Synthesis Imaging in Radio Astronomy II
ASP Conference Series, Vol. XXX, 1998
G.B. Taylor, C.L. Carilli, and R. A. Perley (eds.)

16. Non-Imaging Data Analysis

Timothy J. Pearson
California Institute of Technology, Pasadena, CA, USA, 91125

Abstract. In many types of observation it is impossible or inappropriate to make an image from the visibility data. In this lecture I discuss ways of interpreting visibility data directly, with an emphasis on model-fitting techniques.

1. Introduction

Producing an image with the standard Fourier synthesis and deconvolution procedures is not always the best way to analyze data from a synthesis telescope. In many observations a lot may be learned by inspection of the data in the visibility domain or (u, v) plane. This is, after all, the domain in which the measurements are made, and where errors in the data are easiest to recognize. The Fourier transform involved in imaging spreads errors that are localized in the (u, v) plane throughout the image, so different pixels in an image will have correlated errors, while measurements at different points in the (u, v) plane are largely uncorrelated. Thus quantitative analysis, including estimates of errors in the derived quantities, is often best done in the (u, v) plane. There are some types of observation, particularly those with very sparse (u, v) coverage or poor calibration, in which it is difficult or impossible to make an image, and in these cases it is necessary to interpret the observed visibility data directly. I shall also show that some quantitative astronomical questions can be addressed better in the visibility domain than in the image. For example, in comparing two images made at different times it can be difficult to determine whether apparent changes are due to real changes in the source, or just to differences in the (u, v)-plane sampling and the imaging parameters. It is much more straightforward to compare the measured visibilities directly.

In this Lecture I will discuss only continuum data (single frequency channel, single polarization), although the techniques are readily extended to more complex data sets.

2. Visibility Data

As discussed in earlier lectures, the complex visibility $V(u, v)$ measured on a baseline with coordinates (u, v) is related to the Fourier transform of the sky brightness distribution $I(l, m)$:

$$V(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{A}(l,m) I(l,m) \exp[-2\pi i(ul + vm)]\, dl\, dm. \qquad (16\text{-}1)$$

$\mathcal{A}(l.m)$ is the primary beam response, which can often be ignored when the field of interest is small, as in VLBI. In many cases, it is possible to construct an estimate of the sky brightness distribution by direct inversion of this equation. However, in practice there are many difficulties with this approach which make it preferable to try to interpret the visibility data directly rather than first forming an image.

2.1. Sampling

Measurements of $V(u,v)$ are available at only a finite number of points in a small region of the (u,v) plane. The effect of this is that the reconstructed image is a convolution of the sky brightness with a dirty beam. In extreme cases (such as a single baseline) the dirty beam can have such large and extensive side-lobes that a reliable deconvolution is impossible: too many Fourier components of the image are unconstrained. In such cases an image may be liable to misinterpretation or even impossible to interpret.

2.2. Calibration

The visibility data themselves may be uncalibrated or uncalibratable. For example, the phases of the visibilities may be severely corrupted by the atmosphere or local-oscillator instabilities, or the amplitudes may be dominated by unknown or changing antenna gains. Often this problem can be avoided by using self-calibration, but in extreme cases – again, these cases are often those with sparse (u,v)-plane sampling – self-calibration may fail to converge reliably or the result of self-calibration may be dependent on the assumed starting model. In these cases, we can sometimes proceed by examining the *closure phases* and *closure amplitudes*, which are unaffected by the calibration errors. While these quantities are "good observables," they cannot be used to derive an image by simple Fourier transformation, and they can be rather difficult to interpret. Use of the closure quantities is discussed in Section 4.

2.3. Non-Fourier Imaging

In some cases the simple Fourier transform relationship (eqn. 1) is not a sufficiently accurate representation of the imaging process; for example in wide-field mapping with non-coplanar baselines where the assumptions used in deriving eqn. 1 break down.

2.4. Noise

The observed visibilities are subject to additive noise from the sky, receivers, ground pick-up, etc. To a very good approximation, the noise is gaussian with equal variance in the real and imaginary parts of the visibility. The Fourier inversion of eqn. 1 has the desirable property that the noise in the dirty image is also gaussian, with known covariance. However, if the image is deconvolved with CLEAN, MEM or similar non-linear techniques, the noise properties of the resulting image will be poorly understood and it may be difficult to estimate the uncertainty in a measurement (e.g., of a component flux density) from the image. For quantitative analysis, it is often better to work directly with the visibilities.

3. Inspecting Visibility Data

All users of synthesis telescopes should at some point look at their visibility data. The visibility data are complex quantities sampled at points spread across the (u,v) plane, so it can be difficult to visualize the whole dataset. Useful displays include plots of amplitude or phase as functions of time on a single baseline, and

plots of amplitude and phase versus radius in the (u, v) plane or projected onto a particular direction in the (u, v) plane. Plotting subsets of data involving a single antenna can often identify systematic errors. All of these displays, and more, are readily available in the standard packages (AIPS, AIPS++, DIFMAP). On strong sources, including most that are strong enough for self-calibration, these plots will show structure corresponding to the Fourier transform of the brightness distribution. For weak sources, the structure is usually obscured by the scatter of points due to receiver noise, but suitable averaging of the data (up to the coherence time) may allow source features to be seen.

In order to interpret visibility data, it is essential to have a good understanding of the basic properties of the Fourier transform, which are summarized in the Appendix. With some experience, it is quite easy to recognize features of simple sources by inspection of the visibility data. It is usual to try to represent the source as a sum of simple "components". The *addition theorem* ensures that the visibility function is the sum of the complex visibilities of the individual components. For example, a point source model has a visibility function that has constant amplitude but a phase gradient across the (u, v) plane, the gradient depending on the displacement from the phase center. A double-source model, consisting of two point components, has a visibility function that is the sum of two such functions with different phase gradients; the resultant visibility has a sinusoidal amplitude with wavelength inversely proportional to the double separation. Figure 16-1 is a useful summary that shows how the basic properties of a simple model can be estimated by inspection of the dependence of visibility amplitude and phase on position in the (u, v) plane.

Several simple "component" brightness distributions are frequently used (see Appendix). The most common is the gaussian, which has a simple Fourier transform – another gaussian. The uniform disk and the optically-thin sphere are physically somewhat more plausible models, whose Fourier transforms can be expressed in terms of Bessel functions. However, if the components are only barely resolved, the exact functional form of the brightness distribution is unimportant: all these functions have a similar quadratic dependence on baseline at short baselines. Figure 16-2 shows the dependence of visibility amplitude on baseline length for several brightness distributions, adjusted so that the 50%-visibility points coincide. This figure shows, for example, that at short baselines a gaussian of FWHM 1 arcsec is indistinguishable from an optically thin sphere of diameter 1.8 arcsec.

In order to compare a model with an image derived, for example, by Fourier inversion and self-calibration, the model should be convolved with a point-spread function (or "beam"). As in CLEAN, it is conventional to use a gaussian beam. If the model is made up of gaussian components, this is straightforward: the convolution of two elliptical gaussians is another elliptical gaussian whose parameters can be determined analytically (Wild 1970). (For circular gaussians, the widths just add quadratically.) For the other component types, a numerical convolution must be used.

3.1. An Example

As an example, I will consider a data set from a VLBI observation in which the source 2021+614 was observed in four hour-long "snapshots" on 11 antennas.

