Modified AIC and MDL Model Selection Criteria for Short Data Records

Fjo De Ridder, Rik Pintelon, *Fellow, IEEE*, Johan Schoukens, *Fellow, IEEE*, and David Paul Gillikin

***Abstract*—The classical model selection rules such as Akaike information criterion (AIC) and minimum description length (MDL) have been derived assuming that the number of samples (measurements) is much larger than the number of estimated model parameters. For short data records, AIC and MDL have the tendency to select overly complex models. This paper proposes modified AIC and MDL rules with improved finite sample behavior. They are useful in those measurement applications where gathering a sample is very time consuming and/or expensive.**

***Index Terms*—Akaike information criterion (AIC), finite sample, minimum description length (MDL), model selection.**

I. Problem Statement

An IDENTIFICATION procedure typically consists of estimating the parameters of different models and then selecting the optimal model complexity within that set. Increasing the model complexity will decrease the systematic errors, however, at the same time the model variability increases [1], [2]. Hence, it is not a good idea to select the model with the smallest cost function within the set because it will continue to decrease when more parameters are added. At a certain complexity, the additional parameters no longer reduce the systematic errors but are used to follow the actual noise realization on the data. As the noise varies from measurement to measurement, the additional parameters only increase the model variability. To avoid this unwanted behavior, the cost function is extended with a model complexity term that compensates for the increasing model variability. To summarize, the model selection criterion should be able to detect undermodeling (= too simple model) as well as overmodeling (= too complex model). Undermodeling occurs when the true model does not belong to the considered model set, i.e., unmodeled dynamics and/or nonlinear distortions in linear system identification, or too small a number of sinewaves and/or nonperiodic deterministic disturbances in signal modeling. Overmodeling occurs when the considered model includes the true model and is described by too many parameters.

Manuscript received September 9, 2003; revised April 23, 2004.

This work was supported by the Fund for Scientific Research (FWO-Vlaanderen), the Flemish Government (GOA-IMMI), and the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy programming (IUAP 5), and by the Belgian Federal Science Policy Office, Brussels, Belgium (CALMARS, Contract EV/36/04).

F. De Ridder, R. Pintelon, and J. Schoukens are with the Vrije Universiteit Brussel, Department of Fundamental Electricity and Instrumentation, 1050 Brussels, Belgium (e-mail: federid@pop.vub.ac.be; Rik.Pintelon@vub.ac.be).

D. P. Gillikin is with the Vrije Universiteit Brussel, Department of Analytical and Environmental Chemistry, 1050 Brussels, Belgium.

Digital Object Identifier 10.1109/TIM.2004.838132

Two popular model selection criteria are the Akaike information criterion (AIC) [3], and the minimum description length (MDL) [4], [5]. Under the hypothesis of Gaussian disturbances, they take the form

$$\frac{K(\hat{\theta}(Z),Z)}{N}e^{p(n_{\theta'}N)} \text{ with penalty } p(n_{\theta'}N) = \begin{cases} \frac{2n_\theta}{N} & \text{AIC} \\ \frac{\ln(N)n_\theta}{N} & \text{MDL} \end{cases} \tag{1}$$

when the noise variance is unknown [1], and

$$\frac{K(\hat{\theta}(Z),Z)}{N}\big(1+p(n_{\theta'}N)\big) \tag{2}$$

with $p(n_{\theta'}N)$ defined in (1) for known or prior estimated noise variance(s) [6]. In (1) and (2), $\hat{\theta}(Z)$ are the estimated model parameters

$$\hat{\theta}(Z) = \arg\min_{\theta} K(\theta, Z) \tag{3}$$

$n_\theta = \dim(\theta)$ the number of free parameters in the model, $K(\theta, Z)$ the quadratic-like cost function corresponding to the Gaussian maximum likelihood estimator

$$K(\theta,Z) = \frac{1}{2}e^T(\theta,Z)e(\theta,Z) \tag{4}$$

with $e(\theta, Z)$ the N by 1 vector of the (weighted) residuals of the model equation (= difference between measurements and model), Z the measurements, and $N = \dim(e)$ the number of data points. According to the AIC or MDL[1] criteria (1) or (2), the optimal model complexity is obtained by minimizing (1) or (2) over the given set of models. Note that (1) reduces to (2) if $p(n_{\theta'}N) \ll 1$, which is mostly the case in system identification and signal modeling problems. For these cases, the minimizers of (1) and (2) coincide.

