UNIVERSITY OF ZAGREB, CROATIA
FACULTY OF CHEMICAL ENGINEERING AND TECHNOLOGY

Soft sensor application case study in refinery

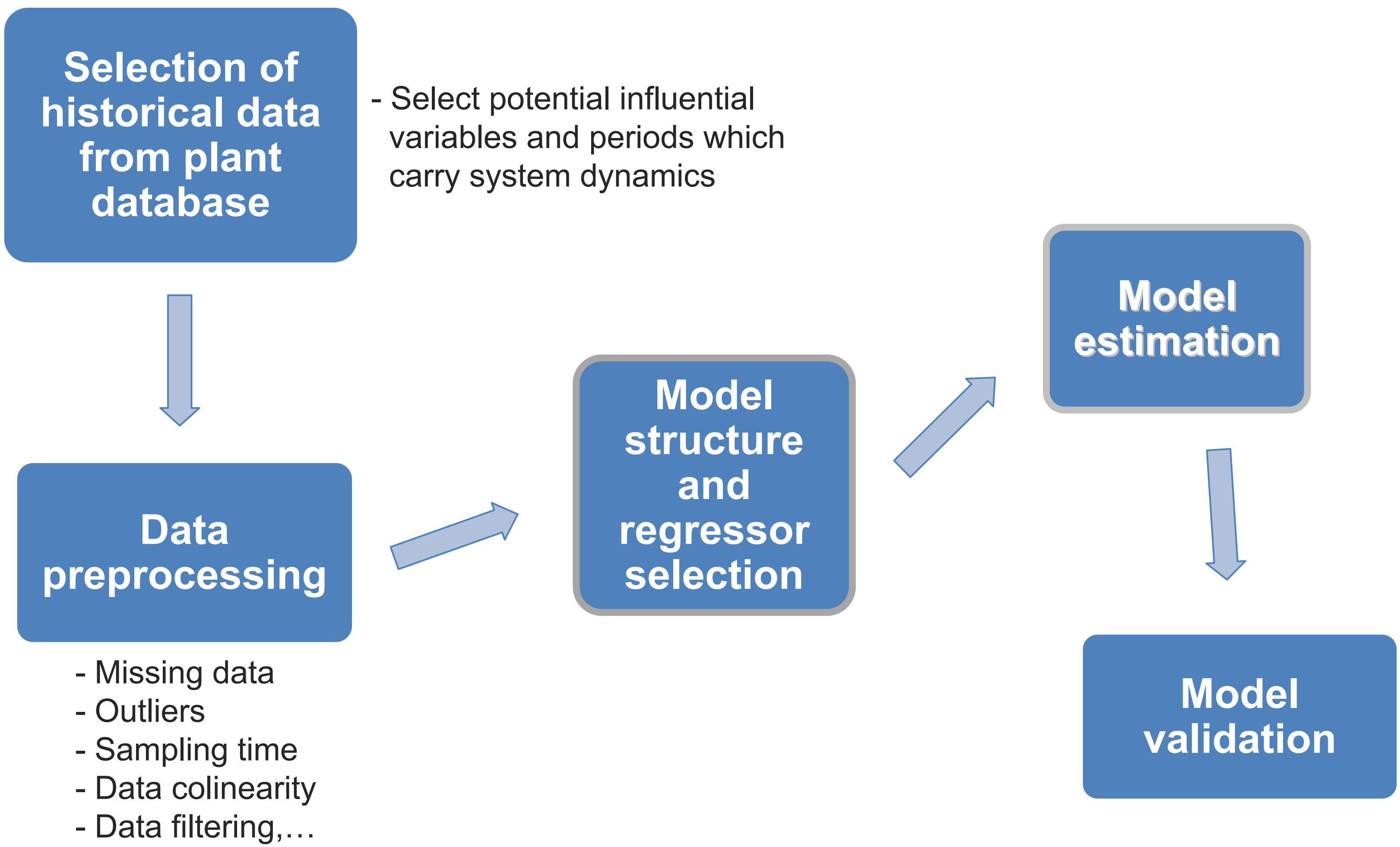
Željka Ujević Andrijić, dipl.chem.ing.

Ph.D. **Nenad Bolf**, assoc. prof.

CONTENTS

Identification procedure

The soft sensor design based on **data-driven** approaches follows this block scheme.



METHODOLOGY FOR SOFT SENSOR DEVELOPMENT

Different model structures can be used to model real systems. In industrial applications, attention is focused on parametric structures i.e. **autoregressive models** in the linear (FIR, ARX, ARMAX) and nonlinear versions (NFIR, NARX,NARMAX).

Problem

During development of dynamic models, set of model parameters need to be predetermined and adjusted in order to obtain the best model performance.

To overcome the selection of the best model order, input delays and other parameters, in an ad hoc manner, genetic algorithms and pattern search are used for determining the best set of parameters.

Applied procedure is shown on the example of FIR model and Output Error model.

