Dealing with Noise

Frederic Gueth, IRAM
Stephane Guilloteau, LAB

Institut de Radioastronomie Millimétrique

iram Institut de Radioastronomie Millimétrique

Outline

Noise & Sensitivities

- Point source sensitivity
- Noise in images
- Brightness sensitivity

Low S/N analysis

- Continuum data
- Line data
- Examples

System Temperature

- Power are expressed in temperatures: $P = k\,T\,\Delta\nu$
- System temperature (= noise)

$T_{ant} = T_{bg}$	cosmic background
$+ T_{sky} \approx \eta_f (1-\exp(-\tau_{atm}) T_{atm}$	sky noise
$+ T_{spill} \approx (1-\eta_f-\eta_{loss}) T_{ground}$	ground noise pickup
$+ T_{loss} \approx \eta_{loss} T_{cabin}$	losses in receiver cabin
$+ T_{rec}$	receiver noise

- Antenna temperature (=source) T_A is the temperature of the equivalent blackbody seen by the antenna (in the Rayleigh Jeans approximation)

System Temperature

- We usually refer the temperatures to a perfect antenna located outside the atmosphere, and single sideband signal:

$$T_{sys} = (1+g)\ e^{\tau_{atm}}\ T_{ant} / \eta_f$$

$$T_A^* = (1+g)\ e^{\tau_{atm}}\ T_A / \eta_f$$

- This antenna temperature T_A^* is weather-independent, and linked to the source flux S by an antenna-dependent quantity only

$$T_A^* = \frac{\eta_a A}{2k} S$$

Noise Equation

- The noise power is T_{sys} and there are $2\ \Delta\nu\ \Delta t$ independent samples to measure a correlation, so the noise is

$$\delta T = \frac{T_{sys}}{\sqrt{2\ \Delta t\ \Delta\nu}}$$

- In terms of flux: $$\delta S = \frac{\sqrt{2}k}{\eta_a A} \frac{T_{sys}}{\sqrt{\Delta t\ \Delta\nu}}$$

- Note: this is $\sqrt{2}$ worse than that of an antenna with the same total collecting area → this sensitivity loss is because we ignore the autocorrelations

Noise Equation

- Noise on one visibility (with efficiencies):

$$\delta S = \frac{\sqrt{2}k}{\eta_a \eta_q \eta_J \eta_P A} \frac{T_{sys}}{\sqrt{\Delta t \, \Delta \nu}}$$

- Noise is uncorrelated from one baseline to another
- There are n(n-1)/2 baselines for n antennas
- So the point source sensitivity is

$$\delta S = \frac{2k}{\eta A} \frac{T_{sys}}{\sqrt{N(N-1) \, t_{int} \, \Delta \nu}}$$

Noise Equation

- Noise on one visibility (with efficiencies):

$$\delta S = \frac{\sqrt{2}k}{\eta_a \eta_q \eta_J \eta_P A} \frac{T_{sys}}{\sqrt{\Delta t}}$$

Average of all visibilities to detect a point source

But we are doing a map, ie a Fourier Transform…

- Noise is uncorrelated from one base
- There are n(n-1)/2 baselines for n a
- So the point source sensitivity is

$$\delta S = \frac{2k}{\eta A} \frac{T_{sys}}{\sqrt{N(N-1)\, t_{int}\, \Delta\nu}}$$

Noise in Images

- The Fourier Transform is a linear combination of the visibilities with some rotation (phase factor) applied. How do we derive the noise in the image from that on the visibilities ?

- Noise on visibilities
  - the correlator gives the same noise (variance) on the real and imaginary part of the complex visibility $\langle \varepsilon_r^2 \rangle = \langle \varepsilon_i^2 \rangle$
  - Real and Imaginary are uncorrelated $\langle \varepsilon_r \varepsilon_i \rangle = 0$

- So rotation (phase factor) has NO effect on noise

$$\varepsilon'_{\mathrm{R}} = \varepsilon_{\mathrm{R}} \cos(\phi) - \varepsilon_{\mathrm{I}} \sin(\phi)$$

$$\varepsilon'_{\mathrm{I}} = \varepsilon_{\mathrm{R}} \sin(\phi) + \varepsilon_{\mathrm{I}} \cos(\phi)$$

$$\langle {\varepsilon'_{\mathrm{R}}}^2 \rangle = \langle \varepsilon_{\mathrm{R}}^2 \rangle \cos^2(\phi) - 2 \langle \varepsilon_{\mathrm{R}} \varepsilon_{\mathrm{I}} \rangle \cos(\phi) \sin(\phi) + \langle \varepsilon_{\mathrm{I}}^2 \rangle \sin^2(\phi) = \langle \varepsilon^2 \rangle$$

$$\langle \varepsilon'_{\mathrm{R}} \varepsilon'_{\mathrm{I}} \rangle = \langle \varepsilon_{\mathrm{R}}^2 \rangle \cos(\phi) \sin(\phi) - \langle \varepsilon_{\mathrm{I}}^2 \rangle \cos(\phi) \sin(\phi) = 0$$

Noise in Images

- Noise can be estimated at the phase center
- In the imaging process, we combine (with some weights) the individual visibilities V_i. At the phase center:

$$I = \sum w_i V_i \,/\, \sum w_i$$

- This is a classical case of noise propagation. If natural weights $w_i = 1/\sigma_i^2$ we have

$$1/\sigma^2 = \sum 1/\sigma_i^2$$

- Which is true anywhere else in the image by application of a phase shift
- So the noise rms in the image is indeed given by:

$$\delta S = \frac{2k}{\eta A} \frac{T_{sys}}{\sqrt{N(N-1)\, t_{int}\, \Delta\nu}}$$

Institut de Radioastronomie Millimétrique

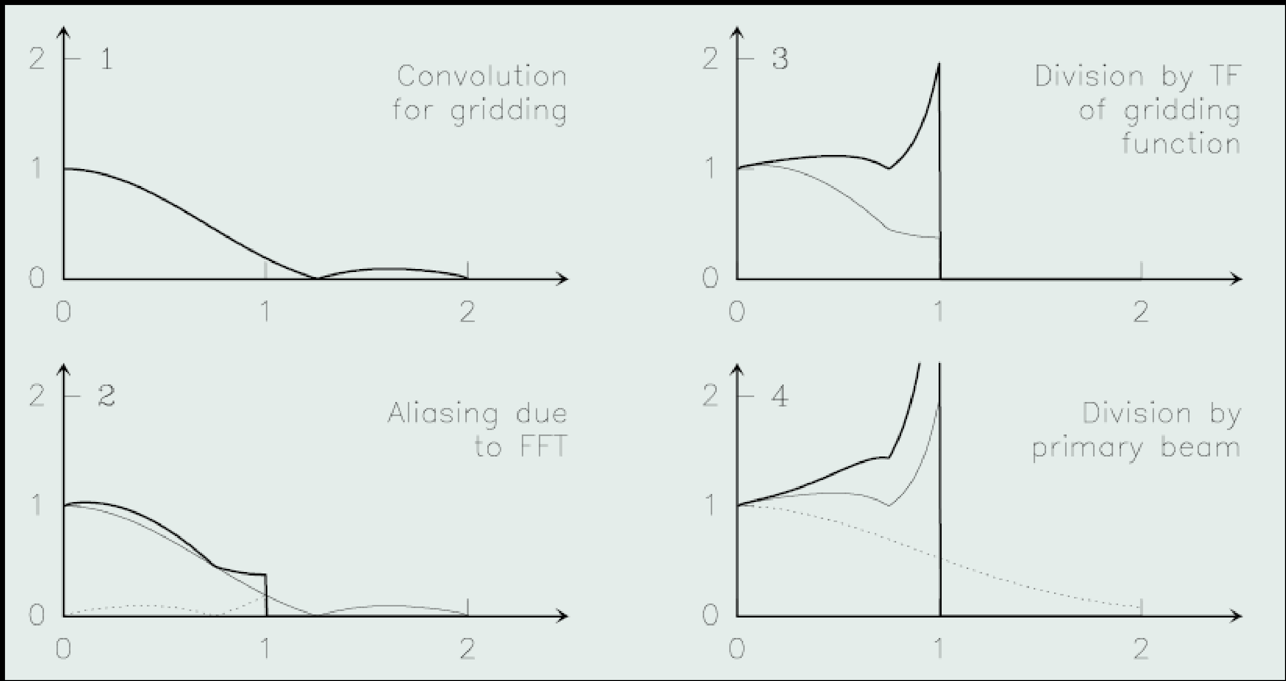
Noise in Images

- When using non-natural weights ($w_i \neq 1/\sigma_i^2$), either as a result of Uniform or Robust weighting, or due to Tapering, the noise (for point sources) increases
  - Robust weighting improves angular resolution
  - Tapering can be used to smooth data
  - Both decrease sensitivity

