15th Triennial World Congress, Barcelona, Spain

A HYBRID GA FOR NOX EMISSION MODELLING IN POWER GENERATION PLANTS

Jian-xun Peng, Steve Thompson, Kang Li

School of Mechanical and Manufacturing Engineering
Queen's University of Belfast, Ashby Building, Stranmillis, Rd.
Belfast BT9 5AH, U.K.
Email: j.peng@qub.ac.uk, steve.thompson@qub.ac.uk, k.li@qub.ac.uk

Abstract: This paper reviews a grey-box (GB) modelling method that uses genetic algorithms (GA). The GA-GB modelling framework is used for finding NOx emission from coal-fired power generation boilers using operator controlled variables. The main contribution of this paper is the inclusion of a distributed elitist scheme within the GA. This enables a local gradient-based optimisation routine to be incorporated within the GA. The new hybrid GA-GB modelling procedure is shown to be able to produce NOx models having similar performance to those of the original method but requires less computational effort. *Copyright © 2002 IFAC*

Keywords: Power station control, Modelling, Genetic algorithms, Gradient methods, Identification.

1. INTRODUCTION

Coal-fired power plants are one of the main sources of NOx emission. NOx contributes significantly to photochemical smog, acidification, ground-level ozone (O_3), eutrophication of water and soils, and has serious implications for public health. For these reasons cleaner coal technologies and the reduction of NOx emissions is a matter of public interest and worldwide legislation.

In order to optimise a power plant's operational conditions, a number of advanced control techniques have been proposed. An ETSU report (1997) describes five advanced boiler control methods that claim to reduce NOx emission. Each could be considered as consisting of two stages. In the first stage some form of plant modelling captures the relationship between the plants operational inputs and the NOx output. In the second stage some form of constrained optimisation is used to manipulate the inputs of the model in order to minimise the NOx output. These values are then presented to the operator (open-loop mode) or in some cases used to automatically adjust the inputs (closed-loop mode). It is claimed that the various techniques will produce NOx reductions between 15% and 25%.

It is obvious that in such a NOx reduction system a good model is vital. Specifically, the model for on-line

purposes must be simple enough to be updated and optimised on-line, and yet complex enough to predict NOx production precisely. Currently artificial neural networks (ANN) (Li and Thompson, 2000), grey box (Li and Thompson, 2001), linear and non-linear identification (Li and Thompson, 1996) techniques may be used.

Li and Thompson (2001) have developed a fundamental grey-box (GB) modelling technique for NOx emission modelling which, when test on an operational power plant, was found to be superior to ANN, linear or non-linear modelling methods. Essentially, the GB method (see section 3) uses the NOx formation equations to identify fundamental terms and their relationship to the plants operational inputs. These terms, individually or when multiplied together, may be combined to produce a model structure. This involves various searches to determine which terms should be used in the model and to determine any coefficients and parameters.

Recently, Peng, et al. (2001) produced a formal modelling framework for the GB method by incorporating a genetic algorithm (GA) search. This paper considers an improvement in this GA search that reduces the computational load during the modelling process while produces models with equivalent predictive capability.

2. GENETIC ALGORITHMS

GA's are recognised for producing globally optimal solutions to complex search problems (multi-modal, non-linear, high dimensional, discontinuous). Also, they are robust and work with raw objectives (Glodberg, 1989). However, the computational time required can be excessive and, due to discrete random sampling over the solution space, a GA will normally only approximately determine the location of the global optimum even in a continuous optimisation problem.

These characteristics when applied to the GB modelling method tend to identify a near optimal solution whilst permitting flexibility in the use of an integrated performance index.

Despite research on genetic structure, solution encoding, advanced genetic operators, and finely tuned genetic parameters it seems unlikely that genetic manipulation alone will simultaneously improved solution accuracy and speed. Hence the interest in combining GA's with other search methods.

Li and Aggarwal (2000) combined a relaxed GA with a local gradient technique. The best GA solution obtained after a small number of generations is handed over to a gradient method that performs local refinement. And an elitist scheme (ES) is employed in the relaxed GA to speed up the global search. However the best GA solution from a small number of generations is not necessarily located in the neighbourhood of the global optimum, and two chromosomes which are very close to each other might locate within the neighbourhood of a same local optimum.

To overcome these problems, a new distributed elitist scheme (DES) is proposed. In the original ES the best chromosomes are preserved at each generation. In the DES selection is based on the best chromosomes and their relative location. That is, each chromosome should preserve a threshold distance from its neighbour. Further, the elitist set is guided using a local gradient-based optimisation routine during the evolution of GA. In this way the best characteristics of the gradient-based techniques (efficiency, fitness) and the GA's (ability of identifying global optimum, robustness) are combined efficiently.