DYNAMIC MODELS IN SYSTEM IDENTIFICATION

Regressors	Linear models	Nonlinear models
u(t-i)	FIR	NFIR
u(t-i), y(t-i)	ARX	NARX, NARXnet
u(t-i), ŷ(t-i)	OE	NOE, HW
u(t-i), y(t-i),e(t-i)	ARMAX	NARMAX
u(t-i), y_s*(t-i),e(t-i)* i e_s*(t-i)*	BJ	NBJ

Regressor	Meaning
u(t-i)	past measured inputs
y(t-i)	past measured outputs
ŷ(t-i)	past model (estimated) outputs
e(t-i) = ŷ(t-i) - y(t-i)	past prediction error
$y_s(t-i)$	past simulated outputs
$e_s(t-i) = \hat{y}_s(t-i) - y(t-i)$	past simulated prediction error

FIR MODEL *(Finite Impulse Response)*

- linear regression over the past measured inputs.

$$\hat{y}(t) = \mathbf{B}(q) \cdot u(t - nk)$$

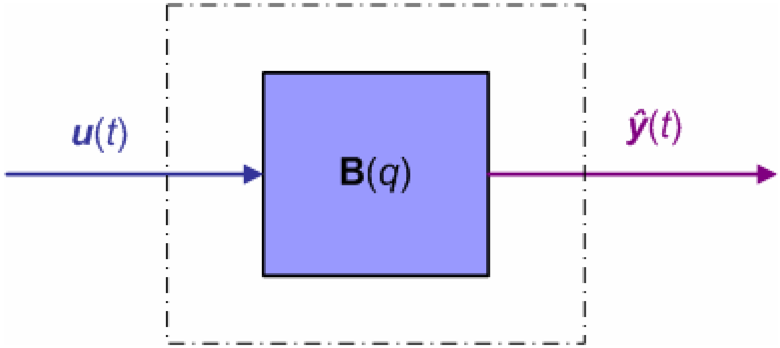
q – time operator

$q^{-1}u(t) \equiv u(t-1)$

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

u(t)	**Input at time *t***
y(t)	**Output at time *t***
nk	**Input delay**
nb	**Number of past inputs**

B is coefficient matrix: n(*y*)×n(*u*);



In case of FIR model with 5 input variables and 1 output, **10** parametars need to be estimated: 5*nk* + 5*nb*

ARX MODEL (*AutoRegressive model with eXogenous inputs*)

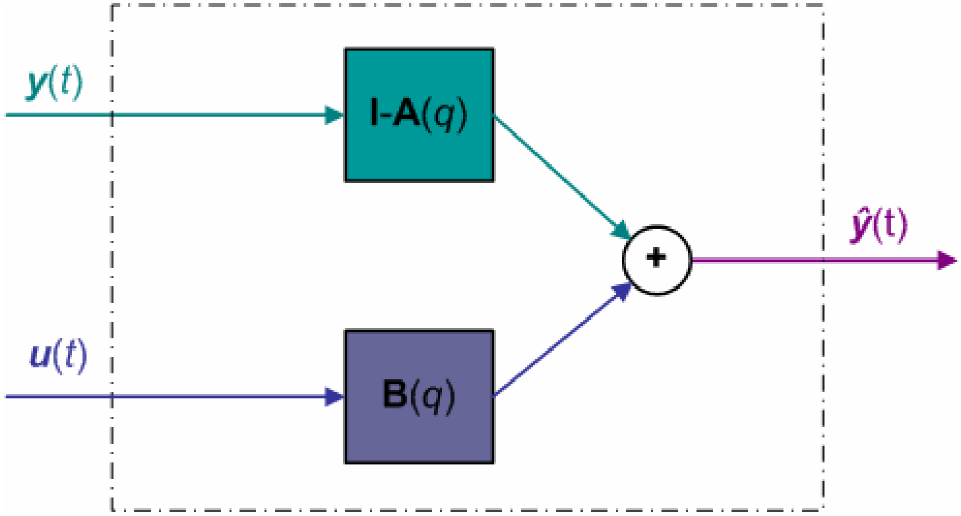
$$\hat{y}(t) = \left[\mathbf{I\text{-}A}(q)\right] y(t) + \mathbf{B}(q) u(t - nk)$$

$$\mathbf{A}(q) = \mathbf{I} + \mathbf{A}_1 q^{-1} + \ \dots \ + \mathbf{A}_{na} q^{-na}$$

A is matrix dimension: n(y)×n(y)

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

na	**Number of past mesured outputs**
nk	Input delay
nb	Number of past inputs



In case of ARX model with 5 input variables and one output, **11** parametars need to be estimated: 5*nk* + 5*nb* + 1*na*

ARMAX MODEL (*AutoRegressive Moving Average with eXogenous inputs)*

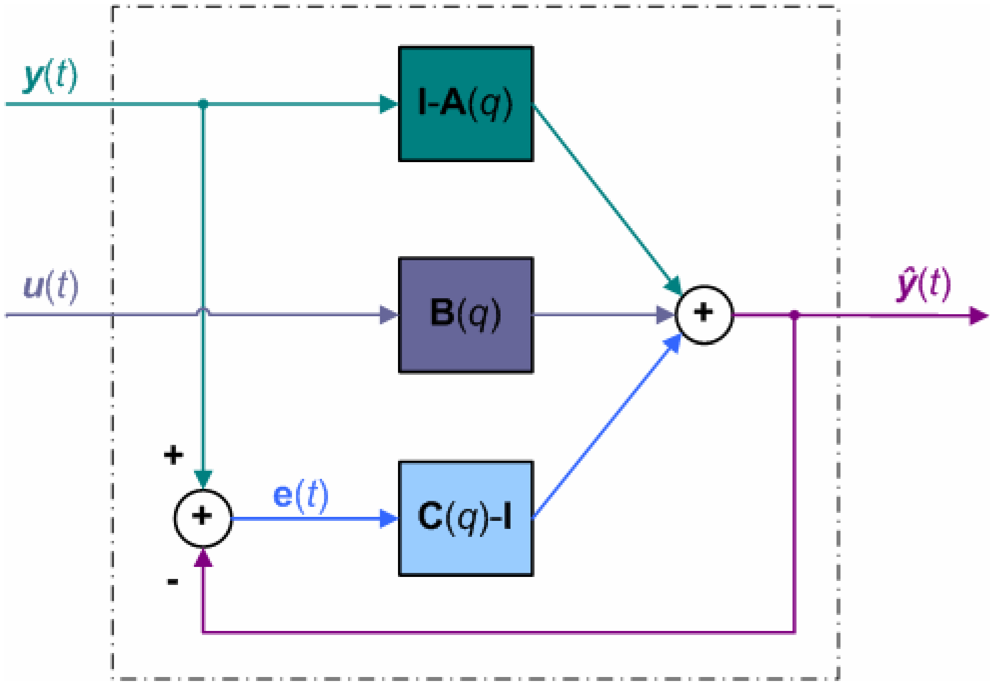
$$\hat{y}(t) = [\mathbf{I} - \mathbf{A}(q)]\, y(t) + \mathbf{B}(q) u(t-nk) + [\mathbf{C}(q) - \mathbf{I}]\, e(t)$$

