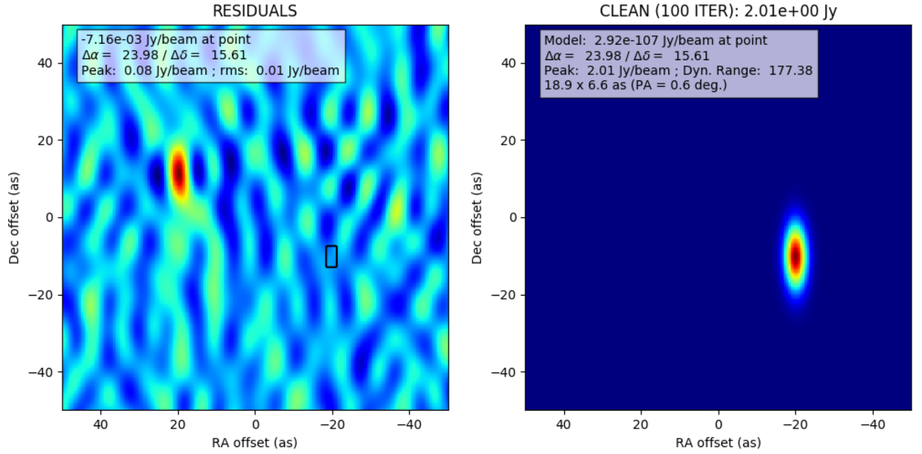
This introduces dynamic-range limitations in the image (i.e., the noise is proportional to the image peak, hence limiting the achievable contrast).

Gains and Dynamic Range



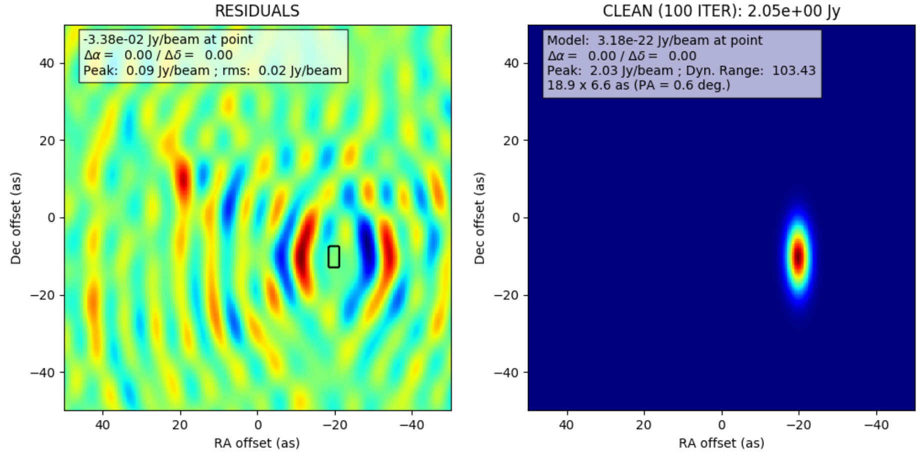
Point source with weak companion (CLEANing perfectly-calibrated data).

Gains and Dynamic Range



Point source with weak companion (CLEANing perfectly-calibrated data).

Gains and Dynamic Range



Point source with weak companion (CLEANing data with calibration issues).

Gain Effects in the Image Plane

How do $\mathcal{G}(u,v)$ and $G(x,y)$ behave?

- Amplitude gain: $\mathcal{G}(u,v)$ is real-valued and even. Hence, $G(x,y)$ is even.

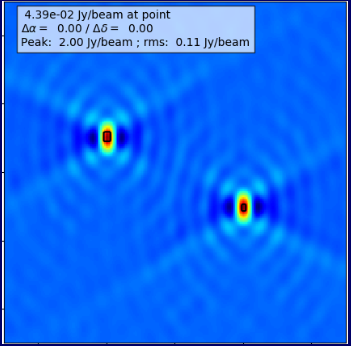
Gain Effects in the Image Plane

How do $\mathcal{G}(u,v)$ and $G(x,y)$ behave?

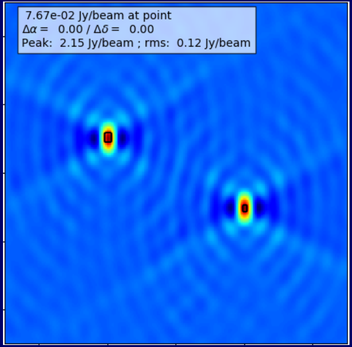
- Amplitude gain: $\mathcal{G}(u,v)$ is real-valued and even. Hence, $G(x,y)$ is even.

