Radio Interferometry: Theory, Techniques and Applications,
IAU Coll. 131, ASP Conference Series, Vol. 19, 1991,
T.J. Cornwell and R.A. Perley (eds.)

APPLICATION OF AN ESTIMATOR-FREE INFORMATION CRITERION (*WIC*) TO APERTURE SYNTHESIS IMAGING

M. ISHIGURO
The Institute of Statistical Mathematics, 4-6-7 Minami Azabu, Minato-ku, Tokyo, 106 Japan

K.I. MORITA
Nobeyama Radio Observatory, National Astronomical observatory, Nobeyama, MinamiSaku Nagano, 384-13

M. ISHIGURO
Nobeyama Radio Observatory, National Astronomical observatory, Nobeyama, Minami Maki-mura, Minami Saku-gun, Nagano-ken, 384-13 Japan

ABSTRACT A statistical criterion for stopping *CLEAN* procedure is proposed. The criterion is called *WIC* and an estimator of Kullback-Leibler information quantity which is a measure of the goodness of statistical models. A numerical example is given.

INTRODUCTION

Visibility Data

Visibility z_j observed at time j can be expressed in terms of "true visibility" defined by the brightness distribution $T(x, y)$ and observation noise ε_j:

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

$$\tilde{T}(u, v) = \int\int \exp\{2\pi i(xu + yv)\} T(x, y) dx dy, \tag{2}$$

where$\{(u_j, v_j)\}$ is a set of points in the u-v plane. If ε_j can be regarded as a random variable, then the vector $z = (z_1, z_2, \ldots, z_N)$ is an N dimensional complex random variable. Let $f(z)$ denote the probability density function (PDF) of z.

Statistical Modeling

Assume that a model of T(x, y) is given by

https://doi.org/10.1017/S0252921100013403 Published online by Cambridge University Press

$$T_M(x,y) = \sum_{m=1}^{M} w_m K(x - x_m, y - y_m), \tag{3}$$

where $K(x,y)$ is a suitably fixed kernel function.

Substituting this equation for $T(x,y)$ in Eq. (1), and assuming $\{\varepsilon_j\}$ be a mutually independent complex normal random numbers with mean 0 and variance σ^2, a statistical model of data z is defined which is parametrized by $(w_1,x_1,y_1),(w_2,x_2,y_2),\ldots,(w_M,x_M,y_M)$ and σ^2. Let us denote this model by $f(z|(w_1,x_1,y_1),(w_2,x_2,y_2),\ldots,(w_M,x_M,y_M),\sigma_2)$.

Image formation

When an estimate $\{(\hat{w}_1,\hat{x}_1,\hat{y}_1),(\hat{w}_2,\hat{x}_2,\hat{y}_2),\ldots,(\hat{w}_M,\hat{x}_M,\hat{y}_M)\}$ is given, it is natural to estimate σ^2 by

$$\hat{\sigma}^2 = \frac{1}{2N}\sum_{j=1}^{N} |z_j - \tilde{T}_M(u_j,v_j)|^2, \tag{4}$$

where $\tilde{T}_M(u,v)$ is defined by substituting estimate $T_M(x,y)$ in Eq.(2). The model of data z, $f(z|(\hat{w}_1,\hat{x}_1,\hat{y}_1),(\hat{w}_2,\hat{x}_2,\hat{y}_2),\ldots,(\hat{w}_M,\hat{x}_M,\hat{y}_M),\hat{\sigma}^2)$ should be an estimate of the true PDF $f(z)$.

INFORMATION CRITERION

The goodness of statistical models is measured by the expected log likelihood defined by

$$ELL = \int f(z)\log f(z|(\hat{w}_1,\hat{x}_1,\hat{y}_1),(\hat{w}_2,\hat{x}_2,\hat{y}_2),\ldots,(\hat{w}_M,\hat{x}_M,\hat{y}_M),\hat{\sigma}^2)dz. \tag{5}$$

A larger value of ELL means a better fit of a model.

Ishiguro & Ishiguro (1980) successfully applied the ELL minimization procedure to the aperture synthesis data analysis. They employed the maximum likelihood method to the model fitting, and used AIC(Akaike, 1973) as the estimator of ELL. The example was the one-dimensional imaging problem. The extension of the proposed method, however, to the two-dimensional imaging was difficult. Two dimensionlal models are too complex to be fitted by the maximum likelihood method and it is a vital condition for the use of AIC that the parameter estimation is done by the maximum likelihood method.

