Improved Model Order Selection in Dynamical System Identification Based on Trend Extraction

Ju Liu [*,**] **Jiayi Zhao** [*] **Chao Shang** [*] **Dexian Huang** [*]

[*] *Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: c-shang@tsinghua.edu.cn).*
[**] *Luoyang Electronic Equipment Test Center, Luoyang 471000, China.*

Abstract: Model order selection plays a crucial role in system identification. The existing model order selection methods rely on finding the balance between fitting error and model complexity. However, when the data contains large noise, the model order obtained by the existing methods may not be reliable. To resolve this issue, we present a new model order selection method based on trend error analysis. By making use of a specific property of trend extraction — the insensitivity against noise, our method improves the accuracy of model order selection. Numerical simulation results show the effectiveness of the proposed method and outperformance over known heuristics under different noise levels.

Keywords: Model order identification, Trend error, Fitting error.

1. INTRODUCTION

System identification is a pragmatic method of constructing the mathematical model of a dynamic system by measuring its input and output signals. It is widely used in the field of industrial control. System identification mainly includes two parts: model structure identification and parameter identification. System models include parametric models and non-parametric models. Unlike non-parametric models, which do not require the prior knowledge of model structure, parametric models cannot identify the model's parameters without knowing its structure. Thus, model structure identification is the basis of the identification of parameter models. Specifically, for single-input-single-output (SISO) systems, model order identification is central to deciding the model structure (Herpen et al. (2011); Yang et al. (2013); Varanasi and Jampana (2016); Besanon et al. (2018); Liu et al. (2023a)).

The existing order identification methods of parameter models mainly include the F-test method (Fu et al. (2018)), Akaike information criterion (AIC) (Ninomiya (2005)) and final prediction error criterion (Weerts et al. (2018); Liu et al. (2021)). The F-test method deducts a reasonable model order by testing how the fitting error changes when the hypothetic model order increases. If the fitting error changes significantly at one point, then the corresponding model order of that point is the estimated model order. AIC defines a function composed of the likelihood function and the model order, and the model order is determined by minimizing the function. The prediction error method defines a criterion function of the prediction error. Similarly, the order that minimizes the criterion function is selected as the model order. The essence of these methods is to select the model order according to the change of fitting error. However, when the noise corruption is heavy, the order selected by these methods may not be accurate.

Considering that the data trend is an effective information that can reflect the change of data and is less affected by noise, we proposed a model order identification method based on trend error analysis. First, we extract the trend of the model output to obtain the trend error of the predicted value. Then, the holistic error composed of trend error and fitting error is defined. Finally, the most appropriate model order is selected according to the curve of holistic error changing with model order. To test the validity of the method, we use a numerical case of auto-regressive moving average with extra input (ARMAX) model to carry out the experiment. The results show that the proposed method performs better than the existing statistical-testing methods and information-theoretic methods, which are based on fitting error analysis.

The remainder of this article unfolds as follows. Section 2 revisits the existing order identification methods and trend extraction methods. In Section 3, an order identification method based on trend error analysis is proposed. The performance of the proposed method on numerical examples is investigated in Section 4, followed by final conclusions.

2. PRELIMINARIES

2.1 The identification of ARMAX models

Identifying an SISO system involves finding the number of model parameters and accurately estimating the value of each parameter. In industrial processes, the ARMAX model has been widely used to describe SISO systems. The ARMAX model can be described as follows :

⋆ This work was supported by National Key R&D Program of China (2022YFB3304703).

$$A(z^{-1})y(k) = B(z^{-1})u(k) + D(z^{-1})v(k) \quad (1)$$

where $u(k)$ and $y(k)$ are the input and output respectively; $v(k)$ is a Gaussian white noise with a mean of 0 and a variance of σ_v^2. The delay operator polynomials $A(z^{-1})$, $B(z^{-1})$ and $D(z^{-1})$ of the model can be described as follows:

$$\begin{cases} A(z^{-1}) := 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_{n_a} z^{-n_a} \\ B(z^{-1}) := b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{n_b} z^{-n_b} \\ D(z^{-1}) := 1 + d_1 z^{-1} + d_2 z^{-2} + \cdots + d_{n_d} z^{-n_d} \end{cases} \quad (2)$$

where n_a, n_b and n_d are the model orders. We define $n = n_a + n_b + n_d$. Eq. (1) can be described as the least square form:

$$\boldsymbol{y}_L = \boldsymbol{H}_L \boldsymbol{\theta} + \boldsymbol{e}_L, \quad (3)$$

where L is the data length and

$$\begin{cases} \boldsymbol{y}_L = [y(1), y(2), \cdots, y(L)]^\top \\ \boldsymbol{e}_L = [e(1), e(2), \cdots, e(L)]^\top \\ \boldsymbol{H}_L = \begin{bmatrix} y(0) & \cdots & y(1-n_a) & u(0) & \cdots & u(1-n_b) \\ y(1) & \cdots & y(2-n_a) & u(1) & \cdots & u(2-n_b) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y(L-1) & \cdots & y(L-n_a) & u(L-1) & \cdots & u(L-n_b) \end{bmatrix} \\ \boldsymbol{\theta} = [-a_1, \cdots, -a_{n_a}, b_1, \cdots, b_{n_b}]^\top \\ e(k) = v(k) + d_1 v(k-1) + \cdots + d_{n_d} v(k-n_d) \end{cases} \quad (4)$$

The purpose of system identification is to estimate the parameters $a_1, \cdots, a_{n_a}, b_1, \cdots, b_{n_b}, d_1 \cdots, d_{n_d}$ from the data. Representative identification methods of ARMAX models include extended least squares (Ding and Ding (2008)), maximum likelihood (ML) (Febrianti et al. (2021); Madsen (2009); Michael et al. (2004)), prediction error method (PEM) (Borjas and Garcia (2005); Nguyen and Ohtsu (2000); Casas et al. (2002)) and so on. In the extended least squares method, the objective is to minimize the criterion function $J(\boldsymbol{\theta})$:

$$J(\boldsymbol{\theta}) = (\boldsymbol{y}_L - \boldsymbol{H}_L \boldsymbol{\theta})^\top (\boldsymbol{y}_L - \boldsymbol{H}_L \boldsymbol{\theta}) \quad (5)$$

In the ML method, the goal is to maximize the likelihood function, while PEM aims at minimizing the conditional expectation of error. However, these methods are build upon the premise that the orders (n_a, n_b, n_d) of the model are precisely known. In practice, however, it is non-trivial to attain the model order in advance. Thus, an accurate estimation of model order is the prerequisite for a successful identification of dynamic systems.

2.2 The criteria for order determination

The change of fitting error plays an important role in the existing criteria for model order determination, such as in the F-test method and AIC. Basically, the residual variance V_n tends to decrease with the increase of the order n, and the corresponding order at the inflection point of V_n's change curve indicates a proper choice of the model order. However, due to the noise effect, the uncertainty of the fitting error will increase, making it difficult to determine the inflection point. Therefore, the inflection point is generally determined by the F-test method (Kabaila (2005)). This method introduces a t-statistic for two candidate orders n_1 and n_2:

$$t(n_1, n_2) = \frac{V_{n_1} - V_{n_2}}{V_{n_2}} \frac{L - 2n_2}{2(n_2 - n_1)} \quad (6)$$

where $n_2 > n_1$. When the noise $v(k)$ is Gaussian distributed, the t-statistic is known to follow an F-distribution, i.e., $t \sim F(2(n_2 - n_1), L - 2n_2)$. Let $n_2 = n_1 + 1$, it then follows that:

$$t(n, n+1) = \frac{V_n - V_{n+1}}{V_{n+1}} \frac{L - 2n - 2}{2} \sim F(2, L - 2n - 2) \quad (7)$$

n_0 is denoted as the true order of the model. If the null hypothesis $\mathrm{H}_0 : n \geq n_0$ holds, then the residual variance V_{n+1} will not decrease evidently as compared to V_n. Therefore, by setting the risk level α, one can test whether the null hypothesis H_0 holds:

$$\begin{cases} t(n, n+1) > t_\alpha, \text{ Reject } H_0 \\ t(n, n+1) \leq t_\alpha, \text{ Accept } H_0 \end{cases} \quad (8)$$

where the threshold can be decided as $t_\alpha = F(2(n_2 - n_1), L - 2n_2)$. However, one has to to manually set the risk level α, and a different α may lead to a considerable divergence in the selected order.

