Željka Ujević Andrijić, Romano Karlović, Nenad Bolf, Ivana Šarlija

ISSN 0350-350X
GOMABN 52, 2, 122-133
Izvorni znanstveni rad / Original scientific paper

MODELS FOR CONTIONUOS ESTIMATION OF BENZENE IN REFORMATE

Abstract

Due to environmental regulations and production requirement the benzene content in fuels need to be limited. Therefore, it is necessary to continuously monitor the benzene content in light and heavy reformate. As the process analyzers that measure the benzene content in reformate, are often out of service, models of soft sensor are developed for the continuous estimation of benzene content. Soft sensors are developed using linear identification methods and global optimization methods. The development of Finite Impulse Response (FIR)) model and Output Error (OE) model are presented. To overcome the problem of selecting the best model order for multiple input models, by trial and error, genetic algorithms (GA) was used which makes the development of the soft sensors more systematic. Developed models show a satisfactory match with analyzer data on a validation data set. Models are implemented on the fractionation plant for the estimation of benzene content in light reformate.

Key words: soft sensor, linear dynamic model, identification

Introduction

One of the common problems that occur in refinery process monitoring and control is the impossibility of continuous measurement and analysis of key process parameters, in particular the composition of process streams and product characteristics. Cost of the implementation and maintenance of on-line analyzers can be very high. As an alternative, it is possible on the basis of easily measurable quantities to calculate the states of difficult to measure quantities by determining their functional connections. For this purpose, area of inferential measurement and control, and application of software sensors is developed.

Development of software sensors is an area of great interest whereby, mostly on the basis of empirical models, the states of the process that is difficult or impossible to measure can be predicted. The soft sensing technique, developed in recent years, has become a widely used solution. It utilizes variables measurable on-line to predict the product quality variable through certain modeling approaches, such as mechanism modeling, statistical modeling, artificial intelligence modeling, etc.

As distributed control system (DCS) is installed in most chemical plants, many process variables can be measured and stored in real time (Ma et al., 2009). A database containing historical data enables engineers to build soft sensors with a goal of producing reliable estimates of unmeasured data.

Typical soft sensor design procedure is (Fortuna et al., 2007):

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

Differently structured models can be used to model real systems. One possible approach is to start with a simple model structure and gradually increase its complexity. Choosing the optimal model structure and regressors is crucial for soft sensor performance (Kadlec et al., 2009).

In the field of industrial applications, the focus of attention is on parametric (polynomial) model structures. To estimate polynomial models, the model order must be predetermined. Model order can be defined as a number of coefficients for each polynomial included in the selected model structure. Dead time – given by the number of samples before the output corresponds to the input – must also be specified. Depending on the model structure, mentioned and an additional set of parameters may need to be adjusted. In a case of multiple input models, predetermining the set of parameters can become a very complex task. A cumbersome trial and error procedure is therefore commonly applied. To overcome the problem of selecting the best model order, as well as the delays of each input and other configurable parameters, genetic algorithms (GA) were used.

The primary application of soft sensors developed in this case study is a temporary substitution of measuring equipment, either during maintenance or other periods of unavailability.

In this work, Finite Impulse Response (FIR) and Output Error (OE) models were considered. FIR and OE models don't require past samples of the measured output (variable inferred by the soft sensor) when using validation data. Hence, FIR and OE model can be used for back-up of measuring devices.

Model identification

Linear dynamical models are in many cases sufficient for real-life applications (Ljung, 1999). One of the most used linear parametric models is the FIR model which presents the linear regression over the past samples of measured input signals:

$$\hat{y}(t) = \sum_{i=1}^{nu} \mathbf{B}_i(q) \cdot u_i(t - nk_i). \quad (1)$$

$$\hat{y}(t) = \mathbf{B}_1(q) \cdot u_1(t - nk_1) + ... \\ ... + \mathbf{B}_2(q) \cdot u_2(t - nk_2) + \mathrm{K} + \mathbf{B}_{nu}(q) \cdot u_{nu}(t - nk_{nu}) \quad (2)$$

where:

q – the time-shift operator,

$$\mathbf{B}_1(q) = b_{1,1} + b_{1,2}q^{-1} + \mathrm{K} + b_{1,nb_1}q^{-nb_1+1} \ , \quad (3)$$

$$\mathbf{B}_2(q) = b_{2,1} + b_{2,2}q^{-1} + \mathrm{K} + b_{2,nb_2}q^{-nb_2+1} \ , \quad (4)$$

