Optimum grouping in a modified genetic algorithm for discrete-time, non-linear system identification

H Jamaluddin[1], **M F Abd Samad** [2]*, **R Ahmad**[1], and **M S Yaacob**[1]
[1]Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, Johore, Malaysia
[2]Faculty of Mechanical Engineering, Universiti Teknikal Malaysia Melaka, Malacca, Malaysia

The manuscript was received on 29 September 2006 and was accepted after revision for publication on 24 July 2007.

DOI: 10.1243/09596518JSCE362

Abstract: The genetic algorithm approach is widely recognized as an effective and flexible optimization method for system identification. The flexibility of a genetic algorithm allows various strategies to be applied to it. One of the strategies applied is the modified genetic algorithm which relies on, among other things, the separation of the population into groups where each group undergoes mutual recombination operations. The strategy has been shown to be better than the simple genetic algorithm and conventional statistical method, but it contains inadequate justification of how the separation is made. The usage of objective function values for separation of groups does not carry much flexibility and is not suitable since different time-dependent data have different levels of equilibrium and thus different ranges of objective function values. This paper investigates the optimum grouping of chromosomes by fixed group ratios, enabling more efficient identification of dynamic systems using a NARX (Non-linear AutoRegressive with eXogenous input) model. Several simulated systems and real-world time-dependent data are used in the investigation. Comparisons based on widely used optimization performance indicators along with outcomes from other research are used. The issue of model parsimony is also addressed, and the model is validated using correlation tests. The study reveals that, when recombination and mutation are used for different groups, equal composition of both groups produces a better result in terms of accuracy, parsimony, speed, and consistency.

Keywords: model structure selection, NARX model, parameter estimation, system identification, correlation test, genetic algorithm

1 INTRODUCTION

The growing field of system identification has brought much attention to high-end computation for accuracy, ease of calculation, and speed. It comprises four distinctive stages: data acquisition, model structure selection, parameter estimation, and model validity tests. The model structure selection stage is crucial in determining the form of model structure suitable to explain the problem at hand. The issue brings wide recognition to the study of computational systems such as genetic algorithms (GAs).

The GA approach introduced in 1975 [**1**] and has since been studied and developed for many applications [**2–4**]. It has been used in parallel with neural network modelling solutions such as the group data handling network [**5**], the fuzzy neural network [**6**], the Gaussian–Hopfield neural network [**7**], and the recurrent neural network [**8**]. Other types of modelling solution used are block-oriented models [**9**], the Volterra function [**10**], and fuzzy rules [**11**]. Applications of GAs in system identification with direct application to control designs are also available in the literature [**12–14**]. Mitsukura *et al.* [**15**] demonstrated the use of GAs for the combination of functions, including exponential and trigonometric functions, but this has an excessively long chromosome representation. In some of these applications, the model structures are defined beforehand.

In reference [**16**], a GA is used in estimating the parameters of difference functions in the form of ARMAX (AutoRegressive Moving Average with

* *Corresponding author: Faculty of Mechanical Engineering, Universiti Teknikal Malaysia Melaka, Ayer Keroh, Malacca 75450, Malaysia. email: mdfahmi@utem.edu.my*

eXogenous input) and a cubic non-linear function. The model structures in the work are, however, predefined. Ahmad *et al.* [**17**] apply a GA to ARMAX and NARMAX (Non-linear AutoRegressive Moving Average with eXogenous input) models, but argue on the phenomenon of premature convergence.

In spite of the successes, much modification and investigation are still needed and seem feasible as the GA relies on many different parameters. Among its control parameters are population size, recombination operators, crossover probability, mutation probability, mating preferences, and selection strategies [**18**]. The refinement of its control parameters and their interdependence can ensure efficiency of the search in terms of solution optimality, search speed, convergence, algorithm robustness, etc.

Some researchers suggest refinement of search by a hybrid algorithm or the simultaneous usage of different operators to attain quicker convergence to the optimum solution. This can be performed by grouping of the population. Perry *et al.* refer to the groups as species and use different types of crossover and mutation operators[**19**]. The grouping approach as a selection strategy was shown by Ahmad *et al.* [**20**] to perform better than a simple GA in terms of stability and to be equally good or may be even better than an orthogonal least-squares (OLS) algorithm [**20**]. Grefenstette's experiment with an elitist strategy, which also requires grouping, was shown to perform better than a pure strategy in online evaluation [**21**].

Although a GA enables quick evaluation of model structures, the particular issue of parsimony of model structure also arises. This issue is addressed by considering the significance of the terms and variables. Various approaches are available such as the use of locally linear and cross-bilinear models [**22**], penalty functions [**23**], and information criteria such as Akaike's information criterion, the B-information criterion, and the ϕ-information criterion [**24**]. It allows the complexity of the structure to be reduced, gaining a parsimonious model of acceptable accuracy level. The penalty function is a straightforward approach when searching model structures with a GA [**23**].

This paper uses a GA as a model structure selection tool for system identification and also proposes a modified GA by focusing on the suitable proportion of exploitation and exploration of search space. The paper demonstrates that a certain fixed ratio regarding the grouping of population requires less computation time (fewer evolution generations) but is still able to produce an accurate and parsimonious model when a NARX (Non-linear AutoRegressive with eXogenous input) model is used. The application of a NARX model, however, does not restrict the potential of the method for other model representation with slight modification to the structure representation [**25**, **26**].

The organization of this paper is as follows. Section 2 explains the NARX model, as a model structure representation, and the parameter estimation method. Section 3 describes the GA and the modified GA in more detail. This section also lists the performance indicators and correlation tests used as model validity tests in the study. The simulations and results of the study are given in section 4 and discussed in section 5. Section 6 concludes the findings of the study.

2 NARX MODEL

The applicability of the GA applies to both linear and non-linear models. There are wide choices of linear and non-linear models to represent input–output relationships [**25**]. A common model structure representation for a linear discrete-time system is the ARX (AutoRegressive with eXogenous input) model written as

$$y(t) = a_1 y(t-1) + \cdots + a_{n_y} y(t-n_y) + b_0 u(t) + b_1 u(t-1) + \cdots + b_{n_u} u(t-n_u) + e(t) \quad (1)$$

where $y(t)$, $u(t)$, and $e(t)$ are the output, input, and noise respectively at time t; n_y and n_u are the orders of lag for output and input respectively, and $a_1, \ldots, a_{n_y}$, b_0, and $b_1, \ldots, b_{n_u}$ are the coefficients, or parameters, of the model. Non-linear models give much richer possibilities in describing systems and have better flexibility when inferring from a finite dataset. The non-linear version of the ARX model is the NARX (Non-linear ARX) model. When a time delay exists, it is written as

$$y(t) = F_*^l[y(t-1), y(t-2), \ldots, y(t-n_y), u(t-d), \ldots, u(t-d-n_u+1), e(t)] \quad (2)$$

which is also a generalization of the linear difference equation. In the above equation, $F_*^l[\cdot]$ is a non-linear function of u and y, d is the time delay, l is the degree of non-linearity, and other notations are the same as before. Model structure selection refers to the process of determining the lags of input, n_u, output, n_y, and time delay, d, from the information of input, u, and output, y, sequences for a discrete-time model [**24**]. The aim in model structure selection mainly is to determine the significant terms to be included in a system model.

Before the parameters of the model can be estimated using the least-squares method, the model has to be transformed into a linear regression model as follows

$$y(t) = \boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta} + e(t), \qquad n_y \leqslant t \leqslant N \tag{3}$$

where $\boldsymbol{\theta}$ is the parameter vector, $\boldsymbol{\phi} = [\phi_1 \quad \phi_2 \quad \cdots \quad \phi_L]^{\mathrm{T}}$ is the regressor vector, e is the value of noise or disturbance, L is the number of significant variables and terms, which also determines the size of the parameter vector, and N is the number of data.