Alternatively, in AIC the following information-theoretic objective is considered (Ninomiya (2005)):

$$\mathrm{AIC}(n) = -2\log L(\hat{\boldsymbol{\theta}}_{\mathrm{ML}}) + 2n, \quad (9)$$

where $L(\hat{\boldsymbol{\theta}}_{\mathrm{ML}})$ is the likelihood function. AIC can be interpreted as the sum of two terms; the first term is a measure of model fit, as $L(\hat{\boldsymbol{\theta}}_{\mathrm{ML}})$ is mainly affected by the fitting error. The second term is a penalty for the number of model parameters, The order n of the model represents the number of parameters. The purpose of the AIC is to find a balance between the fitting error and the model complexity. Similar to AIC, there are other information-theoretic such as the Schwarz Criterion (SC)(Lee (1995)) to select the model order. However, the increase of the noise variance will lead to the uncertainty of the fitting error. Thus, the accuracy of the AIC or SC will be affected by the noise.

3. MODEL ORDER SELECTION BASED ON TREND ERROR ANALYSIS

In open-loop industrial processes, the input does not change at every moment but is piecewise constant. Therefore, the output will show the characteristics of piecewise monotonicity. We illustrate this by using an example of actual process industry data. The example is the historical data from a dichloroaniline production process in China. The two main variables are manipulate variable (MV) and process variable (PV). PV is usually the output in industrial processes. MV and PV are shown in Fig. 1. As the figure shows, there is a significant change in MV only between the 200th and 300th points. Correspondingly, PV changes monotonically when MV is changing. This feature provides a possibility of detouring the inaccurate results in real-world scenarios which are filled with uncertain noises. Instead of imprecisely analyzing the prediction error based on the real output and the predicted value, we use the trend extraction of the output to help judge the prediction accuracy. Starting from this point, we propose a new method of selecting model order based on the trend error in this paper.

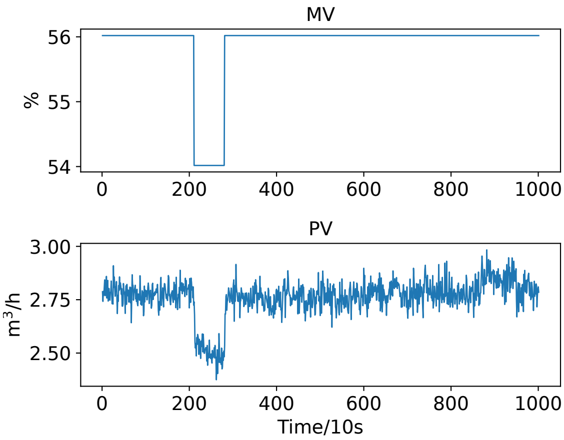


Fig. 1. MV and PV of the process

3.1 Output trend extraction

The central idea of trend extraction is to split a data trajectory into multiple consecutive episodes, where data points in each episode are locally approximated by simple symbolic representations. Common options of symbolic representations can be roughly classified into shape constrained splines Villez et al. (2013); Villez (2015) and low-order polynomials Keogh and Pazzani (1998); Zhou and Ye (2016); Keshani and Masoumi (2021); Liu et al. (2023b), e.g. affine functions. Consider a scalar-valued time series

$$\mathbf{y} = [y(1), y(2), \cdots, y(L)]^\top \tag{10}$$

of length L. To divide $\mathbf{y}$ into $M > 1$ segments, we define $M+1$ segmentation points $\{s_0, s_1, \cdots, s_{M-1}, s_M\}$ in total, where $s_0 = 0$ and $s_M = L$, and $\{s_1, \cdots, s_{M-1}\}$ need to be suitably chosen. The qth-order polynomial is used to approximately capture the trend of $\mathbf{y}$ in each segment and remove the noise:

$$\hat{y}(t) = \beta_0 + \beta_1 t + \cdots + \beta_q t^q, \tag{11}$$

where $\hat{y}(t)$ is the approximation of $y(t)$, $t = 1, \cdots, L$, and $\beta_0, \beta_1, \cdots, \beta_q$ are polynomial coefficients. Eq. (11) can be further expressed as:

$$\hat{\mathbf{y}} = [\hat{y}(1), \hat{y}(2), \cdots, \hat{y}(L)]^\top = \mathbf{T}\boldsymbol{\beta}, \tag{12}$$

