

# Non-Gaussian System Identification Based on Improved Estimation of Distribution Algorithm

Jinglin Zhou, Yiqing Jia, Han Zhang, Jing Wang, Haijiang Zhu

College of Information Science & Technology, Beijing University of Chemical Technology, Beijing, 100029, China

E-mail: [jinglinzhou@mail.buct.edu.cn](mailto:jinglinzhou@mail.buct.edu.cn); [jwang@mail.buct.edu.cn](mailto:jwang@mail.buct.edu.cn); [zhuhj@mail.buct.edu.cn](mailto:zhuhj@mail.buct.edu.cn)

**Abstract:** In this paper, an improved estimation of distribution algorithm (EDA) is proposed and applied to the identification of ARMA model parameters. The system parameter identification problem is transformed into the optimization problem in high dimensional parameter space. Based on the traditional EDA algorithm, the parameters of preliminary estimation and data selection are added to improve the speed of searching and optimization precision. Because the mean and the variance are not enough to describe the uncertainty of non-Gaussian system, the entropy is regarded as fitness value to achieve the parameter identification of non-Gaussian system. Finally, an example of improved EDA identification is given to illustrate the effectiveness of the proposed approach.

**Key Words:** Non-Gaussian System, EDA, Parameter Estimation

## 1 Introduction

Considering the unpredictability of the control system model under non-Gaussian disturbance in advance, the estimation of the system model from the data of daily operation becomes difficult and challenging to some extent. Solving the problem of the accuracy of model identification requires an appropriate criterion, which is an indicator of system model identification. The model identification based on the mean square error (MSE) criterion and the corresponding least squares method has been widely applied to the model parameter identification of Gaussian system. However, in practice the realization of disturbance may not confirm to the Gaussian distribution, the information gathered by the low order statistics is not sufficient for the non-Gaussian system parameter identification. Unlike Gaussian variables, all distributions of random variables are included in the first and second moments. So the MSE method is inadequate to cope with the non-Gaussian system identification. To solve the stochastic systems with non-Gaussian disturbance, the stochastic distribution control theory was proposed [1]. Then, a more general measure of uncertainty, namely the entropy, should be used to characterize the uncertainty of the tracking error. It is known that all the stochastic information is included in the complete probability density function (PDF). Therefore, by minimizing the entropy instead of the mean square error, all higher order moments (not only the second one) can be minimized. Entropy has been widely used in information, thermodynamics and control theory [2], [3], [4]. In [5], entropy was used as a measure for the average information contained in the PDF of the estimation errors using the filters.

In recent years, with the development of science and deepening of the advanced control theory research,

optimization methods and intelligent control theory are in-depth application to the control field. Some intelligent optimization algorithms, such as genetic algorithms [6],[7], neural networks [8],[9], particle swarm optimization [10],[11], differential evolution algorithms [12] and distribution estimation algorithms [13] have been used in system parameter identification, which are characterized by concise programming and facile application. These methods can be used to deal with the nonlinear and non-Gaussian system parameter identification, which is superior to the traditional optimization algorithm for parameter identification of linear system and Gaussian system. Thus, the non-Gaussian system identification based on the intelligent optimization algorithms is valuable. In this paper, the estimation of distribution algorithm is improved to enhance its global optimization ability and to accelerate the convergence speed. Then, it is applied to ARMA model parameter optimization estimation. Based on the initial parameter space, the optimal ARMA model parameters and unknown noise distribution are obtained by global search. The paper is organized as follows: In Section 2, review the traditional EDA algorithm and point out the defects. In Section 3, the improvement measures are put forward. In Section 4, a simulation case is conducted.

