Proceedings of the 6th World Congress on Intelligent Control
and Automation, June 21 - 23, 2006, Dalian, China

A Multi-objective Genetic Programming/ NARMAX Approach to Chaotic Systems Identification

Pu Han, Shiliang Zhou and Dongfeng Wang
Department of Automation, North China Electric Power University, Baoding, 071003, China
zsl0zsl@163.com

***Abstract* - A chaotic system identification approach based on genetic programming (GP) and multi-objective optimization is introduced. NARMAX (Nonlinear Auto Regressive Moving Average with exogenous inputs) model representation is used for the basis of the hierarchical tree encoding in GP. Criteria related to the complexity, performance and chaotic invariants obtained by chaotic time series analysis of the models are considered in the fitness evaluation, which is achieved using the concept of the non-dominated solutions. So the solution set provides a trade-off between the complexity and the performance of the models, and derived model were able to capture the dynamic characteristics of the system and reproduce the chaotic motion. The simulation results show that the proposed technique provides an efficient method to get the optimum NARMAX difference equation model of chaotic systems.**

Index Terms* - *Genetic programming, Multi-objective optimization, NARMAX models, Chaotic time series analysis, Chaotic system identification.

I. INTRODUCTION

Chaos provide a mechanism and explanation that simple deterministic laws can generate complex behaviors, but chaotic system identification remains a difficult task due to its dissipation and sensitivity to initial conditions properties, and no initial information is generally available about the system structure.

The nonlinear difference-equation model known as the NARMAX model can provide a very concise representation for nonlinear systems.

Up to now, several orthogonal estimators have been developed for identifying polynomial NARMAX systems. These methods have provided an efficient combination of structure selection with parameter estimation. However, because the error reduction depends on the order in which the candidate terms are orthogonal in the regression function, this may produce misleading information regarding the significance of individual terms. A forward regression orthogonal algorithm[1] has been proposed in order to overcome the problem presented by the orthogonal least squares methods. Nevertheless, this algorithm does not guarantee a minimal structure for the model, thus providing a suboptimal solution.

Parameters estimated by Orthogonal Least Squares (OLS) minimize residual variance, but can't guarantee the corresponding model is parsimonious and has similar characteristic of its attractor. That is to say, the derived model should satisfy, in some optimal fashion, a number of objectives involving, for example, performance and model complexity. Here, we describe a multi-objective optimization approach to nonlinear system identification based on polynomial NARMAX representations, and introduce structure selection method for polynomial NARMAX models which take advantage of the GP and the OLS algorithm. And we discussed the use of chaotic invariants for chaotic system identification in a multi-objective framework. Our work has drawn lessons from forefathers' research results (More details refer [2][3][4][5]).

The paper is organized as follows. In Section 2 the polynomial NARMAX models, in Section 3 the Orthogonal Least Squares are presented, in Section 4 a modified GP is presented which is suitable for polynomial NARMAX models, and Section 5 presents a brief introduction of multi-objective optimization and the definition of Pareto-optimality. Finally in Section 6 the application examples are shown.

II. POLYNOMIAL NARMAX MODELS

The NARMAX (Non-linear Auto Regressive Moving Average with eXogenous inputs) model represents the extension of the well-known ARMAX model to the nonlinear case, and is defined as:

$$y(k) = F^l(y(k-1),\cdots,y(k-n_y),u(k-1),\cdots, u(k-n_u),e(k-1),\cdots,e(k-n_e)) + e(k) \quad (1)$$

where $y(k)$is the output, $u(k)$ are the controlled inputs (i.e., exogenous variables), $e(k)$ is the noise input, and $\{e(k)\}$ is assumed to be a white sequence. n_y , n_u and n_e are their associated maximum lags, and $F(\bullet)^l$ is some unknown nonlinear function of degree l. If $e(k)$ is unknown, equation (1) is rewritten as:

$$y(k) = F^l(y(k-1),\cdots,y(k-n_y),u(k-1),\cdots,u(k-n_u), e(k-1),\cdots,e(k-n_e)) + \varepsilon(k) \quad (2)$$

