Very Long Baseline Interferometry and the VLBA
ASP Conference Series, Vol. 82, 1995
J. A. Zensus, P. J. Diamond, and P. J. Napier (eds.)

Printed June 13, 1995

CHAPTER 14

Non-Imaging Data Analysis

TIMOTHY J. PEARSON
CALIFORNIA INSTITUTE OF TECHNOLOGY

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

Contents

14.1 Introduction

In an ideal world, astronomers would be able to regard a VLBI array as a black-box imaging device: they would schedule their observations, wait for the correlation to be completed, pass their data through a standard analysis package, and out would come an image. But this ideal is far from being reached with current VLBI observations, and producing a standard image is not always the best way to analyze VLBI data. In this chapter I want to try to convince you that there is a lot to 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. There are some types of observation 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.

I will restrict myself to continuum data (single frequency channel, single polarization), although the techniques are readily extended to more complex data sets. I will also ignore astrometric and geodetic applications, which are discussed in chapters 19 and 18.

14.2 Visibility Data

As discussed in earlier chapters, the complex visibility $V(u, v)$ measured on a baseline with coordinates (u, v) is the Fourier transform of the sky brightness distribution $I(x, y)$:

$$V(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y) \exp[2\pi i(ux + vy)]\, dx\, dy\,. \tag{14.1}$$

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.

14.2.1 Sampling

Measurements of $V(u, v)$ are available at only a finite number of points in a small region of the (u, v) plane. As described in chapter 1, 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 sidelobes 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.

14.2.2 Closure Quantities

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. In most cases, however, we can form 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. The closure phase (Jennison 1958; Rogers et al. 1974) is the sum of the visibility phases around a triangle of three baselines:

$$\Psi_{lmn}(t) = \phi_{lm}(t) + \phi_{mn}(t) + \phi_{nl}(t)\,, \tag{14.2}$$

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 and 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:

$$\text{Bispectrum} = V(u,v)V(u',v')V(-u-u',-v-v')\,. \tag{14.3}$$

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 and Little 1960; Readhead et al. 1980). This is the ratio of visibility amplitudes on baselines between four antennas arranged so that antenna gains cancel:

$$\frac{|V_{kl}|\cdot|V_{mn}|}{|V_{km}|\cdot|V_{ln}|}\,. \tag{14.4}$$

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 chapters, but if the data are sparse it is often better to interpret the closure quantities directly using the methods outlined in this chapter.

14.2.3 Non-Fourier Imaging

In some cases the simple Fourier transform relationship (eqn. 14.1) 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 equation 14.1 break down.

14.2.4 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 equation 14.1 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, the maximum entropy method (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.

14.3 Model Fitting

14.3.1 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 equation 14.1 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 (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 optimization method is the *maximum-likelihood* method.

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.

14.3.2 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.

In VLBI data analysis, model fitting is typically used for the following:

1. Checking and adjusting amplitude calibration. Good unresolved calibration sources for VLBI observations are difficult to find, 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 not close to a point source, such as an almost-equal double source. A good starting model is also advisable for the Schwab–Cotton global fringe-fitting algorithm used in AIPS task FRING (Schwab and 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 and 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."

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 (see chapters 19 and 18) is often resolved structure in the beacon 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.

14.4 Practical Model Fitting

14.4.1 Simple Models

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 Appendix A. 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 14.1 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 B). The most common is the gaussian, which owes its fame to its