## 2 Estimation of Distribution Algorithm

The concept of Estimation of Distribution Algorithm (EDA) was first proposed by the scholars in 1996 [14]. In the following few years, the algorithm developed rapidly, which became the hotspot in the field of evolutionary computation. There is a significant difference between the EDA algorithm and the genetic algorithm (GA). EDA predicts the optimal region by searching spatial sampling and statistical learning, and then produces excellent new individuals, which differs from GA uses crossover and mutation to generate new individuals. Compared with GA's evolutionary approach based on micro-level of gene, EDA adopts the evolutionary method of macro-level based on search space and has stronger global searching ability and faster convergence speed. The probability distribution model is used to describe

---

\* This work was supported in part by NSFC (Grant No. 61473025, and 61573050), the open-project grant funded by State Key Laboratory of Management and Control for Complex System at Institute of Automation, CAS (20160107).

the distribution of candidate solutions in the search space. The statistical model is used to construct a probability model to describe the solution distribution from the macroscopic point of view and then, the probability model is randomly sampled to generate new population of evolution. Under this situation, the population is initialized with distribution estimation algorithm. In each iteration the fitness value of each individual is obtained and sorted according to the evaluation function. Then a part of individuals with better fitness value are selected to make up the advantage group. And the probability distribution model which the subject obeyed is estimated according to the dominant group, such as the Gaussian distribution model. Furthermore, a random sampling of the estimated probability distribution model is used to generate some new individuals to replace some of the individuals with poor fitness values in the initial population to form a new generation population. When satisfying the iteration termination condition, the iteration of this algorithm will be terminated and finally the optimal outcomes obtained by employing EDA are the best fitness value of the current population.

The specific process shown as follows.

Step 1: Initialize the population.

Step 2: Select the dominant group.

Step 3: Construct the probability model.

Step 4: Random sampling.

Step 5: Generate new populations.

Step 6: Judge the termination condition. If yes, end; otherwise, turn to Step 2.

However, it is worth noting that process of parameter identification is still affected by many factors in practical applications of the estimation of distribution algorithm.

1) The setting of the initialization parameter space triggers a great impact on the recognition effect. In the case of unknown parameters, the initial value will affect the convergence rate if it is too large, especially for high-dimensional space cannot guarantee uniform sampling. Therefore, before initializing the parameter space, a preliminary estimation of the model parameters is required to ensure that the parameter space covers the real parameters as much as possible.

2) Owing to the non-uniformity of the high-dimensional space sampling, randomly generated seeds are likely to deviate from the real parameters even diverge the system. In order to obtain more appropriate seeds, the abnormal value needs to be filtered out before the fitness value is calculated to accelerate the convergence rate.

3) The error entropy is regarded as the fitness value. When the entropy is calculated, the entropy of two generations cannot be subtracted or the minimum entropy value is not unique because of the mean shift. In order to avoid this situation, a unified standard is needed to make the entropy calculation more normalized.

From the above discussions it is clear that the estimation of distribution algorithm should be improved.

### 3 Improvement Measures

#### 3.1 Preliminary Estimation

Since the initialization parameter space has a pivotal role on the parameter identification, it is necessary to make a

preliminary estimation of the model parameters before initializing the population to ensure that the parameter space covers the real parameters as much as possible.

For non-Gaussian systems, there is no better parameter initialization method at present and so we adopt the method of parameter initialization for Gaussian system. The parameter estimation method of ARMA model is used to initialize the range of parameters to be estimated. In order to illustrate the process at length, we will elaborate the parameter estimation method of ARMA model for the Gaussian systems summarized as below.

Consider an ARMA time series described as follows.

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

where  $y(k)$  is the output of the system, the disturbance  $v(k)$  is an uncorrelated random noise sequence with zero mean and variance  $\sigma_v^2$ .

And

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

defined

$$\theta = [a_1, a_2, \cdots, a_{n_a}, d_1, d_2, \cdots, d_{n_d}]^T \quad (3)$$

In this paper, a recursive maximum likelihood (RML) [15] estimation method is used to estimate the parameters. Thus the maximum likelihood problem of the model parameter is to find the parameter  $\theta$  to make the objective function reach the minimum.

$$J(\theta) \Big|_{\hat{\theta}_{ML}} = \frac{1}{2} \sum_{k=1}^L v^2(k) \Big|_{\hat{\theta}_{ML}} = \min \quad (4)$$

