

# Support Vector Machine-based Soft Sensors in the Isomerisation Process<sup>+</sup>



This work is licensed under a Creative Commons Attribution 4.0 International License

S. Herceg,\* Ž. Ujević Andrijić, and N. Bolf

Faculty of Chemical Engineering and Technology,  
University of Zagreb, Department of  
Measurements and Process Control,  
Savska c. 16/5a, 10 000 Zagreb, Croatia

<https://doi.org/10.15255/CABEQ.2020.1825>

Original scientific paper  
Received: May 20, 2020  
Accepted: November 29, 2020

This paper presents the development of soft sensor empirical models using support vector machine (SVM) for the continual assessment of 2,3-dimethylbutane and 2-methylpentane mole percentage as important product quality indicators in the refinery isomerisation process. During the model development, critical steps were taken, including selection and pre-processing of the industrial process data, which are broadly discussed in this paper. The SVM model results were compared with dynamic linear output error model and nonlinear Hammerstein-Wiener model. Evaluation of the developed models on independent data sets showed their reliability in the assessment of the component contents. The soft sensors are to be embedded into the process control system, and serve primarily as a replacement during the process analysers' failure and service periods.

## Key words:

support vector machine, soft sensor, isomerisation process, process analyser

## Introduction

Process analysers, used for measurement of key process variables, are often weak links in refinery plants. Their long analysis time, tendency of failure, and high price usually make them impractical and unprofitable. Soft sensors that enable real-time prediction of key product properties occur as an alternative to process analysers.

Rarely used as first principle modelling and more often as data-driven mathematical models, soft sensors can well describe dynamics of complex industrial processes<sup>1</sup>.

This paper presents data-driven soft sensors which have common steps in the development procedure: selection of real process data from plant history database, data pre-processing, determination of a model structure and regressors, model estimation and validation<sup>2</sup>.

The support vector machine is a popular method for soft sensor model development presented by Vapnik<sup>3</sup> as part of a general learning theory. The method has attractive features, such as the ability to learn well with only a very small number of free

parameters, robustness, and computational efficiency compared to several other methods<sup>4</sup>.

The method is widely used for nonlinear system identification required in the process industry.

The application of SVM has been described in many published papers over the past few years. Meng *et al.*<sup>5</sup> developed data-driven soft sensor based on twin support vector regression for cane sugar crystallisation. Ibrahim *et al.*<sup>6</sup> used SVM and surrogate column models for a novel optimisation-based design of crude oil distillation units. Lv *et al.*<sup>7</sup> proposed SVM-based model for puerarin extraction. Shokri *et al.*<sup>8</sup> developed SVM model for the prediction of the content of hydrogen sulphide in the hydrotreatment (HDT) refinery process. Support vector machine is presented in the papers by Xu *et al.*<sup>9</sup> where least squares support vector machine (LS-SVM) is used for gas flow measurements, while Cheng and Liu<sup>10</sup> used LS-SVM to propose online soft sensor for product quality monitoring in propylene polymerisation process. Some earlier works should be mentioned, such as the paper by Yan *et al.*<sup>11</sup> where SVM was introduced during soft sensor modelling for light gas oil freezing point assessment in the distillation process, as well as the paper by Li *et al.*<sup>12</sup> who developed the model for kerosene dry point assessment based on least squares support vector machine (LS-SVM).

The research and application of soft sensors on an isomerisation process are still rare. Lukec *et al.*<sup>13</sup> proposed application of a software analyser for online

<sup>+</sup>This paper was presented at the Meeting of Young Chemical Engineers 2020 at the Faculty of Chemical Engineering and Technology, University of Zagreb

\*Corresponding author: Srećko Herceg: [hsrecko@gmail.com](mailto:hsrecko@gmail.com); [sherceg@fkit.hr](mailto:sherceg@fkit.hr)

estimation of isopentane content in the deisopentaniser column top product, in the feed treatment section for the isomerisation process, while Xianghua *et al.*<sup>14</sup> presented the model for para-xylene content estimation at an isomerisation unit reactor outlet.

In our research, development of soft sensor empirical models based on SVM for the continual assessment of mole percentage of 2,3-dimethylbutane and 2-methylpentane in the product streams of the refinery isomerisation process is presented. These components directly affect the octane number of the isomerate – the product of the isomerisation process. The model development procedure and the model results are presented and discussed.

## Material and methods

In this section, SVM, along with the dynamic output error and Hammerstein-Wiener model structure are briefly explained. A particular part is dedicated to the description of the refinery isomerisation process, while soft sensor model development is explained in detail.

### SVM model structure

The basic idea of support vector machine can be expressed as shown in Fig. 1. Input space objects, separated with a complex non-linear curve, are mapped (rearranged) into a so-called feature space, where the objects are linearly separable, i.e., an optimal separable curve can be found<sup>15</sup>.