Given that the model structure and consequently the vector of significant regressors has already been defined, the estimated parameters $\widehat{\boldsymbol{\theta}}$ can be obtained using least-squares methods [**25**, **26**].

The number of terms or regressors in a NARX model, L, is calculated as follows

$$L = M + 1 \tag{4a}$$

Here

$$M = \sum_{i=1}^{l} n_i \tag{4b}$$

where l is the degree of non-linearity, and

$$n_i = \frac{n_{i-1}(n_y + n_u + i - 1)}{i} \tag{4c}$$

where $n_0 = 1$, with n_y and n_u as in model (1).

Suppose a NARX system is known to have a non-linearity order of two, an order of lag for input $n_u = 2$, an order of lag for output $n_y = 2$, and a time delay $d = 0$. The number of possible terms in the model is found to be 15, along with the inclusion of a constant term. Since the decisions on the terms are either inclusion or omission, simple binomial theorems apply. Therefore, in a model consisting of L terms, the search space is $2^L - 1$ (which in the above example is 32 767).

3 GA AND MODIFIED GA

3.1 GA

The GA conforms to the metaphor of natural biological evolution by applying the principle of survival of the fittest. With this metaphor, it operates as a stochastic global search method in producing better approximations to a solution of system identification [**1**, **3**]. It emphasizes genetic encoding of potential input–output relationship solutions of a system into chromosomes and applies genetic operators to these chromosomes. Some early work, theories, and background on the GA can be found in references [**1**] and [**27**].

A chromosome is a string of code used as a representation of the solution. Each position in the string is called a gene. There are, however, only two main genetic operators, crossover and mutation, each providing different styles of manipulation of the string solutions towards building the optimum solution. The crossover operator operates by 'mating' two chromosomes so that portions of their structures or hyperplanes are exchanged. In this exchange of information, new structures called offspring are formed. The crossover probability or rate, p_{c}, determines the number of pairs in the population that are forced to mate. The mutation operator acts by changing a portion of the structure individually. It involves the targeting of certain positions of a structure and changing the value held by the gene. The mutation probability or rate, p_{m}, indicates the number of bits in a population that is being mutated.

The general procedure of the GA consists of a population of chromosomes where the operators act on them, creating a new population. The old population is replaced so that the new creation populates or is passed forward to the following generation. During this process, the fitness of each chromosome is evaluated on the basis of a specified function. The fitness of a chromosome can be defined subjectively in terms of the objective function (OF) value. With this evaluation, chromosomes are selected for the next generation. The creation and evaluation of new solutions are performed until a user-defined termination criterion is reached, such as the maximum number of generations. Details of GA operators and parameters can be found in references [**3**] and [**23**]. The general flow chart of the GA is given in Fig. 1.

3.2 Proposed modified GA

A modification to the simple GA, known as the modified genetic algorithm (MGA), presented in reference [**20**], takes another step in exploring the potential of the GA by dividing the population into groups according to their fitness values, where each group undergoes a different set of genetic operations. It has been shown to perform better than the simple GA and the OLS algorithm.

The success of grouping in reference [**20**] over the simple GA emphasizes the division into good, second best, and bad groups, while reference [**21**] recognizes the importance of the elitist strategy along with

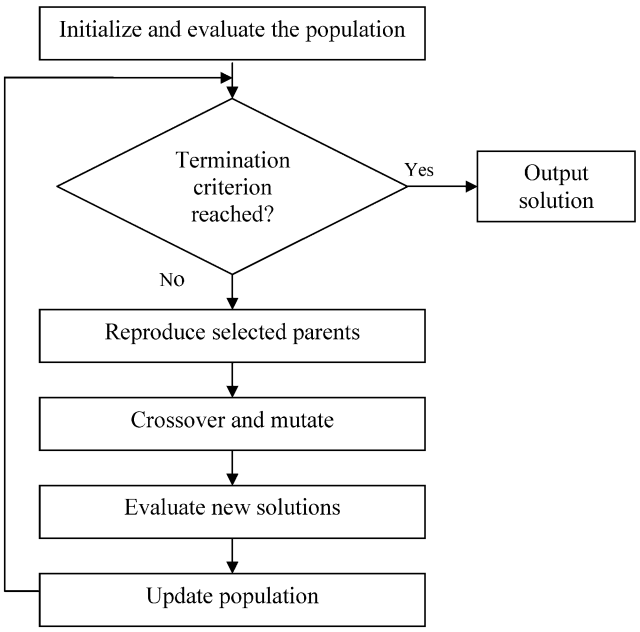

Fig. 1 Simple GA flow chart

other conventional routines. The proposed algorithm considers all of the above by emphasizing four different groups. The flow chart for its implementation is given in Fig. 2. Two of the groups (best and bad) consist of single individuals, while the other two (acceptable and ordinary) are actual groups with varied proportions.

Listed from the fittest to the weakest, the groups are as follows.

1. *Best*. The best chromosome is also known as the elitist. Conforming to what is known as the elitist strategy, the single best individual of the generation is preserved or reproduced for the next generation without any changes to its structure. It preserves the structure or schemata so that direct comparison can be made between it and other manipulated chromosomes of the next generation.
2. *Acceptable*. The acceptable group has the second highest fitness values and undergoes only mutation. The rationale behind this strategy is that the chromosomes in this group are expected to have optimum hyperplanes owing to high fitness levels. Therefore, in order to prevent schemata disruption (the complete disruption of its solution structure), the crossover process is not allowed.
3. *Ordinary*. This group consists of the next fittest chromosomes, and they undergo the conventional sequence in recombination – crossover and mutation. The importance of inducing mutation in this sequence is to increase the variability of the population, thus maintaining high exploration of search space.
4. *Bad*. The bad chromosome, having the lowest fitness value in the population, is compared against a newly initialized chromosome and replaced if the replacement is better.

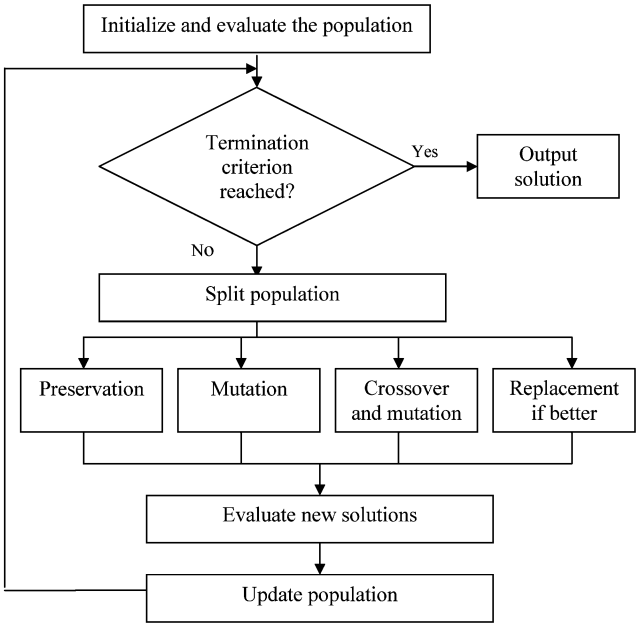

Fig. 2 MGA flow chart

The importance of the division is exceptionally clear in the appointment of the type of genetic operators to the groups. Although other researchers suggest mutation as a 'background operator', recent findings reveal that mutation is also useful as a search operator [**28**]. The separation of the population into groups raises the question of the optimum ratio for separation of the acceptable and ordinary groups. Previous usage of setting the separation point on the basis of a fixed value of the objective function is only suitable when it is known that the OF values of time-dependent data spread across the fixed value.

Previous research using a fixed value for separation is exemplified as follows. By running the algorithm for the first generation, a fixed value can be readily selected relative to the whole range of OF values in that generation. For a minimization problem, when the fixed value is ten, all chromosomes having an OF value higher than ten are classified as ordinary and bad, while chromosomes having an OF value lower than ten fall into the acceptable and best categories.