3. GA-GB NOX EMISSION MODELLING

The GB method is based on a set of candidate terms (Thompson and Li, 2001), derived from known NOx formation mechanisms. For each of the n_T terms the relationship between NOx production and the coal-fired power station's operational conditions are established. Let these terms be denoted by:

$$j_i = j_i\left(u_1, u_2, \cdots, u_{n_u}; c_{i1}, c_{i2}, \cdots, c_{in_{ci}}\right), i = 1,2,\cdots,n_T \qquad (1)$$

where j_i's are linear or non-linear functions of the boilers operational inputs, $u_1, u_2, \cdots, u_{n_u}$, (and would include such things as the mass flow of coal (m_f), mass flow of primary air (m_{pa}), mass flow of secondary air (m_{sa}), burner tilt position (q), etc.). The i'th term j_i is a function of the inputs (u) and parameters (c). Examples of such terms are; $(m_f)^{c1}$, $\exp(c_2/m_f)$, $(m_{pa})^{c3}$, $(m_f)^{c4}(m_{pa})^{c5}$, etc.,

Using a selection vector **S** and a parameter vector **C**, potential models are constructed by picking terms from a selection set containing time sequences of

measured NOx values and the candidate terms defined in (1). The selection vector $\mathbf{S}=[s_1, s_2, \cdots s_m]^T$, $s_i \in [0, n_s-1], i=1,2,\cdots,m$, is a vector of integer numbers indicating the position of the *m* items chosen from the original selection set and $\mathbf{C}=[c_1, c_2, \cdots c_{nc}]^T$ is a vector of real numbers containing the parameters associated with the selected terms. The model obtained from **S** and **C** takes the NARX (Nonlinear AutoRegressive with eXogeneous) form:

$$y(t)=q_0+\sum_{i=1}^{n_y} q_i y\left(t-k_{yi}\right)+\sum_{i=n_y+1}^{m} q_i w_i(t)+e(t) \qquad (2)$$

where n_y is the number of selected terms in the NOx output, k_{yi}'s are time lags, $w_i(t)$ is the *i*'th selected candidate term. The *m* regressive coefficients, $q_i, i=0,1,2,\cdots,m$, are determined by least-squares optimisation over a set of training data.

To assess a constructed model, an overall model performance index was proposed in (Peng *et al*., 2001) as follows:

$$V(\mathbf{S};\mathbf{C})=w_T V_T(\mathbf{S};\mathbf{C})+w_V V_V(\mathbf{S};\mathbf{C})+w_L V_L(\mathbf{S};\mathbf{C}) \qquad (3)$$

where

$$V_T(\mathbf{S};\mathbf{C})=\frac{w_0}{N_T}\sum_{\Omega_T}\left[\hat{y}(t)-y(t)\right]^2 \qquad (4)$$

$$V_V(\mathbf{S};\mathbf{C})=\sum_{i=1}^{n_v}\frac{w_i}{N_i}\sum_{\Omega_{Vi}}\left[\hat{y}(t)-y(t)\right]^2 \qquad (5)$$

$$V_L(\mathbf{S};\mathbf{C})=\frac{w_0}{N_T}\sum_{\Omega_T}\left[\hat{y}_L(t)-y(t)\right]^2+ +\sum_{i=1}^{n_v}\frac{w_i}{N_i}\sum_{\Omega_{Vi}}\left[\hat{y}_L(t)-y(t)\right]^2 \qquad (6)$$

are performance indices of training, validation and long-term prediction, respectively, w_T, w_V and w_L are corresponding weighting factors, $\hat{y}$ and $\hat{y}_L$ are respectively the one-step-ahead and long-term prediction of output *y*:

$$\hat{y}(t)=q_0+\sum_{i=1}^{n_y} q_i y\left(t-k_{yi}\right)+\sum_{i=n_y+1}^{m} q_i w_i(t), \qquad (7)$$

$$\hat{y}_L(t)=q_0+\sum_{i=1}^{n_y} q_i \hat{y}_L\left(t-k_{yi}\right)+\sum_{i=n_y+1}^{n} q_i w_i(t), \qquad (8)$$

where Ω_T denotes the set of training data with N_T samples, Ω_{Vi} the *i*'th validation data set, and n_v is the number of validation sets. $N_i, i=1,2,\cdots,n_V$ are the number of samples in the *i*'th validation set. $w_i, i=0,1,\cdots,m$ are user chosen weights indicating the importance attached to the corresponding data set of Ω_T or Ω_{Vi}'s.

