TEDI - International Interdisciplinary Journal of Young Scientists from the Faculty of Textile Technology

Optimizing configurable parameters of model structure using genetic algorithms

Ž. Ujević Andrijić*, N. Bolf*, T. Rolich**

**Faculty of Chemical Engineering and Technology, University of Zagreb*
Department of Measurements and Process Control, Savska c. 16/5A, 10000 Zagreb
(e mail: zujevic@fkit.hr)

***Faculty of Textile Technology, University of Zagreb*
Department of Basic Natural and Technical Sciences, Prilaz Baruna Filipovića 28A, 10000 Zagreb
(e mail: tomislav.rolich@ttf.hr)

Abstract: Fractionation product properties of crude distillation unit (CDU) need to be monitored and controlled through feedback mechanism. Due to inability of on-line measurement, soft sensors for product quality estimation are developed. Soft sensors for kerosene distillation end point are developed using linear and nonlinear identification methods. Experimental data are acquired from the refinery distributed control system (DCS) and include on-line available continuously measured variables and laboratory assays. In present work development of AutoRegressive Moving Average with eXogenous inputs (ARMAX) and Nonlinear AutoRegressive model with eXogenous inputs (NARX) are presented. To overcome the problem of selecting the best model set parameters by trial and error procedure, genetic algorithms were used for optimizing the best set of model parameters. Genetic algorithms were approved to be suitable method for optimizing ARMAX and NARX model structure in a way to find the best fits for given parameters range. Based on developed soft sensors it is possible to estimate fuel properties continuously by embedding model in DCS on site as well as applying the methods of inferential control.

1. INTRODUCTION

Process industry is nowadays faced with the ever-growing requirements for product quality enhancement. Hence, there is a need for the continuous monitoring of large number of process variables and properties.

In various process plants there are variables that are difficult to measure online (viscosity, density, boiling point, color, etc.) or it is only possible to get infrequent laboratory measurements. Those properties are often of great importance for process industry because they may have high influence on the final product quality.

Laboratory analyses can be time consuming and influenced by human factor. Solution of this problem can be found in application of the soft sensors – estimators that estimate infrequent-measured variables on the basis of easy measured variables, such as temperature, pressures, flows, etc. (Dam and Saraf, 2006). As DCS is installed in most chemical plants, many process variables can be measured and stored in real time (Ma *et al*., 2009). That historical database enables engineers to build soft sensors with the goal to produce real-time reliable estimates of unmeasured data.

Typical soft sensor design procedure is presented as follows (Fortuna *et al*., 2007):

1. Selection of historical data from plant database
2. Outlier detection, data filtering
3. Model structure and regressor selection
4. Model estimation
5. Model validation

Step 3, i.e. selection of the optimal model structure is crucial for the soft sensor performance (Kadlec *et al*., 2009). Linear and nonlinear autoregressive types of models are often used for developing soft sensor model. Nonlinear auto regression models, such as nonlinear ARX and Hammerstein-Wiener models, can use tree-partition networks, wavelet networks and multi-layer neural network as nonlinear estimators. Set of parameters need to be adjusted in order to obtain the best performance of each models. Parameters of those kinds of models are often selected in an ad hoc manner. MATLAB system identification (SID) toolbox is one of the most used tools for development of dynamic models. SID toolbox allows defining optimal zero and model pole order for given rang only for ARX (autoregressive) model, while for others cumbersome trial and error procedure should be applied. To overcome the problem of selecting best model order and delays of each input and other configurable parameters, by trial and error procedure, genetic algorithms were used for optimizing the best set of parameters. Applied procedure is shown on the example of ARMAX and nonlinear ARX model.