…

$$\mathbf{B}_{nu}(q) = b_{nu,1} + b_{nu,2}q^{-1} + \mathrm{K} + b_{nu,nb_{nu}}q^{-nb_{nu}+1} \ , \quad (5)$$

nu – the number of system inputs,
$\hat{y}(t)$ – model output (predicted output) at time t,
$u_i(t)$ – i-th input at time t,
nk_i – the time delay for the i-th input,
nb_i - number of past input values.
Polynomial coefficients of $\mathbf{B}i(q)$ can be determined using the least-square optimization method or instrumental variable method.
The block diagram shown on Figure 1 represents the FIR model structure.

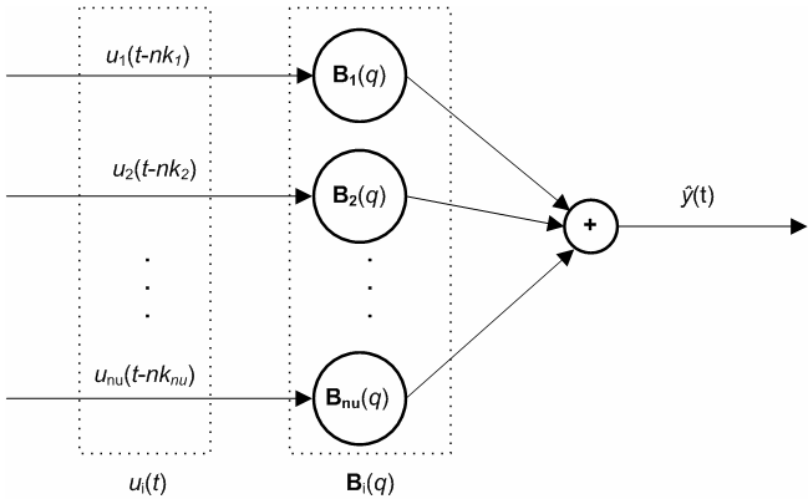

Figure 1: Block diagram of the FIR model

A somewhat extended variant of FIR model is the OE model.

The OE model predictor, in shortened form, is given by:

$$\hat{y}(t) = \mathbf{B}(q)u(t-nk) + \left[\mathbf{I} - \mathbf{F}(q)\right]\hat{y}(t) \quad (6)$$

$$\mathbf{F}(q) = \mathbf{I} + \mathbf{F}_1 q^{-1} + \mathbf{F}_2 q^{-2} + \ ... \ + \mathbf{F}_{nf} q^{-nf} \ , \quad (7)$$

where *nf* is the maximum number of past predicted output.

The block diagram shown on Figure 2 represents the structure of OE model.

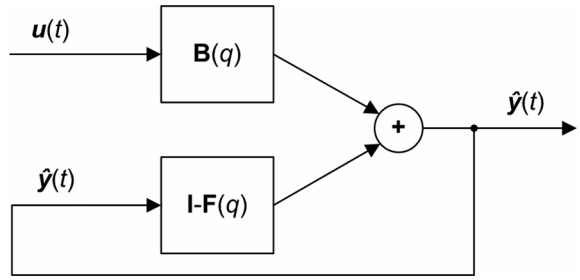

Figure 2: Block diagram of the OE model

Model orders (i.e. the parameters *na*, *nb* and *nk*) of the FIR and OE models are determined by GA. GA is global optimizing technique with a high potential for finding the global optimum of the fitness (objective) function. GA's 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, 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 used in the next iteration of the algorithm. The algorithm terminates when a maximum number of generations has been produced, or a satisfactory fitness level has been reached (Affenzeller et al., 2009).