EXAMPLE: 6h with WSRT at 45° Declination (done with APSYNSIM):

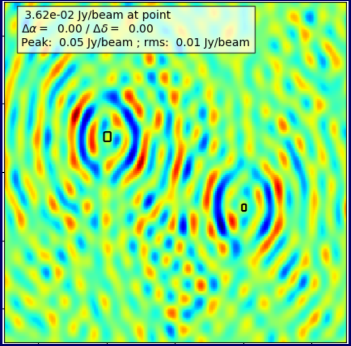
Perfect calibration



50% Amplitude in one antenna



CLEAN residuals



Gain Effects in the Image Plane

How do $\mathcal{G}(u,v)$ and $G(x,y)$ behave?

- Small Phase gain: $\mathcal{G}(u,v)$ is imaginary and odd. Hence, $G(x,y)$ is odd.

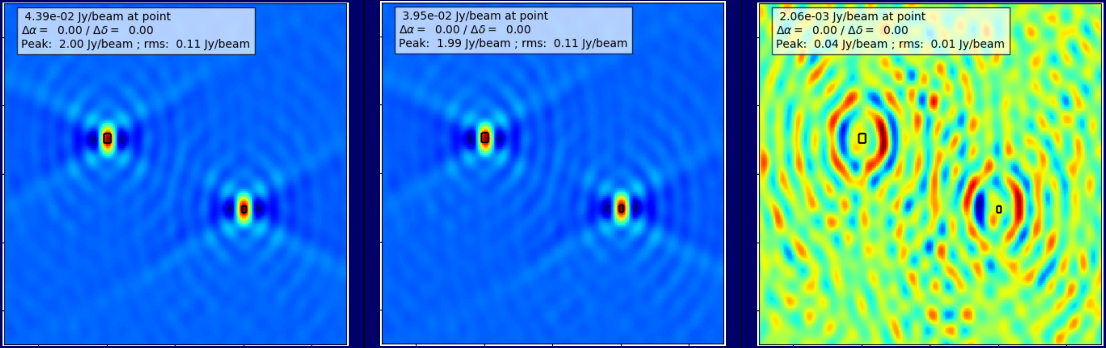
Gain Effects in the Image Plane

How do $\mathcal{G}(u,v)$ and $G(x,y)$ behave?

- Small Phase gain: $\mathcal{G}(u,v)$ is imaginary and odd. Hence, $G(x,y)$ is odd.

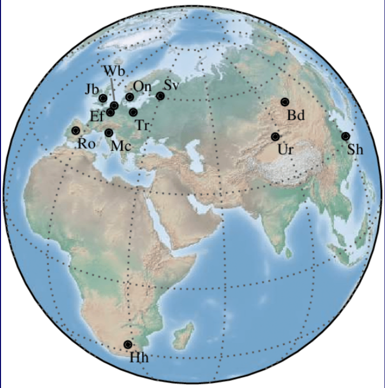
EXAMPLE: 6h with WSRT at 45° Declination (done with APSYNSIM):

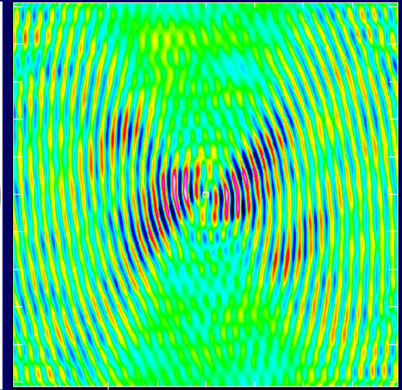
Perfect calibration	20° Phase in one antenna	CLEAN residuals



Quizzzz

Which EVN baseline and gain dominates these residuals?