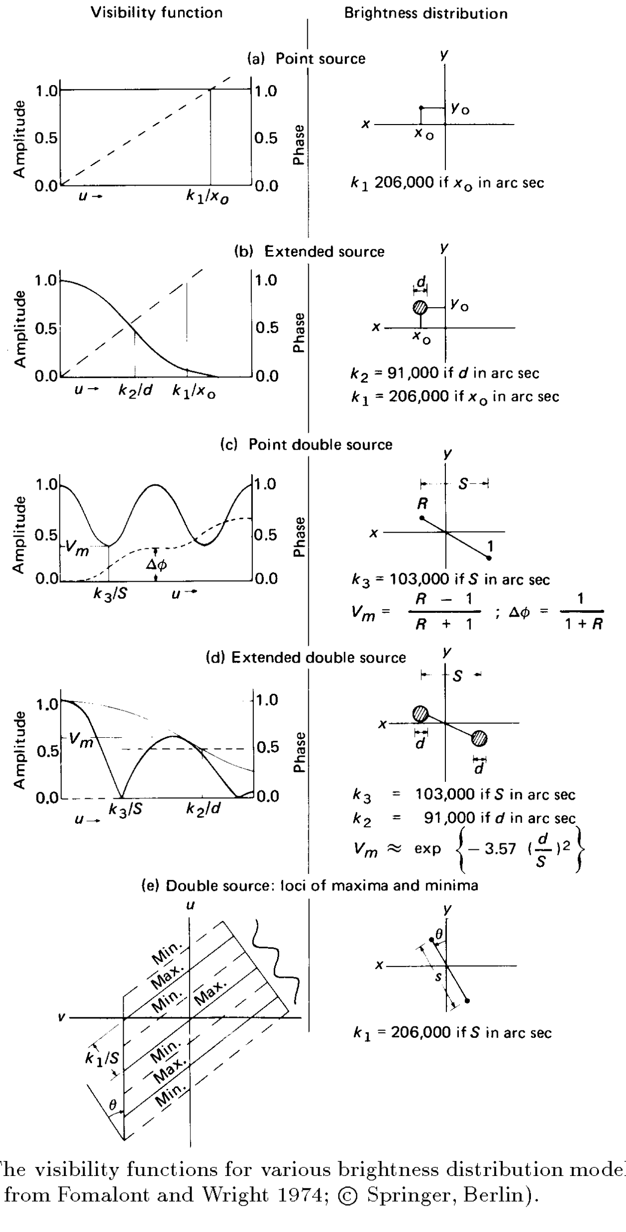

Figure 14.1: The visibility functions for various brightness distribution models (reproduced with permission from Fomalont and Wright 1974; © Springer, Berlin).

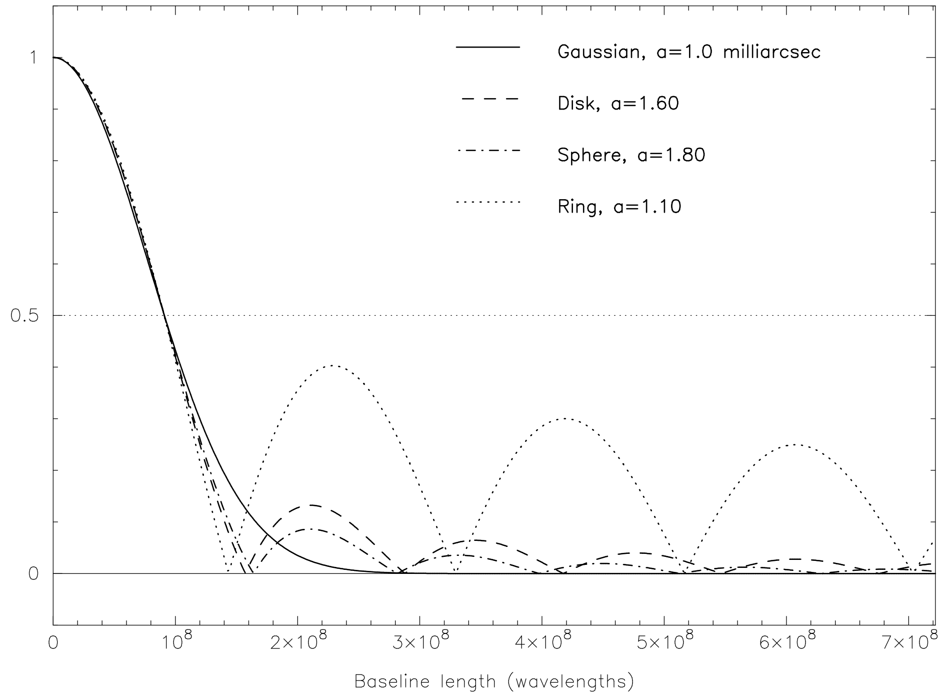

Figure 14.2: Dependence of visibility amplitude on baseline length for four different circularly symmetric brightness distributions. Analytical expressions for these functions are given in Appendix B. The scale-size a for each distribution has been adjusted so that the visibility of each drops to 50% at the same baseline.

simple Fourier transform. 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 14.2 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 milliarcsec is indistinguishable from an optically thin sphere of diameter 1.8 milliarcsec.