- Deconvolution
  - Dirty image in Jy/(dirty beam) – ill-defined unit
  - Deconvolved image in Jy/(clean beam)

Noise in Images

- Gridding introduces a convolution in UV plane, hence a multiplication in image plane
- Aliasing folds the noise back into the image
- Gridding Correction enhances the noise at edge
- Primary beam Correction even more...



iram
Institut de
Radioastronomie
Millimétrique

Bandwidth Effects

- The correlator channels have a non-square shape, i.e. their responses to narrow band and broad band signals differ.
- Hence the noise equivalent bandwidth $\Delta\nu_N$ is not the channel separation $\Delta\nu_C$, neither the effective resolution $\Delta\nu_R$
- These effects are of order 15-30 % on the noise.
- In practice, $\Delta\nu_N > \Delta\nu_C$, i.e. adjacent channels are correlated.
- Noise in one channel is less than predicted by the Noise Equation when using the channel separation as the bandwidth.
- But it does not average as $\sqrt{n_c}$ when using n_c channels...
- When averaging $n_c \gg 1$ i.e. many channels, the bandpass becomes more or less square: the effective bandwidth becomes $n_c\ \Delta\nu_C$.
- Consequence: There is no (simple) exact way to propagate the noise information when smoothing in frequency.
- Consequence: In GILDAS software, it is assumed $\Delta\nu_N = \Delta\nu_C = \Delta\nu_R$, and a $\sqrt{n_c}$ noise averaging when smoothing

Brightness sensitivity

- Extended source sensitivity?
- We use brightness temperatures, as measured in a solid angle Ω (= beam)

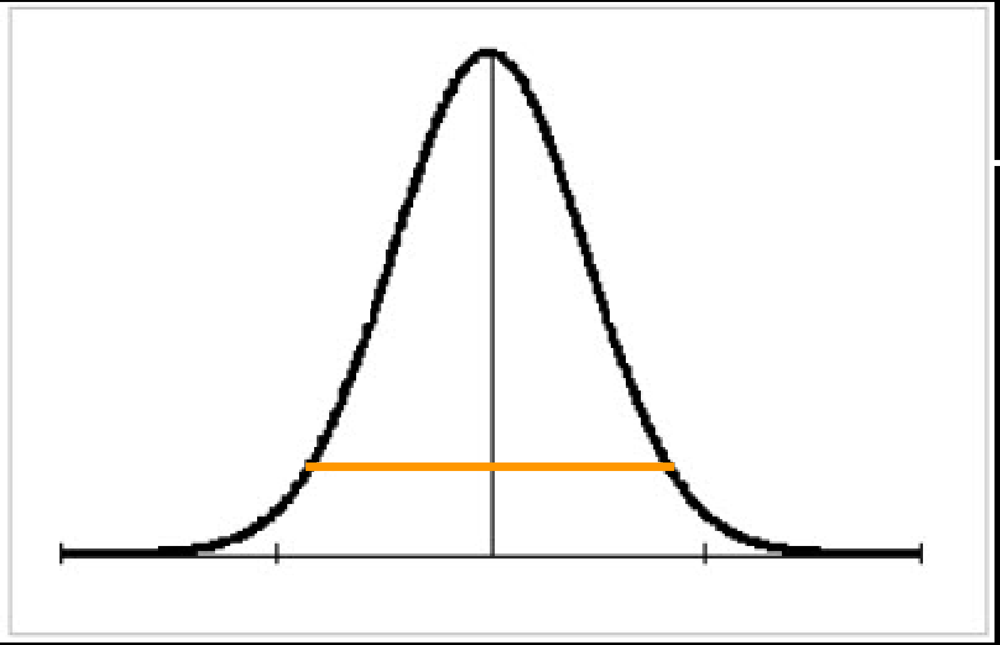
$$T = \frac{\lambda^2}{2k\,\Omega} S = \frac{\lambda^2}{2\,k} \frac{4ln(2)}{\pi\,\theta_1\theta_2} S$$

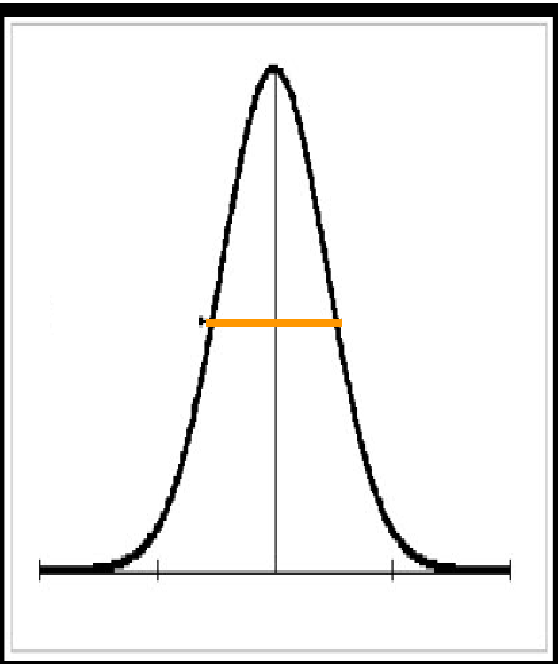
- So the brightness temperature rms is:

$$\delta T = \frac{2ln(2)\lambda^2}{k\,\pi} \frac{1}{\theta_1\theta_2} \delta S$$

Brightness s

- Temperature = for a source filling the beam
- Brightness temperature depends on the beam size
- Beam x Temperature = flux





Sensitivities

- Point-source sensitivity (Jy/beam) does not depend on the angular resolution
- Brightness sensitivity (Kelvin) does depends on the angular resolution θ

$$\delta S = \frac{2k}{\eta A} \frac{T_{sys}}{\sqrt{N(N-1)\, t_{int}\, \Delta\nu}}$$

$$\delta T = \frac{2 ln(2) \lambda^2}{k\,\pi} \frac{1}{\theta_1 \theta_2} \delta S$$

Sensitivities

- Point-source sensitivity (Jy/beam) does not depend on the angular resolution
- Brightness sensitivity (Kelvin) does depends on the angular resolution θ

BRIGHTNESS SENSITIVITY DEPENDS ON THE ANGULAR RESOLUTION

Sensitivities

Example 1:

- At 1'' resolution, my source has been detected with 20 σ in only 30 min, so this will be easy to map it at 0.1''

Sensitivities

Example 1:

- At 1’’ resolution, my source has been detected with 20 σ in only 30 min, so this will be easy to map it at 0.1’’
- Really?
  - Increase resolution by 10 means reducing brightness sensitivity by 100

Sensitivities

Example 1:

- At 1'' resolution, my source has been detected with 20 σ in only 30 min, so this will be easy to map it at 0.1''
- Really?
  - Increase resolution by 10 means reducing brightness sensitivity by 100
  - Need 10000 times more integration time to reach same brightness sensitivity, i.e. 5000 hours ~ 7 months, full-time
  - Time $\propto$ resolution^4 for a given sensitivity…
  - If we relax sensitivity by a factor 5 (4 σ detection), still need 400 times more integration time = 200 h

Sensitivities

Example 2:

- ALMA accepts projects for a given angular resolution (e.g. 1'')
- But observes with 0.8''
- Same integration time? Brightness rms increased by 1.5
  - Yes, but then, I can smooth the image, right?
  - Yes, will get 1'' resolution, but not the same brightness rms (because smoothing = downweighting long baselines = reducing integration time)
- Same brightness sensitivity? Integration time increased by 2.25 (time $\propto$ resolution^4)

Sensitivities

Conclusions: do not forget

$$\delta T \propto \frac{1}{\theta^2 \sqrt{t_{int}\,\Delta\nu}}$$

- Planning observation often means compromizing sensitivity/time/resolutions
- Mapping sources at (very) high angular resolution is extremely time-consuming and reserved to very bright sources

Low Signal to Noise

- A nice case
  - Observers advantage: don't have to worry about bandpass & flux calibration…
  - Theorists advantage: the data is always compatible with your favorite model

- A necessary challenge
  - mm interferometry is (almost) always sensitivity limited
  - so a careful analysis is necessary: when is a source detected? which parameters can be derived?