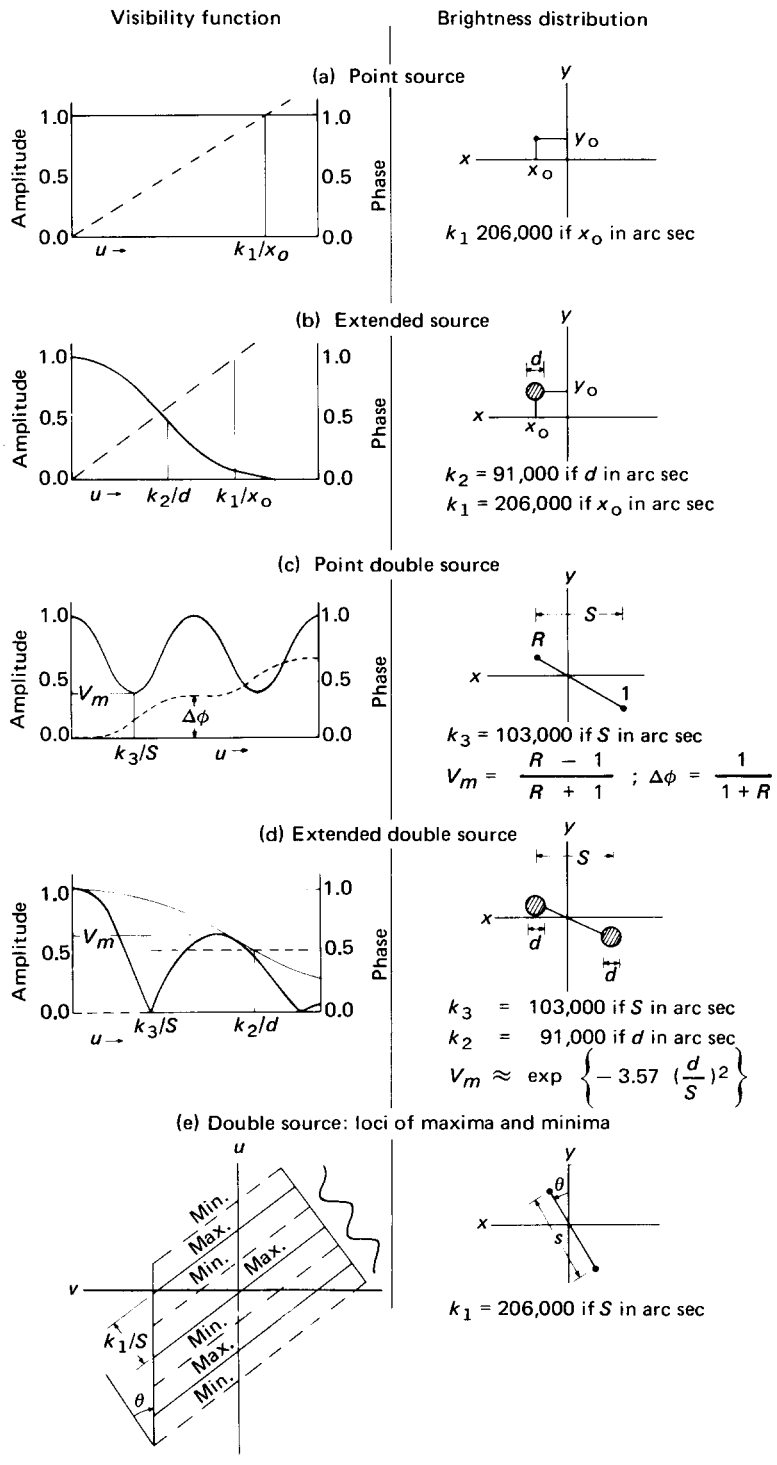

Figure 16-1. The visibility functions for various brightness distribution models (reproduced from Fomalont & Wright 1974; © Springer, Berlin).

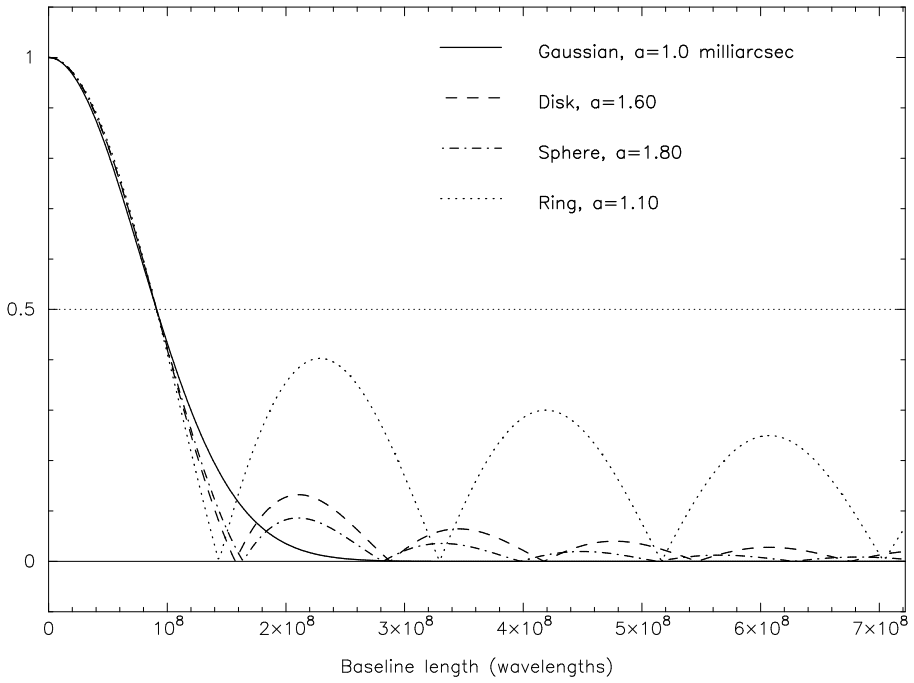

Figure 16-2. Dependence of visibility amplitude on baseline length for four different circularly symmetric brightness distributions. Analytical expressions for these functions are given in the Appendix. The scale-size a for each distribution has been adjusted so that the visibility of each drops to 50% at the same baseline.

There are many ways of examining the visibility data from such an observation, and I encourage you to use all the tools at your disposal to do so. Some of the most useful projections of the two-dimensional complex data set are shown in Figure 16-3, which was obtained with the program DIFMAP (Shepherd 1997). A simple plot of the (u, v) coverage (Fig. 16-3a) shows the spatial frequencies to which the observation is sensitive. It can also be useful to encode visibility amplitude information by color or symbol size. The graph of amplitude versus projected baseline length (Fig. 16-3b) in this case shows that the source is resolved (not dominated by a point component of constant amplitude) and has subcomponents that beat against each other to give a wide range of visibility amplitudes at each baseline. By projecting onto a line in the (u, v) plane (Fig. 16-3c) we can see that to first approximation the source is an equal double; after adjusting the projection angle to make the minima line up, we find that the visibility takes on the canonical form shown in Fig. 1d. By comparison with Fig. 1d, we find an approximate *starting model*: two equal components, each of 1.25 Jy, separated by about 6.8 mas in p.a. 33°. From the upper envelope, the size of each component is about 0.8 mas (gaussian FWHM).