The MDL rule has a much better reputation than the AIC rule: The MDL criterion gives strongly consistent estimates ($N \to \infty$) of the order of autoregressive moving average (ARMA) noise processes [7], while the AIC criterion has a strong tendency to select models that are too complex (see [2], and Section III). Both criteria are nevertheless inappropriate for short data records where the number of data samples N is not much larger than the number of parameters n_θ. Indeed, selecting a model with gives $n_\theta = N$ given $K(\hat{(\theta)}(Z), Z) = 0$,

[1]In case of errors-in-variables (EIV) problems $\ln(N)$ in the MDL penalty (1) is replaced by $\ln(\text{rank}(C_n))$, where C_n is the covariance matrix of the noise on measurements Z (see [2], Chapter 17). For example, for single input single output systems within an EIV framework, $\text{rank}(C_n) = 2N$. Note that for multivariable EIV problems $\dim(Z) > \dim(\varepsilon)$.

and both AIC and MDL are exactly zero. Hence, the global minimum of (1) and (2) is attainted by the most complex models, which is of course totally undesirable. It explains why AIC and MDL select far too complex models for short data records (see Section III). The same phenomenon has been observed in (non)linear regression problems with unknown noise variance [8]–[10] and autoregressive time series modeling [9], [11]. In [11], tailor made finite sample AIC criteria for autoregressive-order selection are proposed, while in [8] and [9], a corrected AIC rule for general (non)linear regression problems is given. The corrected AIC rule [8], [9] for problems with unknown noise variance has the form

$$\mathrm{AIC_c}: \frac{K(\hat{\theta}(Z),Z)}{N}e^{p_c(n_{\theta'}N)} \text{ with penalty } p_c(n_{\theta'}N) = \frac{2(n_\theta+1)}{N-n_\theta-2} \quad (5)$$

Similar to the classical AIC criterion (1), $\mathrm{AIC_c}$ has also the tendency to select overly complex models (see [10] and Section III). Therefore, the following improvement of $\mathrm{AIC_c}$ (5) has been proposed in [10]

$$\mathrm{AIC_u}: \frac{K(\hat{\theta}(Z),Z)}{N-n_\theta}e^{p_c(n_{\theta'}N)} \quad (6)$$

with $p_c(n_{\theta'}N)$ defined in (5). According to the sample size, $\mathrm{AIC_c}$ (very small) or $\mathrm{AIC_u}$ (moderate to large) is preferred [10]. Which model selection criterion to use also depends on the intended application: physical interpretation or prediction. In case of physical interpretation, the criterion that best selects the true model should be preferred, while in case of prediction, the criterion that minimizes the prediction error on validation data should be chosen. In the latter case, the optimal model order may strongly depend on the sample size [11]. In this paper, we focus on the first problem, i.e., the physical interpretation of the data. Note that no finite sample criteria are available for estimation problems with known or prior estimated noise variance(s).

This paper proposes 1) the MDL equivalent of $\mathrm{AIC_c}$ (5) for estimation problems with unknown noise variance, and 2) small sample AIC and MDL rules for estimation problems with known noise variance(s) (see Section II). The performance of the new rules is compared with the existing criteria on three simulation examples (Section III) and one real measurement example (Section IV).

II. Modified AIC and MDL Criteria

A. New Model Selection Rules

When the noise variance is unknown, the new MDL rule has the form

$$\mathrm{MDL_c}: \frac{K(\hat{\theta}(Z),Z)}{N}e^{p_c(n_{\theta'}N)} \text{ with penalty } p_c(n_{\theta'}N) = \frac{\ln(N)(n_\theta+1)}{N-n_\theta-2} \quad (7)$$

while for known or prior estimated noise variance(s) we propose

$$\frac{K(\hat{\theta}(Z),Z)}{N-n_\theta}(1+p_s(n_{\theta'}N)) \text{ with penalty } p_s(n_{\theta'}N) = \begin{cases} \frac{2n_\theta}{N-n_\theta}, & \mathrm{AIC_s} \\ n_\theta\frac{\ln(N)}{N-n_\theta}, & \mathrm{MDL_s} \end{cases} \quad (8)$$

The rationale for deriving (7) and (8) is given in the next section.

B. Rationale

The new $\mathrm{MDL_c}$ rule (7) is derived from the corrected $\mathrm{AIC_c}$ rule (5), by analogy of AIC and MDL in (1). One could derive an $\mathrm{MDL_u}$ rule exactly the same way, starting from the $\mathrm{AIC_u}$ rule (6), but the latter is retained because it has the tendency to select too simple models.