In which  $\hat{\theta}_{ML}$  is the estimate of  $\theta$ , and  $v(k)$  satisfying

$$v(k) = [D(z^{-1})]^{-1} [A(z^{-1})y(k)] \quad (5)$$

If  $v(k)$  is Taylor-expanded at the  $\hat{\theta}_{ML}$  point, it can be approximated as

$$v(k) \approx v(k \Big|_{\hat{\theta}_{ML}}) + \frac{\partial v(k)}{\partial \theta} \Big|_{\hat{\theta}_{ML}} (\theta - \hat{\theta}_{ML}) \quad (6)$$

Then the filter value of  $y(k)$  and  $\hat{v}(k)$  is  $y_f(k)$  and  $\hat{v}_f(k)$ , respectively.

$$\begin{cases} y_f(k) = [\hat{D}(z^{-1})]^{-1} y(k) \\ \hat{v}_f(k) = [\hat{D}(z^{-1})]^{-1} \hat{v}(k) \end{cases} \quad (7)$$

$$\begin{aligned} h_f(k) &= [-y_f(k-1), \cdots -y_f(k-n_a), \\ &\quad \hat{v}_f(k-1), \cdots \hat{v}_f(k-n_d)]^T \end{aligned} \quad (8)$$

The recursive maximum likelihood estimation algorithm can be described as formula (9).

Based on the estimation of the parameters of the Gaussian system, we can get the estimate of variance  $\hat{\sigma}_v$  and the estimated value of the parameter  $\hat{\theta}$ , and then we use the  $\hat{\theta} \pm 3\hat{\sigma}_v$  as the initialized range of the parameter space.

$$\begin{cases}
\hat{\theta} = \hat{\theta}(k-1) + K(k)\hat{v}(k) \\
K(k) = P(k-1)h_f^T(k)[h_f^T(k)P(k-1)h_f(k) + I]^{-1} \\
P(k) = [I - K(k)h_f^T(k)]P(k-1) \\
\hat{v}(k) = y(k) - h^T(k)\hat{\theta}(k-1) \\
h(k) = [-y(k-1), \dots, -y(k-n_a), \\
\quad \hat{v}(k-1), \dots, \hat{v}(k-n_c)]^T \\
h_f(k) = [-y_f(k-1), \dots, -y_f(k-n_a), \\
\quad \hat{v}_f(k-1), \dots, \hat{v}_f(k-n_c)]^T \\
y_f(k) = [\hat{D}(z^{-1})]^{-1}y(k) \\
\hat{v}_f(k) = [\hat{D}(z^{-1})]^{-1}\hat{v}(k)
\end{cases} \quad (9)$$

### 3.2 Data Selection

Since the range of parameter estimation and parameter space dimension are relatively large, the data generated by using the uniform distribution of each subspace directly may result in large estimation bias. In the identification of the distribution estimation algorithm, we adopt the unknown parameters as seeds. In order to obtain appropriate seeds, adding the screening conditions is an effective measure to eliminate the seeds with large deviations.

Because the mean of the noise is assumed to be zero, the mean of the residuals can be regarded as the screening condition. It can be described as follows.

$$\begin{aligned}
e_k &= y_k - h^T \theta \\
\text{mean}(e_k) &< \varepsilon \quad (\varepsilon \rightarrow 0)
\end{aligned} \quad (10)$$

If the mean value is less than  $\varepsilon$ , it will be retained as good seed and becomes a member of the new parameter space.

### 3.3 Selection of Fitness Value

If  $X$  is a random variable with length  $M$  discrete values, and corresponding distribution probability  $P = (p_1, p_2, \dots, p_M)$ , then the Renyi's entropy denoted by  $H_\alpha(X)$  is

$$H_\alpha(X) = \frac{1}{1-\alpha} \ln \sum_{i=1}^M p_i^\alpha \quad (11)$$

In this paper, we take the error entropy as the fitness value, when the error entropy reaches the minimum while the parameter reaches the optimum.

It can be expressed as

$$\Phi^{opt} = \arg \min_{\Phi \in \Psi} H(e) \quad (12)$$

where  $\Psi$  denotes the parameters space.