where $\boldsymbol{\beta} = [\beta_0, \beta_1, \cdots, \beta_q]^\top$, $\mathbf{T}$ is the time index matrix with entries $T_{i,j} = i^{j-1}$, $1 \leq i \leq L$, $1 \leq j \leq q$. For a given number of segments M, we need to find the optimal segmentation points to minimize the overall fitting error $f(M) = \min \sum_{m=1}^{M} \ell(\hat{\mathbf{y}}_m, \mathbf{y}_m)$, where $\ell(\hat{\mathbf{y}}_m, \mathbf{y}_m)$ is the error loss function for the mth episode. To determine the number M of segments, the bilateral criterion is adopted in the GPTE method (Zhou et al. (2017)):

$$M^* = \arg \max_{1 \leq M \leq M_{\max}} \frac{\log[\frac{f(M-1)}{f(M)}] - \log[\frac{f(M)}{f(M+1)}]}{\log[\frac{f(M-1)}{f(M)}] + \log[\frac{f(M)}{f(M+1)}]}. \tag{13}$$

A wide interval $[1, M_{\max}]$ is necessary for optimally searching for M^*. Since the GPTE method is easy to implement and has been widely used in processing industrial data, it is used in this paper for the sake of trend extraction.

3.2 Trend error and holistic error

We define the trend error TE as as the error between the predicted value and the trend of the real output:

$$\mathrm{TE}(\hat{\boldsymbol{y}}, \boldsymbol{y}) := \|\hat{\boldsymbol{y}} - \bar{\boldsymbol{y}}\|^2, \tag{14}$$

where $\hat{\boldsymbol{y}}$ is the predicted value, and $\bar{\boldsymbol{y}}$ is the trend of the real output. The holistic error HE consists of the fitting error and the trend error:

$$\mathrm{HE}(\hat{\boldsymbol{y}}, \boldsymbol{y}) := c_1\|\hat{\boldsymbol{y}} - \boldsymbol{y}\|^2 + c_2 \mathrm{TE}(\hat{\boldsymbol{y}}, \boldsymbol{y}), \tag{15}$$

where $c_1 > 0$ and $c_2 > 0$ are coefficients of fitting error and trend error, respectively. The sum of c_1 and c_2 is typically set to 1. The larger c_2 is, the more the holistic error focuses on the trend error. When the output $\boldsymbol{z}$ contains noise, c_2 should be set close to 1.

3.3 A new approach to model order selection

Since we do not know the order of the model in advance, we need to traverse all possible order combinations. For a given order combination, we use the maximum likelihood method to estimate the parameters and obtain the prediction value $\hat{\boldsymbol{y}}$. According to the predicted value and the real output value, the fitting error can be obtained. The two steps of selecting the model order combination are as follows.

First, we extract the trend $\bar{\boldsymbol{y}}$ of the output. Second, we calculate the trend error $\mathrm{TE}(\hat{\boldsymbol{y}}, \boldsymbol{y})$ and the holistic error $\mathrm{HE}(\hat{\boldsymbol{y}}, \boldsymbol{y})$ under each order combination and draw the curve of the holistic error along with the changing hypothetic order. We find the inflection point of the holistic error change curve and select the corresponding order as the model order.

The rationale of the proposed method for selecting the model order is as follows. When the model order is lower than the real model order, the fitting error and trend error manifest an obvious decrease with the increase of the order. When the model order is equal to the real order, the fitting error and trend error will reach a relatively small value, which is not 0 because of the noise. With the increase of the model order, the fitting error may still be greatly reduced, but the trend error will not change significantly since the trend of the predicted value is robust to noise.

4. CASE STUDY

We consider the identification of the following ARMAX system:

$$\begin{aligned} y(k) =& 1.7y(k-1) - 0.7y(k-2) + u(k-1) + 0.5u(k-2) \\ &+ v(k) - 0.3v(k-1) + 0.2v(k-2). \end{aligned} \tag{16}$$

Eq. (16) illustrates that the real order of the model is $n_a = n_b = n_d = 2$. The model input is an M sequence, which is a pseudo-random signal. Its characteristic polynomial is: $F(s) = s^6 \oplus s^5 \oplus 1$. The data length is $L = 1200$. In the experiment, we set the standard deviation σ to 0.6, 0.8 and 1 respectively, and compare the order identification results of various methods. For a given order combination (n_a, n_b, n_d), we use the maximum likelihood method to estimate the model parameters and obtain the predicted value $\hat{\boldsymbol{y}}$. In the proposed method, we first use the GPTE method to extract the trend of $y(k)$; the parameter $M_{\max}$ is set to 50, and the trend error is obtained. Second, we calculate the holistic error. The coefficients of the holistic error are selected as $c_1 = 0.2$ and $c_2 = 0.8$, that is, we take the trend error as the main part of the holistic error. Each order(n_a, n_b, or n_d) varies from 1 to 4. In order to facilitate the comparison of the comprehensive error under different order combinations, we introduce the order sum n, which is defined as:

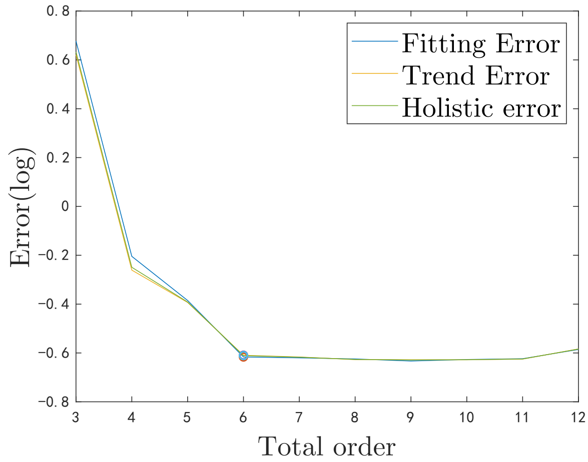
$$n = n_a + n_b + n_d \tag{17}$$

Next, we find the minimum holistic error under each order sum and draw the change curve of the holistic error. Then we find the inflection point and take the corresponding order sum as the order sum of the model. Finally, under the determined order sum, we select the order combination that minimizes the holistic error as the model order. The variation of the fitting error and the holistic error with the order sum is shown in Fig. 2.

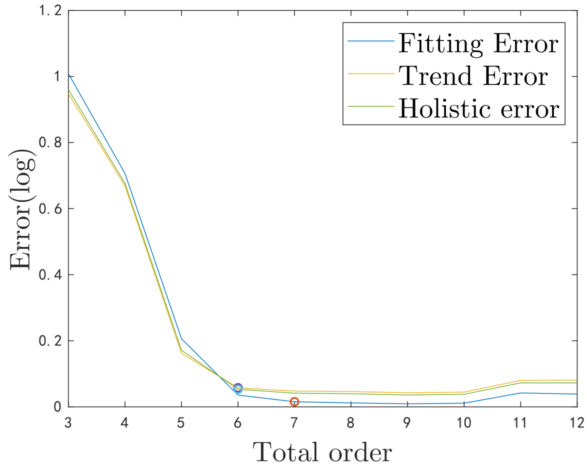
Fig. 2 demonstrates that when the noise standard deviation $\sigma = 0.6$, the inflection points of the fitting error and the holistic error curve clearly appear at the order sum $n = 6$. With the increase of σ, the inflection point of the fitting error variation curve appears at the order sum $n = 7$, while the inflection point of the holistic error variation curve is still at $n = 6$. Especially in the case of $\sigma = 1$, when the order sum of the model exceeds the real order sum $n = 6$, the fitting error is still greatly reduced with the increase of n. However, the holistic error is not significantly reduced. In this case, the method based on fitting error analysis will estimate the inaccurate order sum ($n = 7$).

After determining the order sum as 6, the order combination (n_a, n_b, n_d) is selected as the combination that minimizes the holistic error. The results of fitting error, trend error and holistic error are shown in Table 1. The results in Table 1 illustrate that no matter σ is 0.6,0.8 or 1, the holistic error is the smallest when the order combination is $n_a = n_b = n_d = 2$. Therefore, the final model order combination is selected as $(n_a, n_b, n_d) = (2, 2, 2)$.

