Non-Imaging Data Analysis

Greg Taylor

University of New Mexico

Spring 2006

Tenth Summer Synthesis Imaging Workshop
University of New Mexico, June 13-20, 2006

Outline

- Introduction
- Inspecting visibility data
- Model fitting
- Some applications
  - Superluminal motion
  - Gamma-ray bursts
  - Gravitational lenses
  - The Sunyaev-Zeldovich effect

3

Introduction

- Reasons for analyzing visibility data
  - Insufficient (*u,v*)-plane coverage to make an image
  - Inadequate calibration
  - Quantitative analysis
  - Direct comparison of two data sets
  - Error estimation
    - Usually, visibility measurements are independent gaussian variates
    - Systematic errors are usually localized in the (*u,v*) plane
- Statistical estimation of source parameters

G. Taylor, Summer Synthesis Imaging Workshop 2006

4

Inspecting Visibility Data

- Fourier imaging

$$V(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{A}(l,m) I(l,m) \exp[-2\pi i(ul+vm)]\, dl\, dm$$

- Problems with direct inversion
  - Sampling
    - Poor (*u,v*) coverage
  - Missing data
    - e.g., no phases (speckle imaging)
  - Calibration
    - Closure quantities are independent of calibration
  - Non-Fourier imaging
    - e.g., wide-field imaging; time-variable sources (SS433)
  - Noise
    - Noise is uncorrelated in the (*u,v*) plane but correlated in the image

G. Taylor, Summer Synthesis Imaging Workshop 2006

5

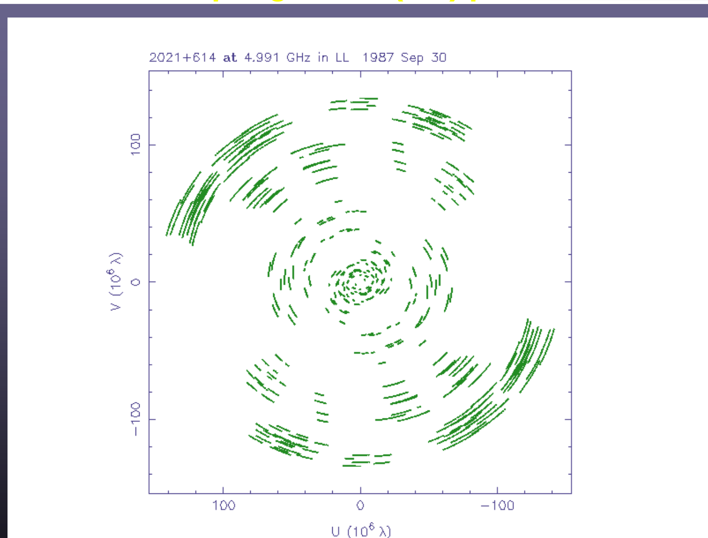
Inspecting Visibility Data

- Useful displays
  - Sampling of the (*u*,*v*) plane
  - Amplitude and phase *vs*. radius in the (*u*,*v*) plane
  - Amplitude and phase *vs*. time on each baseline
  - Amplitude variation across the (*u*,*v*) plane
  - Projection onto a particular orientation in the (*u*,*v*) plane
- Example: 2021+614
  - GHz-peaked spectrum radio galaxy at *z*=0.23
  - A VLBI dataset with 11 antennas from 1987
  - VLBA only in 2000