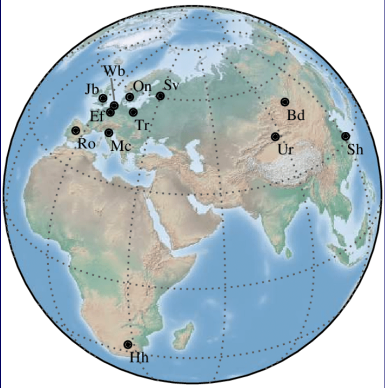


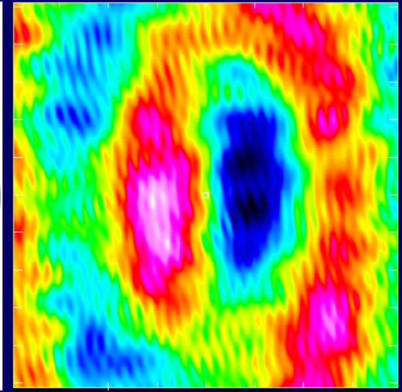


- **A.** Jb - On. Amplitude gain.
- **B.** Jb - Ro. Phase gain.
- **C.** Sh - Ro. Amplitude gain.
- **D.** Hh - Mc. Phase gain.

Quizzzzz

Which EVN baseline and gain dominates these residuals?

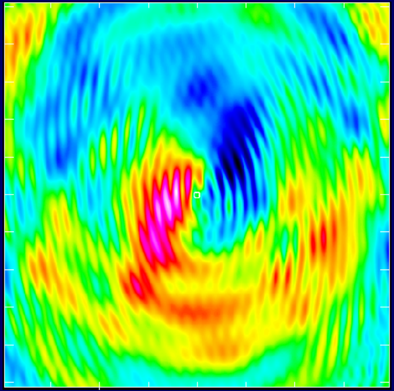




- **A.** Jb - Wb. Amplitude gain.
- **B.** Ro - Tr. Amplitude gain.
- **C.** Jb - Wb. Phase gain.
- **D.** Hh - Ur. Phase gain.

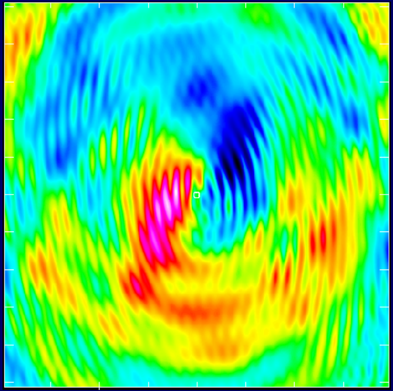
Reality is usually worse

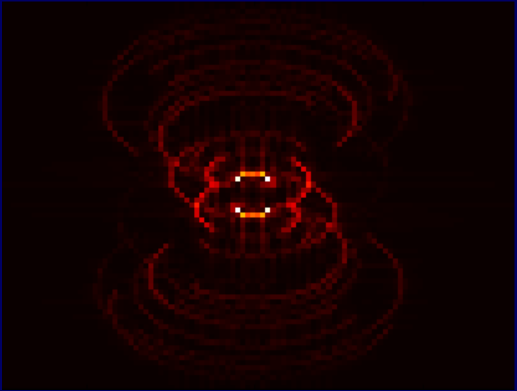
How can we identify the bad gains in the general case?



Reality is usually worse

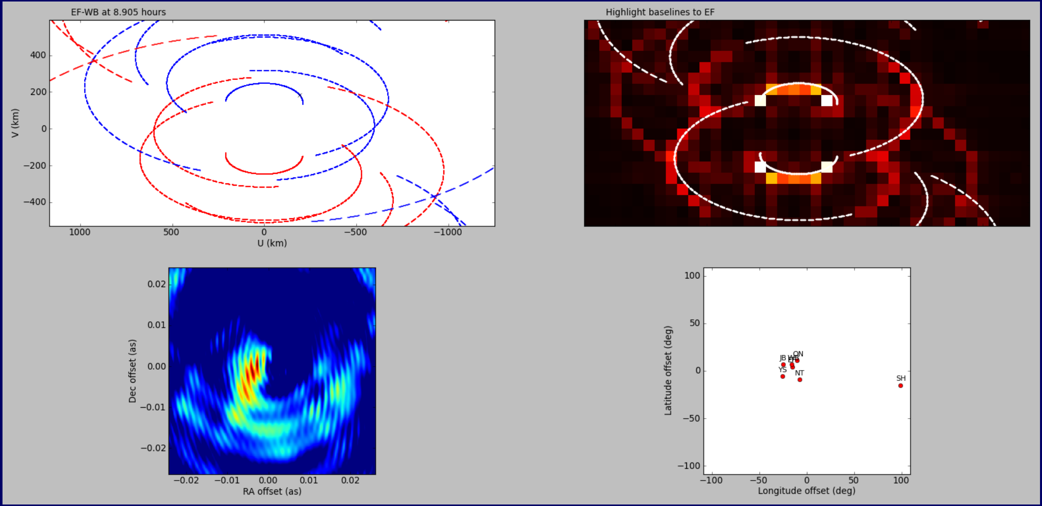
How can we identify the bad gains in the general case?



