Journal of Engineering and Applied Sciences 12 (Special Issue 1): 5660-5665, 2017
ISSN: 1816-949X

Comparison of Information Criterion on Identification of Discrete-Time Dynamic System

Md. Fahmi Abd. Samad and Abd. Rahman Mohd. Nasir
Faculty of Mechanical Engineering, Universiti Teknikal Malaysia Melaka,
76100 Durian Tunggal, Malacca, Malaysia

Abstract: Information criterion is an important factor for model structure selection in system identification. It is used to determine the optimality of a particular model structure with the aim of selecting an adequate model. A good information criterion not only evaluate predictive accuracy but also the parsimony of model. There are many information criterions those are widely used such as Akaike Information Criterion (AIC) corrected Akaike Information Criterion (AICc) and Bayesian Information Criterion (BIC). Another information criterion suggesting use of logarithmic penalty, named as Parameter Magnitude-based Information Criterion (PMIC) was also introduced. This study presents a study on comparison between AIC, AICc, BIC and PMIC in selecting the correct model structure for simulated models. This shall be tested using computational software on a number of simulated systems in the form of discrete-time models of various lag orders and number of term/variables. As a conclusion, PMIC performed in optimum model structure selection better than AIC, AICc and BIC.

Key words: Akaike information criterion, Bayesian information criterion, model structure selection, parameter magnitude-based information criterion, system, software

INTRODUCTION

System identification is the field of approximating dynamic system models using experimental data (Saleem *et al*., 2015). Its basic idea is to compare the time dependent responses of the actual system and identified model based on a performance function, hereby, referred to information criterion, giving a measure of how well the model response fits the system response (Alfi and Fateh, 2010). The procedure of identification can be divided into several distinctive steps. These steps are data acquisition, model structure selection, parameter estimation and model validity tests (Keesman, 2011). An identification procedure typically consists in estimating the parameters of different models and next 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 (Ridder *et al*., 2005).

When selecting a model structure, two considerations need to be evaluated. One is model accuracy and the other one is model parsimony known as variance and bias (Ljung, 1999):

$$f(J) = Var(J)+Bias(J)$$

Hence, it is not a good idea to select the model with the smallest variance 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. To compensate, having too high a model's complexity such loss function or information criterion is extended with a bias term. In brief, the model selection criterion should be able to detect undermodelling (= too simple model) as well as overmodelling (= too complex model). Undermodelling occurs when the true model does not belong to the considered model set. For example, unmodelled dynamics and/or non-linear distortions in linear system identification or too small a number of sine waves and/or nonperiodic deterministic disturbances in signal modelling. Overmodelling occurs when the considered model includes the true model and is described by too many parameters (Ridder *et al*., 2005). Therefore, a study on the effectiveness of information criterion is warranted.

In this study, the effectiveness of Akaike Information Criterion (AIC) (Akaike, 1974) corrected Akaike Information Criterion (AICc) (Anderson *et al*., 1994), Bayesian Information Criterion (BIC) (Schwarz, 1978) and Parameter Magnitude-based Information Criterion (PMIC), (Samad and Fahmi, 2013) are compared by testing on four simulated dynamic models in the form of difference equations model. These models are linear and non-linear

Corresponding Author: Md. Fahmi Abd. Samad, Faculty of Mechanical Engineering, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Malacca, Malaysia

autoregressive models with exogenous input (ARX and NARX) (Ljung, 1999). The benefit of using simulated models is the presence of an opportunity to compare the final model directly with the true model.

System identification: System identification is a pre-requisite to analysis of a dynamic system and design of an appropriate controller for improving its performance. The more accurate the mathematical model identified for a system the more effective will be the controller designed for it (Zibo and Naghdy, 1995). Often in order to deal with the bias-variance trade-off, the loss function or information criterion is augmented with a penalty term intended to guide the search for the "optimal" relationship penalizing undesired regressors where regressors refer to possible terms and variables identified from model order and linearity specifications. Regularized estimation has been widely applied also in the context of system identification (Prando *et al.*, 2015). Several strategies have been proposed to avoid over-parameterization while utilizing all the data for training the model. The most popular strategy is to minimize a theoretically derived formula or criterion which includes a goodness-of-fit index and a penalty factor for model complexity (Xiao and Mukkamala, 2005). System identification can be framed as an optimization problem:

$$\hat{\theta} = \arg\min_{\theta} J_F(\theta, D_N)$$

Where:

$J_F(\theta, D_N)$ = Measure how well the model described by parameter
θ = Describes the measure data

A widely used variation of the estimation criterion includes a so-called 'regularization term' in the loss function to be minimized that is:

$$\hat{\theta} = \arg\min_{\theta} J_F(\theta, D_N) + J_R(\theta, n)$$

Where:

θ = Estimated by trading-off the data fitting term
$J_F(\theta, D_N)$ = The regularization term
$J_R(\theta, n)$ = Which act as a penalty to penalize certain parameters vectors θ which describe 'unlikely' systems (Prando *et al.*, 2015)

In today's literature, various types of models are proposed for system modelling such as linear autoregressive with exogenous input (ARX) Model and nonlinear autoregressive with exogenous input (NARX) Model (Ljung, 1999).

Information criterion: Model complexity selection is the sub-problem of model selection (Kristinsson and Dumont, 1992). Parsimony, working hypotheses and strength of evidence are three principles that regulate the ability to make inferences (Kristinsson and Dumont, 1992). To overcome this, many information criterions were developed such as AIC, AICc, BIC and PMIC. Estimation of the Kullback-Leibler information is the key to deriving the AIC which was the first model selection criterion to gain widespread acceptance (Burnham and Anderson, 2002). Example of another widely used criterion is BIC which is based, respectively on Bayesian and coding theory. AIC is written as:

$$AIC = n \ln \frac{RSS}{n} + 2k$$

Where:

n = The number of observations
RSS = The Residual Sum of Square is the maximized value of the likelihood function for the estimated model
k = The number of parameters in the statistical model (Akaike, 1974)

Maximize value of the likelihood function for the estimated Model (RSS) can be defined in equation as:

$$L = \sum_{t=k}^{N} \varepsilon^2(t) = \sum_{t=k}^{N} \left(y(t) - \hat{y}(t)\right)^2$$

Where:

$\varepsilon(t)$ = The residual
$\hat{y}(t)$ and $y(t)$ = The k-step-ahead predicted output and actual output value at time t, respectively
N = The number of data

The k-step-ahead prediction is used when the value of k depends on the output's smallest lag order in the selected model structure which in turn depends on the variables selected by the search method.

Although, AIC has proven to be widely applicable, it can have serious deficiencies. Indeed, AIC was designed as an approximately unbiased estimator of the expected Kullback-Leibler divergence between the generating model and the fitted approximating model under the assumption that the true model is correctly specified or overfitted (Seghouane and Amari, 2007). When the sample size is small or when the number of fitted parameters is a moderate to large fraction of the sample size, AIC becomes a strongly negatively biased estimate of the Kullback-Leibler divergence and leads to the choice of over parameterized models (Burnham and Anderson, 2002).

From Anderson *et al.* (1994) improvement has been made to AIC called corrected Akaike Information Criterion (AICc). Different approaches have been made to improve AIC by correcting its penalty term. One such approach is to asymptotically evaluate the penalty term as precisely as possible to provide better estimates of the model order (Burnham and Anderson, 2002). AICc can be written as:

$$\text{AICc} = n\ln\frac{\text{RSS}}{n} + \frac{2k(k+1)}{n-k-1}$$

The BIC is a likelihood criterion penalized by the model complexity. The penalty in BIC for additional parameters is known to be stronger than that of the AIC. The BIC is an asymptotic result derived under the assumptions that the data distribution is in the exponential family (Schwarz, 1978). BIC is defined as:

$$\text{BIC} = n\ln\frac{\text{RSS}}{n} + k\ln(n)$$

According to Samad and Fahmi (2013), Parameter Magnitude-based Information Criterion (PMIC) evaluates the bias contribution by the sum of squared residuals while the variance contribution is calculated by a Penalty Function (PF). This written as follows:

$$\text{PMIC} = \left(\sum_{t=k}^{N}(y(t)-\hat{y}(t))^2\right) + \text{PF}$$

where, PF = ln n, n is The number of terms satisfying ($|a_j|$<penalty+1 whilst $|a_j|$represents the absolute value of the parameter for term j and penalty is a fixed value termed penalty function parameter.