G. Taylor, Summer Synthesis Imaging Workshop 2006

6

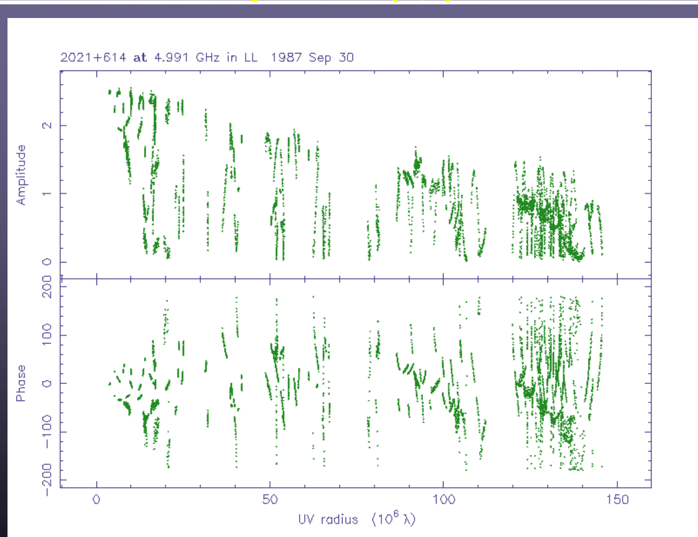
Sampling of the (u,v) plane



G. Taylor, Summer Synthesis Imaging Workshop 2006

7

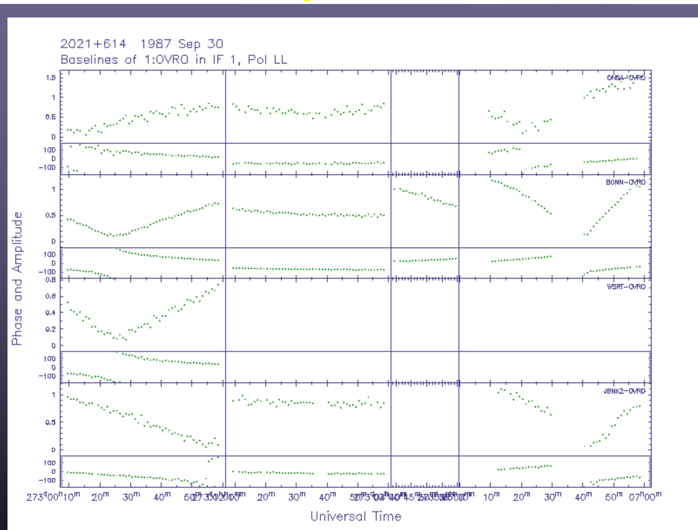
Visibility versus (u,v) radius



G. Taylor, Summer Synthesis Imaging Workshop 2006

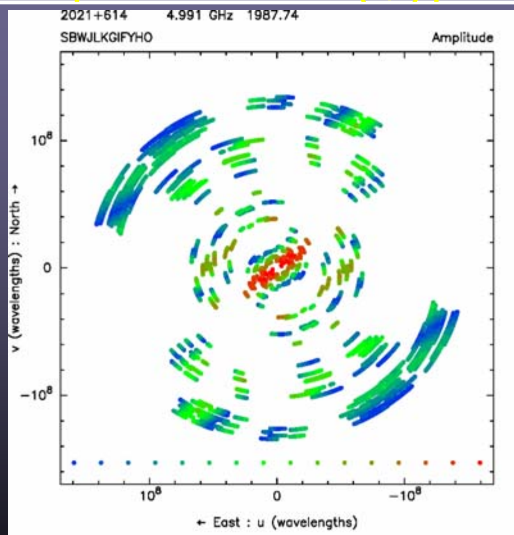
8

Visibility versus time

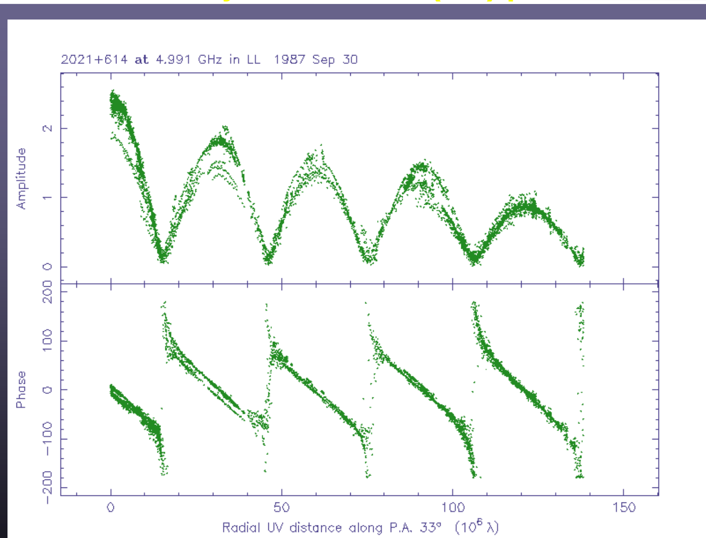


G. Taylor, Summer Synthesis Imaging Workshop 2006

Amplitude across the (u,v) plane



Projection in the (u,v) plane



Properties of the Fourier transform

- See, e.g., R. Bracewell, *The Fourier Transform and its Applications* (1965).

- Fourier Transform theorems
  - Linearity
    - Visibilities of components add (complex)
  - Convolution
  - Shift
    - Shifting the source creates a phase gradient across the (*u*,*v*) plane
  - Similarity
    - Larger sources have more compact transforms

Fourier Transform theorems

$$F(u,v) = \mathrm{FT}\{f(x,y)\}$$

i.e.,

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\exp[2\pi i(ux+vy)]\,dx\,dy$$

Linearity

$$\mathrm{FT}\{f(x,y)+g(x,y)\} = F(u,v)+G(u,v)$$

Convolution

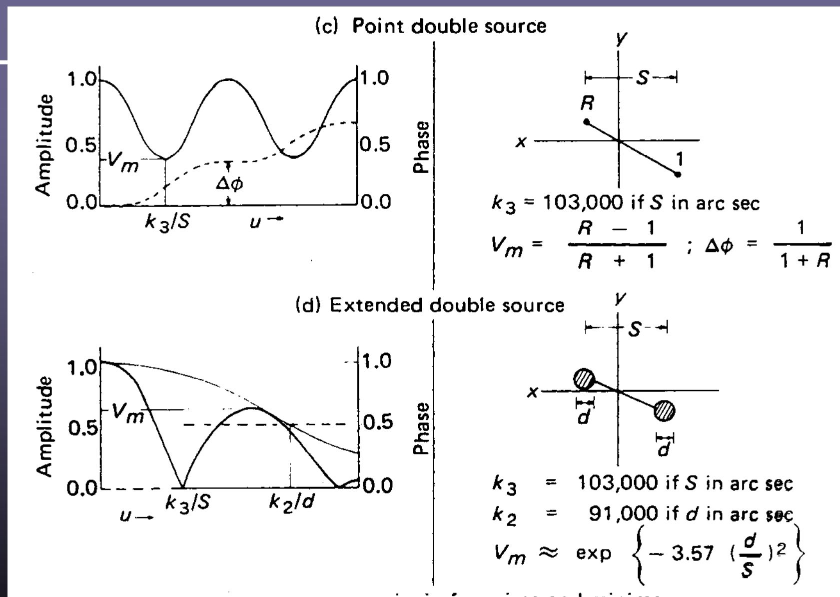
$$\mathrm{FT}\{f(x,y)\star g(x,y)\} = F(u,v)\cdot G(u,v)$$