Continuum : detection

- do not resolve the source
- get the best absolute position (optical, previous obs, …)
- use UV_FIT (fit in the uv plane) to determine the S/N ratio
- what is the position accuracy?

< 1/10th of beam

- need > 3σ to claim detection
- fix the position
- use an appropriate source size

About the beam

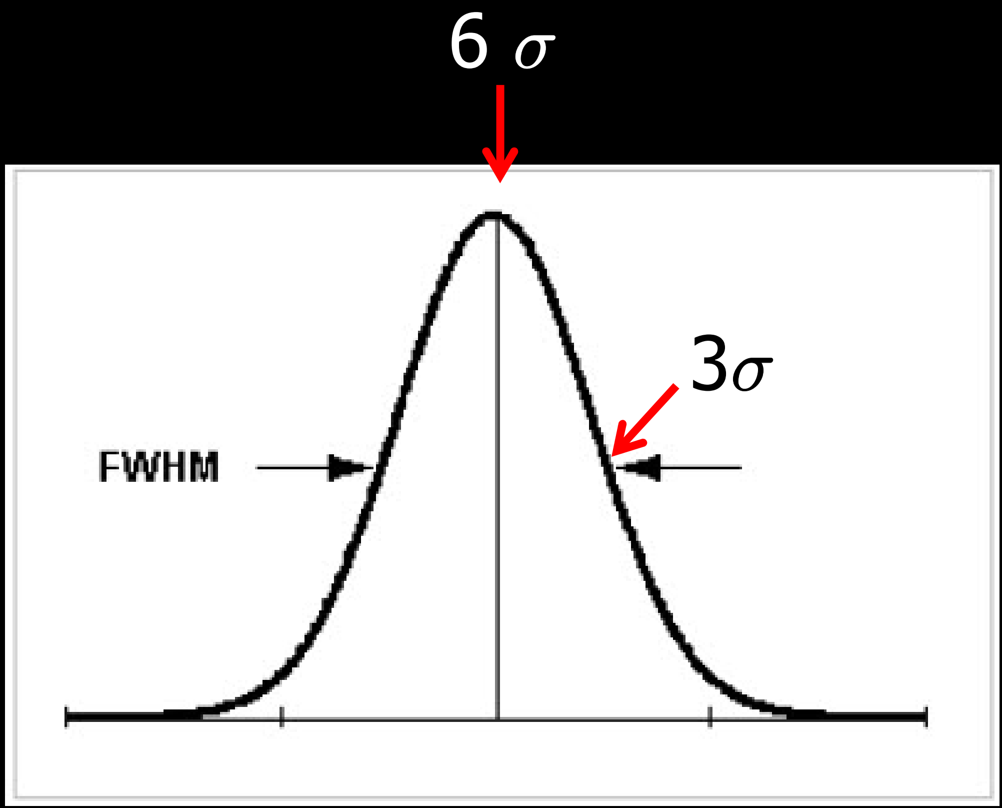
- need > 4σ for detection
- do not fix the position
- use an appropriate source size

Unknown

- need 5σ signal for detection
- make an image to locate
- use as starting point
- do not fix the position
- use an appropriate source size

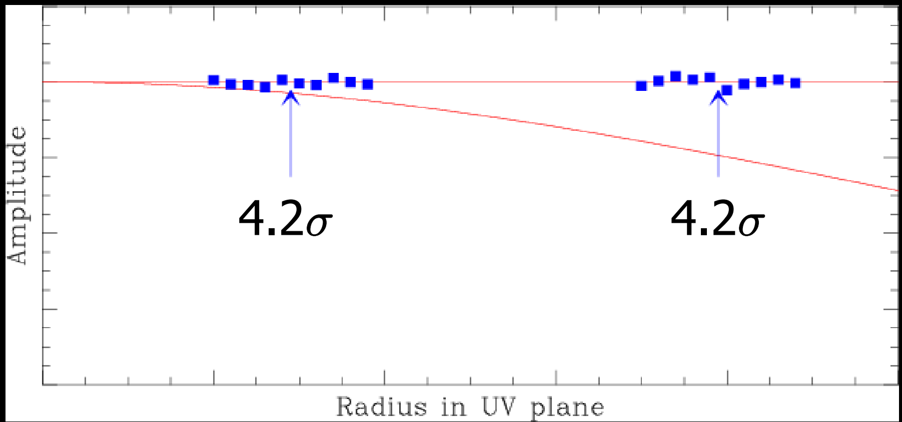
Continuum: source size

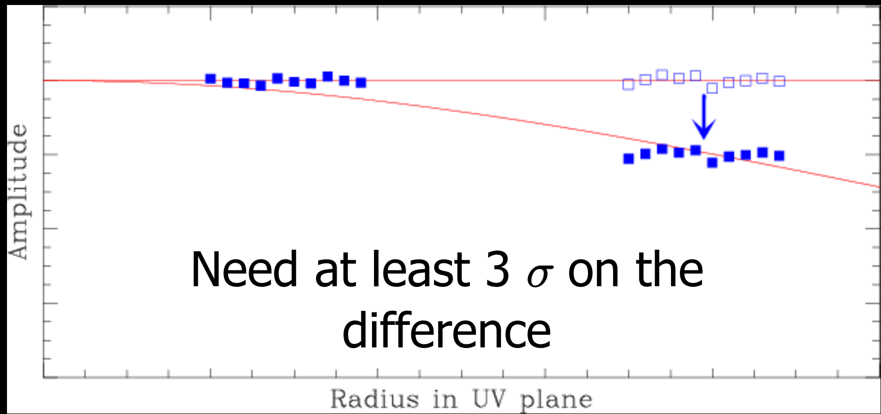
- With SNR < 6 σ , cannot measure any source size