When computing the probability, we assume that  $X$  is sampled at a fixed time and  $N$  valid data are collected. All of these points fall within the interval  $S = [W_1, W_M]$ ,  $W_1$  and  $W_M$  are the left and right boundaries respectively. We divide the interval uniformly into  $M$  consecutive and non-overlapping subintervals  $(s_1 \cup s_2 \cup s_3 \dots \cup s_M)$ . The central value of each interval is denoted as  $w_i (i = 1, 2, \dots, M)$ . And  $s_{i(1 < i < M)} = [w_i - \delta, w_i + \delta], (w_i - \delta = w_{i-1} + \delta, w_i + \delta = w_{i+1} - \delta)$ . Then statistics the number of sampled data falling into each

subinterval,  $N = (l_1, l_2, \dots, l_M)$  and these frequencies are approximated as the probability of  $w_i (i = 1, 2, \dots, M)$ , the corresponding probability distribution is as follows,

$$P_i = \left( \frac{l_1}{N}, \frac{l_2}{N}, \dots, \frac{l_M}{N} \right) = (p_1, p_2, \dots, p_M), \quad (i = 1, 2, \dots, M) \quad (13)$$

In the EDA identification process, we take the error entropy as the fitness value. Since the selection of seeds during the iteration is random, the range of residuals for each generation may be different. If the interval is divided into  $M$  sub-intervals to calculate the value of the entropy, the central value of the deviation will cause that the calculation of entropy is not uniform and the entropy between adjacent generations cannot be subtracted or cannot even converge under such circumstances. In order to solve this problem, we set a uniform interval range and partition the subinterval with the same interval so that the center value is fixed, which ensures the consistency of the entropy calculation.

### 3.4 Algorithm Summary

Combined with EDA algorithm and the proposed improvement measures, the improved EDA algorithm is summarized as follows.

(1) Preliminary estimation. Rough estimation parameters are obtained by recursive maximum likelihood method. Then the task of initializing the parameters space can be accomplished with a uniform distribution using the initial value of parameter as the value range.

(2) Screen seeds. From the first generation randomly select  $R$  group seed from the parameter space and calculate the mean value of the residuals. Remove the seeds whose mean value is more than  $\varepsilon$ , and keep the seeds whose mean value is less than  $\varepsilon$ . If the number of reserved seeds is less than  $R$ , resampling is performed until the number of seeds retained is not less than  $R$ , and then the seeds reserved are collected in new parameter spaces  $\Psi^{(l)}(Q)$  ( $Q \geq R$ ).

(3) Calculate fitness. Randomly create  $R$  individuals from the parameter space  $\Psi^{(l)}(Q)$ ,  $A_l = \{\Phi^{(l)}(1), \Phi^{(l)}(2), \dots, \Phi^{(l)}(R)\}$ . For each parameter vector in  $A_l$ , estimate the corresponding error entropy based on the training data set  $\{y_i\}$ .

(4) Select  $N$  superior individuals  $B_l = \{\Phi^{(l)}, \Phi^{(2)}, \dots, \Phi^{(N)}\}$  based on the cost function and estimate probability density function based on the statistical information extracted from the selected  $N$  individuals.

(5) Calculate the estimated average value of the selected  $N$  parameters and establish the probability models.

(6) Set  $l \leftarrow l + 1$ , and resample  $R$  individuals from the updated PDF.

(7) Go to step2 until the stopping criterion is met.

This process can be represented by Figure 1.