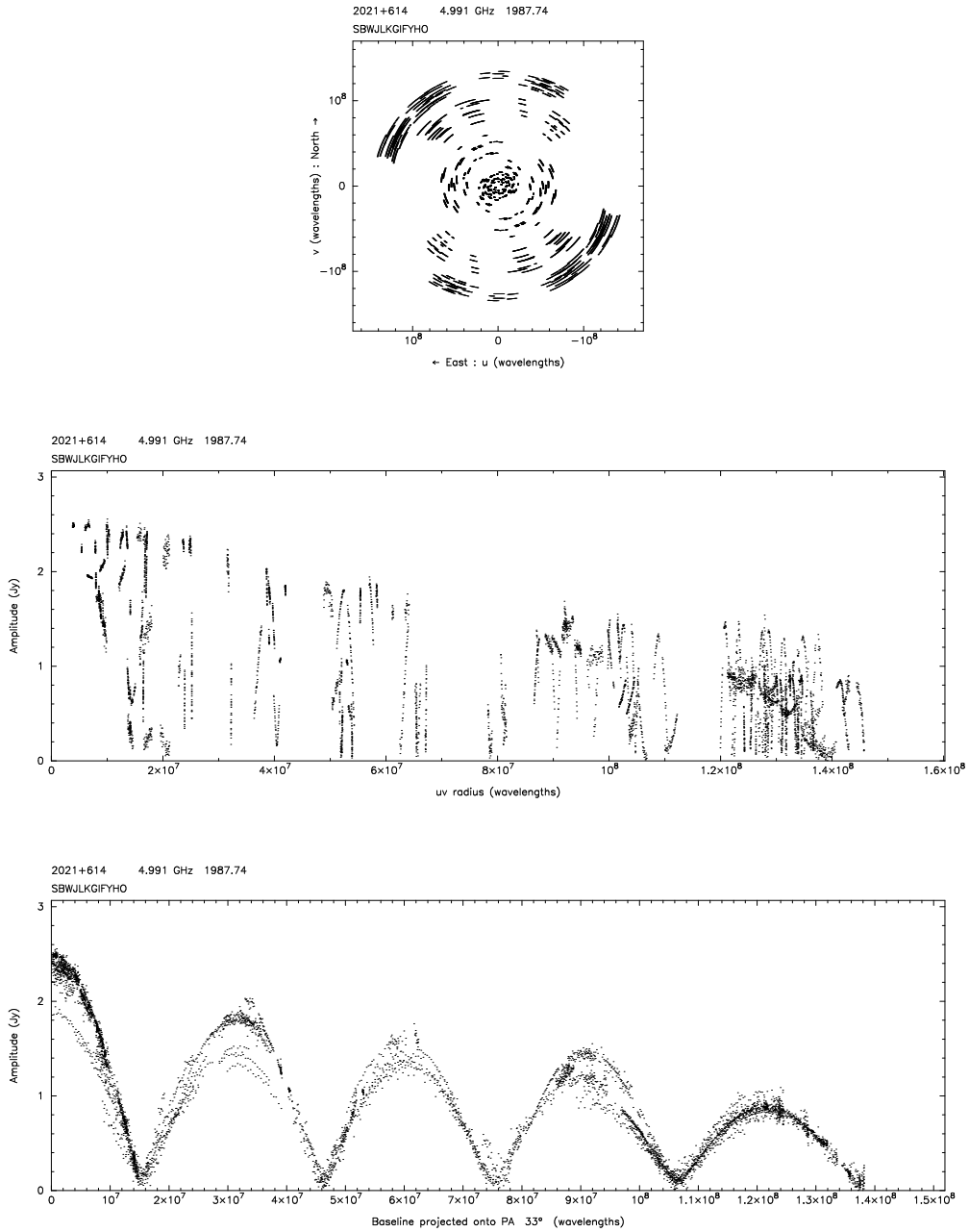

Figure 16-3. Observed visibility data from a 5-GHz Mark-II VLBI observation of the radio galaxy 2021+614 in 1987 (Conway et al. 1994). (a) The sampling of the u, v plane. (b) Visibility amplitude as a function of radius in the u, v plane (projected baseline). (c) Amplitude as a function of the component of baseline length projected in position angle 33°.

4. The Closure Quantities

The closure phase (Jennison 1958; Rogers et al. 1974) is the sum of the visibility phases around a triangle of three baselines:

$$\Psi_{lmn}(t) = \phi_{lm}(t) + \phi_{mn}(t) + \phi_{nl}(t), \tag{16-2}$$

where $\phi_{lm}(t)$ is the visibility phase on the baseline between antennas l and m at time t. In this sum, antenna-based phase errors cancel (Pearson & Readhead 1984). The closure phase is the argument of the *bispectrum* of the sky brightness distribution, which is the triple product of complex visibilities on a closed triangle of baselines:

$$\text{Bispectrum} = V(u,v)V(u',v')V(-u-u',-v-v'). \tag{16-3}$$

This makes it clear that the closure phase is a function of four variables (two positions in the (u,v) plane). In practice this four-dimensional space is always poorly sampled, and the relationship of the observed closure phases to the sky brightness distribution is far from intuitive. However, in an array of N antennas with $N(N-1)/2$ baselines, the fraction of the visibility phase information that is available from the closure phases is $(N-2)/N$.

Another "good observable" is the closure amplitude (Twiss, Carter, & Little 1960; Readhead et al. 1980). This is the ratio of visibility amplitudes on baselines between four antennas arranged so that antenna gains cancel:

$$\frac{|V_{kl}| \cdot |V_{mn}|}{|V_{km}| \cdot |V_{ln}|}. \tag{16-4}$$

The closure amplitude is a function of six variables (three points in the (u,v) plane). The fraction of the amplitude information available from the closure amplitudes is $(N-3)/(N-1)$.

In many cases it is possible to construct an image from the closure quantities by the iterative self-calibration procedures discussed in other lectures, but if the data are sparse it is often better to interpret the closure quantities directly using the methods outlined in this lecture.

Several programs are available for inspection of closure phases and closure amplitudes, but because they are functions of 4 or 6 variables, it is rarely possible to learn much about the structure of the source simply by inspection of the closure quantities. An exception is a double source which will show large jumps in the closure phase where the amplitude on one baseline has a minimum.

5. Model Fitting

5.1. Imaging as an Inverse Process

Synthesis imaging is a member of the general class of *inverse problems*. In such problems, we understand the *forward problem*, which in our case means that if we knew the true sky brightness we could calculate the measured quantities (complex visibilities or closure quantities) using eqn. 1 or a suitable generalization, but we want to invert this process to estimate the sky brightness from the

measurements. There is a wide literature on inverse problems which emphasizes the difficulties of determining whether there is a unique solution and of devising a stable algorithm that will find the solution (e.g., Parker 1977; Press et al. 1992).

One technique that is generally applicable to inverse problems is *model fitting*. In this technique, one makes a parametric model of the sky brightness distribution and uses the imaging equations to calculate the expected measurements. One then adjusts the parameters of the model to get the "best fit" model. When the error statistics of the observations are understood, the appropriate statistical technique to estimate the parameters of the model is the *maximum likelihood* method. For gaussian errors, this is the same as the *least-squares* method. Press et al. (1992, chapter 15, or chapter 14 in the first edition) give a more extended discussion of least-squares fitting that I strongly recommend to anyone who uses a least-squares program; it is of course essential reading for anyone planning to write such a program. Another useful reference is the book by Bevington & Robinson (1992, chapter 8); the first edition of this book (Bevington 1969, chapter 11) is still useful for Fortran programmers.