The starting point for deriving (8) are the AIC and MDL rules assuming 1) that the variance(s) of the disturbing noise source(s) is (are) known, and 2) that no model errors are present

$$\begin{aligned} &\mathrm{AIC}: \quad \frac{1}{N}(K(\hat{\theta}(Z),Z)+n_\theta) \\ &\mathrm{MDL}: \quad \frac{1}{N}\left(K(\hat{\theta}(Z),Z)+\frac{n_\theta}{2}\ln(N)\right) \end{aligned} \quad (9)$$

(see [1], [2], and [6]). In the absence of model errors, the expected value of the global minimum of the maximum likelihood cost function equals

$$\mathrm{E}\{K(\hat{\theta}(Z),Z)\} = (N-n_\theta)/2 \quad (10)$$

(see [2], Theorem 17.12). Hence, on the average, the value between brackets of AIC (9) will be

$$\begin{aligned} K(\hat{\theta}(Z),Z)+n_\theta &\approx \frac{N-n_\theta}{2}+n_\theta \\ &\approx \frac{N-n_\theta}{2}\left(1+\frac{2n_\theta}{N-n_\theta}\right) \\ &\approx K(\hat{\theta}(Z),Z)\left(1+\frac{2n_\theta}{N-n_\theta}\right) \end{aligned} \quad (11)$$

Making similar calculations for MDL we get

$$\frac{K(\hat{\theta}(Z),Z)}{N}(1+p_s(n_{\theta'}N)) \quad (12)$$

where $p_s(n_{\theta'}N)$ is defined in (8).

Next, recall that the factor $K(\hat{\theta}(Z),Z)/N$ in the classical (1) and corrected (5) AIC criteria is an estimate of the variance σ^2 of the innovations (= driving white noise source of the disturbance)

$$\hat{\sigma}^2 = \frac{2}{N}K(\hat{\theta}(Z),Z) = \frac{1}{N}\varepsilon^T(\hat{\theta}(Z),Z)\varepsilon(\hat{\theta}(Z),Z) \quad (13)$$

(see [1], [8], and [9]). In the absence of modeling errors n_θ linear dependencies exist among the residuals $\varepsilon(\hat{\theta}(Z),Z)$ (see [2, Ch.

17]) and, therefore, an improved variance estimate is obtained as

$$\hat{\sigma}^2 = \frac{2}{N - n_\theta} K(\hat{\theta}(Z), Z) \tag{14}$$

Replacing $K(\hat{\theta}(Z), Z)/N$ by $K(\hat{\theta}(Z), Z)/(N - n_\theta)$ in (12) finally gives (8).

Although (8) has been derived in the absence of model errors (overmodeling), it is also used in the presence of model errors (undermodeling). In that case $(2K(\hat{\theta}(Z), Z))/(N - n_\theta)$ is the sum of the innovations variance σ^2 and the remaining modeling errors. Note that a similar reasoning has been applied in [1] for deriving AIC (1).

III. Simulation Examples

Three simulation examples are given: a time domain signal modeling problem, a frequency domain system identification problem, and noise modeling problem.

A. *Signal Modeling Example*

The goal in this example is to simultaneously estimate the harmonic content (fundamental frequency and the complex amplitudes) and the time base distortion parameters of the following signal model $s_m(n, \theta)$

$$s(n, \theta) = \sum_{k=1}^{h} A_k \cos(k\omega t_n T_s) + A_{k+h} \sin(k\omega t_n T_s), \quad \text{for } n = 0, 1, \ldots, N-1 \tag{15}$$

where A is a vector of length $2h$ with unknown amplitudes, ω is the unknown fundamental angular frequency, and T_s is the sample period. The time instances, t_n, are given by

$$t_n = n + \sum_{l=1}^{b/2} B_l \cos(2\pi n l T_s) + B_{l+b/2} \sin(2\pi n l T_s) \tag{16}$$

where B is a vector of length b with unknown time base distortion parameters. Note that the number of time base distortion parameters, b, and the number of harmonics, h, are unknown. The signal parameters $\theta = [\omega, A^T, B^T]^T$ are estimated by minimizing

$$K(\theta, Z) = \frac{1}{2} \sum_{n=0}^{N-1} \left(s(n) - s_m(n, \theta)\right)^2 \tag{17}$$

with respect to θ, where $s(n)$ is the observed signal (see [12] and [13] for the details).

The chosen parameter values in the Monte Carlo simulation are $\omega = 2\pi 5$ year^{-1}, $A = [1, 0, 0, 1]^T (h = 2)$, $B = [1, 1, 0, 0, 0, 1]^T (b = 6)$, and $N = 50$. Zero mean white Gaussian noise was added such that the signal-to-noise ratio (SNR) of the measurements equals 2 or 1.5. Note that the variance of the disturbing noise is unknown, and, hence, has to be estimated. The optimal model order (h, b) is selected by minimizing AIC and MDL (1), AIC_c (5), AIC_u (6), and MDL_c (7) over the set $h = 1, \ldots, 5$ and $b = 2, 4, \ldots, 20 (n_\theta = b + 2h + 1)$.