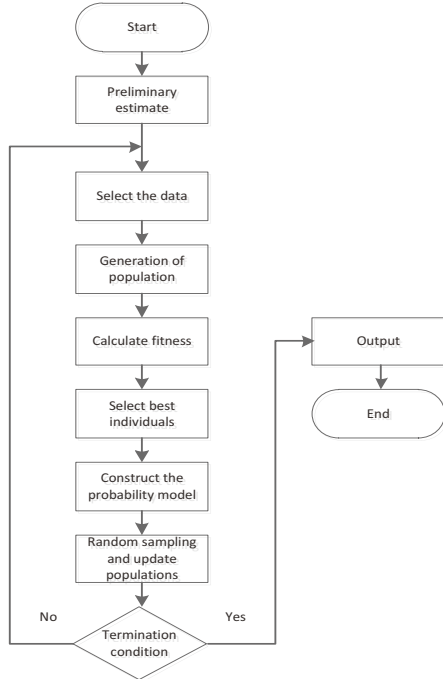


Fig 1. The process of the improved distribution estimation algorithm

#### 4 Simulation

In this section two experiments are given using the improved EDA method and the results are compared with the EDA introduced in the section 2. In order to illustrate the effectiveness of the improved method, consider the following system

$$y(k) = \frac{1 + 1.5z^{-1} + 0.9z^{-2}}{1 - 1.7z^{-1} + 0.7z^{-2}} v(k) \quad (14)$$

The simulation is carried out under different noise disturbance. In all testing cases, the parameters of the two methods are  $N=80$ ,  $R=200$ . And a Gaussian model with diagonal covariance matrix is utilized as the probabilistic model. Let  $H^{(l)} = H(\bar{e}^{(l)})$  denote the error entropy obtained at  $l$  based on the estimated average parameters. The stopping criteria of two methods can be set to  $|H^{(l)} - H^{(l-1)}| < 0.001$ . The selection criteria of the improved EDA algorithm is  $\bar{e}_k < 0.1$ .

(1)  $v(k) \sim N(0, 0.255^2)$ , after using the RML algorithm, the preliminary estimate of model parameters are  $\hat{\theta} = [-1.7510, 0.7520, 0.9201, 0.0572]^T$ .

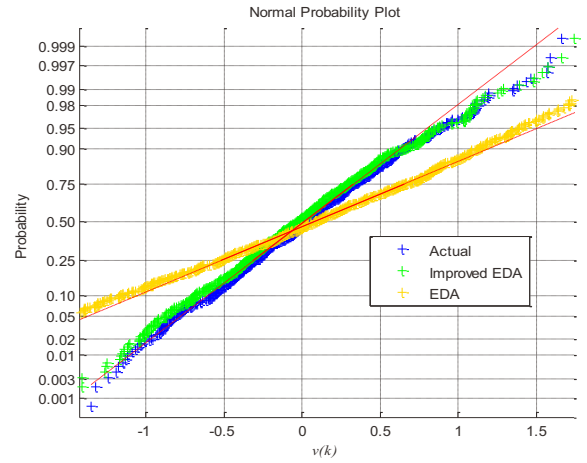


Fig 2. The normal probability plots of the actual and estimated disturbances of the system with Gaussian disturbance.

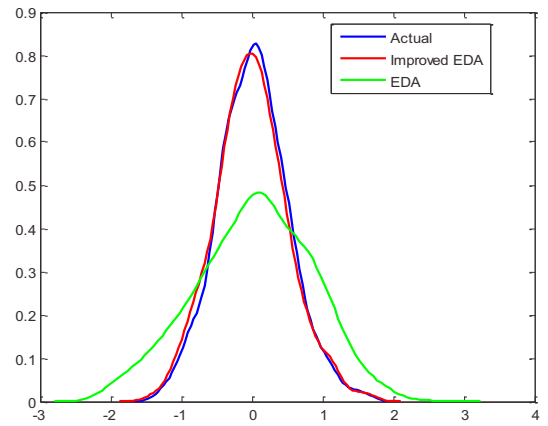


Fig 3. PDF of the actual and estimated disturbance using improved EDA algorithm and EDA algorithm.