There are three steps involved in model fitting: (1) Design a *model* defined by a number of adjustable parameters; (2) Choose a *figure-of-merit* function; (3) Adjust the parameters to *minimize the merit function*. The goals are to obtain: (1) Best-fit values for the parameters; (2) A measure of the goodness-of-fit of the optimized model (relative to the measurement errors); (3) Estimates of the uncertainty in the best-fit parameters.

As an example, consider a model of the N observed visibilities $V_i(u,v)$ (the same technique is applicable to the closure quantities). The model, $F(u,v)$, depends on a number M of parameters a_j (typically there will be 6 parameters per model "component": component flux density, two sky coordinates, angular size, axial ratio, and orientation). The model is intended to reproduce the observations within their uncertainty, i.e.,

$$V(u,v) = F(u,v;a_1,\ldots,a_M) + \text{noise}. \tag{16-5}$$

The *likelihood* of the model is the probability of obtaining the data, assuming that the model is correct. If the noise in the observations is gaussian, so that each visibility measurement V_i has an associated standard deviation σ_i, the likelihood is:

$$L \propto \prod_{i=1}^{N} \left\{ \exp\left[-\frac{1}{2} \left(\frac{V_i - F(u_i,v_i;a_1,\ldots,a_M)}{\sigma_i} \right)^2 \right] \right\}. \tag{16-6}$$

The conventional method of statistical estimation is the *maximum likelihood method*: choose the values of the parameters that maximize L. This is equivalent to minimizing $-\log L$, or minimizing

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{V_i - F(u_i,v_i;a_1,\ldots,a_M)}{\sigma_i} \right)^2 \tag{16-7}$$

Thus the maximum likelihood method is, in this case, the method of least squares: we must minimize χ^2, the weighted sum of squares of the deviations

between the data and the model. Note that this is strictly applicable only if the data have gaussian errors.

If the errors are gaussian, then near the minimum χ^2 follows the chi-square distribution with $\nu = N - M$ degrees of freedom (N data points, M parameters). The expected value of χ^2 is ν with standard deviation $\sqrt{2\nu}$. A measure of the goodness of fit of the optimized model is the *reduced chi-square* $\chi^2/(N-M)$ which should be close to 1 for a good fit. Large values of the reduced chi-square indicate a bad fit, while values much smaller than 1 indicate too good a fit, perhaps because the errors σ_i have been overestimated.

Note that the measured quantities V_i need not be the complex visibilities (or amplitudes and phases): one might for example choose to use the closure quantities if the data set is poorly calibrated.

Model fitting has many desirable properties. It can (in principle) take into account all the details of the measurement process, which need not be a simple Fourier transform; and because it operates in the domain of observation, where the statistics of the measurement process are well understood, it can estimate the statistical uncertainties in the parameters of the best-fit model, which may be the astronomically important quantities. However, it also has serious problems: it may be difficult to choose a suitable parameterization of the model, the solutions are not unique, and it can be much, much slower than the conventional Fourier inversion and deconvolution methods.

5.2. Uses of Model Fitting

Fitting a model to the visibility data is not the appropriate technique to use for every observation, of course. If the primary objective is an image, and there are sufficient data to form a reliable image, then use the standard inversion and deconvolution techniques. Model fitting is very useful, however, for interpreting sparse or uncalibrated data, and for quantitative analysis. In many cases model fitting and conventional imaging will both be useful and can supplement each other. Model fitting is most useful, of course, when the brightness distribution can be represented accurately by a model with a small number of parameters. Examples include fields containing a small number of unresolved sources, where the parameters are the positions and flux densities of the sources; barely resolved sources, where the parameters are position, flux density, and angular size (if the source is not more than 50% resolved on the longest baseline, three parameters – major axis, minor axis, and position angle – are sufficient to characterize the brightness distribution: see Figure 16-2); and sources consisting of a small number of barely-resolved components.

Model fitting is typically used for the following:

1. Checking and adjusting amplitude calibration. Good unresolved calibration sources are difficult to find, especially for VLBI observations, and frequently a calibrator will have some structure. However in many cases the calibrator can be represented accurately by a one- or two-component model. The parameters of such a model can be estimated by model fitting using the best-calibrated baselines of the array, and then the model can be used to calibrate the other antennas. This technique has proved particularly useful on arrays including antennas with a wide range of sensitivities. It is advisable to check the calibration by using two or more calibrators.

2. The hybrid-mapping and self-calibration methods for making images from poorly calibrated or uncalibrated data all require a starting model. Frequently a point-source model can be used, but I have found that the self-calibration process always converges faster if a good starting model is used. This is particularly important for sources that are very dissimilar from a point source, such as almost-equal double sources. A good starting model is also advisable for the Schwab–Cotton global fringe-fitting algorithm used in AIPS task FRING (Schwab & Cotton 1983); indeed if the source is not close to a point source, this algorithm may fail to find fringes altogether if a point-source model is used. One technique that I have found effective is to make a crude image by the standard self-calibration technique starting from a point source, and use the image to deduce a starting model for least-squares model fitting, the result of which is used in turn as the starting point for another round of self-calibration (Biretta, Moore, & Cohen 1986). Use of model fitting in this way can be regarded as the application of another constraint in the self-calibration process; in addition to enforcing positivity and finite support, we are requiring the image to be "simple" and "smooth." Using model fitting instead of CLEAN to refine the self-calibration model thus shares some of the advantages of the Non-Negative Least Squares algorithm (Briggs 1995).

3. If the source is simple enough to be accurately represented by a parametric model, then model fitting provides accurate estimates of the model parameters, such as component flux densities and separations, with error estimates. Estimates derived in this way directly from the visibility data will almost always be more reliable than parameters estimated from a deconvolved image. Model fitting is also a deconvolution process: the component size estimates or positional uncertainties may be much smaller than the beam, but they can still be reliable if the signal-to-noise ratio is high enough. The error estimates should provide a guide to the reliability of the deconvolution. The drawback of this approach is that the form of the chosen model may not be correct, and almost certainly will not be unique; it is necessary to explore a range of different model types.

4. Model fitting can be used to improve the behavior of the commonly used imaging techniques CLEAN and MEM. For example, for an extended source CLEAN often gives better results if a large part of the extended structure can be modeled and subtracted before deconvolution; MEM can give better results if dominant point sources are modeled and subtracted.

5. The limiting factor in astrometric and geodetic VLBI is often resolved structure in the reference sources. Rather than try to image each source from the astrometric data themselves (which are usually not planned for optimum imaging), it may be better to use a simple parametric model of each source adjusted to fit the complete ensemble of observations; if necessary, the parameters can be adjusted to take into account smoothly varying source structure.

