Extracting information from images and uv-data

Andrei Lobanov

MPIfR Bonn

European Radio Interferometry School
Bonn, 13/09/2007

Noise and measurement

- Assessing image quality

- Noise in interferometry
  -- “irreducible” factors
  -- “controllable” factors

- Practical estimates of the expected noise in an image

- Quantifying the brightness distribution in an image
  -- information and its limits in interferometric images
  -- representing the observed structure; model fit
  -- error estimates for model fits

Assessing image quality

- **Fit to baseline visibilities:**
  -- should represent adequately amplitudes, phases, and phase closures.

- **Residual flux:**
  -- should be distributed smoothly, with a nearly zero mean and comparable positive and negative amplitudes.

- **Noise in the final image:**
  -- should have a Gaussian distribution;
  -- should be approaching the thermal noise level, corrected for the bandwidth and time-average smearing and modified by self-calibration.

Noise in interferometric images

$V(u,v)$ – visibility data; $\epsilon(u,v)$ – associated errors
$I(l,m)$ – image brightness distribution; $\mathbf{F}$ – Fourier transform operator.

- **Additive errors**, $V + \epsilon \rightleftharpoons I + \mathbf{F}\,\epsilon$
  -- *system noise, interference, cross-talk, baseline-dependent errors*
- **Multiplicative errors**, $V\,\epsilon \rightleftharpoons I * \mathbf{F}\,\epsilon$
  -- *uv-coverage, gain calibration errors, atmospheric and ionospheric errors*
- **Convolution errors**, $V * \epsilon \rightleftharpoons I\,\mathbf{F}\,\epsilon$
- **Position dependent errors**
  -- *pointing errors, bandwidth and time-averaging smearing*

Noise in interferometric images

During data processing, the noise is modified by

- Gridding and convolution
- Averaging in frequency and time
- Editing
- Tapering and weighting
- Deconvolution (CLEAN-ing)
- Self-calibration

Thermal noise

S – source: total flux density

A, $T_{\rm sys}$, $\eta_{\rm a}$, ϵ – antenna: Area, System temperature, Efficiency, and Errors (phase [radians], amplitude [fractional])

N, $\eta_{\rm c}$ – array: Number of antennas, Correlator efficiency

$t_{\rm int}$, $t_{\rm obs}$, $\Delta\nu$ – observation: Integration time, Observing time, Observing band-width

1. Dynamic range of the observation: $$D \approx \frac{1}{\epsilon}\sqrt{\frac{N(N-1)}{2}}$$

2. Thermal errors in the antenna phase and amplitude $$\epsilon_{\rm th} = \frac{\sqrt{2}k_{\rm B}T_{\rm sys}}{A\,\eta_{\rm a}S\sqrt{t_{\rm obs}\Delta\nu}}$$

3. Thermal noise on a baseline $$\Delta S_{\rm m} = \frac{\epsilon_{\rm th}}{\eta_{\rm c}}S$$

4. Thermal noise in the image $$\Delta\, I_{\rm m} = \Delta S_{\rm m}\sqrt{\frac{2\,t_{\rm int}}{t_{\rm obs}N(N-1)}}$$

Bandwidth smearing

ν_0, $\Delta\nu$ — observing frequency and bandwidth; finite $\Delta\nu/\nu_0$; averaging across $\Delta\nu$: $u_0 = u_\nu(\nu_0/\nu)$, $v_0 = v_\nu(\nu_0/\nu)$ — chromatic aberration.

Effects:

Consider Gaussian bandpass and circular Gaussian tapering.

1. Beam degradation across the image:

$$B_\mathrm{D}(\Delta\theta, \theta_0) = \frac{1}{\sqrt{\beta_{\nu\theta}}} \exp\left(-\frac{\alpha_\theta^2}{\beta_{\nu\theta}}\right),$$

with

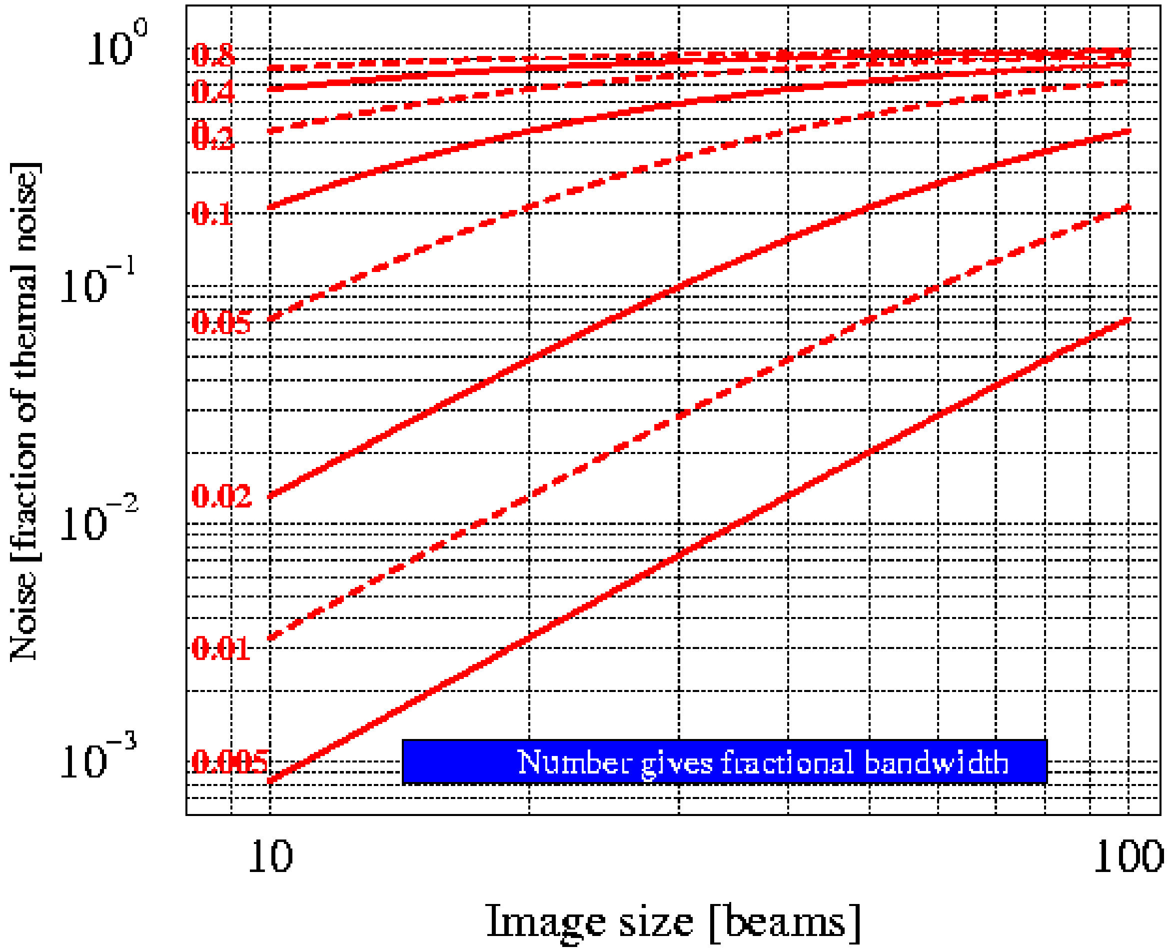
$$\alpha_\theta = 2\ln 2 \frac{\Delta\theta}{\theta_\mathrm{HPBW}}, \quad \beta_{\nu\theta} = 1 + \left(\frac{\Delta\nu}{\nu_0}\frac{\theta_0}{\theta_\mathrm{HPBW}}\right)^2$$