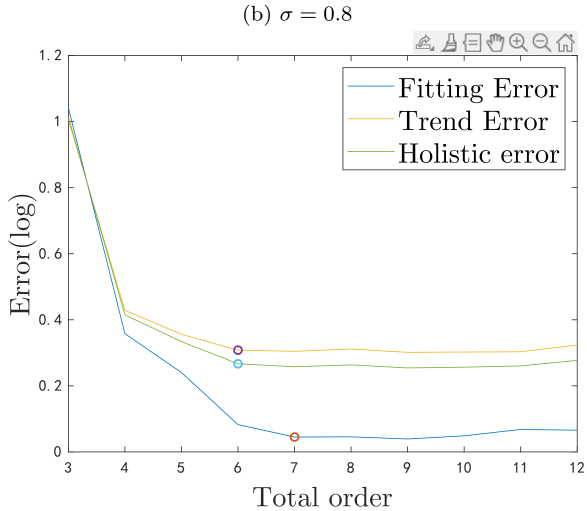
Next, we use the F-test method to select the model order of ARMAX. We define the loss function as $LV_{n_a=n_b,n_d}$. For the case of $n_a = n_b$, $LV_{n_a=n_b,n_d}$ and the statistic t can be firstly calculated. Then the order n_a and n_b are selected. Finally, the order n_d is determined. Taking $\sigma = 0.6$ as an example, the loss function $LV_{n_a=n_b,n_d}$ and statistics t when $n_a = n_b$ are shown in Table 2. When the risk level α is 0.05, the threshold t_α is 3.00. When $(n_a, n_b, n_d) = (3, 3, 1)$, statistics $t(n_a, n_a + 1; n_d) = 2.448 < 3.00$. Therefore, we select $n_a = n_b = 3$ and then calculate $t(n_d, n_d+1; n_a = n_b)$ when n_d ranges from 1 to 4. When $n_d = 1$, $t(n_d, n_d + 1; n_a = n_b = 3) = -6.665 < 3.00$, so the choice of the order combination is $(n_a, n_b, n_d) = (3, 3, 1)$; When $\alpha = 0.1$, the threshold $t_\alpha = 2.307$; When $(n_a, n_b, n_d) = (2, 2, 2)$, statistics $t(n_a, n_a + 1; n_d) = -4.732 < 2.307$. Thus, $n_a = n_b = 2$. When $n_d = 2$, $t(n_d, n_d + 1; n_a = n_b = 2) = -1.497 < 2.30$, thus the order combination is $(n_a, n_b, n_d) = (2, 2, 2)$. We use the same method to compare the loss function $LV_{n_a=n_b,n_d}$ and the statistic t when σ is 0.8 or 1. When the risk level $\alpha = 0.05$ or 0.1, in the case of $\sigma = 0.8$, the selected order combination is $(n_a, n_b, n_d) = (1, 1, 3)$; in the case of $\sigma = 1$, the selected order combination is $(n_a, n_b, n_d) = (2, 2, 3)$. Therefore, the results of the F-test method may be greatly affected by the selection of risk levels and noise.



(a) $\sigma = 0.6$



(b) $\sigma = 0.8$



(c) $\sigma = 1$

Fig. 2. The change of the fitting error, the trend error and the holistic error.

Finally, AIC is used to select the model order. For the ARMAX model, the AIC function is described as follows:

$$\mathrm{AIC}(n_a, n_b, n_d) = -2L\log\hat{\sigma}_v^2 + 2(n_a + n_b + n_d) \tag{18}$$

For different order combinations (n_a, n_b, n_d), the values of the corresponding AIC functions are calculated respectively. Taking $\sigma = 0.6$ as an example, the results are shown in Table 3. It can be seen that when $(n_a, n_b, n_d) = (3, 3, 3)$, the AIC value is the smallest, which

Table 1. Fitting error (FE), Trend Error (TE) and Holistic Error (HE)

Model order (n_a, n_b, n_d)	$\sigma = 0.6$			$\sigma = 0.8$			$\sigma = 1$		
	FE	TE	HE	FE	TE	HE	FE	TE	HE
$(1, 1, 4)$	1.5459	1.4390	1.4604	2.0965	1.9693	1.9947	2.1622	2.3509	2.3131
$(1, 2, 3)$	1.026	0.9640	0.9763	1.5640	1.4904	1.5051	1.5935	1.7162	1.6917
$(1, 3, 2)$	0.8332	0.8105	0.8150	1.3586	1.3739	1.3709	1.4455	1.5679	1.5434
$(1, 4, 1)$	0.8008	0.8063	0.8052	1.3886	1.4367	1.4271	1.5880	1.8496	1.7973
$(2, 1, 3)$	0.7196	0.6867	0.6933	1.2092	1.1697	1.1776	1.2460	1.3775	1.3512
$(2, 2, 2)$	0.5398	0.5436	**0.5429**	1.0362	1.0592	**1.0546**	1.0863	1.3688	**1.3123**
$(2, 3, 1)$	0.6469	0.6482	0.6480	2.0972	2.0959	2.0962	1.3830	1.5848	1.5444
$(3, 1, 2)$	0.6164	0.6017	0.6047	1.2533	1.2025	1.2127	1.1558	1.3607	1.3197
$(3, 2, 1)$	0.5529	0.5511	0.5515	1.0605	1.0647	1.0638	1.2081	1.4226	1.3797
$(4, 1, 1)$	0.6403	0.6306	0.6325	1.2101	1.1907	1.1946	1.2267	1.4888	1.4364

