Dealing with NOISE

Part I: Noise in general

Part II: Low Signal to Noise case

Stéphane GUILLOTEAU

Observatoire de Bordeaux

IRAM Millimeter Interferometry School 4
Nov 2004

System Temperature

- The output power of the receiver is linked to the **Antenna System Temperature** by:

$$P_N = \gamma k T_{ant} \Delta\nu \tag{1}$$

- When looking at a source, the output power becomes $P_N + P_a$ where

$$P_a = \gamma k T_a \Delta\nu \tag{2}$$

- T_a is called the **antenna temperature** of the source.
- This is not a purely conventional definition.
  It can be demonstrated that P_a is the power the receiver(+antenna) would deliver when observing a blackbody (filling its entire beam pattern) at the physical temperature T_a.
- Thus, T_{ant} is the temperature of the "equivalent" blackbody seen by the antenna (in the Rayleigh Jeans approximation)

System Temperature

- So, T_{ant} , is given by (just summing the input powers...)

$$
\begin{aligned}
T_{ant} &= T_{bg} \quad \texttt{cosmic background} \\
&+T_{sky} \sim \eta_f (1 - e^{-\tau_{atm}}) T_{atm}, \quad \texttt{sky noise} \\
&+T_{spill} \sim (1 - \eta_f - \eta_{loss}) T_{ground}, \quad \texttt{ground noise pickup} \\
&+T_{loss} \sim \eta_{loss} T_{cabin}, \quad \texttt{optical losses in the receiver cabin} \\
&+T_{rec} \quad \texttt{receiver noise}
\end{aligned} \tag{3}
$$

- Note that this is a broad-band definition. It is a **DSB** (Double Side Band) noise temperature

System Temperature

- Many astronomical signals are narrow band. g being the image to signal band gain ratio, the equivalent DSB signal giving the same antenna temperature as a pure SSB signal is only
$$P_{DSB} = (1 \times P_{SSB} + g \times 0)/(1+g)$$
- We usually refer the **system temperature** and **antenna temperature** to a perfect antenna ($\eta_f = 1$) located outside the atmosphere, and **single sideband signal**:
$$T_{sys} = \frac{(1+g)}{\eta_f} e^{\tau_{atm}} T_{ant} \qquad (4)$$
$$T_A^* = \frac{(1+g)}{\eta_f} e^{\tau_{atm}} T_a$$
- this **antenna temperature** T_A^* is weather independent, and is linked to the source flux by an antenna dependent quantity only:
$$T_A^* = \frac{\eta_a A}{2k} S_\nu \qquad (5)$$

The Noise Equation

- The noise power is T_{sys}, the signal is T_A^*, and there are $2\Delta\nu\Delta t$ independent samples to measure a correlation product in a time Δt, so the Signal to Noise is

$$\mathcal{R}_{sn} = \frac{T_{sys}}{T_A^*}\sqrt{2\Delta\nu\Delta t} \tag{6}$$

- The noise on a single baseline is thus

$$\Delta S = \frac{\sqrt{2}kT_{sys}}{\eta_a A\sqrt{\Delta\nu\Delta t}} \tag{7}$$

- this is $\sqrt{2}$ *less* than that of a *single* antenna in total power
- but $\sqrt{2}$ *worse* than that of an antenna with the *same total collecting area*
- this *sensitivity loss* is because we ignore the autocorrelations

The Noise Equation

- Quantization must be accounted for

$$\Delta S = \frac{\sqrt{2} k T_{sys}}{\eta_q \eta_a A \sqrt{\Delta\nu \Delta t}} \tag{8}$$

with η_q the quantization efficiency (0.93 for the 2-bit, 4-level correlator).

