Chapter 19

Low Signal-to-noise Analysis

Stéphane Guilloteau

guillote@iram.fr

IRAM, 300 rue de la Piscine, F-38406 Saint Martin d'Hères, France

Low Signal to Noise (S/N) is unfortunately a frequent situation in mm astronomy. Since S/N increases only as the square root of the integration time, it may not be feasible at all to go from a case of low S/N to a case of even "decent" S/N. Thus the astronomer has to worry about interpreting properly low S/N data. The risk of over-interpretation in such cases is to be considered seriously. This lecture will give you some hints, as well as point out some traps which must be avoided.

19.1 Continuum Source

Let us start with a continuum source to simplify. The basic source parameters are position (x,y), flux density S_ν, and size. To determine the first 3 parameters, the best strategy is to avoid resolving the source. Since the position errors are proportional to the beam size, one should thus try to match the source size to the beam size. Having *a priori* information on the source position (by other observations, e.g. an optical image) will help to get a better accuracy on the source flux.

19.1.1 Flux measurement

- **Accurate Position:** If the position is known to better than $1/10^{\text{th}}$ of the beam, you should then use `UV_FIT` with the position fixed to determine the flux and the noise. In doing so, you should use an appropriate fixed source size, based on any *a priori* information you have (as you are following the "best strategy", this source size should be smaller than the beam size). In such a case, you only have one free parameter, the flux, and a 3σ signal is sufficient to claim a detection.

- **Rough Position** If the source position accuracy is not sufficient, you need to measure the position

Rule 1	**Do not resolve the source**		
Rule 2	**Get the best absolute position before**		
Rule 3	**Use `UV_FIT` to get the parameters and their errors**		
	a priori position accuracy		
	< 0.1 Beam	$\simeq$ Beam	Any
Minimum signal	$3\ \sigma$	$4\ \sigma$	$5\ \sigma$
Position	fixed	free	free, (make an image)
Source size	fixed	fixed	fixed

Table 19.1: Recipes to use `UV_FIT` to measure the flux of a weak source

also. As you now have more parameters to derive, a higher S/N will be required to do so. When the position uncertainty is about the beam size, a $4\ \sigma$ signal will be required to get a firm detection. Use `UV_FIT` to measure the flux and source position, with the same fixed source size as before.

- **Unknown Position** When the source position is really unknown, a $5\ \sigma$ signal may become necessary to claim a detection. To locate the source position, make an image first. Cleaning is usually not required at that level, unless the sidelobe level is higher than the noise to signal ratio. Then, use `UV_FIT` to measure the source flux, position and their associated errors, always using the same fixed source size as before.

Note that in all cases, the source size being used should be at least equal to the effective seeing of the observations, even if the source is actually a point source.

Table 19.1 summarizes the procedure to be followed. Once you have done your best in determining the source parameters, they remain to be properly interpreted. As a rule of thumb, remember that **All fluxes for detected weak sources are biased by 1 to 2 σ.** The only exception is when the source position and size is known *a priori.* The reason for the bias is very intuitive. Assume you have observed just enough to get a 3σ detection. A positive noise peak will bring that up to a 4σ value, a negative noise peak down to 2σ, which you will consider as a non-detection.

19.1.2 Other parameters

The other source parameters (position & size) require higher signal to noise to be determined. The position accuracy is the synthesized beam size divided by the S/N ratio. Hence, to get a position accuracy to 25 % of the beam size, at least a 4σ detection is required.

The above limitations are valid for a point source. If the source is not expected to be small enough, additional complications occur. If you have performed the experiment according to the guidelines given before (i.e. avoiding resolving the source), the source size may be just about the beam size. In such circumstances, no source size at all can be estimated with current mm interferometers if the detection is less than $6\,\sigma$. To convince yourself, let us perform a simple thought experiment. Assume we have detected a source at the $6\,\sigma$ level. Take this $6\,\sigma$ signal, and divide the observations in two equal (in sensitivity) data sets, one containing only the shortest baselines, the other ones only the longest baselines. Each subset has a $\sqrt{2}$ times higher noise level, and the error on the flux difference between these two data sets is 2 times the original noise level. Assume that the shortest baselines give us twice more flux than the longest one. In such a case, we would in fact have a better detection ($6.4\,\sigma$) with the short baselines only, but the difference flux is only measured with $3\,\sigma$. Such an experiment is not optimal from the detection point of view, since we would have obtained a better result ($6.4\,\sigma$) by observing only half of the time... Table 19.2 summarizes the corresponding numbers, and indicates that the minimum detection level to resolve a source at the 4σ level is $7.1\,\sigma$.

