Monte Carlo Image Analysis in Radio Interferometry

Urvashi R.V. and T.J. Cornwell

September 24,2004

Abstract

Image analysis, such as component fitting of radio interferometric images has traditionally been based on likelihood techniques applied to deconvolved images. The analysis usually ignores uncertainties arising from fitting components to extended emission as well as from the process of deconvolution itself. One would therefore like to estimate the properties of components representing the entire emission present in the raw, dirty image. In practice, this is not feasible given the large dimensionality of the parameter space. We present an intermediate approach in which a Bayesian image analysis is performed to fit elliptical gaussian components to sub-regions of the dirty image, taking full account of the point spread function. Our method samples the posterior distribution to estimate the relative probabilities and uncertainties associated with the number of components and their parameters. This information can augment the process of object detection and characterization and we compare the performance of this approach to the standard methods.

1 Introduction

Image analysis in radio interferometry often involves component fitting applied to previously deconvolved images. Combinations of gaussian, disc or point source flux component models are used to approximate the emission patterns seen in a deconvolved image. There are two sources of uncertainty associated with this approach which traditional methods generally ignore.

1. Incomplete sampling of the spatial coherence function leads to an invisible distribution which makes deconvolution a non-linear inverse problem with no exact solution. Algorithms that use different approximations can converge to slightly different results[1]. Moreover, the only estimate of error due to deconvolution is the residual image which holds no information about the possible existence of multiple solutions with the same goodness-of-fit.

[1] CLEAN,MEM,MS-CLEAN,ASP-CLEAN are each suited to different emission patterns

2. There is an uncertainty in the process of estimating the parameters of overlapping flux components in extended emission. The objective function involved in fitting a linear combination of elliptical gaussians is often multi-modal and any likelihood technique (unconstrained and constrained minimization) is subject to this uncertainty.

A Monte Carlo approach to this form of image analysis has the potential of providing information about relative probabilities and uncertainties associated with the number of flux components and their parameters by sampling the posterior distribution associated with the Bayesian formulation of deconvolution.

Due to the uncertainty arising from the process of deconvolution, one would like to estimate the properties of components representing the entire emission present in the raw dirty image. In practice however, a pixel based Monte Carlo method is not feasible given the large dimensionality of the parameter space. A scale sensitive representation of an image as a collection of gaussian, disk or point source flux components $I^M = \Sigma_i P_i$ can control this problem. An additional advantage of thus incorporating the flux component model into the image representation is that the deconvolution and the component fitting will then be treated together and the resulting uncertainty estimates will reflect the entire process applied to the raw dirty image.

We present the MC-FIT algorithm, in which an image is represented as a collection of elliptical gaussian flux components, and Monte Carlo sampling is performed on the number of flux components and their parameters. A-priori information is provided in the form of probability distribution functions for each type of parameter and the χ^2 goodness of fit criterion is used as the likelihood function. We used BayeSys[2] as the Monte Carlo sampling engine. Tests on simulated synthesis data as well as real data are presented.

2 Bayesian Inference

2.1 Bayes theorem

Bayes theorem can be used to formulate the problem of deconvolution in terms of a multi-dimensional conditional probability distribution. The posterior distribution of the parameters is given as

$$P(I^M|D,M) \;=\; \frac{P(D|I^M,M)P(I^M|M)P(M)}{P(D)} \;\alpha\; P(D|I^M,M)P(I^M|M) \qquad (1)$$

where $I^M = \Sigma_i P_i$ denotes the model image as a collection of flux components, D denotes the observed data and M denotes a-priori information about the parameters.

[2] BayeSys and Massinf, John Skilling, 2004.

$P(D|I,M)$ is the likelihood distribution, and is given by

$$P(D|I,M) \;=\; e^{-\frac{1}{2}\chi^2} \tag{2}$$

$$\chi^2 \;=\; \Sigma\left(V^{obs} - S.V^M\right)^2 \tag{3}$$

where V^{obs} represents the observed visibilities, S is the visibility sampling function and $V^M \;=\; F.I^M$ represents the Fourier transform of the model image I^M. Maximizing the likelihood function is equivalent to an unconstrained minimization of χ^2.