2. Reduction in the peak response: $$R_{\Delta\nu} = \frac{I}{I_0} = \frac{1}{\sqrt{\beta_{\nu\theta}}}$$

Bandwidth smearing

3. Average contribution to image noise in an image of size $\Omega_{\rm I}$:

$$\langle \sigma_{\Delta\nu} \rangle = \frac{\Delta I_{\rm m}}{\Omega_{\rm I}} \int_0^{\Omega_{\rm I}} R_{\Delta\nu} {\rm d}\theta = \frac{\Delta I_{\rm m}}{\nu' \Omega'_{\rm I}} {\rm arcsinh}(\nu' \Omega'_{\rm I}) \,,$$ where $\Omega'_{\rm I} = \Omega_{\rm I}/\theta_{\rm HPBW}$, and $\nu' = (\Delta\nu)/\nu_0$.



Time average smearing

τ_a – averaging time: $(u,v)|_{t_0-0.5\tau_\mathrm{a}}^{t_0+0.5\tau_\mathrm{a}} \Rightarrow (u,v)|_{t_0}$

Effects:

1. Largest detectable structure on a baseline B_L

$$\theta_\mathrm{max} = \frac{c}{\nu_\mathrm{obs}\omega_\mathrm{e}\tau_\mathrm{a}B_\mathrm{L}}$$

$$\omega_\mathrm{e} = 7.25 \times 10^{-5}\ \mathrm{rad/sec}$$

2. Average baseline amplitude reduction over a 12–hour period, for a sky location (l,m)

$$\langle R^B_{\tau_\mathrm{a}} \rangle = \langle I/I_0 \rangle \approx 1 - \frac{\pi^2}{12\theta^2_\mathrm{max}}(l^2 + m^2 \sin^2\delta)$$

3. Average amplitude reduction in an image location offset by θ from the phase–tracking center

$$\langle R_{\tau_\mathrm{a}} \rangle = 1 - \frac{\alpha\pi^2}{12}\omega_\mathrm{e}^2\tau_\mathrm{a}^2\left(\frac{\theta}{\theta_\mathrm{HPBW}}\right)^2,$$

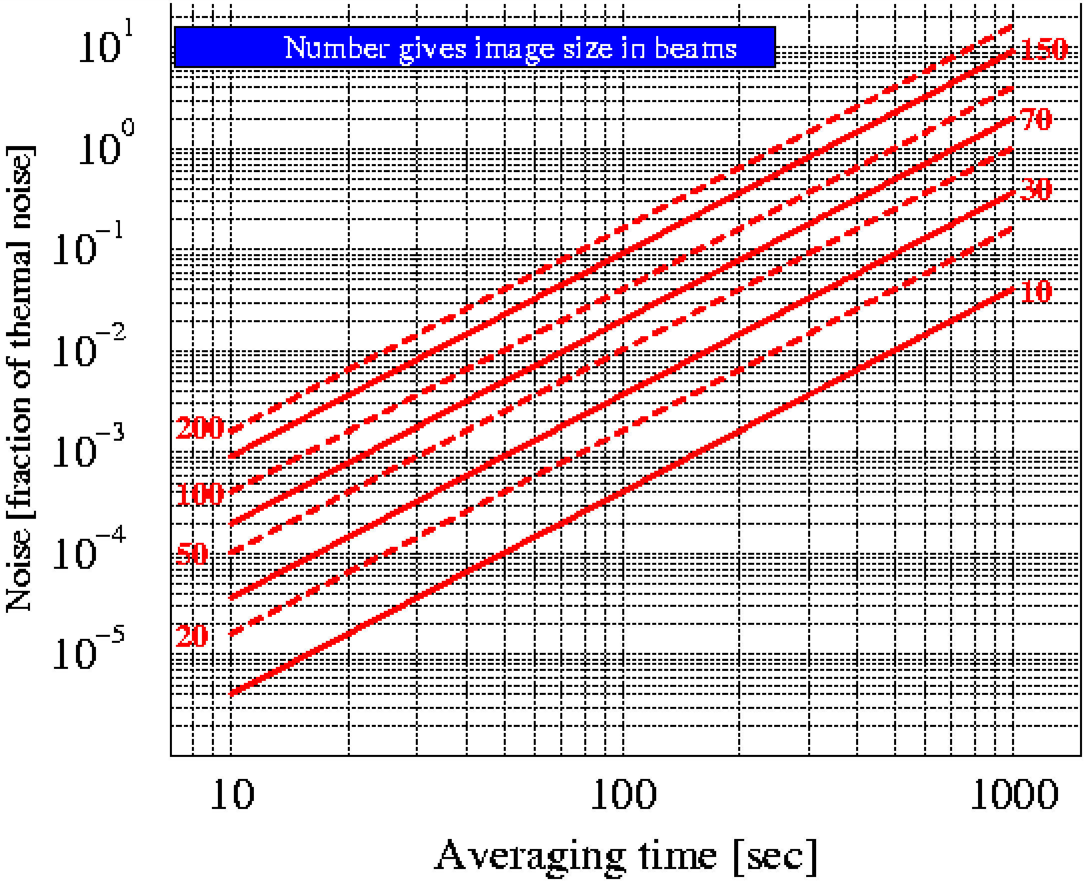
with $\alpha = 4(\ln 2)/\pi^2$, for a circular uv–coverage with Gaussian taper.