The polynomial NARMAX model is represented as:

$$y(t) = \hat{y}(k) + \varepsilon(k) \quad (3)$$

$$\hat{y}(k) = \sum_{i=1}^{M} \theta_i x_i(t) \quad (4)$$

where $x_i(t)$includes all the output, input and noise terms as well as all possible combinations up to degree l and up to time k-1. The coefficients of such terms are in the parameter vector $\boldsymbol{\theta}$. $\varepsilon(k)$ is the residual at time k.

$$M = \sum_{i=0}^{l} n_i, n_0 = 1 \quad (5)$$

$$n_i = n_{i-1}\frac{(n_y + n_u + n_e + i - 1)}{i}, i = 1\cdots l \quad (6)$$

$$x_1(t) = 1 \tag{7}$$

$$x_i(t) = \prod_{j=1}^{p} y(t-n_{yj}) \prod_{k=1}^{q} u(t-n_{uk}) \prod_{m=1}^{r} e(t-n_{em}) \tag{8}$$

$$i = 2, \cdots, n, p, q, r \geq 0, 1 \leq p+q+r \leq l \tag{9}$$

$$1 \leq n_{yi} \leq n_y, 1 \leq n_{uk} \leq n_u, 1 \leq n_{em} \leq n_e \tag{10}$$

The compact matrix form corresponding to polynomial NARMAX models (3) is

$$Y = X\theta + \varepsilon \tag{11}$$

Where $\boldsymbol{Y}$ =[y(1), y(2), …, y(n)]$^{\mathrm{T}}$, $\boldsymbol{X}$ (k)= [x_1(k), x_2(k), …, x_n(k)], $\boldsymbol{\theta}$ =[θ_1(k) θ_2(k) , …, θ_n(k)].

In this sort of model representation, the complexity of the identification process increases with the degree of nonlinearity and lag of the input, output and cross-coupled terms. And it has the appealing feature to be linear-in-the-parameters, so that a straight implementation of least-squares techniques can be applied.

When the system to identify is nonlinear a direct estimation based on (2) generally leads to an over-parameterized model. If the values of n_y, n_u, n_e and l are increased to obtain a good accuracy, an excessively complex model will result together with a numerical ill-conditioning. A procedure is needed to select terms from the large set of candidates to provide a parsimonious model.

III. Orthogonal Least Squares

The OLS algorithm is an effective algorithm that allows each coefficient in the model to be estimated. At the same time, it provides an indication of the contribution that each term makes to the system output, using the error reduction ratio (ERR) which is a measure of the decrease in the variance of output by a given term.

The LS method minimizes the square error between measured and calculated output to get optimal parameters

$$\chi^2 = \sum_{k=1}^{N} \left(y(k) - \hat{y}(k) \right)^2 \tag{12}$$

Where N is the number of data-points. Hence the optimal parameter vector $\boldsymbol{\theta}$, where the χ^2 is minimal, can be calculated by LS method:

$$\boldsymbol{\theta} = \left(X^T X \right)^{-1} XY \tag{13}$$

Assume that the regression matrix $\boldsymbol{X}$ can be orthogonally decomposed as $\boldsymbol{X}$ = $\boldsymbol{WA}$, where $\boldsymbol{A}$ is an M×M triangular matrix ($\boldsymbol{A}_{i,j}$=0, if i>j) and $\boldsymbol{W}$ is an N×M matrix with orthogonal columns in the sense that $\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}$= $\boldsymbol{D}$ diagonal matrix. (N is the length of $\boldsymbol{Y}$ vector, M is the number of regressors.) After that one can calculate the OLS auxiliary parameter vector $\boldsymbol{g}$ as

$$\boldsymbol{g} = \boldsymbol{D}^{-1}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{Y} \tag{14}$$

The output variance ($\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}$)/n can be explained as

$$Y^T Y = \sum_{i=1}^{M} g_i^2 w_i^T w_i + e^T e \tag{15}$$

Where $\boldsymbol{g}_i$ is the corresponding element of the OLS solution vector.

Thus the error reduction ratio, [err]$_{\mathrm{i}}$ of F_i term can be expressed as

$$\left[eer \right]_i = \frac{g_i^2 w_i^T w_i}{Y^T Y} \tag{16}$$

This ratio offers a simple mean for ordering the terms, and it can be easily used to select the significant model terms.