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 the other component types, a numerical convolution must be used.

14.4.2 Inspecting Visibility Data

As an example of practical model fitting, I will consider a data set from a Mark II VLBI observation from 1987 in which the source 2021+614 was observed in four hour-long "snapshots" on 11 antennas. 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 14.3, which was obtained with the program DIFMAP in the Caltech package (Shepherd, Pearson and Taylor 1994). A simple plot of the (u, v) plane coverage (fig. 14.3*a*) 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. 14.3*b*) 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. 14.3*c*) 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 figure 14.1*d*. By comparison with figure 14.1*d*, 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).

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 14.1 and figure 14.4, with agreement factors of 1.3 and 1.5.

14.4.3 Programs for Model Fitting

Several programs are available for model fitting. In the Caltech VLBI package (Pearson 1991) there is the program MODELFIT, originally written by George Purcell and Richard Simon. This will do a least-squares fit of a model to visibility amplitudes and closure phases, which is appropriate if the data have good amplitude calibration but poor phase calibration. In a data set with a large number of antennas, however, it is inefficient to explicitly calculate all the closure phases, and a better approach is to combine model-fitting of visibility amplitudes and phases with a self-calibration process to estimate the complex antenna gains. 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. However, the data must be calibrated, and for poorly calibrated data self-calibration and