Time average smearing

4. Average contribution to image noise, for an image of size $\Omega_{\rm I}$

$$\langle\sigma_{\rm avg}\rangle = \frac{\Delta I_{\rm m}}{\Omega_{\rm I}}\int_0^{\Omega_{\rm I}}\langle R_{\tau_{\rm a}}\rangle {\rm d}\theta = \Delta I_{\rm m}\left(1 - \frac{\ln 2}{9}\omega_{\rm e}(\tau_{\rm a}\Omega'_{\rm I})^2\right)\,,$$

where $\Omega'_{\rm I} = \Omega_{\rm I}/\theta_{\rm HPBW}$



Self-calibration

1. Antenna phase and amplitude errors due to self-calibration on a timescale $\tau_{\rm cor}$

$$\epsilon_{\rm g}(\tau_{\rm cor}) = \frac{\Delta S_{\rm m}(\tau_{\rm cor})}{S\sqrt{N(N-1)}} = \frac{\Delta I_{\rm m}}{S}\sqrt{\frac{t_{\rm obs}}{2\,\tau_{\rm cor}}}$$

2. Combined antenna errors

$$\epsilon_\tau = \epsilon_{\rm th}\sqrt{1 + \frac{t_{\rm obs}}{\tau_{\rm cor} N(N-1)}}$$

3. Expected contribution by self-calibration to the noise level in the image

$$\sigma_{\rm g} = \Delta I_{\rm m}\sqrt{\frac{t_{\rm obs}}{\tau_{\rm cor} N(N-1)}}$$

4. Expected dynamic range in the self-calibrated image

$$D_{\rm g} = \frac{S}{\Delta I_{\rm m}\sqrt{1 + \frac{t_{\rm obs}}{\tau_{\rm cor} N(N-1)}}}$$

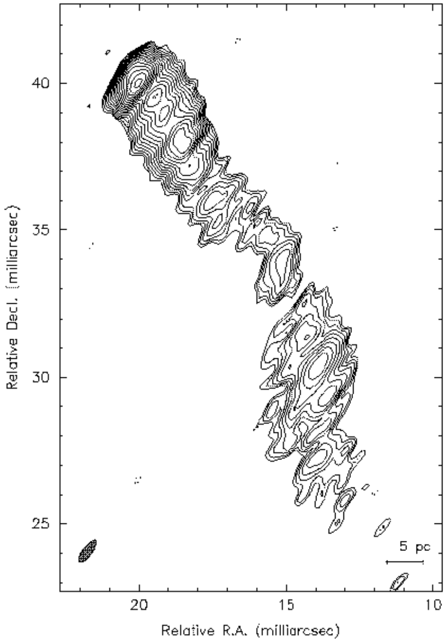
Example: An observation lasting for $t_{\rm obs} = 12$ hours, with a homogeneous array of 10 telescopes; self-calibration on a timescale of $\tau_{\rm cor} = 1$ minute. The resulting expected noise level is $I_{\rm g} \approx 3\, I_{\rm m}$.

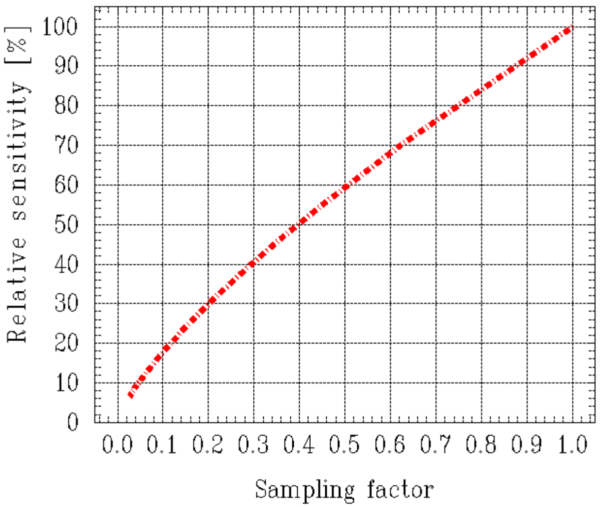
Gridding and convolution

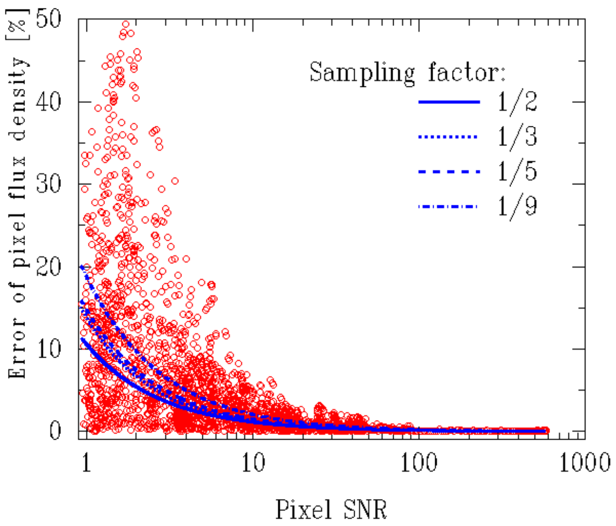
- *uv-coverage is incomplete and irregular:* needs to be projected onto a regular (rectangular) grid
- *convolution:* is used for interpolating between the observed distribution (u_0, v_0) and gridded distribution (u_k, v_k).
- convolution function is chosen so that it suppresses the responses to sources outside the image (a suppression factor of ~100 is typically reached).
- *Simple recipe:* avoid undersampling and bright sources near the edges of the image
  -- then most of the noise due to convolution can be removed during deconvolution, and remaining (irreducible) addition to the image noise should be within $\sim 0.01 \Delta I_{\rm m}$

Effect of uv-sampling

8600-km baseline																
ν	structure size (mas)								sampling interval (min.)							
[GHz]	1	3	5	10	20	30	50	100	10	15	20	25	30	35	40	50
43.2	19	6.5	4.0	2.0	1.0	0.5	0.3	0.2	1.9	1.3	1.0	0.8	0.7	0.6	0.5	0.4
22.2	36	12	7.0	3.5	1.8	1.2	0.7	0.4	3.5	2.4	1.8	1.4	1.2	1.0	0.9	0.7
15.1	55	18	11	5.5	2.7	1.8	1.1	0.6	5.5	3.8	2.7	2.2	1.8	1.6	1.4	1.1
8.4	*110*	38	22	11	5.5	4.0	2.5	1.0	11	7.0	5.5	4.4	3.8	3.2	2.7	2.2
5.0	*160*	55	32	16	8.0	5.5	3.0	1.5	17	11	8.0	6.5	5.5	4.6	4.0	3.2
2.3	*360*	*120*	70	35	18	12	7.0	3.5	35	24	18	14	12	10	9.0	7.0
1.6	...	*180*	*110*	55	27	18	11	5.5	55	38	27	22	18	16	14	11
0.6	...	...	*280*	*140*	70	46	28	14	140	95	70	55	47	40	35	28
0.3	...	...	...	*240*	*120*	80	50	25	240	160	120	100	80	70	60	50
	maximum sampling interval (min.)								largest detectable structure (mas)							







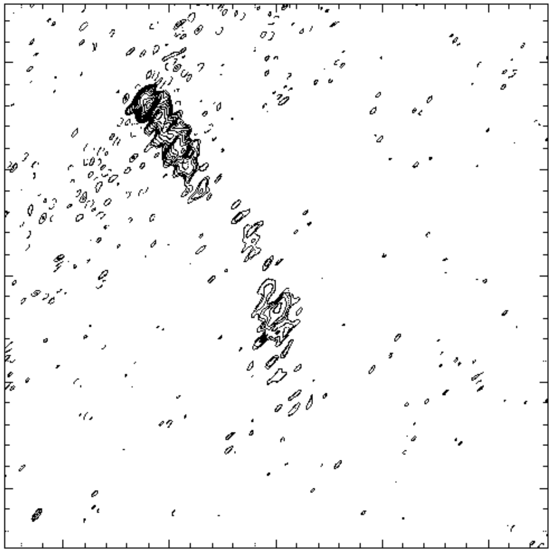


Image dimensions

Image dimensions are set by the requirements for a successful gridding and by:

1. the observed emission distribution
2. the ranges of observing uv–coverage: $u_{\rm min} \le u \le u_{\rm max}$, $v_{\rm min} \le v \le v_{\rm max}$.