IV. Genetic Programming for Polynomial NARMAX Models

In this section we propose a structure selection method for Polynomial NARMAX Models. This method is based on the Genetic Programming and OLS algorithm. An advantage of this approach is that fewer assumptions have to be made regarding the final form of the model as the algorithm appeared to be able to automatically select the appropriate input variables as well as discover the model structure and parameters simultaneously (this process is known as symbolic regression). This means that GP has advantages over other algorithms as it can perform optimization at a structural level.

Some constraints is introduce to crossover operator, mutation operator, generating tree, So that every individual in GP expresses a corresponding NARMAX model. These constraints and definition of function set, definition of terminal set is explained as follows:

1） Binary tree is used to express individuals, terminal set T contained the following arguments {x_1(k),…, x_n(k)}, consists of all linear terms in the output, input and noise (if it is the case of NARMAX structures) with maximum lags defined by n_y , n_u and n_e, respectively. Function set F is defined as {+,*}.

2） In the process for generating trees, if one inner node is "*" type, its subtree can't include "+" type node.

3） While implementing crossover operator, types of two nodals chosen at random must meet the following rule: they both ara "+" type node, both are "*" type node or one is leaf node the other is "*" type node. This is shown in Fig. 1.

4） While implementing mutation operator, if mutation node is "*" type, then inner nodes of subtree generated at random can only be "*" type. If mutation node or its father node is "*" type, then subtree generated at random must meet rule 1. This is shown in Fig. 2.

5） Pruning operator: Before OLS evaluation, If leaf node's numbers of every type are all the same in T_i and T_j subtree, then delete T_j subtree. And after OLS evaluation if corresponding [err]$_i$ of T_i obtained by OLS is smaller than a certain threshold then delete T_j subtree and romove redundant "+" type node. Pruning operator implement before calculation of the fitness value of the tree and after crossover operator and mutation operator. This is shown in Fig. 3.

6） If the model expressed by z_i and z_j are identical in structure, fitness of z_j is equal to fitness of z_i, so only need to calculate fitness of one of them.

Based on above-mentioned rules, the model expression can be generated through the tree's preorder traversal. set i=0, implement preorder traversal to individual in GP, when meet non "+" type node, perform preorder traversal to this subtree and regard traversal result as F_i term of NARMAX model,

then marked this subtree with tag S_i. i=i+1, continue traversal. Suppose that the traversal result is "F_1+F_2+…+F_q", so the corresponding polynomial NARMAX model of this subtree is:

$$\hat{y}(k) = \theta_0 + \theta_1 F_1 + \theta_2 F_2 + \cdots + \theta_q F_q \tag{17}$$

This is shown in Fig. 4. With the help of the OLS evaluation, the parameters of the model represented by the tree are estimated. After that, this new individual proceeds on its way in the GP algorithm (fitness evaluation, selection, etc.).

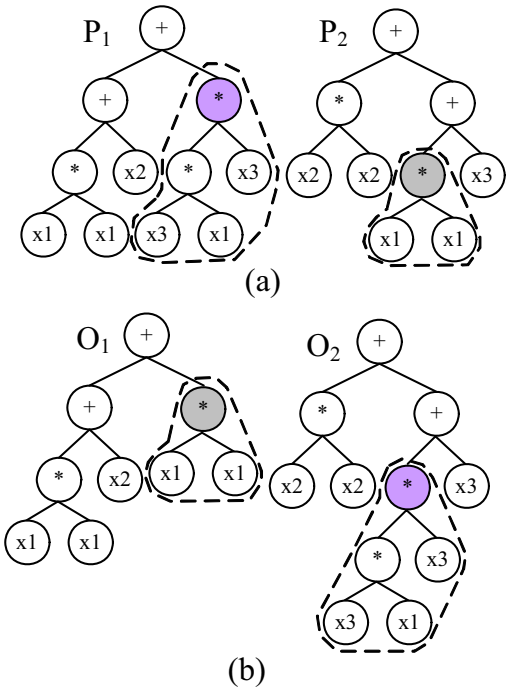


Fig. 1 GP crossover: (a) Random selection of a subtree from P_1 and P_2. (b) Offspring O_1 and O_2 . Dashed lines enclose subtrees to be exchanged. A place worthy of remark is that the type of crossover points must be matched.