This simulation was run 100 times to show its average performance. Fig. 1 and Table I show the number of times that AIC and MDL (1), AIC_c (5), AIC_u (6) and MDL_c (7) selected a certain model order. The following observations can be made.

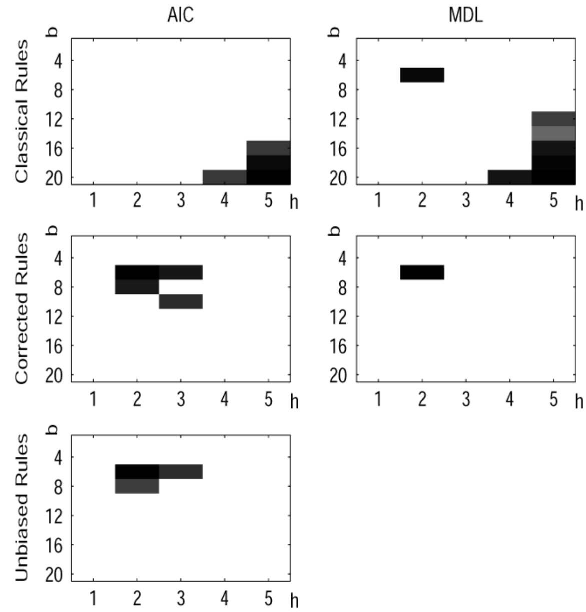

Fig. 1. Graphical visualization of the AIC (left) and MDL (right) values of the simulation $(\text{SNR} = 2)$. The first line gives the result of the old rules (1), the second line of AIC_c (5) and the new MDL_c (7), and the bottom line of AIC_u (6). Notice the contraction of the results around the true model ($h = 2$ and $b = 6$).

TABLE I
Results Signal Modeling Simulation Example for Two SNRs

SNR = 2	classical rules (1) AIC	classical rules (1) MDL	AIC_c (5)	AIC_u (6)	new rule MDL_c (7)
% selection true model (*h*, *b*) = (2, 6)	0%	10%	69%	86%	98%
mean(*h*, *b*) ± std(*h*, *b*)	(5.0, 19.4) ± (0.2, 1.4)	(4.4, 16.6) ± (1.0, 5.1)	(2.3, 6.7) ± (0.8, 1.7)	(2.1, 6.3) ± (0.4, 1.0)	(2.0, 6.0) ± (0.0, 0.3)
SNR = 1.5	**classical rules (1) AIC**	**classical rules (1) MDL**	**AIC_c (5)**	**AIC_u (6)**	**new rule MDL_c (7)**
% selection true model (*h*=2, *b*=6)	0%	4%	59%	83%	63%
mean(*h*, *b*) ± std(*h*, *b*)	(4.9, 19.2) ± (0.4, 1.4)	(4.6, 17.6) ± (0.8, 3.5)	(2.5, 7.3) ± (0.9, 2.5)	(2.1, 6.2) ± (0.6, 1.2)	(2.0, 5.2) ± (0.3, 1.6)

1) The undermodeling ($h = 1$ and $b = 2, 4$) is clearly detected by all model selection criteria, which is not the case for the overmodeling ($h > 2$ and/or $h > 6$).
2) The classical AIC and MDL rules (1) often overestimate the model order: the average selected model orders are too high.
3) When the SNR is high (≥ 2 in this simulation) the new MDL_c rule outperforms AIC_u, while for low SNR values AIC_u performs better.
4) As could be expected for the classical rules (1), AIC selects higher orders than MDL. The same is valid for AIC_c (5) with respect to MDL_c (7).

B. System Identification Example

The frequency domain system identification example concerns the identification of a discrete-time, finite impulse response (FIR) system $G(z^{-1})$

$$G(z^{-1},\theta)=\sum_{r=0}^{R} g_r z^{-r} \quad \text{with } \theta=[g_0,g_1,\ldots,g_R]^T \tag{18}$$

from noisy observations of the input/output discrete Fourier transform (DFT) spectra $U(k), Y(k)$

$$\begin{aligned} Y(n) &= Y_0(n)+N_Y(n)\\ U(n) &= U_0(n)+N_U(n). \end{aligned} \tag{19}$$

$U_0(n), Y_0(n)$ stand for the true unknown input/output DFT spectra and $N_U(n), N_Y(n)$ for the input/output errors. For a given order R, the model parameters are estimated by minimizing the maximum likelihood cost function [2] [see (20)]. with σ_Y^2, σ_U^2, and σ_{YU}^2 the known noise (co-)variances of the disturbing noise sources N_Y and N_U; $z_k=\exp(j2\pi n/N)$; N the number of frequencies in the band $[0,1)$ (normalized with respect to f_s); $\mathrm{Re}(\,)$ the real part of; and where the overline stands for the complex conjugate. Since $z_{N-n}=\bar{z}_n, Y(N-n)=\overline{Y(n)}$ and similarly for $U(n)$ and $\sigma_{YU}^2(n)$, (20) contains only N independent real residuals: $N/2-1$ complex $(n=1,2,\ldots,N/2-1)$ and two real $(n=0,N/2)$. Hence, the AIC and MDL model selection rules (2) should be applied with N and $n_\theta=R+1$. Because of the errors-in-variables framework (all observations are noisy), the MDL rules should be used with $\ln(2N)$ instead of $\ln(N)$ (see footnote 1 and [2]).