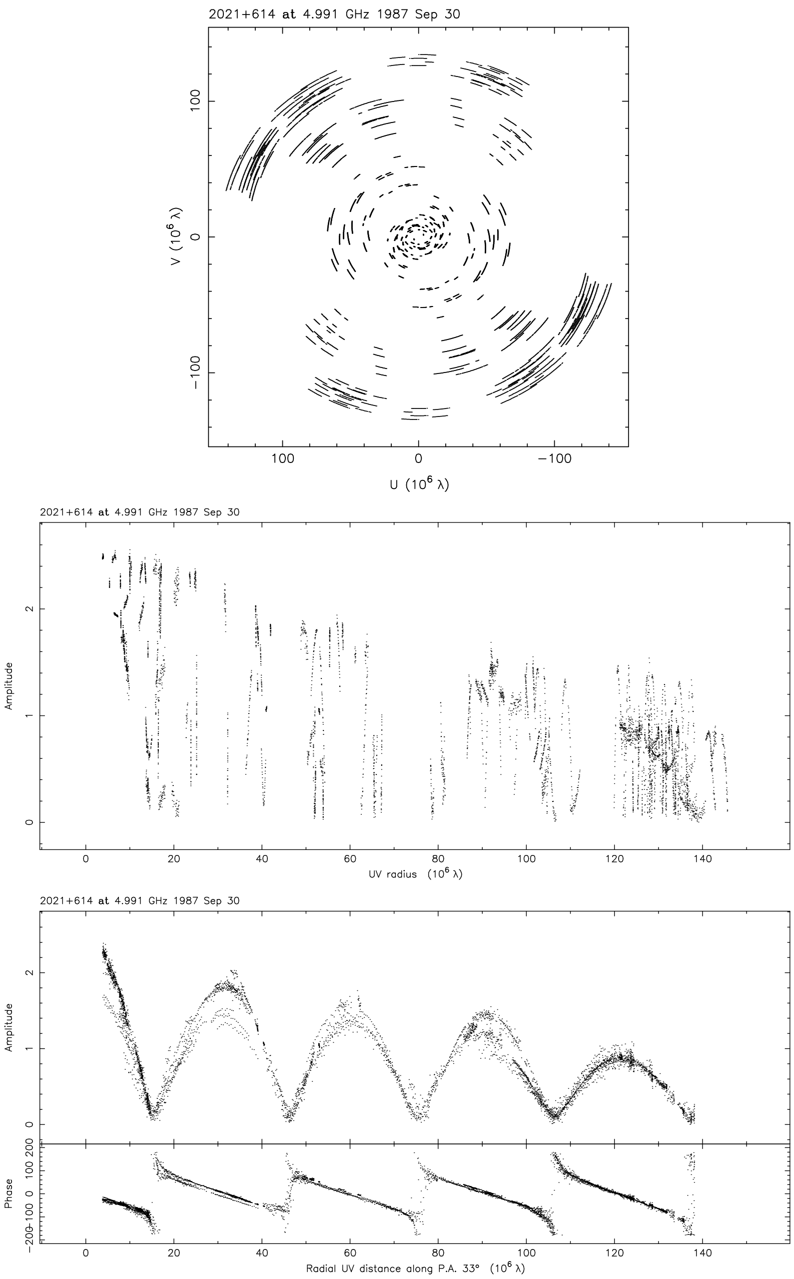

Figure 14.3: Observed visibility data from a 5-GHz Mark-II VLBI observation of the radio galaxy 2021+614 in 1987 (Conway et al. 1994). (a) The sampling of the (u, v) plane. (b) Visibility amplitude as a function of radius in the (u, v) plane (projected baseline). (c) Amplitude and phase as functions of the component of baseline length projected in position angle 33°.

Table 14.1: Model Parameters

Flux (Jy)	Radius (mas)	θ (°)	Axis (mas)	Ratio	ϕ (°)
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

Notes: The parameters of the gaussian components are flux, radius, position angle θ, major axis, axial ratio, and position angle of major axis ϕ.

fitting must be alternated in an iterative process; this is the approach taken for model fitting in the Caltech program DIFMAP (Shepherd, Pearson and Taylor 1994). In MODELFIT, UVFIT, and DIFMAP the model is defined as a sum of gaussian or optically-thin-sphere components (see Appendix B); MODELFIT and DIFMAP also support a number of other component types, though these are generally less useful.

Similar least-squares model-fitting can be applied in the image plane (though this is outside the scope of this chapter). 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.

14.5 Least-squares Fitting Algorithms

In this section I will summarize the method of least squares that is usually used for model fitting, and draw attention to some of its pitfalls. 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 and 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-

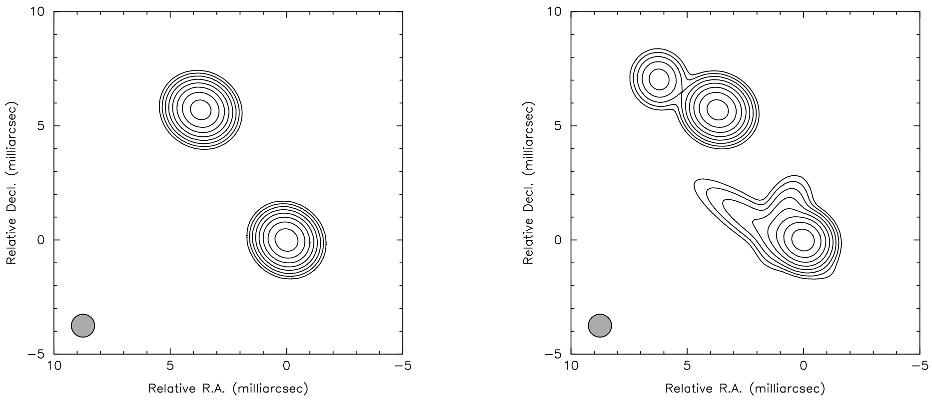

Figure 14.4: Contour maps of two gaussian models of 2021+614, convolved with a circular gaussian beam of FWHM 1 milliarcsec. Contours are drawn at 0.5, 1, 2, 4, ..., 64% of the peak. *Left*: best two-component model; *right*: best five-component model.