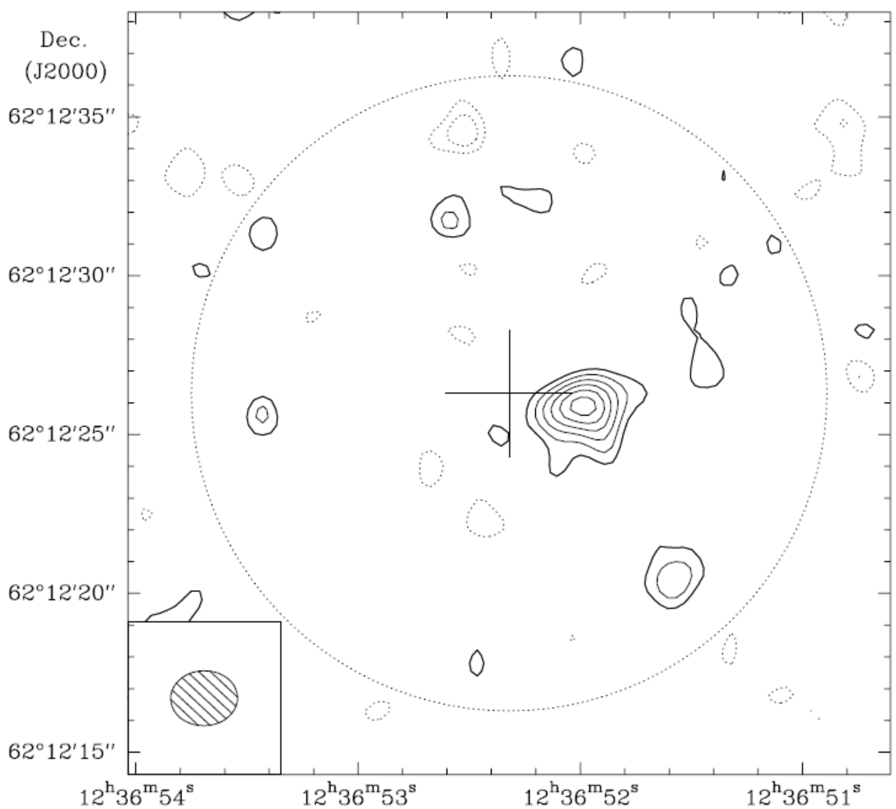


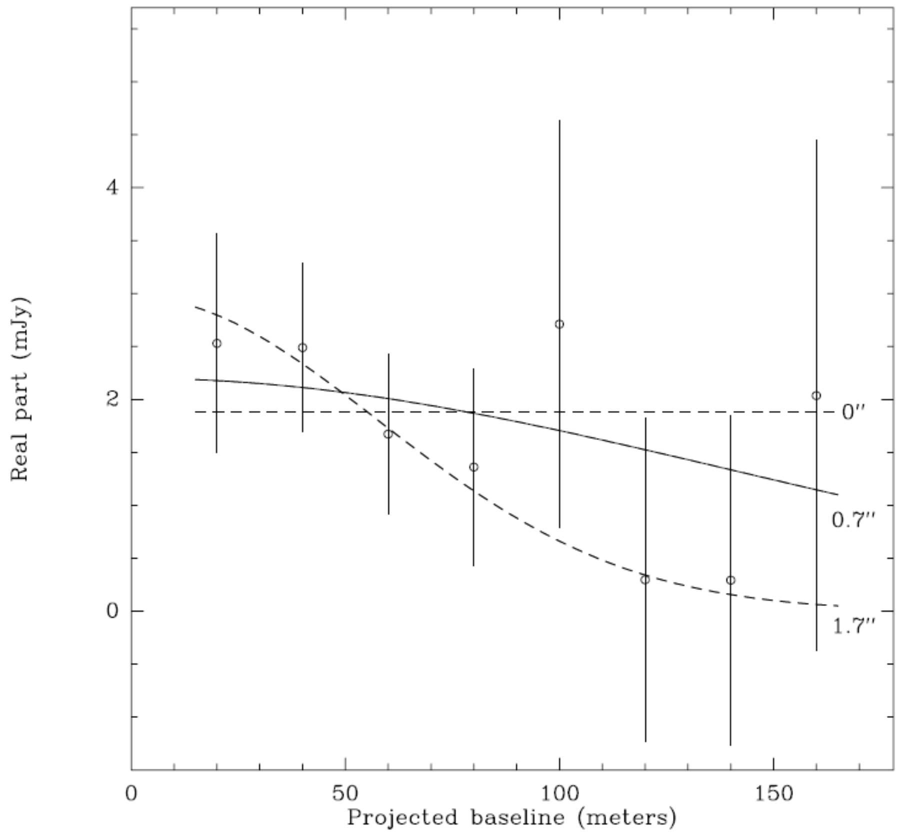
Continuum: source size

- With SNR < 6 σ , cannot measure any source size
  - divide data in two subsets: shortest baselines on one side, longest on another
  - each subset gets a 4.2 σ error on mean flux
  - error on the difference is then just 3 σ









Example: HDF source (Downes et al. 1998)

7 σ detection of the strongest source in the Hubble Deep Field. Note that contours are *visually cheating* (start at 2 σ but with 1 σ steps).

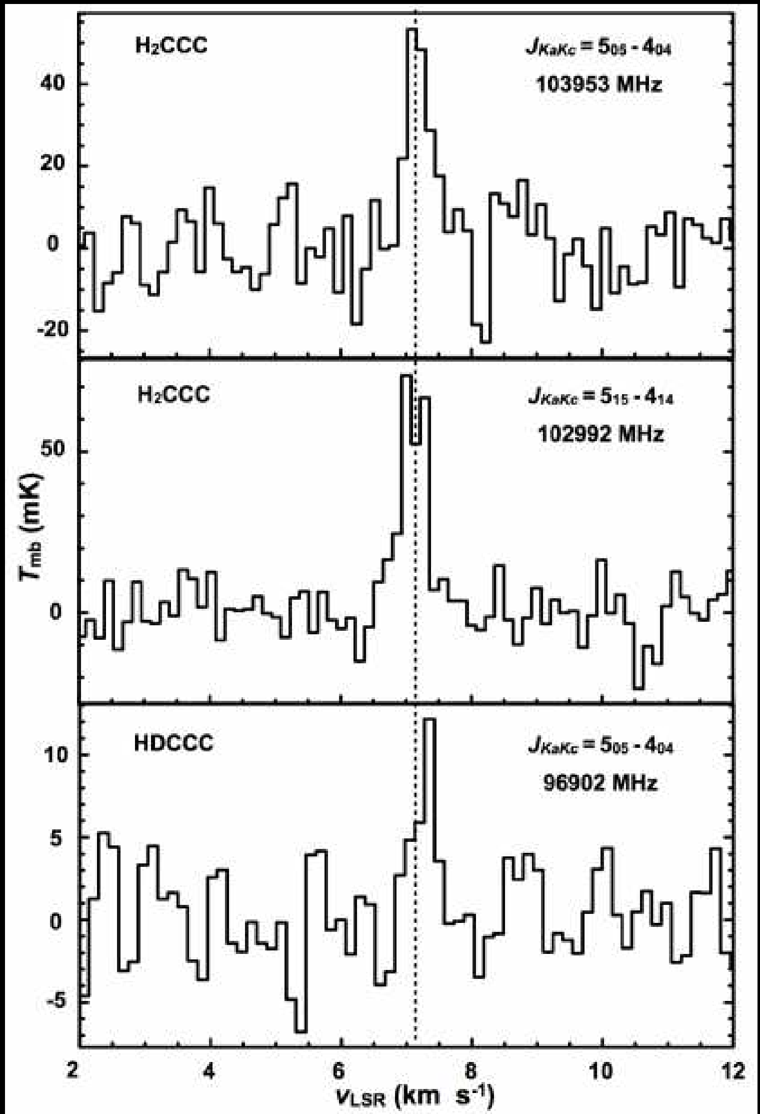
Attempts to derive a size. Size can be as large as the synthesized beam... Note that the integrated flux increases with the source size.

Line: things get worse…

- Line velocity unknown: observer will select the brightest part of the spectrum ➔ bias
- Line width unknown: observer may limit the width to brightest part of the spectrum ➔ another bias
- If position is unknown, it is determined from the integrated area map (or visibilities) made from the tailored line window specified by the astronomer. This gives a biased total flux.

- Any speculated extension will increase the total flux, by enlarging the selected image region (same effect as the tailored line window).
- Net result = 1 to 2 σ positive bias on integrated line flux.
- *Things get really messy if a continuum is superposed to the weak line...*

Line: things get worse…



Line

Point source or unresolved source (< 1/3rd of the beam)

- Determine position, e.g. from continuum if available, or from integrated line map if not, or from other data
- Derive line profile by fitting point or small (fixed size, fixed position) source into UV data for each channel
- Gives you a flux as function of velocity/frequency
- Fit this spectrum by Gaussian (with or without constant baseline offset, depending on whether the continuum flux is known or not)

Line

Extended sources, and/or velocity gradient

- Fit multi-parameter (6 for an elliptical gaussian) source model for each spectral channel into UV data
- Consequence : signal in each channel should be >6 σ to derive any meaningful information
- Strict minimum is 4 σ (per line channel) to get flux and position for a fixed size Gaussian
- Velocity gradients not believable unless even better signal to noise is obtained per line channel...