1.1 Model identification

One of the most used linear form which enables modelling of additive disturbance is ARMAX model:

$$\mathbf{A}(q)y(k)=\mathbf{B}(q)u(k)+\mathbf{C}(q)\xi(k) \qquad (1)$$

where:

$y(k)$ presents output at time k and $u(k)$ is input at time k.

$$\mathbf{A}(q) = \mathrm{I} + \mathbf{A}_1 q^{-1} + \mathbf{A}_2 q^{-2} + \ldots + \mathbf{A}_{\mathrm{na}} q^{-\mathrm{na}}$$

na is maximal number of past values of the output and q is time-shift operator.

$$\mathbf{B}(q) = \mathbf{B}_1 q^{-1} + \mathbf{B}_2 q^{-2} + \ldots + \mathbf{B}_{\mathrm{nb}} q^{-\mathrm{nb}}$$

nb is maximal number of past values of the input.

$$\mathbf{C}(q) = \mathrm{I} + \mathbf{C}_1 q^{-1} + \mathbf{C}_2 q^{-2} + \ldots + \mathbf{C}_{\mathrm{nc}} q^{-\mathrm{nc}}$$

nc is maximal number of past values of the disturbance signal.

ξ is white-noise disturbance value.

Block diagram shown on Fig. 1 represents the ARMAX model structure.

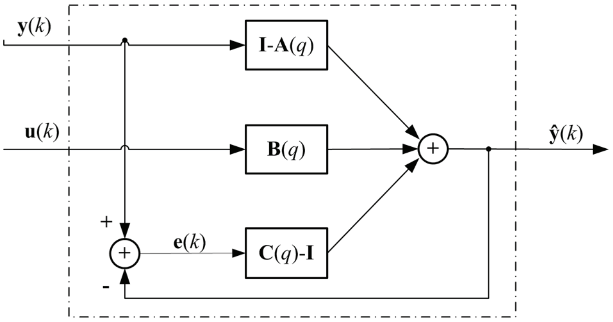

Figure 1. Block diagram representing the ARMAX model structure

Block diagram shown on Fig. 2 represents the structure of a nonlinear ARX model.

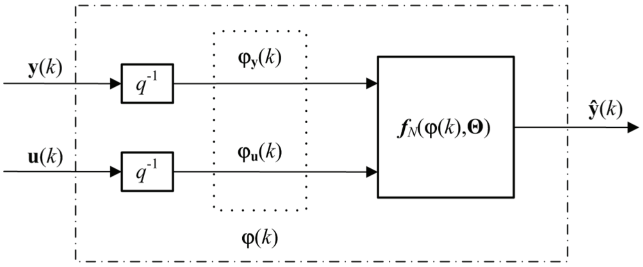

Figure 2. Block diagram representing the structure of a nonlinear ARX model

Most generalized NARX (Nonlinear ARX) model is derived by applying nonlinear regression over the past measured input and output samples:

$$y(k) = f\left(y^{k\text{-}1}, u^{k\text{-}1}\right) + \xi(k) \qquad (2)$$

NARX model predictor is given by:

$$\hat{y}(k) = f_N\left(\varphi(k),\Theta\right) = f_N\left(\left[\varphi y(k), \varphi u(k)\right],\Theta\right) \qquad (3)$$

NARX is a feedforward model, i.e. its regressors do not depend on model parameters so NARX model structure enables simple appliance of static neural networks for approximation of nonlinear function $f_N(\varphi, \Theta)$.

Most nonlinearity estimators represent the nonlinear function as a summed series of nonlinear units, such as wavelet networks or sigmoid functions. In our research as a nonlinear estimator the sigmoid network is used.

Sigma function is given by form:

$$\kappa(s) = \frac{1}{1+e^{-s}} \qquad (4)$$

The network is presented with equation:

$$g(x) = \sum_{k=1}^{n} \alpha_k \kappa\left(\beta_k\left(x-\gamma_k\right)\right) \qquad (5)$$

where β_k is a raw vector such that $\beta_k(x - \gamma_k)$ is a scalar, and n is a number of units.

On the Fig. 3 general prediction model that uses static neural network for identification is presented.