Equivalently, in the image domain,

$$\chi^2 \;=\; \Sigma\left(I^{D^2} - I^M\left[I^D + I^R\right]\right) \tag{4}$$

where I^D is the dirty image and I^R is the residual image.

$P(I^M|M)$ is the prior distribution which provides a measure of how well the model image I^M conforms to known information about the component parameters and serves to bias the posterior distribution towards a priori information. A well known prior is the entropy prior, which is pixel-based and gives the probability of a model image I^M based on a measure of distance between I^M and an a-priori image I^P. Another kind of prior suitable for a parameterized flux component based image representation involves separate probability distributions for each type of parameter. For instance, position parameters of the flux components could have a uniform distribution within the region of interest, and scale parameters could have a non-uniform distribution reflecting the actual observed distribution of scale in a typical image.

The most probable image can be obtained by maximizing the posterior distribution function. This is equivalent to a constrained minimization of χ^2 in a search space whose dimensionality is given by the total number of component parameters being fitted.

2.2 Monte Carlo Sampling of $P(I|D,M)$: BayeSys

Various techniques exist for efficiently sampling a probability distribution in a multi-dimensional parameter space. Markov Chain Monte Carlo techniques generate samples by following a Markov Chain through the parameter space based on a transition matrix whose limiting distribution is the target posterior distribution.

BayeSys is an application that samples the posterior distribution associated with an object represented as a collection of atoms, an atomic prior, and a suitable goodness of fit criterion. BayeSys implements an MCMC algorithm along with selective annealing and has the following features.

1. A BayeSys object is a mixture model comprised of a collection of flux components called Atoms, each represented by a set of parameters. Prior information can be supplied in the form of distribution functions per parameter.

2. Multiple parallel sample streams, called ensembles are allowed to communicate so that they can catalyse each other's progress and guard against any sequence being trapped in a local maxima.

3. There are several sampling engines used to create, destroy and move the atoms around efficiently by varying their parameters. Generality of the sampling engines is achieved by restricting the parameters to lie within a U[0,1] hypercube. These samples are then transformed according to their individual prior probability distributions to obtain sets of parameters that can be used along with the goodness of fit criterion to evaluate the posterior distribution.

4. Annealing is achieved by initializing the sequence to sample the prior distribution and gradually increasing the influence of the likelihood function. Annealing is usually stopped (and the sequence is said to have converged) when the likelihood and prior are equally weighted and the posterior distribution is being sampled. Further annealing will result in only the likelihood distribution being sampled.

3 MC-FIT Algorithm

We present the MC-FIT algorithm, which combines the two stages of component based image analysis (deconvolution and component fitting) and uses the Bayesian formulation of deconvolution to obtain estimates (with associated uncertainties) for parameters of extended flux components present in the dirty image. The main features of this algorithm are as follows.

1. An image is represented as a sum of elliptical gaussian flux components, each described by six parameters (two for position, two for scale, one for amplitude and one for position angle). The practical advantage of such a representation is three-fold. It allows for a scale sensitive representation of an image, is well suited to component fitting based image analysis, and conforms to the object representation framework used by BayeSys.

2. The MC-FIT algorithm uses component based priors. The distribution functions for the position parameters are chosen to be uniform between the bounds of the image (or region of interest). The distribution functions for the scale parameters are chosen to be non-uniform and of the form Ase^{-Bs} where s represents the scale. This approximates the distribution of scales observed in a typical image, where the majority of features correspond to small scales and a comparitively smaller number of large scale features exist. The distribution for the amplitude is chosen to be uniform within a range, and the position angle (defined here as the angle between the horizontal axis and the first specified axis of the ellipse) is allowed to vary uniformly between 0 and 90 degrees.

3. Sampling on the number of components is achieved by defining a distribution for the number of components (uniform or poisson) within a specified range.

4. Image based priors based on entropy, emptiness (and positivity constraints), can be applied by including them during the evaluation of the posterior distribution function.