The modelling problem is then turned into a minimisation problem:

$$\min_{\mathbf{S},\mathbf{C}} V(\mathbf{S};\mathbf{C}) \qquad (9)$$

That is find a selection vector **S** and a corresponding parameter vector **C** in the solution space, which minimise the modelling performance index.

When using GA's to solve this type of mixed optimisation problem (9), parameters **S** and **C** may be represented using chromosomes (Peng *et al*., 2001). That is in each chromosome, the first *m* integer genes corresponding to **S** and the rest are float-point genes encoding all the term-contained parameters in all the candidate terms (1). When applying genetic operation on the integer genes, the same integer gene value is not permitted in an individual chromosome.

To evaluate a chromosome, which corresponds to a potential practical model, a selection vector **S** is constructed with its *m* integer genes, and a parameter vector **C** is formed with its floating-point genes. The chromosome can then be evaluated with objective function (3).

The method used to identify the optimal model is:

a) Initialisation of the population.
   Randomly create an initial population with *N* chromosomes.

b) Evaluation.
   Evaluate each chromosome of the population by the overall model performance function (3).

c) Ranking and selection.
   The model performance index is mapped into the chromosome fitness by a ranking function. Selection is the process of determining the number of times a particular chromosome is chosen for reproduction. The selected chromosomes are put into a mating pool.

d) Crossover and mutation.
   Both crossover and mutation are performed in the mating pool. Crossover is the basic GA operator for producing new chromosomes. Like its counterpart in nature, crossover produces new individuals that have genetic material obtained from both parents'. A uniform crossover scheme is proposed for integer genes.

For floating-point genes the offspring are produced from intermediate recombination:

$$\begin{cases} G_{o1} = G_{p1} + a\,(G_{p2} - G_{p1}) \\ G_{o2} = G_{p2} + a\,(G_{p1} - G_{p2}) \end{cases}$$

where G_{o1} and G_{o2} are the offspring genes obtained from the parents genes G_{p1} and G_{p2}, α is a scaling factor uniformly generated for each at random over a given interval $[L_c, U_c]$.

Mutation is randomly applied but with a low probability (mutation rate P_m). For integer genes, a random integer number is generated uniformly over interval $[0, n_s - 1]$ to replace the mutated gene. For floating-point genes, a random number generated uniformly over interval $[-r_m R, r_m R]$ is added to the gene, with R being the variation range of the corresponding gene and r_m being a given mutation step ratio.

e) Replacement.
After crossover and mutation the mating pool replaces the current population. A new generation is then generated. In this paper, to accelerate and improve the search of the base GA procedure it is combined with an elitist scheme. That is the best N_e chromosome(s) is not replaced by the offspring. In this case the size of the mating pool is set to N-N_e, where N is the population size, N_e is a specified number of elitists.

f) Termination of the GA.
If a termination criterion is satisfied, the evolution process is exited. Otherwise, the evolutionary process is continued from step (b). A common practice is to terminate the GA after a specified number of generations.

4. IMPLIMENTATION

4.1 Design of the DES

A conventional elitist scheme is designed to preserve the fittest chromosomes at each generation (Man *et al.*, 1999). In this way, the evolution process will not disrupt potential high-performance building blocks, with the potential to form a better string. However, if the N_e elitists are similar (close to each other in the solution space), the building blocks contained in the elitists may be repeated and reduce the effectiveness of the scheme.

To overcome this problem a threshold distance, s_s between the N_e elitists is introduced as follows:

a) Sort the population according to the fitness of the chromosomes and let the initial elitist set be empty.

b) If the size of the elitist set equals N_e, stop the elitist selection process. Else continue.

c) From the original population choose the next best chromosome not yet considered and test that its distance from each chromosome already in the elitist set is not less than s_s. If this is the case add the chromosome to the elitist set and loop to step b). Otherwise, repeat step c).

By careful selection of the threshold distance s_s, a diverse population is maintained and more high-performance building blocks are likely to be preserved.

When the elitist set is identified, each of the elitist chromosomes is optimised by a local gradient-based optimisation routine.

4.2 Implementation of gradient-based optimisation

The distributed elitists, say $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_q\}$, identified from the population at each generation can be considered as a set of q peaks. Further, the q peaks can be considered as being located in the neighbourhoods of several local optima in the multimodal solution space. Therefore, a local gradient-based optimisation procedure can be employed to search for the true local optima.