- Noise is uncorrelated from one baseline to another
- there are $n(n-1)/2$ baselines for n antennas
- thus the **point source** sensitivity is

$$\Delta S = \frac{2 k T_{sys}}{\eta_q \eta_a A \sqrt{n(n-1)\Delta\nu \Delta t}} = \frac{\mathcal{J} T_{sys}}{\eta_q \sqrt{n(n-1)\Delta\nu \Delta t}} \tag{9}$$

since

$$\mathcal{J} = \frac{2k}{\eta_a A}$$

is the Jy/K conversion factor of one antenna

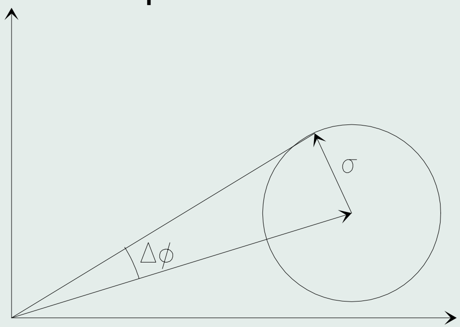
Noise on Amplitude and Phase

- Noise properties for 1 baseline vary with Signal-to-Noise ratio
- On the amplitude & flux density

$$S \ll \sigma \quad \begin{cases} \sigma_A & \simeq \sigma\sqrt{2-\frac{\pi}{2}}\left(1+\left(\frac{S}{2\sigma}\right)^2\right) \\ <S> & \simeq \sigma\sqrt{\frac{\pi}{2}}\left(1+\left(\frac{S}{2\sigma}\right)^2\right) \end{cases} \tag{10}$$

$$S \gg \sigma \quad \begin{cases} \sigma_A & \simeq \sigma \\ <S> & \simeq S \end{cases} \tag{11}$$

- On the phase



$$S \ll \sigma \quad \left\{\sigma_\phi \simeq \frac{\pi}{\sqrt{3}}\left(1-\sqrt{\frac{9}{2\pi^3}}\frac{S}{\sigma}\right)\right. \tag{12}$$

$$S \gg \sigma \quad \left\{\sigma_\phi \simeq \frac{\sigma}{S}\right. \tag{13}$$

- Source detection is much easier on the *phase* than on the *amplitude*, since for $S/N \sim 1$, $\sigma_\phi = 1$ radian $= 60^\circ$.

Noise in Images: preamble

- 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 *complex* (or *spectral*) correlator gives the same variance on the real and imaginary part of the complex visibility, $\langle \varepsilon_{\mathrm{R}}^2 \rangle = \langle \varepsilon_{\mathrm{I}}^2 \rangle = \langle \varepsilon^2 \rangle$
  - noise in Real and Imaginary parts are uncorrelated $\langle \varepsilon_{\mathrm{R}} \varepsilon_{\mathrm{I}} \rangle = 0$
- Effect of rotation: **NONE**
  any phase factor (rotation) applied to the complex visibility still result in the same properties on the variance of the real and imaginary parts, because $\cos^2(\phi) + \sin^2(\phi) = 1$

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

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

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

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

Noise in Imaging: first order

- In the imaging process, we combine (with some **weights**) the individual visibilities V_i. At the phase center:

$$I = \left(\sum w_i V_i\right) / \left(\sum w_i\right) \tag{14}$$

- Assuming a point source at the phase center, $V_i = V + \varepsilon_{\mathrm{R}i}$

$$I = \left(\sum w_i (V + \varepsilon_{\mathrm{R}i})\right) / \left(\sum w_i\right) \tag{15}$$

  where $\varepsilon_{\mathrm{R}i}$ is the (real part) of the noise.

- thus the expectation of $I = V$, since $\langle \varepsilon_{\mathrm{R}i} \rangle = 0$
- since $\langle \varepsilon_{\mathrm{R}i} \varepsilon_{\mathrm{R}j} \rangle = 0$ the variance of I is

$$\sigma^2 = \langle I^2 \rangle - \langle I \rangle^2 = \frac{\sum w_i^2 \langle \varepsilon_{\mathrm{R}i}^2 \rangle}{(\sum w_i)^2} \tag{16}$$

- using $\langle \varepsilon_{\mathrm{R}i}^2 \rangle = \sigma_i^2$ and the **natural weights** $w_i = 1/\sigma_i^2$, we find as expected

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

- At any other point in the image, the same remains true, since only a phase factor is applied to combined all visibilities together.