Process description

70-85% of the benzene in the refinery processes is contributed by reformate from catalytic reforming process. The catalytic reformate is fractioned into light and heavy reformate and the benzene-rich fraction in splitter columns. Although benzene has a high octane number and high calorific values, the benzene content in light reformate need to be reduced to 2,7 mas.%, which is less than 2 vol %. This is due the fact that benzene is a precursor for the formation of cyclohexane in the process of isomerization, and thus an undesirable component of gasoline (low octane number). Also, European emission standards (such as EURO IV and EURO V) for vehicle exhaust emission and MSAT (Mobile Source Air Toxics) regulations limit the amount of benzene in gasoline, due to the hazardous environmental impact.

Hence, the continuous measurement of benzene content in reformate is necessary. Problem with continuous assessment by process analyzers is that such devices besides regular maintenance need also to be often maintained due to contamination of the sampling system and also frequent calibration is desirable.

An overview of a fractionation reformate plant with the variables used for soft sensor development is given on Figure 4. Fractionation of light reformate (top product) takes place in column C1. Fractionation of the benzene-rich fraction (top product) from heavy reformate (bottom product) takes place in column C2.

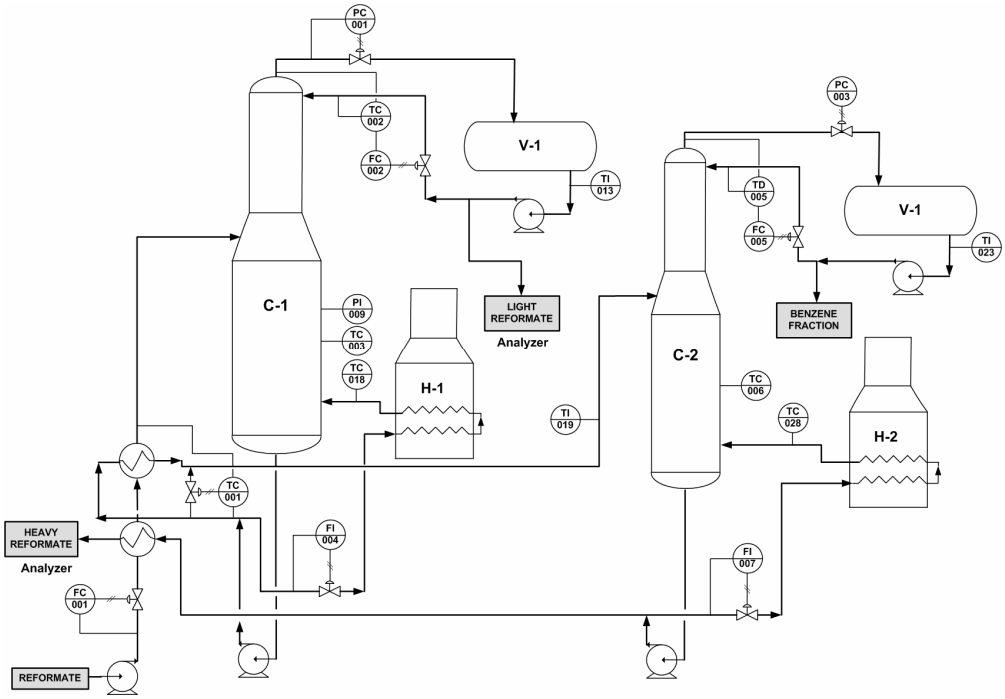

Figure 4: Fractionation reformate plant

On the basis of consultations with the experts from the plant, and appliance of Pearson's coefficients, PCA and PLS analyses, the following continuously measured variables have been chosen as input variables of the soft sensor for the estimation of benzene content in light reformate:

C1 inlet stream temperature, TC-001;
C1 column bottom temperature, TC-018;
C1 column temperature, TC-003;
C1 column pressure, PI-009 and pumparound flowrate, FC-002.