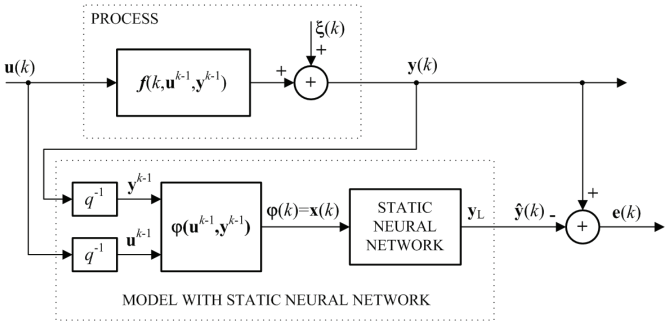

Figure 3. Identification procedure with static neural network

After choosing model structure, optimal configurable parameters values of ARMAX and nonlinear ARX model are determined by genetic algorithms. A genetic algorithm (GA) is an optimization technique used to find exact or approximate solutions to different problems.

Genetic algorithms are implemented in a computer simulation in which a population of chromosomes of candidate solutions to an optimization problem evolves toward better solutions. The evolution starts from a population of randomly generated individuals. In each generation, the fitness of every individual is evaluated, multiple individuals are stochastically selected from the current population based on their fitness and modified (recombined and mutated) to form a new population. The new population is then used in the next iteration of the algorithm. Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population.

2. PROCESS DESCRIPTION

The CDU processes the crude oil entering a refinery. Since the CDU is the first unit in the sequence of refinery processing, it is crucial that the quality of fractionation products (unstabilized naphtha, heavy naphtha, kerosene, light gas oil, heavy gas oil), is monitored and controlled (Cerić, 2006; Chatterjee and Saraf, 2004).

Heavy naphtha, petroleum, and light gas oil fractions are used for blending of diesel fuel. Thereby, very important product property is end boiling point (D95). Section of the column for diesel fuel production with variables used for soft sensor development is given on Fig. 4.

The following variables have been chosen as input variables of soft sensor model for distillation end point:

- column top temperature (T_{TOP}), TR-6104;
- kerosene temperature – 23rd tray (T_K), TR-6197;
- light gas oil temperature – 19th tray (T_{LGO}), TR-6198;
- heavy gas oil temperature–14th tray (T_{HGO}),TR-6199;
- pumparound temperature (T_{PA}), TR-6103 and
- pumparound flow rate (F_{PA}), FI-6130.

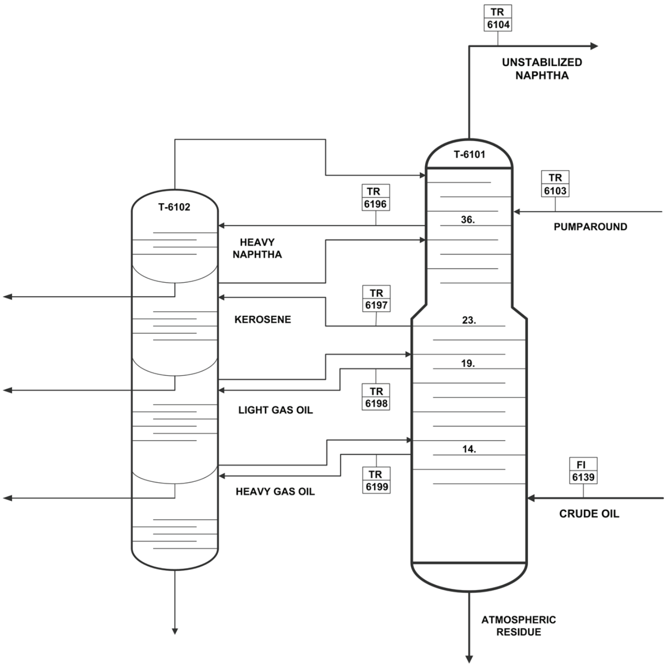

Fig. 4. Crude distillation column section with diesel fuel products

3. SOFT SENSOR MODEL DEVELOPMENT

Plant data have been obtained in the period from January 2009 to September 2009. The laboratory assays of end boiling point are carried out four times a day.

Data preprocessing included detecting and outlier removal, generating additional data by Multivariate Adaptive Regression Splines (MARSplines) algorithm (Salvatore *et al.*, 2009). Data were also detrended for development of ARMAX model.

MATLAB System identification toolbox and Genetic algorithm and Direct Search toolbox were used through this work. Data from historical database was divided into two sets: 60% for modelling (estimation) data set and 40% for validation data set. After the model structure is chosen optimal model structure parameters need to be determined. Configurable parameters of proposed ARMAX and nonlinear ARX structure are shown in Table I.