5.3. Practical Model Fitting

I now return to the example dataset introduced in Section 3.1. The simple model that we derived by visual inspection is shown in Table 1. We can now use a least-squares model-fitting program to improve this model. This data set has been calibrated in amplitude but not in phase, so we fit to the visibility amplitudes and closure phases. The initial "eyeball" model has agreement factors (square root of reduced chi-square) of 5.9 (amplitudes) and 4.2 (closure phases). After adjusting the parameters the agreement factors are improved to 3.9 and 3.1. Further improvement requires the introduction of more components; after some perseverance, I was able to obtain the 5-component model shown in Table 16-1 and Figure 16-4, with agreement factors of 1.3 and 1.5.

Table 16-1. Model Parameters

Flux	Radius	Theta	Axis	Ratio	Phi
Starting ("eyeball") model					
1.250	0.000	0.0	0.80	1.00	0.0
1.250	6.800	33.0	0.80	1.00	0.0
Best 2-component model					
1.185	0.000	0.0	0.81	0.76	45.1
1.141	6.788	32.9	0.89	0.79	54.0
Best 5-component model					
1.008	0.000	0.0	0.70	0.73	46.7
1.094	6.786	32.9	0.89	0.75	56.1
0.142	1.169	45.2	1.79	0.25	-17.7
0.128	9.378	41.4	0.78	0.77	19.0
0.120	1.893	72.6	3.89	0.00	55.3

The parameters given for each component are: flux density (Jy), position of center (polar coordinates relative to phase center, with radius in milliarcsec), FWHM major axis (milliarcsec), axial ratio (minor/major), and position angle of major axis.

5.4. Programs for Model Fitting

Several programs are available for model fitting using the complex visibilities. In the AIPS package, the task UVFIT does a least-squares fit of a model to complex visibility data (real part and imaginary part); this is preferable to fitting to amplitude and phase for reasons discussed below. SLIME (Flatters 1998) is an add-on AIPS task that combines model fitting with a graphical editor. Martin Shepherd's program DIFMAP also provides a convenient graphical interface for model fitting and data inspection. All of these programs allow the model to be defined as a sum of gaussian, optically-thin sphere, or other components (see the Appendix). If the data are poorly calibrated, model fitting and self-calibration must be alternated in an iterative process, in order to derive both the model and the calibration parameters.

In some cases, a better approach to uncalibrated data would be to fit to the closure quantities directly, or to amplitudes and closure phases if only the phases are badly calibrated. Because there can be a very large number of closure

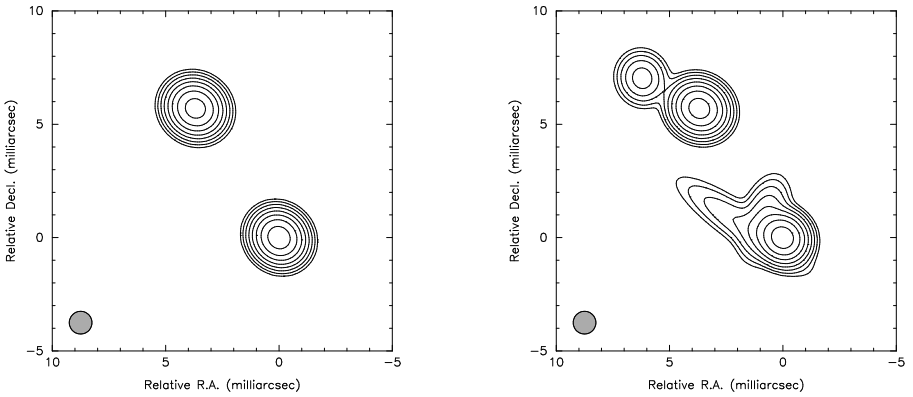

Figure 16-4. Contour maps of two gaussian models of 2021+614, convolved with a circular gaussian beam of FWHM 1 milliarcsec. Contours are drawn at 0.5, 1, 2, 4, ..., 64% of the peak. *Left*: best two-component model; *right*: best five-component model.

quantities (although they are not all independent), this is less well supported by existing programs. One program, MODELFIT, originally written by George Purcell and Richard Simon, is available in the Caltech VLBI package (Pearson 1991); however, this program is now showing its age and cannot cope with modern multichannel datasets.

Similar least-squares model fitting can be applied in the image plane. The AIPS task JMFIT can fit one or more gaussian components to a region of an image. This can sometimes be a useful way to derive a starting model for (u, v)-plane fitting. Image-plane fitting is most useful for estimating the parameters of a source when there is complex emission elsewhere in the field. However, a proper error analysis, taking into account correlations between pixels, is difficult.

5.5. Limitations of the Least-Squares Method

Note that when we use the least-squares method, we are making the following assumptions:

1. The model is actually a good fit to the data. This should be tested by checking that the reduced chi-square is close to 1.

2. The errors are actually gaussian. This is true if the data are the real and imaginary parts of the observed visibility, but not if the data are visibility amplitudes, closure phases, or closure amplitudes, which do not have gaussian distributions except in the limit of high signal-to-noise ratio. The least-squares method is frequently (incorrectly) used with non-gaussian data, but it must be used with considerable caution: the results may be biased and any estimates of the errors on the fitted parameters may be wrong.

3. The errors are known. In most cases, an accurate estimate of the errors in the real and imaginary parts of the visibilities should be available from the system temperatures, from the statistics of bit counting in the correlator, or from the scatter of points within an integration. However, this information is often discarded during data reduction and may be difficult to recover.

4. There are no systematic (calibration) errors. Systematic errors that are not removed in the calibration can in principle be estimated as additional parameters in the model; e.g., one could regard the antenna gain factors as unknown parameters to be estimated. When data have been self-calibrated in amplitude or phase before model fitting, additional parameters (antenna gains or phases) have been estimated from the data. This reduces the number of degrees of freedom. If this is not taken into account when calculating the reduced chi-square, then the model will appear to fit the data better than it should (reduced chi-square too small).

5. The errors are uncorrelated. This should be the case for additive noise in the observed visibilities, but it is not true, for example, of errors in the closure phases or closure amplitudes if non-independent closure quantities are included in the fit.

5.6. Least-Squares Algorithms

The goal of any least-squares algorithm is to find the global minimum of χ^2 in the M-dimensional parameter space. At any minimum, the derivatives of χ^2 with respect to the parameters will be zero,

$$\nabla\chi^2 = \frac{\partial\chi^2}{\partial a_k} = 0. \tag{16-8}$$

If the model $F(u, v; a_i, \ldots, a_M)$ is a linear function of the parameters a_i, this set of M equations can be solved by standard matrix-inversion methods. In most cases, however, the model is a non-linear function of the parameters and an iterative technique must be used.

One simple method is the "grid search" (Bevington & Robinson 1992): select starting trial values for the parameters, and adjust each parameter in turn (keeping the others fixed) to minimize χ^2 with respect to that parameter. The disadvantages of this method are that it can be very slow, and it is not obvious by how much one should change each parameter at each trial.

A more efficient method is the "gradient search": at each step increment all the parameters to move in the direction of the gradient of χ^2 in parameter space. The gradient of χ^2 is usually calculated numerically by evaluating the change in χ^2 for small increments of each parameter; thus one step in the gradient search may require many more model evaluations than one step in the grid search, but it should take one closer to the minimum. Alternatively, $\nabla\chi^2$ can be determined directly if analytical expressions for the derivative of the model with respect to each parameter are available (DIFMAP uses this method: all the standard model component shapes have analytical derivatives). The magnitude of the step to

take along the gradient can be estimated by examining the second derivative of χ^2, the *Hessian* matrix

$$\nabla^2 \chi^2 = \frac{\partial^2 \chi^2}{\partial a_k \partial a_l} \tag{16-9}$$

The most commonly used algorithm is a refinement of this simple gradient search known as the Levenberg-Marquardt method; for details, see Press et al. (1992) and Bevington & Robinson (1992). More sophisticated algorithms are also available (e.g., Bunch, Gay, & Welsch 1993).