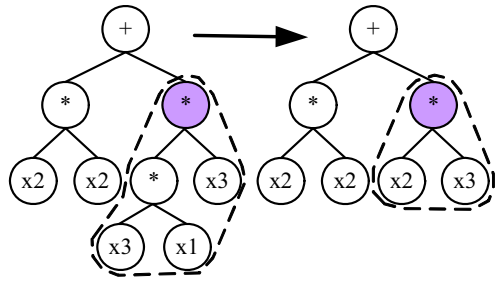
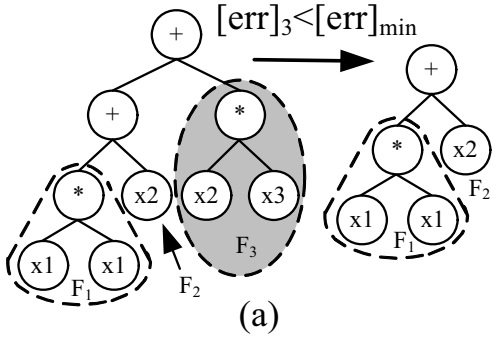


Fig. 2 GP mutation



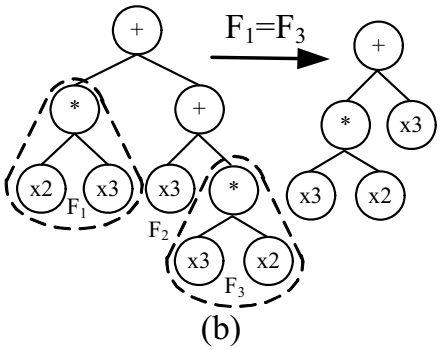
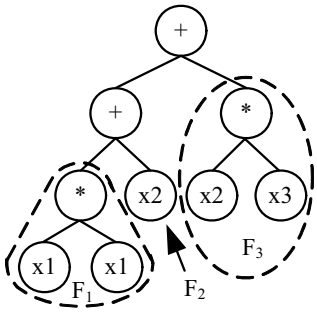


Fig. 3 Pruning operator: (a) Remove redundant terms; (b) Remove terms of whose [err] is smaller than threshold.



$$\hat{y}(k) = \theta_0 + \theta_1 F_1 + \theta_2 F_2 + \theta_3 F_3$$

$$F_1 = x_1^2, F_2 = x_2, F_3 = x_2 x_3$$

Fig 4. GP representation of NARMAX model

V MULTI-OBJECTIVE OPTIMIZATION

In this section, a multiobjective approach is employed as a means for aggregating information of several kinds in the parameter estimation. Multi-objective problem optimization can be stated as follows[8]:

$$\max\{J(\boldsymbol{\theta})\} \quad \boldsymbol{\theta} \in \boldsymbol{D} \tag{18}$$

Where

$$J(\boldsymbol{\theta}) = [J_1(\boldsymbol{\theta}), J_2(\boldsymbol{\theta}) \cdots J_q(\boldsymbol{\theta})]^T \tag{19}$$

$$\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_q] \tag{20}$$

$\boldsymbol{D}$ shows the feasible region, $J_i(\boldsymbol{\theta})$(i=1, 2,…q) is objective function, q is number of optimization objective.

It is the fundamental of non-dominated approach that objects is mapped to fitness function directly, and the efficient solutions are searched by comparing the dominance relationship of objectives, the fitness function is defined based on non-dominant ordering, and keeping diversity in front porch is a key problem of this method.

Non-dominated individual: $\boldsymbol{Z}$ is current population, to a giving individual $z_0 \in \boldsymbol{Z}$, z_0 is termed as nondominated individual, if there has not any other individual $z \in \boldsymbol{Z}$ make the below expression feasible

$$J_k^z > J_k^{z_0}, \exists k \in \{1,2,\cdots q\} \wedge J_l^z \geq J_l^{z_0}, \forall l \in \{1,2,\cdots,q\}$$

Where J_k^z , $J_k^{z_0}$ denote the kth objective function value of individual z, z_0 respectively.

