FITTING A GAUSSIAN MODEL TO APERTURE SYNTHESIS DATA BY AKAIKE'S INFORMATION CRITERION (AIC)

MASATO ISHIGURO
The Research Institute of Atmospherics, Nagoya University
Toyokawa, 442 Japan

MAKIO ISHIGURO
The Institute of Statistical Mathematics, Tokyo, Japan

ABSTRACT

The problem of determining the optimum number of components in fitting a Gaussian model to aperture synthesis data is considered. As a measure of the badness of the fitted model, we propose the use of Akaike's Information Criterion (AIC).

1. INTRODUCTION

The output of an aperture synthesis telescope is a set of sampled values of the visibility function which is the Fourier transform of the sky brightness distribution. If the visibility function is completely sampled along the coordinate of spatial frequency under the circumstance of high signal-to-noise ratio, the true sky brightness distribution is obtained by a simple Fourier inversion process. In many practical cases, however, measurement of the visibility function is limited by a small number of antennas and by the measurement errors.

When the measured visibilities are incomplete and noisy, the synthesis map estimated through the direct Fourier inversion is disturbed by undesirable sidelobe effects, which make the physical interpretation of the structure very difficult. Various reconstruction techniques to cope with the problem of estimating a reliable brightness distribution from a limited set of visibility data have been developed. Some of these are model fitting (Fomalont, 1968), CLEAN (Högbom, 1974), and Maximum Entropy Method (Gull and Daniell, 1978) and they have been used successfully for some class of brightness distribution suitable for each reconstruction technique.

We have taken a new look at the problem of model fitting. A model brightness distribution is usually represented by a set of simple-shaped

C. van Schooneveld (ed.), Image Formation from Coherence Functions in Astronomy, 277–286.

components. We use a Gaussian model because it has the non-negative property and it is convenient for the Fourier transform. For mathematical simplicity, we considered the one-dimensional case. So far, the goodness of fit has been evaluated in a least square sense, and the number of components has been determined rather arbitrarily and/or empirically. If the number of degrees of freedom is too large compared with that of the independent observations, the results of fitting are satisfactory but are very sensitive to the noise, and *vice versa*.

As a measure of the badness of the fitted model, we propose the use of Akaike's Information Criterion (AIC). We can select the optimum model on the basis of the Minimum AIC Estimate (MAICE). In the following sections, the basic principle of the method is described and some numerical examples are illustrated.

2. STATISTICAL NATURE OF THE OBSERVATION

Let $T^*(x)$ be the true brightness distribution. If the observation noise is additive, N observations of the visibility function are expressed by

$$z_j = \tilde{T}^*(u_j) + n_j \qquad (j=1,2,\ldots,N), \tag{1}$$

where $\tilde{T}^*(u)$ is the Fourier transform of $T^*(x)$, and $\{u_j\}$ are the spatial frequencies at which the $\{z_j\}$ are measured. The $\{n_j\}$ are the noises. If $n^N = (n_1,n_2,\ldots,n_N)$ is a random vector whose probability density function (PDF) is $g^*(n^N)$, the PDF of the random vector $z^N=(z_1,z_2,\ldots,z_N)$ is given by

$$f^*(z^N) = g^*(y^N) \tag{2}$$

where y^N is a N-vector whose j-th element is

$$y_j = z_j - \tilde{T}^*(u_j) \qquad (j=1,2,\ldots,N).$$

3. MODELING

As an estimate of $T^*(x)$, we adopt the Gaussian model:

$$T_M(x|\theta^M) = \sum_{m=1}^{M} w_m \frac{1}{\sqrt{2\pi}\,\sigma_m} \exp\{-\frac{(x-\mu_m)^2}{2\sigma_m^2}\}, \tag{3}$$

where

$$\theta^M = (w_1,\sigma_1,\mu_1,w_2,\sigma_2,\mu_2,\ldots,w_M,\sigma_M,\mu_M)$$

is the parameter vector of the model to be adjusted. Taking the Fourier transform of Equation (3), the visibility function is modeled as

https://doi.org/10.1017/S0252921100074984 Published online by Cambridge University Press

$$\tilde{T}_M(u|\theta^M) = \sum_{m=1}^{M} w_m \exp\{-\tfrac{1}{2}(2\pi u\sigma_m)^2 + i2\pi u\mu_m\}. \quad (4)$$