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

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

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

C matrix : n(y)×n(e);



na	Number of past mesured outputs
nk	Input delay
nb	Number of past
nc	**Number of past prediction error.**

In case of ARMAX model with 5 inputs and one output, **12** parameters need to be estimated: 5*nk* + 5*nb* + 1*na*+1*nc*

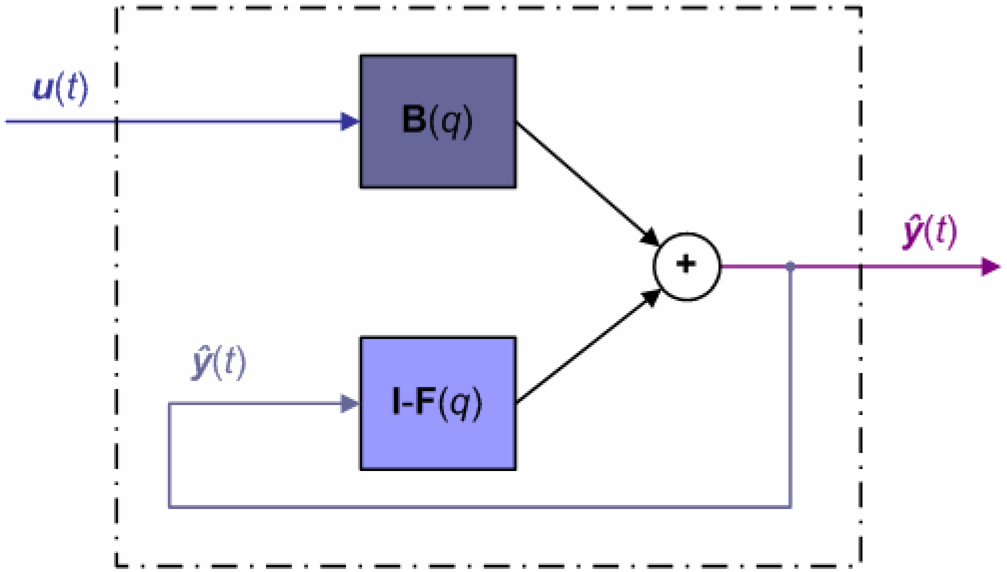
OE MODEL (*Output Error*)

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

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

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

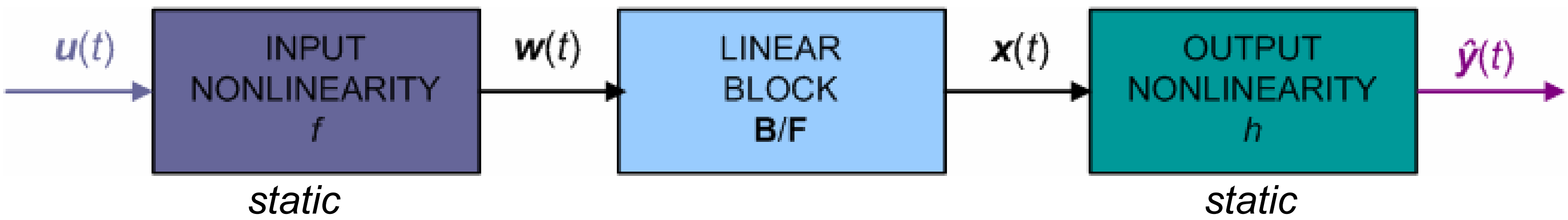
F matrix : n(ŷ)×n(ŷ)



nk	Input delay
nb	Number of past inputs
nf	**Number of past predicted output**

In case of OE model with 5 inputs and 1 output, **15** parameters need to be estimated: 5*nk* + 5*nb* + 5*nf*

HW model (Hammerstein-Wiener)