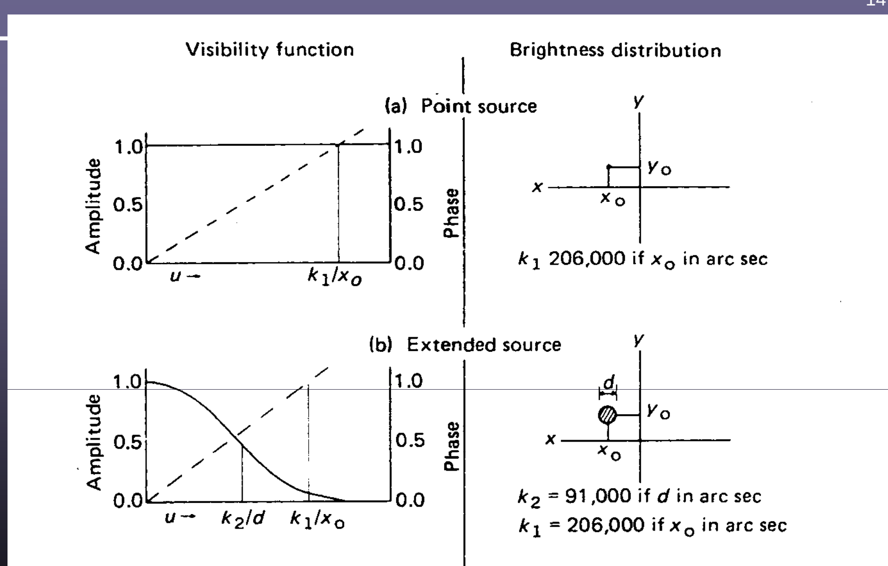
Shift

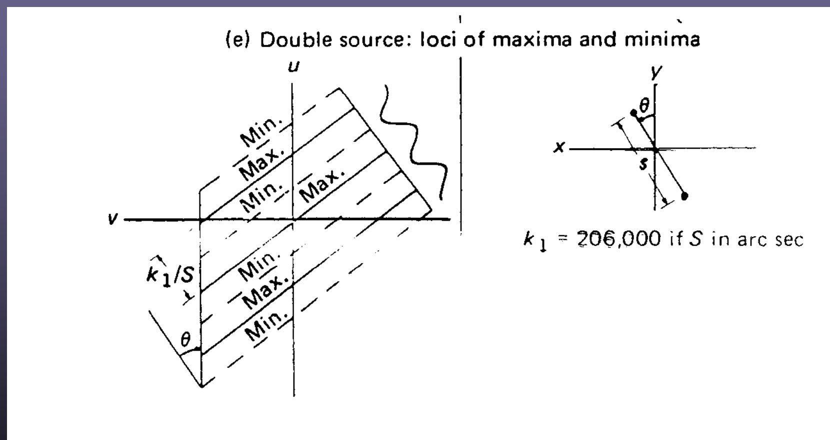
$$\mathrm{FT}\{f(x-x_i,y-y_i)\} = F(u,v)\,\exp[2\pi i(ux_i+vy_i)]$$

Similarity

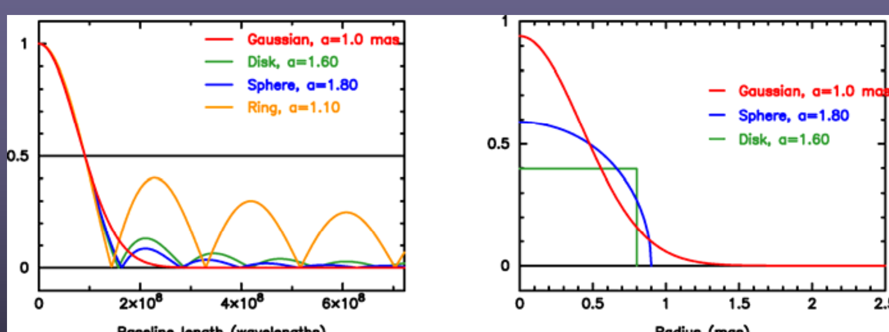
$$\mathrm{FT}\{f(ax,by)\} = \frac{1}{|ab|}F\left(\frac{u}{a},\frac{v}{b}\right)$$


(c) Point double source
Amplitude
Phase
1.0
0.5
0.0
V_m
Δφ
k_3/S
u →
y
x
S
R
1
k_3 = 103,000 if S in arc sec
V_m = (R − 1)/(R + 1) ; Δφ = 1/(1 + R)
(d) Extended double source
d
k_2/d
k_3 = 103,000 if S in arc sec
k_2 = 91,000 if d in arc sec
V_m ≈ exp {− 3.57 (d/S)^2}


Visibility function
Brightness distribution
(a) Point source
Amplitude
Phase
1.0
0.5
0.0
u →
k_1/x_o
y
x
y_o
x_o
k_1 206,000 if x_o in arc sec
(b) Extended source
d
k_2/d
k_2 = 91,000 if d in arc sec
k_1 = 206,000 if x_o in arc sec



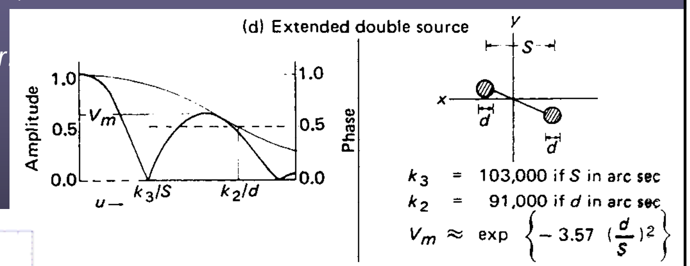
Simple models

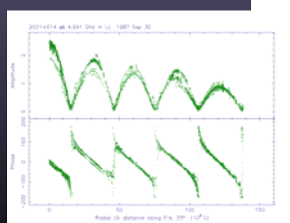


- Visibility at short baselines contains little
- information about the profile of the source.

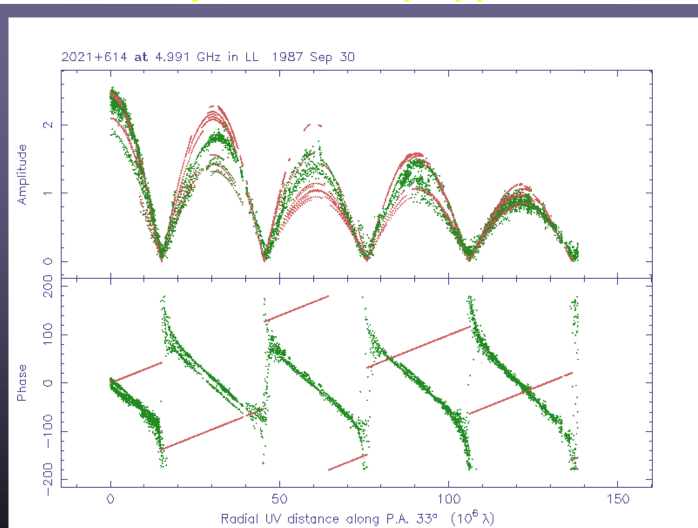
Trial model

- By inspection, we can derive a simple model:
- Two equal components, each 1.25 Jy, separated by about 6.8 milliarcsec in p.a. 33º, each about 0.8 milliarcsec in diameter (Gaussian FWHM)
- *To be refined later*





Projection in the (u,v) plane



Closure Phase and Amplitude: closure quantities

- Antenna-based gain errors

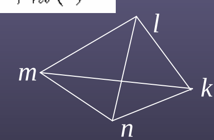
$$V_{kl} \equiv |V_{kl}| \exp(i\phi_{kl}) = g_k g_l V_{kl}^{\mathrm{true}} \exp(i\phi_k) \exp(-i\phi_l)$$