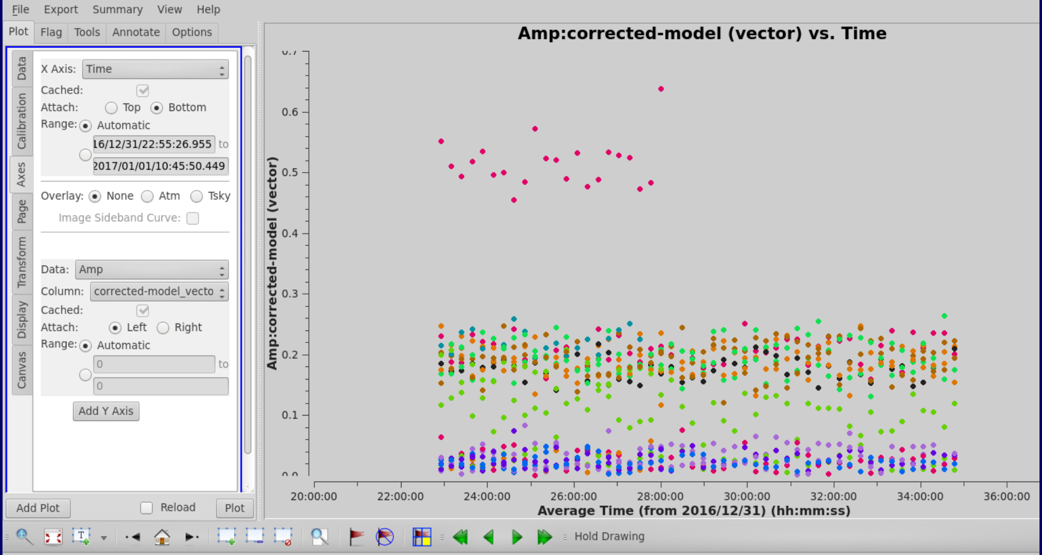


The residual image encodes the spatial frequencies of $G(x,y)$!

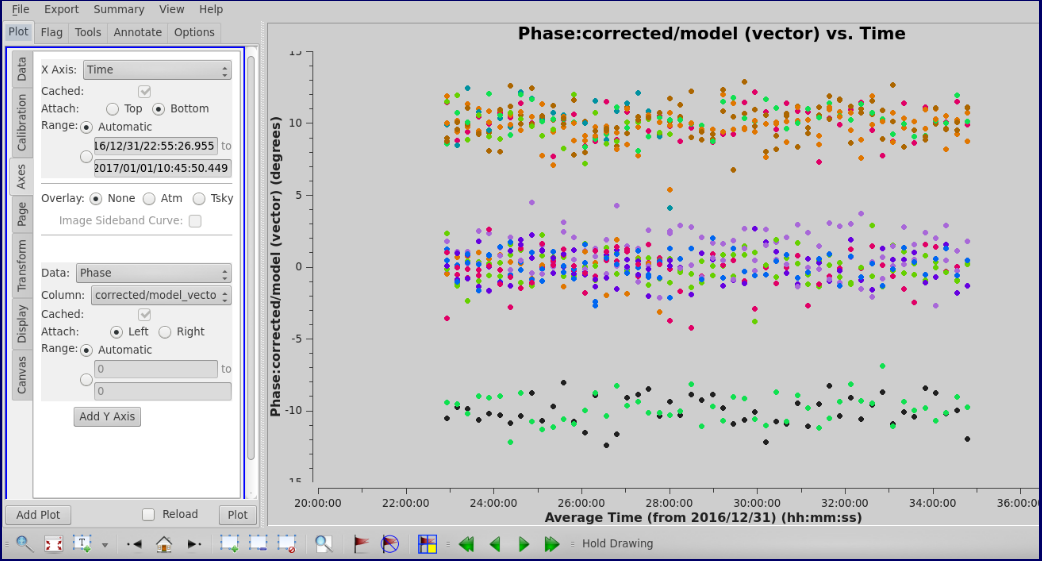
The “checkres” task (Nordic ARC Node)



The plotms task



The plotms task



So, what do we do?