Table 2. The loss function and statistics when $\sigma = 0.6$

n_d	1				2			
$n_a = n_b$	1	2	3	4	1	2	3	4
$LV_{n_a=n_b,n_d}$	2362.471	816.366	645.676	643.035	1994.673	647.803	652.978	642.117
$t(n_a, n_a+1; n_d)$	1132.544	157.822	**2.448**	-	1243.324	-4.732	10.081	-
$n_a = n_b$	3				2			
n_d	1	2	3	4	1	2	3	4
$LV_{n_a=n_b,n_d}$	645.676	652.978	637.195	723.768	816.366	647.803	649.434	740.214
$t(n_d, n_d+1; n_a = n_b)$	**-6.665**	14.763	-71.291	-	155.084	**-1.497**	-73.094	-

is -741.602. Therefore, the order combination selected by AIC is $(n_a, n_b, n_d) = (3, 3, 3)$. We use the same method to calculate the the AIC value when $\sigma = 0.8$ and 1. When $\sigma = 0.8$, the selected order combination is $(n_a, n_b, n_d) = (4, 3, 2)$; when $\sigma = 1$, the selected order combination is $(n_a, n_b, n_d) = (3, 2, 4)$.

The results demonstrate that the F-test method and AIC are greatly affected by noise when selecting the model order. On the contrary, the proposed order selection method can effectively reduce the impact of noise, and assures the accuracy of the selected model order.

After selecting the model order, we estimate the parameters by the maximum likelihood method. The estimation results are shown in Table 4. The results show that the maximum likelihood method can accurately estimate the model parameters under different noise levels.

5. CONCLUSION

In this paper, we proposed a model order selection method based on trend error criterion. Different from the existing statistical-testing methods and information-theoretic methods, the proposed method selects the model order according to the change of holistic error, which consists of fitting error and trend error. The efficiency of the proposed method is tested on a numerical example of the ARMAX model. Compared with the existing statistical-testing methods and information-theoretic methods, the proposed method is less susceptible to noises, and the selected model order is more accurate.

REFERENCES

Besanon, G., Voda, A., Becq, G., and Machado, M.M.P. (2018). Order and parameter identification for a non-integer-order model of an EEG system. *IFAC-PapersOnLine*, 51(15), 772–777.

Borjas, S.D.M. and Garcia, C. (2005). Modeling of FCC using identification methods based on prediction error and sub-space. *IEEE Latin America Transactions*, 2(2), 108–113.

Casas, R.A., Bitmead, R.R., Jacobson, C.A., and Johnson, C.R. (2002). Prediction error methods for limit cycle data. *Automatica*, 38(10), 1753–1760.

Ding, J. and Ding, F. (2008). The residual based extended least squares identification method for dual-rate systems. *Computers & Mathematics with Applications*, 56(6), 1479–1487.

Febrianti, R., Widyaningsih, Y., and Soemartojo, S. (2021). The parameter estimation of logistic regression with maximum likelihood method and score function modification. *Journal of Physics: Conference Series*, 1725(1), 012014.

Fu, G., Zhang, L., Fu, J., Gao, H., and Jin, Y. (2018). F test-based automatic modeling of single geometric error component for error compensation of five-axis machine tools. *Springer London*, (9), 4493–4505.

Herpen, R.V., Oomen, T., and Bosgra, O. (2011). A robust-control-relevant perspective on model order selection. In *American Control Conference*.

Kabaila, P. (2005). Assessment of a preliminary f-test solution to the behrens-fisher problem. *Communications in Statistics. A, Theory and Methods*, 34(2), p.291–302.