The true FIR systems $G_0(z^{-1})$ consists of the 20 first samples [$k=19$ in (18)] of the impulse response of a sixth-order digital Chebyshev filter with a pass band ripple of 0.1 dB and a cutoff frequency of 0.1 (normalized with respect to the sampling frequency f_s). One thousand runs of a Monte Carlo simulation were performed with $N=80$. For each independent run (over the frequency k), circular complex normally distributed errors (see [2]) with standard deviations $\sigma_Y=\sigma_U=0.05$, and $\sigma_{YU}=0$ are added to the true input/output spectra $U_0(k)=1$ and $Y_0(n)=G_0(z_n^{-1})U_0(n)$; and the model parameters θ are estimated for model orders $R=14,15,\ldots,24$. The optimal model order R is selected by minimizing the classical (2), and the new (8) AIC and MDL rules over the set $R=14,15,\ldots,24 (n_\theta=R+1)$.

Table II shows the number of times a model of order R has been selected for the different model selection criteria. The following observations can be made.

1) The undermodeling $(R=14,15,\ldots,18)$ is clearly detected by all model selection criteria, which is not the case for the overmodeling $(R=20,21,\ldots,24)$.

TABLE II
NUMBER OF TIMES AN FIR MODEL (18) OF ORDER n HAS BEEN SELECTED (TRUE MODEL ORDER = 19), AND MEAN VALUE AND STANDARD DEVIATION OF THE SELECTED MODEL ORDER

model order R	**classical rules (2)**		**new rules (8)**	
	AIC	**MDL**	**AIC$_s$**	**MDL$_s$**
14	0	0	0	0
15	0	0	0	0
16	0	0	0	0
17	0	0	0	0
18	0	0	0	0
19	*356*	*607*	*845*	*923*
20	125	136	78	49
21	103	85	45	18
22	112	55	16	4
23	130	55	10	4
24	174	62	6	2
mean value	21.1	20.0	19.3	19.1
standard deviation	1.9	1.5	0.79	0.51

2) The classical AIC and MDL rules (2) clearly overestimate the model order.

3) The new AIC$_s$ and MDL$_s$ rules (8) outperform: The number of times the correct model order is selected is higher, the mean value of the selected orders is closer to the true value, and the standard deviation of the selected orders is smaller.

4) As could be expected for the classical rules (2), AIC selects higher orders than MDL. The same is valid for the new rules (8).

It can be concluded that the new MDL$_s$ rule (8) outperforms on this example.

C. Noise Modeling Example

Consider an autoregressive (AR) noise model of order n

$$y(t)+a_1y(t-1)+a_2y(t-2)+\cdots + a_ny(t-R)=e(t) \tag{21}$$

where $e(n)$ is a zero mean white noise source with variance σ^2. The parameters of the AR model $a_1,a_2,\ldots,a_n$ are estimated form N observations $y(0),y(1),\ldots,y(N-1)$ (no observations of $e(t)$ are available). Since the data record is

$$K(\theta,Z)=\sum_{n=0}^{N-1}\frac{\left|Y(n)-G\left(z_n^{-1},\theta\right)U(n)\right|^2}{\sigma_Y^2(n)+\sigma_U^2(n)\left|G\left(z_n^{-1},\theta\right)\right|^2-2\mathrm{Re}\left(\sigma_{YU}^2(n)\overline{G\left(z_n^{-1},\theta\right)}\right)} \tag{20}$$

short, it is necessary to also estimate the initial conditions $y(-1), y(-2), \ldots, y(-n)$ of the AR process, and the vector of the noise model parameters equals

$$\theta = [a_1, a_2, \ldots, a_R, y(-1), y(-2), \ldots, y(-R)]^T. \quad (22)$$

This can be done either in the time domain [1] or in the frequency domain [2] leading to exactly the same results [14]. The estimate $\hat{\theta}$ is found by minimizing

$$K(\theta, Z) = \frac{1}{2}\sum_{t=0}^{N-1}\left|\left(1+\sum_{r=1}^{R} a_r q^{-r}\right) y(t)\right|^2 \quad (23)$$

with respect to θ, where q^{-1} is the backward shift operator. For the AR-modeling example (21), the performance of the small sample AIC and MDL rules (5) to (7) can also be compared to the combined information criterion (CIC)

$$\frac{K(\hat{\theta}(Z), Z)}{N-R} e^{p_{\text{CIC}}(R,N)} \quad (24)$$

where

$$p_{\text{CIC}}(R,N) = \max\left(\prod_{i=0}^{R}\frac{1+v(i)}{1-v(i)} - 1, 3\sum_{i=0}^{R} v(i)\right) \quad \text{and}$$
$$v(i) = 1/(N+2-2i) \quad (25)$$

for the least squares estimator (23) (see [11]).