Soft sensor development

Input data was obtained from the plant database over a continuous period of three weeks, with a sampling time of five minutes according to the process dynamics. Data preprocessing included detecting missing data and outlier detection and removal (3 sigma rule). The output data (benzene, vol. %) were taken from the period when the on-line analyzer was properly worked, with a sampling time of twenty minutes. The number of each input data (sampled every five minutes) must corresponds to the number of output data, thus requiring additional output data. This was generated by the Multivariate Adaptive Regression Splines algorithm. Data was also detrended (linear trend was removed from data - using the least squares method and subtracting these values from the measured values) and filtered by Loess filter (*locally weighted scatterplot smoothing*). After data preprocessing, the modeling data set for the estimation and the independent validation data set included 4500 samples and 1500 samples respectively. After data preprocessing, MATLAB System identification toolbox and Global Optimization toolbox were used for the development of soft sensor models.

A. Optimizing model parameters by GA

Since that the search space (solution space) of finding the optimal parameters values are very big, GA method was used. Configurable parameters (of FIR and OE model) and their ranges are shown in Table 1. Parameter ranges have been chosen based on operators' experience of observed process dynamics and preliminary investigation.

Table 1: Configurable parameters of FIR and OE model

Parameter	Parameter description	Minimum	Maximum
nb	No. of past input samples (there are five *nb* in each model according to five inputs).	1	8
nk	Input delay (there are five *nk* according to five inputs).	0	15
nf	No. of past predicted output.	1	5

Table 2 shows the total number of parameters which need to be determined for each model with five inputs. Search space, i.e. the total number of possible combinations of parameters *nb*, *nk* and *nf* is also presented. The aim of GA is to find one combination of parameters among the billions of possible combinations of parameters that will best (or as close to the best) describe the process dynamics.

Table 2: Number of parameters in model and search space

Model	Number of parameters	Search space
FIR	$5nb + 5nk = 10$	$8^5 \cdot 16^5 = 3{,}4360 \cdot 10^{10}$
OE	$5nb + 5nk + 5nf = 15$	$8^5 \cdot 16^5 \cdot 5^5 = 1{,}0737 \cdot 10^{14}$

Important GA parameters, presented in Table 3, have been chosen based on experience, as well as rational calculating time. In the GA evolution, each individual chromosome in the population represents a set of possible model orders. In the proposed GA, 34 individuals have been created in each generation by using a crossover procedure, 15 by using a mutation procedure, and 1 is an elite individual (i.e., with the lowest fitness function value from the previous generation). The algorithm terminated when 60 generations (itarations) had been produced.

Table 3: GA parameters

Parameter	Value / property
Population size	50
Number of generation	60
Function evaluation	3000
Selection	Stochastic uniform
Crossover	Scattered
Mutation	Uniform
Mutation probability rate	0,1
Fitness scaling	Rank
Number of elite individuals	1
Crossover fraction	0,7

B. Model evaluation criteria and Fitness Function Assignment

Models are evaluated based on statistical measures like FIT, FPE and RMS value. The fitness function (FIT) of model is calculated as follows:

$$\mathrm{FIT} = \left(1 - \frac{\sqrt{\sum_{i=1}^{n} \left(\hat{y}_i - y_i \right)^2}}{\sqrt{\sum_{i=1}^{n} \left(y_i - y_m \right)^2}} \right) \cdot 100 \tag{8}$$

y is the measured output, $\hat{y}$ is the simulated or predicted model output, and y_m is the mean of y. 100% corresponds to a perfect FIT. Akaike's Final Prediction Error (FPE) criterion represents the compromise between model accuracy, expressed by the accuracy of estimated parameters, and model complexity. According to Akaike's theory, the most accurate model has the smallest FPE (Verhaegen i Verduit, 2007).

FPE is defined by the following equation:

$$\mathrm{FPE} = V\left(1 + {}^{2d}\!/_{n}\right), \tag{9}$$

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:

$$V = \det\left(\frac{1}{n}\sum_{1}^{n}\varepsilon\left(t,\theta_n\right)\left(\varepsilon\left(t,\theta_n\right)\right)^{T}\right), \tag{10}$$

where θ_n represents the estimated parameters and ε is output error.