We also assume that the noise n^N has the PDF:

$$g_N(n^N|\sigma) = \prod_{j=1}^{N} \frac{1}{2\pi\sigma^2} \exp\{-\frac{|n_j|^2}{2\sigma^2}\}. \quad (5)$$

Equation (5) is derived from the assumption that the real and the imaginary parts of $\{n_j\}$ are statistically independent samples from the same normal distribution with zero mean and standard deviation σ. From (2) and (5) we obtain the expression for the joint PDF of z^N as

$$f_M(z^N|\theta^M,\sigma) = \prod_{j=1}^{N} \frac{1}{2\pi\sigma^2} \exp\{-\frac{|z_j - \tilde{T}_M(u_j|\theta^M)|^2}{2\sigma^2}\}. \quad (6)$$

4. MEASURE OF THE BADNESS OF THE FITTED MODEL

If the number of Gaussian components M of the model (3) is specified, the parameters of (6) can be estimated by the maximum likelihood method. Let $(\hat{\theta}^M, {}_M\hat{\sigma})$ be the maximum likelihood estimate (MLE). Then estimates of $T^*(x)$ and $f^*(z^N)$ are given by $T_M(x \mid \hat{\theta}^M)$ and $f_M(z^N \mid \hat{\theta}^M, {}_M\hat{\sigma})$ respectively.

As a measure of the badness of the fitted model, we propose the use of Akaike's Information Criterion (AIC), which is defined by

$$\text{AIC} = -2 \times \log(\text{maximum likelihood of a model}) + 2 \times (\text{number of parameters to be adjusted}). \quad (7)$$

This criterion is introduced by Akaike (1973,1977) as an estimate of minus twice the entropy of the true PDF with respect to the MLE of PDF. In our case, the entropy and AIC are given by

$$B\{f^*(.);f(.\ |\hat{\theta}^M, {}_M\hat{\sigma})\} = \int f^*(z^N) \log \frac{f(z^N \mid \hat{\theta}^M, {}_M\hat{\sigma})}{f^*(z^N)} dz^N, \quad (8)$$

and

$$\text{AIC}(M) = -2\log\{f_M(z^N \mid \hat{\theta}^M, {}_M\hat{\sigma}) + 2(3M+1), \quad (9)$$

https://doi.org/10.1017/S0252921100074984 Published online by Cambridge University Press

respectively. Although it is impossible to maximize (8) without knowing the true PDF $f^*(.)$, (9) can be minimized by a suitable choice of M. Thus AIC provides a criterion for the optimum choice of the number of Gaussian components. The estimated model which gives the minimum of AIC is called the Minimum AIC Estimate (MAICE).

5. NUMERICAL PROCEDURE

To maximize (6) for given z^N, it is sufficient to minimize

$$-\log\{f_M(z^N|\theta^M,\sigma)\} = N\log 2\pi + N\log\sigma^2 + \frac{1}{2\sigma^2}\sum_{j=1}^{N}|z_j-\tilde{T}_M(u_j|\theta^M)|^2. \quad (10)$$

If the function

$$\phi(\theta^M) = \sum_{j=1}^{N}|z_j-\tilde{T}_M(u_j|\theta^M)|^2 \quad (11)$$

is minimized by $\hat{\theta}^M$, the minimum value of (10) is given by

$$-\log\{f_M(z^N|\hat{\theta}^M,{}_M\hat{\sigma})\} = N\log 2\pi + N\log_M\hat{\sigma}^2 + N, \quad (12)$$

where ${}_M\hat{\sigma}^2$ is given by

$${}_M\hat{\sigma}^2 = \frac{1}{2N}\ \phi(\hat{\theta}^M). \quad (13)$$

Now the problem is reduced to the minimization of the non-negative function $\phi(\theta^M)$. Since the minimization of (11) leads to a problem of solving a set of nonlinear equations, it must be solved iteratively, starting with a suitable initial guess.

It is easy to evaluate the function $\phi(\theta^M)$ and its gradient for given θ^M. It is also not difficult to obtain a rough estimate of the Hessian matrix of second partial derivatives. Considering the situation, we propose the use of Davidon's Variance Algorithm for minimization (Davidon, 1968, 1969), which is essentially a modified Newton-Raphson algorithm, where the direct computation of variance, the inverse of the Hessian, is replaced by successive correction of the estimated variance.

