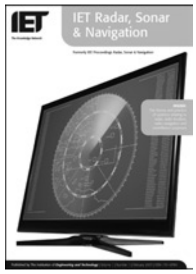
Published in IET Radar, Sonar and Navigation
Received on 7th March 2007
Revised on 19th October 2007
doi: 10.1049/iet-rsn:20070102



ISSN 1751-8784

# A comparative study of model selection criteria for the number of signals

*P. Chen*[1] *T.-J. Wu*[2] *J. Yang*[3]

[1]*Department of Mathematics, Syracuse University, Syracuse, NY 13244, USA*
[2]*Department of Statistics, National Cheng-Kung University, Tainan, Taiwan 70101, Republic of China*
[3]*Department of Electrical Engineering, Syracuse University, Syracuse, NY 13244, USA*
*E-mail: tjwu@stat.ncku.edu.tw*

**Abstract:** The performance of six existing model selection criteria is compared, which are commonly used in time series and regression analysis, when they are applied to the problem of the number of signals in the multiple signal classification (MUSIC) method. The five criteria are Akaike Information Criterion (AIC), Hannan and Quinn Criterion, Bayesian Information Criterion (BIC), Corrected AIC (AICc) and the recently introduced Vector Corrected Kullback Information Criterion (KICvc) and Weighted-Average Information Criterion (WIC). The general form of the above information criteria consists of a log likelihood function expressed in terms of the eigenvalues of the sample covariance matrix and a unique penalty term. In our estimation procedure, the number of signals is obtained by minimising each of the above criteria. Several simulated data sets, including a linear antenna array data set, are adopted for the comparison purpose. The authors show that, in simple MUSIC additive white noise model, for small sample size *n*, WIC performs nearly as well as AICc and outperforms other criteria, and for moderately large to large *n*, WIC performs nearly as well as BIC and outperforms other criteria. Therefore when the authors are not certain of the relative sample size, WIC may be a practical alternative to any criterion. The main purpose is to draw the attention and interests of signal processing researchers to adopt more recent statistical model selection criteria, such as WIC, in general signal processing problems.

## 1 Introduction

An accurate estimate of the number of signals is essential in many signal processing problems such as in the direction of arrival (DOA) estimation by a smart antenna system, in the poles retrieval of a system response and in image processing. The observed or measured data can be modelled as the superposition of a finite number of signals with an additive noise. A widely used estimation method for such problem is the multiple signal classification (MUSIC) [1]. MUSIC is a signal subspace method, which covers the techniques for multiple source localisation by using the eigen-structure of the covariance matrix of the measured data. Under the assumption that the incoming signals and the noise are uncorrelated, the number of signals is equal to the cardinality of the orthogonal basis for the subspace spanned by the signal vectors in the presence of noise, that is, the corresponding number of dominated eigenvalues of the covariance matrix of the measured data. Therefore the number of signals can be determined by identifying the number of dominated eigenvalues of the covariance matrix of the measured data above a threshold.

In the MUSIC method, the observed data $\boldsymbol{x}(t_1)$, $\boldsymbol{x}(t_2)$ , . . . , $\boldsymbol{x}(t_n)$ are $n$ independent snapshots of the following $p$-dimensional process ($p$ stands for the number of sensors)

$$\boldsymbol{X}(t) = \sum_{i=1}^{q} \boldsymbol{A}(\theta_i)s_i(t) + \boldsymbol{n}(t) \tag{1}$$

thus, at each time instance t, $\boldsymbol{X}(t)$ is a $p \times 1$ vector and is a superposition of multiple signals and a random noise, where $\boldsymbol{A}(\theta_i)$ is a $p \times 1$ complex vector that is independent of $t$ and whose functional form is known, but the scalar parameter $\theta_i$ associated with the $i$th signal is unknown and normally needs to be estimated (e.g the DOA of each radar signal), $s_i(t)$ is a known complex wavefront of the $i$th signal, $\boldsymbol{n}(t) = [n_1(t)\ n_2(t)\ \cdots\ n_p(t)]^{\mathrm{T}}$ is a $p \times 1$ additive noise vector and $q$ is the number of signals we want to estimate. There are some assumptions in MUSIC. First, the signals $s_1(t), s_2(t), \ldots, s_q(t)$ are stationary, ergodic Gaussian processes with a positive definite covariance matrix $\boldsymbol{S}$. Secondly, the noise process $\boldsymbol{n}(t)$ is independent of the signals and is a complex, stationary, ergodic Gaussian process with zero mean and covariance matrix $\sigma^2\boldsymbol{I}$. Here $\sigma^2$ is a measure of the noise power and $\boldsymbol{I}$ is the $p \times p$ identity matrix. Equation (1) can be re-written in matrix form

$$\boldsymbol{X}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{n}(t) \tag{2}$$

where $\boldsymbol{s}(t) = [s_1(t)\ \ s_2(t)\ \ \cdots\ \ s_q(t)]^{\mathrm{T}}$ and $\boldsymbol{A} = [\boldsymbol{A}(\theta_1)\ \boldsymbol{A}(\theta_2)\cdots\boldsymbol{A}(\theta_q)]$ is a $p \times q$ matrix parameterised by $\theta_1, \theta_2, \ldots, \theta_q$.