Line

- Do not believe velocity gradient unless proven at a 6 σ level in each channel. Remember that position accuracy per channel is the beamwidth divided by the signal-to-noise ratio...
- Do not believe source size unless S/N > 10 (or better)
- Expect line widths to be very inaccurate
- Expect integrated line intensity to be positively biased by 1 to 2 σ
- Even more biased if source is extended
- *These biases are the somehow analogous of the Malmquist bias*

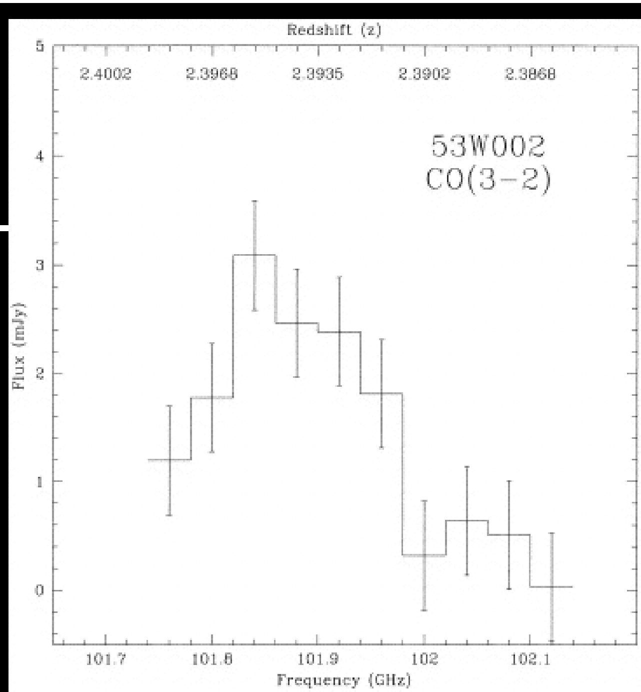
Examples

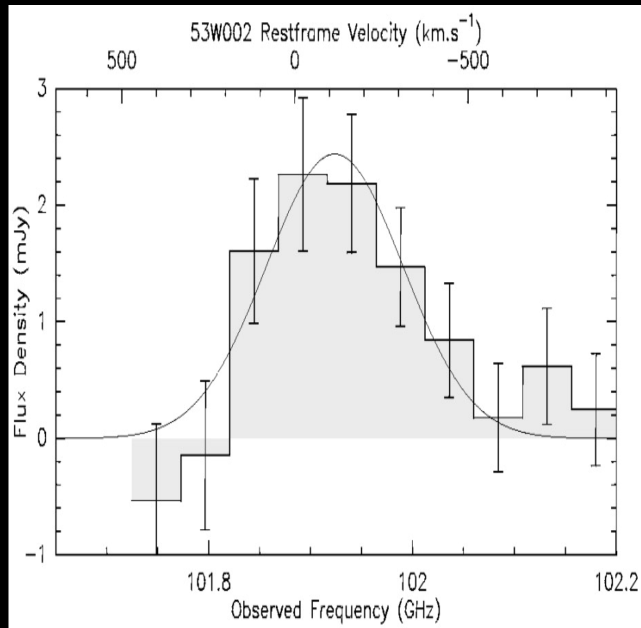
- Examples are numerous, specially for high redshift CO, e.g. 53 W002 :

  - OVRO (S. et al. 1997) claims an extended source, with velocity gradient. Yet the total line flux is 1.5 § 0.2 Jy.km/s i.e. (at best) only 7 ¾ .

  - PdBI (A. et al. 2000) finds a line flux of 1.20 § 0.15 Jy.km/s, no source extension, no velocity gradient, different line width and redshift.

  - Note that the line fluxes agree within the errors…

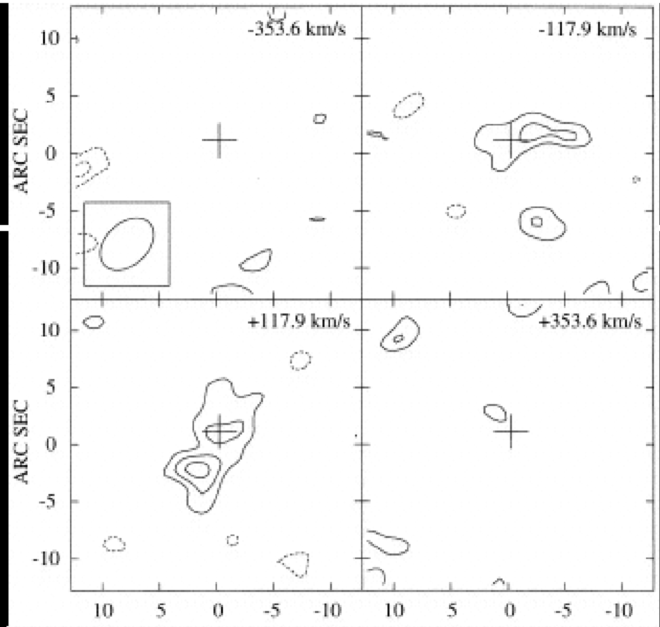


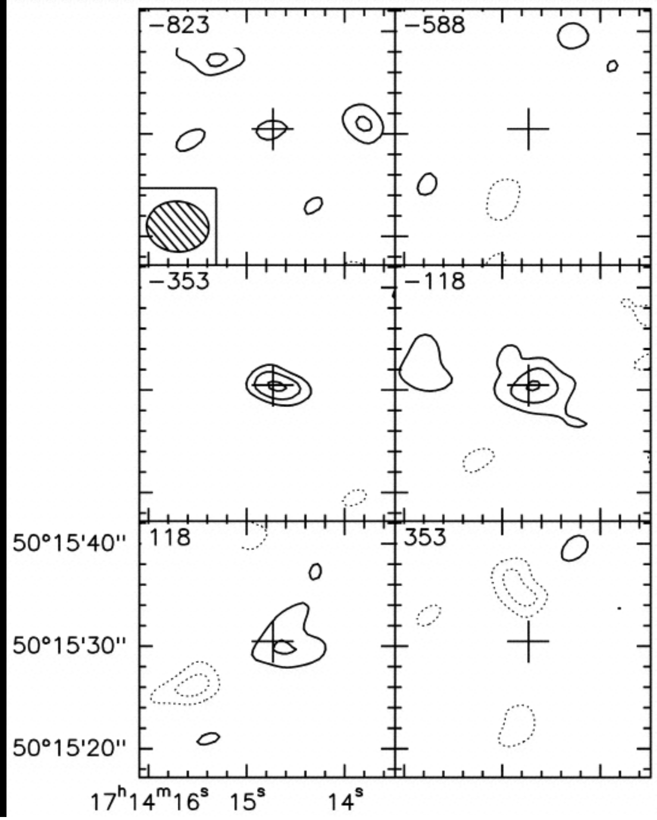


Examples

Remark(s)

- But the images (contours) look convincing !
- Answer : beware of visually confusing contours which start at 2 σ (sometimes even 3) but are spaced by 1 σ