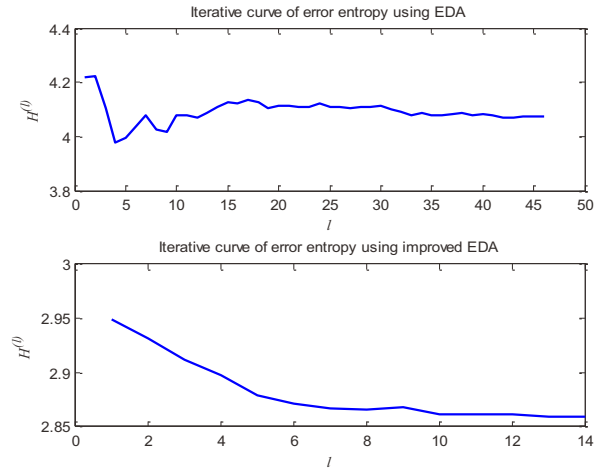


Fig 4. The iterative curve of  $H^{(l)}$  using improved EDA and EDA algorithm.

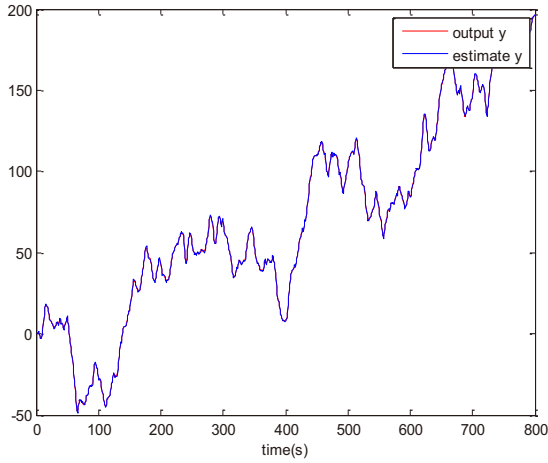


Fig 5. System output curve and its fitting curve under Gaussian disturbance

Table 1: The results of system parameters identification under Gaussian disturbance are obtained according to two methods.

	$a_1$	$a_2$	$b_1$	$b_2$	$\sigma_v^2$
true value	-1.7	0.7	1.5	0.9	0.2474
EDA	-0.4651	-0.5384	1.4831	0.9104	0.6676
improved EDA	-1.7739	0.7736	1.3185	0.6914	0.2732

In the first case, the system with the Gaussian disturbance is identified by the two methods. As shown in Fig.4, after 105 seconds and 15 generations, the error entropy for the population is converged by using the improved EDA algorithm. The estimated variance of the noise sequence is 0.2732, which is close to the real variance 0.2474. To make comparisons, the original EDA identification is also applied here. After 90 seconds and 47 generations, the error entropy for the population is converged, and the estimated variance of the noise sequence is 0.6676. The normal probability plot is a graphical technique to identify substantive departures from normality. If the set of real numbers follows a normal distribution, the normal probability plot will be a straight line. The Fig.2 shows that the both results are Gaussian distributions, but the result of improved EDA algorithm is closer to the true distribution than the original method. For better clarity, regard the disturbance and error sequence as a continuous random variable with their respective nonnegative probability density function. And the kernel density function is used to fit the PDF in Fig.3. Although the original EDA algorithm also estimates the Gaussian distribution, it is clear that there is a big gap from the true distribution. Fortunately, as shown in the Fig.3 the result of improved EDA identification is almost completely coincident with the actual distribution. Organize the above simulation results into Table 1. It can be summarized that the improved EDA method can improve the accuracy of parameter identification and avoid the premature convergence of EDA after the rough estimation parameters are obtained.

$$(2) v(k) \sim f(x; \alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

where  $\alpha = 2$ ,  $\beta = 9$ , after using the RML algorithm, the preliminary estimate of model parameters are  $\hat{\theta} = [-1.7502, 0.7498, 0.7729, -0.0484]^T$ .

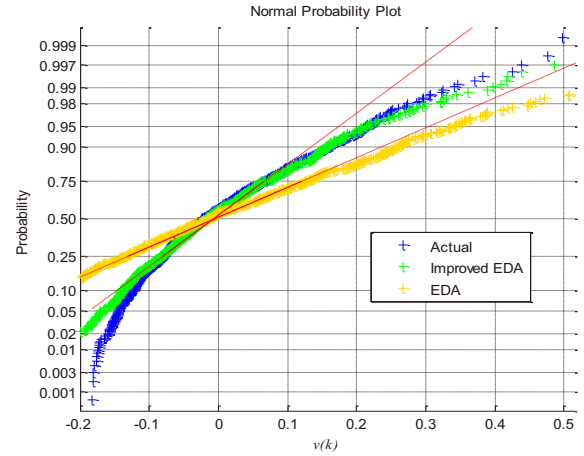


Fig 6. The normal probability plots of the actual and estimated disturbances of the system with  $\beta$  disturbance.