Table I. Configurable parameters of ARMAX and nonlinear ARX model

Parameter label	Parameter meaning
na	No. of past output terms used to predict the current output.
nb	No. of past input terms used to predict the

	current output (there are six nb in this model according to six inputs).
nk	Delay from input to the output in terms of the number of samples (there are six nk in this model according to six inputs).
nc	No. of past values of the disturbance signal (only for ARMAX model).
n	No. of nonlinear units of sigmoid network (only for nonlinear ARX model).

There are 14 configurable parameters in both proposed models which are all positive integers. For multiple-input systems, nb and nk are presented with matrix form where the i-th row element corresponds to the i-th input, and j-th column element corresponds to the j-th possible value. Parameters na, nc and *n* are row vectors where i-th element corresponds to the i-th input.

Table II. Minimal and maximal values of configurable parameters

Parameter	Min value	Max value
na	0	8
nb for all six inputs	0	8
nk for all six inputs	0	4
nc	0	8
n	1	12

Minimal and maximal configurable parameters values of ARMAX and nonlinear ARX are shown in Table II. They are chosen on the basis of rational complexity of model structure and rational calculating time.

Genetic algorithm parameters from Genetic Algorithm and Direct Search Toolbox are presented in Table III. Those parameters have been chosen based on investigation experiments and experience. In each generation 9 individuals have been created with a crossover procedure, 9 individuals have been created with a mutation procedure, and 1 individuals are elite individuals (individuals with lowest value of fitness function from previous generation).

Search space of configurable parameters for ARMAX model is equal to: $9*9^6*5^6*9$ = 672605015625. Search space of configurable parameters for NARX model is equal to: $9*9^6*5^6*12$ = 896806687500.

The parameters of the GA used for optimization of ARMAX and NARX model are presented in Table III.

Table III. Genetic algorithm parameters for ARMAX and NARX model

Parameter	Value / property
Population size	20
Number of generation	50
Selection	Stochastic uniform
Crossover	Scattered
Mutation	Uniform
Mutation probability rate	0.5
Fitness scaling	Proportional
Number of elite individuals	1
Crossover fraction	0.5

Models are evaluated based on *FIT* value (Matlab, 2009). The fitness function used in the present study is calculated, as follows:

$$FIT = \left(1 - \frac{\sqrt{\sum_{i=1}^{n} \left(\hat{y}_i - y_i \right)^2}}{\sqrt{\sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2}} \right) \cdot 100 \tag{6}$$

y is the measured output, $\hat{y}$ is the simulated or predicted model output, and $\bar{y}$ is the mean of y. 100% corresponds to a perfect fit, and 0% indicates that the fit is no better than guessing the output to be a constant ($\hat{y} = \bar{y}$).

Akaike's Final Prediction Error (*FPE*) criterion is also used for evaluating models. According to Akaike's theory, the most accurate model has the smallest *FPE*. (Ljung, 1999; Matlab, 2009).

FPE is defined by the following equation:

$$FPE = V\left(1 + {}^{2d}\!/\!{}_{N}\right) \tag{7}$$

where V is the loss function, d is the number of estimated parameters, and N is the number of values in the estimation data set.

The loss function V is defined by the following equation:

$$V = \det\left(\frac{1}{N} \sum_{1}^{N} \varepsilon\left(t, \theta_{\mathrm{N}}\right) \left(\varepsilon\left(t, \theta_{\mathrm{N}}\right) \right)^{T} \right) \tag{8}$$

where θ_{N} represents the estimated parameters and ε is output error.

4. RESULTS AND DISCUSSION

Optimized ARMAX model structure parameters with best achieved *FIT* value are presented in following matrix form:

na = [3]

nb = [4 6 6 3 5 6]

nk = [1 2 0 0 0 1]

nc = [1]

Fig. 5 shows plot of the best and mean value of the fitness function for ARMAX model. ARMAX model properties are presented in Table IV.