Depending on the level of equilibrium of time-dependent data over finite time, different time-dependent data give different ranges of OF values with regards to the population. The range may also change throughout evolution. The selection of a fixed

value for separation needs to be avoided since it is important not to differentiate the quality of a chromosome favourably based on the OF value but with the position in the population after sorting based on the OF value.

In order to avoid this situation, the number of chromosomes allocated to the groups in the proposed algorithm are determined by different fixed ratios before an optimum ratio based on predefined performance indicators can be selected. The settings of the ratios are explained in section 4.

3.3 Application of GA to structure selection of the NARX model

Following the example in section 2, where the nonlinearity of a system was of order two, the order of lag for input $n_u=2$, the order of lag for output $n_y=2$, and the time delay $d=0$, the variables are $y(t-1)$, $y(t-2)$, $u(t-1)$, and $u(t-2)$, and the terms are constant a_1 and the multiplications of variables.

The output, $y(t)$, for the system is represented by

$$\begin{aligned} y(t) = {} & a_1 + a_2 y(t-1) + a_3 y(t-2) + a_4 u(t-1) \\ & + a_5 u(t-2) + a_6 y^2(t-1) + a_7 y(t-1) y(t-2) \\ & + a_8 y(t-1) u(t-1) + a_9 y(t-1) u(t-2) \\ & + a_{10} y^2(t-2) + a_{11} y(t-2) u(t-1) \\ & + a_{12} y(t-2) u(t-2) + a_{13} u^2(t-1) \\ & + a_{14} u(t-1) u(t-2) + a_{15} u^2(t-2) + e(t) \end{aligned} \quad (5)$$

In a binary-represented GA, the variables and the terms are represented by the genes of the chromosome as bit 1 for existence and bit 0 for omission. Therefore, chromosome [110 100 001 000 100] represents the model

$$\begin{aligned} y(t) = {} & a_1 + a_2 y(t-1) + a_4 u(t-1) \\ & + a_9 y(t-1) u(t-2) + a_{13} u^2(t-1) \end{aligned} \quad (6)$$

The model is completed by estimation of the parameters a_1, a_2, a_4, a_9, and a_{13} by parameter estimation methods.

3.4 Model validation

3.4.1 Performance indicators

In order to evaluate the accuracy of a structure effectively, a performance measure is needed. The indicator helps in determining the best structure among the possible solutions in the population. Another indicator is used in the iterations to determine the group for model structures. Since the algorithm relies on optimum grouping of predetermined population size, an indicator of the algorithm convergence rate (in terms of the required number of generations) is also needed.

Three performance indicators are used in this paper.

1. *Error index of the best chromosome*. The error index (EI) refers to the square root of the sum of squared errors of the best chromosome, namely the elitist in the final generation, divided by the sum of the output squared. The calculation of EI is as follows

$$EI = \sqrt{\frac{\sum [y(t) - \hat{y}(t)]^2}{\sum y^2(t)}} \quad (7)$$

   where $\hat{y}(t)$ is the one-step-ahead predicted output obtained from the least-squares method. When several trials are made, the mean value is calculated.
2. *OF value of the best chromosome*. The OF is used as a measure of fitness of the chromosome where it is set that a highly fit individual has a low OF value, and *vice versa*. This indicator is the evaluation function repeatedly used in the simulation to decide the group to which a chromosome belongs before the next stage proceeds. Since it is also of high importance that a solution is at a suitable parsimony, a penalty function is integrated in the calculation. The logarithmic penalty function is observed in Akaike's B-information criterion [**24**] and shown to be better than other forms of penalty function [**23**]. Here, the OF is as follows

$$OF = \left\{ \sum_i^N [y_i(t) - \hat{y}_i(t)]^2 \right\} + \log(n) \quad (8)$$

   where n is the number of insignificant regressors plus one. Insignificant regressors refer to regressors with an estimated parameter less than or equal to a specified penalty value agreed upon for trade-off with the level of accuracy. The selection of the penalty value requires a trial-and-error approach based on the knowledge that the number of selected significant regressors is inversely proportional to the penalty value. The significant regressors refer to regressors that correspond to high-valued parameters

$$\textit{Number of selected regressors} \propto \frac{1}{\textit{penalty value}}$$

   In the comparison between different ratios, it refers to the OF value of the elitist that is selected

by the end of evolution. Like the EI, when several trials are used, the mean value is presented.

3. *Generation count.* The generation count refers to the generation number in the evolution at which the elitist in the final generation first appears. With the assumption that each generation requires the same amount of time, this indicator allows the comparison of computation time between different ratios needed to identify the best chromosome. Two statistical calculations are given:
   (a) mean from a number of trials, $\bar{x}$

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{J} \tag{9}$$

   where x is the value of the generation count of a trial and J is the number of trials;
   (b) standard deviation from a number of trials, s

$$s = \sqrt{\frac{\sum_{i=1}^{J} (x_i - \bar{x})^2}{J-1}} \tag{10}$$

   The mean decides the speed of the ratio while the standard deviation detects the level of consistency or reliability of the ratio.

The run of the algorithms is made several times for a fair conclusion on its performance. By running the algorithms several fixed times, the element of dependence on its initial population is effectively reduced.

These performance indicators are observed in the order as listed above. The order is made to preserve the priorities of observing algorithm performance, beginning from accuracy, parsimony, speed, and lastly consistency. This is because an accurate solution is easier to achieve than parsimonious solutions. Then, effective ratios are differentiated by speed, then consistency.

3.4.2 *Correlation tests*

The final step crucial in system identification is the model validity test. A model can only be accepted as valid once it is proven that the selected terms and variables do not contribute bias to its accuracy. This can be done by applying correlation tests to ensure that no other significant terms and/or variables are omitted from the model. A non-linear model requires more tests than a linear model since it contains polynomials of variables. It is valid if it fulfils the following conditions [**29**]

$$\phi_{\varepsilon\varepsilon}(\tau) = \frac{E[\varepsilon(t)\varepsilon(t+\tau)]}{E[\varepsilon^2(t)]} = \delta(\tau), \qquad \tau = 0 \tag{11a}$$

$$\phi_{u\varepsilon}(\tau) = \frac{E[u(t)\varepsilon(t+\tau)]}{\sqrt{E[u^2(t)\varepsilon^2(t)]}} = 0, \qquad \forall\tau \tag{11b}$$

$$\phi_{\varepsilon(\varepsilon u)}(\tau) = \frac{E[\varepsilon(t)\varepsilon(t-1-\tau)u(t-1-\tau)]}{\sqrt{E[\varepsilon^2(t)]E[\varepsilon^2(t)u^2(t)]}} = 0, \qquad \tau \geqslant 0 \tag{11c}$$

$$\phi_{u^{2'}\varepsilon}(\tau) = \frac{E[(u^2(t) - \overline{u^2})\varepsilon(t+\tau)]}{\sqrt{E[(u^2(t) - \overline{u^2})^2]E[\varepsilon^2(t)]}} = 0, \qquad \forall\tau \tag{11d}$$

$$\phi_{u^{2'}\varepsilon^2}(\tau) = \frac{E[(u^2(t) - \overline{u^2})\varepsilon^2(t+\tau)]}{\sqrt{E[(u^2(t) - \overline{u^2})^2]E[\varepsilon^4(t)]}} = 0, \qquad \forall\tau \tag{11e}$$

where the residual, $\varepsilon(t)$, is calculated by

$$\varepsilon(t) = y(t) - \hat{y}(t) \tag{12}$$

$\hat{y}(t)$ is the one-step-ahead predicted output, and the overbar denotes the time average, so that $\overline{u^2}$ is given by

$$\overline{u^2} = \frac{1}{N}\sum_{t=1}^{N} u^2(t) \tag{13}$$

When analysing a system that is not affected by external input [**30**], the single-dimensional correlation tests to counter the autocorrelations of residuals of low and high order consist of equation (11a) and

$$\phi_{\varepsilon^2\varepsilon^2} = \frac{\sum_{t=1}^{N+\tau} (\varepsilon^2(t) - \overline{\varepsilon^2})}{\sum_{t=1}^{N} (\varepsilon^2(t) - \overline{\varepsilon^2})^2} = \begin{cases} 1, & \tau = 0 \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

where

$$\overline{\varepsilon^2} = \frac{1}{N}\sum_{t=1}^{N} \varepsilon^2(t) \tag{15}$$

The accepted bandwidth for the fit of the model to the system is approximately $\pm 1.96/\sqrt{N}$ when allowed a 95 per cent confidence interval, with N as the number of data points.