Under the basic assumptions, the $p \times p$ covariance matrix $\boldsymbol{R}$ of $\mathbf{x}(t)$ is given by

$$\boldsymbol{R} = \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^{\mathrm{T}} + \sigma^2\boldsymbol{I} \tag{3}$$

where the superscript 'T' denotes the complex conjugate transpose. When the number of wavefronts $q$ is less than the number of array elements $p$, the $p \times p$ matrix $\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^{\mathrm{T}}$ is singular and has a rank $q$, which can be determined by the eigenvalues of $\boldsymbol{R}$. In the complete set of eigenvalues of $\boldsymbol{R}$, the smallest eigenvalue will occur $(p - q)$ times. The $q$ dimensional space spanned by the $q$ signals, which correspond to the $q$ largest eigenvalues of the covariance matrix, is referred as the signal subspace and the $(p - q)$ dimensional subspace spanned by the $(p - q)$ noise eigenvectors as the noise subspace. In practice, one can expect that the multiple smallest eigenvalues of the sample covariance matrix appear in a cluster rather than being all equal because of random errors. The 'spread' on this cluster decreases as more data is processed.

Bartlett [2] and Lawley [3] proposed a hypothesis testing procedure to test a sequence of hypothesis $H_q$ by likelihood ratio statistics with pre-determined thresholds. The hypothesis $H_q$ is defined as

$$H_q\colon\ \lambda_{q+1} = \lambda_{q+2} = \cdots = \lambda_p \tag{4}$$

where $\lambda_{q+1}, \lambda_{q+2}, \ldots, \lambda_p$ are $p - q$ smallest eigenvalues of the covariance matrix of the measured data. The likelihood ratio statistics are functions of $q$, the number of signals. The $q$ value for which the hypothesis $H_q$ is first accepted is chosen to be the estimate of the number of signals. But the actual distribution of the test statistics can only be approximated and a choice of the level of significance is required. Recently, Chen [4] used selection theory and the ratio of individual eigenvalues to the smallest eigenvalue of the sample covariance matrix as the selection statistics to determine the confidence limits for $q$.

Instead of hypothesis testing and interval estimation, Wax and Kailath [5] proposed an approach based on two model selection criteria, Akaike information criterion (AIC) [6] and minimum description length (MDL) [7, 8]. The number of signals for both criteria is determined by simply minimising the value of an information criterion. Since then, more criteria have been proposed for model selection in time series and regression analysis. Three such criteria are Corrected AIC (AICc) [9], Bayesian information criterion (BIC) [7] and weighted-average information criterion (WIC) [10], a weighted average between AICc and BIC. The AICc part of the WIC criterion receives most of the weight for small sample size $n$, and the BIC part receives the most weight for large $n$. At small and large $n$, WIC shows characteristics of AICc and BIC, respectively. At intermediate $n$, WIC performs better than the worst of the two criteria, AICc or BIC, and nearly as well as the best of these two. Therefore WIC is a very robust and stable criterion. (See Section 2.5 herein and [10] for more details.) All the aforementioned criteria are existing benchmarks commonly used in model selection problems in time series and regression analysis. The purpose of this paper is to make comparison of those criteria and provide recommendation as we apply them to the estimation problem of the number of signals. In Section 2, we briefly review those model selection criteria. Section 3 gives a description to our estimation procedure. Numerical comparisons are given through examples in Section 4. Concluding remarks are in Section 5.

## 2 Model selection criteria

Given the observed data and a set of candidate probability models that depend on the parameter vector $\boldsymbol{\Theta}$, a model selection criterion selects the model by minimising a loss function and can be expressed in the following form [11]

$$\text{A model selection criteria} = -2\log L(\boldsymbol{X}|\hat{\boldsymbol{\Theta}}) + P(n, m) \tag{5}$$

where $L(\boldsymbol{X}|\boldsymbol{\Theta})$ is the likelihood function, $\hat{\boldsymbol{\Theta}}$ the maximum likelihood estimate of $\boldsymbol{\Theta}$, $n$ the sample size and $m$ the number of freely-adjusted parameters



doi: 10.1049/iet-rsn:20070102

(degree of freedom) in $\boldsymbol{\Theta}$, which is a function of both $p$ and $q$. The first term is a measure of the goodness-of-fit of the candidate model to the data. The second term serves as a penalty term for over fitting. In the following, we briefly review the criteria used in our study.