6. NUMERICAL EXAMPLES

As a true brightness distribution, three Gaussian components are used. Simulated noisy data are generated by adding zero mean normal random numbers with a standard deviation of 0.1, to the real and imaginary parts of the Fourier transform of the true brightness distribution. In this example, the number of complex visibility data is 50.

We start with equally spaced and identical Gaussians as an initial estimate, the number of which is the same as that of the assumed model. In Table 1., AIC(M), its constituent terms, and $_M\hat{\sigma}$, for one- to six-component models are shown. Constant terms in (12) are omitted.

M	AIC(M)	$-2\log\{f_M(z^N\mid\hat{\theta}^M,{}_M\hat{\sigma})\}$	2(3M+1)	$_M\hat{\sigma}$
1	-359.29	-367.29	8	0.156
2	-393.86	-407.86	14	0.127
3	-421.01*	-441.01	20	0.108
4	-417.23	-443.23	26	0.107
5†	-411.23	-443.23	32	0.107
6	-418.39	-456.39	38	0.100

*MAICE †weight for one of the Gaussians is reduced to zero.

Table 1.

Figure 1. shows the result of MAICE together with that of direct Fourier inversion. The effectiveness of the model fitting by MAICE over the Fourier inversion is very clear. Fitted models for M=2,3,4 and 6 are shown in Figure 2.

We dropped some visibility data to see the applicability of the MAICE procedure when visibility data are incompletely sampled. The results are shown in Figure 3. Numbers of complex visibility data are 38 and 19 for Figures 3.a and 3.b, respectively. The main features of the true brightness distribution are still well reconstructed.

The convergence of the iterative procedure is quadratic, and the computing time for the three component model is less than 6 seconds.

7. CONCLUSION

We have shown by numerical examples that the MAICE procedure is very useful in fitting a Gaussian model to aperture synthesis data. The shapes of the components need not necessarily be Gaussian; other shapes can be used in specific cases. In any case, it is preferable that the component shape has the properties of non-negativity and smoothness. The first and the second terms in the AIC expression represent "badness of fit"

https://doi.org/10.1017/S0252921100074984 Published online by Cambridge University Press