WIC

We propose the use of WIC, an estimator-free information criterion, when the model fitting is not done by the maximum likelihood method. WIC is an Extension of AIC which is defined by

$$\begin{aligned} WIC = & -2 \times \log f(z|(\hat{w}_1, \hat{x}_1, \hat{y}_1), (\hat{w}_2, \hat{x}_2, \hat{y}_2), \ldots, (\hat{w}_M, \hat{x}_M, \hat{y}_M), \hat{\sigma}^2) \\ & + 2 \times E_{z^*} \{ \log f(z^* |(\hat{w}_1^*, \hat{x}_1^*, \hat{y}_1^*), (\hat{w}_2^*, \hat{x}_2^*, \hat{y}_2^*), \ldots, (\hat{w}_M^*, \hat{x}_M^*, \hat{y}_M^*), \hat{\sigma}^{*2}) \\ & - \log f(z|(\hat{w}_1^*, \hat{x}_1^*, \hat{y}_1^*), (\hat{w}_2^*, \hat{x}_2^*, \hat{y}_2^*), \ldots, (\hat{w}_M^*, \hat{x}_M^*, \hat{y}_M^*), \hat{\sigma}^{*2}) \} \end{aligned} \quad (6)$$

where z^* is a simulated data obtained by the resampling technique $\{(\hat{w}_1^*, \hat{x}_1^*, \hat{y}_1^*), (\hat{w}_2^*, \hat{x}_2^*, \hat{y}_2^*), \ldots, (\hat{w}_M^*, \hat{x}_M^*, \hat{y}_M^*)\}$ and $\hat{\sigma}^{*2}$ are estimates of parameters based on the data z^*. E_{z^*} denotes the expectation with respect to the variation of $z*$, which is practically computed by the Monte Carlo method.

A NUMERICAL EXAMPLE

We applied the WIC minization procedure for the choice of the number of $CLEAN$ component.

A set of data was produced, asuming the structure shown in Fig. 1a. The $CLEAN$ process proceeds as shown in Figs.1d through 1f. The WIC values indicate that we should stop around $M = 600$. ELL values which could be computed only when the data is artificially generated supports this choice. Visual inspection also endoses this choice.

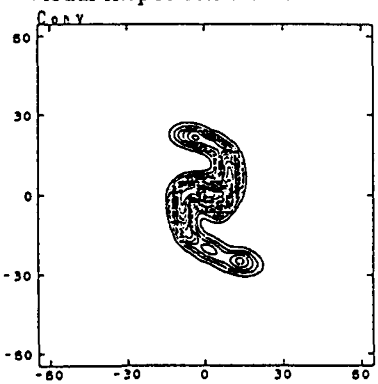

Fig. 1a. Source Model.

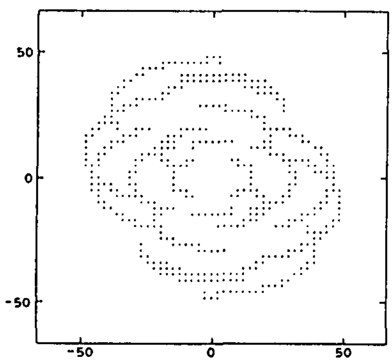

Fig. 1b. $U - V$ plane.

https://doi.org/10.1017/S0252921100013403 Published online by Cambridge University Press

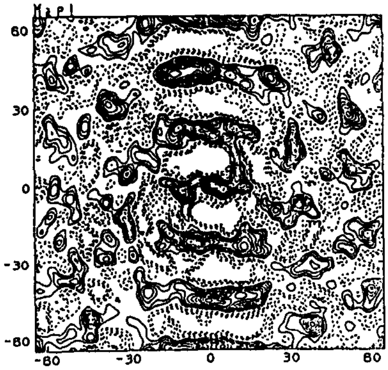

Fig. 1c. Synthesized Map.

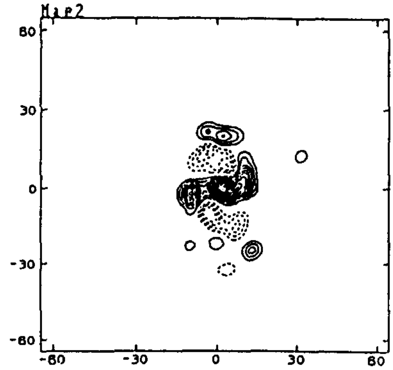

Fig. 1d. Clean Map (M=200).