5. Component fitting can be restricted to certain regions in the image. Prior to the fitting of emission inside a given region, the emission outside the region is removed from influence by masking out and subtracting from the observed data, visibilities corresponding to parts of a previously deconvolved image. This information is then incorporated into the prior distributions of the position parameters by restricting them to follow the mask.

6. The computation of χ^2 in the evaluation of the likelihood distribution is computed using Equations 3 and 4 for synthesis data involving a point spread function. For fitting components directly to images obtained without the use of a point spread function χ^2 is computed as $\chi^2 = \Sigma I^{R^2}$

7. BayeSys provides the user with control points for the actual sampling algorithm used. Choices about the sampling engine to be used, the number of parallel ensembles to run, and the annealing speed, allow the user to tune the sampler to obtain faster convergence and better sampling of the posterior distribution.

Following are the steps by which MCMC samples are generated and collected.

1. Define suitable priors for each of the parameters.

2. Generate a random sample set of parameters from the U[0,1] hypercube, using one of several sampling engines, and based on the current state of the system.

3. Transform the U[0,1] numbers to physical parameters using transfer functions for each of the parameter prior distributions.

4. Compute a model image I^M.

5. Compute the likelihood and an image based prior term (if present), to obtain the corresponding value of the posterior distribution.

6. Update BayeSys annealing parameters based on the result of Step 5.

7. Repeat steps 2 through 6 until annealing is complete.

8. Fix the annealing parameter and collect a large number of samples to form the desired ensemble reflecting the posterior distribution.

9. Analyse this ensemble to obtain (most probable) values for fitted parameters and their associated uncertainties. For a more accurate best-fit, the most probable parameter set can be used to initialize a regular likelihood maximization scheme (MEM + ASP).

4 Tests on Simulated Synthesis Data

The MC-FIT algorithm was applied to a simulated synthesis data set corresponding to a VLA C array observation of a source composed of four overlapping elliptical gaussian features.

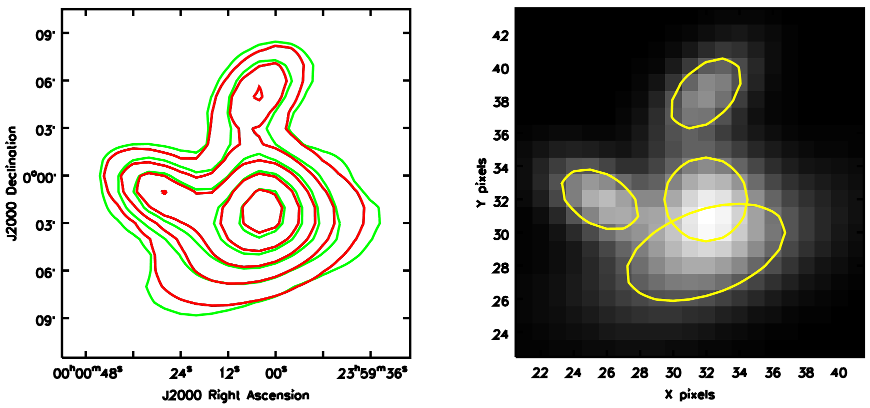

Figure 1: Tests on simulated synthesis data : [Left] original image (red), dirty image (green) with signal to noise ratio = 100. [Right] Most probable model image with 3σ contours for the corresponding set of gaussian components.

Component	x0	y0	Amplitude	σ_x	σ_y	Position Angle
1	5.0	4.5	80	1.0	0.5	20
2	4.0	5.0	100	0.3	0.5	60
3	5.0	6.0	80	0.5	0.3	45
4	5.0	5.0	100	0.5	0.5	45

The true parameters of the elliptical gaussian components used to compute the sample data are given in Table 4. Inspection of Figure 2 shows that all the histograms have peaks at the correct locations.

The value of normalized χ^2 computed using the most probable image obtained via the MC-FIT algorithm, was compared to the values obtained using images from the CLEAN and MEM algorithms, and with images formed from gaussian components fitted to the deconvolved images.