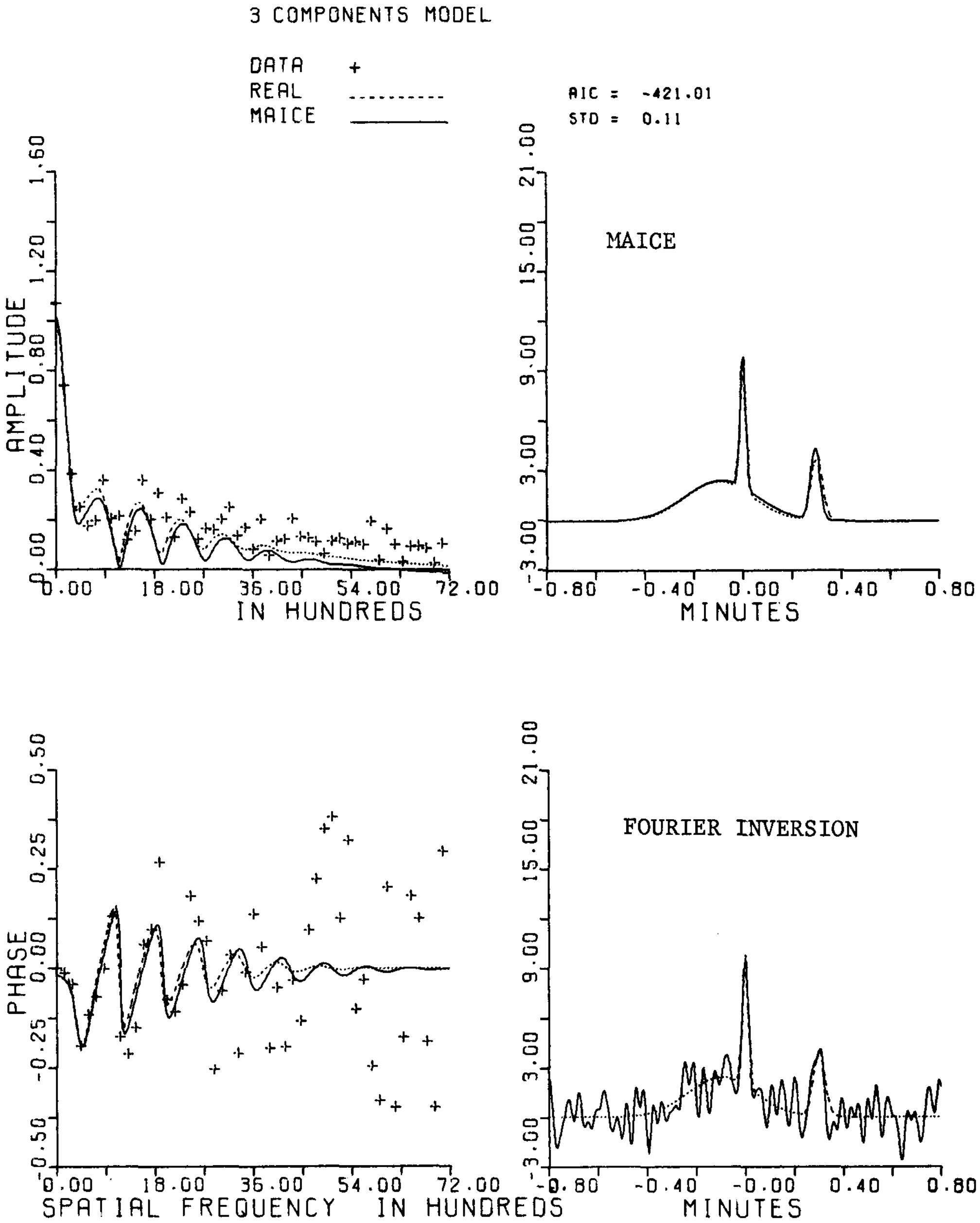

Figure 1. Selected model by MAICE (N=50).

https://doi.org/10.1017/S0252921100074984 Published online by Cambridge University Press