Effects:

Image dimensions must satisfy the following conditions:

pixel size:

$$\Delta l \le \frac{1}{2u_{\rm max}}, \quad \Delta m \le \frac{1}{2v_{\rm max}}$$

number of pixels along each axis:

$$N_l \ge \frac{1}{\Delta l u_{\rm min}}, \quad N_m \ge \frac{1}{\Delta m v_{\rm min}}$$

Relation between the pixel size, $\theta_{\rm pix}$, and the size of the syntesized beam, $\theta_{\rm HPBW}$:

$\theta_{\rm pix} \approx (1/3)\theta_{\rm HPBW}$ limits the dynamic range to $D \le 10^4$

$\theta_{\rm pix} \le (1/6)\theta_{\rm HPBW}$ is sufficient for any $D \le S/(\Delta I_{\rm m})$

Weighting and tapering

A. Lobanov

$V(u_k, v_k),\ k = 1, M$ – measured visibility distribution

It needs to be sampled, weighted, and (normally) evaluated on an $N \times N$ grid.

The resulting distribution

$$V^{\mathrm{W}}(u,v) = \sum_{k=1}^{M} R_k\, T_k\, D_k\, \delta(u - u_k, v - v_k) V(u_k, v_k)$$

R_k – visibility weights

T_k – tapering weight (tapering function); typically, a Gaussian defined in terms of the uv-distance $q = \sqrt{u^2 + v^2}$

$$T_k(q) = \exp(-q^2/2\sigma_{uv}^2)$$

This results in the synthesised beam of $\theta_{\mathrm{HPBW[mas]}} = \dfrac{76.3}{\sigma_{uv[\mathrm{M}\lambda]}}$

D_k – density weighting function.

$D_k = 1$ – natural weighting; minimizes the noise level in the image

$D_k = 1/N_k$ – uniform weighting; minimizes rms of the sidelobes of the synthesised beam

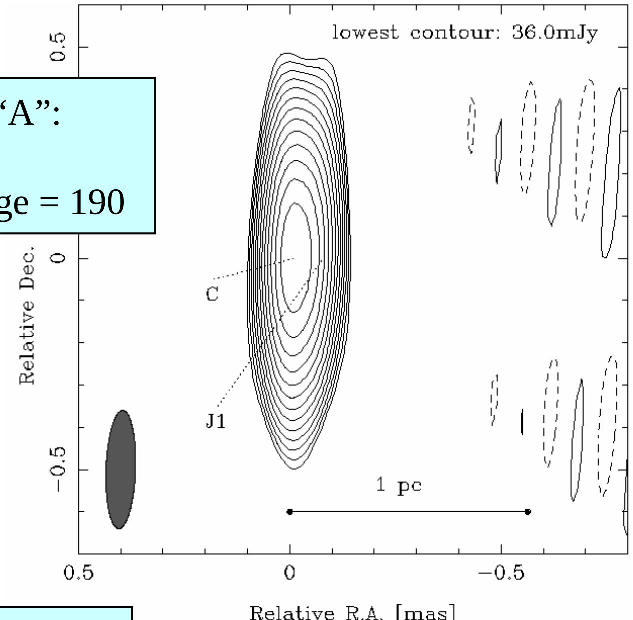
Deconvolution

- has to be used because the visibility data are incomplete... which leads to diffraction patterns in the image plane that cannot be removed either by direct Fourier inversion, or by linear methods
- non-linear deconvolution has to be applied to correct for the diffraction patterns.
- a number of complementary deconvolution algorithms exist (CLEAN, MEM, etc.) that can be applied to specific obser-vational setups and particular brightness distributions.
- a successful deconvolution does not make a significant contribution to the noise level
- problems may arise with the distribution of noise.