### 2.1 Akaike information criterion

When the penalty term in (1) is set to be $2m$, we obtain the famous AIC, which can be expressed as follows

$$\mathrm{AIC} = -2\log L(\boldsymbol{X}|\hat{\boldsymbol{\Theta}}) + 2m \qquad (6)$$

In time series analysis, AIC is asymptotically efficient in the sense that it minimises the one-step-ahead mean squared prediction error. However, the main problem of AIC is that it is not consistent. 'Consistency' means that when the sample size increases to infinity, the estimated number of signals converges to the true number of signals in probability. It has been shown in [7] that AIC tends to overestimate the number of signals when the sample size is large.

### 2.2 Bayesian information criterion

In order to overcome the overestimation problem of AIC, Schwartz [7] and Rissanen [8] proposed a criterion from different viewpoints. Schwartz's approach was based on Bayesian argument, which selects the model with the maximum posterior probability for a given prior probability. Rissanen's approach was based on the information theory which selects the model with the minimum code length. When the sample size is large, both approaches yield the same criterion that can be expressed as

$$\mathrm{BIC} = -2\log L(\boldsymbol{X}|\hat{\boldsymbol{\Theta}}) + m\log n \qquad (7)$$

It is also referred to as MDL, the minimum description length criterion. Comparing the difference between (6) and (7), we can see that the coefficient for $m$ changes from 2 in AIC to log $n$ in BIC, so the estimated number of signals will be smaller by using BIC than by using AIC. It is known that, unlike AIC, BIC is consistent.

### 2.3 Hannan and Quinn criterion

Another choice for the penalty term, proposed by Hannan and Quinn [12], is $2m\log(\log n)$. The Hannan and Quinn criterion (HQ) criterion can be expressed as

$$\mathrm{HQ} = -2\log L(\boldsymbol{X}|\hat{\boldsymbol{\Theta}}) + 2m\log(\log n) \qquad (8)$$

HQ is also consistent.

### 2.4 Corrected AIC

Hurvich and Tsai [9] proposed another information criterion, which is called 'corrected' AIC, to correct the over-fitting nature of AIC. AICc can be written as

$$\mathrm{AICc} = -2\log L(\boldsymbol{X}|\hat{\boldsymbol{\Theta}}) + 2nm/(n-m-1) \qquad (9)$$

When $n$ increases to infinity, the penalty term goes to $2m$. Therefore AICc converges to AIC. This implies that AICc is asymptotically efficient, but not consistent. When $m/n$ is large, AICc has a greater penalty term than AIC and therefore AICc can correct the over-fitting nature of AIC.

### 2.5 Vector corrected Kullback information criterion

Another recently introduced model selection criterion proposed by Seghouane [13] is Vector corrected Kullback information criterion (KICvc), a derivative of original KIC [14]. KICvc can be written as

$$\mathrm{KICvc} = -2\log L(\boldsymbol{X}|\hat{\boldsymbol{\Theta}}) + \frac{np[2q+p+1]}{n-p-q-1} + \frac{np}{n-q-(p-1)/2} + \frac{2pq+p^2-p}{2} \qquad (10)$$

### 2.6 Weighted-average information criterion

Previous studies [10, 15] show that when the rtio $m/n$ is large, AICc tends to outperform other criteria; while BIC tends to give the best estimation when $m/n$ is small. Inspired by this result, Wu and Sepulveda [10] proposed a model selection criterion WIC, which is a weighted average of AICc and BIC. The weights attached to AICc and BIC are proportional to the model order complexity terms of AICc and BIC, respectively. Thus, the AICc part of the criterion receives most of the weight for small $n$, and the BIC part of the criterion receives most of the weight for large $n$. More specifically, WIC is expressed in the following form

$$\begin{aligned}\mathrm{WIC} &= [A/(A+B)]\mathrm{AICc} + [B/(A+B)]\mathrm{BIC} \\ &= -2\log L(\boldsymbol{X}|\hat{\boldsymbol{\Theta}}) + W \qquad (11)\end{aligned}$$

where $A = 2nm/(n-m-1)$ is the penalty term in AICc and $B = m\log n$ is the penalty term in BIC.