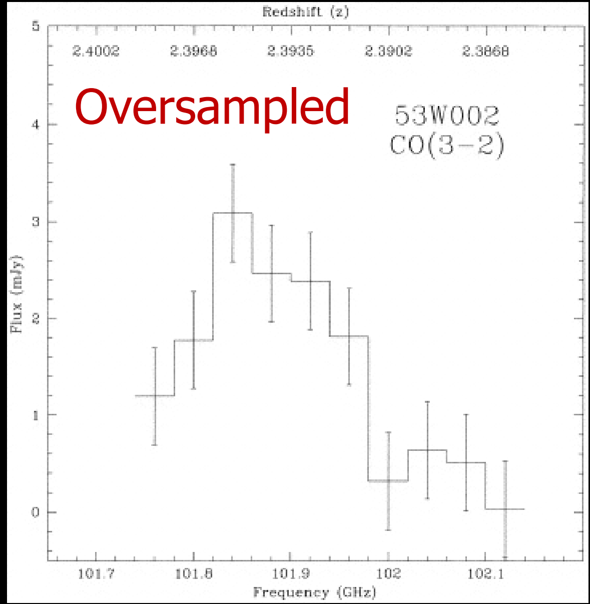


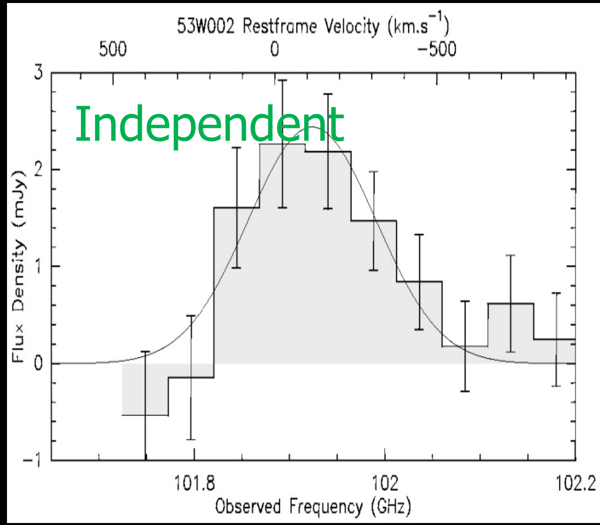
Examples

Remark(s)

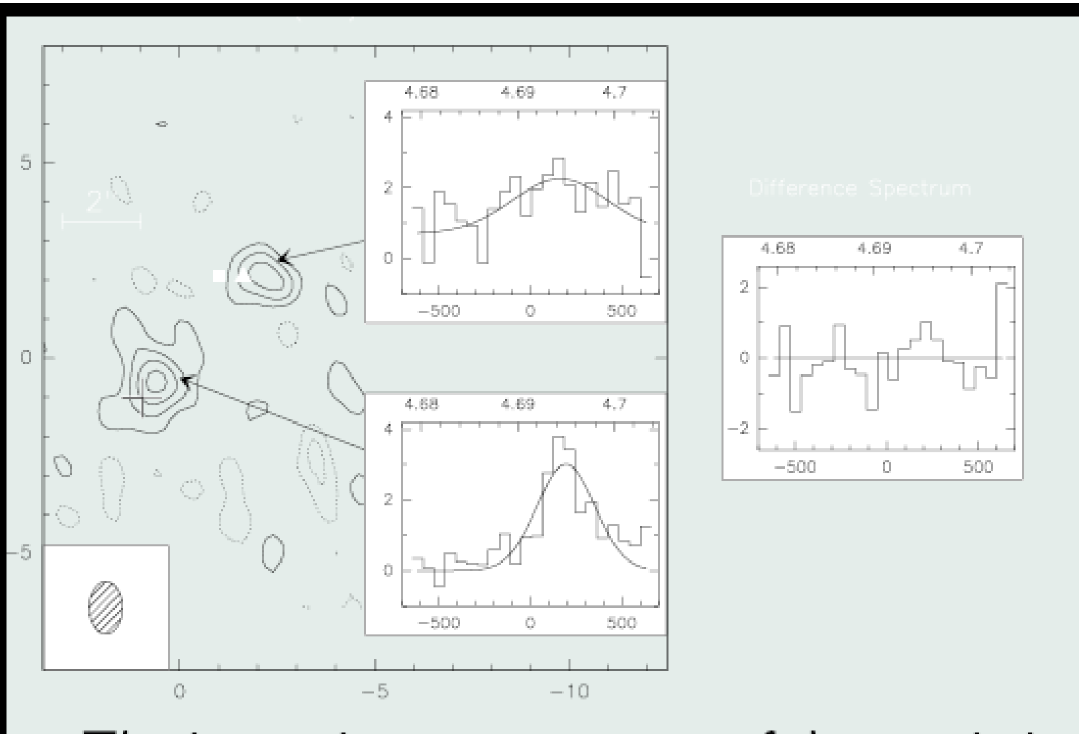
- But the images (contours) look convincing !
- Answer : beware of visually confusing contours which start at 2 σ (sometimes even 3) but are spaced by 1 σ

- But the spectrum looks convincing, too !
- Answer : beware of visually confusing spectra, which are oversampled by a factor 2. The noise is then not independent between adjacent channels.





Example: (no) Velocity Gradients



- Contour map of dust emission at 1.3 mm, with 2 σ contours
- The inserts are redshifted CO(5-4) spectra
- A weak continuum (measured independently) exist on the Northern source
- The rightmost insert is a difference spectrum (with a scale factor applied, and continuum offset removed): No SIGNIFICANT PROFILE DIFFERENCE!
- i.e. no Velocity Gradient measured.

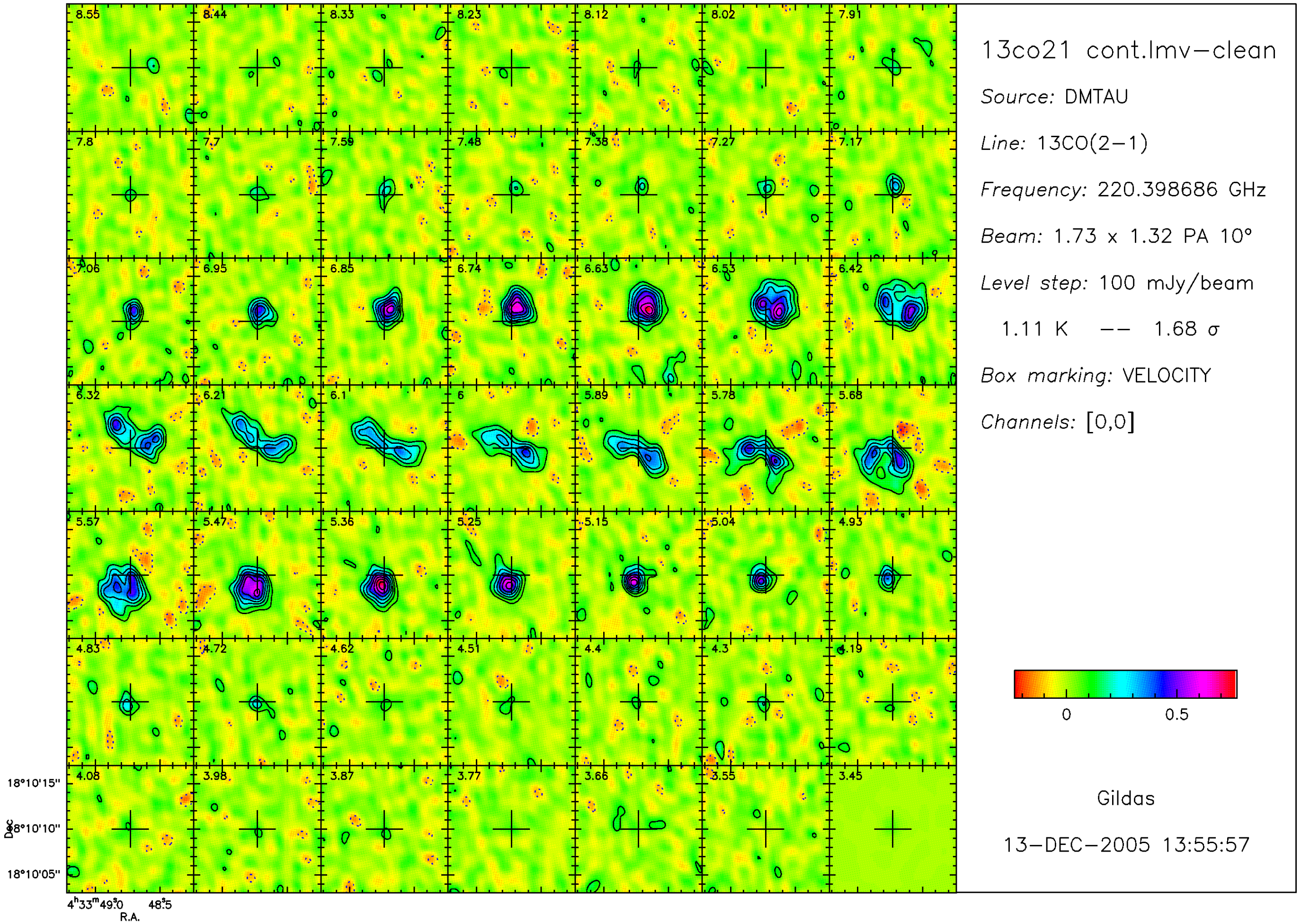
How to analyze weak lines ?