The local optimisation employed in this paper is performed as follows:

$$\mathbf{p}_i(k+1) = \mathbf{p}_i(k) - h\, l_i \nabla f(\mathbf{p}_i(k)), i = 1,2,\cdots,q \quad (10)$$

where, k denotes the iteration variable and $\mathbf{p}_i(0)$ is one of the elitists obtained from the GA search, h is a damping factor dependant on the characteristics of the objective function, step size l_i's are chosen to minimise the norm of the corresponding local gradient $\nabla f(\mathbf{p}_i(k))$ at each iteration.

It should be noted that the objective function (3) is a mixed one with both integer- and real- variables.

Therefore, when performing local optimisation, the integer-type variables, **S**, are fixed as constant. To simplify the notation, the objective with **S** fixed is re-denoted as $f(\mathbf{x})$, where **x** denotes the corresponding parameter vector **C**.

In the neighbourhood of each point $\mathbf{p}_i$, function $f(\mathbf{x})$ can be expanded in Taylor series

$$\begin{aligned} f(\mathbf{p}+\mathbf{r}) &= f(\mathbf{p}) + \sum_i \left.\frac{\partial f}{\partial x_i}\right|_{\mathbf{p}} r_i + \frac{1}{2}\sum_{i,j} \left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{\mathbf{p}} r_i r_j + O(\mathbf{r}^3) \\ &\approx f_0 + \mathbf{b}^T\mathbf{r} + \frac{1}{2}\mathbf{r}^T\mathbf{A}\mathbf{r} \end{aligned} \quad (11)$$

where **p** can be any one of $\mathbf{p}_i$'s in the peak set, and

$$\mathbf{b} = \nabla f\big|_{\mathbf{p}}, [\mathbf{A}]_{ij} = \left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{\mathbf{p}}, i, j = 1,2,\cdots,n \quad (12)$$

where n denotes the dimension of **x**. The matrix **A** is called the *Hessian matrix* of the function at **p**, whose components are the second partial derivatives of the function.

In the approximation of (11), the gradient will change as we move from **p** along some direction **r**, and can easily be calculated according to approximation (11) as

$$\nabla f = \mathbf{b} + \mathbf{A}\cdot\mathbf{r}. \quad (13)$$

In a steepest-descent optimisation algorithm, the solution is moved in the direction of local gradient with step size l as formula (10), i.e. $\mathbf{r} = l\cdot\mathbf{b}$, where l can be any one of l_i's in (10).

For a smoothly continuous function, the local minimum is characterised by zero first order derivatives, or say zero norm gradient. Naturally, we hope that the step size l is chosen such that the local gradient is minimised by each move. Based on this principle, l can be determined as following:

$$\min_l \nabla f^T \nabla f = \min_l [\mathbf{b} + \mathbf{A}\cdot l\cdot\mathbf{b}]^T [\mathbf{b} + \mathbf{A}\cdot l\cdot\mathbf{b}] \quad (14)$$

Thereby, l is determined by letting the derivative of (14) in terms of l be zero as following:

$$l = -\mathbf{b}^T\mathbf{A}\mathbf{b} \big/ \mathbf{b}^T\mathbf{A}^T\mathbf{A}\mathbf{b}. \quad (15)$$

Since the derivative information of the objective function cannot be obtained analytically the first and second order partial derivatives are calculated by use of the two-point finite numerical differential formulae:

$$\begin{cases} \dfrac{\partial f(\mathbf{x})}{\partial x_i} = \dfrac{f(\mathbf{x}+d\mathbf{e}_i) - f(\mathbf{x}-d\mathbf{e}_i)}{d} + O(d^2) \\ \dfrac{\partial^2 f(\mathbf{x})}{\partial x_i^2} = \dfrac{f(\mathbf{x}+d\mathbf{e}_i) + f(\mathbf{x}-d\mathbf{e}_i) - 2f(\mathbf{x})}{d^2} + O(d^2) \end{cases}$$

$$i = 1,2,\cdots,n \quad (16)$$

where d is a differential step, a small positive constant, $\mathbf{e}_i$'s are n-dimensional vectors with the i'th components being unity and the others being zero.

In the presented study, the non-diagonal elements of the Hessian matrix, the mixed second order partial derivatives, are set to be zero and (15) becomes

$$l = -\sum_i b_i^2 a_{ii} \Big/ \sum_i b_i^2 a_{ii}^2 \quad (17)$$

This simplified steepest-descent local optimisation algorithm requires $2n$ additional evaluations of the objective function to perform a one-step move.

5. RESULTS

Comparing its performance with that of the original algorithm tests the effectiveness of the proposed hybrid algorithm. Four sets of data are required to produce and test the NOx emission model. Both modelling programs use data sets 1 and 2 for training and cross validation respectively. Data sets 3 and 4 are unseen and used to test the models. Details of the modelling procedure and data are described in (Peng, *et al*., 2001).