The interpretation of such data is made even more difficult by the fact that if the size is unknown, the error on the total flux increases **quite** significantly. Fig.19.1 shows the detection of a weak high-redshift object in the Hubble Deep Field area [Downes et al 1999] Although the detection is at the 7σ level, the source size is not constrained by these observations, and the total flux becomes uncertain by as much as 40 % when the uncertainty on the source size is included.

		Point Source			Beam Size Source			Minimum Size Resolved		
		Flux	Noise	S/N	Flux	Noise	S/N	Flux	Noise	S/N
(1)	Short baselines	6	1.4	4.2	9	1.4	6.4	10	1.4	7.1
(2)	Long baselines	6	1.4	4.2	3	1.4	2.1	2	1.4	1.4
(3)	Difference	0	2	0	6	2	3	8	2	4
(4)	Mean	6	1	6	6	1	6	6	1	6

Table 19.2: Signal to Noise example for source size measurement. Line (1) indicate the flux measured on short baselines, line (2) on long baselines, line (3) the difference between (1) and (2), and line (4) the average. Three cases are shown: a point source, a source with size similar to the beam, and the smallest source which can be resolved at the 4σ level.

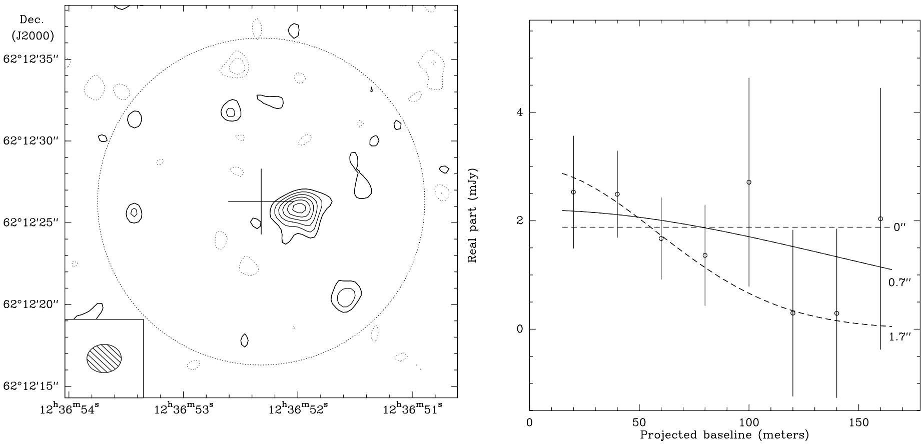

Figure 19.1: Left: 7σ detection of the strongest source in the Hubble Deep Field. Note that the contours are visually misleading (they start at 2 σ but with 1σ steps, given the impression of a much better detection). Right: Attempt to derive a size. Size can be as large as the synthesized beam... Note that the integrated flux increases with the source size.

19.2 Spectral Line Sources

Unfortunately, things get even worse for spectral lines, because the uncertainties on the line width and source velocity add up to the position and size problem. If the source velocity is unknown, the observer will tend to select the brightest part of the spectrum to define the integrated flux. This result in a positive bias on the flux. Furthermore, if the line width is not known, the observe may limit the line to the brightest part of the spectrum, resulting in another bias. This bias is in general positive, since positive noise peaks will be included in the line region, but could be negative for specific line shapes.

If the source position was *a priori* unknown, it is common practice to determine it from the integrated line flux map made using the tailored line window specified by the astronomer. Such a procedure results in a positively biased total flux. Any speculated extension will also increase the total flux, by enlarging the selected image region by selection of positive noise peaks. The net effect is a 1 to 2 σ positive bias on the integrated line flux. Things get really messy if a continuum is superposed to the weak line...

A good strategy is required to minimize these biases. The correct approach to point sources (or sources less than about $1/3^{\mathrm{rd}}$ of the synthesized beam), is to first determine the position (e.g. from continuum data is available, or from the integrated line map if not, or ideally from other data). Once the position is fixed, the line profile can be derived by fitting a point (or small fixed size) source, at fixed position into the UV data. Using this line profile, the total line flux, as well as source velocity and line width, can be derived by fitting an appropriate lineshape, e.g. a Gaussian if no other information is available. In this last step, a constant baseline offset should be added if there is a continuum contribution.