The true AR process has the following parameters: $R = 4; a_1 = -1.9602; a_2 = 2.2454; a_3 = -1.6450; a_4 = 0.5776$ (poles at $0.8e^{\pm 0.12\pi j}$ and $0.95e^{\pm 0.42\pi j}$); and $\sigma = 1$. One thousand runs of a Monte Carlo simulation were performed with $N = 50$. For each run, the model parameters θ and the noise variance σ^2 are estimated for model orders $R = 1, 2, \ldots, 10$. The optimal model order n is selected by minimizing AIC and MDL (1), AIC_c (5), AIC_u (6), MDL_c (7), and CIC (24) over the set $R = 1, 2, \ldots, 10 (n_\theta = 2R)$.

Table III shows the number of times a model of order R has been selected for the different model selection criteria. The following observations can be made.

1) Contrary to the signal modeling (see Section III-A) and the system identification example (see Section III-B), both too simple (undermodeling) and too complex (overmodeling) models are selected.
2) The classical AIC (1) and the CIC (24) rules clearly overestimate the model order.
3) The AIC_u rule (6) performs better than all the other rules: The number of times the correct model order is selected is higher, the mean value of the selected orders is closer to the true value, and the standard deviation of the selected orders is smaller. For $N \geq 90$, the MDL_c rule (7) outperforms. Below this N-threshold, the new MDL_c rule too often selects overly simple models and, hence, the AIC_u rule should be used.

Notes

1) Although CIC less often selects the true model than AIC_c and AIC_u, the mean square error of the power spectra of the selected models are about the same. It shows that the variance of CIC (too complex models) is traded for the bias of AIC_c and AIC_u rules (too simple models).
2) When decreasing the number of samples N in the simulation ($N < 50$), first the AIC_c performs best ($N = 40$), and next the CIC rule ($N = 26$).

TABLE III
NUMBER OF TIMES AN AR MODEL (21) OF ORDER n HAS BEEN SELECTED (TRUE MODEL ORDER = 4), AND MEAN VALUE AND STANDARD DEVIATION OF THE SELECTED MODEL ORDER

model order R	classical rules (1)		AIC_c (5)	AIC_u (6)	CIC (24)	new rule MDL_c (7)
	AIC	MDL				
1	0	0	0	0	0	0
2	0	1	0	1	0	3
3	11	72	40	74	20	167
4	*645*	*881*	*880*	*896*	*763*	*824*
5	106	38	67	25	116	5
6	72	7	10	4	49	1
7	48	1	3	0	21	0
8	49	0	0	0	19	0
9	29	0	0	0	5	0
10	40	0	0	0	7	0
mean value	5.0	4.0	4.1	4.0	4.4	3.8
standard deviation	1.7	0.4	0.4	0.3	1.0	0.4

D. Conclusion

In general, for physical modeling, the new MDL_c (7) (unknown noise variance) and MDL_s (8) (known or prior estimated noise variances) rules outperform on small data records. Below some SNR threshold, the new MDL_c (7) and MDL_s (8) rules have the tendency to select overly simple models and, hence, should be replaced by the existing $\text{AIC}_\text{u}/\text{AIC}_\text{c}$ rules and the new AIC_s (8) rule, respectively. This SNR threshold is data dependent and should be established by simulation. On real data with unknown noise variance, the SNR of the measurements can be calculated *a posteriori* using the improved variance estimate (14). An alternative method for choosing between the different AIC and MDL criteria consists in applying the metaselection concept of [15]. A disadvantage of the method is that S times more models must be estimated, where S is the number of data subsets.

IV. REAL MEASUREMENT EXAMPLE

The actual example is a stable oxygen isotope data record[2] sampled from a the skeleton of a bivalve mollusk (a clam), i.e., a *Saxidomus giganteus* (the number of samples is $N = 63$). These data are discussed in greater detail in [16]. Collecting such a data record of the stable oxygen isotopes in a shell takes

[2]A thick section of the aragonite shell was continuously sampled using a computer-controlled micro drill from the growing tip to half way to the umbo. The carbonate powder ($\pm 100\ \mu$ g) was processed using an automated carbonate device (Kiel III) coupled to a Finnigan Delta+XL. Data were corrected using an internal laboratory standard and are reported relative to V-PDB in conventional notation. Precision with this instrumental set up is generally better than 0.08% [16].