Demonstration ...

$$\sigma^2 = \frac{\sum w_i^2 \langle \varepsilon_{\mathrm{R}i}^2 \rangle}{(\sum w_i)^2} \tag{17}$$

$$= \frac{\sum w_i^2 \sigma_i^2}{(\sum w_i^2)^2} \tag{18}$$

$$= \frac{\sum (\frac{1}{\sigma_i^2})^2 \sigma_i^2}{(\sum \frac{1}{\sigma_i^2})^2} \tag{19}$$

$$= \frac{\sum (\frac{1}{\sigma_i^2})}{(\sum \frac{1}{\sigma_i^2})^2} \tag{20}$$

$$= \frac{1}{\sum \frac{1}{\sigma_i^2}} \tag{21}$$

as expected given the choice of weights...

Noise in Imaging: Weighting and Tapering

- 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
- the increase is given by

$$w_{rms}/w_{mean}$$

where

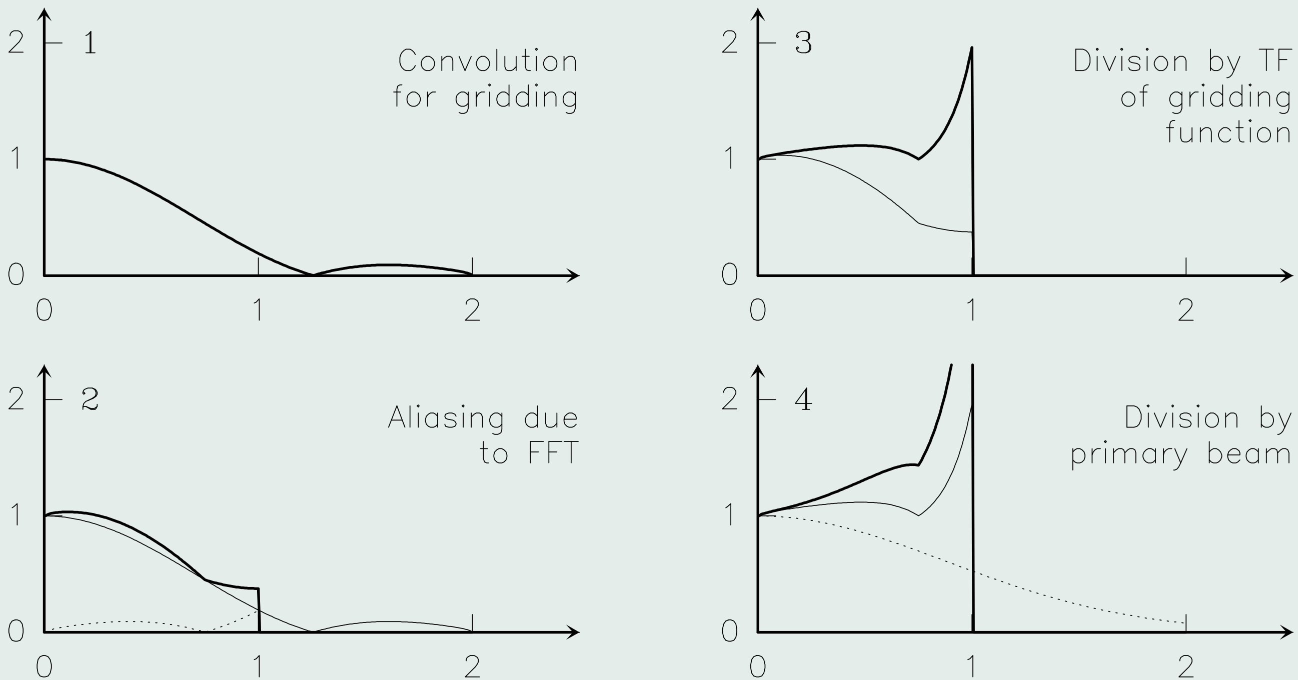
$$w_{rms} = \sqrt{\left(\sum (WT)^2\right)/n}$$

$$w_{mean} = \left(\sum WT\right)/n$$

- **Robust** weighting allows to improve angular resolution, and yet minimize (control) the noise increase
- **Robust** weighting and **Tapering** can allow to control the beam shape.