		Normalized χ^2
1	Gaussian fits to CLEAN restored image	2.023
2	Gaussian fits to MEM restored image	1.996
3	MC-FIT mode image	1.122
4	CLEAN model image	1.113
5	MEM model image	1.114

Table 4 lists values of normalized χ^2 for various model images. Rows 1 and 2 were obtained via the traditional method of fitting gaussian components to deconvolved, restored images. Row 3 represents the corresponding value for the image formed from the most probable set of component parameters produced by MC-FIT. Rows 4 and 5 were computed from the CLEAN and MEM un-restored model images and represent the value of χ^2 obtained as a result of the respective minimization algorithms. This data shows that if an image can be decomposed into elliptical gaussian components, the MC-FIT algorithm is capable of producing a component list that represents the image almost as well as the minimum χ^2 CLEAN and MEM model images.

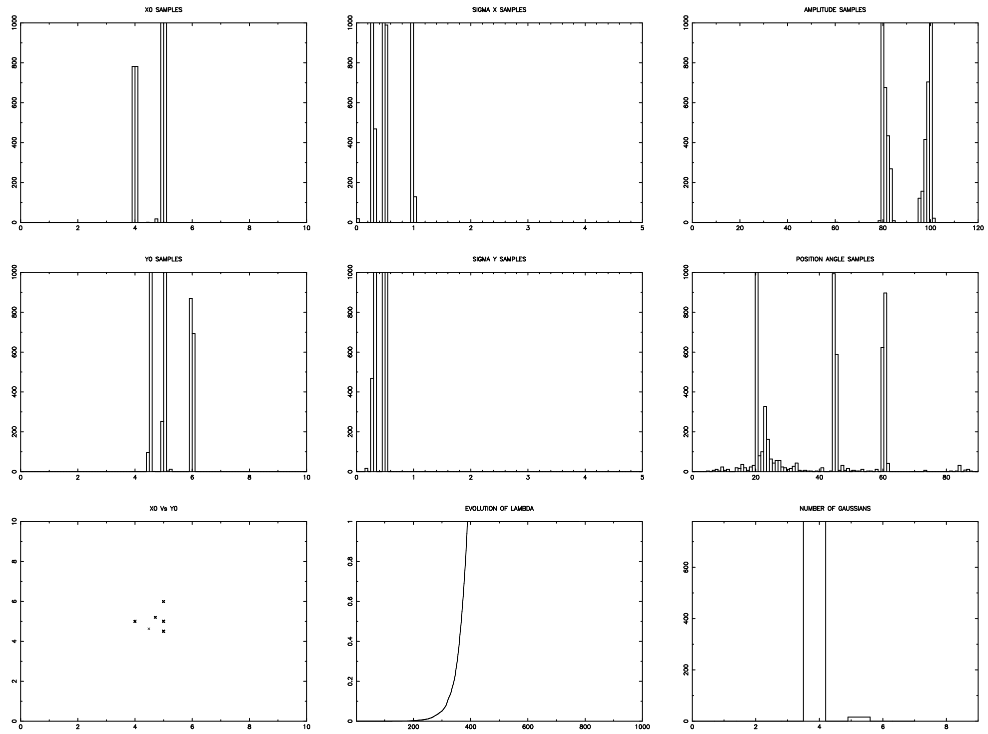

Figure 2: [Rows 1,2] Histograms of the obtained samples for each of the six parameters. [Row 3] (left) Scatter plot of the positions of the centres of the gaussian components, (middle) evolution of the annealing parameter between 0(sampling the prior) and 1(sampling the posterior), as a function of sample iteration number, and (right) histogram of the number of components over all samples.

The effect of the noise level and different priors on the shape of the posterior distribution can be seen using histogram displays similar to Figure 2. Increased noise levels result in a slower convergence via annealing and wider distributions. The use of different priors affect the rate of convergence. In the absense of any significant flux components, the parameter histograms reflect the shape of the prior distributions.