A. Lobanov

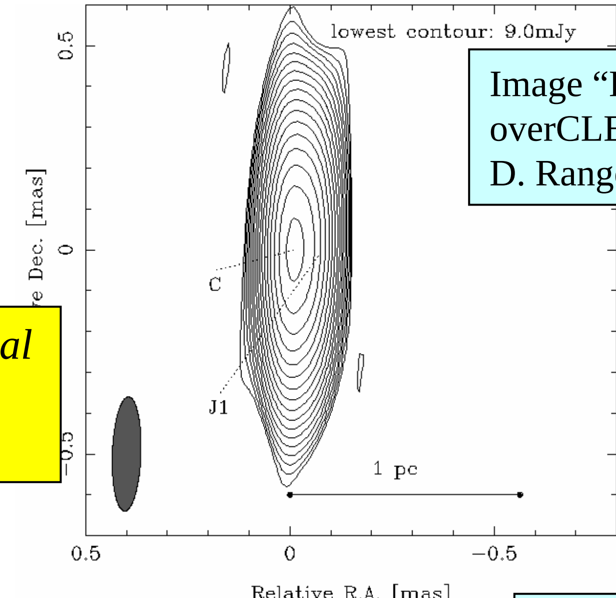
Excessive use of CLEAN

Image "A":
proper
D. Range = 190

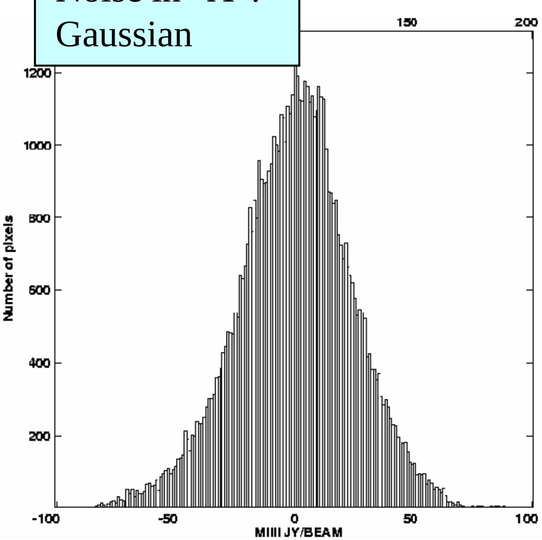


Estimated thermal noise: 16mJy/beam

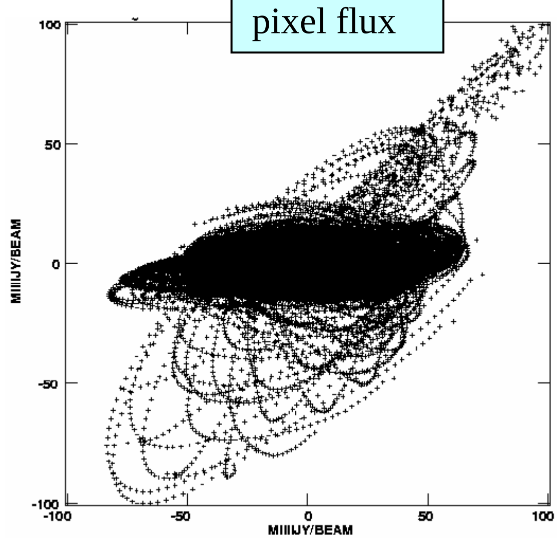
Image "B":
overCLEANed
D. Range = 750



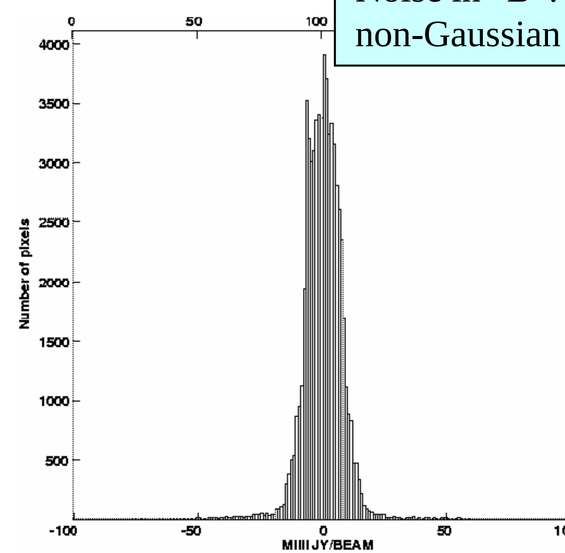
Noise in "A":
Gaussian



"B" vs. "A
pixel flux

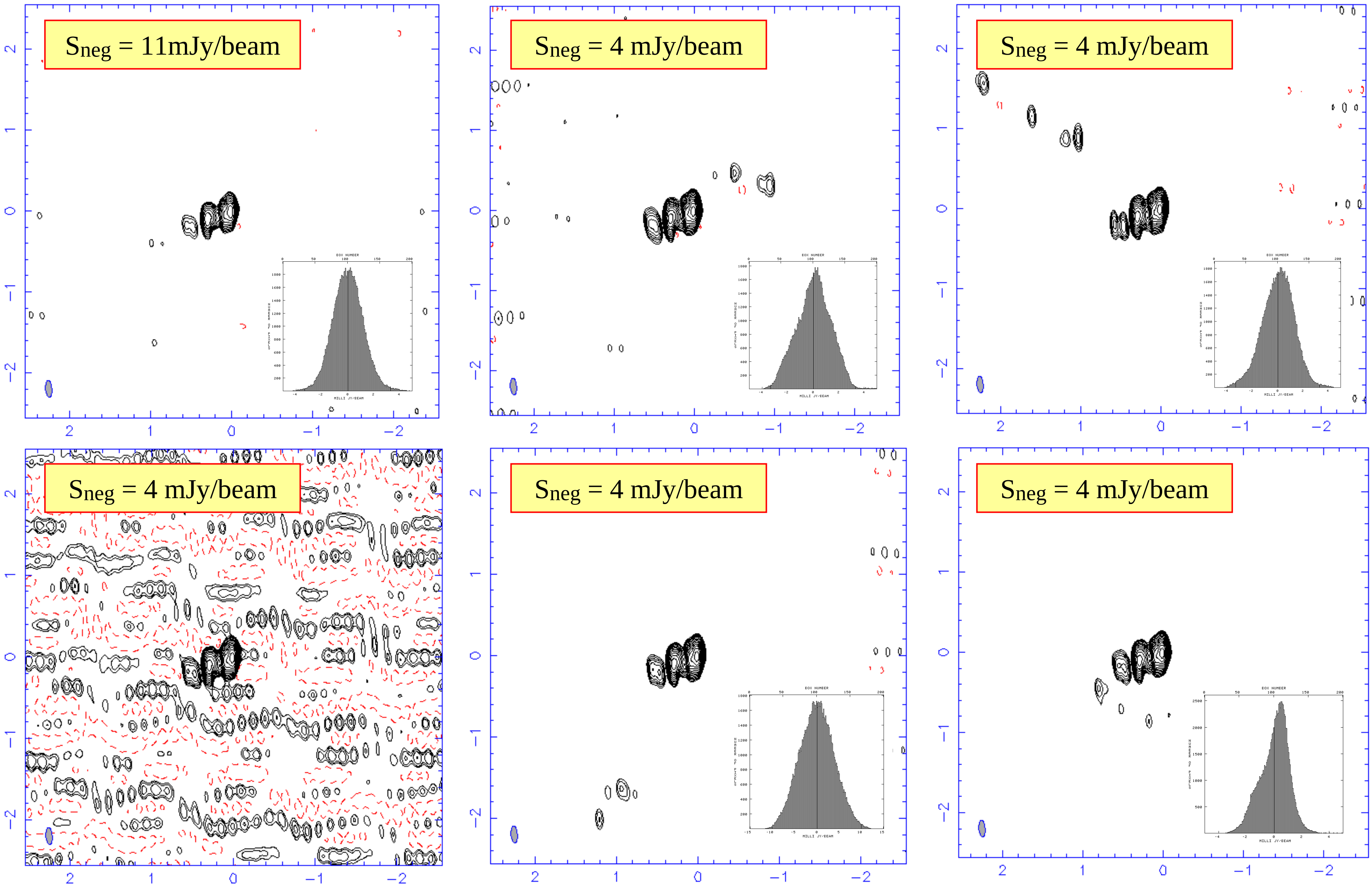


Noise in "B":
non-Gaussian



Excessive use of CLEAN

A. Lobanov



Gaussian noise

Suppose that a residual image (noise image) has an rms σ_r and the maximum absolute flux density $|s_\mathrm{r}|$. For Gaussian noise with a zero mean, the expectation of s_r is

$$|s_\mathrm{r,exp}| = \sigma_\mathrm{r}\left[\sqrt{2}\ln\left(\frac{N_\mathrm{pix}}{\sqrt{2\pi}\sigma_\mathrm{r}}\right)\right]^{1/2},$$

where N_pix is the total number of pixels in the image.