Noise in Imaging: second order

- **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...



Extended Source Sensitivity

- This is problematic. Here is the usual approach:
- We use **brightness temperature** for extended sources
- Use the flux to brightness conversion factor

$$S = \frac{2kT_b\Omega_s}{\lambda^2} = \frac{2kT_b\pi\theta_s^2}{4ln(2)\lambda^2}$$

for a synthesized beam of solid angle Ω_s (Gaussian of FWHM θ_s)

- Since from the antenna equation $\Omega_A A_{eff} = \lambda^2$, the flux noise equation

$$\Delta S = \frac{2kT_{sys}}{\eta_q A_{eff}\sqrt{n(n-1)\Delta\nu\Delta t}}$$

gives the brightness noise equation

$$\Delta T_b = \frac{\Omega_A}{\Omega_s}\frac{T_{sys}}{\eta_q\sqrt{n(n-1)\Delta\nu\Delta t}} = \left(\frac{\theta_p}{\theta_s}\right)^2\frac{T_{sys}}{\eta_q\sqrt{n(n-1)\Delta\nu\Delta t}}$$

which is just a simple "beam dilution" formula applied to the standard noise for one antenna in total power, and accounting for n antennas.

Extended Source Sensitivity

- Brightness Noise Equation

$$\Delta T_b = \left(\frac{\theta_p}{\theta_s}\right)^2 \frac{T_{sys}}{\eta_q \sqrt{n(n-1)\Delta\nu\Delta t}}$$

- The previous formula is right only for sources just filling one synthesized beam.
- For more extended sources, it is **not** appropriate to count the number of synthesized beams n_b and divide by $\sqrt{n_b}$.
- This only gives a lower limit...
- **Why ?**
  - Averaging n_b beams is equivalent to smoothing
  - This is equivalent to tapering, i.e. to ignore the longest baselines...
  - This increases the noise ...
- Moreover, for very extended structures, **missing flux** may become a problem.

Noise in Imaging: 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

A parte: Reweighting in Frequency ?

- The receiver bandpass is not flat: T_{sys} depends on ν
- Hence the weights depend on the channel number i
- When synthesizing broad band data, should we take the weights into account ?
- For **pure** continuum data
  - **Yes**: it improves S/N
  - But: ill-defined equivalent central frequency, and undefined equivalent detection bandwidth
- For **line** data
  - **No**: could degrade S/N is line shape is not consistent with the weights
  - **No**: undefined bandwidth: does not allow to compute a *integrated line flux* ($\int S_\nu(\nu) d\nu$)
- In practice: not implemented in current GILDAS software

Noise in Imaging: Decorrelation

- Each visibility is affected by a random atmospheric phase ϕ
- Assuming a point source at the phase center, $V_i = Ve^{i\phi_i} + \varepsilon_{Ri}$

$$I = (\sum w_i(Ve^{i\phi_i} + \varepsilon_{Ri}))/(\sum w_i) \tag{22}$$

- the expectation of I is now only $Ve^{-(\Delta\phi)^2/2}$
- the noise does not change
- but the signal to noise is decreased
- the Signal is spread around the source (*seeing*)
- So the effect is different for an extended source...
- This may limit the **Dynamic range**, and the effective noise level may be much higher than the thermal noise

Estimating the Noise

- The **weights** are used to give a prediction of the noise level in the images.
- Displayed by `UV_MAP`
- Carried on in the image headers (`aaa%noise` variable for an image displayed with `GO MAP, GO NICE` or `GO BIT`)
- but does not handle properly the noise equivalent bandwidth
- neither the effects of decorrelation...
- `GO RMS` will compute the rms level on the displayed image. May be biased by the source structure
- `GO NOISE` will plot an histogram of image values, and fit a Gaussian to it to determine the noise level. Will be less biased than `GO RMS`.
- Both `GO NOISE` and `GO RMS` will include dynamic range effects (i.e. give you the "true" noise of your image, rather than the theoretical).