5 Tests on Real Data

5.1 MC-FIT with G192.16-3.84 non-synthesis data

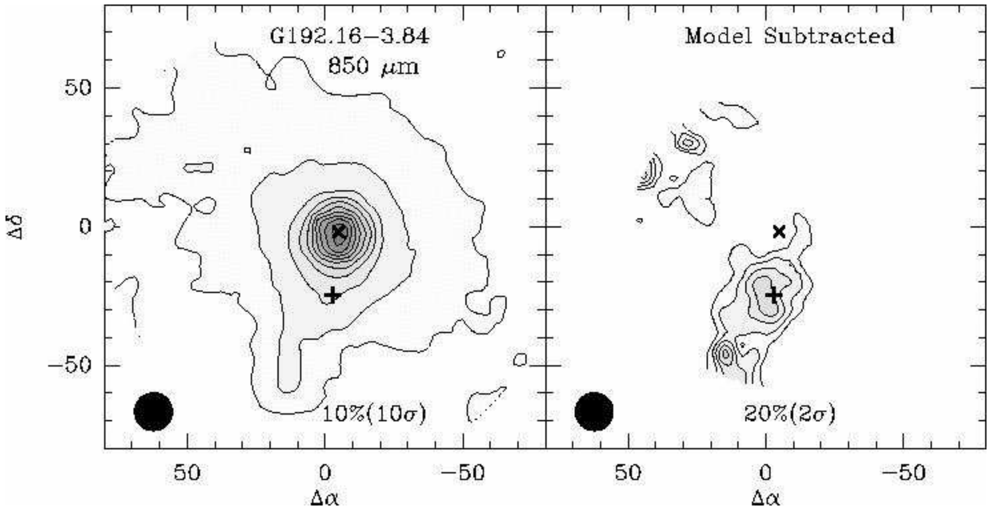

Figure 3: G192.16-3.84 - dust continuum map. (Image credits:Ref)

A test was performed on a dust continuum map of the Early B protostar G192.16-3.84 shown in Figure 3. Evidence from NH_3 data shows the existence of a low-mass protostellar core southwest of the main protostar. The signal to noise ratios of the main peak and the peak of the off-centre core, in the dust continuum map are 7 and 1.4 respectively, making it difficult for an accurate detection. The authors had to model the main core and subtract it out to reveal with more clarity the position of the off-centre core.

To obtain error estimates on the location and shape of the off-centre core, we ran the MC-FIT algorithm on the central 64x64 pixels of this image (Figure 4), and obtained the histograms shown in Figure 5.

The histograms and plots in Figure 5 confirm the existence of a peak at the location of the low-mass protostellar core. A comparison of the heights and widths of the histogram peaks for the position parameters shows respectively the relative uncertainty in the existence of the peaks and a measure of the uncertainty in the actual position of the peaks. An analysis of the scale parameter histograms shows that the off-centre core is compact. The amplitude histogram peaks at values that correspond to the correct peak amplitudes in the image (Figure 4).