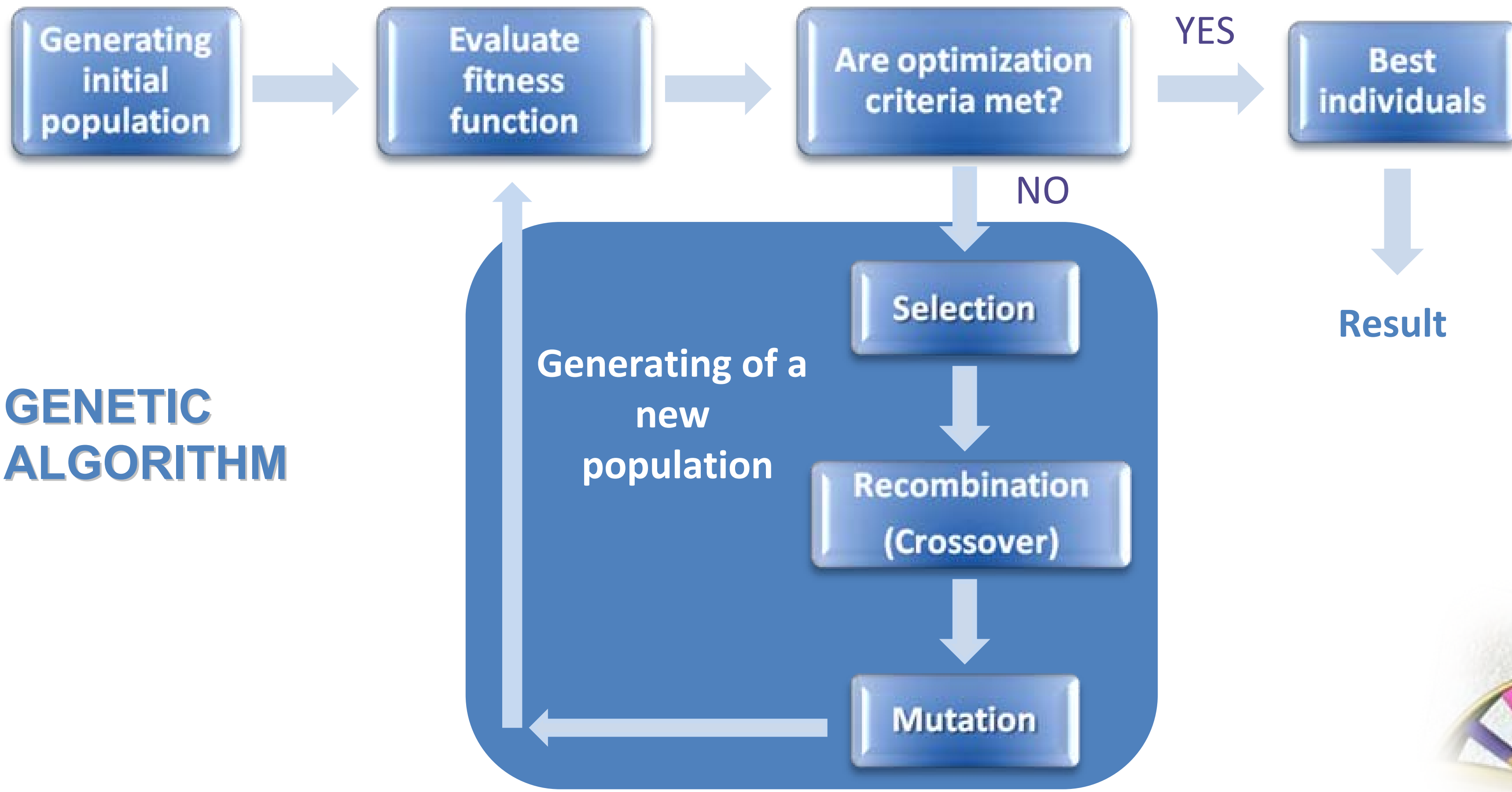
- ***w***(*t*) = ***f***(***u***(*t*)) – nonlinear function transforming input data *u*(t).
- ***x***(*t*) = (**B/F**)***w***(*t*) – linear transfer function. **B** i **F** are polynomials of the linear **OE** model.
- ***y***(t) = ***h***(***x***(*t*)) – nonlinear function that maps the output of the linear block to the system output.

Nonlinearity can be configured as a summed series of nonlinear units **(*n*)**, such as tree-partition networks, wavelet networks, multi-layer neural network, piecewise linear or sigmoid functions.

In case of HW model with 5 input variables and one output, **21** parametar need to be estimated: 5***nk*** + 5***nb*** + 5***nf*** + 6***n***

- Next step is finding the optimal model order by GA i DS method.

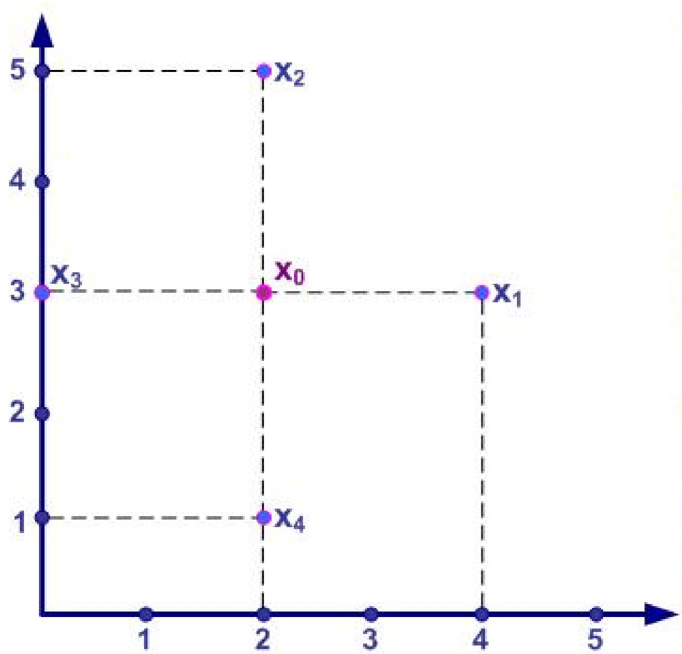
Genetic algorithm and pattern (direct) search are optimizing techniques that are used for finding global minimum of the objective (fitness) function.