Quality of the residual noise is given by $\zeta_\mathrm{r} = s_\mathrm{r}/s_\mathrm{r,exp}$.

$\zeta_\mathrm{r} \to 1$ for the residual noise approaching Gaussian noise.

$\zeta_\mathrm{r} > 1$ indicates that not all the structure has been adequately recovered.

$\zeta_\mathrm{r} < 1$ implies that the image model has an excessively large number of degrees of freedom.

$\kappa_\sigma = \exp(|\ln\zeta_\mathrm{r}|) - 1$ gives the relative deviation (in units of the Gaussian standard deviation, σ) of the measured noise distribution from the ideal Gaussian noise.

Expected noise in image

- Suppose that
  - -- data are properly sampled and image dimensions are properly set
  - -- natural weighting is chosen
  - -- no strong uv-tapering is applied
  - -- no significant errors are introduced during self-calibration and deconvolution
- The resulting noise in the image should approach

$$\Delta I_{real} = \left[\Delta I_{\rm m}^2 + \sigma_{\rm g}^2 + \langle\sigma_{\rm avg}\rangle^2 + \langle\sigma_{\Delta\nu}\rangle^2\right]^{1/2}$$

If the measured noise $\Delta I_{\rm meas}$ in the image:

$\Delta I_{\rm meas} > \Delta I_{\rm real}$ — there are still perhaps some strides to be made

$\Delta I_{\rm m} \le \Delta I_{\rm meas} \le \Delta I_{\rm real}$ — should be good enough to call it a day

$\Delta I_{\rm meas} < \Delta I_{\rm m}$ — hope that not too many things have gone out of hands

Practical suggestions

Imaging doctrine in thirteen (sixteen) words:

The less you have to do to your data, the better you (and the data!) are.

In other words, one should try to minimize the number of operations required to achieve the expected noise level in an image.

The shortest path is usually the right one.

If the reduction process does not converge, some problem has most likely occurred early on.

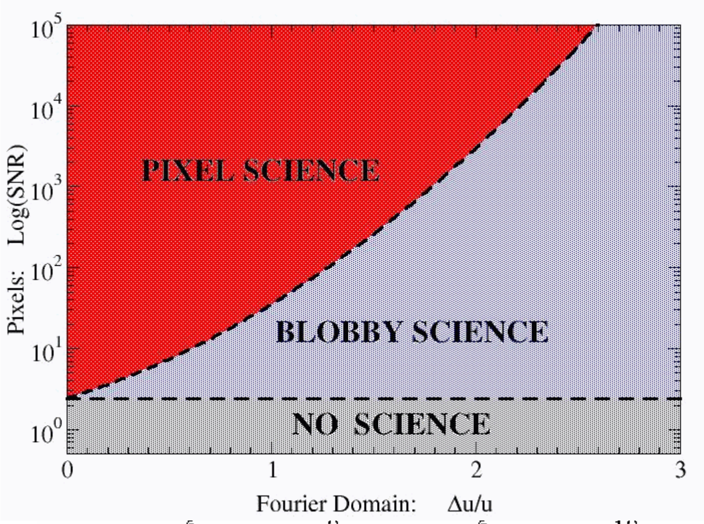
Information in images

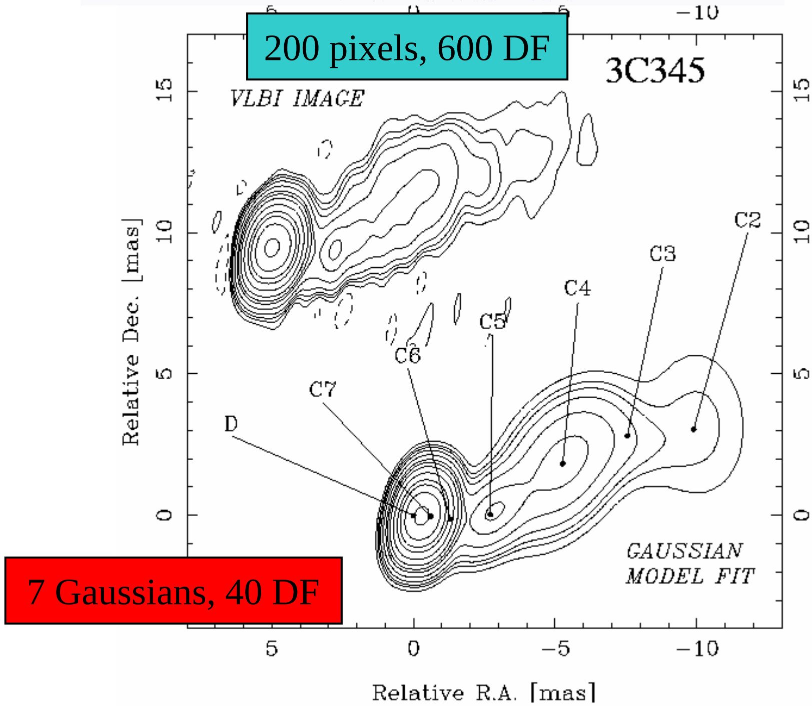
- Ideally, information can be extracted from each pixel... but:
  -- pixel fidelity?
  -- physical significance?

- Fitting by a set of a priori defined shapes (gaussians, discs, etc.) is a common remedy:

  -- it provides a viable description of the structure observed

  -- but reduces the number of degrees of freedom of the description





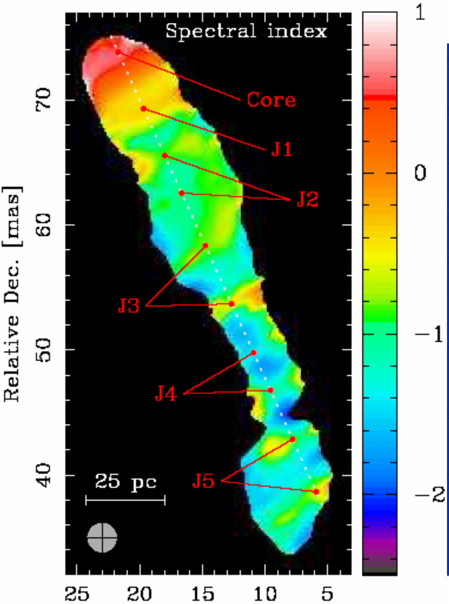
Information in images

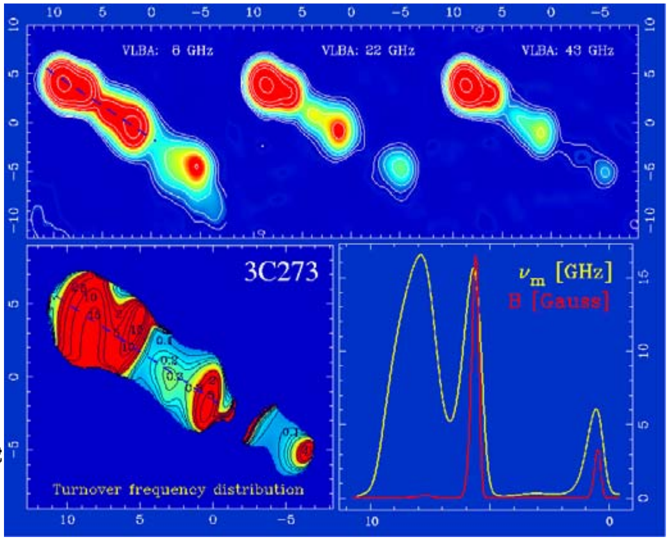
- Examples of pixel-based extraction of information:

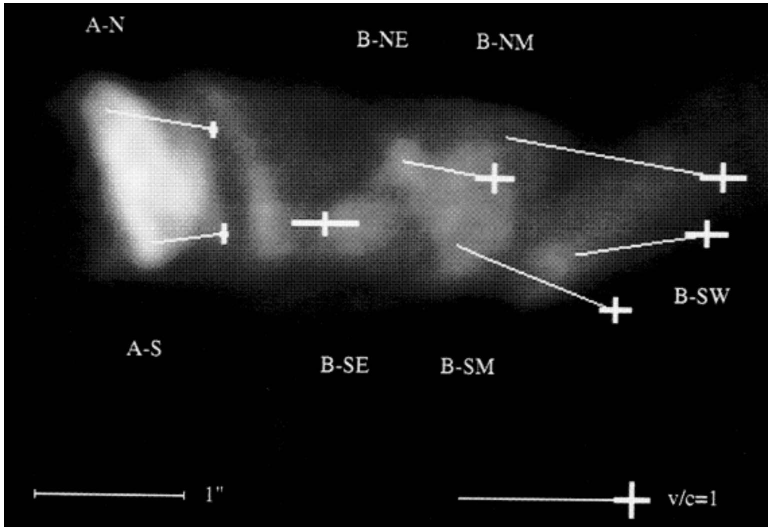
  -- spectral index imaging

  -- turnover frequency imaging

  -- 2D correlations between pixels in different images







Fitting a priori defined patterns

- Fitting measurements with a priori defined patterns is a general approach to inverse problems:
  -- design a model with a number of adjustable parameters
  -- use the model to predict measurements
  -- choose a figure-of-merit function to quantify deviation between model predictions and measurements
  -- adjust the parameters to minimize the merit function

- Goals:
  -- best-fit values for the parameters and their uncertainties
  -- a measure of the goodness-of-fit of the optimized model

- Types of fitting:
  -- fitting in the image plane (IMFIT)
  -- fitting in the Fourier domain – model fitting (UVFIT, MODELFIT, DIFMAP)

Model fitting

Optimization of the fit is achieved by maximizing the likelihood of the model.
The model:

$$V(u,v) = F(u,v;a_1,...,a_M) + \text{noise}$$

The likelihood of the model (assuming that the noise is gaussian):

$$L \propto \prod_{i=1}^{N} \left\{ \exp\left[-\frac{1}{2} \left(\frac{V_i - F(u_i, v_i; a_1, ..., a_M)}{\sigma_i} \right)^2 \right] \right\}$$

Maximizing the likelihood is equivalent to minimizing χ^2:

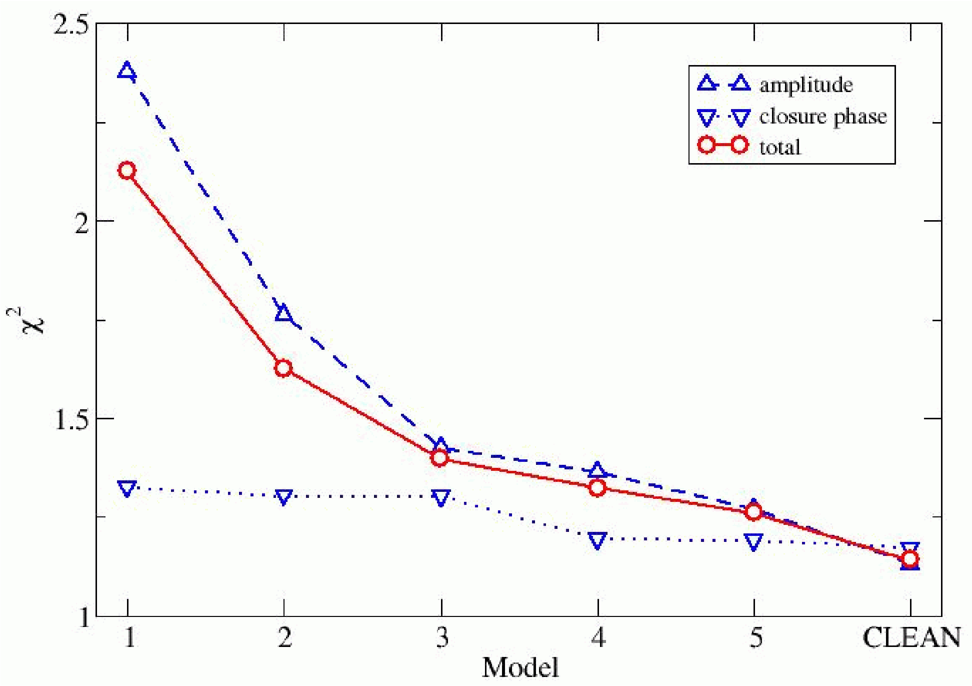
$$\chi^2 = \sum_{i=1}^{N} \left(\frac{V_i - F(u_i, v_i; a_1, ..., a_M)}{\sigma_i} \right)^2$$

Least-squares algorithms are applied for the minimization.

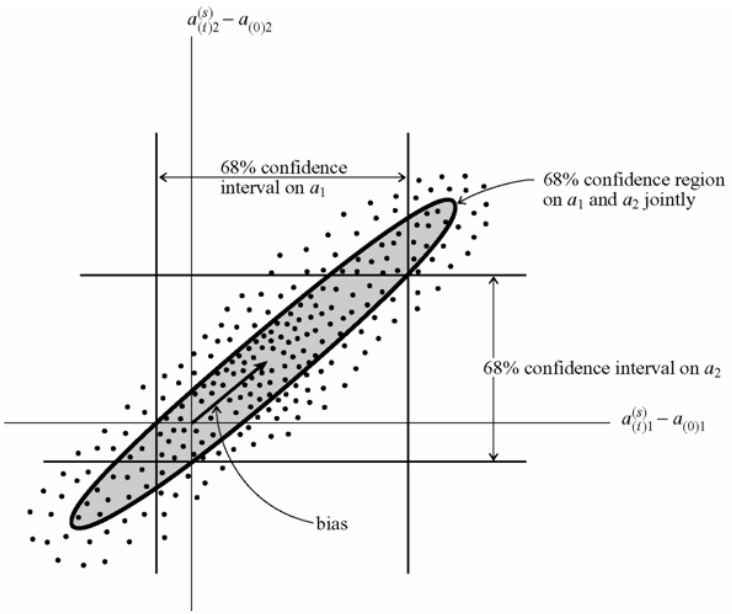
Problems with least squares

- Chance of finding a local minimum instead of the global one
- Slow convergence along the axes representing poorly-constrained parameters
- Choosing the right number of parameters