Table 3. The AIC value of different order combinations when $\sigma = 0.6$

n_d	1				2			
n_a \ n_b	1	2	3	4	1	2	3	4
1	818.864	355.434	-18.117	-254.620	617.791	137.452	-207.015	-406.087
2	-237.311	-452.257	-510.756	-487.772	-342.833	-727.789	-720.396	-713.178
3	-395.182	-699.021	-729.735	-729.892	-568.670	-647.230	-714.240	-703.904
4	-523.017	-730.047	-733.468	-730.653	94.492	-707.110	-718.199	-730.368
n_d	3				4			
n_a \ n_b	1	2	3	4	1	2	3	4
1	555.729	42.287	-322.182	-489.555	534.729	9.083	-363.309	-527.186
2	-382.889	-722.772	-723.325	-711.279	-78.344	-563.766	-491.532	-553.124
3	-414.309	-647.064	**-741.602**	-732.195	-169.440	-499.130	-586.727	-638.672
4	-291.114	-708.923	-724.231	-726.242	-237.333	-614.103	-667.916	-678.827

Table 4. Parameters estimation results under different noise levels

Parameter	a_1	a_2	b_1	b_2	d_1	d_2
True value	-1.7	0.7	1	0.5	-0.3	0.2
$\sigma = 0.6$	-1.684	0.684	0.968	0.529	-0.320	0.159
$\sigma = 0.8$	-1.682	0.684	1.022	0.509	-0.315	0.229
$\sigma = 1$	-1.678	0.678	0.928	0.556	-0.266	0.224

Keogh, E.J. and Pazzani, M.J. (1998). An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. In *KDD*, volume 98, 239–243.

Keshani, S. and Masoumi, N. (2021). The art of piecewise linear approximation in MMSE estimator for most accurate and fast frequency extraction in DIFM receivers. *IEEE Transactions on Instrumentation and Measurement*, 70, 1–9.

Lee, C.B. (1995). Estimating the number of change points in a sequence of independent normal random variables. *Statistics & Probability Letters*, 25(3), 241–248.

Liu, Q., Shang, C., and Huang, D. (2021). Efficient low-order system identification from low-quality step response data with rank-constrained optimization. *Control Engineering Practice*, 107, 104671.

Liu, Q., Shang, C., Liu, T., and Huang, D. (2023a). Efficient relay autotuner of industrial controllers via rank-constrained identification of low-order time-delay models. *IEEE Transactions on Control Systems Technology*, 31(4), 1787–1802.

Liu, Z., Chen, X., Xu, J., Zhang, C., Xu, L., Guo, N., Zhao, L., and Yan, R. (2023b). High-accurate robust total variation denoising algorithm with adjustable exponential upper bound function for micro-thrust measurement. *IEEE Transactions on Instrumentation and Measurement*, 72, 1–18.

Madsen, L. (2009). Maximum likelihood estimation of regression parameters with spatially dependent discrete data. *Journal of Agricultural, Biological, and Environmental Statistics*, 14, 375–391.

Michael, Lokshin, Zurab, and Sajaia (2004). Maximum likelihood estimation of endogenous switching regression models. *The Stata Journal*, 4(3), 282–289.

Nguyen, D.H. and Ohtsu, K. (2000). An adaptive optimal autopilot using the recursive prediction error method. *IFAC Proceedings Volumes*, 33(21), 185–190.

Ninomiya, Y. (2005). Information criterion for Gaussian change-point model. *Statistics & Probability Letters*, 72(3), 237–247.

Varanasi, S.K. and Jampana, P.V. (2016). Parameter estimation and model order identification of LTI systems. *IFAC-PapersOnLine*, 49(7), 1002–1007.

Villez, K. (2015). Qualitative path estimation: A fast and reliable algorithm for qualitative trend analysis. *AIChE Journal*, 61(5), 1535–1546.

Villez, K., Venkatasubramanian, V., and Rengaswamy, R. (2013). Generalized shape constrained spline fitting for qualitative analysis of trends. *Computers & Chemical Engineering*, 58, 116–134.

Weerts, H.H., Van den Hof, P.M., and Dankers, A.G. (2018). Prediction error identification of linear dynamic networks with rank-reduced noise. *Automatica*, 98, 256–268.

Yang, C., Le, Bouquin Jeannes, R., Bellanger, J.J., and Shu, H. (2013). A new strategy for model order identification and its application to transfer entropy for EEG signals analysis. *IEEE Transactions on Biomedical Engineering*, 60(5), 1318–1327.

Zhou, B. and Ye, H. (2016). A study of polynomial fit-based methods for qualitative trend analysis. *Journal of Process Control*, 37, 21–33.

Zhou, B., Ye, H., Zhang, H., and Li, M. (2017). A new qualitative trend analysis algorithm based on global polynomial fit. *AIChE Journal*, 63(8), 3374–3383.