GENETIC ALGORITHM

DIRECT (Pattern) SEARCH

- Direct search algorithm searches a set of points, called a mesh, around the current point—the point computed at the previous step of the algorithm.
- The mesh is formed by adding the current point to a scalar multiple of a set of vectors called a pattern.
- If the pattern search algorithm finds a point that improves the objective function at the current point, the new point becomes the current point



$x_0 = [2\ \ 3]$

2N

v1 = [1 0]
v2 = [0 1]
v3 = [-1 0]
v4 = [0 -1]

$d_m = 2$

$x_1 = x_0 + d_m * v_1 = [2\ \ 3] + 2*[1\ \ 0] = [4\ \ 3]$
$x_2 = x_0 + d_m * v_2 = [2\ \ 3] + 2*[0\ \ 1] = [2\ \ 5]$
...

search → **succesfull** **dm’ = dm/2**

search → **Not succesfull** **dm’ = 2*dm**

CASE STUDY – Fractionation of reformate

Catalytic reformate is fractioned into light and heavy reformate in reformate splitter columns.

Benzene content in refomate need to be reduced to 1%, due to:

- benzene is precursor for formation of cyclohexane in isomerization process, which is undesireable component of gasoline (low octane number).
- EPA MSAT II requirements

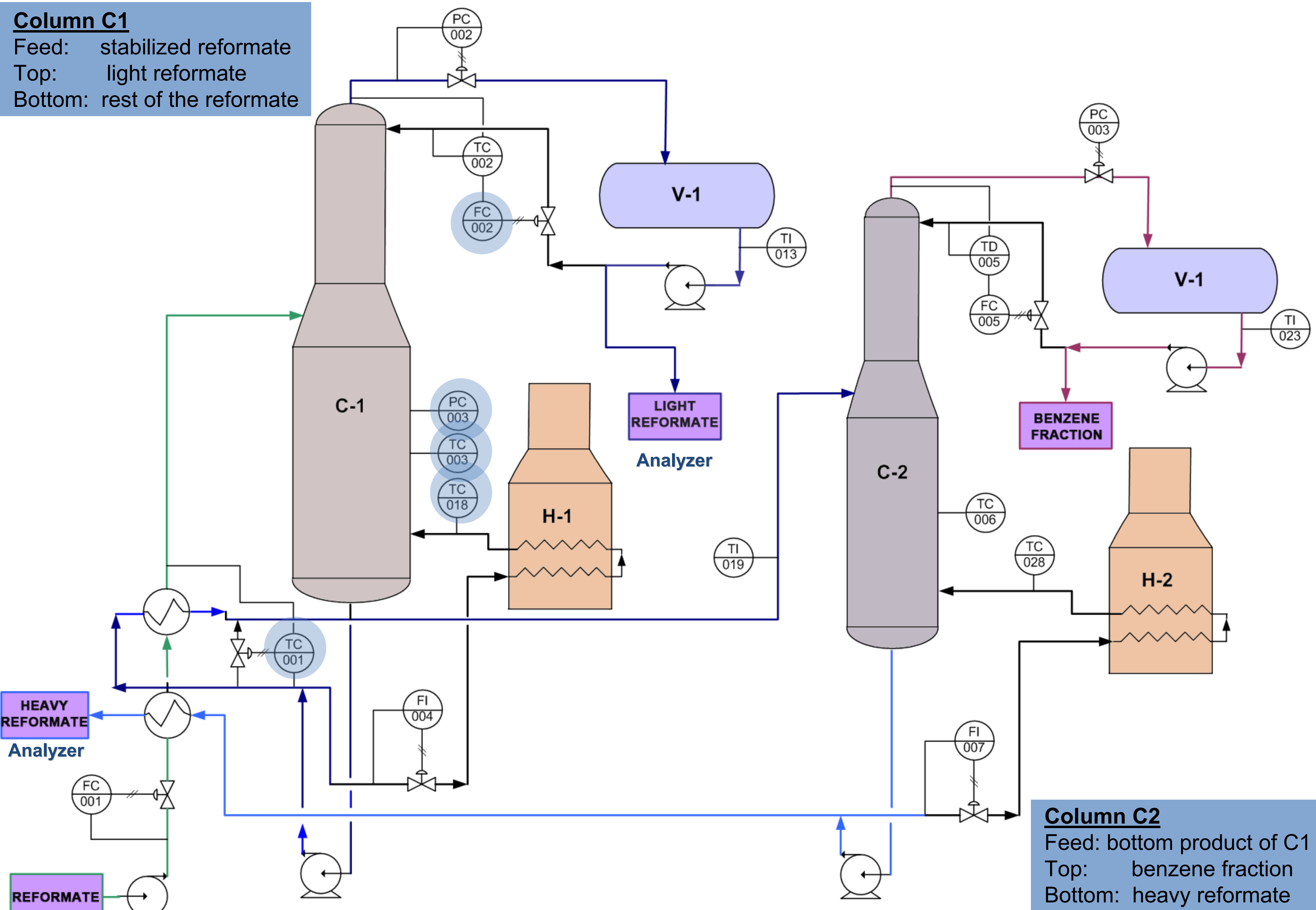
→ Continuous measurement of benzene is necessary.
Problem: benzene analyzer is often out of service or under maintenance.