Conclusions

- mm interferometry is not so difficult to understand
- even if you don't, the noise equation is all you need
- the noise equation

$$\Delta T_{\rm b} = \frac{T_{\rm sys}}{\eta n \sqrt{\Delta \nu t}} \left(\frac{\theta_{\rm P}}{\theta_{\rm S}} \right)^2 \qquad (23)$$

  allows you to check quickly if a source of given brightness $T_{\rm b}$ can be imaged at a given angular resolution $\theta_{\rm S}$ and spectral resolution $\Delta\nu$ (n is the number of antennas, $\theta_{\rm P}$ their primary beam width, and η an efficiency factor of order 0.5)
- $T_{\rm sys}$ is easy to guess: the simplistic value of 1 K per GHz of observing frequency is a good enough approximation in most cases.
- and **you know** $T_{\rm b}$ because you know the physics of your source!
- that is (almost) all you need to decide on the feasibility of an observation...

Part II: Low Signal to Noise

When is a source detected ?

What parameters can be derived ?

Low S/N: Continuum source

- Rule 1: do not resolve the source
- Rule 2: get the best absolute position before
- Rule 3: Use UV_FIT to determine the signal to noise ratio.
- if position accuracy better than 1/10th of beam
  - a 3 σ signal is sufficient to claim a detection.
  - Fix the position.
  - Use an appropriate source size.
- if position accuracy is about the beam
  - a 4 σ signal will be needed.
  - Do not fix the position.
  - Use an appropriate source size.
- if position is unknown
  - a 5 σ signal will be needed.
  - make an image to locate it.
  - Do not fix the position.
  - Use an appropriate source size.

