

Dealing with Noise



Stéphane GUILLOTEAU

Laboratoire d'Astrophysique de Bordeaux Observatoire Aquitain des Sciences de l'Univers

I - Theory & Practice of noise

II – Low S/N analysis





1. Basic Theory

- 1. Point source sensitivity
- 2. Noise in images
- 3. Extended source sensitivity
- 4. Available Tools
- 2. Low S/N analysis
 - 1. Continuum data
 - 2. Line data
 - 3. Examples
 - 4. Advanced tricks: filtering & stacking





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

$$\mathsf{P}_{\mathsf{N}}$$
 = γ k $\mathsf{T}_{\mathsf{ant}} \varDelta \nu$

On source, the power is P_N + P_a with

$$P_a = \gamma \ k \ T_a \ \Delta \ \nu$$

- 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



T_{ant} is given by (just summing powers...)

- $\begin{array}{ll} \mathsf{T}_{\mathsf{ant}} = \mathsf{T}_{\mathsf{bg}} & \operatorname{cosmic} \mathsf{background} \\ & + \mathsf{T}_{\mathsf{sky}} & \approx \eta_f \left(1{\text{-exp}}({\text{-}}\tau_{\mathsf{atm}}) \, \mathsf{T}_{\mathsf{atm}} \right) \\ & + \mathsf{T}_{\mathsf{spill}} & \approx \left(1{\text{-}}\eta_f{\text{-}}\eta_{\mathsf{loss}} \right) \mathsf{T}_{\mathsf{ground}} \\ & + \mathsf{T}_{\mathsf{loss}} & \approx \eta_{\mathsf{loss}} \, \mathsf{T}_{\mathsf{cabin}} \\ & + \mathsf{T}_{\mathsf{rec}} & \operatorname{in} \operatorname{receiver} \mathsf{noise} \end{array}$
- This is a broad-band definition. It is a DSB (Double Side Band) noise 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:

$$T_{sys} = (1+g) \exp(\tau_{atm})T_{ant} / \eta_{f}$$
$$T_{A}^{*} = (1+g) \exp(\tau_{atm})T_{a} / \eta_{f}$$

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

 $T_A^* = \eta_a A S_\nu / 2k$





• 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

$$R_{sn} = (2\Delta\nu \Delta t)^{1/2} T_A^* / T_{sys}$$

On a single baseline, the noise is thus

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

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



Noise Equation



With quantization

• Where $\mathcal{J} = \frac{2k}{n A}$

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

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

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

is the Jy/K conversion factor of one antenna





- For 1 baseline, this varies with Signal to Noise ratio
- On Amplitude



$$S \ll \sigma \begin{cases} \sigma_A \simeq \sigma \sqrt{2 - \frac{\pi}{2}} \left(1 + \left(\frac{S}{2\sigma}\right)^2 \right) \\$$

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

- On Phase
- Source detection is much easier on the phase than on the amplitude, since for S/N = 1, σ_{ϕ} = 1 radian = 60°.





- 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 $\langle \epsilon_r^2 \rangle = \langle \epsilon_i^2 \rangle = \langle \epsilon_i^2 \rangle$
 - > Real and Imaginary are uncorrelated $<\epsilon_r \epsilon_i > = 0$
- So rotation (phase factor) has NO effect on noise

$$\begin{split} \varepsilon_{\rm R}' &= \varepsilon_{\rm R} \cos(\phi) - \varepsilon_{\rm I} \sin(\phi) \\ \varepsilon_{\rm I}' &= \varepsilon_{\rm R} \sin(\phi) + \varepsilon_{\rm I} \cos(\phi) \\ \langle \varepsilon_{\rm R}'^2 \rangle &= \langle \varepsilon_{\rm R}^2 \rangle \cos^2(\phi) - 2 \langle \varepsilon_{\rm R} \varepsilon_{\rm I} \rangle \cos(\phi) \sin(\phi) + \langle \varepsilon_{\rm I}^2 \rangle \sin^2(\phi) = \langle \varepsilon^2 \rangle \\ \langle \varepsilon_{\rm R}' \varepsilon_{\rm I}' \rangle &= \langle \varepsilon_{\rm R}^2 \rangle \cos(\phi) \sin(\phi) - \langle \varepsilon_{\rm I}^2 \rangle \cos(\phi) \sin(\phi) = 0 \end{split}$$