In practice, you are likely to run into a number of problems whatever algorithm you use:

1. Finding the global minimum. It is very easy to get stuck in a local minimum of χ^2 which may be far from the global minimum (where the reduced chi-square should be close to 1). The better the starting point, the more likely you are to find the global minimum, which is why you should spend some time refining the starting model by inspection of the visibilities before using a least-squares program. The grid search method can sometimes find its way out of a local minimum where the gradient search method gets stuck.

2. Slow convergence. Often the minimum χ^2 lies in a wide flat valley, where changes to some of the parameters make little change to χ^2. Gradient search should converge faster than grid search in this case. However, this is often a symptom of a poorly-constrained model where parameters are not independent. For example, if data are available in only a limited range of (u, v) distance, a strong, wide component may be difficult to distinguish from a weaker, more compact component. In such cases it is best to constrain some of the parameters to their *a priori* values and let the program adjust the others; but remember that the result is not a unique solution.

3. Constraints. A physically realistic model will have positive component flux densities, and the model-fitting program should apply such constraints. Other constraints may be artificial ones imposed by the formulation of the model; e.g., if the component position is specified by polar coordinates (r, θ), r should be positive, and θ is poorly constrained when r is small. A better approach might be to choose orthogonal coordinates $(x, y) = (r\cos\theta, r\sin\theta)$ as the adjustable parameters. Similar considerations apply to major axis, axial ratio, and component position angle: the algorithm often finds its best solution at a boundary of parameter space with a zero axial ratio (this is apparent in the model in Table 1).

4. Choosing the right number of parameters or components. Model fitting does not have a unique solution, and with a sufficient number of variable parameters many equally good solutions can be found. One should not introduce more parameters than are necessary to obtain a reduced chi-square close to unity. Thus if a circular gaussian component is a good fit, it is not necessary to adjust the axial ratio and position angle. The appropriate statistical test for determining whether additional parameters

actually improve the fit is the F-test (Bevington & Robinson 1992, chapter 11). However, such tests are not very useful if the assumption that the data points have independent gaussian noise of known magnitude is not valid.

5.7. Error Estimation

Press et al. (1992, section 15.6) give a good account of methods for determining confidence limits on the estimated model parameters. Most model-fitting programs determine the curvature of the χ^2 surface around the minimum, from which a covariance matrix for the fitted parameters can be determined. It is a good idea to look at the covariance matrix to see which parameters are well constrained and how the parameters are correlated. However, it is dangerous to use the covariance matrix for direct estimates of the uncertainties of the parameters, because the assumptions that go into the theory of least-squares are frequently violated. A better approach is to use contours of constant chi-square around the minimum to define confidence limits on the parameters, i.e., find the region of parameter space in which

$$\chi^2 < \chi^2_{\rm min} + \Delta\chi^2. \tag{16-10}$$

The choice of $\Delta\chi^2$ depends on the required *confidence level* and the number of parameters estimated; e.g., for 68.3% confidence and one parameter, $\Delta\chi^2 = 1$; for 90% confidence and 6 parameters, $\Delta\chi^2 = 16.8$. One must distinguish confidence intervals on a set of parameters considered jointly from confidence intervals on a single parameter. In the latter case, the goal is to determine how bad the fit gets as the specified parameter is changed from its optimum value, while adjusting the other parameters to compensate. The 68.3% confidence range for a single parameter can be found by projecting the contour $\Delta\chi^2 = 1$ onto the axis corresponding to that parameter.

It is important to remember that these theoretical confidence limits will not apply if the data are not gaussian or not independent, or if the best-fit model is not in fact a good fit. It is a good idea to test these assumptions by other methods. The Monte-Carlo method involves using a model of the sky brightness distribution to generate simulated data sets observed under similar conditions to the real observations. By applying the model-fitting procedure to these data sets, and comparing the results with the input model, one can get some idea of the uncertainties involved. The uncertainties are likely to be larger for real data sets which may contain errors of unknown origin that cannot be simulated. Another approach is to obtain multiple data sets by repeated observation, or by dividing a data set into subsets.

6. Some Applications

6.1. Superluminal Motion

An important application of visibility analysis is in the detection and measurement of *changes* in the brightness distribution of a source. For example, in superluminal sources we compare the relative positions of components at different epochs in order to measure an apparent expansion speed. Indeed, superluminal

motion was first discovered, in the quasar 3C 279, in VLBI observations on a single baseline: sharp minima in the visibility were seen to move toward the origin in the (u, v) plane, corresponding to an expansion of a two-component model of the brightness distribution (Whitney et al. 1971).

Even with good (u, v) coverage, using a direct comparison of images made with the standard self-calibration and deconvolution procedures it can be difficult to disentangle real changes in the source from differences between the images due to calibration errors or differences in (u, v) coverage. A better approach is to compare the visibilities or closure quantities directly, using model fitting as a technique for overcoming the differences in (u, v) coverage. First find a model that is a good fit to the data obtained in one observation (A, say). Usually this will involve imaging and self-calibrating the data set and using the image as a guide to choosing the model. The model can then be transformed to the (u, v) plane and compared directly with the measured visibility samples from the other observation (B). Of course this technique should only be used to interpolate between small differences in (u, v) coverage, not to extrapolate to very different regions of the (u, v) plane. Calibration errors can be circumvented by comparing the closure quantities: significant changes in closure phases or closure amplitudes should be a reliable indicator of real changes in the source. If differences between the two observations are detected in this way, the model can be used to suggest a physical interpretation of the changes between A and B: for example, which components have changed, and are the changes in component flux density, size, or location?

Conway et al. (1994) used this technique to compare observations of the radio galaxy 2021+614 made in 1982 and 1987 (the data presented in Figure 16-3). Although there are significant differences in visibility between the two data sets on the few baselines in common, images made independently from the two data sets are very similar; but by a careful visibility analysis using model fitting, Conway et al. detected an expansion of 69 ± 12 microarcsec, corresponding to a sub-luminal expansion at $0.13c$ (assuming a Hubble constant of 100 km s^{-1} Mpc^{-1}). Note that the estimated uncertainty (due to thermal noise) is much smaller than the beam size (about 1 milliarcsec); this is due to the high signal-to-noise ratio of the two bright components. The greatest remaining uncertainty is the effect of calibration errors.

A similar technique has been used to measure the expansion speeds and distances of planetary nebulae using the VLA (Masson 1986; Kawamura & Masson 1996).