Solution of that problem can be found in applicaton of:

SOFT SENSOR FOR THE ESTIMATION OF BENZENE CONTENT

FRACTIONATION REFORMATE PLANT

Column C1
Feed: stabilized reformate
Top: light reformate
Bottom: rest of the reformate



Column C2
Feed: bottom product of C1
Top: benzene fraction
Bottom: heavy reformate

DATA PREPROCESSING

Selection of historical data from plant database	
Selection of proper sampling time	• *Shannon theorem*
Outlier detection and removal	• 3-sigma method; Hampel identifier • Median
Missing value replacement	• Spline or MAR Spline
Data filtering	• Filter LOESS
Selection of input variables	• Correlation test; F test; PCA, PLS; • Interview with plant experts/operators
Detrending data	• Remove means • Remove trends
Data scaling	• Scaling in limits +-1 • Scaling such as mean= 0, st. dev. =1

MODEL DEVELOPMENT

SELECTION OF MODEL ORDER AND INPUT DELAY

(*na*,*nb*,*nc*,*nd*,*nf*,*nk*,*n*)

- **Genetic algorithm**
- **Pattern search**

simultaneously

ESTIMATION OF MODEL POLYNOMAL COEFFICIENTS

(A, B, C, D, F)

- Gauss-Newton
- Levenberg-Marquardt
- Combination of these methods

Objective function: $\frac{1}{N}\sum_{t}^{N}(y(t)-\hat{y}(t|\theta))^2$

Configurable model parameters (for 5 inputs and 1 output)

Parameter	PARAMETER MEANING	MIN	MAX
na	**No. of past output terms used to predict the current output** (ARX, ARMAX, NARX)	1	9
nb	**No. of past input** (there are 5 *nb* in this model according to 5 inputs). (for all models)	1	9
nk	**Delay from input to the output in terms of the number of samples (5 *nk* in this model according to 5 inputs)**. (for all models)	0	15
nc	**No. of past values of the prediction error** (ARMAX, BJ).	1	8
nf	**No. of past prediction output used to predict the current output.**(OE, BJ, HW) (5 *nf*)	1	5
nd	**No. of past simulated prediction output error** (ARMAX, BJ).	1	5
n	**No. of nonlinear units of network** (NARX i HW) (6 *n* for HW)	1	12

- Search parameter space for FIR model: $9^5 * 16^5 = 6,2*10^{10}$
- Search parameter space for HW model: $9^5 * 5^5 * 16^5 * 12^6 = 5,8e*10^{20}$

Models are evaluated on the basis of:

FIT, ***FPE*** (Final Prediction Error) **and** ***RMS*** (Root mean square).

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

$$FPE = V\left(1 + \frac{2d}{N}\right)$$

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

d – number of estimated parameters
N – number of values in estimation data set

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

$$|\bar{e}| = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \hat{y}_i - y_i \right|$$

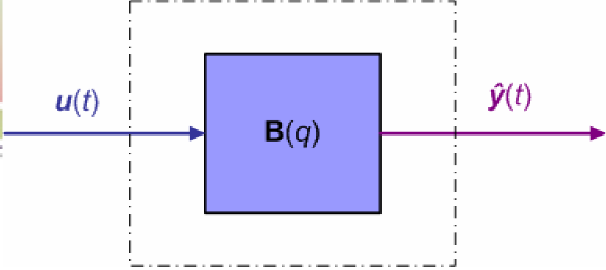
Objective function is set as the multi-objective function, because preliminary investigation showed that *FIT* and *FPE* are not always correlated.

$$\text{Obj. function} = (100 - FIT) + 1000 \cdot FPE + 100 \cdot RMS$$

-Weighted sum method

Each criterion is assigned a weighting value, and objective function is linear combination of all weighted criteria.

RESULTS - light reformate



FIR model

nb = [5 5 8 4 1];   number of past input samples
nk = [10 6 15 7 15]; input delay

fit = 78.48
RMS = 0.053   – root mean square error
E = 0.041   – mean absolute error
Emax = 0.241 – maximal error

Sampling time: 5 min

***u*(t-1)** is input value at 1st step backward;
***u*(t-2)** is input value at 2nd step backward,....

$$y(t) = B_1 \cdot u_1(t) + B_2 \cdot u_2(t) + B_3 \cdot u_3(t) + B_4 \cdot u_4(t) + B_5 \cdot u_5(t)$$

$B_1 = -0.0002163 \cdot u_1(t) + 0.005531 \cdot u_1(t-1) - 0.008785 \cdot u_1(t-2) + 0.01119 \cdot u_1(t-3) - 0.01045 \cdot u_1(t-4)$

$B_2 = 0.01259 \cdot u_2(t-6) + 0.01661 \cdot u_2(t-7) + 0.01648 \cdot u_2(t-8) + 0.005704 \cdot u_2(t-9) + 0.02052 \cdot u_2(t-10)$

$B_3 = 0.03637 \cdot u_3(t-15) - 0.02215 \cdot u_3(t-16) + 0.02651 \cdot u_3(t-17) - 0.008521 \cdot u_3(t-18) - 0.02022 \cdot u_3(t-19) + 0.06112 \cdot u_3(t-20) - 0.05342 \cdot u_3(t-21) + 0.05196 \cdot u_3(t-22)$