4 SIMULATION STUDY

Simulated systems and real-world data are used in the search for the optimum grouping for the MGA. The simulated systems are of second- and third-degree non-linearity. A high degree of non-linearity exhibits high complexity and requires careful selection of regressors. Their model structures are predefined

when generating data. This enables immediate identification of the order of lags and direct comparison between the correct model structure and the one identified by the GA.

The time-dependent data are the Wölfer sunspot and gas furnace data exhibiting an *a posteriori* modelling requirement [**31**]. The selection of the order of lags for these data and specifications are explained in their respective subsections. When used with real data, a high-order polynomial function is believed to be able to yield better modelling [**32**].

In a MGA, the optimum group ratio has to be decided. Five trials were made for each of five selected ratios implemented for each system and time-dependent dataset. The selection of five trials has been found to be adequate to represent the natural behaviour of the algorithm. The representation and operations used in the algorithm are those of the conventional simple GA, also known as the canonical GA: binary representation, one-point crossover, and bit-flipping mutation [**33**]. The selection method used is individual selection, where the fitness of individuals controls their future potential as parents, while the mating preference is mating of two chromosomes with close OF values. This mating preference is similar to positive assortive mating [**3**].

The investigation is made using a separation ratio. Fixed ratios are applied so that the number of chromosomes in each group is maintained. It has to be noted that, although there are four groups in the population, two groups are of primary concern: acceptable and ordinary. After the evaluation of individual chromosomes has been done, two chromosomes are labelled as *best* and *bad*. This leaves a reduced population of size

$$\textit{Separation population} = \textit{popsize} - 2 \tag{16}$$

The reduced population is split into two by the following ratios.

1. $Acc = 1 : Ord = 0$.
2. $Acc = 3 : Ord = 1$.
3. $Acc = 1 : Ord = 1$.
4. $Acc = 1 : Ord = 3$.
5. $Acc = 0 : Ord = 1$ (representative of a simple GA).

where *Acc* is the number of chromosomes in the acceptable group and *Ord* is the number in the ordinary group. These ratios are referred to as ratios 1 to 5. For comparative intention, however, it is interesting to note that ratio 5 is representative of a simple GA where, putting aside all supplementary strategies, all chromosomes are required to undergo crossover and mutation in the indicated sequence.

Unless mentioned otherwise, the specification of the algorithm is fixed for all trials, simulated models, and time-dependent data. The population size, *popsize*, is set to 200, the maximum generation is 100, the crossover probability $p_c = 0.6$, the mutation probability $p_m = 0.7/\text{chromosome length}$, and the penalty value is 0.001.

4.1 Simulated model 1

The first simulated system is generated from the following model (model 1)

$$y(t) = 0.5y(t-1) + 0.3u(t-2) + 0.3y(t-1)u(t-1) + 0.5u^3(t-1) + e(t)$$

The input lag $n_u = 2$, the output lag $n_y = 1$, and the non-linearity $l = 3$. The inputs, disturbances, and number of data are as follows: the input, $u(t)$, is a random white signal in the range $[-1, 1]$; the noise, $e(t)$, is a random white noise in the range $[-0.01, 0.01]$; the number of data generated $N = 500$.

The model contains 20 regressors, resulting in a search space of size 1048 575. By knowing the regressor selection sequence written in the computer program, the correct chromosome for the model can be written as [010 101 000 000 000 010 00].

In this experiment, all ratios provide chromosomes with the same EI as the optimum solution, valued at 0.0161. However, all trials of ratios 1 to 4 give the same low OF values of the best chromosomes, at 0.017 30, where most trials needed less than half of the maximum generation. Ratio 5 gives an average of 0.017 32, and this failure is an obvious indication of poor performance. Ratio 4 was found to have the lowest mean and standard deviation of the generation count. The next best ratio appears to be ratio 3, then ratio 2. Numerically, ratio 4 provides a 51 per cent quicker and 63 per cent more consistent result than its closest competitor.

Figure 3 gives the graphical results in one of the trials of ratio 4. Figure 3(a) gives the sum of squared errors of the whole population throughout the generation. The convergence trend of population error is quite unclear owing to the large population size, and this is also seen in the other ratios. The search for the solution reached its optimum at the 19th generation (Fig. 3(b)). This is faster than most of the trials of other ratios. The best structure contains nine regressors, including all the correct regressors. The reason for this non-parsimonious model selection is that the penalty value set beforehand in the algorithm is too low, causing the number of selected regressors to rise. This can be overcome by applying different

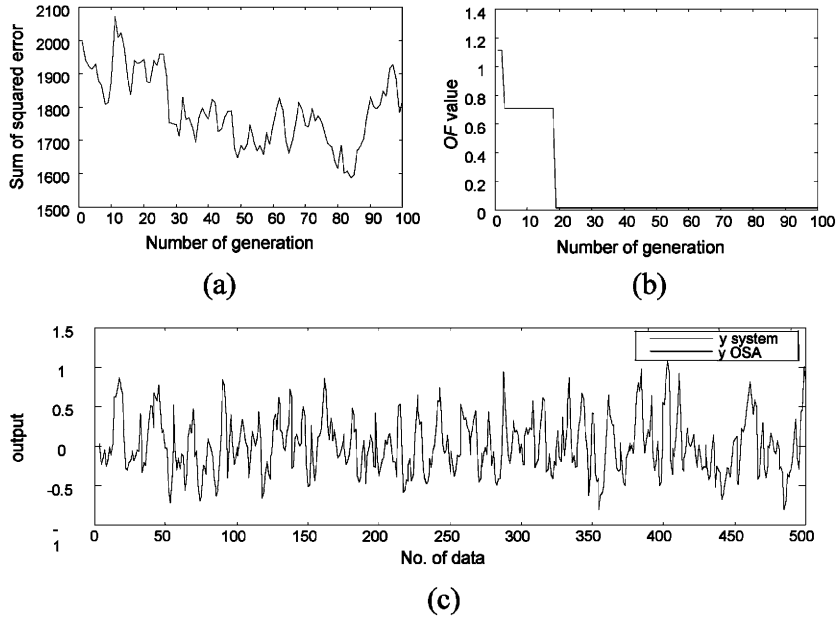

Fig. 3 Results of simulated model 1 (ratio 4): (a) sum of squared errors of population versus number of generations; (b) OF value of best chromosome versus number of generations; (c) system output and one-step-ahead output prediction versus number of data

values of penalty before a suitable level of parsimony is accepted. By increasing the penalty value to 0.1, thereby penalizing bits with a parameter value less than 0.1, the best chromosome was found to be exactly the same as the correct chromosome which is [010 101 000 000 000 010 00].

The identification of the best chromosome is followed by model validity tests to confirm that no bias is contained in the model. This is presented in Fig. 4 for the correct chromosome using a penalty value of 0.1, where it can be seen that all points of the selected τ value satisfy equations (11a) to (11e) within the confidence bandwidth.