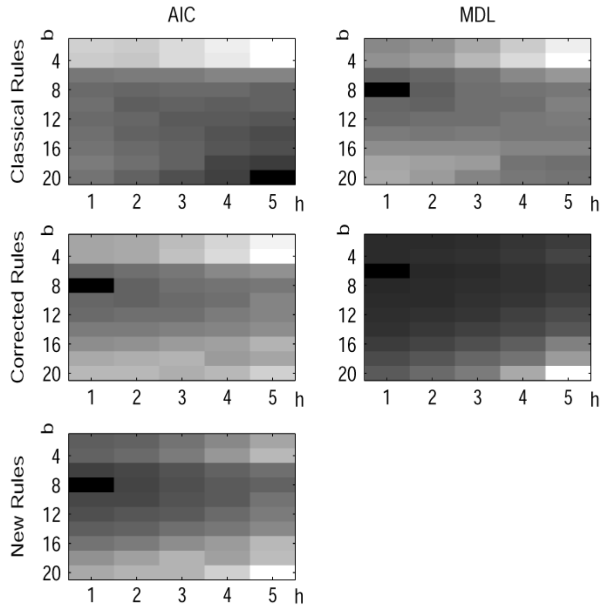

Fig. 2. Graphical visualization of the AIC (left) and MDL (right) values of the real example: black are the lowest values and white are the highest. The first line gives the result of the old rules (1), the second line of $\mathrm{AIC_c}$ (5) and the new $\mathrm{MDL_c}$ (7), and the bottom line of $\mathrm{AIC_u}$ (6).

about 1.5 days work of manual sampling and another 1.5 days for automatically analyzing the samples. This signal is a proxy for the sea surface temperature and the fluctuations have an annual periodicity, which is disturbed by growth rate variations. The annual periodicity can be modeled by (15) and the variations in growth rate by (16). Again, we will try to estimate the number of harmonics, h, and the number of time base distortion parameters, b, necessary to model the signal properly. The most complex model consists of five harmonics and 20 time base distortion parameters. Fig. 2 illustrates the results found by the different criteria, i.e., the classical, corrected and new rules of the AIC and MDL criterion for problems with unknown noise variance. The classical AIC selected the most complex model within the model set $(h = 5, b = 20)$, while the classical MDL criterion was able to select a reasonable model complexity $(h = 1, b = 8)$. The $\mathrm{AIC_c}$ and $\mathrm{AIC_s}$ rules both selected a model with $h = 1, b = 8$. The new $\mathrm{MDL_c}$ criterion selected a model with $h = 1, b = 6$. Both models corresponded to a SNR value of three. A Monte Carlo simulation with this SNR value and a model with $h = 1, b = 6$ was run 50 times and showed that the model selection was performed above the MDL threshold, which means that the new $\mathrm{MDL_c}$ rule (7) is to be preferred above the $\mathrm{AIC_c}$. We conclude the following.

1) While the old MDL rule (1) selected a reasonable model, this is not the case for the old AIC rule (1).
2) The $\mathrm{AIC_c}$ (5) and $\mathrm{AIC_u}$ (6) rules select the same model complexity. This is not in contradiction with the simulation results of Table I, where it can be seen that both rules select the same model with high probability.
3) The new $\mathrm{MDL_c}$ rule (7) selected a slightly less complex model than $\mathrm{AIC_c}$ and $\mathrm{AIC_u}$. Thus, independent of the rules, the AIC selects more complex models than the MDL criterion.

V. Conclusion

Based on an intuitive reasoning, modified AIC and MDL model selection rules for short data records have been derived. Such criteria were not available for estimation problems with known or prior estimated noise model. Based on the simulation results and the real measurement data, the following conclusion can be drawn: No single model selection criterion outperforms under all circumstances. For sufficiently high SNRs or sample sizes, the new $\mathrm{MDL_c}$ (7) and $\mathrm{MDL_s}$ (8) rules outperform. Below some SNR (or sample size) threshold, they should be replaced by the existing $\mathrm{AIC_c}$ (5) or $\mathrm{AIC_u}$ (6) rules and the new $\mathrm{AIC_s}$ (8) rule, respectively. In AR modeling, the CIC (24) rule is also a potential candidate. The choice between the different finite sample criteria can be based on the SNR or sample size threshold (to be established by simulation) or by the meta model selection approach in [15]. The classical rules (1) and (2) are a special case of the new rules (7) and (8). With an increasing number of samples the new rules (7) and (8) converges asymptotically to the old rules (1) and (2). So, the new rules can be used for short as well as for long data records.