The models produced by the original GA-GB software (referred to as GAB) and the hybrid GA-GB software (HGAB) are used to produce the one-step-ahead and long-term NOx emission prediction results shown in Fig. 1. The error in this figure is defined by:

$$Error = \left(\sum_\Omega [\hat{y}(t) - y(t)]^2 \Big/ N_\Omega \right)^{1/2} \quad (18)$$

where $\hat{y}$ in the above equation is replaced by equation (7) for one-step-ahead and by (8) for long-term prediction. This figure shows that the predictive performance of both models is similar.

Fig. 2 shows the computational efficiency of the new hybrid GA-GB software. From (16), $2n$ function evaluations are required for each step in the local search. Consequently, the numbers of function evaluations for both the algorithms are calculated as follows:

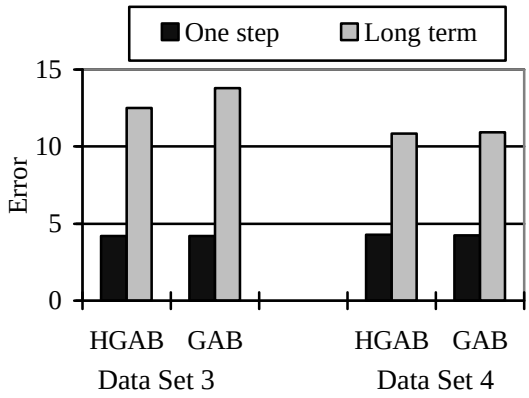

Fig. 1. Comparison of prediction errors over data set 3 and 4.

$FE(\text{HGAB}) = NG + 2nqT = 7040$

$FE(\text{GAB}) = NG = 30000$

The algorithm parameters in this experiment are tuned through test runs to produce the best results and fixed as follows:

HGAB: N=40, G=50, q=2, T=60.
GAB: N=200, G=150.

where T is the number of local search iterations in the HGAB algorithm, G is the number of generations in the GA and N is the population size.

6. CONCLUSIONS

A modelling technique for finding NOx emission from coal-fired power generation boilers using grey box modelling methods and genetic algorithms has been reviewed. It has been shown that by including a distributed elitist scheme within the GA a local gradient-based optimisation routine may be incorporated within the GA. The new hybrid GA-GB modelling procedure is able to produce NOx models of similar predictive performance to those of the original method but require less computational effort.

ACKNOWLEDGEMENTS

Acknowledgement is made to the British Coal Utilisation Research Association and the UK Department of Trade and Industry for a grant in aid of this research but the views expressed are those of the authors, and not necessarily those of BCURA or the Department of Trade and Industry.

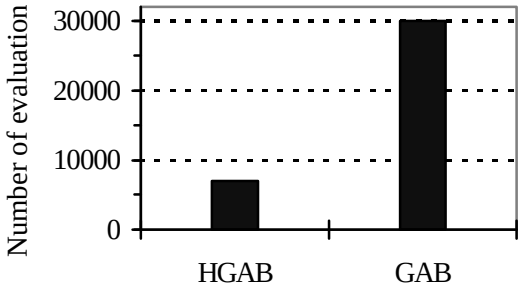

Fig. 2. Comparison of computational load

REFERENCES

ETSU (1997). *Coal R&D programme*. Technology status report: NOx control for puliverised coal-fired power plant, ETSU, Harwell.

Glodberg, David E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Publishing Company, Inc.

Li, K. and Thompson, S. (2000). Developing NOx Emission Model for a Coal-fired Power Generation Plant Using Artificial Neural Networks. UKACC International Conference on CONTROL 2000, Cambridge, 4-7 Sept.

Li, K. and Thompson, S. (2001). Fundamental Grey-box Modelling. European Control Conference, Porto, Portugal, pp.3648-3653.

Li, N. and Thompson, S. (1996). A Simplified Non-linear Model of NOx Emissions in a Power Station Boiler. IFAC Control'96, Exeter UK, **2**, pp 1016-102.

Li, F. and Aggarwal, R.K. (2000). Fast and Accurate Power Dispatch Using a Relaxed Genetic Algorithm and a Local Gradient Technique. *Expert Systems with Application*, **19**, pp.159-165.

Man, K.F., Tang, K.S. and Kwong, S. (1999). *Genetic Algorithms*. Springer-Verlag London Limited.

Peng, J., Li, K. and Thompson, S. (2001). GA based software for power generation plant NOx emission modelling. 6th Int. Conf. Technologies and Combustion for A Clean Environment, **2**, pp.81-887, Porto, Portugal.