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.,

$$V(u,v) = F(u,v;a_1,\ldots,a_M) + \text{noise}\,. \tag{14.5}$$

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:

$$L \propto \prod_{i=1}^{N} \left\{ \exp\left[-\frac{1}{2} \left(\frac{V_i - F(u_i,v_i;a_1,\ldots,a_M)}{\sigma_i} \right)^2 \right] \right\}. \tag{14.6}$$

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

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{V_i - F(u_i,v_i;a_1,\ldots,a_M)}{\sigma_i} \right)^2 . \tag{14.7}$$

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.

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).

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 agreement factor 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. For VLBI data, the errors in the real and imaginary parts of the visibilities can be estimated from the statistics of bit counting in the correlator, from $T_{\rm sys}$, or from the scatter of points within an integration.

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.

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.

14.5.1 Non-Linear 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,

$$\nabla\chi^2 = \frac{\partial\chi^2}{\partial a_k} = 0\,. \tag{14.8}$$

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 and 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. The magnitude of the step to take along the gradient can be estimated by examining the second derivative of χ^2, the *Hessian* matrix

$$\nabla^2\chi^2 = \frac{\partial^2\chi^2}{\partial a_k \partial a_l}\,. \tag{14.9}$$

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 and Robinson (1992).

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: a failing of the Caltech MODELFIT program is that it often finds its best solution at a boundary of parameter space with a zero axial ratio (this is apparent in the model in table 14.1).

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 and 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.

14.5.2 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

$$\chi^2 < \chi^2_{\min} + \Delta\chi^2. \tag{14.10}$$

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.

14.6 Comparison of Observations

An important application of model fitting 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. One way to attempt this is by direct comparison of images made with the standard self-calibration and deconvolution procedures. Unfortunately, with this approach it is 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 (Conway et al. 1994).

In order to detect source changes between the two observations, Conway et al. suggest first finding 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?

As an example, I will return to the radio galaxy 2021+614, for which the A data set was presented in section 14.4.2 on page 275. In this case the A data were obtained in 1987 using short tracks on 11 antennas, while the B data were obtained in 1982 using long tracks on 5 antennas (Conway et al. 1994). 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. We started by obtaining a model that fitted the A data well. This is a poor fit to the B data. By adjusting the calibration of the B data, we could improve the fit, but not eliminate the differences. Thus we concluded that the source had really changed, and attempted to characterize the changes by making minimal modifications to the A model to improve the fit to the B data. We first looked for changes in component flux densities without accompanying position shifts, by allowing only the flux densities to vary in model fitting. In this way we found that component 'D' (the first one listed in table 14.1) increased in flux density by 40% while the others increased by about 15%. These changes are consistent with a flux-density scale difference between observations of 15% and a real increase in D of about 25%. This flux-density change does not account for all the differences, however. The modified model is still a bad fit to some of the B visibilities; in particular, it fails to reproduce the location of a sharp amplitude minimum on one baseline. This strongly suggests that the separation of the two strong components has changed. We confirmed this by model fitting: we could not get a good fit by adjusting component flux densities and sizes only, but needed to allow the position of component D to change. We found a shift of 69 ± 12 microarcsec, corresponding to an apparent speed of expansion of $0.13c$ (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: by adjusting the calibration of the B dataset, we can reduce the required shift to 30 microarcsec, but we cannot eliminate it completely. Thus we believe we have detected a real, subluminal motion in the source.

This work was supported by the National Science Foundation under grant AST 91-17100 to the Owens Valley Radio Observatory.

References

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

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

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

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

Conway, J. E., Myers, S. T., Pearson, T. J., Readhead, A. C. S., Unwin, S. C., & Xu, W. 1994. "Evidence for more than one class of parsec-scale radio double in AGN". *Astrophys. J.* **425**, 568.

Fomalont, E. B., & Wright, M. C. H. 1974. "Interferometry and aperture synthesis". In G. L. Verschuur and K. I. Kellermann (Eds.), *Galactic and Extragalactic Radio Astronomy*. Berlin: Springer, p. 256.

Jennison, R. C. 1958. "A phase sensitive interferometer technique for the measurement of the Fourier transforms of spatial brightness distributions of small angular extent". *Monthly Notices Roy. Astron. Soc.* **118**, 276.