Perform a statistical analysis (e.g. χ^2, or other statistical test) comparing model prediction to observations, i.e. visibilities

- Physical model of the source, with milited number of free parameters
- Predict visibilities
  - *The GILDAS software offer tools to compute visibilities from an image / data cube (task UV_FMODEL)*
- Beware of various subtle effect, eg primary beam, correlated (original) channels
- Appropriate statistical tests to constrain input parameters
- This can actually provide a better estimate of the noise level than the prediction given by the weights.

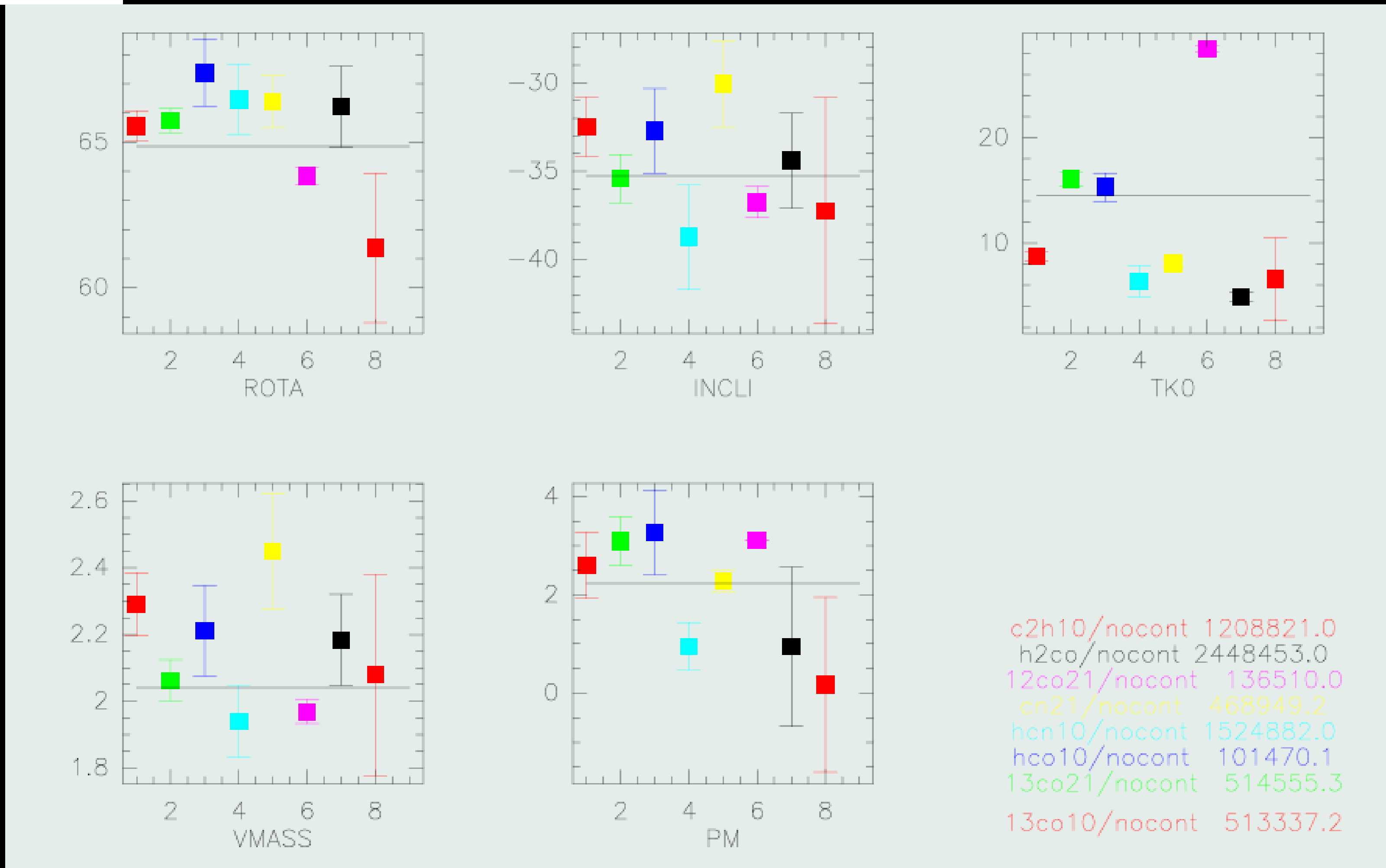
iram
Institut de Radioastronomie Millimétrique

Example of Analysis



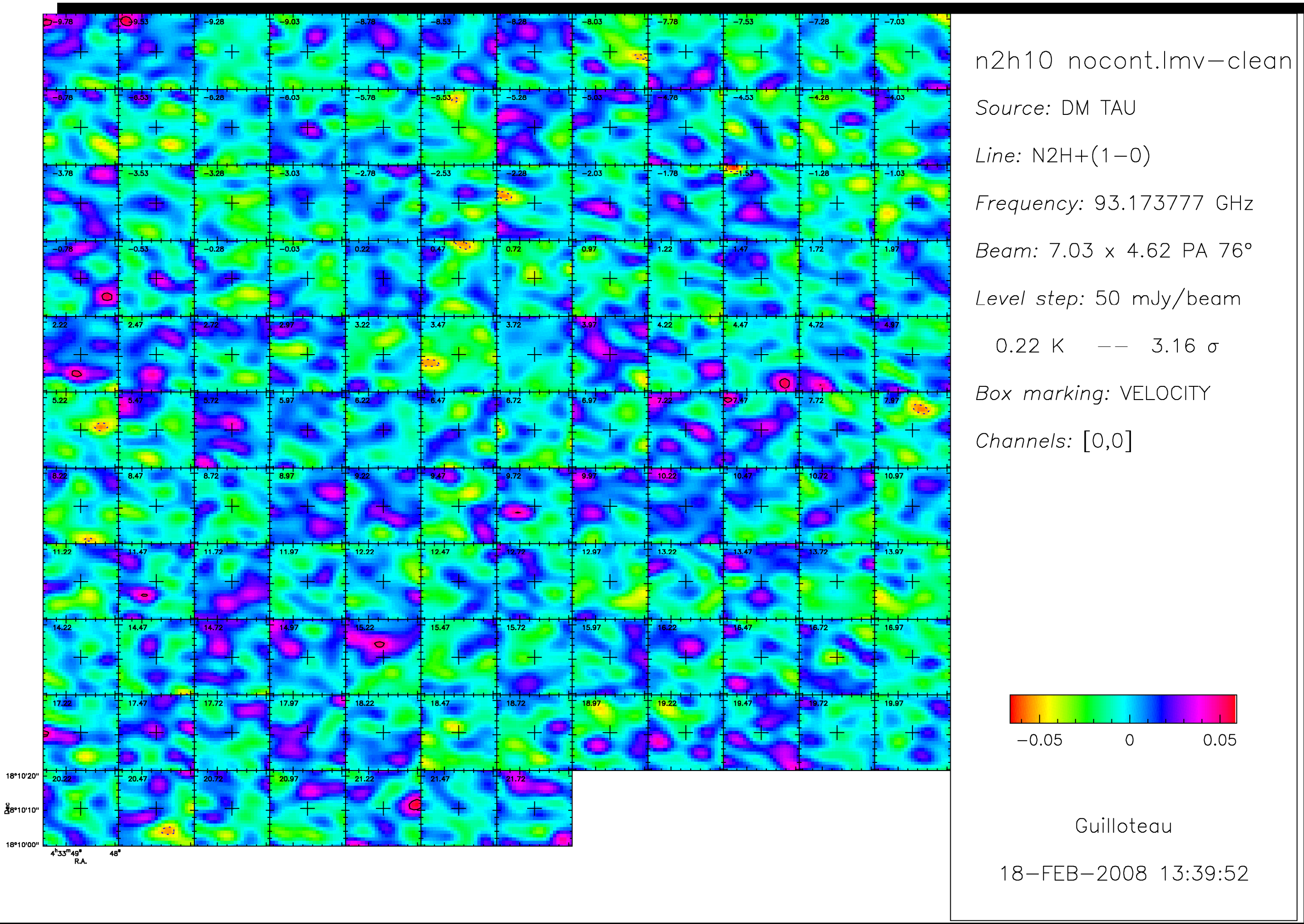
A typical data cube showing 13CO emission in a protoplanetary disk. It has quite decent S/N, and one can recognize the rotation pattern of a Keplerian disk