Root mean square error (RMS) as frequently used criteria for model evaluation is also presented in the results:

$$\mathrm{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{11}$$

FIT should be as high as possible, and FPE and RMS as low as possible.

Results also include the mean absolute error, e_{MAE} that is given by:

$$e_{\mathrm{MAE}} = \frac{1}{n}\cdot\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right|. \tag{12}$$

The definition of chromosome fitness function is an important element of a GA (Dam i Saraf, 2007). Since preliminary investigation showed that FIT and FPE are not correlated, but both are frequently used criteria for model evaluation, they are integrated in one fitness function *y*, as follows:

$$y = \left(100 - \mathrm{FIT}\right) + 1000\cdot\mathrm{FPE} + 100\cdot\mathrm{RMS} \tag{13}$$

The weighted sum method for multi-objective criteria is used for setting the fitness function. Each criterion is assigned a weighting value (estimated from the previous investigation), and fitness function is a linear combination of all weighted criteria (Venkataraman, 2009; Deb, 2009).

Results

Table 4 shows measurement range and basic statistical measures of input variables and output variable.

Table 4: Descriptive statistics of input and output data

	No. of samples	Mean	Median	Min	Max	Stand. deviation
Input 1, °C	6000	107,40	100,98	89,83	125,35	10,43
Input 2, °C	6000	175,49	175,49	168,80	177,65	0,48
Input 3, °C	6000	125,95	125,74	120,74	138,76	1,66
Input 4, bar	6000	2,02	2,02	1,96	2,09	0,032
Input 5, t/h	6000	41,60	41,31	28,65	62,74	4,01
Output, vol%	6000	1,20	1,27	0,50	1,89	0,29

Figure 5 shows 6000 continuously measured input data and analyzer's output data collected from the process history data base. All results were calculated for a new set of validation data, containing 1500 data. Parameters *nb*, *nf* and *nk* were estimated by minimizing the objective function (13) using the GA procedure, as was discussed earlier. Coefficients of polynomial **B**i(q) and **F**(q) were estimated using the optimization methods integrated with the MATLAB System Identification Toolbox.