6.2. Gravitational Lenses

A gravitational lens can split the image of a background source into two or more components. If the background source varies in intensity, there will be a time delay between the corresponding variations in the individual image components. Together with the redshifts of the lens and background source, and a model of the gravitational potential of the lens, the time delay can be used to determine the linear scale of the lens, and hence its distance by comparison with the angular scale. This is a powerful technique for measuring the Hubble constant. Many of the best lensed systems are radio sources and the images can be monitored with the VLA or MERLIN to detect variations. It is necessary to measure the

component flux densities with high accuracy ($< 1\%$) over a long period, although it may be impossible to use the same (u, v) coverage for every observation. This is an ideal case for model fitting: there are few free parameters (flux densities of a few point sources of known position, but not well resolved from each other), and we need reliable error estimates. In some cases it may be necessary to include unrelated field sources in the model. This technique has been used, for example, to estimate intercomponent time delays in the lens system B1608+656 (Myers et al. 1995; Fassnacht et al., in preparation).

6.3. The Sunyaev-Zeldovich Effect

In the Sunyaev-Zeldovich effect, inverse Compton scattering of photons of the cosmic microwave background by hot electrons in clusters of galaxies shifts the photons to higher energies. The result is a decrement in the background intensity at centimeter wavelengths (e.g., Birkinshaw 1998). The magnitude of the decrement is proportional to the integral of the electron density through the cluster. The electron density can be estimated from X-ray observations, so this gives a measure of the linear depth of the cluster. Comparison with the angular size of the cluster then gives an estimate of the distance and the Hubble constant. Several groups have made synthesis observations of clusters of galaxies to measure the decrement (e.g., Carlstrom, Joy, & Grego 1996 used the Owens Valley millimeter array; Jones et al. 1993 used the Ryle Telescope at Cambridge). The typical angular scale is a few arc minutes, so short baselines are needed.

Typically clusters are modeled with an isothermal β model, for which the expected profile in the decrement is

$$f(r) = f_0 \left(1 + \frac{r^2}{a^2}\right)^{(1-3\beta)/2}, \tag{16-11}$$

where r is the angle from the cluster center and a is a core radius. The visibility, given by a Hankel transform (see the Appendix), is

$$F(\rho) = (2\pi f_0 a^2)\frac{(\pi a\rho)^{n-1}}{\Gamma(n)} K_{1-n}(2\pi a\rho), \tag{16-12}$$

where $n = (3\beta - 1)/2$ and K is the modified Bessel function of the second kind. The visibility decreases exponentially with baseline; e.g., for the typical case $\beta = 2/3$,

$$F(\rho) = (2\pi f_0 a^2)\frac{\exp(-2\pi a\rho)}{2\pi a\rho}. \tag{16-13}$$

The steep increase in visibility to short baselines, and the central "hole" in the (u, v) coverage of all synthesis arrays combine to ensure that a synthesis image will almost always underestimate the magnitude of the decrement. Quantitative analysis requires model fitting to provide estimates of the central decrement and β.

6.4. The Cosmic Microwave Background Radiation

Several groups are conducting or planning interferometric observations of the fluctuations in the microwave background radiation (e.g., the Cambridge Cosmic Anisotropy Telescope: Scott et al. 1996). Although it is possible to make

images, this is another field in which the analysis is best done in the (u, v) plane. The background radiation is expected to show gaussian fluctuations with characteristic angular scales from a few arc minutes to a few degrees. The angular power spectrum of the fluctuations depends sensitively on the cosmological parameters (including the density parameters Ω, Ω_b, and Ω_Λ and the Hubble constant) and an accurate measurement of the power spectrum would be of great value in measuring these parameters. The power spectrum is the square of the Fourier transform of the sky brightness distribution, and is thus closely related to visibility: the expected value of the square of the visibility at (u, v) is proportional to the power spectrum $C(u, v)$ convolved with the square of the Fourier transform of the primary beam $\mathcal{A}(l, m)$. (Here $C(u, v)$ corresponds to the commonly used spherical power spectrum C_l at multipole order $l \approx 2\pi\sqrt{u^2 + v^2}$). Optimum methods to extract the power spectrum from visibility measurements in the presence of noise and possible foreground emission are being developed by several groups (e.g., Hobson, Lasenby, & Jones 1995; Maisinger, Hobson, & Lasenby 1997; White, Carlstrom, & Dragovan 1998).

7. Appendix

7.1. Properties of the Fourier Transform

Given a model brightness distribution $f(l, m)$ in the sky plane, the model visibility $F(u, v)$ is computed by Fourier transformation:

$$F(u, v) = \mathfrak{F}\{f(l, m)\} \tag{16-14}$$

i.e.,

$$F(u, v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(l, m) \exp[-2\pi i(ul + vm)]\, dl\, dm \tag{16-15}$$

The following properties of the Fourier transform are useful for interpreting visibility data. For details, see Bracewell (1986).

Addition Theorem

$$\mathfrak{F}\{f(l, m) + g(l, m)\} = F(u, v) + G(u, v) \tag{16-16}$$

Convolution

$$\mathfrak{F}\{f(l, m) \star g(l, m)\} = F(u, v) \cdot G(u, v) \tag{16-17}$$

Shift Theorem

$$\mathfrak{F}\{f(l - l_i, m - m_i)\} = F(u, v) \exp[-2\pi i(ul_i + vm_i)] \tag{16-18}$$

Similarity Theorem

$$\mathfrak{F}\{f(al, bm)\} = \frac{1}{|ab|} F\left(\frac{u}{a}, \frac{v}{b}\right) \tag{16-19}$$

7.2. Simple Models

Here I give analytic expressions for some of the commonly-used model components (Purcell 1973). All the expressions are for a circularly symmetric component $f(r)$ centered at the origin of the (l, m) coordinate system, with $r = \sqrt{l^2 + m^2}$. The Fourier transform $F(\rho)$ is circularly symmetric in the (u, v) plane, with $\rho = \sqrt{u^2 + v^2}$. The relationship between $f(r)$ and $F(\rho)$ is a Hankel transform:

$$F(\rho) = 2\pi \int_0^\infty f(r) J_0(2\pi r\rho) r\, dr \tag{16-20}$$

By application of the theorems of Section 7.1., it is straightforward to derive the expressions for elliptical components with arbitrary positions and orientations.

Delta Function (Point Source)

$$f(x, y) = \delta(x, y), \tag{16-21}$$
$$F(u, v) = 1. \tag{16-22}$$

Gaussian

$$f(r) = \frac{1}{\sqrt{\pi/4\ln 2}\, a} \exp\left(\frac{-4\ln 2\, r^2}{a^2}\right) \tag{16-23}$$
$$F(\rho) = \exp\left(\frac{-(\pi a\rho)^2}{4\ln 2}\right), \tag{16-24}$$

where a = full width to half-maximum intensity (FWHM).

Uniformly Bright Disk

$$f(r) = \begin{cases} 4/(\pi a^2), & \text{if } r \le a/2 \\ 0, & \text{otherwise} \end{cases} \tag{16-25}$$
$$F(\rho) = \frac{2J_1(\pi a\rho)}{\pi a\rho}, \tag{16-26}$$

where a = diameter.

Optically Thin Sphere (The brightness at each point is proportional to the path length through the sphere.)

$$f(r) = \begin{cases} 6/(\pi a^2)\sqrt{1 - (2r/a)^2}, & \text{if } r \le a/2 \\ 0, & \text{otherwise} \end{cases} \tag{16-27}$$
$$F(\rho) = 3\sqrt{\pi/2} J_{3/2}(\pi a\rho)(\pi a\rho)^{-3/2}$$
$$= \frac{3}{(\pi a\rho)^3}\left[\sin(\pi a\rho) - \pi a\rho\cos(\pi a\rho)\right], \tag{16-28}$$

where a = diameter.