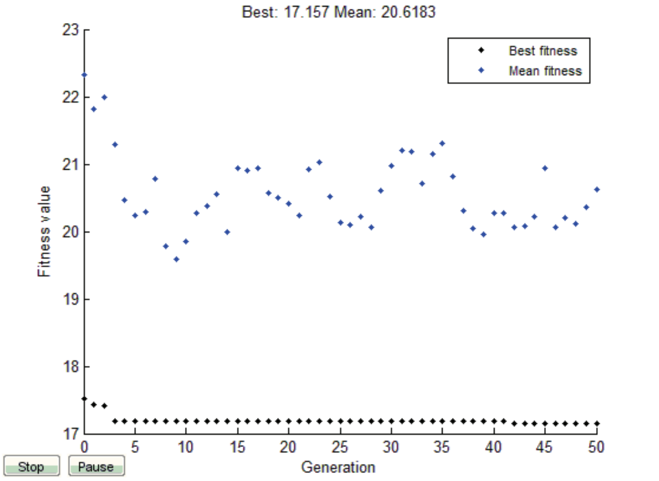

Fig. 5. Plot of the best and mean values of the fitness function at each generation for ARMAX model

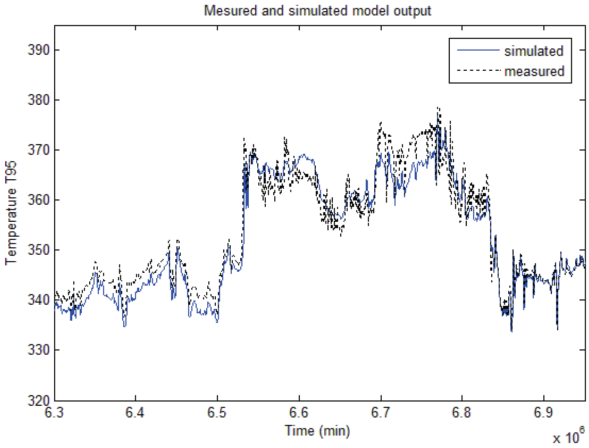

Fig. 6. Comparison between measured and simulated output for validation data, using an ARMAX model

Table IV. ARMAX model properties

No. of inputs	6
No. of outputs	1
V	1,710
FPE	1,716
FIT on validation data (%)	82,84
FIT on 5-step predicted ahead data (%)	87,14

Fig. 6 shows comparison between simulated and measured output for validation data set. It can be noticed that the model output matches the validation data very well.

Due to the complexity and nonlinearity of the distillation process, nonlinear ARX (NARX) model was developed.

Optimized NARX model parameters with best achieved *FIT* value are:

na = [2]

nb = [5 4 1 4 6 1]

nk = [0 0 0 0 2 0]

n = [2]

Fig. 7 shows plot of the best and mean value of the fitness function for NARX model. NARX model properties are presented in Table V.

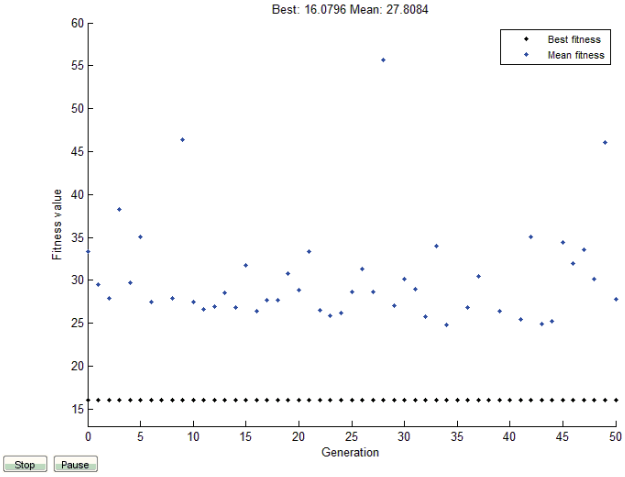

Fig. 7. Plot of the best and mean values of the fitness function at each generation, for NARX model

Table V. NARX model properties

No. of inputs	6
No. of outputs	1
V	1.659
FPE	1.668
FIT on working data / %	86,44
FIT on validation data / %	83,92
FIT on 5-step predicted ahead data (%)	86,32