The penalty function penalises terms with the absolute values of the estimated parameter less than the penalty. The penalty parameter value will be set equal to or slightly lower than the parameter value in cases the smallest tolerable absolute parameter value is known or can be roughly estimated (Samad and Fahmi, 2013).

MATERIALS AND METHODS

Simulation setup: In this simulation, three ARX and a NARX Model are simulated using computer simulation software MATLAB. All models are denoted as Model 1- 4 and each model is further classified as having d.c. level and not having d.c. level. The difference between the two are one having the output average subtracted (hence, has no d.c. level) and the other not subtracted. The following are the models written as linear regression models, its specifications, number of correct regressors and number of possible regressors.

Model 1:

$$y(t) = 0.1y(t-2) + 0.3u(t-1) + 0.8u(t-3) + e(t)$$

Specification: l = 1, assumed maximum output order, n_y = 3, maximum input order, n_u = 3. Number of correct regressor = 3 out of 7 (if d.c. level is assumed present) or 6 (if d.c. level is assumed absent). Number of possible model =127 (with d.c. level) or 63 (without d.c. level).

Model 2:

$$y(t) = 0.1y(t-1) + 0.4y(t-5) - 0.3u(t-3) + 0.5u(t-4) + e(t)$$

Specification: l = 1, assumed maximum output order, n_y = 5, assumed maximum input order n_u = 5. Number of correct regressor = 4 out of 11 (if d.c. level is assumed present) or 10 (if d.c. level is assumed absent). Number of possible model = 2047 (with d.c. level) or 1023 (without d.c. level).

Model 3:

$$y(t) = 0.1y(t-2) + 0.3y(t-4) - 0.3y(t-7) + 0.2u(t-3) + 0.3u(t-5) + e(t)$$

Specification: l = 1, assumed maximum output order, n_y = 7, assumed maximum input order n_u = 7. Number of correct regressor = 5 out of 15 (if d.c. level is assumed present) or 14 (if d.c. level is assumed absent). Number of possible model = 32767 (with d.c. level) or 16383 (without d.c. level).

Model 4:

$$y(t) = 0.3y(t-2) - 0.4u(t-2) + 0.1y(t-1)y(t-2) + 0.1y(t-1)u(t-2) - 0.3u(t-2)u(t-2) + e(t)$$

Specification: l = 2, assumed maximum output order, n_y = 2, assumed maximum input order n_u= 2. Number of correct regressor = 5 out of 15 (if d.c. level is assumed present) or 14 (if d.c. level is assumed absent). Number of possible model = 32767 (with d.c. level) or 16383 (without d.c. level).

The input u(t) is generated from a random uniform distribution in the interval [-1, 1] to represent white signal, while noise e (t) is generated from a random uniform distribution [-0.01, 0.01] to represent white noise. The 5 hundred data points are generated from all models. All models are penalized by AIC, AICc, BIC and PMIC in order to select a correct model. Throughout the simulation, the penalty parameter value for PMIC was set in accordance to the recommendation by Samad and Fahmi (2013), specifically as 0.1.

RESULTS AND DISCUSSION

In this study, comparisons are made between AIC, AICc, BIC and PMIC for all models based on their selected model. The selected models are based on the minimization of respective criterions. Table 1-4 show the results. Models with d.c. level are denoted as Model 1a-4a while models without d.c. level are denoted as Model 1b-4b. The simulated model is denoted as S.M. For brevity the variables and terms are not included but numbered.

Figure 1 shows the comparison of number of regressors selected between AIC, AICc, BIC and PMIC. From the tables and figure AIC, AICc and BIC showed the same outcome but PMIC performed differently. As can be seen, for all linear models (ARX Model) PMIC performed well by selecting the same model as the simulated models while AIC, AICc and BIC cannot select the correct model. Either under-modelling or over-modelling occurred in the case of the ARX Models.