Ring (The brightness is zero except on the circumference.)

$$f(r) = \frac{1}{\pi a}\delta(r - a/2), \quad (16\text{-}29)$$

$$F(\rho) = J_0(\pi a \rho), \quad (16\text{-}30)$$

where a = diameter.

Acknowledgments. This work was supported by the National Science Foundation under grants AST 9117100, AST 9420018, and AST 9802989.

References

Bevington, P. R. 1969, *Data Reduction and Error Analysis for the Physical Sciences*, (New York: McGraw-Hill).

Bevington, P. R. and Robinson, D. K. 1992, *Data Reduction and Error Analysis for the Physical Sciences*, (New York: McGraw-Hill), 2nd edition.

Biretta, J. A., Moore, R. L., and Cohen, M. H. 1986, The evolution of the compact radio source in 3C 345. I. VLBI observations, *Astrophys. J.*, **308**, 93–109.

Birkinshaw, M. 1998, The Sunyaev-Zel'dovich effect, *Physics Reports*, in press.

Bracewell, R. N. 1986, *The Fourier Transform and its Applications*, (New York: McGraw-Hill), 2nd edition.

Briggs, D. S. 1995, High fidelity deconvolution of moderately resolved sources, PhD thesis, New Mexico Institute of Mining and Technology.

Bunch, D. S., Gay, D. M., and Welsch, R. E. 1993, Algorithm 717: subroutines for maximum likelihood and quasi-likelihood estimation of parameters in nonlinear regression models, *ACM Trans. Mathematical Software*, **19**, 109–130.

Carlstrom, J. E., Joy, M., and Grego, L. 1996, Interferometric imaging of the Sunyaev-Zeldovich effect at 30 GHz, *Astrophys. J.*, **456**, L75–L78.

Conway, J. E., Myers, S. T., Pearson, T. J., Readhead, A. C. S., Unwin, S. C., and Xu, W. 1994, Evidence for two classes of parsec-scale radio double source in active galactic nuclei, *Astrophys. J.*, **425**, 568–581.

Flatters, C. 1988, `http://www.aoc.nrao.edu/~cflatter/slime.html`.

Fomalont, E. B. and Wright, M. C. H. 1974, Interferometry and aperture synthesis, in *Galactic and Extragalactic Radio Astronomy*, ed. G. L. Verschuur and K. I. Kellermann, (Berlin: Springer), 256–290.

Hobson, M. P., Lasenby, A. N., and Jones, M. E. 1995, A Bayesian method for analysing interferometer observations of cosmic microwave background fluctuations, *Mon. Not. R. Astr. Soc.*, **275**, 863–873.

Jennison, R. C. 1958, A phase sensitive interferometer technique for the measurement of the Fourier transforms of spatial brightness distributions of small angular extent, *Mon. Not. R. Astr. Soc.*, **118**, 276–284.

Jones, M., Saunders, R., Alexander, P., Birkinshaw, M., Dillon, N., Grainge, K., Hancock, S., Lasenby, A., Lefebvre, D., Pooley, G., Scott, P., Titterington, D., and Wilson, D. 1993, An image of the Sunyaev-Zel'dovich effect, *Nature*, **365**, 320–323.

Kawamura, J., & Masson, C. 1996, Distances to planetary nebulae BD +30°3639 and NGC 6572, *Astrophys. J.*, **461**, 282–287.

Maisinger, K., Hobson, M. P., and Lasenby, A. N. 1997, A maximum entropy method for reconstructing interferometer maps of fluctuations in the microwave background radiation, *Mon. Not. R. Astr. Soc.*, **290**, 313–326.

Masson, C. R. 1986, Angular expansion measurement with the VLA: The distance to NGC 7027, *Astrophys. J.*, **302**, L27–L30.

Myers, S. T., Fassnacht, C. D., Djorgovski, S. G., Blandford, R. D., Matthews, K., Neugebauer, G., Pearson, T. J., Readhead, A. C. S., Smith, J. D., Thompson, D. J., Womble, D. S., Browne, I. W. A., Wilkinson, P. N., Nair, S., Jackson, N., Snellen, I. A. G., Miley,

G. K., de Bruyn, A. G., and Schilizzi, R. T. 1995, 1608+656: a quadruple-lens system found in the CLASS gravitational lens survey, *Astrophys. J.*, **447**, L5–L8 .

Parker, R. L. 1977, Understanding inverse theory, *Ann. Rev. Earth Planet. Sci.*, **5**, 35–64.

Pearson, T. J. 1991, Caltech VLBI analysis programs, *BAAS*, **23**, 991–992. (See `http://astro.caltech.edu/~tjp/citvlb/`).

Pearson, T. J. and Readhead, A. C. S. 1984, *ARAA*, **22**, 97.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. 1992, *Numerical Recipes*, (Cambridge: Cambridge University Press), 2nd edition.

Purcell, G. H. 1973, The structure of compact radio sources at 609 MHz. PhD thesis, California Institute of Technology.

Readhead, A. C. S., Walker, R. C., Pearson, T. J., and Cohen, M. H. 1980, Mapping radio sources with uncalibrated visibility data, *Nature*, **285**, 137–140.

Rogers, A. E. E., Hinteregger, H. F., Whitney, A. R., Counselman, C. C., Shapiro, I. I., et al. 1974, The structure of radio sources 3C 273B and 3C 84 deduced from the "closure" phases and visibility amplitudes observed with three-element interferometers, *Astrophys. J.*, **193**, 293–301.

Schwab, F. R. and Cotton, W. D. 1983, Global fringe search techniques for VLBI, *Astron. J*, **88**, 688–694.

Scott, P. F., Saunders, R., Pooley, G., O'Sullivan, C., Lasenby, A. N., Jones, M., Hobson, M. P., Duffett-Smith, P. J., and Baker, J. 1996, Measurements of structure in the cosmic background radiation with the Cambridge Cosmic Anisotropy Telescope, *Astrophys. J*, **461**, L1–L4.

Shepherd, M. C. 1997, Difmap: an interactive program for synthesis imaging, in *Astronomical Data Analysis Software and Systems VI*, ed. G. Hunt & H. E. Payne, ASP Conference Series, **125**, 77–84. (See `ftp://astro.caltech.edu/pub/difmap`.)

Twiss, R. W., Carter, A. W. L., and Little, A. G. 1960, Brightness distribution over some strong radio sources at 1427 Mc/s, *Observatory*, **80**, 153–159.

White, M., Carlstrom, J. E., and Dragovan, M. 1998, Interferometric observation of cosmic microwave background anisotropies, *Astrophys. J.*, in press (astro-ph/9712195)

Whitney, A. R., Shapiro, I. I., Rogers, A. E. E., Robertson, D. S., Knight, C. A., Clark, T. A., Goldstein, R. M., Marandino, G. E., and Vandenberg, N. R. 1971, Quasars revisited: rapid time variations observed via very-long-baseline interferometry, *Science*, **173**, 225.

Wild, J. P. 1970, De-convolution of barely resolved radio sources mapped with aerial beams of elliptical cross section, *Aust. J. Physics*, **23**, 113–115.