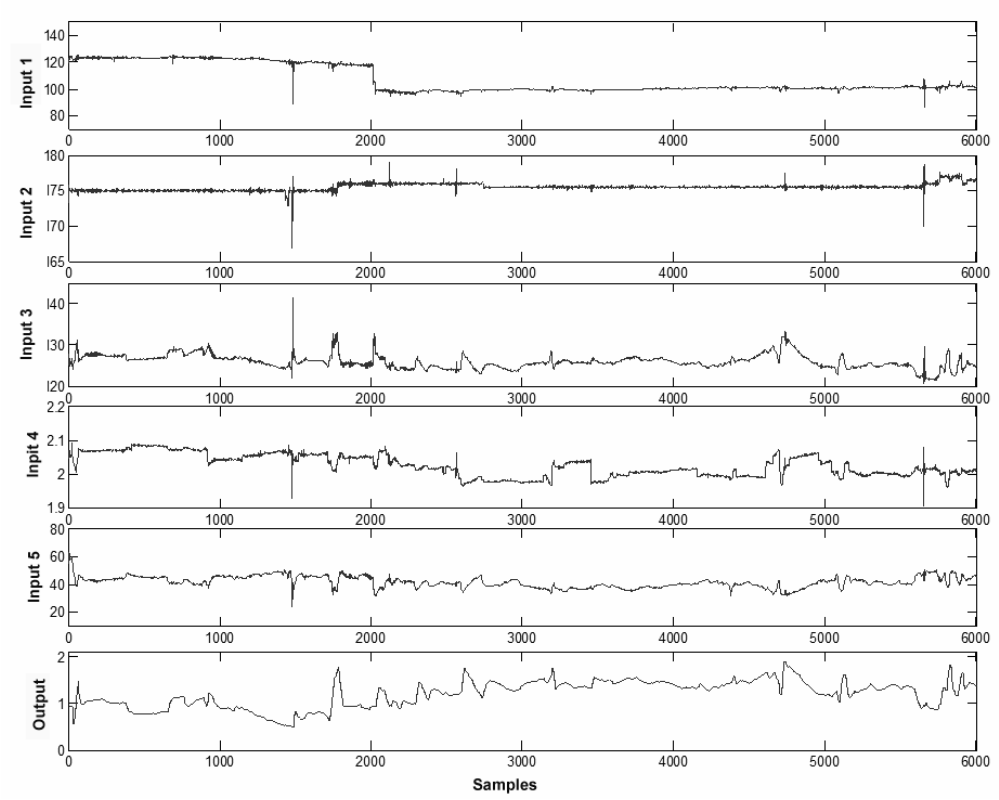

Figure 5: Input and output variables

Model order parameters of the best achieved models are presented in the vector form in table 5. Properties of the FIR and OE model are shown in Table 6.

Table 5: Estimated model orders

	FIR	OE
nb	[6 7 8 8 8]	[6 1 2 3 3]
nk	[5 3 15 5 15]	[1 8 7 9 13]
nf	/	[1 1 2 2 1]

Table 6: Model properties

	FIT	FPE	RMS	e_{MAE}
FIR	78,895	0,0064	0,0517	0,0391
OE	90,267	0,0023	0,0239	0,0171

Although FIR model has a relatively simple model structure, it gives satisfactory results on a validation data set. Compared to the FIR model, the OE model gives better results, with a lower model order (lower *nb*) as was expected, since the past predicted output values are taken. Figures 6 and 7 show the comparison between model and measured output for the validation data set of FIR and OE model, respectively. It can be seen that all model outputs match very good for the validation data. A graphical comparison and the corresponding FIT, FPE and RMS values show that analyzer and model OE data are in accurate agreement.

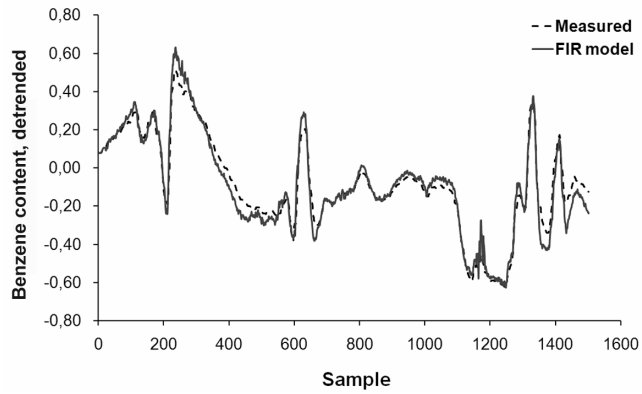

Figure 6: Comparison between analyzer data and FIR model output data

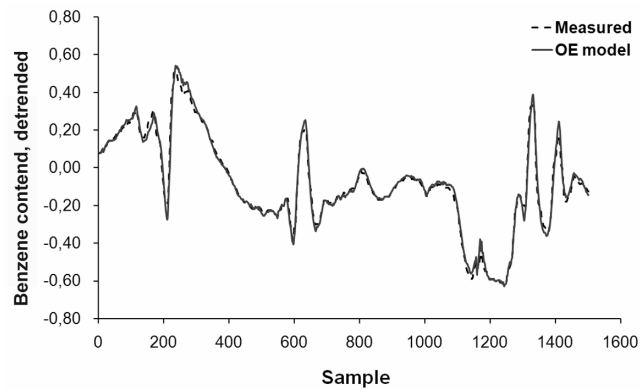

Figure 7: Comparison between analyzer data and OE model output data

OE model obviously gave better results, but OE model depends on past model output, hence, can be unstable. In the case when one input is unavailable for any reason or for interruption period of plant operation, it takes some time for OE to start to correctly calculate the output. Therefore, besides OE model, FIR model should be applied on-site as more reliable model with still satisfied results.