References

[1] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.
[2] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*. New York: IEEE Press, 2001.
[3] H. Akaike, “A new look at the statistical model identification,” *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.
[4] J. Rissanen, “Modeling by shortest data description,” *Automatica*, vol. 14, pp. 465–471, 1978.
[5] G. Schwarz, “Estimating the dimension of a model,” *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.
[6] J. Schoukens, Y. Rolain, and R. Pintelon, “Modified AIC rule for model selection in combination with prior estimated noise models,” *Automatica*, vol. 38, no. 5, pp. 903–906, 2002.
[7] E. J. Hannan, “The estimation of the order of an ARMA process,” *Ann. Stat.*, vol. 8, no. 5, pp. 1071–1081, 1980.
[8] N. Sugiura, “Further analysis of the data by Akaike’s information criterion and the finite corrections,” *Comm. Statist.*, vol. A7, pp. 13–26, 1978.
[9] C. M. Hurvich and C. L. Tsai, “Regression and time series model selection in small samples,” *Biometrica*, vol. 76, pp. 297–307, 1989.
[10] A. McQuarrie, R. Shumway, and C. L. Tsai, “The model selection criterion $\mathrm{AIC_u}$,” *Statistics Probability Lett.*, vol. 34, no. 3, pp. 285–292, 1997.
[11] P. M. T. Broersen, “Finite sample criteria for autoregressive order selection,” *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3550–3558, Dec. 2000.
[12] F. De Ridder, R. Pintelon, J. Schoukens, D. P. Gillikin, L. André, W. Baeyens, and F. Dehairs, “Decoding nonlinear growth rates in biogenic archives,” *Geochem. Geophys. Geosyst.*, submitted for publication.
[13] F. De Ridder, R. Pintelon, J. Schoukens, and A. Verheyden, “Two models for characterizing a nonharmonic time base distortion,” *IEEE Trans. Instrum. Meas.*, to be published.
[14] R. Pintelon and J. Schoukens, “Time series analysis in the frequency domain,” *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 206–210, Jan. 1999.
[15] X. De Luna and K. Skouras, “Choosing a model selection strategy,” *Scand. J. Statist.*, vol. 30, pp. 113–128, 2003.
[16] D. P. Gillikin, H. Ulens, F. De Ridder, M. Elskens, E. Keppens, W. Baeyens, and F. Dehairs, “Environmental and biological controls on oxygen and carbon isotopes in the aragonitic bivalve Saxidomus giganteus: Implications for palaeoclimate studies,” *Palaeogeography Palaeoclimatology Palaeoecology*, submitted for publication.

Fjo De Ridder was born in Leuven, Belgium, on March 30, 1976. He received the degree in chemistry, option physical chemistry, in July 1999 from the Free University of Brussels (VUB), Brussels, Belgium.

In October 2000, he joined the department ELEC of the VUB. Currently, he is working as a Ph.D. research assistant. His research is focused on applications of system identification techniques in the field of environmental chemistry (microanalysis of minor and trace elements in biogenic calcite).

Rik Pintelon (M'90–SM'96–F'98) was born in Gent, Belgium, on December 4, 1959. He received the degree of electrical engineer (burgerlijk ingenieur), the degree of doctor in applied sciences, and the qualification to teach at university level (geaggregeerde voor het hoger onderwijs) from the Vrije Universiteit Brussel (VUB), Brussels, Belgium, in July 1982, January 1988, and April 1994, respectively.

From October 1982 to September 2000, he was a researcher of the Fund for Scientific Research—Flanders at the VUB. Since October 2000, he has been a Professor at the VUB in the Electrical Measurement Department (ELEC). His main research interests are in the field of parameter estimation/system identification, and signal processing.

Johan Schoukens (M'90–SM'92–F'97) was born in Belgium in 1957. He received the degree of engineer and the degree of doctor in applied sciences from the Vrije Universiteit Brussel, Brussels, Belgium, in 1980 and 1985, respectively.

He is currently a Professor at the Vrije Universiteit Brussel. His main research interests are in the field of system identification for linear and nonlinear systems.

David Paul Gillikin was born in Syosset, NY, on April 2, 1970. He received the B.Sc. degree in geology from the State University of New York, New Paltz, in 1994 and the M.Sc. degree in marine science from the Vrije Universiteit Brussel (VUB), Brussels, Belgium, in 2000.

Since May 2001, he has been involved in the Belgian Federal Science Policy Office project: CALMAR's (www.vub.ac.be/calmar) as a Ph.D. degree candidate. His main research interests involve the validation and/or calibration of bivalves as environmental archives.