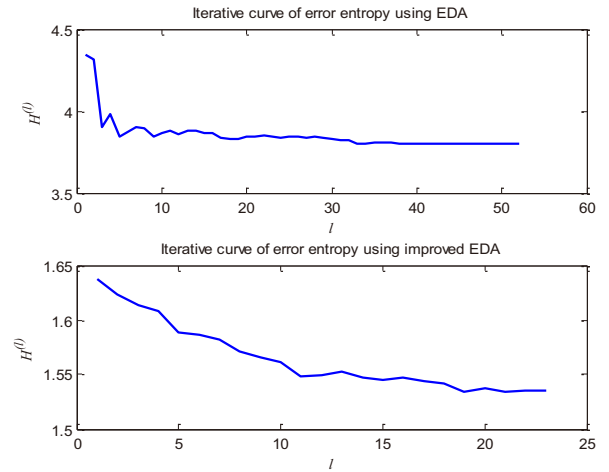


Fig 7. The iterative curve of  $H^{(l)}$  using improved EDA and EDA algorithm.

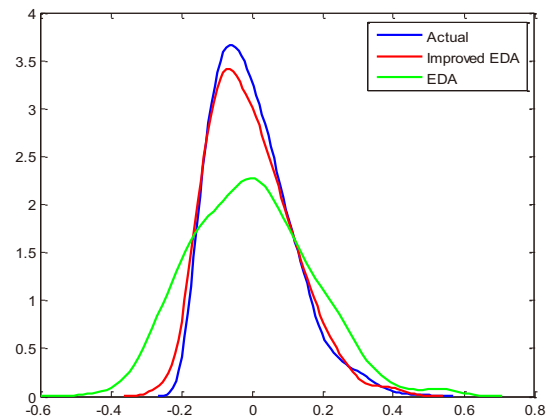


Fig 8. PDF of the actual and estimated  $\beta$  disturbance using improved EDA algorithm and EDA algorithm.



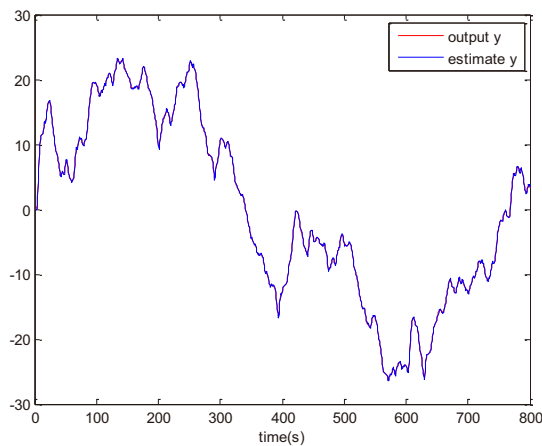


Fig 9. System output curve and its fitting curve under the  $\beta$  disturbance.

Table 2: The results of system parameters identification under  $\beta$  disturbance are obtained according to two methods.

	$a_1$	$a_2$	$b_1$	$b_2$	$\sigma_v^2$
true value	-1.7	0.7	1.5	0.9	0.0128
EDA	-0.3044	-0.6892	1.6657	0.9463	0.0578
improved EDA	-1.7154	0.7189	1.2706	0.6243	0.0139