For extended sources, which may be affected by velocity gradient, one has to fit a multi-parameter (6 for an elliptical gaussian) source model for each spectral channel into the UV data. As a consequence, the signal in each channel should be at least $6\,\sigma$ to derive any meaningful information. The strict minimum is 4σ (per line channel...) to get flux and position for a fixed size source. Velocity gradients are not believable unless even better signal to noise is obtained per line channel !. . . Moreover, for narrow lines, most correlators produce spectral channels which are not independent; the correlation between adjacent channels should be taken into account when analyzing velocity gradients.

To sum up the weak spectral line problems:

- Do not believe velocity gradient unless proven at a 5 σ level. This 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
- or even more biased if the source is extended

These biases are the analogous of the Malmquist bias.

Unfortunately, examples for such problems are numerous, especially for high redshift CO lines. The $z = 2.8$ galaxy 53 W 002 was detected in CO with the OVRO interferometer by [Scoville et al 1997], who claim an extended source, with velocity gradient. The published images (contour maps) look convincing, but this is biased by the chosen visual representation, with contours starting at 2σ but spaced by 1σ. This creates the visual impression of higher S/N ratio. The published spectra also looks convincing, but are presented as a fully sampled spectra (i.e. channel width equal to twice the channel separation). Although this is the proper way to present a complete information, the astronomer's eye is not accustomed to such a presentation, and the astronomer tends to interpret the data as if the channels were independent, thereby underestimating the noise. Yet the total line flux is 1.5 ± 0.2 Jy.km/s i.e. (at best) only 7 σ, and thus, according to the discussion presented above, no extension/gradient should be measurable. Indeed, using the IRAM interferometer, [Alloin et al 2000] find 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, with the second determination just 1σ below the first one, as expected for an initially biased result...

Another example of visually misleading result is shown in Fig.19.2. Although the two spectra appear different, there is a weak continuum (which was measured independently) on the Northern source. Once the continuum offset and a scale factor have been applied, the lack of visible structure in the difference

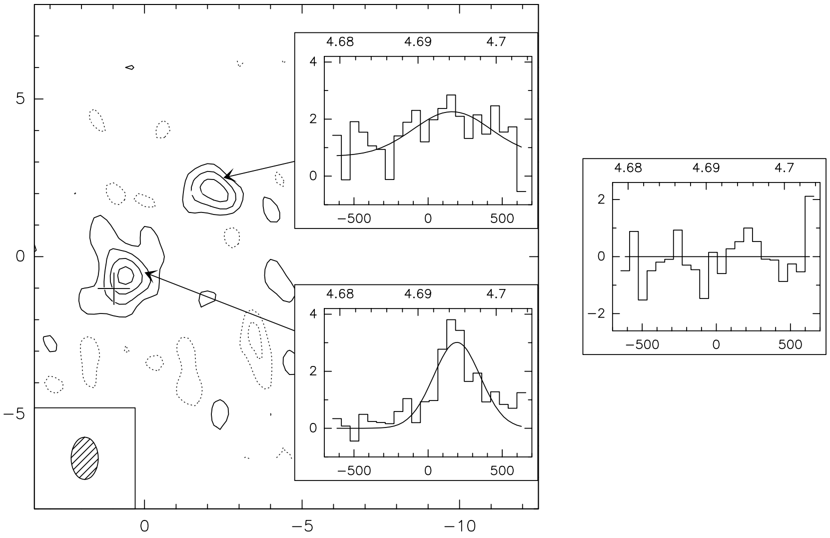

Figure 19.2: Example of search for a Velocity Gradient: BR 1202-0725. The image is a contour map of dust emission at 1.3 mm, with 2 σ contours. The inserts are redshifted CO(5-4) spectra from the indicated directions. A weak continuum (measured `independently`) exist on the Northern source. The rightmost insert is **a** difference spectrum (with a scale factor applied, and continuum offset removed) (Cox, Guélin, Guilloteau & Omont, in preparation).

spectrum shows that both line profiles are indistinguishable, i.e. that there is no measurable velocity gradient. These two sources could be lensed images of the same galaxy...