The penalty term $W$ in WIC can be written as

$$W = (A^2 + B^2)/(A + B) = \frac{[2\,nm/(n-m-1)]^2 + (m \log n)^2}{2\,nm/(n-m-1) + m \log n} \quad (12)$$

From (12) we see that as $n \to \infty$, $A/(A+B) \to 0$ and $B/(A+B) \to 1$. Hence WIC is asymptotically equivalent to BIC as $n \to \infty$ and, consequently, WIC is consistent. On the other hand, as $n \to \mathrm{m}+1$, $A/(A+B) \to 1$ and $B/(A+B) \to 0$. Hence WIC is asymptotically equivalent to AICc as $n \to \mathrm{m}+1$. Wu and Sepulveda also proved that in the class of penalty terms $W_r = (A^{r+1} + B^{r+1}A^r + B^r)$, $r = 1$ is the optimal value in the sense of minimising the overall discrepancy $d(r)$ of $W_r$ from $A$ and $B$, where

$$d(r) = (A - W_r)^2 + (B - W_r)^2 = [\mathrm{AICC} - \mathrm{WIC}(r)]^2 + [\mathrm{BIC} - \mathrm{WIC}(r)]^2, \quad r \geq 1 \quad (13)$$

and

$$\mathrm{WIC}(r) = -2 \log L(\boldsymbol{X}|\hat{\boldsymbol{\Theta}}) + W_r \quad (14)$$

## 3 Estimation procedure

In the MUSIC method described in (1)–(3), the observed $n$ independent measurements are complex Gaussian distributed, and [5] the maximum of the log likelihood function $\log \hat{L}q$ (which was previously denoted by $\log L(\boldsymbol{X}\,|\hat{\boldsymbol{\Theta}})$, see Section 2) can be written as a function of the eigenvalues of the sample covariance matrix (which is an estimate of $\boldsymbol{R}$ in (3))

$$\log \hat{L}q = n\left\{\sum_{i=q+1}^{p} \log l_i - (p-q) \log\left[\sum_{i=q+1}^{p} \frac{l_i}{(p-q)}\right]\right\}$$
$$q = 0, 1, 2, \ldots, p-1 \quad (15)$$

where $l_i$ ($i = 1, \ldots, p$) is the $i$th largest eigenvalue of the sample covariance matrix. The number of freely-adjusted parameters is $m(q, p)$. That is, the degree of freedom of the space spanned by the signal vectors, is given by

$$m(q,p) = q(2p-q) + 1 \quad q = 0, 1, 2, \ldots, p-1 \quad (16)$$

In view of (14), (15) and (6)–(10), we obtain the following equations

$$\mathrm{AIC}(q) = -2 \log \hat{L}q + 2[q(2p-q)+1] \quad (17)$$

$$\mathrm{BIC}(q) = -2 \log \hat{L}q + [q(2p-q)+1] \log n \quad (18)$$

$$\mathrm{HQ}(q) = -2 \log \hat{L}q + 2[q(2p-q)+1] \log(\log n) \quad (19)$$

$$\mathrm{AICc}(q) = -2 \log \hat{L}q + \frac{2n[q(2p-q)+1]}{n-p-q-1} \quad (20)$$

$$\mathrm{KICvc}(q) = -2 \log \hat{L}q + \frac{np[2q+p+1]}{n-p-q-1} + \frac{np}{n-q-(p-1)/2} + \frac{2pq+p^2-p}{2} \quad (21)$$

$$\mathrm{WIC}(q) = -2 \log \hat{L}q + \frac{[2n(q(2p-q)+1)]^2 + [(n-q(2p-q)-2)(q(2p-q)+1)\log n]^2}{2n[(q(2p-q)+1)(n-q(2p-q)-2)] + [n-q(2p-q)-2]^2[q(2p-q)+1]\log n} \quad (22)$$

The procedure for estimating the number of signals is summarised as follows:

1. Collect the data vectors $\boldsymbol{x}(t_1), \boldsymbol{x}(t_2), \ldots, \boldsymbol{x}(t_n)$.

2. Compute the sample covariance matrix.

3. Order the eigenvalues $l_1 > l_2 > , \ldots, > l_p$ of the sample covariance matrix.

4. Apply (17)–(22) to compute the model selection criteria.

5. Take the minimiser, denoted by $\hat{q}$ of a criterion (over the range $q = 0, 1, 2, \ldots, p-1$) as the estimate of the number of signals $q$ based on that criterion.

An advantage of this procedure is that no statistical hypothesis testing is involved. Therefore no subjective significance level is required in determining the number of signals.