In the second case, by using the improved EDA method, the system with the non-Gaussian disturbance is identified after the 24 iterations with 59 seconds. Compared with EDA method ended in 53 generations with 100 seconds, the number of iterations is reduced and the convergence speed is accelerated shown in Fig.7. The normal probability plot is also used here to check the identification effect. It is noticed that the disturbance in the tails of the plots bends away from the straight line shown in Fig.6, which illustrates the effectiveness of the algorithm for non-Gaussian distribution identification. But it is clear that the result of the improved EDA method is less deviate from the actual value than the previous method. The Fig.8 shows the PDF of actual and estimated disturbance. Green line depicts that although the EDA algorithm terminates, the effect is not optimal. And through the improved EDA algorithm, the estimation effect shown in Fig.8 has been significantly improved. As shown in Table 2, the real variance of the disturbance is 0.0128, and the estimated values by the two methods are 0.0139 and 0.0578, respectively. From the above description, it can be seen that the improved EDA method enhances the effectiveness of parameter identification for non-Gaussian systems.

The Fig.5 and Fig.9 show that the parameters obtained by this method are optimal and have good fitting effect.

## 5 Conclusions

This paper briefly introduces an improved method based on the traditional identification method with the estimation of distribution algorithm. The simulation results show that the proposed algorithm can accelerate the convergence speed and enhance the recognition effect through parameter

preliminary estimation and data filtering. For the system with non-Gaussian noise, the parameter estimation could achieve a relatively high level of data accuracy and at the same time, the convergence process becomes more stable and rapid. Before initializing the parameter space, a preliminary estimation and data filtering of the model parameters are required to ensure that the parameter space is closer to the real parameters and improve the algorithm's computational efficiency. The error entropy is calculated more standardized by the fixed interval length and the subinterval length. This paper provides a new method for the study of non-Gaussian system identification. In the future, the identification of non-Gaussian nonlinear systems will be further studied and discussed.

## References

- [1] H. Wang, *Bounded Dynamic Stochastic Systems: Modeling and Control*, Springer-Verlag, London, UK, 2000.
- [2] X. B. Feng and K. A. Loparo, Active probing for information in control system with quantized state measurements: A minimum entropy approach, *IEEE Trans. Automat. Control*, vol. 42, no. 2, pp. 216–238, Feb. 1997.
- [3] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.
- [4] A. Renyi, *A Diary on Information Theory*. New York: Wiley, 1987.
- [5] L. Guo and H. Wang, Minimum entropy filtering for multivariate stochastic systems with non-gaussian noises, *IEEE Trans. Automat. Control*, vol. 51, no. 4, pp. 695–700, Apr. 2006.
- [6] Goldberg D E. *Genetic Algorithms in Search, Optimization and Machine Learning*. 1989, xiii.
- [7] Kristinsson K, Dumont G A. System identification and control using genetic algorithms. *IEEE Transactions on Systems Man & Cybernetics*, 1992, 22(5):1033-1046.
- [8] Lu Y, Sundararajan N, Saratchandran P. Performance evaluation of a sequential minimal radial basis function (RBF) neural network learning algorithm. *IEEE Transactions on Neural Networks*, 1998, 9(2):308-318.
- [9] Ye G, Li W, Wan H. Study of RBF Neural Network Based on PSO Algorithm in Nonlinear System Identification. *International Conference on Intelligent Computation Technology and Automation IEEE*, 2015:852-855.
- [10] Eberhart R C, Shi Y. Particle swarm optimization: Development, applications and resources. *Evolutionary Computation, Proceedings of the 2001 Congress on*. 2001:81-86 vol. 1.
- [11] Trelea I C. The particle swarm optimization algorithm: convergence analysis and parameter selection. *Information Processing Letters*, 2003, 85(6):317-325.
- [12] Qin A K, Huang V L, Suganthan P N. Differential evolution algorithm with strategy adaptation for global numerical optimization. *IEEE Transactions on Evolutionary Computation*, 2009, 13(2):398-417.
- [13] Hauschild M, Pelikan M. An introduction and survey of estimation of distribution algorithms. *Swarm and Evolutionary Computation*, 2011, 1(3):111-128.
- [14] Larranaga P, Lozano J A. *Estimation of distribution algorithms: A new tool for evolutionary computation*. Boston: Kluwer Press, 2002.
- [15] Tjalling J. Ypma, *Historical development of the Newton-Raphson method*, Society for Industrial and Applied Mathematics, 1995.