Besides, at Model 4b, AIC, AICc and BIC selected a wrong model however, the number of regressor was not too far different from the right number. This is when AIC, AICc and BIC selected a model with 6 regressors out of 14 while simulated model has 5 regressors. Over-modelling occurred for these criterions in the case of Model 4a. PMIC almost able to select the right model on Model 4a with only one regressor different which is 0.1y (t-1) u (t-1) instead of 0.1y (t-1)u (t-2).PMIC chose too complex model with 11 regressors out of 14 in the case of Model 4b.

Table 1: Results on Model 1

	Regressor No.						
Model 1a	1	2	3	4	5	6	7
S.M	-	-	0.1	-	0.3	-	0.8
AIC, AICc, BIC	-0.02	-	-	0.1	-	-	-
PMIC	-	-	0.1	-	0.3	-	0.8
Model 1b							
S.M	-	0.1	-	0.3	-	0.8	-
AIC, AICc, BIC	0.02	-	-	-	-0.02	-	-
PMIC	-	0.1	-	0.3	-	0.8	-

Table 2: Results on Model 2

	Regressor No.										
Model 2a	1	2	3	4	5	6	7	8	9	10	11
S.M	-	0.1	-	-	-	0.4	-	-	-0.3	0.5	-
AIC, AICc, BIC	-0.6	-	-0.06	-	-	0.3	0.01	0.01	-0.3	-	-
PMIC (θ)	-	0.1	-	-	-	0.4	-	-	-0.3	0.5	-
Model 2b											
S.M	0.1	-	-	-	0.4	-	-	-0.3	0.5	-	-
AIC, AICc, BIC	-	-	-0.02	-	-	0.02	-	-0.3	-	-	-
PMIC (θ)	0.1	-	-	-	0.4	-	-	-0.3	0.5	-	-

Table 3: Results on Model 3

	Regressor No.														
Model 3a	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
S.M	-	-	0.1	-	0.3	-	-	-0.3	-	-	0.2	-	0.3	-	-
AIC, AICc, BIC	-0.2	-	0.7	-	-	-	-	-	-	-	-	0.02	-	-	-
PMIC (θ)	-	-	0.1	-	0.3	-	-	-0.3	-	-	0.2	-	0.3	-	-
Model 3b															
S.M	-	0.1	-	0.3	-	-	-0.3	-	-	0.2	-	0.3	-	-	-
AIC, AICc, BIC	-0.01	1.3	-0.01	0.1	-0.1	-0.4	-0.04	-	0.01	0.2	-	-	0.01	-0.4	-
PMIC	-	0.1	-	0.3	-	-	-0.3	-	-	0.2	-	0.3	-	-	-

Table 4: Results on Model 4

	Regressor No.														
Model 4a	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
S.M	-	-	0.30	-	-0.4	-	0.1	-	-	0.1	-	-	-	-	-0.3
AIC, AICc, BIC	-1.0	-	0.14	0.1	-	0.01	0.1	-0.01	-	-	-	-0.1	-	0.01	-
PMIC	-	-	0.30	-	-0.4	-	0.1	-	0.1	-	-	-	-	-	-0.3
Model 4b															
S.M	-	0.3	-	-0.4	-	0.1	-	-	0.1	-	-	-	-	-0.3	-
AIC, AICc, BIC	-	0.2	-	0.2	0.01	-	-	0.14	-	0.06	-	-0.02	-	-	-
PMIC	-0.04	0.4	-	0.2	0.04	0.1	0.01	0.10	0.04	0.01	-	0.100	-	-0.2	-