4.2 Simulated model 2

The second simulated system is generated from the following model (model 2)

$$y(t) = 0.5y(t-1) + 0.4u(t-1) - 0.2u(t-3) + 0.3y(t-1)u(t-1) + 0.1y(t-2)y(t-2) + 0.4u^2(t-3) + e(t)$$

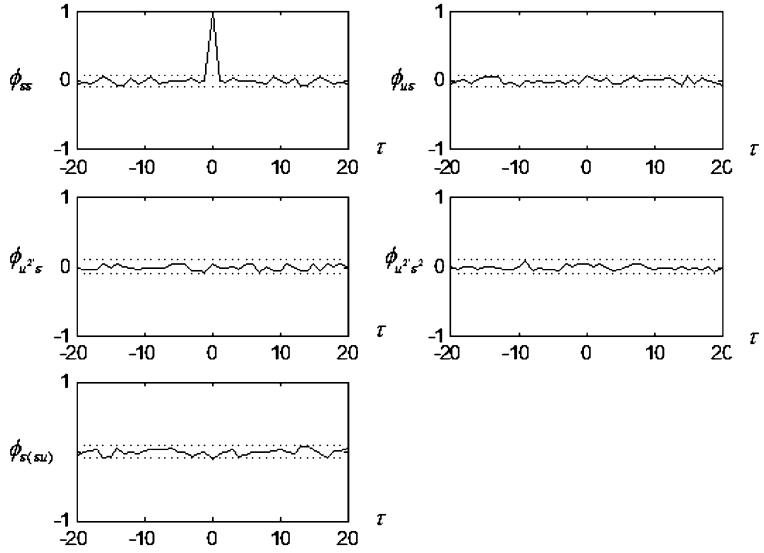

Fig. 4 Model validity tests for simulated model 1

The input lag $n_u = 3$, the output lag $n_y = 2$, and the non-linearity $l = 2$. It contains 21 regressors with 2 097 151 overall candidate solutions in its search space. The correct chromosome for the model is [010 101 001 001 000 000 001]. The input region, disturbance region, and the settings of the algorithm are the same as in simulated model 1.

In this test, ratios 1 and 5 are unable to produce a solution with a low EI like the other 'winning' ratios, that is, 0.0131. This eliminates the potential of ratios 1 and 5 as they require more generations before a newer solution with a lower EI can be selected. Comparisons regarding the parsimony of models with OF values reveal that the 'winning' ratios give equally parsimonious models with OF = 0.016 30. Ratio 2 gives the lowest mean and lowest standard deviation of the generation count, followed by ratio 3. The differences are 27 per cent in speed and 39 per cent in consistency.

Figure 5 gives the graphical results of one of the trials of ratio 2. The convergence trend of the whole population seen in Fig. 5(a) is smoother than in simulated model 1 as the *popsize*-to-search space ratio is smaller. The actual number of the generation count in this sample run is 27 (Fig. 5(b)), with the best chromosome containing 20 regressors including all the correct variables and terms. By increasing the penalty value to 0.1 to increase parsimony, the elitist was represented by the correct chromosome, [010 101 001 001 000 000 001].

Figure 6 shows the model validity tests for the elitist found from a penalty value of 0.1. All tests reveal that the structure is within the confidence bandwidth.

4.3 Wölfer sunspot time series data

The Wölfer sunspot time series data consist of the average number of sunspots on the sun measured annually [**31**]. It is a suitable and commonly used time series exhibiting discrete-time behaviour and equilibrium over a long period. Since the series does not have input, it involves only regressions among past output terms at equal time intervals. Applying a term selection algorithm combined with the OLS and error reduction ratio approach, Wei *et al.* [**22**] identify the suitable lag order for the output to be nine. The identification by Wei *et al.* [**22**] was made using a linear model, namely the ARX model. The variables identified as significant are [$y(t-1)$, $y(t-9)$, $y(t-2)$, constant], and when this is represented in binary coding of length ten it is [111 000 000 1].

To avoid pure coincidence, the Wölfer sunspot time series data are tested by applying $n_y = 10$ and $l = 1$. This results in 11 regressors and 2047 candidate solutions. Since the number of regressors is smaller, the value of *popsize* is reduced to 30. All other specifications are as in the simulated systems. There are 288 output values collected from the year 1700 to 1987, and the data are normalized to the standard interval [0, 1].

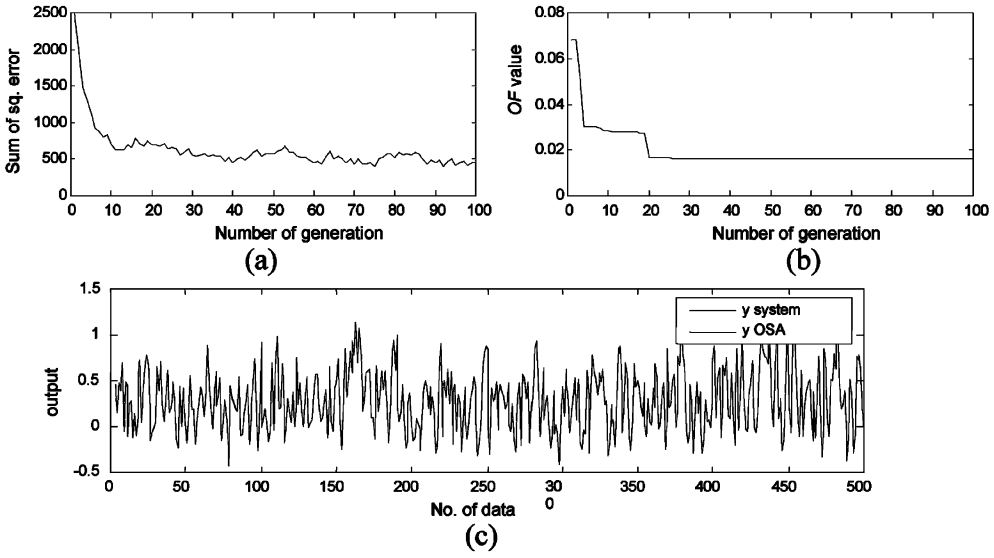

Fig. 5 Results of simulated model 2 (ratio 2): (a) sum of squared errors of population versus number of generations; (b) OF value of best chromosome versus number of generations; (c) system output and one-step-ahead output prediction versus number of data

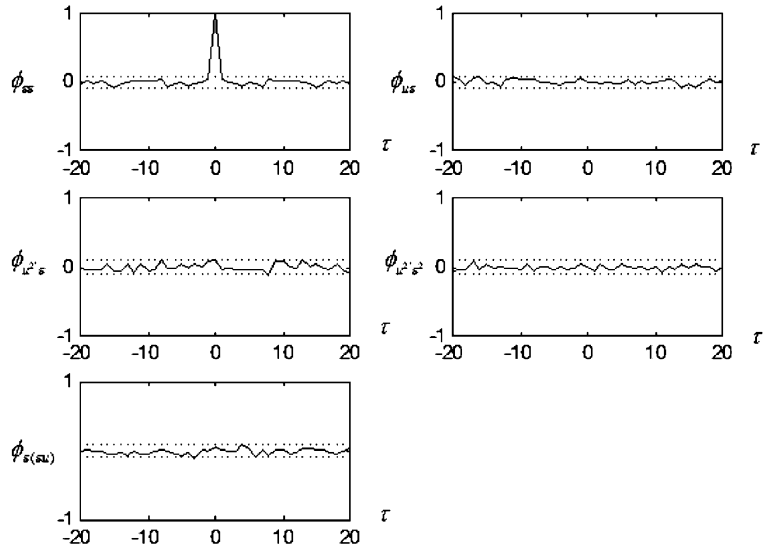

Fig. 6 Model validity tests for simulated model 2

In this test, all ratios give the same value of the EI, 0.2341. The OF values are also the same at 1.6883. This suggests a comparison of the generation count. Ratio 3 gives the lowest mean value, with a 12 per cent difference from the second best ratio, ratio 2. Ratio 2 gives the lowest standard deviation value, with a 25 per cent difference from the second best ratio, ratio 3.