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

$$I = (\Sigma w_i V_i) / \Sigma w_i$$

- for a point source at phase center, $V_i = V + \epsilon_{Ri}, \, \epsilon_{Ri}$ being the real part of the noise

$$I = (\Sigma w_i (V + \varepsilon_{Ri})) / \Sigma w_i$$

- So its expectation is I = V, as $< \epsilon_{Ri} > = 0$
- As $\langle \epsilon_{Ri} | \epsilon_{Rj} \rangle = 0$, its variance is

$$\sigma^2 = \langle I^2 \rangle - \langle I \rangle^2 = = (\Sigma w_i^2 \langle \epsilon_{Ri}^2 \rangle) / (\Sigma w_i)^2$$

- Now using $<\epsilon_{Ri}{}^2>=\sigma_i{}^2\,$ and the natural weights w_i = 1/ $\sigma_i{}^2$ we have

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

Which is true anywhere else in the image by application of a phase shift





When using non-natural weights (w_i # σ_i²), either as a result of Uniform or Robust weighting, or due to Tapering, the noise (for point sources) increases by w_{rms} / w_{mean}

 $w_{\text{rms}} = ((\Sigma(WT)^2)/n)^{1/2}$ $w_{\text{mean}} = (\Sigma WT)/n$

- Robust weighting improves angular resolution
- Tapering can be used to smooth data



Noise in Imaging



- 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...







- 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



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

- This is right only for sources just filling one synthesized beam θ_s .
- For more extended sources, it is not appropriate to count the number of synthesized beams n_b and divide by √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.





- 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 $\Delta \nu_{\rm N}$ is not the channel separation $\Delta \nu_{\rm C}$, neither the effective resolution $\Delta \nu_{\rm R}$
- These effects are of order 15-30 % on the noise.
- In practice, $\Delta \nu_{\rm N} > \Delta \nu_{\rm C}$, 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 $\Delta \nu_{\rm N} = \Delta \nu_{\rm C} = \Delta \nu_{\rm R}$, and a $\sqrt{n_{\rm c}}$ noise averaging when smoothing





- 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
 - ➢ so, may be: it depends on your scientific case...
 - > Weighting could take into account a spectral index, for example...
- For line data
 - No: could degrade S/N if the line shape is not consistent with the weights
 - No: undefined bandwidth: does not allow to compute an integrated line flux
- In practice: not implemented in current GILDAS software. Could be useful for specific weak source searches. See "Optimal Filtering" later





- Each visibility is affected by a random atmospheric phase
- Assuming a point source at the phase center,

$$= V e^{i\phi_i} + \varepsilon_{Ri}$$

$$I = (\sum w_i (V e^{i\phi_i} + \varepsilon_{Ri})) / (\sum w_i)$$

- the expectation of / is now only $V e^{-(\Delta \phi)^2/2}$
- 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.
- The result depends on the source structure.
- There is so far no good simulation tool to evaluate the importance of this effect. It is not fully random at Plateau de Bure...



Estimating the Noise



- The weights are used to give a prediction of the noise level in the images.
- Predictions displayed by UV_MAP and UV_STAT
- Carried on in the image headers (aaa1%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).



Noise on Mosaics...



- GO NOISE does (yet) not work on mosaics...
- Because noise is NOT uniform on mosaics...
- $J = \Sigma B_i F_i / \Sigma B_i^2$
- Let us define $W = \sum B_i^2$
- If we instead use $L = J / W^{1/2}$
 - > The noise on L is uniform (provided all fields had similar noise) of value σ_L
 - It corresponds to the noise at the most sensitive place in the mosaic
 - > L/σ_L is a signal-to-noise image
- Valid also for 1 field mosaic...