Efficient solution: To a giving $\boldsymbol{\theta}^* \in \boldsymbol{D}$, it is known as

efficient solution, if and only if there has not any other $\boldsymbol{\theta} \in \boldsymbol{D}$ which meet the following inequality $J_k(\boldsymbol{\theta}) > J_k(\boldsymbol{\theta}^*), \exists k \in \{1,2,\cdots,q\} \wedge J_k(\boldsymbol{\theta}) \geq J_k(\boldsymbol{\theta}^*), \forall l \in \{1,2,\cdots,q\}$.

Fitness Assignment: A rank-based fitness assignment strategy is adopted in this work. First, all non-dominated individuals in generation t are marked rank 1, and these individuals are moved aside; Then, all non-dominated individuals in the remaining individuals are marked rank 2, and they are removed; Repeat above-mentioned course, until all individuals are marked with rank.

Definition of distance between two individuals: The distance in decoding space between individual z_i and z_j in population $\boldsymbol{Z}$ is defined as follows:

$$d_{ij}=\mathrm{d}(z_i, z_j).$$

where $d\left(z_i, z_j\right)=\sum_{k=1}^{l} R_k$, and R_k denotes different terms of degree k in z_i and z_j.

For example, if

$$z_i = x_1x_2 + x_2^2 + x_2x_3 + x_1^3, \quad z_j = x_1x_3 + x_1^2 + x_2x_3 + x_1x_2^2$$

then, $d(z_i, z_j)$=2+1=3.

Sharing function:

$$Sh(d_{ij}) = \begin{cases} 1-\left(d_{ij}/\sigma_{share}\right)^{\alpha}, & d_{ij} < \sigma_{share} \\ 0, & d_{ij} \geq \sigma_{share} \end{cases} \tag{21}$$

where α is a constant, σ_{share} is niche radius. Shared fitness of a individual is the quotient of its original fitness divided by its niche number

$$f_i' = f_i / m_i \tag{22}$$

The niche count of giving individual i is the sum of sharing function value on whole population.

$$m_i = \sum_{j=1}^{pop\_size} Sh(d_{ij}) \tag{23}$$

The derived model with good performance and acceptable complexity is convenient for application and analysis. And to chaotic system, if the derived model and original system are identical in chaotic invariants, they have similar chaotic attractor. So a number of objectives involving performance, model complexity, chaotic dynamics have to be taken into account. In this paper, residual variance and correlation coefficient are taken as objectives of model performance, Because of unpredictable feature of chaotic system, long time prediction error criterion is not regarded as objective. Model size (node number of tree) is adopted as objective of model complexity, and defined ND_2, and NK_1 as $NLLE$ as objective of chaotic dynamics.

$$NLLE^i = 1/(\left|\lambda_{\max}^i - \lambda_{\max}^0\right| + 1) \tag{24}$$

$$NK_1^i = 1/(\left|K_1^0 - K_1^i\right| + 1) \tag{25}$$

$$ND_2^i = 1/(\left|D_2^0 - D_2^i\right| + 1) \tag{26}$$

Where $\lambda_{\max}^0$ and $\lambda_{\max}^i$ are the largest Lyapunov exponents, K_1^0 and K_1^i are the Kolmogorov entropys, D_2^0 and D_2^i are the correlation dimensions, of original system and derived model, respectively. Largest Lyapunov exponent, Kolmogorov entropy and correlation dimension quantify property of chaotic attractor from aspects of unordered degree, the self-semilarity of fractal, sensitivity to initial conditions, respectively, and there has fast and robust algorithms to calculate this invariants. When calculate $\lambda_{\max}^i$, D_2^i, K_1^i, the time serials are generated by iterative computations from a initial value, and the initial value is obtained by introducing small variation to a random selected point of the original time serials.

VI EXPERIMENTS AND RESULTS

Based on our experiments we found that with the parameters given in Table 1 the GP is able to find good solutions for various problems.