Observing the graphical results of ratio 3 given in Fig. 7, the convergence trend is unclear (Fig. 7(a)), as is the case for simulated model 1. The elitist or best chromosome was detected from the fifth generation (Fig. 7(b)). Although the value of the maximum generation is high, the method always converges in far fewer generations, the slowest in ratio 3 being 13 generations. With a low penalty value, the algorithm picks up all regressors as significant. Although the model accuracy is better, the chromosome consists of all regressors, neglecting the importance of parsimony. An additional test applying a penalty value of 0.1 selects chromosome [011 000 000 10] which is similar to the finding by reference [**22**].

The result of model validation using the parsimonious model found by applying a penalty value of 0.1 is shown in Fig. 8, where the first test is successful while the second test gives out-of-boundary values

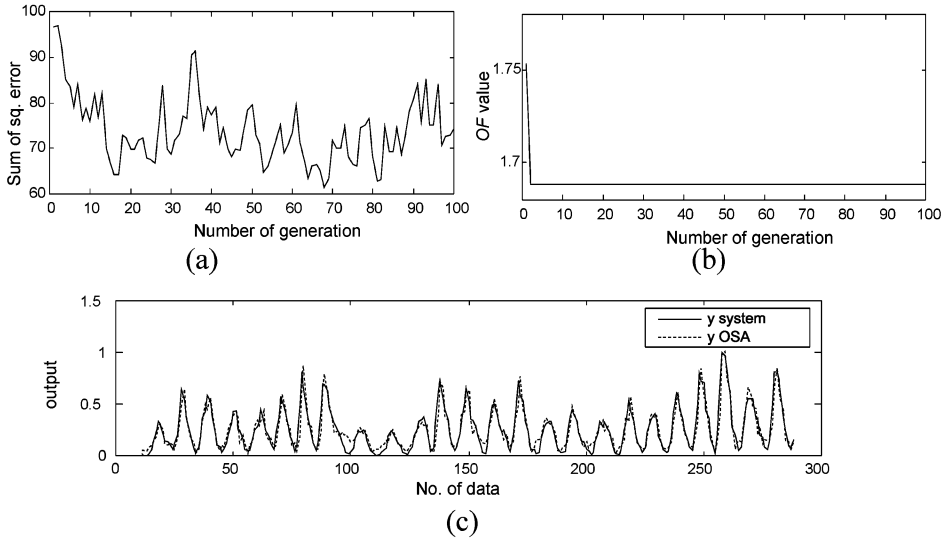

Fig. 7 Results of Wölfer sunspot time series data (ratio 3): (a) sum of squared errors of population versus number of generations; (b) OF value of best chromosome versus number of generations; (c) system output and one-step-ahead output prediction versus number of data

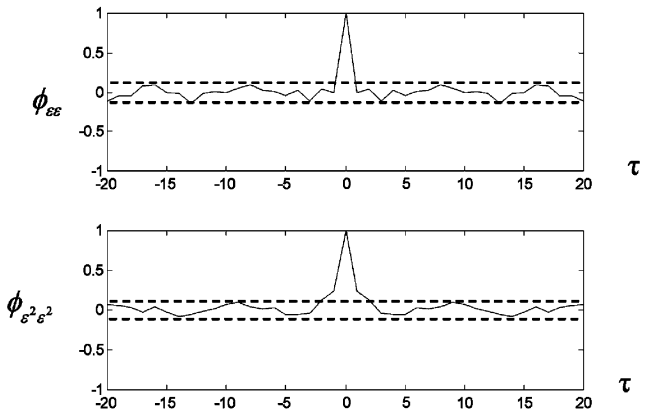

Fig. 8 Model validity tests for Wölfer sunspot time series data using a linear model structure (a non-linearity of one)

when applying $\tau = 0.1$ and $\tau = -0.1$. This may be inherent from the selection of linear modelling. It suggests that a higher-order model is required.

Another trial was made, applying a non-linearity of three. The order of lag of output $n_y = 9$, since linear modelling has demonstrated that the system is better identified with a high lag value. This correponds to 220 regressors/genes being contained in each chromosome. The population size, *popsize*, is set to 100, the penalty value is 1.5, and the other settings of the algorithm are the same as those for a non-linearity of one. The high penalty value selected in the algorithm is due to ensuring parsimony for the high quantity of parameters associated with a high number of regressors. The results were tested using the correlation tests to ensure model validity, and this is shown in Fig. 9 where all lag points are within confidence bands.

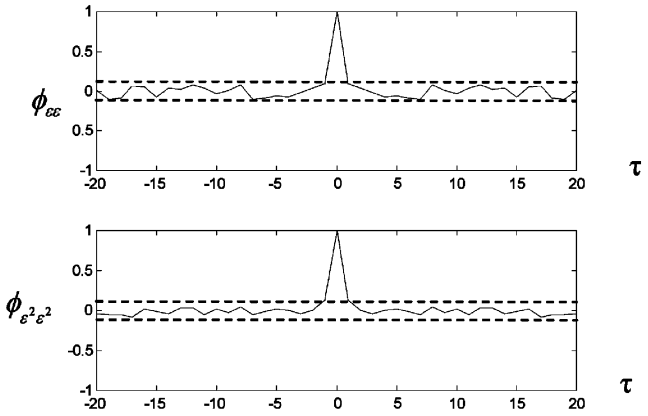

Fig. 9 Model validity tests for Wölfer sunspot time series data using a non-linear model structure (a non-linearity of three)

4.4 Gas furnace data of Box *et al.* [31]

These time series data were obtained by Box *et al.* [**31**] from actual process plant data consisting of a discrete stochastic input series of gas feed rate in cubic feet per minute and an output series of carbon dioxide concentration in the outlet gas. There are 296 pairs of input–output data sampled at an interval of 9 s. After several tests, a satisfactory specification of non-linearity, input lag, and output lag is used for this study: output lag = 2; input lag = 2; non-linearity = 2.

With this specification, the number of regressors amounts to 15 and the search space is 32 767. Owing to the small number of regressors, the *popsize* is set to 100 while other specifications are kept the same as in the simulated systems.

All the ratios provide the same EI value of 0.0046. The objective OF values of all the ratios are also the same at 18.1491. This means that the solutions all have equal levels of error and parsimony. Regarding the generation count, ratio 3 reveals the lowest mean value, followed by ratio 5, with a difference of 13 per cent. The lowest standard deviations are those of ratios 2 and 5. Their values are close, and the other ratios are at least 40 per cent more inconsistent.

Figure 10 gives the graphical results of one of the trials of ratio 3, where an unstable situation is observed at the beginning of evolution before convergence of population error is observed (Fig. 10(a)). The generation number when the best chromosome was first found is nine (Fig. 10(b)). In this simulation, the algorithm selects all regressors as significant owing to the low penalty value. Increasing the penalty value to 0.1 gives [111 101 011 011 111] as the best chromosome. A further increment in penalty value to 0.5 yields [111 010 010 000 000] as the best chromosome.

Figure 11 shows the result when the chromosome of penalty value 0.5 is tested. All tests are satisfied except $\phi_{u\varepsilon}$, which indicates that the process estimate is biased. In a linear system, however, when the residual is correlated with the input, the residuals are also autocorrelated even if the noise model is correct (when $\phi_{u\varepsilon}(\tau) \neq 0\ \forall\tau$, $\phi_{\varepsilon\varepsilon}(\tau) \neq \delta(\tau)$) [**29**]. However, since this is analysed non-linearly, the slight inadequacy may be inherent from wrong selection of lag orders or non-linearity when these parameters are set.