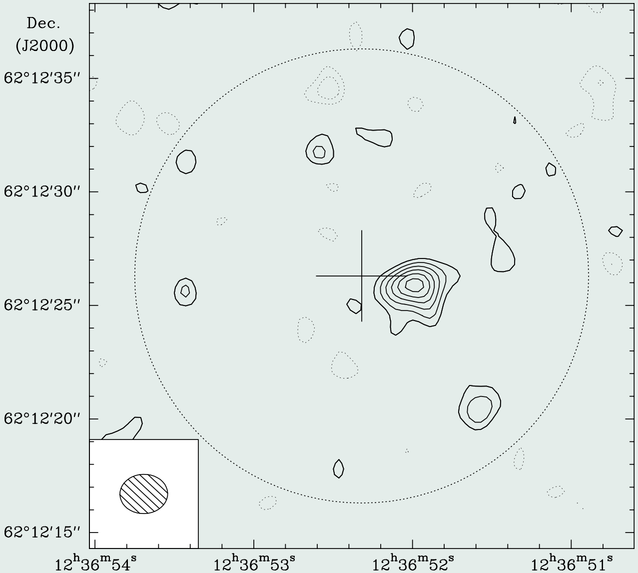
Continuum source parameters

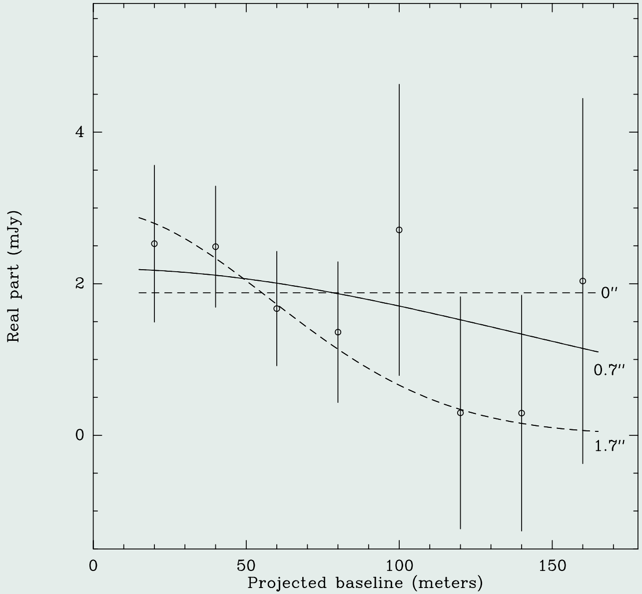
- **Rule of thumb**

  All fluxes are biased by 1 to 2 σ

- If position is free, flux is biased by 1 σ
- at least 4σ to get a position to 25 % of beam size
- With $< 6\sigma$, cannot measure any source size !
  - divide data in two, shortest baselines on one side, longest on another. Each subset get a 4.2σ error on mean flux.
  - Error on the difference is then just 3σ, i.e. any difference must be larger than 33 % to be significant
  - Mean baseline length ratio for the subsets is 3.
  - No smooth source structure can give a visibility difference larger than 30 % on such a baseline range ratio.
- If size is free, σ on flux increases **quite** significantly.

Example: HDF source





Left: 7σ detection of the strongest source in the Hubble Deep Field. Note that contours are *cheating* (start at 2 σ but with 1σ steps).
Right: Attempt to derive a size. Size can be as large as the synthesized beam... Note that the integrated flux increases with the source size.

Line sources

- **Things get even worse for spectral lines**
- Line velocity unknown: observer will select the brightest part of the spectrum $\longrightarrow$ bias
- Line width unknown: observer may limit the width to brightest part of the spectrum $\longrightarrow$ 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 !.
- These biases are all positive (noise is added to signal).
- 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...**

The correct approach

- Point source or unresolved source ($< 1/3$ of the beam)
  - Determine position (e.g. from 1.3 mm 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 spectral data
  - Fit line profile by Gaussian (with or without constant baseline offset, depending on whether the continuum flux is known or not)
- 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 $> 6\sigma$ to derive any meaningful information.
  - Strict minimum is 4σ (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 !...

Conclusions: for weak spectral lines

- Do not believe velocity gradient unless proven at a 5 σ level. Requires a S/N larger than 6 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 analogous of the Malmquist bias**

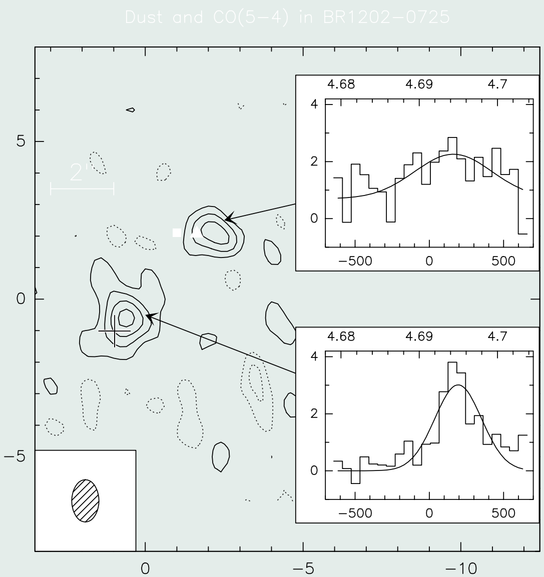
How to analyze weak lines ?

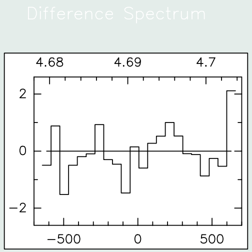
- Perform a statistical analysis (e.g. χ^2, or other statistical test) comparing **model prediction** to **observations, i.e. VISIBILITIES**
- The GILDAS software offer tools to compute visibilities from an image / data cube (`UV_FMODEL`)
- Beware that (original) channels are correlated ($\Delta\nu_N > \Delta\nu_C$)
- Appropriate statistical tests can actually provide a better estimate of the noise level than the prediction given by the weights.
- Up to you to develop the model adapted to your science case (and select the proper statistical tool for your measurement).