TABLE 1
PARAMETERS OF GP IN THE APPLICATION EXAMPLES

Population size	100
Maximum Generation	100
Type of selection	roulette-wheel
Type of mutation	subtree-mutation
Type of crossover	one-point (2 parents)
Type of replacement	elitist
Generation gap	0.9
Probability of crossover	0.7
Probability of mutation	0.3
Termination Criterion	Maximum Generation
Generation method	Ramped half-and-half
Max depth of individual	7

The reconstruction delay τ, embedding dimension m can be gotten using phase space reconstruction technology. So input-output data can generate from sampling data according to τ and m. Considering that most chaotic systems are governed by their internal low-dimensional laws, and the embedding dimensions m estimated from time series tend to be larger than the modeling dimension necessary, so we set $n_y = m$ in the experiment.

In order to test the new metric, it is applied to three different kinds of discrete chaotic system: Logistic map, Cubic map and Low-Order Predator-Prey Chaotic Polynomial System. The length of the series is 5000 so that it is long enough to estimate the chaotic invariants. The time series from different chaotic systems were generated and then white Gaussian noise was added as measurement noise, then noisy time series at particular SNR values were created. Here, SNR is 30dB.

1）Logistic map

$$X_{n+1} = AX_n(1 - X_n)$$

Where A=4, X_0=0.1. Largest Lyapunov exponent, Kolmogorov entropy and correlation dimension are as below: $\lambda \approx 0.6931$, D_2 =1.0, $K_1 \approx 0.935$. The optimal model structure achieved by our approach is

$$X_{n+1} = \theta_1 X_n + \theta_2 X_n^2$$

Where θ_1 =4.0000, θ_2=-4.0000, and θ_1 =3.9563, θ_2=-3.9779 when measurement noise was added. Chaotic invariants of

the model series are λ≈0.6927, $D_2 \approx 0.9873$ and $K_1 \approx 0.926$. $MSE = 0.0299$.

2）Cubic map

$$X_{n+1} = AX_n(1 - X_n^2)$$

Where A=3, X_0=0.1. λ≈1.0986, D_2 =1.0, $K_1 \approx 1.481$. The optimal model structure is

$$X_{n+1} = \theta_1 X_n + \theta_2 X_n^3$$

Where θ_1 =3.0000, θ_2=-3.0000, and θ_1 =3.0138, θ_2=-3.0141, when measurement noise was added. Chaotic invariants are λ ≈ 1.086, D_2 ≈0.9873 and K_1 ≈1.473. $MSE = 0.0256$.

3）Low-Order Predator-Prey Chaotic Polynomial System

$$X_{n+1} = -0.3401 - 1.316X_{n-1} + 0.8376X_n + 0.6050X_{n-1}^2$$
$$-0.6284X_nX_{n-1} + 0.2889X_nX_{n-1}^2$$

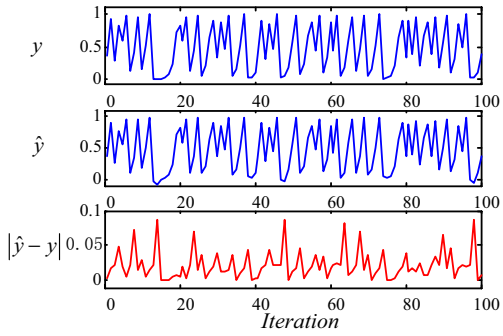
Where $X_0 = 0.1$, $X_1 = 0.1$, λ ≈ 0.057, $D_2 \approx 1.2255$, K_1≈0.087. The optimal model structure is

$$X_{n+1} = \theta_0 + \theta_1 X_{n-1} + \theta_2 X_n + \theta_3 X_{n-1}^2 + \theta_4 X_n X_{n-1} + \theta_5 X_n X_{n-1}^2$$