- Closure phase **(bispectrum phase)**

$$\Psi_{lmn}(t) = \phi_{lm}(t) + \phi_{mn}(t) + \phi_{nl}(t)$$

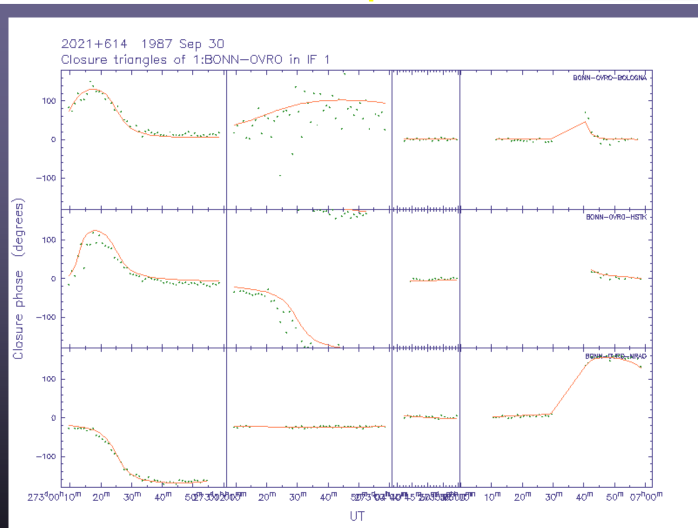
- Closure amplitude

$$\frac{|V_{kl}| \cdot |V_{mn}|}{|V_{km}| \cdot |V_{ln}|}$$



  - Closure phase and closure amplitude are unaffected by antenna gain errors
  - They are conserved during self-calibration
  - Contain (*N*–2)/*N* of phase, (*N*–3)/(*N*–1) of amplitude info
  - Many non-independent quantities
  - They do not have gaussian errors
  - No absolute position or flux info

Closure phase



Model fitting

- Imaging as an Inverse Problem
  - In synthesis imaging, we can solve the **forward problem**: given a sky brightness distribution, and knowing the characteristics of the instrument, we can predict the measurements (visibilities), within the limitations imposed by the noise.
  - The **inverse problem** is much harder, given limited data and noise: the solution is rarely unique.
  - A general approach to inverse problems is **model fitting**. See, e.g., Press et al., *Numerical Recipes*.
    1. Design a model defined by a number of adjustable parameters.
    2. Solve the forward problem to predict the measurements.
    3. Choose **a figure-of-merit** function, e.g., rms deviation between model predictions and measurements.
    4. Adjust the parameters to **minimize the merit function**.
  - Goals:
    1. Best-fit values for the parameters.
    2. A measure of the goodness-of-fit of the optimized model.
    3. Estimates of the uncertainty of the best-fit parameters.

Model fitting

- Maximum Likelihood and Least Squares
  - The model: $V(u,v) = F(u,v;a_1,\ldots,a_M) + \text{noise}$
  - The likelihood of the model (if noise is gaussian):

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

  - Maximizing the likelihood is equivalent to minimizing chi-square (for gaussian errors):

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

  - Follows chi-square distribution with *N* – *M* degrees of freedom. Reduced chi-square has expected value 1.

23

Uses of model fitting

- Model fitting is most useful when the brightness distribution is simple.
  - Checking amplitude calibration
  - Starting point for self-calibration
  - Estimating parameters of the model (with error estimates)
  - In conjunction with CLEAN or MEM
  - In astrometry and geodesy
- Programs
  - AIPS UVFIT
  - Difmap (Martin Shepherd)

G. Taylor, Summer Synthesis Imaging Workshop 2006

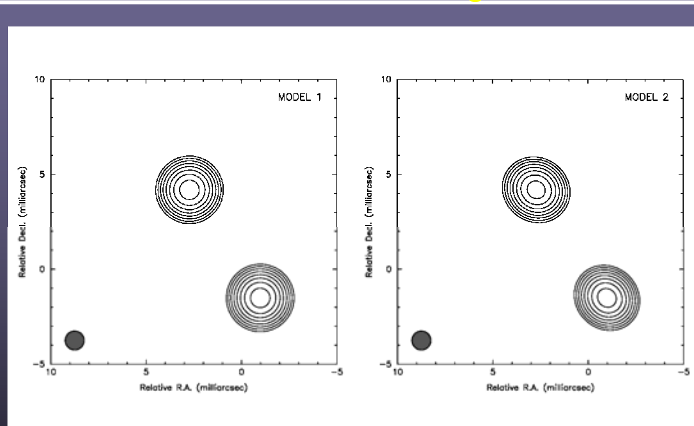
24

Parameters

- Example
  - Component position: (*x,y*) or polar coordinates
  - Flux density
  - Angular size (e.g., FWHM)
  - Axial ratio and orientation (position angle)
    - For a non-circular component
  - 6 parameters per component, plus a "shape"
  - This is a conventional choice: other choices of parameters may be better!
  - (Wavelets; shapelets* [Hermite functions])
    - * Chang & Refregier 2002, ApJ, 570, 447

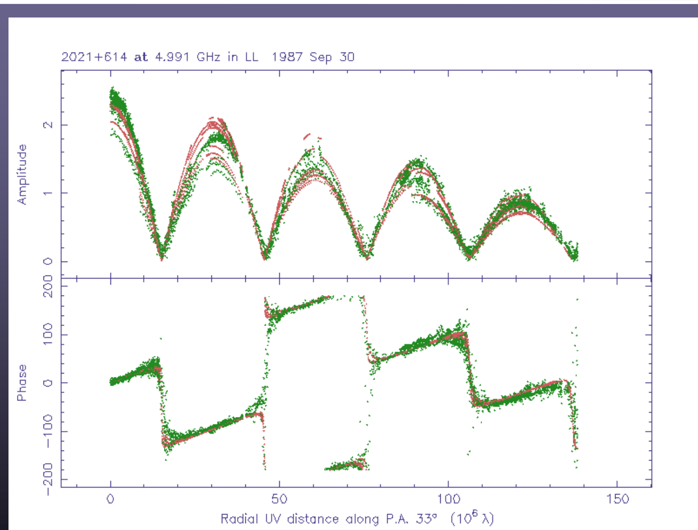
G. Taylor, Summer Synthesis Imaging Workshop 2006