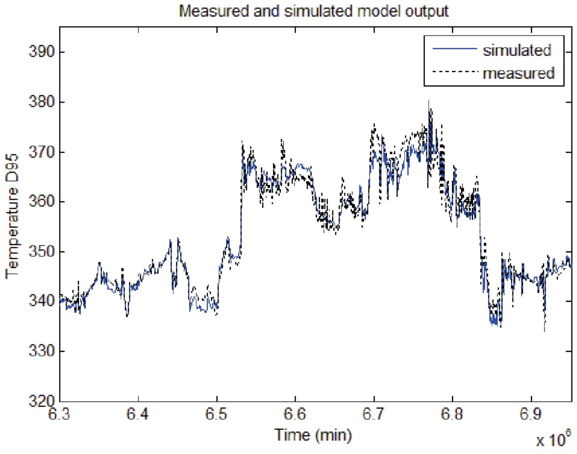

Fig. 8. Comparison between measured and simulated output for validation data, using NARX model with sigmoid network

Fig. 8 shows comparison between simulated and measured output for validation data. From graphical comparison, corresponding *FIT* and *FPE* value it can be observed that experimental and model data show satisfactory agreement.

In presented work only ARMAX and NARX model development is shown because the emphasis is given to application of genetic algorithms for estimating the optimal parameters of autoregressive models. Thereby, cumbersome "guesswork" of parameters is avoided. Developed code is quite general and with some modifications it can be easy adapted for development of other autoregressive identification models. In this work, for fitness function was chosen *FIT* value that is common used measure for evaluation and comparing the model quality. However, fitness function can be some other measure that is used for model evaluation, for example *FPE* or loss function. Moreover, developed code can be adjusted to find optimal parameters with simultaneously maximizing *FIT* and minimizing *FPE*. Presented work showed that evolutionary algorithms can be successfully used for tuning NARX and ARMAX model parameters.

5. CONCLUSION

In this work, dynamic models for estimation end boiling point were developed. Linear and nonlinear models are developed using identification methods. Chosen linear and nonlinear models show satisfactory matching with experimental data, thus improved that can be employed as the soft sensors for on-line estimation and prediction of the key product properties of the crude distillation unit. Using the present procedure it was shown that evolutionary algorithms can be satisfactory applied for optimizing configurable parameters of autoregressive models.

6. REFERENCES

1.) T. Chatterjee, D. N. Saraf, On-line estimation of product properties for crude distillation units Journal of Process Control, Volume 14, 2004, Pages 61-77.

2.) M. Dam, D.N. Saraf, Design of neural networks using genetic algorithm for on-line property estimation of crude fractionator products Computers & Chemical Engineering, Volume 30, 2006, Pages 722-729

3.) B. Kadlec, B. Gabrys, S. Strandt, Data-driven Soft Sensors in the process industry Computers and Chemical Engineering, Volume 33, 2009, Pages 795-814.

4.) M. Ma, J. Ko, S. S. Wang, M. Wu, S. Jang, S. Shieh, D. S. Wong, Development of adaptive soft sensor based on statistical identification of key variables Control Engineering Practice, Volume 17, 2009, Pages 1026–1034

5.) L. Fortuna, S. Graziani, A. Rizzo, M.G. Xibilia, Soft Sensors for Monitoring and Control of Industrial Processes (Advances in Industrial Control*)*, Springer, London, 2007

6.) L. Ljung, System Identification: Theory for the User, 2nd ed., Prentice Hall, New Jersey, 1999

7.) L. Salvatore, M. Bezerra de Souza, M.C.M.M. Campos, Design and Implementation of a Neural Network Based Soft Sensor to Infer Sulfur Content in a Brazilian Diesel Hydrotreating Unit Chemical Engineering Transaction, Volume 17, 2009, Pages 1389-1394.

8.) Matlab The Language of Technical Computing, www.mathworks.com, 2009

9.) E. Cerić, Petroleum – Processes and products, INA and Kigen, Zagreb, (in Croatian), 2006.