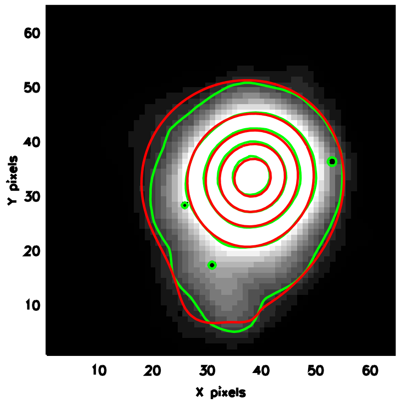

Figure 4: Central 64x64 pixels of the dust-continuum map

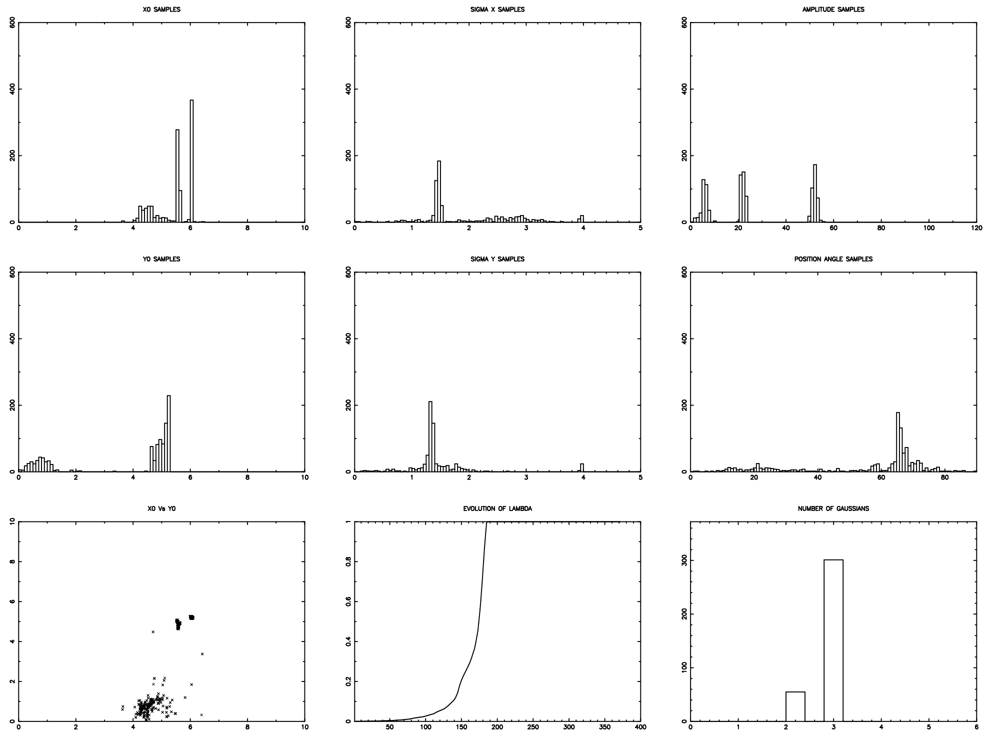

Figure 5: Sample Histograms - G192.16-3.84 dust continuum detection of a low-mass protostellar core Southwest of the main core. A normalized χ^2 of 1.77 was obtained. The histograms for the position parameters show relative probabilities and uncertainties for the two peaks. The central core is slightly elongated and has been modelled by two gaussian components whose total amplitude matches the value at the peak in Figure 4. The scale histograms suggest that the off-centre core is relatively compact.

5.2 MC-FIT with 3C273 synthesis data

A second test was performed on 3C273 synthesis data. Here χ^2 was computed using Equation 4. Figure 6 shows the dirty image, and the unrestored model image corresponding to the parameter set whose probability was the maximum of all the samples. The histograms in Figure 7 show that the algorithm has found the central core and the

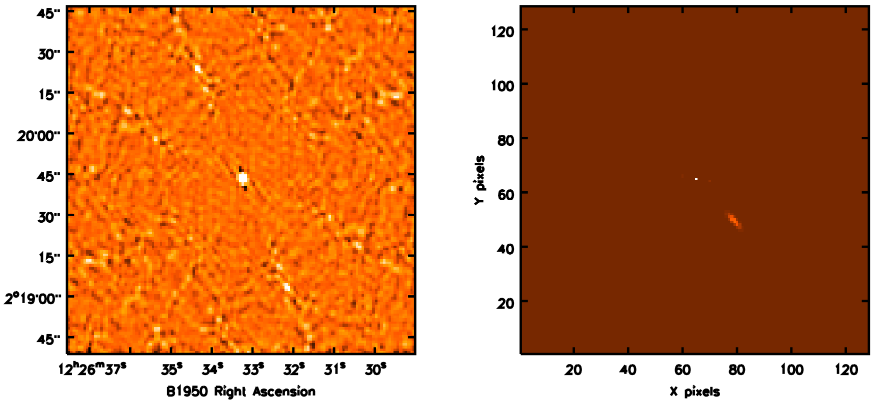

Figure 6: 3C273 : (left) dirty image, (right) MC-FIT mode image

extended radio lobe. Again, relative heights and widths of the parameter histograms give estimates of the uncertainty in the position, amplitude and shape of the gaussian components. The scale histograms shows the existence of a fainter elongated component. The amplitude histogram peaks at 30Jy which is the known peak flux for 3C273. The algorithm was allowed to create samples of upto five components, and it found that two components were sufficient to accurately describe the image.