Example of Analysis



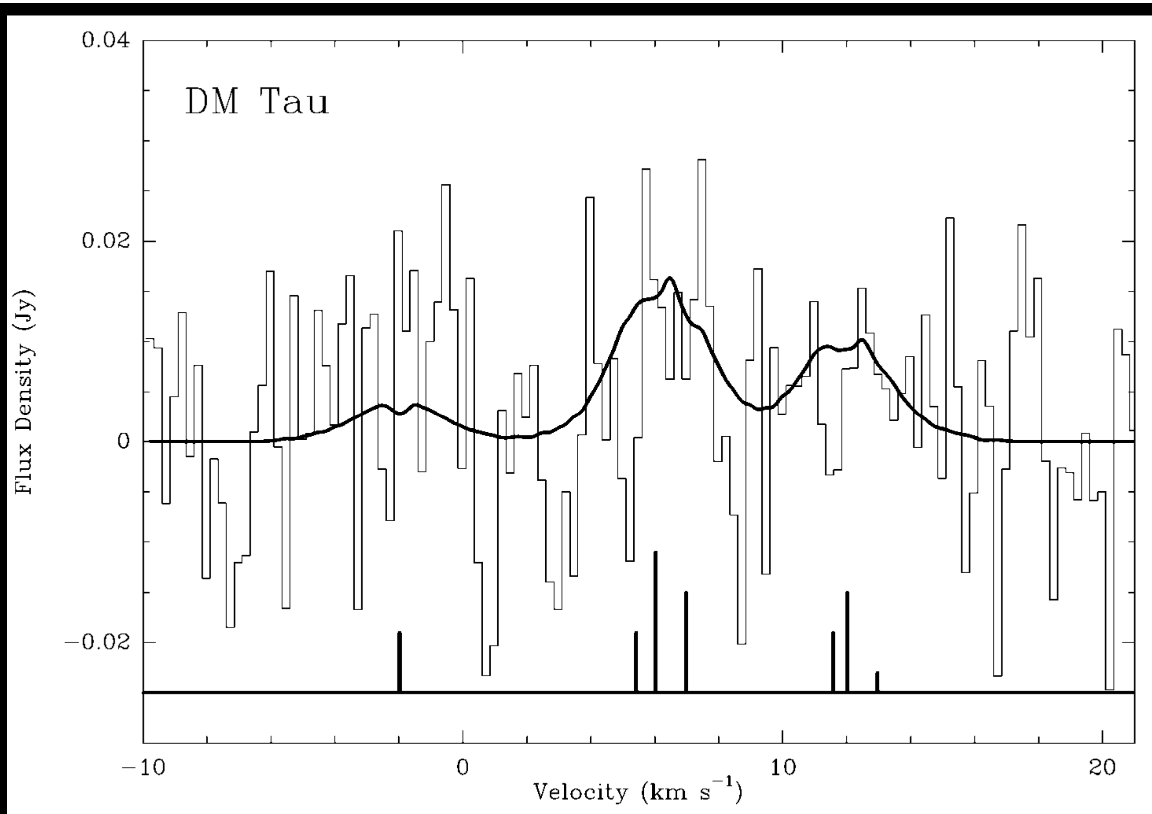
χ^2 analysis in the UV plane (5 disk parameters, for 8 disks)

Example of Analysis



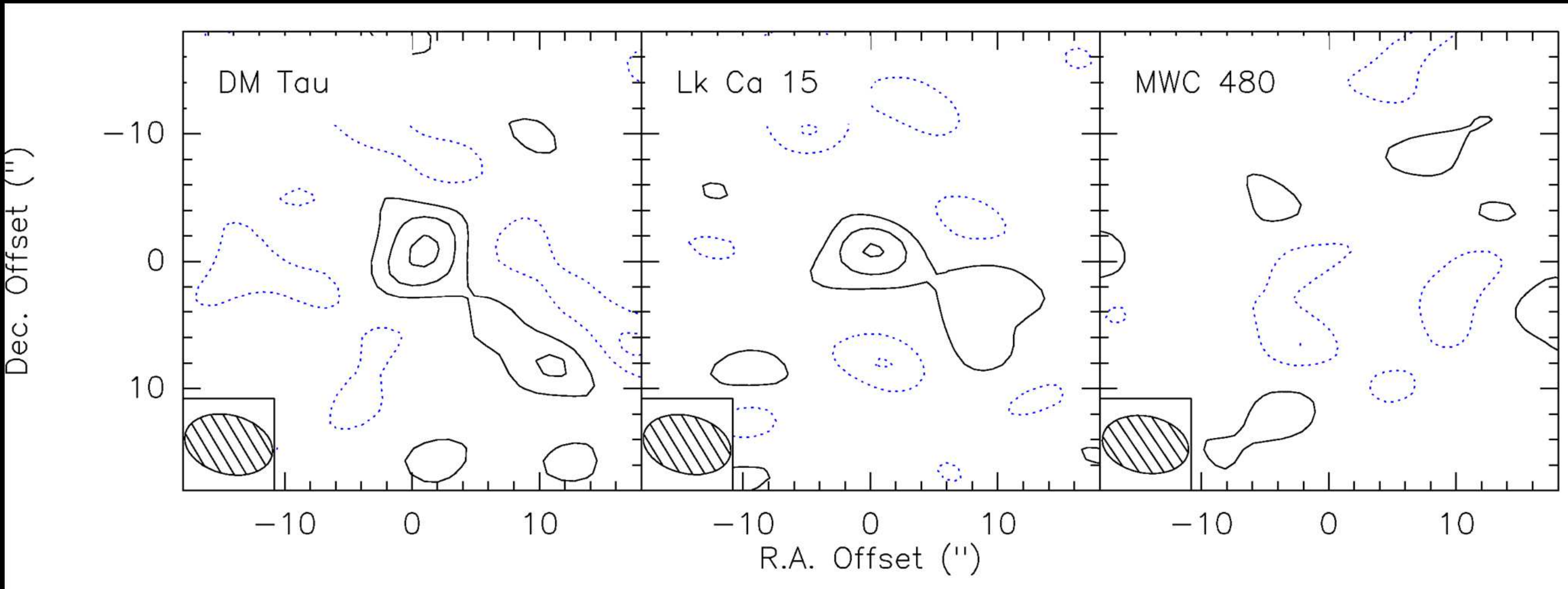
A (really) low Signal to Noise image of the protoplanetary disk of DM Tau in the main group of hyperfine components of the N_2H^+ 1-0 transition.

Example of Analysis



Best fit integrated profile for the N_2H^+ 1-0 line, derived from a χ^2 analysis in the UV plane, using a line radiative transfer model for proto-planetary disks, assuming power law distributions, and taking into account the hyperfine structure (Dutrey et al. 2007).

Example of Analysis



- Maps of the integrated N_2H^+ 1-0 line emission, using the best profile derived from the χ^2 analysis in the UV plane as a (velocity) smoothing kernel (optimal filtering).
- 7 σ detection for DM Tau, 6 σ detection for LkCa 15, beam is 7x4.6''