5 DISCUSSION

Three out of four data samples used in the paper show that, with a sufficient number of generations, all ratios are able to produce solutions with an

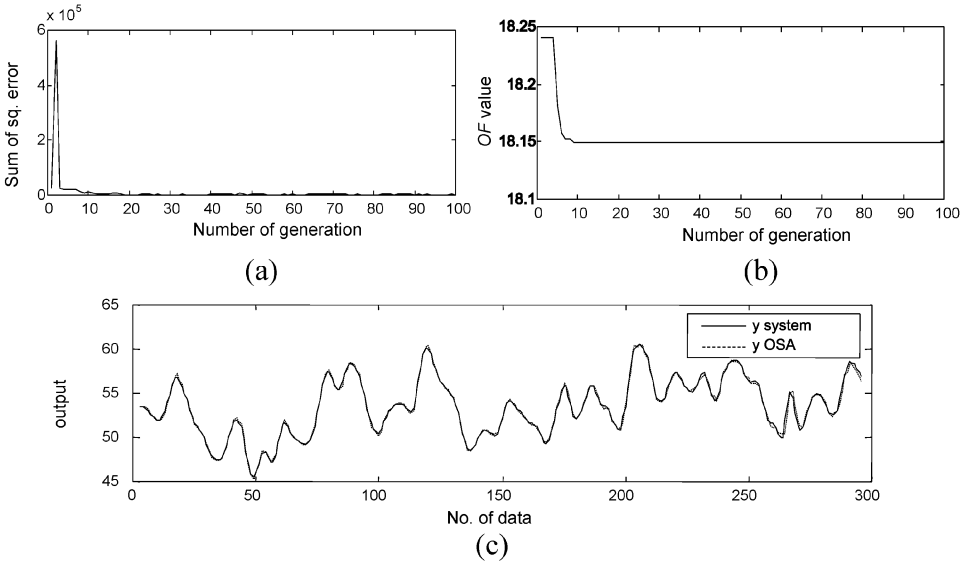

Fig. 10 Results of gas furnace time series data (ratio 3): (a) sum of squared errors of population versus number of generations; (b) OF value of best chromosome versus number of generations; (c) system output and one-step-ahead output prediction versus number of data

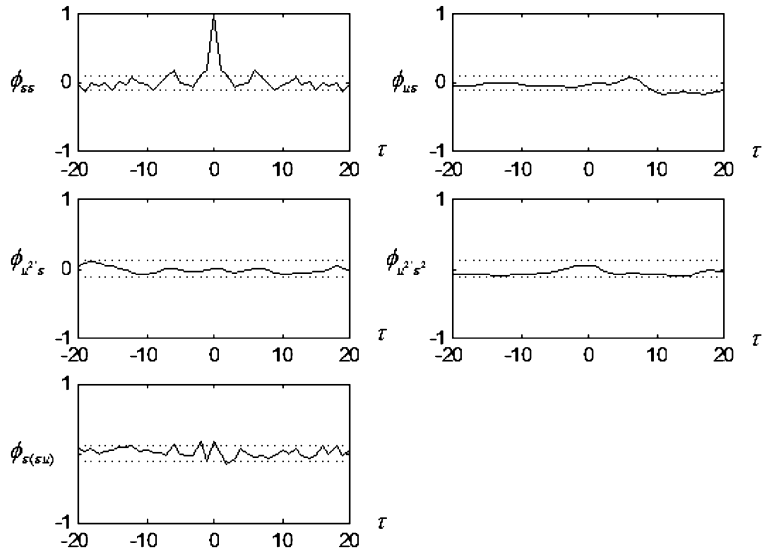

Fig. 11 Model validity tests for gas furnace time series data

equal level of the EI. In some of the trials using ratios 1 and 5 in simulated model 2, the final solution has a high EI. Regarding the OF, although ratio 5 has a low EI in simulated model 1, its solution is non-parsimonious compared with other ratios. Thus far, ratios 2 to 4 are more promising in producing a more accurate and parsimonious solution. The maximum number of generations set for the ratios seems to play an important role that may affect the whole optimization outcome.

The results of the mean and standard deviation of the generation count are presented again in Fig. 12 and analysed. Without affecting previous observations, trials that theoretically need more generations to produce accurate and parsimonious solution are set to 100 for the generation count in the graph. They are connected by smoothed lines in an attempt to define the trend throughout the ratios. This also means that the mean graph (Fig. 12(a)) must be looked at first before any further refinement

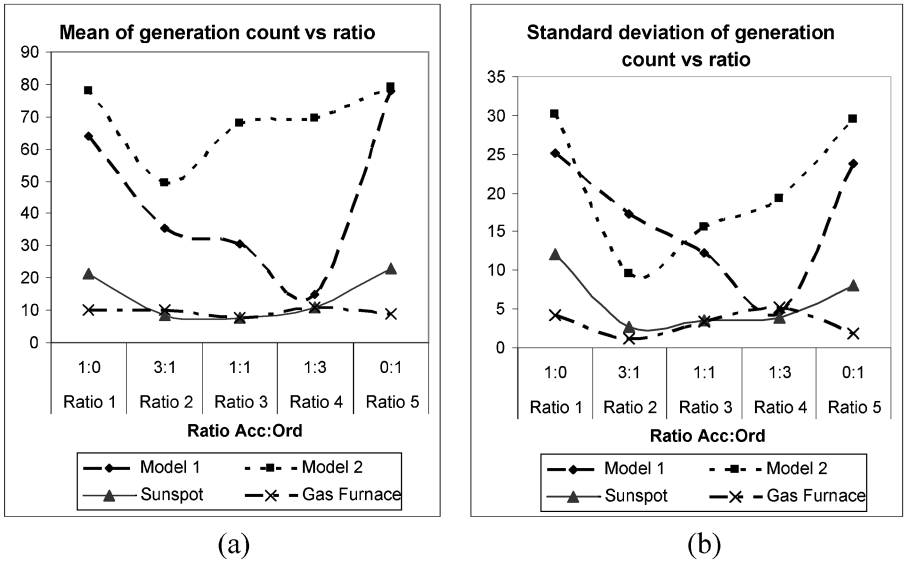

Fig. 12 (a) Mean of generation count versus ratio and (b) standard deviation of generation count versus ratio

of the conclusion using the standard deviation graph. The simulated systems clearly persist on ratio 4 for simulated model 1 and on ratio 2 for simulated model 2. Their trends as the generation count increases are alike. Observing the results of the Wölfer sunspot time series data seems to show that ratio 3 is the best ratio in terms of mean. As for the gas furnace time series data, ratio 3 again gives a better performance than even neighbouring ratios.

Meanwhile, in the standard deviation graph, generally, when the ratio is the best for the data, its consistency is also better than that of other ratios. For the real-world data, to be exact, consistency is seen in the neighbouring ratio, ratio 2. Overall, study of the simulated systems and real-world time series data gives all three middle ratios as having the smallest means and standard deviations, resulting in a valley in the middle. By using a ratio that consistently needs a minimum number of generations, wastage in computational time can be avoided.

6 CONCLUSIONS

The modification made to the conventional GA by dividing the population into groups to undergo mutual recombination operations is able to speed up identification of the optimum solution by requiring fewer generations. As long as this solution is found, the model parameters are approximated accurately by using the least-squares method. The higher convergence rate allows more effective use of computation time and costs than using the simple GA. In the comparison of the correct ratio, the study reveals that the ratio should be somewhere between 1 : 3 and 3 : 1 in terms of acceptable group size to ordinary group size. As a conclusion, it is safe and optimistic to say that an equal size – ratio 3 or 1 : 1 – should be selected on the basis of the outcome of Fig. 12 where the lowest points tend to be in the middle. A suitable selection of penalty value provides a more parsimonious model. An alternative approach to the study is to use the same initial population every time a trial is made. A different setting to the algorithm is also expected to yield a better overall conclusion.

ACKNOWLEDGEMENTS

The authors wish to acknowledge all the support of Universiti Teknologi Malaysia and Universiti Teknikal Malaysia Melaka throughout this research, especially for the UTeM-SLAB sponsorship.

REFERENCES

1 **Holland, J. H.** *Adaptation in natural and artificial systems*, First edition, 1975 (The University of Michigan); 1st MIT Press edition, 1992 (Massachusetts Institute of Technology, Ann Arbor).
2 **Zalzala, A. M. S.** and **Fleming, P. J.** (Eds) *Genetic algorithms in engineering systems*, 1997 (The Institution of Electrical Engineers, London, UK).
3 **Goldberg, D. E.** *Genetic algorithms in search, optimization and machine learning*, 1989 (Addison-Wesley Publishing Company, Reading, Massachusetts).