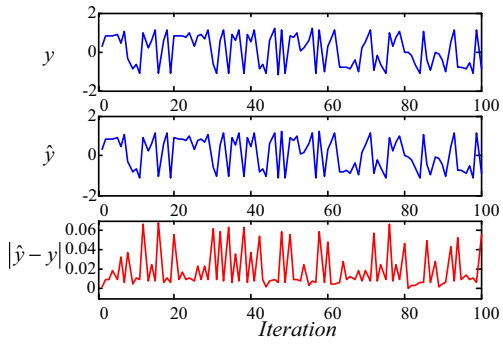
Where θ_0=-03401, θ_1=-1.3162, θ_2=0.8376, θ_3=0.6050, θ_4=-0.6284, θ_5=0.2889, and θ_0=-0.3428, θ_1=-1.297, θ_2=0.8254, θ_3=0.6061, θ_4=-0.6276, θ_5=0.2877 at an SNR of 30dB. λ ≈ 0.063, $D_2 \approx 1.231$, $K_1 \approx 0.093$, $MSE = 0.0468$.

The model's output, actual output and the differences between them of above three chaotic systems are illustrated in Fig 5 (a), (b) and (c) respectively. And prediction length are 2, 1, 10 step(s), respectively.

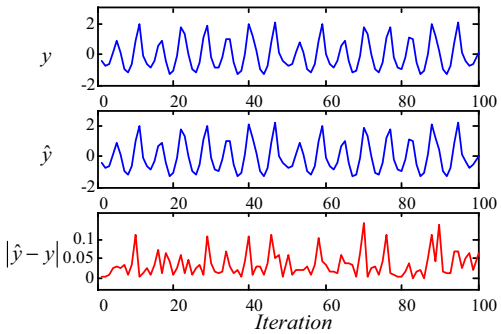
The experiment results on known chaotic series prove that the method can find appropriate models successfully. And derived models have similar system invariants as the measured series indications, which show the effectiveness of this method on revealing system dynamics.



(a) Logistic map



(b) Cubic map



(c) Low-Order Predator-Prey chaotic polynomial system

Fig 5. Model's output, actual output and the differences between them

VII CONCLUSIONS

This paper proposes a new method for chaotic system identification based on polynomial NARMAX representation and mutiobjective GP. The method uses GP to generate polynomial NARMAX models represented in tree structure. The flexibility of the GP approach permits the examination of different NARMAX model by exploiting the dynamic size, shape and content of the tree. The incorporation of a multi-objective approach has enabled the separate consideration of different objectives related to model complexity, model performance and chaotic dynamic. It can get simpler model expressions and model get by this approach was able to reproduce attractors which resemble the geometry of the original data.

REFERENCES

[1] Chen S, Billings, S.A. and Luo, W., Orthogonal Least Squares Methods and Their Application to Non-linear System Identification, Int. J. Control, 1989, vol. 50, no. 5, pp. 1873-1896.

[2] Iba, Hitoshi; Sato, Taisuke; de Garis, Hugo Source, Systemidentification approach to genetic programming: Proceedings of the IEEE Conference on Evolutionary Computation, ICEC, v /1, 1994, pp. 401-406

[3] K. Rodríguez-Vázquez, C. M. Fonseca, and P. J. Fleming , Identifying the Structure of NonLinear Dynamic Systems Using Multiobjective Genetic Programming, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, vol. 34, no. 4, Jul, 2004, pp. 531-546

[4] Force Identification of Dynamic Systems Using Genetic Programming, Yang, Y.W.; Wang, C.; Soh, C.K. Source: International Journal for Numerical Methods in Engineering, vol. 63, no. 9, Jul 7, 2005, pp. 1288-1312

[5] Genetic programming for the identification of nonlinear input-output models, Madar, Janos; Abonyi, Janos; Szeifert, Ferenc Source: Industrial and Engineering Chemistry Research, vol. 44, n 9, Apr 27, 2005, pp. 3178-3186

[6] S.A. Billings, M. Korenberg, and S. Chen. Identification of Nonlinear output-affine systems Using an Orthogonal Least-squares Algorithm. International Journal of Systems Science, vol. 19, no. 8, 1988, pp. 1559–1568.

[7] Chen S, Billings S A, and Luo W. Orthogonal Least Squares Methods and their Application to Non-linear System Identification. International Journal of Control, vol. 50, no. 5, 1989, pp. 1873–1896.

[8] CHANKONG, V. , HAIMES, Y. Y., Multi-objective Decision Making: Theory and Methodology, New York: North-Holland Press, 1983