-- a tedious task; requires proper statistical modelling
-- *F*-test can be used, but this depends on estimating accurately the DOF number of the dataset (number of independent samples)



Error estimation



- Errors are determined by a boundary of constant χ^2 (taken at a desired confidence level) in the multidimensional space of all parameters fit

  -- approximate method: Fisher matrix

$$V_{ij} = \mathrm{cov}[a_i, a_j]; \qquad V_{ij}^{-1} = -\frac{\partial^2 \ln L}{\partial a_i \, \partial a_j}, \quad \text{which gives} \quad \sigma^2(a_i) = V_{ii}^{-1}$$

  -- Fisher matrix often becomes degenerate
  -- Monte Carlo methods can be used generallly
  -- correlation between model fit parameters (e.g. flux density and size) may cause problems

Error estimates

- An analytical (first order) approximation can be given to relate uncertainties of the fit parameters to SNR of detection of a given model fit component

Component: $S_{\rm tot}$ – total flux density; $S_{\rm peak}$ – peak flux density; $\sigma_{\rm rms}$ – post-fit rms; d – size; r - radial distance; θ – position angle. **Uncertainties:**

$$\sigma_{\rm peak} = \sigma_{\rm rms}\left(1 + \frac{S}{\sigma_{\rm rms}}\right)^{1/2}$$

$$\sigma_{\rm tot} = \sigma_{\rm peak}\left(1 + \frac{S_{\rm tot}^2}{S_{\rm peak}^2}\right)^{1/2}$$

$$\sigma_d = d\,\frac{\sigma_{\rm peak}}{S_{\rm peak}}$$

$$\sigma_r = \frac{1}{2}\sigma_d$$

$$\sigma_\theta = {\rm atan}\left(\frac{\sigma_{\rm r}}{r}\right)$$

For large r, bandwidth smearing should be taken into account: $\sigma_{\rm tot}$ are $\sigma_{\rm peak}$ are multiplied by $1/R_{\Delta\nu}$; σ_d and σ_r are multiplied by $B_{\rm D}/{\rm HPBW}$.

Noise and resolution

- Noise and SNR play a fundamental role.

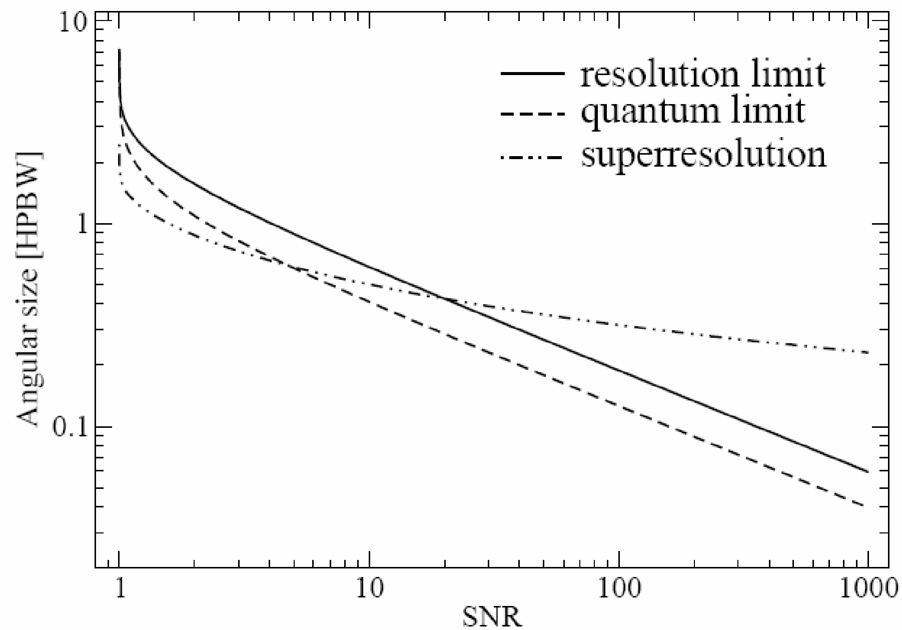
Minimum resolvable size
$$d_{\rm min} = \frac{2^{1+\beta/2}}{\pi}\left[\pi\, a\, b\, \ln 2\, \ln\left(\frac{SNR}{SNR-1}\right)\right]^{1/2}$$

Maximum detectable size
$$d_{\rm max} = \frac{2^{1+\beta/2}}{\pi}\left[\pi\, a\, b\, \ln 2\, SNR^{2/(1+\beta)}\, \ln SNR\right]^{1/2}$$

a, b – axes of resolving beam, β – weighting function ($\beta = 0$ – natural weighting; $\beta = 2$ – uniform weighting)

These limits approach the fundamental quantum resolution limit.

The quantity $1/d_{\rm min}$ evaluated over all pixels gives an equivalent of the total information content in an image.



Summary

- Noise and information in interferometric images are affected by a number of „irreducible“ and „controllable“ factors (bandwidth and time-average smearing, gridding, convolution, tapering and weghting, deconvolution and self-calibration).
- The net effect of these factors must be evaluated and understood, in order to be able to produce and analyze high-quality images.
- Information can be extracted from every pixel of an image, but this sets extremely high requirements on image fidelity.
- Analytical and numerical methods are available to quantify the information in images by fitting *a priori* defined patterns of brightness distribution.

Further reading

- Most of the material covered in this lecture can be found in the latest version of the „NRAO Summerschool Book“:

1. Taylor, G. B., Carilli, C. L., Perley, R. A. (eds.), Synthesis imaging in radio astronomy II, ASP Conf. Ser., v. 180 (ASP: San Francisco) (1999)
2. *ibid.*, Chapter 13: “High dynamic range imaging” *(R.A. Perley)*
3. *ibid.*, Chapter 14: “Image analysis” *(E.B. Fomalont)*
4. *ibid.*, Chapter 15: “Error recognition” *(R.D. Ekers)*
5. *ibid.*, Chapter 16: “Non-imaging data analysis” *(T.J. Pearson)*
6. *ibid.*, Chapter 18: “Bandwidth and time-average smearing” *(A.H. Bridle & F.R. Schwab)*
7. *ibid.*, Chapter 33: “Noise and interferometry” *(V. Radhakrishnan)*

Check individual chapters for further references