Practical model fitting: 2021

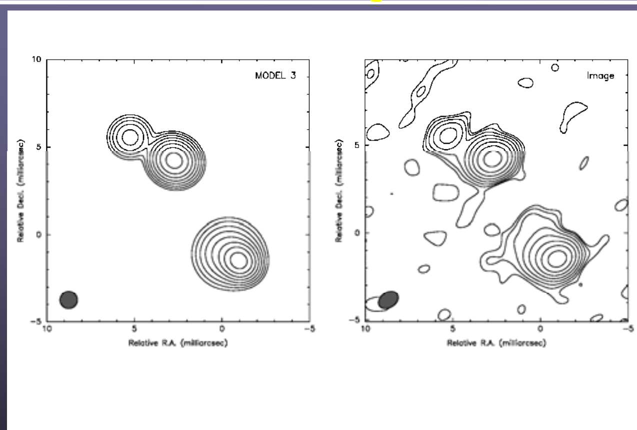


! Flux (Jy)	Radius (mas)	Theta (deg)	Major (mas)	Axial ratio	Phi (deg)	T
1.15566	4.99484	32.9118	0.867594	0.803463	54.4823	1
1.16520	1.79539	-147.037	0.825078	0.742822	45.2283	1

2021: model 2

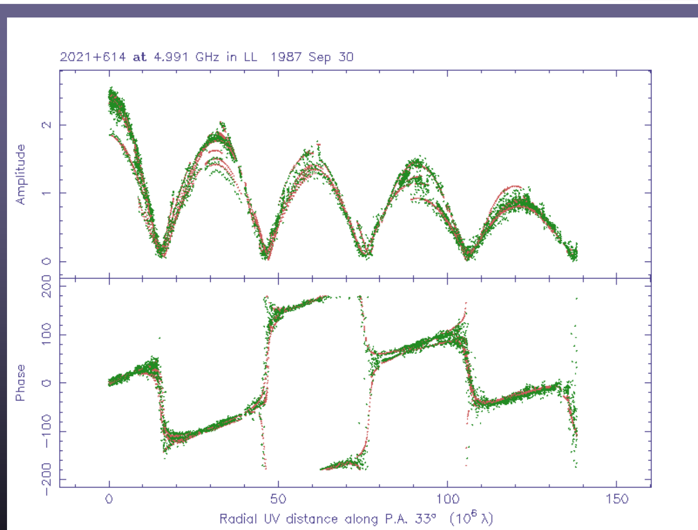


Model fitting 2021



! Flux (Jy)	Radius (mas)	Theta (deg)	Major (mas)	Axial ratio	Phi (deg)	T
1.10808	5.01177	32.9772	0.871643	0.790796	60.4327	1
0.823118	1.80865	-146.615	0.589278	0.585766	53.1916	1
0.131209	7.62679	43.3576	0.741253	0.933106	-82.4635	1
0.419373	1.18399	-160.136	1.62101	0.951732	84.9951	1

2021: model 3



29

Limitations of least squares

- Assumptions that may be violated
  - The model is a good representation of the data
    - Check the fit
  - The errors are Gaussian
    - True for real and imaginary parts of visibility
    - Not true for amplitudes and phases (except at high SNR)
  - The variance of the errors is known
    - Estimate from T_{sys}, rms, etc.
  - There are no systematic errors
    - Calibration errors, baseline offsets, etc. must be removed before or during fitting
  - The errors are uncorrelated
    - Not true for closure quantities
    - Can be handled with full covariance matrix

G. Taylor, Summer Synthesis Imaging Workshop 2006

30

Least-squares algorithms

- **At the minimum, the derivatives**
- **of chi-square with respect to the**
- **parameters are zero**

$$\nabla \chi^2 = \frac{\partial \chi^2}{\partial a_k} = 0$$

- **Linear case: matrix inversion.**
- Exhaustive search**: prohibitive with**
- **many parameters (~ 10^M)**
- Grid search**: adjust each parameter by a**
- **small increment and step down hill in search for minimum.**
- Gradient search**: follow downward gradient toward minimum, using numerical or analytic derivatives. Adjust step size according to second derivative**

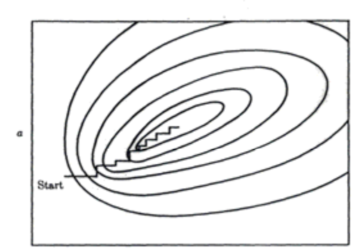


FIGURE 11-3 Tortuous path of grid search in two-parameter space.

$$\nabla^2 \chi^2 = \frac{\partial^2 \chi^2}{\partial a_k \partial a_l}$$

– For details, see *Numerical Recipes*.

G. Taylor, Summer Synthesis Imaging Workshop 2006

31

Problems with least squares

- Global versus local minimum
- Slow convergence: poorly constrained model
  - Do not allow poorly-constrained parameters to vary
- Constraints and prior information
  - Boundaries in parameter space
  - Transformation of variables
- Choosing the right number of parameters: does adding a parameter significantly improve the fit?
  - Likelihood ratio or *F* test: use caution
    - Protassov et al. 2002, ApJ, 571, 545
  - Monte Carlo methods

G. Taylor, Summer Synthesis Imaging Workshop 2006

32

Error estimation

- Find a region of the *M*-dimensional parameter space around the best fit point in which there is, say, a 68% or 95% chance that the true parameter values lie.
- Constant chi-square boundary: select the region in which

$$\chi^2 < \chi^2_{\rm min} + \Delta\chi^2$$

- The appropriate contour depends on the required confidence level and the number of parameters estimated.
- Monte Carlo methods (simulated or mock data): relatively easy with fast computers
- Some parameters are strongly correlated, e.g., flux density and size of a gaussian component with limited (*u*,*v*) coverage.
- Confidence intervals for a single parameter must take into account variations in the other parameters ("marginalization").

G. Taylor, Summer Synthesis Imaging Workshop 2006

Mapping the likelihood

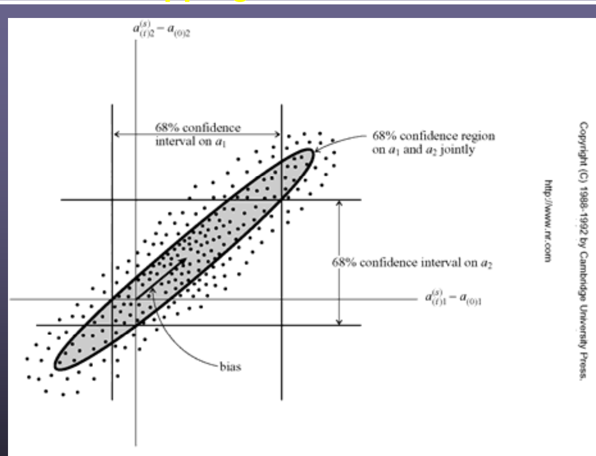


Figure 15.6.3. Confidence intervals in 1 and 2 dimensions. The same fraction of measured points (here 68%) lies (i) between the two vertical lines, (ii) between the two horizontal lines, (iii) within the ellipse.