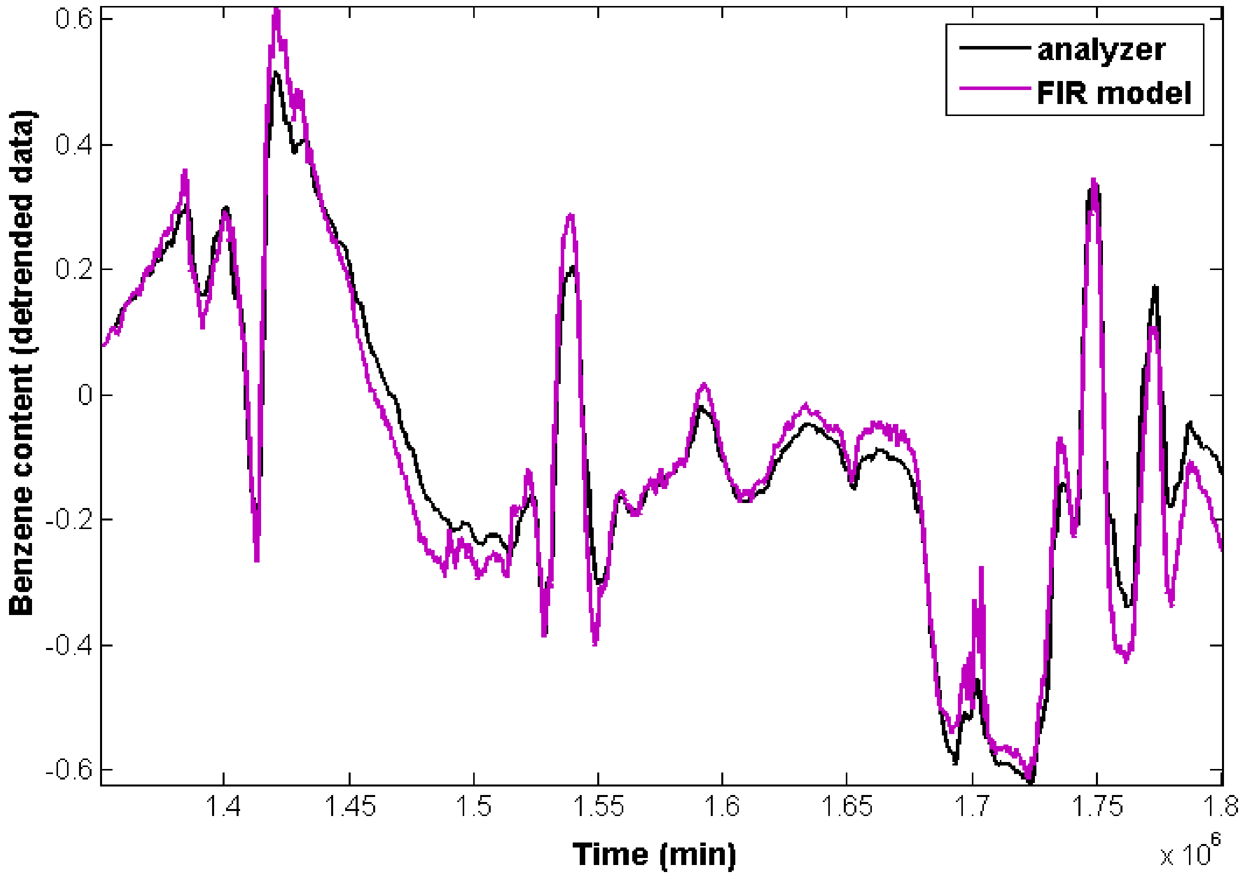
$B_4 = -1.178 \cdot u_4(t-7) - 0.5427 \cdot u_4(t-8) + 0.09258 \cdot u_4(t-9) - 2.639 \cdot u_4(t-10)$

$B_5 = -0.02445 \cdot u_5(t-15)$

RESULTS

Light reformate

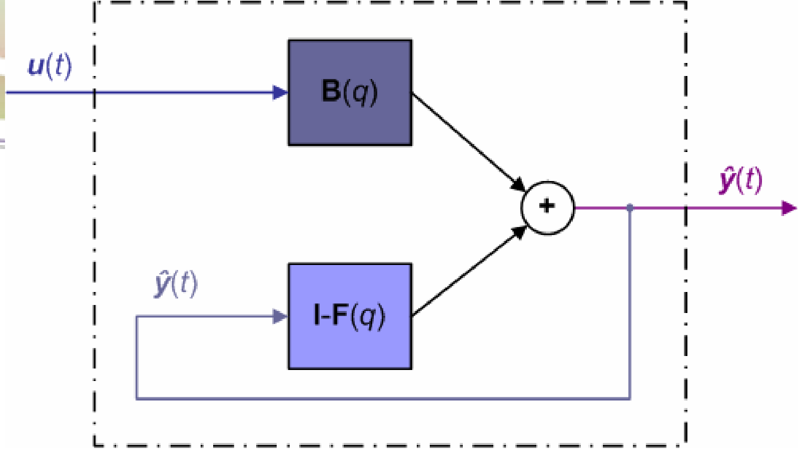
FIR model



Comparison between measured and FIR model benzene content

RESULTS - light reformate

OE model



nb = [4 4 7 6 5]; number of past input
nf = [1 3 1 1 2]; number of past predicted output
nk = [8 1 11 10 14]; input delay

fit = 89.981
RMS =0.025
E = 0.018 – mean absolute error
Emax = 0.118 – maximal error

$$\hat{y}(t) = [B_1 \cdot u_1(t) + B_2 \cdot u_2(t) + B_3 \cdot u_3(t) + B_4 \cdot u_4(t) + B_5 \cdot u_5(t)] + \\ + [(1-F_1)\cdot\hat{y}(t) + (1-F_2)\cdot\hat{y}(t) + (1-F_3)\cdot\hat{y}(t) + (1-F_4)\cdot\hat{y}(t) + (1-F_5)\cdot\hat{y}(t)]$$