6 Discussion

A Bayesian approach to component fitting in the analysis of images formed from synthesis and non synthesis data, can thus be used to obtain values of flux component parameters along with the associated uncertainty estimates. The running time of this algorithm is a directly proportional to the number of components being fitted ($O(N)$), the number of ensembles ($O(N)$), and the image size ($O(N^2 log_2 N)$ for an $N \times N$ image). The number of iterations required to reach convergence depends on the type of prior information provided to the algorithm, the signal to noise ratio of the data, the area under the clean beam, and the complexity of the brightness distribution.

Further work would involve the implementation of the MC-FIT algorithm as a usable tool in AIPS++, parallelizing the implementation, and providing additional ways of analysing and interpreting the parameter ensembles.

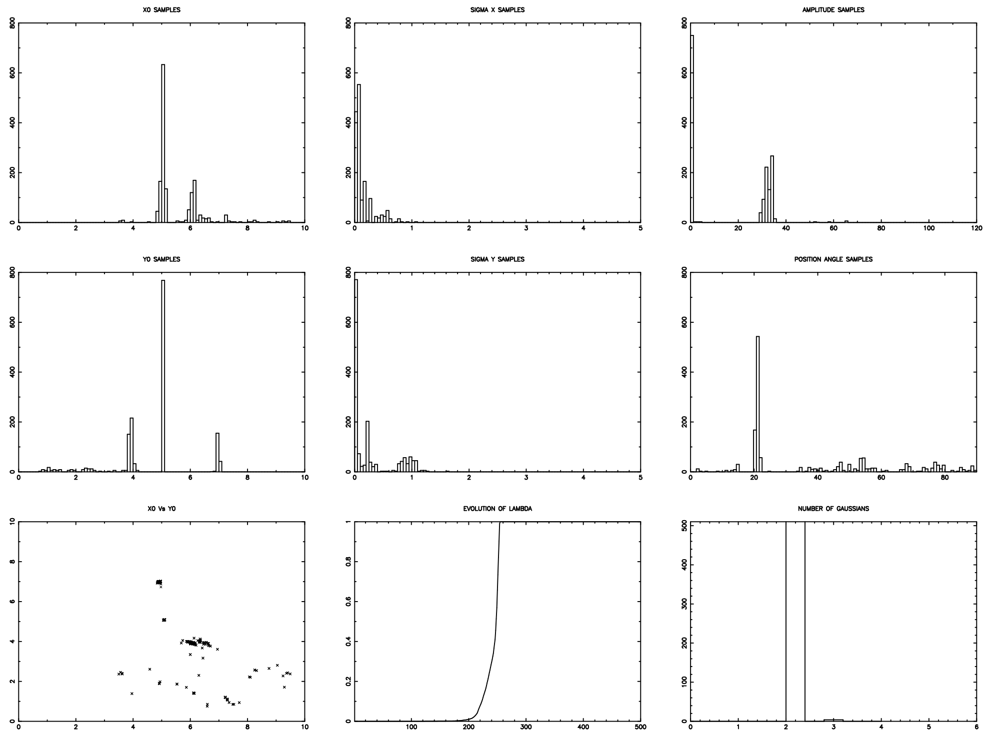

Figure 7: 3C273 : Sample Histograms

7 References

1. *A Bayesian approach to discrete object detection in astronomical data sets*, M. P. Hobson, C. McLachlan, Mon. Not. R. Astron. Soc. 338,765-784, (2003).

2. *BayeSys and Massinf*, John Skilling, Maximum Entropy Data Consultants Ltd(est 1981), February 2004.

3. *Lectures in Synthesis Imaging*,ASP Conference Series, Volume 180, 1999.

4. *Bayesian data analysis*, Gelman et al, CRC Press.

5. *Monte Carlo Methods in Bayesian Computation*, Chen et al.

6. *The circumstellar environment of the Early B Protostar G192.16-3.84 and the Discovery of a Low-Mass Protostellar Core*, D.S.Shepherd et al, Astro-phxxx

7. AIPS++ documentation : http://www.aips2.nrao.edu