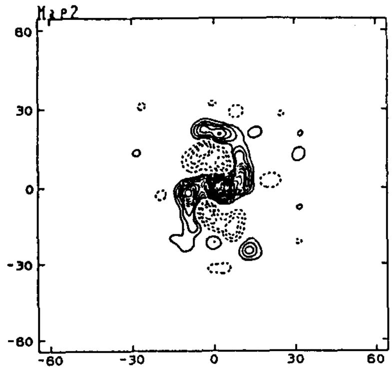

Fig. 1e. Clean Map (M=600).

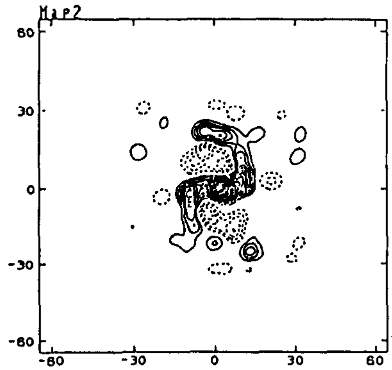

Fig. 1f. Clean Map (M=1000).

TABLE II Analysis of complex data

M	WIC	$-2ELL$
100	29647.9	29643.5
200	29424.2	29411.0
300	29334.6	29312.8
400	29298.2	29269.0
500	29284.2	29248.1
600	29280.6	29238.2
700	29281.6	29235.0
800	29284.3	29235.3
900	29287.5	29235.2
1000	29291.1	29237.2

https://doi.org/10.1017/S0252921100013403 Published online by Cambridge University Press

ACKNOWLEDGEMENTS

This work is partly supported by the Grant-in-Aid for Scientific Research provided by The Ministry of Education, Science and Culture, and carried out under the ISM Cooperative Research Program(90-ISM·CRP-45) of the Institute of Statistical Mathematics.

REFERENCES

Akaike, H. 1973, in *Proc. 2nd Inter. Symp. on Information Theory* ed B. N. Petyov and F. Csai, Akademiai Kiado, Budapest, pp.267-281.
Efron, B. 1979, *The Annals of Statistics*, **7**, 1, pp.1-26.
Sakamoto, Y., Ishiguro, M. & Kitagawa, G. 1986, *Akaike Information Criterion Statistics*, D.Reidel Publishing Company, Dordrecht/Tokyo.
Schwartz, U. J. 1977, *Main Journal.*
Wong, W. H. 1983, *JASA*, **78**, 382, pp.461-463.

APPENDIX: A SHORT DICTIONARY OF STATISTICAL TERMS

AIC: An estimate of Information Criterion proposed by Akaike. It is defined by

$$AIC = -2\log max_\theta f(x \mid \theta) + 2\{number\ of\ free\ parameters\},$$

where f is a statistical model of data x with parameter θ. It is a measure of the badness of the statistical models.

Expected log likelihood: Let $f(x)$ and $f(x|\theta)$ denotes the true *PDF* of data x and its model, respectively. Then the expected log likelihood is defined by

$$\int f(x)\log f(x|\theta)dx.$$

The larger value of the expected log likelihood means a better fit of the model to the real *PDF*.

Log likelihood: Let $f(x|\theta)$ be a statistical model. If x is fixed at the value of given data, f can be regarded as the function of parameter θ only. This function is called *likelihood function* of θ. *Log likelihood function* is the logarithm of likelihood function.

MLE: Maximum likelihood estimate. The estimated parameter of a statistical model obtained by maximizing the (log) likelihood function.

Resampling: When we have a set of data, and need to generate similar set of data, there are two possible way. One is to fit a *PDF* model to the data and generate simulation data using this fitted *PDF*. The other way is the "resampling". Suppose that we have the set of data

$$\{d_1, d_2, \dots d_N\}$$

then the resampled data is obtained by randomly chosing N values from this set of data replacing the chosen value every time. Formal expression of the resampled data could be

$$\{d_{j(1)}, d_{j(2)}, \dots d_{j(N)}\},$$

where $\{j(1), j(2), \dots, j(N)\}$ is a sequence of independent random numbers uniformly distributed on the interval $[1, N]$.

Statistical Model: Usually a parametrized family of probability (density) function.