$B_1 = -0.008728 \cdot u_1(t-8) + 0.012 \cdot u_1(t-9) + 0.005511 \cdot u_1(t-10) - 0.00901 \cdot u_1(t-11)$

$B_2 = -0.001627 \cdot u_2(t-1) - 0.002993 \cdot u_2(t-2) + 0.01591 \cdot u_2(t-3) - 0.01126 \cdot u_2(t-4)$

$B_3 = 0.08615 \cdot u_3(t-11) - 0.1606 \cdot u_3(t-12) + 0.1453 \cdot u_3(t-13) - 0.07753 \cdot u_3(t-14) + 0.002973 \cdot u_3(t-15) + 0.01294 \cdot u_3(t-16) - 0.008005 \cdot u_3(t-17)$

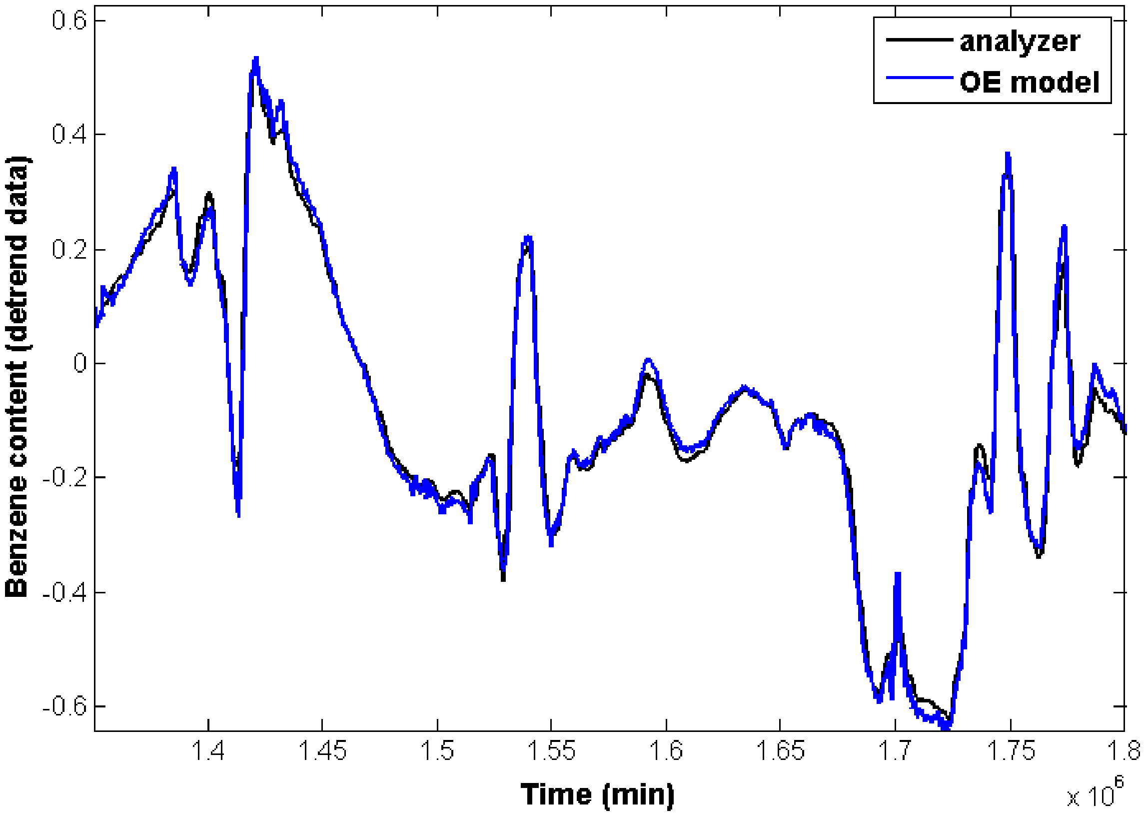
$B_4 = -1.227 \cdot u_4(t-10) + 0.6051 \cdot u_4(t-11) - 1.099 \cdot u_4(t-12) + 2.177 \cdot u_4(t-13) - 0.5003 \cdot u_4(t-14) - 0.3804 \cdot u_4(t-15)$

$B_5 = -0.01169 \cdot u_5(t-14) + 0.02872 \cdot u_5(t-15) - 0.01685 \cdot u_5(t-16) - 0.00725 \cdot u_5(t-17) + 0.007095 \cdot u_5(t-18)$

Light reformate

OE model

$F_1 = 1 - 0.9977 \cdot \hat{y}(t-1)$
$F_2 = 1 - 1.045 \cdot \hat{y}(t-1) - 0.8704 \cdot \hat{y}(t-2) + 0.9157 \cdot \hat{y}(t-3)$
$F_3 = 1 - 0.9946 \cdot \hat{y}(t-1)$
$F_4 = 1 - 0.8507 \cdot \hat{y}(t-1)$
$F_5 = 1 - 1.9 \cdot \hat{y}(t-1) + 0.9005 \cdot \hat{y}(t-2)$



Comparison between measured and OE model benzene content

SUMMARY

Soft sensors for the on-line estimation of benzene content in light and heavy reformate were developed.

Procedure of development soft sensor is shown on the example of FIR and OE model.

It was shown that genetic algorithms can be applied for optimizing configurable parameters of autoregressive models.

Developed models showed satisfactory matching with experimental data, thus improved that can be employed as the soft sensors for the on-line estimation.