So, what do we do?

SELF-CALIBRATION! Iterative hybrid imaging!

So, what do we do?

SELF-CALIBRATION! Iterative hybrid imaging!

But remember: *with great power comes great responsibility*

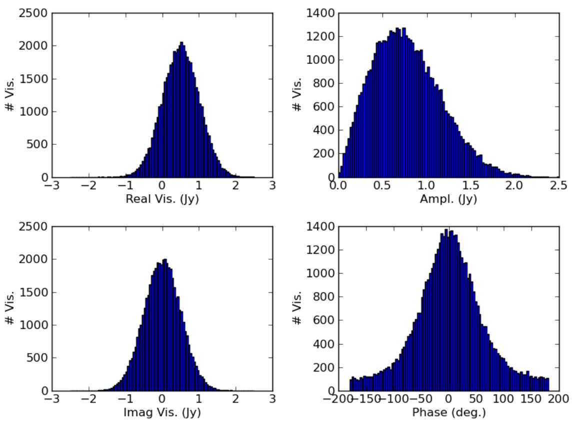
(see next slides)

BAD DATA

Bad data is worse than no data at all.
Alan Marsher

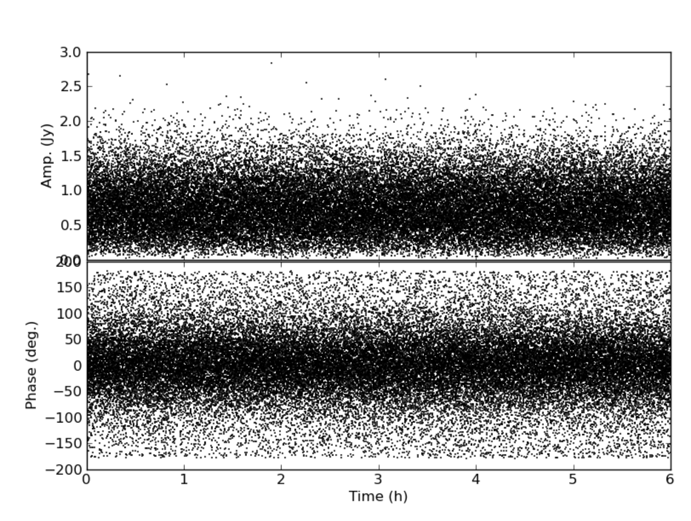
Noise in the visibilities

- Assuming a correct calibration, the real and imaginary parts of the visibilities have Gaussian and independent thermal noise.
- Hence, the amplitudes and phase do **NOT** have neither Gaussian nor independent noises.



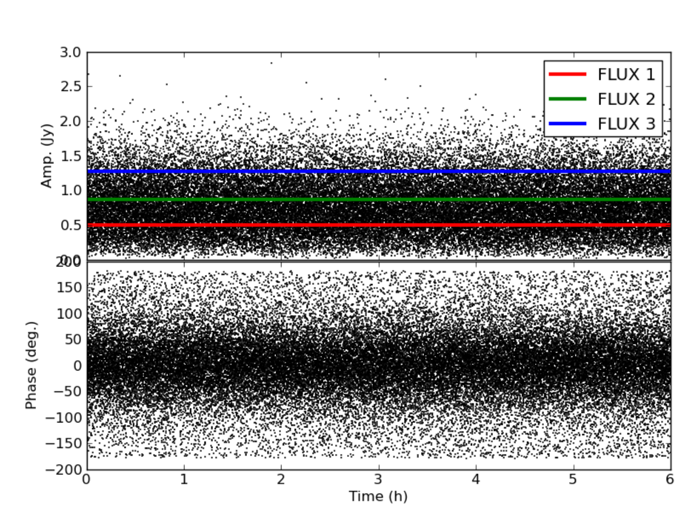
Quizzzzz

What is the flux density of the point source in these observations?