## 4 Numerical examples

In this section we apply the above procedure to numerical examples.



doi: 10.1049/iet-rsn:20070102

**Table 1** The result of AIC (p = 7. n = 100)

| $\hat{q}=$ / $q=$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 0 | 0.9274 | 0.0677 | 0.0042 | 0.0007 | 0 | 0 | 0 |
| 1 | 0 | 0.9088 | 0.0836 | 0.0063 | 0.0011 | 0.0002 | 0 |
| 2 | 0 | 0 | 0.8995 | 0.0878 | 0.0107 | 0.0014 | 0.0006 |
| 3 | 0 | 0 | 0 | 0.8861 | 0.0959 | 0.0133 | 0.0047 |
| 4 | 0 | 0 | 0 | 0 | 0.8670 | 0.1092 | 0.0238 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0.8633 | 0.1367 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

**Table 2** The result of HQ (p = 7. n = 100)

| $\hat{q}=$ / $q=$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0.9970 | 0.0030 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0.9950 | 0.0050 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0.9930 | 0.0070 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0.9780 | 0.0210 | 0.0010 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0.9670 | 0.0330 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

*Example 1:* In this example, the data is generated from a random number generator with complex Gaussian distribution $CN_p(0, \text{diag}[1, \ldots, 1, \delta, \ldots, \delta])$. The number of channels $p$ is 7. The number of $\delta$'s equals to the number of signals $q$ ($q$ is from 0 to 6) and the number of 1's is equal to the $p - q$. The ratio of the signal eigenvalue (the eigenvalue corresponding to the signal subspace) to non-signal eigenvalue (the eigenvalue corresponding to noise subspace) $\delta$ is 21, which corresponds to signal to noise ratio (SNR) of 10 dB. The sample size $n$ is 100. We set the values for $\delta$ and $q$ (from 0 to 6) and simulate 10 000 repetitions in each case. Then we calculate the relative frequency that a criterion select $\hat{q} = i$ when the true value of $q$ is $j$ for $i, j = 0, \ldots, p - 1$. The same configuration was used in [5]. Table 1–6 list such relative frequencies for AIC, HQ, BIC, AICc, KICvc and WIC, respectively.

The diagonal elements of the above tables are plotted in Fig. 1. Those values represent the approximated (or simulated) probability of correct estimation for each model selection criterion. The results for AIC, HQ, AICc, BIC, KICvc and WIC are plotted, respectively, by the solid line, dotted line, dash-dot line, dashed line, dashed line with dot marker and a line with star marker. For example in Tables 5 and 6, when the true number of signals is q = 1, the probabilities of correctly estimating q are 0.9420 and 1, respectively, for KICvc and WIC while when the true number of signals is 5, the probabilities of correctly estimating q are 1 and 0.9970, respectively, for KICvc and WIC.

**Table 3** The result of BIC (p = 7. n = 100)

| $\hat{q}=$ / $q=$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0.9920 | 0.0080 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

**Table 4** The result of AICc (p = 7. n = 100)

| $\hat{q}$ = / $q$ = | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 0 | 0.9860 | 0.0140 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0.9980 | 0.0020 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

**Table 5** The result of KICvc (p = 7. n = 100)

| $\hat{q}$ = / $q$ = | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 0 | 0.8390 | 0.1600 | 0.0010 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0.9420 | 0.0570 | 0.0010 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0.9910 | 0.0090 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0.9950 | 0.0050 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

**Table 6** The result of WIC (p = 7. n = 100)

| $\hat{q}$ = / $q$ = | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0.9970 | 0.0030 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

Another word, the probabilities of erroneously identifying the number of signals for the case when the true number of signals is q = 1 are 0.0580 and 0, respectively, for KICvc and WIC. The same comparison method will be used in Examples 2 and 3 below. From Fig. 1, we see that AIC and HQ are worse than the other three criteria. KICvc is similar

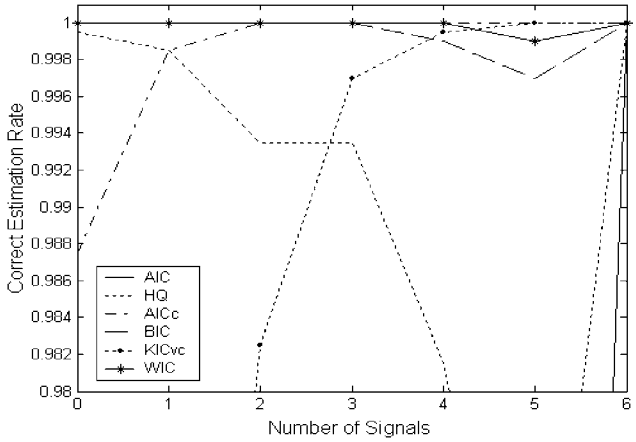



**Figure 1** *Comparison among model selection criteria (p = 7, n = 100)*

but significantly worse than AICc when the number of signal is small. AICc gives a perfect result when the number of signals is large. On the contrary, BIC is the best when the number of signals is small. This result agrees with the conclusion of the previous study given by [10]. The WIC, as a compromise between AICc and BIC, provides the most reliable result (either the best or a very strong second, while other criteria vary more in ranking) over the entire possible number of signals, a desirable feature, since in practice we don't know how many signals there are in the observed data.