- Press et al., *Numerical Recipes*

Applications: Superluminal motion

- Problem: to detect changes in component positions between observations and measure their speeds
  - Direct comparison of images is *bad*: different (*u,v*) coverage, uncertain calibration, insufficient resolution
  - Visibility analysis is a *good* method of detecting and measuring changes in a source: allows “controlled super-resolution”
  - Calibration uncertainty can be avoided by looking at the closure quantities: have they changed?
  - Problem of differing (*u,v*) coverage: compare the same (*u,v*) points whenever possible
  - Model fitting as an interpolation method

Superluminal motion

- Example 1: Discovery of superluminal motion in 3C279 (Whitney et al., Science, 1971)

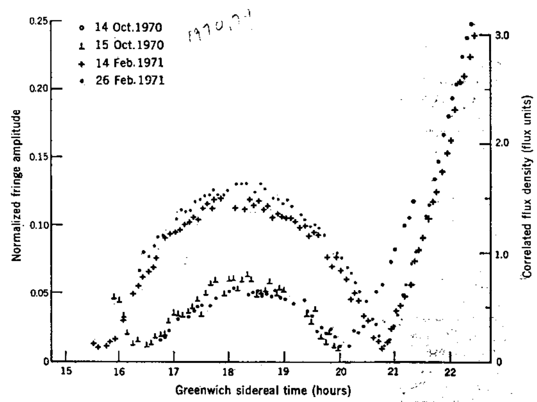


Fig. 1. Fringe-amplitude data from observations of 3C 279 with the Goldstone-Haystack interferometer. Each point is based on 110 seconds of integration.

225

Superluminal motion

- 1.55 ± 0.3 milliarcsec in 4 months: $v/c = 10 \pm 3$

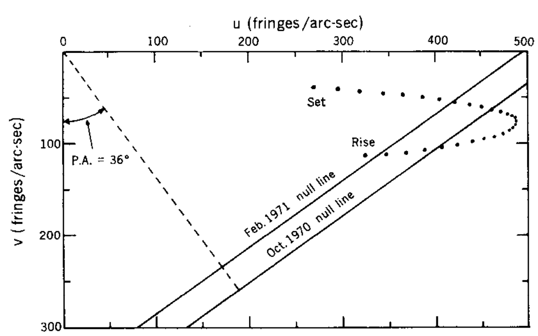
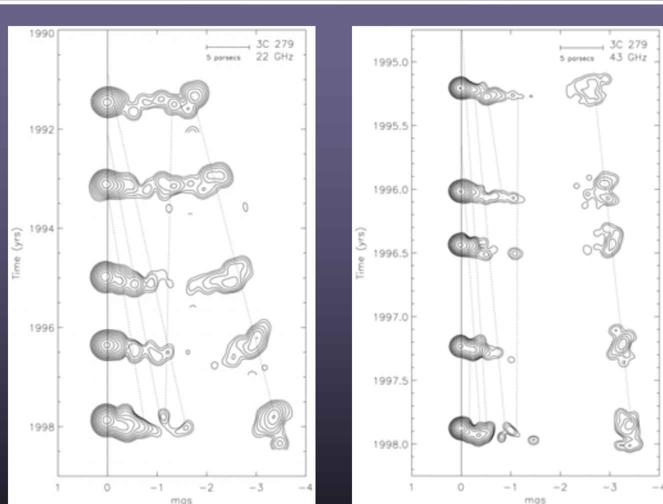


Fig. 4. The *u-v* plane (*1*) representation of the Goldstone-Haystack observations of 3C 279. The dotted curve shows the interferometer resolution at 15-minute intervals from 15 hours 30 minutes to 22 hours 30 minutes Greenwich sidereal time. The solid lines connect the times at which nulls were observed in October 1970 and in February 1971. The distances from the origin to the solid lines are inversely proportional to the separations of the components of the putative double source at the two times of observation. [The position angle (*P.A.*) was assumed to remain constant.]

227

3C279 with the VLBA



- Wehrle et al. 2001, ApJS, 133, 297

Demo

- Switch to Difmap and demo model-fiting on VLBA data

Applications: Expanding sources

- Example 2: changes in the radio galaxy 2021+614 between 1987 and 2000
  - We find a change of 200 microarcsec so v/c = 0.18
  - By careful combination of model-fitting and self-calibration, Conway et al. (1994) determined that the separation had changed by 69 ± 10 microarcsec between 1982 and 1987, for *v/c* = 0.19