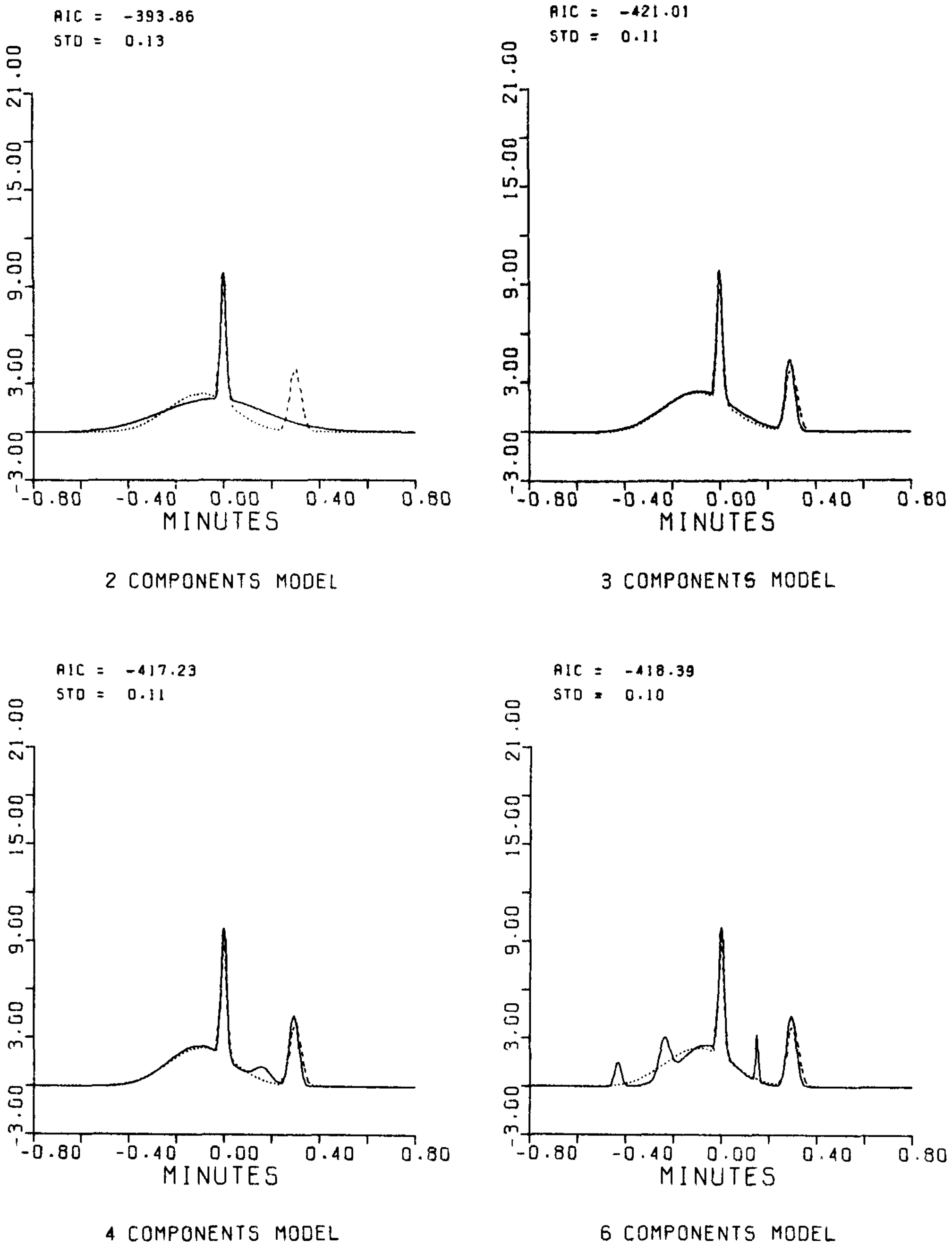

Figure 2. Fitted models for M = 2, 3, 4 and 6. AIC is minimum for M=3.