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

Pearson, T. J. 1991. "Caltech VLBI analysis programs". *Bull. Am. Astron. Soc.* **23**, 991.

Pearson, T. J., & Readhead, A. C. S. 1984. "Image formation by self-calibration in radio astronomy". *Ann. Rev. Astron. Astrophys.* **22**, 97.

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

Purcell, G. H. 1973. *The Structure of Compact Radio Sources at 609 MHz*. Ph. D. thesis, California Institute of Technology.

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

Rogers, A. E. E., Hinteregger, H. F., Whitney, A. R., Counselman, C. C., Shapiro, I. I., Wittels, J. J., Klemperer, W. K., Warnock, W. W., Clark, T. A., Hutton, L. K., Marandino, G. E., Rönnäng, B. O., Rydbeck, O. E. H., & Niell, A. E. 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.

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

Shepherd, M. C., Pearson, T. J., & Taylor, G. B. 1994. "DIFMAP: an interactive program for synthesis imaging". *Bull. Am. Astron. Soc.* **26**, 987.

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

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

A Properties of the Fourier Transform

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

$$F(u, v) = \mathrm{FT}\{f(x, y)\}\,, \tag{A1}$$

i.e.,

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\exp[2\pi i(ux+vy)]\,dx\,dy\,. \tag{A2}$$

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

A.1 Addition Theorem

$$\mathrm{FT}\{f(x,y)+g(x,y)\} = F(u,v)+G(u,v) \tag{A3}$$

A.2 Convolution

$$\mathrm{FT}\{f(x,y)\star g(x,y)\} = F(u,v)\cdot G(u,v) \tag{A4}$$

A.3 Shift Theorem

$$\mathrm{FT}\{f(x-x_i,y-y_i)\} = F(u,v)\,\exp[2\pi i(ux_i+vy_i)] \tag{A5}$$

A.4 Similarity Theorem

$$\mathrm{FT}\{f(ax,by)\} = \frac{1}{|ab|}F\left(\frac{u}{a},\frac{v}{b}\right) \tag{A6}$$

B Simple Models

In this Appendix 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 (x,y) coordinate system, with $r=\sqrt{x^2+y^2}$. 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:

$$F(\rho) = 2\pi\int_0^{\infty} f(r)J_0(2\pi r\rho)r\,dr\,. \tag{B7}$$

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

B.1 Delta Function (Point Source)

$$f(x,y) = \delta(x,y), \tag{B8}$$

$$F(u,v) = 1. \tag{B9}$$

B.2 Gaussian

$$f(r) = \frac{1}{\sqrt{\pi/4 \ln 2}a} \exp\left(\frac{-4 \ln 2\, r^2}{a^2}\right) \tag{B10}$$

$$F(\rho) = \exp\left(\frac{-(\pi a \rho)^2}{4 \ln 2}\right), \tag{B11}$$

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

B.3 Uniformly Bright Disk

$$f(r) = \begin{cases} 4/(\pi a^2), & \text{if } r \le a/2 \\ 0, & \text{otherwise} \end{cases} \tag{B12}$$

$$F(\rho) = \frac{2J_1(\pi a \rho)}{\pi a \rho}, \tag{B13}$$

where a = diameter.

B.4 Optically Thin Sphere

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

$$f(r) = \begin{cases} 6/(\pi a^2)\sqrt{1-(2r/a)^2}, & \text{if } r \le a/2 \\ 0, & \text{otherwise} \end{cases} \tag{B14}$$

$$\begin{aligned} F(\rho) &= 3\sqrt{\pi/2}J_{3/2}(\pi a \rho)(\pi a \rho)^{-3/2} \\ &= \frac{3}{(\pi a \rho)^3}\left[\sin(\pi a \rho) - \pi a \rho \cos(\pi a \rho)\right], \end{aligned} \tag{B15}$$

where a = diameter.

B.5 Ring

(The brightness is zero except on the circumference.)

$$f(r) = \frac{1}{\pi a}\delta(r - a/2), \tag{B16}$$

$$F(\rho) = J_0(\pi a \rho), \tag{B17}$$

where a = diameter.