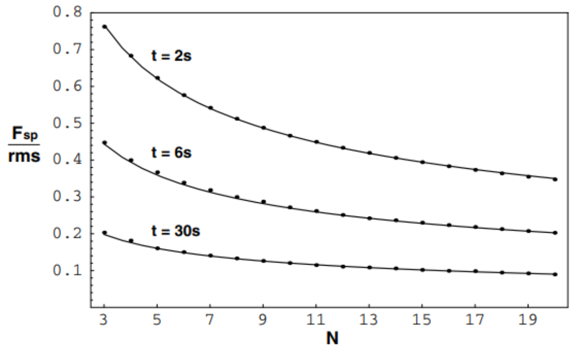


Quizzzzz

What is the flux density of the point source in these observations?



Beware of noisy selfcal!

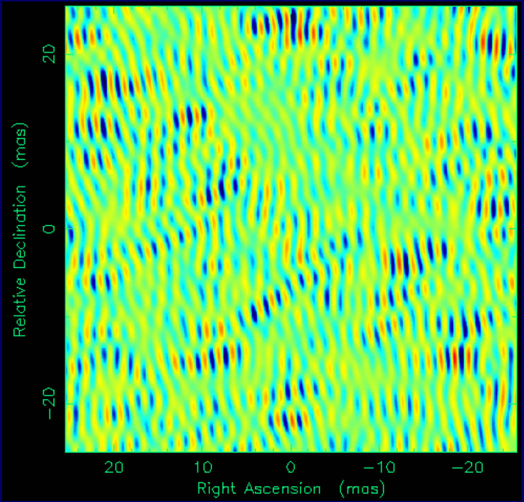


Self-calibrating on noisy data can create spurious sources with a flux density, F_{sp}, similar to the rms of the visibilities, ρ! (e.g., Martí-Vidal & Marcaide 2008).

$$F_{sp} \sim 0.7\rho\sqrt{\frac{t_{\mathrm{int}}}{t_{\mathrm{sol}}}}N_{\mathrm{ant}}^{-0.4}$$

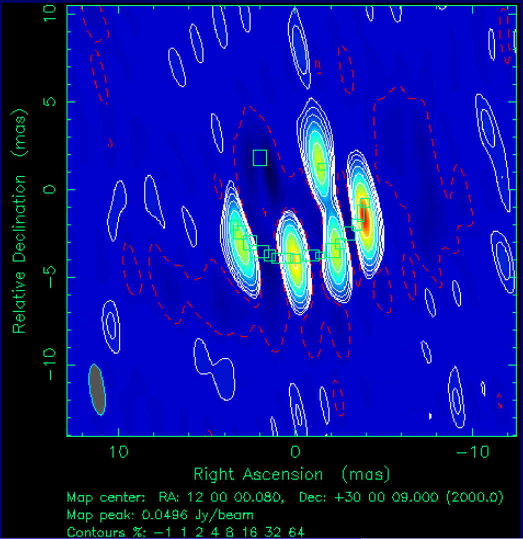
A crazy self-calibration

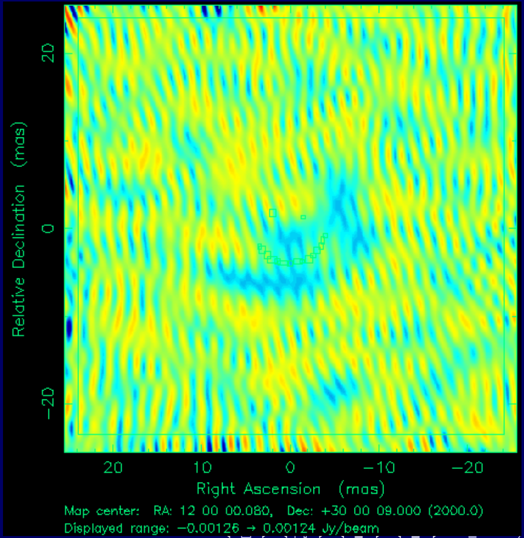
Pure thermal noise observed with the EVN (30° declination, 6h with optimum elevations).



A crazy self-calibration

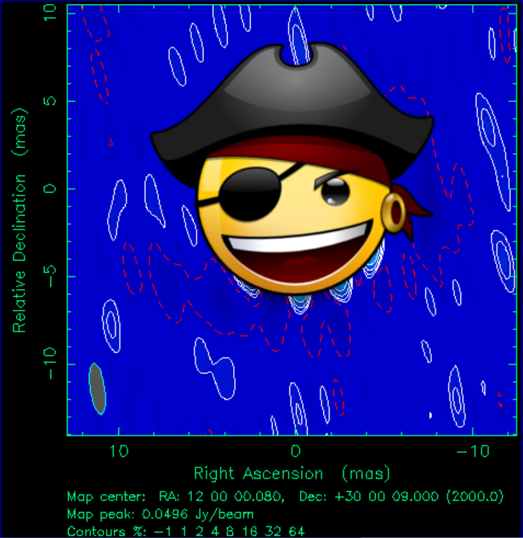
Only phase self-calibration with an a-priori model. Pretty “harmless”, ain’t it?