Examples

- Examples are numerous, specially for high redshift CO.
- e.g. 53 W002 :
  - OVRO (Scoville et al. 1997) claims an extended source, with velocity gradient. Yet the total line flux is 1.51 ± 0.2 Jy.km/s i.e. (at best) only 7 σ.
  - PdBI (Alloin 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...
- Remark(s)
  - But the images (contours) look convincing !
  - Answer : beware of "cheating" contours which start at 2 σ (sometimes even 3), but are spaced by 1 σ
  - But the spectrum looks convincing, too !
  - Answer : beware of "cheating" spectra, which are oversampled by a factor 2. The noise is then not independent between adjacent channels.

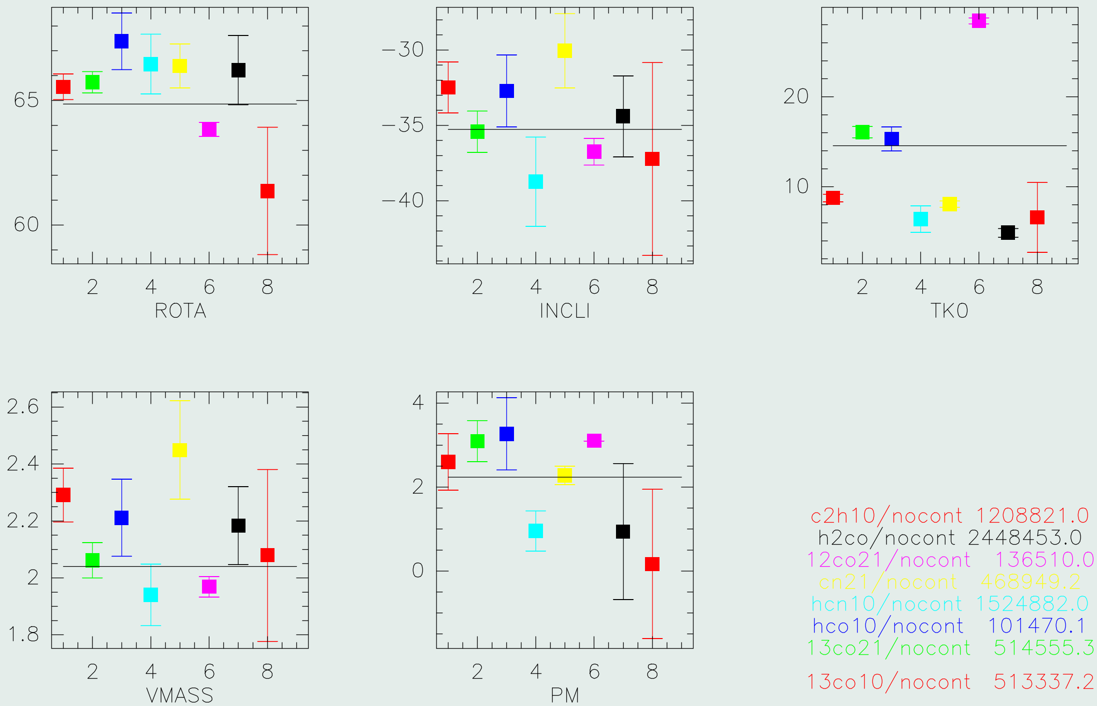
Example of Velocity Gradient: BR 1202-0725





- The image is a contour map of dust emission at 1.3 mm, with 2 σ contours
- The inserts are redshifted CO(5-4) spectra from the indicated directions
- 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.

Example of Analysis with Noise: DM Tau



- Error bars derived from a χ^2 analysis in the UV plane, using a line radiative transfer model for proto-planetary disks.