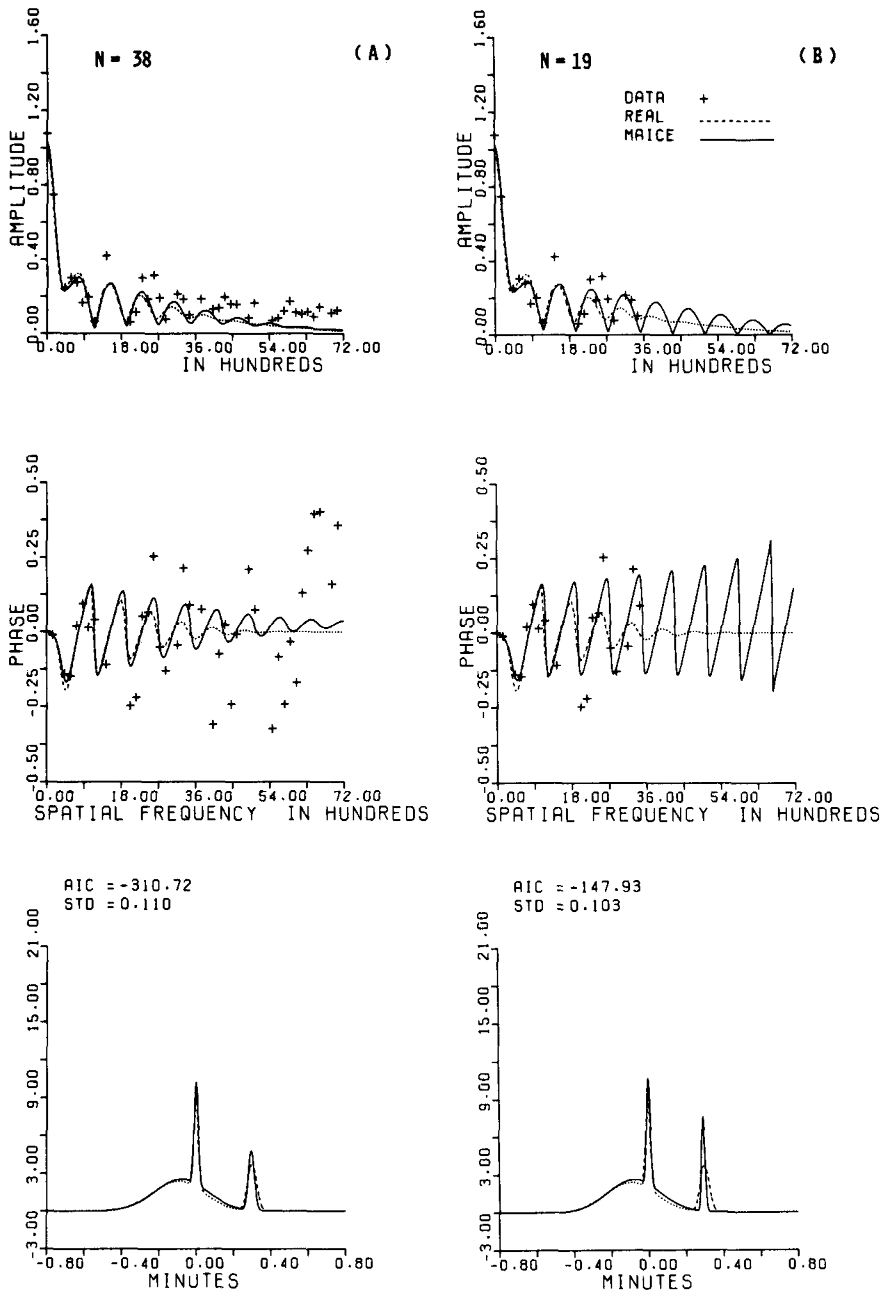

Figure 3. Results of MAICE for incompletely sampled visibility data. Numbers of complex visibilities are; (A) N=38, and (B) N=19.

https://doi.org/10.1017/S0252921100074984 Published online by Cambridge University Press

and "unreliability of model" respectively. The model selected by MAICE is the most parsimonious one. If we have *a priori* information on the brightness distribution, visibility function, and noise statistics, it should be included in the model to reduce the number of degrees of freedom as far as possible. Although we have concentrated our discussion on the model fitting in aperture synthesis, application of the MAICE procedure to other imaging systems is straightforward.

ACKNOWLEDGEMENTS

We would like to thank Dr. H. Akaike of the Institute of Statistical Mathematics for his valuable suggestions and discussions. We are also grateful to Miss. F. Tada for her help in programming.

REFERENCES

Akaike, H.: 1973, *Information Theory and an Extension of the Maximum Likelihood Principle, in 2nd International Symposium on Information Theory* (B.N. Petrov and F. Csaki, eds) pp.207-281, Akademiai Kiado, Budapest.

Akaike, H.: 1977, *On Entropy Maximization Principle, in APPLICATION OF STATISTICS* (P.R. Krishnaiah, ed.) pp.27-41, North-Holland Publishing Company.

Davidon, W.C.: 1968, *Computer J.*, vol.10, pp.406-410.

Davidon, W.C.: 1969, *Optimization* (Fletcher, R. ed.) pp.13-20, Academic Press, London.

Fomalont, E.B.: 1968, *Astrophys. J. Suppl.* 15, pp.203-273.

Högbom, J.A.: 1974, *Astron. & Astrophys. Suppl.* 15, pp.417-426.

Gull, S.F. and Daniell, G.J.: 1978, *Nature*, vol.272, pp.686-690.

DISCUSSION

Comment Y. BIRAUD
How to determine the number of degrees of freedom - or the number of Gaussian components - knowing the noise characteristics?
Reply M. ISHIGURO
We did not do experiments for a wide range of signal to noise ratio. If we know the noise characteristics, the parameters of the noise distribution may be fixed and need not be estimated.

Comment R.H.T. BATES
Is it not dangerous to model the "true" brightness distribution as a sum of Gaussians?

Reply M. ISHIGURO
It is dangerous to use smooth functions to fit true brightness distributions which have sharp boundaries. It is recommended to test several different classes of models. Selection of the optimum model can be done by MAICE.

Comment J.J. WITTELS
It is particularly self-deceptive to create a circular-Gaussian test model for an inversion program that models the data by superposing circular Gaussians. Very few astronomical sources display such a high degree of cylindrical symmetry.
Reply M. ISHIGURO
For a two dimensional example, we used elliptical Gaussians. So the number of parameters is (7M + 1).

Comment J.J. WITTELS
Would you comment on the impression that although, in your example, the model appeared to agree with the calculated curve for the test source, it did not appear to agree very well with the data. This situation is particularly disturbing since agreement with the data (that is both the minimum point to point scatter as well as the absence of systematic disagreement) is the primary criterion for selecting an acceptable model. In this vein, have you tried blind tests of your method?
Reply M. ISHIGURO
The brightness distribution can be modeled by many other suitable functions. We can select most objective ones from a set of possible models by the MAICE procedure.

Comment U.J. SCHWARZ
In relation to the question by Wittels, I would like to point out that the discrepancy between observations and model seems to me - at least for low intensities of the model - to be only apparent. In a display of the modulus of the complex visibility the noise will always give a positive contribution, whereas the model correctly has small or zero amplitude.

https://doi.org/10.1017/S0252921100074984 Published online by Cambridge University Press