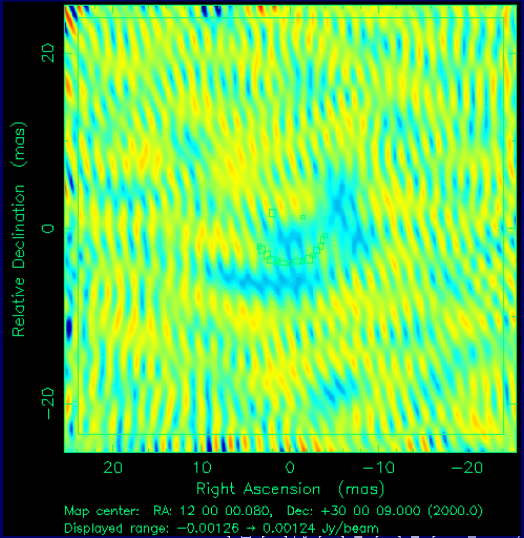




A crazy self-calibration

Only phase self-calibration with an a-priori model. Pretty “harmless”, ain’t it?





BAD SOURCES

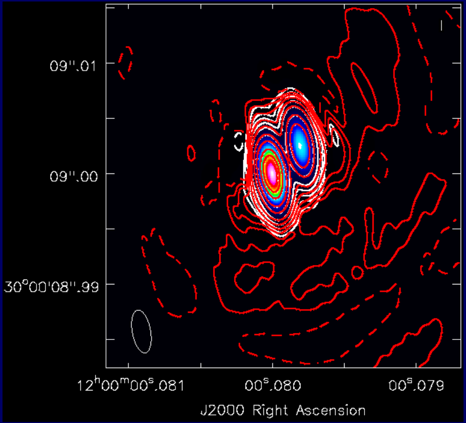
Increasing the SNR by averaging

- We can do averaging either in time or frequency (or both).
- Averaging in uv-space is a *convolution-like* operation (substitutes tracks and lines by points).
- From the convolution theorem, the effects in the image plane are "product-like".
  - Time Smearing and Bandwidth Smearing (baseline-dependent smearing of source components far from the phase center).

$$\mathrm{FoV(arcsec)} = \max\left[11\left(\frac{t_{\mathrm{int}}}{1\,\mathrm{s}}\right)^{-1}\;,\;4.8\left(\frac{\Delta\nu}{1\,\mathrm{MHz}}\right)^{-1}\right]\times\left(\frac{B_{\mathrm{max}}}{10^4\,\mathrm{km}}\right)^{-1}$$

Source frequency dependence

- For large bandwidths, the source's spectral distribution can limit the dynamic range.



Double source with different spectral indices ($\alpha = 0.0$ for the core; $\alpha = -1.0$ for the extension). In **red**, CLEAN in normal mode; In **white**, CLEAN in multi-frequency-synthesis (mfs) mode.

SUMMARY

- Images are limited by uv-coverage and antenna gains. Both have convolution-like effects.
- If there is an incorrect calibration, it is possible to identify the bad gains from the residuals:
  - Find out the bad spatial frequencies in the residual image and identify the bad points (e.g., the "checkres" task).
  - Check the residual visibilities and look for outliers: phases in data/model or amplitudes in data−model (the "plotms" task).
- Once the bad data are identified, we can correct them with self-calibration. BUT, special care has to be taken with noisy data.
- Even if the data are perfectly calibrated, we can still introduce artifacts, e.g.:
  - Smearing effects from visibility averaging.
  - Dynamic-range limitation by source spectral distribution in wideband observations.

THANKS TO OUR SPONSORS:

CASA VLBI

JUMPING JIVE
Joint Institute for VLBI
ERIC

RadioNet

THIS EVENT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENTS 730562 (RADIONET) AND 7308844 (JUMPING JIVE)