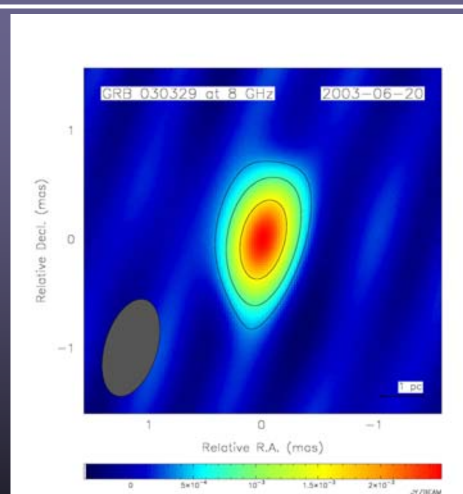
GRB030329

June 20, 2003

t+83 days

Peak ~ 3 mJy
Size 0.172 +/- 0.043 mas
0.5 +/- 0.1 pc
average velocity = 3c

Taylor et al. 2004



VLBA+Y27+GBT+EB+AR+WB = 0.11 km^2

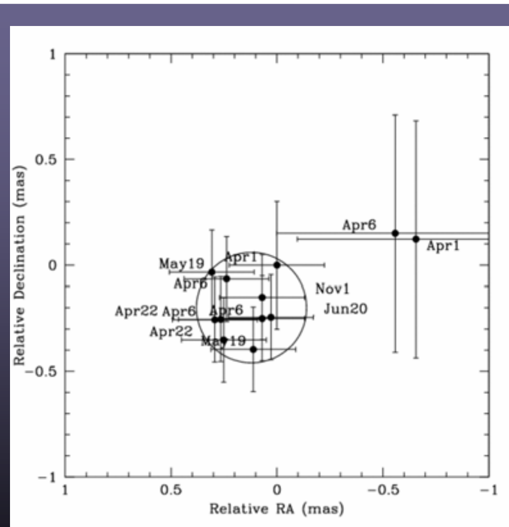
41

GRB 030329

Proper motion limits

RA = -0.02 +/- 0.80 mas/yr
DEC = -0.44 +/- 0.63 mas/yr

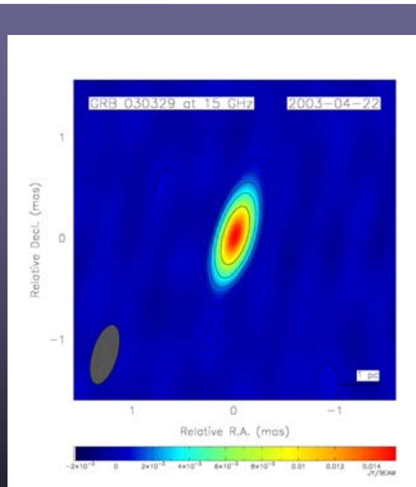
motion < 0.28 mas in 80 days

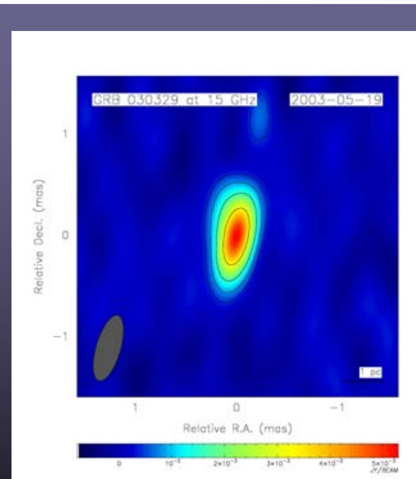


G. Taylor, Summer Synthesis Imaging Workshop 2006

42

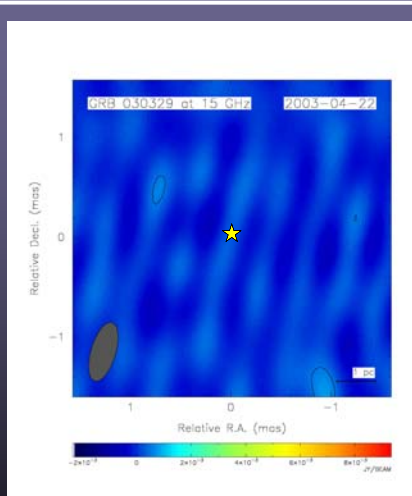
GRB030329

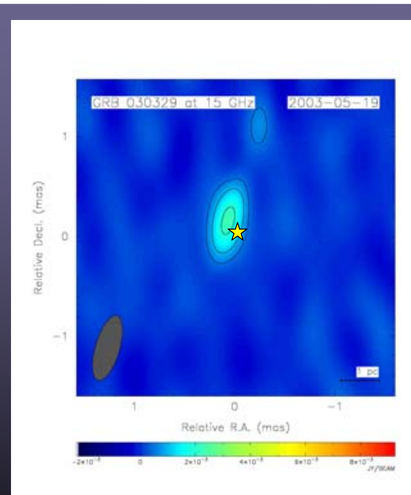




G. Taylor, Summer Synthesis Imaging Workshop 2006

GRB030329 subtracted

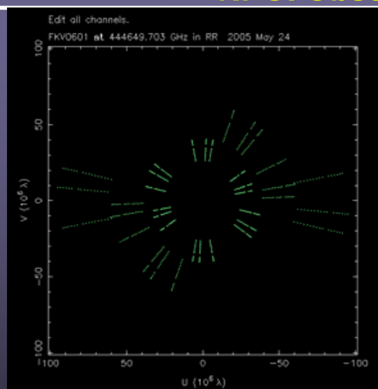


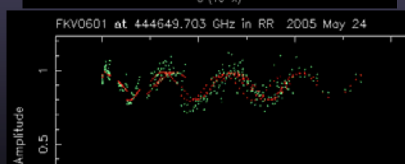


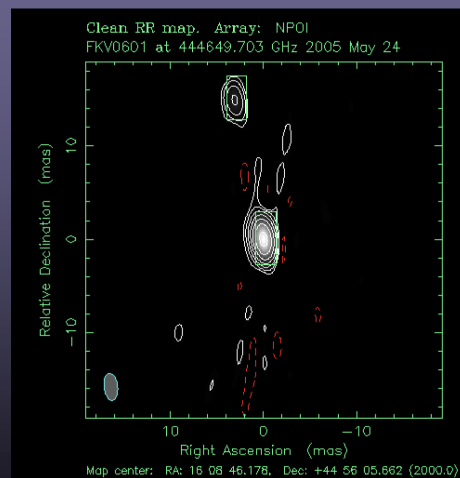
Applications: A Binary Star

- Binary Stars
  - Many stars are in binary systems
  - Orbital parameters can be used to measure stellar masses
  - Astrometry can provide direct distances via parallax and proper motions.
- Application of model fitting
  - Optical interferometry provides sparse visibility coverage
  - Small number of components
  - Need error estimates.
- Example: NPOI observations of Phi Herculis (Zavala et al. 2006)
  - Multiple observations map out the orbit