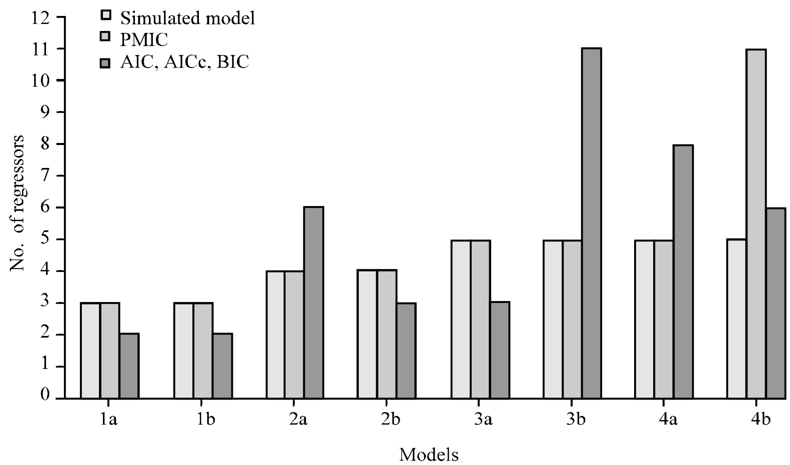

Fig. 1: A comparison of regressor number selected between AIC, AICc, BIC and PMIC for all models

However, PMIC capability in selecting the correct model can be considered much better than AIC, AICc and BIC.

CONCLUSION

From this simulation, AIC, AICc and BIC did not perform well when selecting both linear and non-linear model (ARX and NARX Model). PMIC proved that it can select better models than AIC, AICc and BIC in order to choose the correct linear model (ARX Model). However, in simulation of non-linear model (NARX Model), PMIC almost picked the right model on Model 4a but over-modelling occurred when selecting Model 4b, a NARX Model assuming no d.c. level. The reason that PMIC did not perform well was maybe due to unsuitable penalty parameter value. More rigorous analysis could be made to identify such weakness and therefore improve the capability of PMIC.

ACKNOWLEDGEMENTS

The researchers would like to acknowledge all the support from Universiti Teknikal Malaysia Melaka and Ministry of Higher Education Malaysia for research grant FRGS/1/2015/TK03/UTEM/02/16.

REFERENCES

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control, 19: 716-723.

Alfi, A. and M.M. Fateh, 2010. Parameter identification based on a modified PSO applied to suspension system. J. Software Eng. Appl., 3: 221-229.

Anderson, D.R., K.P. Burnham and G.C. White, 1994. AIC model selection in overdispersed capture-recapture data. Ecol., 75: 1780-1793.

Burnham, K.P. and D.R. Anderson, 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2nd Edn., Springer, New York, USA., ISBN-13: 9780387953649, Pages: 488.

Keesman, K.J., 2011. System Identification: An Introduction. Springer, Berlin, Germany, ISBN:978-085729-521-7, Pages: 322.

Kristinsson, K. and G.A. Dumont, 1992. System identification and control using genetic algorithms. IEEE. Trans. Syst. Man Cybern., 22: 1033-1046.

Ljung, L., 1999. System Identification: Theory for the User. 2nd Edn., Prentice Hall PTR, London, UK., ISBN-10: 9780136566953, Pages: 609.

Prando, G., G. Pillonetto and A. Chiuso, 2015. The role of rank penalties in linear system identification. IFAC. Pap. Online, 48: 1293-1300.

Ridder, D.F., R. Pintelon, J. Schoukens and D.P. Gillikin, 2005. Modified AIC and MDL model selection criteria for short data records. IEEE. Trans. Instrum. Meas., 54: 144-150.

Saleem, A., B. Taha, T. Tutunji and A.A. Qaisia, 2015. Identification and cascade control of servo-pneumatic system using particle swarm optimization. Simul. Modell. Pract. Theory, 52: 164-179.

Samad, A. and M. Fahmi, 2013. Effect of penalty function parameter in objective function of system identification. Intl. J. Automot. Mech. Eng., 7: 940-954.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat., 6: 461-464.

Seghouane, A.K. and S.I. Amari, 2007. The AIC criterion and symmetrizing the kullback leibler divergence. IEEE. Trans. Neural Networks, 18: 97-106.

Xiao, X., Y. Li and R. Mukkamala, 2005. A model order selection criterion with applications to cardio-respiratory-renal systems. IEEE. Trans. Biomed. Eng., 52: 445-453.

Zibo, Z. and F. Naghdy, 1995. Application of genetic algorithms to system identification. Proceedings of the IEEE International Conference on Evolutionary Computation, November 29-December 1, 1995, IEEE, Wollongong, New South Wales, ISBN:0-7803-2759-4, pp: 777-782.