4 **Gen, M.** and **Cheng, R.** *Genetic algorithm and engineering design*, 1997 (John Wiley, New York).

5 **Sakaguchi, A.** and **Yamamoto, T.** A study on system identification using GA and GMDH network. In Proceedings of the 29th Annual Conference of the IEEE Industrial Electronics Society, *IECON'03*, Roanoke, Virginia, 2–6 November 2003, vol. 3, pp. 2387–2392.

6 **Yen, T. G., Kang, C. C.,** and **Wang, W. J.** A genetic based fuzzy-neural networks design for system identification. In Proceedings of the IEEE International Conference on *Systems, man and cybernetics*, Hawaii, 10–12 October 2005, vol. 1, pp. 672–678.

7 **Tan, X.** and **Yang, H.** The optimization of nonlinear systems identification based on genetic algorithms. In Proceedings of the International Conference on *Computational intelligence and security*, Gvangzhai, People's Republic of China, November 2006, vol. 1, pp. 266–269.

8 **Juang, J.** Application of genetic algorithm and recurrent network to nonlinear system identification. In Proceedings of 2003 IEEE Conference on *Control applications*, Istanbul, Turkey, 23–25 June 2003, vol. 1, pp. 129–134.

9 **Hatanaka, T., Uosaki, K.,** and **Koga, M.** Evolutionary computation approach to block oriented nonlinear model identification. In Proceedings of the Fifth Asian Control Conference, Melbourne, Australia, 20–23 July 2004, vol. 1, pp. 90–96.

10 **Abbas, H. M.** and **Bayoumi, M. M.** Volterra-system identification using adaptive real-coded genetic algorithm. *IEEE Trans. Syst., Man and Cybernetics, Part A*, 2006, **36**(4), 671–684.

11 **Tcholakian, A. B., Martins, A., Pacheco, R. C. S.,** and **Barcia, R. M.** Fuzzy system identification through hybrid genetic algorithm. In Proceedings of the 1997 Annual Meeting of the North American Fuzzy Information Processing Society, *NAFIPS'97*, Syracuse, New York, 21–24 September 1997, pp. 428–433.

12 **Pereira, D. S.** and **Pinto, J. O. P.** Genetic algorithm based system identification and PID tuning for optimum adaptive control. In Proceedings of 2005 IEEE/ASME International Conference on *Advanced intelligent mechatronics*, California, USA, 24–28 July 2005, pp. 801–806.

13 **Lee, K. Y.** and **Syed Mohamed, P.** A real-coded genetic algorithm involving a hybrid crossover method for power plant control system design. In Proceedings of 2002 Congress on *Evolutionary computation, CEC'02*, Piscataway, New Jersey, 12–17 May 2002, vol. 2, pp. 1069–1074.

14 **Iwasaki, M., Miwa, M.,** and **Matsui, N.** GA-based evolutionary identification algorithm for unknown structured mechatronic systems. *IEEE Trans. Ind. Electronics*, 2005, **52**(1), 300–305.

15 **Mitsukura, Y., Fukumi, M., Akamatsu, N.,** and **Yamamoto, T.** A system identification method using a hybrid-type genetic algorithm. In Proceedings of 41st SICE Annual Conference, *SICE 2002*, Osaka, Japan, 5–7 August 2002, vol. 3, pp. 1598–1602.

16 **Duong, V.** and **Stubberud, A. R.** System identification by genetic algorithm. IEEE Aerospace Conference Proceedings, Big Sky, Montana, 2002, vol. 5, pp. 2331–2337.

17 **Ahmad, R., Jamaluddin, H.,** and **Hussain, M. A.** Multivariable system identification for dynamic discrete-time nonlinear system using genetic algorithm. In Proceedings of the IEEE International Conference on *Systems, man and cybernetics*, Hammamet, Tunisia, 6–9 October 2002, vol. 5.

18 **Eshelman, L. J.** Genetic algorithms. In *Evolutionary computation 1: Basic algorithms and operators* (Eds T. Bäck, D. B. Fogel, and Z. Michalewicz), 2000, ch. 8 (IOP Publishing Ltd).

19 **Perry, M. J., Koh, G. C.,** and **Choo, Y. S.** Modified genetic algorithm approach to structural identification. The Third International Conference on Structural stability and dynamics, Orlando, USA, 19–22 June 2005.

20 **Ahmad, R., Jamaluddin, H.,** and **Hussain, M. A.** Model structure selection for a discrete-time nonlinear system using genetic algorithm. *Proc. Instn Mech. Engrs, Part I: J. Systems and Control Engineering*, 2004, **218**, 85–98.

21 **Grefenstette, J. J.** Optimization of control parameters for genetic algorithms. *IEEE Trans. Syst., Man and Cybernetics*, 1986, **16**(1), 122–128.

22 **Wei, H. L., Billings, S. A.,** and **Liu, J.** Term and variable selection for non-linear system identification. *Int. J. Control*, 2004, **77**(1), 86–110.

23 **Michalewicz, Z.** *Genetic algorithms + data structures = evolution programs*, Second edition, 1994 (Springer-Verlag, Berlin–Heidelberg).

24 **Veres, S. M.** *Structure selection of stochastic dynamic systems*, 1991 (OPA (Amsterdam) BV).

25 **Ljung, L.** *System identification: theory for the user*, 2nd edition, 1999 (Prentice-Hall PTR, New Jersey).

26 **Johansson, R.** *System modeling and identification*, 1993 (Prentice-Hall, Englewood Cliffs, New Jersey).

27 **Holland, J. H.** Genetic algorithms and the optimal allocation of trials. *SIAM J. Computing*, 1973, **2**(2), 88–105.

28 **Bäck, T., Fogel, D. B., Whitley, D.,** and **Angeline, P. J.** Mutation operators. In *Evolutionary computation 1: Basic algorithms and operators* (Eds T. Bäck, D. B. Fogel, and Z. Michalweicz), 2000, ch. 32 (IOP Publishing Ltd).

29 **Billings, S. A.** and **Voon, W. S. F.** Correlation based model validity tests for non-linear models. *Int. J. Control*, 1986, **44**(1), 235–244.

30 **Billings, S. A.** and **Zhu, Q. M.** Nonlinear model validation using correlation tests, *Int. J. Control*, 1994, **60**(6), 1107–1120.

31 **Box, G. E. P., Jenkins, G. M.,** and **Reinsel, G. C.** *Time series analysis forecasting and control*, 1994 (Prentice-Hall, Englewood Cliffs, New Jersey).

32 **Chen, S.** and **Billings, S. A.** Representations of non-linear systems: the NARMAX model. *Int. J. Control*, 1989, **49**(3), 1013–1032.

33 Yao, X. Evolutionary computation: a gentle introduction. In *Evolutionary optimization* (Eds R. Sarker, M. Mohammadian, and Y. Xin), 2002, ch. 2 (Kluwer Academic Publishers, Boston).

APPENDIX

Notation

a_i	parameter value
Acc	number of chromosomes in the acceptable group
d	response time delay
$e(t)$	system noise at time t
J	number of the trial
l	degree of non-linearity
L	number of regressors (including a constant)
n	number of insignificant regressors
n_e, n_u, n_y	order of lag for noise, input, and output respectively
N	number of data
Ord	number of chromosomes in the ordinary group
$popsize$	population size
p_c, p_m	crossover and mutation probability
$u(t)$	system input at time t
x	value of generation count
$y(t)$	system output at time t
$\hat{y}(t)$	one-step-ahead predicted output at time t
$\varepsilon(t)$	residual at time t
$\boldsymbol{\theta}$	parameter vector
$\hat{\boldsymbol{\theta}}$	vector of estimated parameters
$\boldsymbol{\phi}$	regressor vector