NPOI Observations of Phi Her



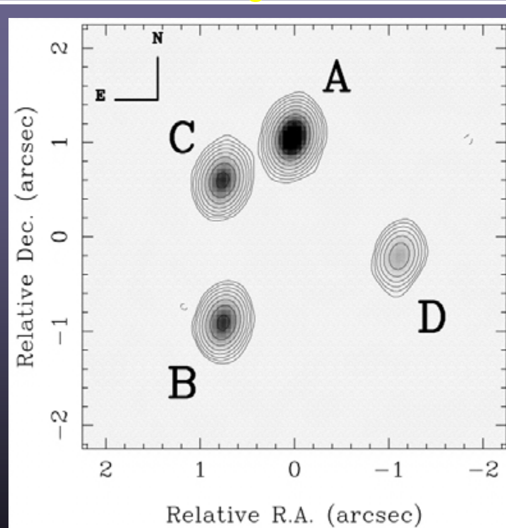




Applications: Gravitational Lenses

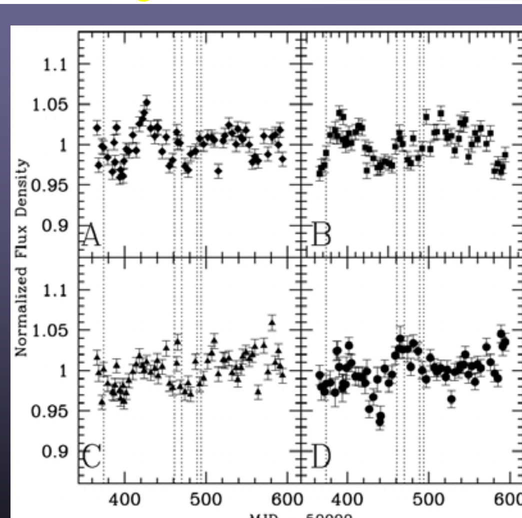
- Gravitational Lenses
  - Single source, multiple images formed by intervening galaxy.
  - Can be used to map mass distribution in lens.
  - Can be used to measure distance of lens and H_0: need redshift of lens and background source, model of mass distribution, and a **time delay**.
- Application of model fitting
  - Lens monitoring to measure flux densities of components as a function of time.
  - Small number of components, usually point sources.
  - Need error estimates.
- Example: VLA monitoring of B1608+656 (Fassnacht et al. 1999, ApJ)
  - VLA configuration changes: different HA on each day
  - Other sources in the field

VLA image of 1608



1608 monitoring results

- B – A = 31 days
- B – C = 36 days
- H_0 = 75 ± 7 km/s/Mpc

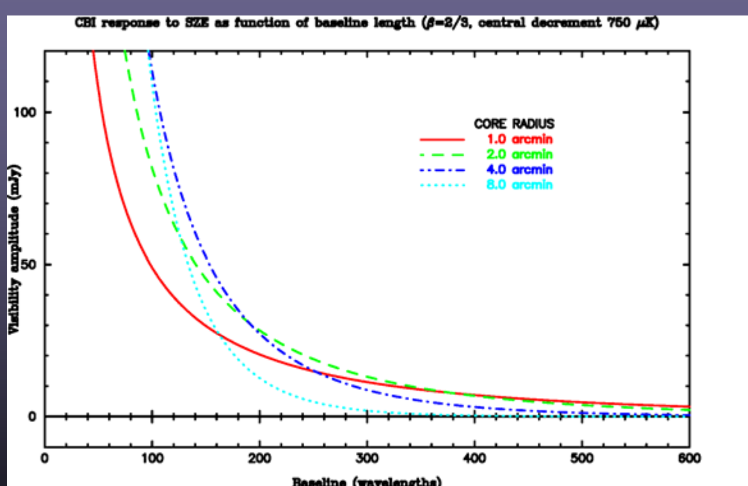


Koopmans et al. 2003

Applications: Sunyaev-Zeldovich effect

- The Sunyaev-Zeldovich effect
  - Photons of the CMB are scattered to higher frequencies by hot electrons in galaxy clusters, causing a negative brightness **decrement**.
  - Decrement is proportional to integral of electron pressure through the cluster, or electron density if cluster is isothermal.
  - Electron density and temperature can be estimated from X-ray observations, so the linear scale of the cluster is determined.
  - This can be used to measure the cluster distance and H_0.
- Application of model fitting
  - The profile of the decrement can be estimated from X-ray observations (beta model).
  - The Fourier transform of this profile increases exponentially as the interferometer baseline decreases.
  - The central decrement in a synthesis image is thus highly dependent on the (*u*,*v*) coverage.
  - Model fitting is the best way to estimate the true central decrement.

SZ profiles



SZ images

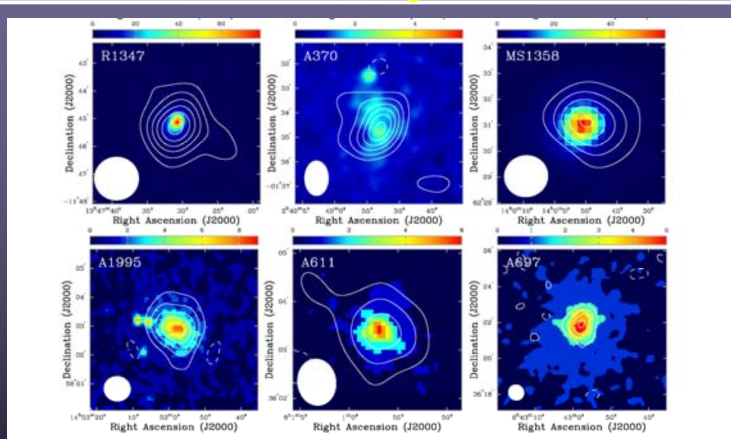


FIG. 2.—SZE (contours) and X-ray (color scale) images of each cluster in our sample. Negative contours are shown as solid lines. The contours are multiples of 2 σ and the FWHM of the synthesized beams are shown in the bottom left corner. The X-ray color scale images are raw counts images smoothed with Gaussians with $\sigma = 15''$ for PSPC data and $\sigma = 5''$ for HRI data. There is a color scale mapping for the counts above each image. The 30 GHz image statistics are summarized in Table 4.

Reese et al. 2002

Summary

- For simple sources observed with high SNR, much can be learned about the source (and observational errors) by inspection of the visibilities.
- Even if the data cannot be calibrated, the **closure quantities** are good observables, but they can be difficult to interpret.
- Quantitative data analysis is best regarded as an exercise in **statistical inference**, for which the maximum likelihood method is a general approach.
- For gaussian errors, the ML method is the **method of least squares**.
- Visibility data (usually) have uncorrelated gaussian errors, so analysis is most straightforward in the (*u,v*) plane.
- Consider visibility analysis when you want a quantitative answer (with error estimates) to a simple question about a source.
- Visibility analysis is inappropriate for large problems (many data points, many parameters, correlated errors); standard imaging methods can be much faster.

Further Reading

- http://www.nrao.edu/whatisra/mechanisms.shtml
- http://www.nrao.edu/whatisra/
- www.nrao.edu

- Synthesis Imaging in Radio Astronomy
- ASP Vol 180, eds Taylor, Carilli & Perley