The models are implemented on advanced application module within the distributed control system (DCS). Benzene content is calculated based on dynamic polynomial model using Honeywell Control Language program. Calculated output values are stored in process history database (PHD) from where they are available to plant operators as numeric values or trends. Currently, the validity of the applied model cannot be fully tested at the plant, because the benzene process analyzer (whose purpose now performs model) is for a long time out of service. Therefore, the results obtained by the model, are compared with laboratory analyzes carried out one-two times a day. As it can be seen from the Figure 8, implemented FIR model gave satisfied results to the comparison with the laboratory assays. In future, it remains to do the fine-tuning of the model as soon as process on-line analyzers data become available.

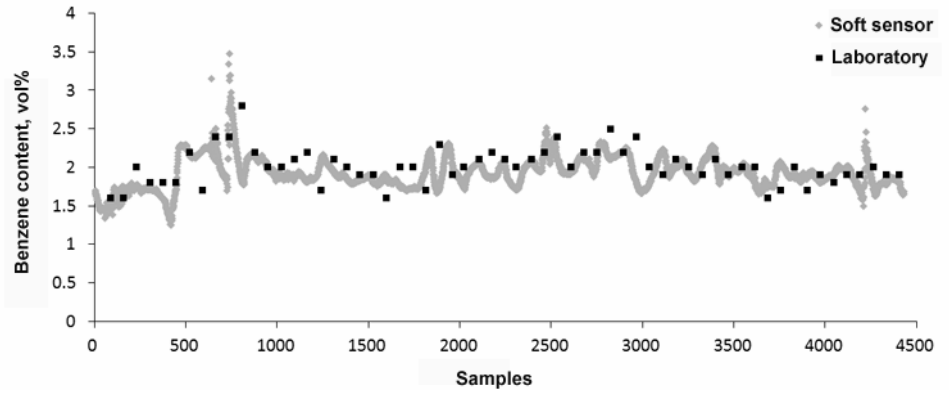

Figure 8: Comparison between laboratory measured and FIR model output data

Conclusion

Linear dynamic models for the estimation of benzene content of light reformate were developed. To avoid a trial and error procedure, GA method was proposed for model order selection, which makes the development of soft sensors more systematic. Chosen models show a satisfactory match with experimental data, thus proving their usefulness as soft sensors for the on-line estimation of benzene content in light reformate. Using the described procedure, it was shown that GA can be satisfactorily applied for optimizing configurable parameters of input-output polynomial models. Both models can be successfully employed as the soft sensor for the on-line prediction of benzene content.

Literature

1. 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 Eng. Pract. 2009, 17, 1026.
2. 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.
3. B. Kadlec, B. Gabrys, S. Strandt, *Data-Driven Soft Sensors in the Process Industry*, Comput. Chem. Eng. 2009, 33, 795.
4. L. Ljung, *System Identification: Theory for the User*; Prentice Hall: New Jersey, 1999.
5. M. Affenzeller, S. Winkler, S. Wagner, A. Beham, *Genetic Algorithms and Genetic Programming - Modern Concepts and Practical Applications*; Taylor & Francis Group: Boca Raton, 2009.
6. M. Verhaegen, V. Verduit, *Filtering and System Identification*; Cambridge University Press: New York, 2007.
7. M. Dam, D. N. Saraf, *Design of neural networks using genetic algorithm for on-line property estimation of crude fractionator products*, Comput. Chem. Eng. 2006, 30, 722.
8. P. Venkataraman, *Applied optimization with MATLAB programming*; John Wiley and Sons: New York, 2009.
9. K. Deb, *Multi-objective optimization using evolutionary algorithms*; John Wiley and Sons: New York 2009.

Authors

Željka Ujević Andrijić[1], Romano Karlović[2], Nenad Bolf[1], Ivana Šarlija[2]
[1] University of Zagreb, Faculty of Chemical Engineering and Technology, Department of Measurement and Process Control, Zagreb, Croatia
e-mail: zujevic@fkit.hr
[2] INA Petroleum Industry, Refinery Rijeka Sector, Process Automation, Kostrena

Received
01.10.2012.

Accepted
05.02.2013.