Figs. 2 and 3 plot the result when the sample size $n$ is changed to 120 and 80, respectively. As expected, the performance of WIC is getting closer to that of BIC when $n$ increases and to that of AICc when $n$ decreases. Figs. 1–3 together show that, throughout

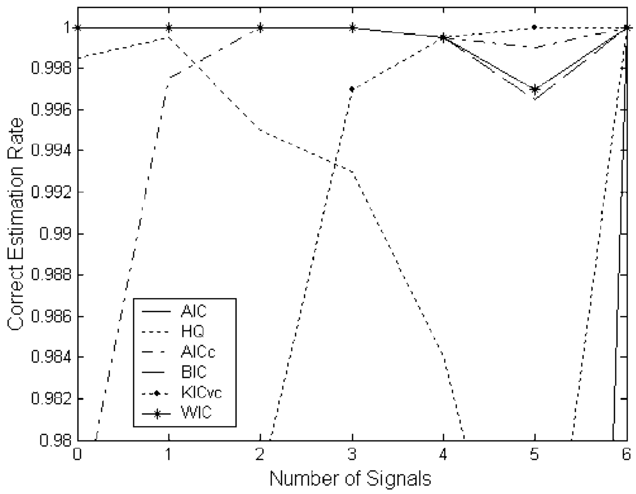



**Figure 2** *Comparison among model selection criteria (p = 7, n = 120)*



doi: 10.1049/iet-rsn:20070102

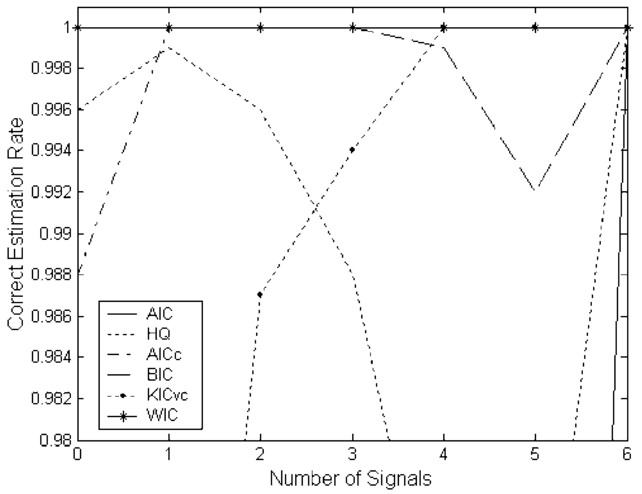



**Figure 3** *Comparison among model selection criteria (p = 7, n = 80)*

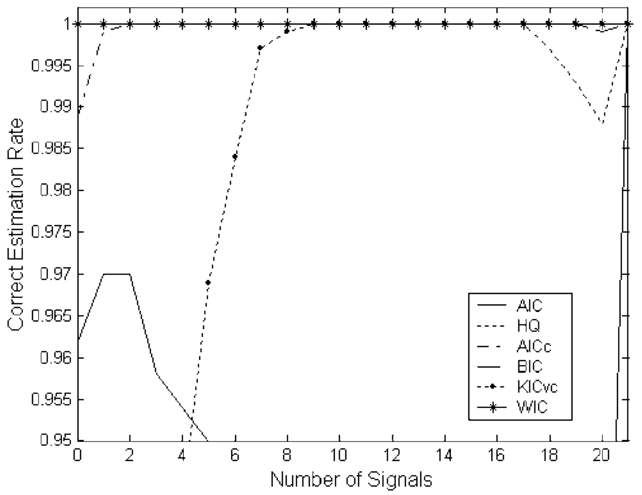



**Figure 4** *Comparison among model selection criteria (p = 22, n = 800)*

all $n$ considered, WIC is the most robust and stable one among all criteria included in the study.

*Example 2:* In this example, we consider the case that the number of channels $p$ is 22 and sample size $n$ is 800. The data is generated from a complex Gaussian distribution $CN_p(0, diag[1, \ldots, 1, \delta, \ldots, \delta])$ Again, the number of $\delta$'s equals to the number of signals $q$ (but here $q$ is from 0 to 21), the number of 1's equals to $p - q$ and $\delta$ is still 21 (i.e. the SNR is still 10 dB). The same configuration was also used in [5]. The results of comparing the model selection criteria are plotted in Fig. 4. It can be seen that AIC is still the worst among the criteria. The result of HQ is much better than that in Example 1. Moreover, BIC and WIC are tied for the best over the entire range of possible number of signals. This is of no surprise, because the sample size is so large that the asymptotics has taken effect for the consistent criteria HQ, BIC and WIC.