Conclusions



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

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

- allows you to check quickly if a source of given brightness T_b can be imaged at a given angular resolution θ_S and spectral resolution $\Delta \nu$ (n is the number of antennas, θ_P their primary beam width, and η an efficiency factor of order 0.5 – 0.8, and t the integration time...)
- T_{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_b because you know the physics of your source!
- that is (almost) all you need to decide on the feasibility of an observation...





- When is a source detected ?
- What parameters can be derived ?



Low Signal to Noise



A nice case

- > Observers advantage
 - You 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
 - > But with proper analysis, you may still invalidate (falsify) some model/theory
- So let us see...



Low S/N -- Continuum



- Rule 1: do not resolve the source
- Rule 2: get the best absolute position before
- Rule 3: Use UV_FIT to determine the S/N ratio
- Rule 4: the 3-4-5 rule about position accuracy
- < 1/10th of beam

About the beam

- Fix the position
- Use an appropriate source size
- $>3 \sigma$ signal for detection $>4 \sigma$ signal for detection
 - Do not fix the position
 - Use an appropriate source size

Unknown

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





- Sources of unknown positions have fluxes biased by 1 to 2 σ
- Free position 1 σ bias
- Position accuracy = beam/(S/N ratio)
- With < 6 σ , cannot measure any source size
 - > divide data in two, shortest baselines on one side, longest on another. Each subset get a 4.2 σ error on mean flux.
 - > Error on the difference is then just 3 σ , i.e. any difference must be larger than 33 % to be significant
 - Mean baseline length ratio for the subsets is at best 3.
 - No smooth source structure can give a visibility difference larger than 30 % on such a baseline range ratio.
- If size is free, error on flux increases quite significantly



Example: HDF source





 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). 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 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 !.
- All these biases are 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...





- Point source or unresolved source (< 1/3rd 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
 - > 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)



- 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...





- 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



- 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 1.5 ± 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...







- 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 1.5 \pm 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 visually confusing contours which start at 2 σ (sometimes even 3), but are spaced by 1 σ





- 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 1.5 ± 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 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.







- e.g. 53 W002 :
 - > OVRO (Scoville 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 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 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.











- 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.

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 (task UV_FMODEL)
- Beware that (original) channels are correlated ($\Delta \nu_{\rm N} > \Delta \nu_{\rm C}$)
- 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).
- GILDAS even provides minimization tools: the ADJUST command (but with no guarantee of suitability to your case, though. Expertise recommended !).







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







 A typical data cube from which the previous parameters were derived. It has quite decent S/N, and one can recognize the rotation pattern of a Keplerian disk





A (really) low Signal to Noise image of the protoplanetary disk of DM Tau in the main group of hyperfine components of the N₂H⁺ 1-0 transition.
It really looks like absolute nothing... but a treasure is hidden inside the noise!





- Best fit integrated profile for the N₂H⁺ 1-0 line, derived from a χ² 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.
- The observed spectrum is the integrated spectrum over a 6x6" area (from the Clean or Dirty image, does not really matter). The noise is about 11 mJy.







- Signal-to-noise maps of the integrated N₂H⁺ 1-0 line emission, using the best profile derived from the χ² analysis in the UV plane as a (velocity) smoothing kernel (optimal filtering).
- 7 σ detection for DM Tau, 6 σ detection for LkCa 15
- Nothing for MWC 480





- ALMA is only 7 times more sensitive than PdB (at 3mm, better ratio at higher frequencies)
- on the N₂H⁺ case, it will (in a mere 8 hours), obtain a peak 10 σ detection per channel, which is quite good, but will barely "see" the weakest hyperfine components.
- but if the resolution is increased just to 2", the S/N will drop by a factor 3 (in this favorable case, as the structure remain unresolved in one direction...)
- and a search for the ¹⁵N substitute remain beyond (reasonable) reach !.
- This is a simple molecule. Things a little more complex, e.g. HCOOH, HC₃N will be tough
- you can transpose this example for extragalactic studies



Optimal Filtering



- Changing the frequency dependence of weights and signal to adjust for a continuum spectral index
- Convolve by expected line profile for blind line search
- If line profile unknown, convolve by several possible ones, and see if one convolution leads to a significant signal