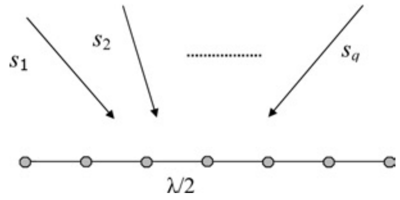



**Figure 5** *Linear antenna array*

*Example 3:* We consider a linear array antenna problem, as depicted in Fig. 5. The antenna elements are uniformly spaced by a half-wavelength $\lambda/2$. The received data are $n$ independent snapshots of

$$\boldsymbol{X}(t) = \sum_{i=1}^{q} \boldsymbol{A}(\theta_i)\mathrm{e}^{-j\eta_i(t)} + \boldsymbol{n}(t) \qquad (23)$$

where $\boldsymbol{A}(\theta_i) = \left[1 \;\; \mathrm{e}^{-j\pi\sin\theta_i} \;\; \cdots \;\; \mathrm{e}^{-j(p-1)\pi\sin\theta_i}\right]^{\mathrm{T}}$, $j = \sqrt{-1}$ and $\eta_i(t)$ is a random phase uniformly distributed over $(0, 2\pi)$ for $i = 1, \ldots, q$. The DOA of the signals is equally spaced between 0 and $\pi$. In radar signal processing, the complex Gaussian random variable $s_i(t)$ in (1) can be expressed as $s_i(t) = a(t)\mathrm{e}^{-j\eta_i(t)}$, where $a(t)$ follows a Rayleigh and $\eta_i(t)$ a uniform. Thus ours is a special case where $s_i(t) = \mathrm{e}^{-j\eta_i(t)}$ for all $i$. The amplitudes of all signals are the same but the phases are different. The SNR is chosen randomly between 0 and 3 dB. The number of antenna elements is $p = 7$. The rest of the simulation setup is the same as those in the last example. Figs. 6 and 7 plot the results of comparing model selection criteria at the sample size 100 and 120, respectively. From the figures we may draw the same conclusion as in Examples 1 and 2. That is, WIC is the most reliable criterion over the entire range of possible number of signals.

# 5 Concluding remarks

In this paper we compare the performance of five model selection criteria for the problem of estimating the number of signals in an additive white noise MUSIC model. We simulate data from several signal processing MUSIC models and compare the performance of AIC, HQ, BIC, AICc, KICvc and WIC. Our simulation result shows that WIC, combining the strengths of AICc and BIC, is a very reliable and stable criterion. Specifically, it shows that WIC performs nearly as well as AICc and outperforms other criteria when the number of signals is relatively large (or, equivalently, when the sample size is relatively small), and WIC performs nearly as well as BIC and outperforms other criteria when the number of signals is relatively small (or, equivalently, when the sample size is relatively large). In cases where we are not certain about the

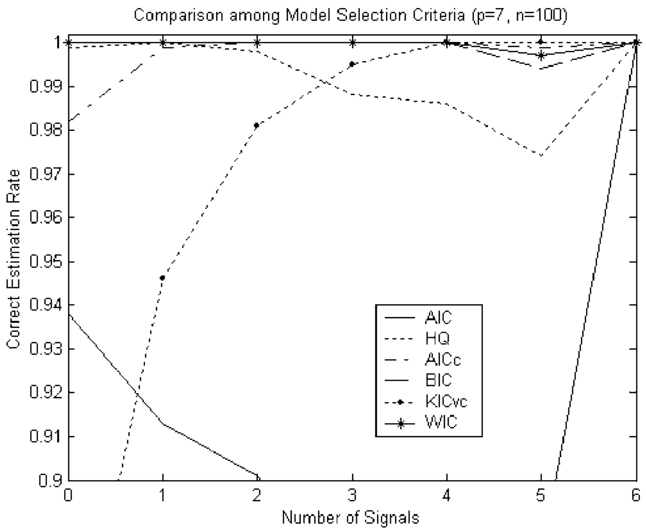



**Figure 6** *Comparison among model selection criteria for antenna array (n = 100)*

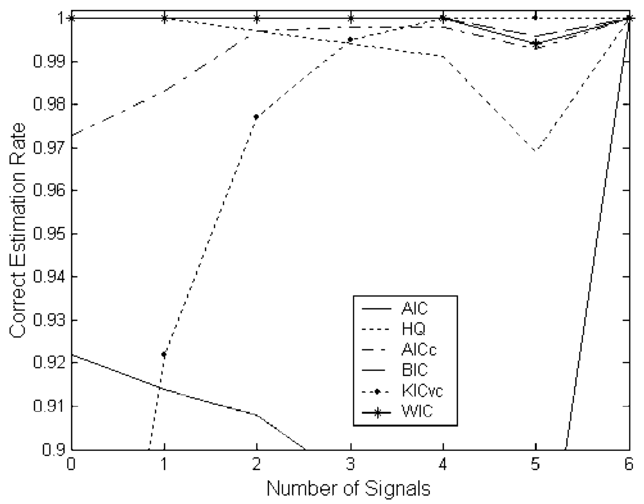



**Figure 7** *Comparison among model selection criteria for antenna array (n = 120)*

relative magnitude of the number of signals, WIC provides a practical alternative to any single criterion.

From our experience in working with MUSIC [4, 16, 17], it appears the sample size $n$ should be fairly large for the asymptotic normality to take effect. We choose to use $n = 100$ and 120 for $p = 7$ and $n = 800$ for $p = 22$, respectively, as they were comparable to the sample sizes used in [4, 16, 17]. We realise that our examples are rather trivial in sensor terms. On the other hand, the main feature of this work is to illustrate, through the specific MUSIC model, that model selection criteria for time series analysis and regression analysis are also suitable in signal processing problems. Our main purpose is that the results of this paper can draw the attention and interests of signal processing researchers to adopt more recent statistical model selection criteria, such as WIC, in general signal processing problems.

Another interesting paper on model selection criteria is written by Stoica and Selen [18]. We have also performed smaller sample simulations such as $n = 14$ (double the degrees of freedom) and $n = 20$ for Example 1. The differences among those six criteria are mixed, unclear and insignificant.

## 6 Acknowledgment

The authors extend their thanks to the editor and the referees for their valuable comments which led to considerable improvement of the article.

## 7 References

[1] SCHMIDT R.O.: 'Multiple emitter location and signal parameter estimation'. Proc. RADC Spectral Estimation Workshop, Rome, NY, 1979, pp. 243–258

[2] BARTLETT M.S.: 'A note on the multiplying factors for various $C^2$ approximations', *J. R. Stat. Soc. B.*, 1954, **16**, pp. 296–298

[3] LAWLEY D.N.: 'Tests of significance of the latent roots of the covariance and correlation matrices', *Biometrika*, 1956, **43**, pp. 128–136

[4] CHEN P.: 'A selection procedure for estimating the number of signal components', *J. Stat. Plan. Inference.*, 2002, **105**, (2), pp. 299–311

[5] WAX M., KAILATH T.: 'Detection of signals by information theoretic criteria', *IEEE Trans. Acoust. Speech Signal Process.*, 1985, **33**, (2), pp. 387–392

[6] AKAIKE H.: 'A new look at the statistical model identification', *IEEE Trans. Autom. Control.*, 1974, **19**, pp. 716–723

[7] SCHWARTZ G.: 'Estimating the dimension of a model', *Ann. Stat.*, 1978, vol. 6, pp. 461–464

[8] RISSANEN J.: 'Modeling by shortest data description', *Automatica*, 1978, **14**, pp. 465–471

[9] HURVICH C.M., TSAI C.L.: 'Regression and time series model selection in small samples', *Biometrika*, 1989, **76**, pp. 297–307

[10] WU T.J., SEPULVEDA A.: 'The weighted average information criterion for order selection in time series and regression models', *Stat. Prob. Lett.*, 1998, **39**, pp. 1–10

[11] PENA D., TIAO G.C., TSAY R.S.: 'A course in time series analysis' (John Wiley & Sons, 2001)

[12] HANNAN E.J., QUINN B.G.: 'The determination of the order of autoregression', *J. R. Stat. Soc. B*, 1979, **41**, pp. 190–195

[13] SEGHOUANE A.K.: 'Multivariate regression model selection from small sample using Kullback's symmetric divergence', *Signal Process.*, 2006, **66**, pp. 2074–2084

[14] KULLBACK S., LEIBLER R.A.: 'On information and sufficiency', *Ann. Math. Stat.*, 1951, **22**, pp. 76–86

[15] LUTKEPOHL H.: 'Comparison of criteria for estimating the order of a vector autoregressive process', *J. Time Ser. Anal.*, 1985, **6**, pp. 35–52

[16] CHEN P., WICKS M.C., ADVE R.S.: 'Development of a procedure for detecting the number of signal in a radar measurement', *IEE Proc., Radar Sonar Navig.*, 2001, **148**, (4), pp 219–226

[17] CHEN P., WICKS M.C.: 'Development of a lower confidence limit for the number of signals', *IEEE Trans. Signal Process.*, 2003, **51**, (6), pp. 1449–1456

[18] STOICA P., SELEN Y.: 'Model-order selection: a review of information criterion rules', *IEEE Signal Process. Mag.*, 2004, **21**, (4), pp. 36–47