Mathematically it can be represented by Eq. (1), where the feature space linear regression function is a solution to the nonlinear regression problem:

$$f(\mathbf{x}) = (\mathbf{w} \cdot \Phi(\mathbf{x})) + b \quad (1)$$

where  $\mathbf{x}$  is the input vector,  $\mathbf{w}$  is the load vector,  $b$  is a non-variable value,  $\Phi(\mathbf{x})$  is a “feature” function, and  $(\mathbf{w} \cdot \Phi(\mathbf{x}))$  is the scalar product in the “feature” space. In order to obtain a model, the optimisation problem of so-called structural risk minimisation

principle should be solved. Employing the commonly used  $\varepsilon$ -intensive cost function and inserting an adjusting constant  $C$ , problem, which is being optimised, from Eq. (1) we obtain:

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (2)$$

with the following constraints:

$$\begin{aligned} (\mathbf{w} \cdot \mathbf{x}_i + b) - y_i &\leq \varepsilon + \xi_i \\ y_i - (\mathbf{w} \cdot \mathbf{x}_i + b) &\leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0; \quad i = 1, 2, \dots, n \end{aligned} \quad (3)$$

where the adjusting constant,  $C$ , is a “penalty” factor of the model complexity  $\|\mathbf{w}\|^2$ , while  $\varepsilon$  is the parameter of the  $\varepsilon$ -intensive cost function and represents the tube radius located around the function  $f(\mathbf{x})$  (Fig. 2)<sup>15</sup>.

Deviation outside the  $[\varepsilon, -\varepsilon]$  region denotes the forecast error, represented by formulation  $C \sum_{i=1}^n (\xi_i + \xi_i^*)$ , using the slack variables  $\xi$  and  $\xi^*$ .

The points on the surface and outside the  $\varepsilon$ -tube are called support vectors (SV). The percentage of SVs affects the model accuracy – as the percentage of SVs decreases, a more flattened model is obtained and vice versa<sup>15</sup>.

The solution of the optimization problem expressed in Eq. (2) is presented by the equation:

$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha - \alpha^*) K(\mathbf{x}, \mathbf{x}_i) + b \quad (4)$$

where  $K(\mathbf{x}, \mathbf{x}_i)$  is a kernel function, and  $\alpha, \alpha^*$  are Lagrange multipliers. Radial basis function (RBF) is the most used kernel function, and is defined as:

$$K(\mathbf{x}, \mathbf{x}_i) = \exp(-\gamma \|\mathbf{x} - \mathbf{x}_i\|^2) \quad (5)$$

where  $\gamma$  is the free parameter of RBF. Kernel function “avoids” cumbersome mathematical operations that take up a lot of computational time<sup>15</sup>.

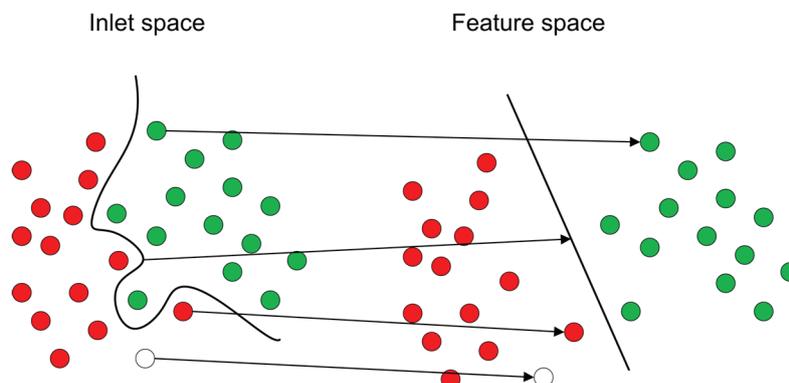
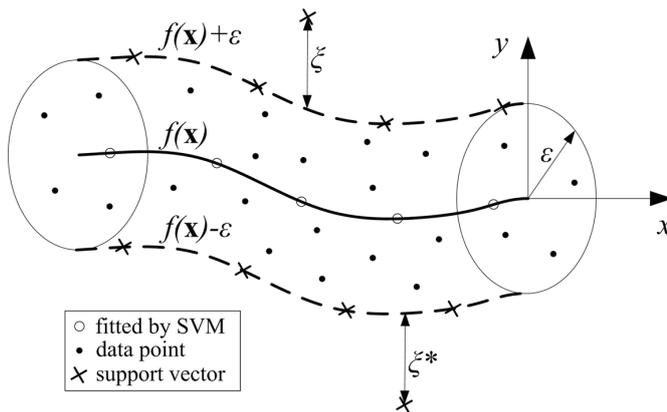


Fig. 1 – Basic idea of SVM

Fig. 2 – Graphical representation of SVM<sup>15</sup>

### Dynamic linear OE and nonlinear HW model structure

Dynamic linear and nonlinear models are concisely explained by Ljung<sup>16</sup>.

OE model is the most complex linear dynamic model usually used for soft sensor model development. The model predictor is

$$\hat{y}(t) = [1 - \mathbf{F}(q)]\hat{y}(t) + \mathbf{B}(q)u(t - nk) \quad (6)$$

where  $\mathbf{B}(q) = \mathbf{B}_1 + \mathbf{B}_2q^{-1} + \dots + \mathbf{B}_{nb}q^{-nb+1}$  is polynomial matrix by  $q^{-1}$  dimensions  $n(y) \times n(u)$ ,  $nb$  is the number of past process inputs, and  $nk$  is the input time delay expressed by the number of samples.  $\mathbf{F}(q) = 1 + \mathbf{F}_1q^{-1} + \mathbf{F}_2q^{-2} + \dots + \mathbf{F}_{nf}q^{-nf}$  is polynomial matrix by  $q^{-1}$  dimensions  $n(\hat{y}) \times n(\hat{y})$ ,  $nf$  is the number of past outputs predicted by the model.

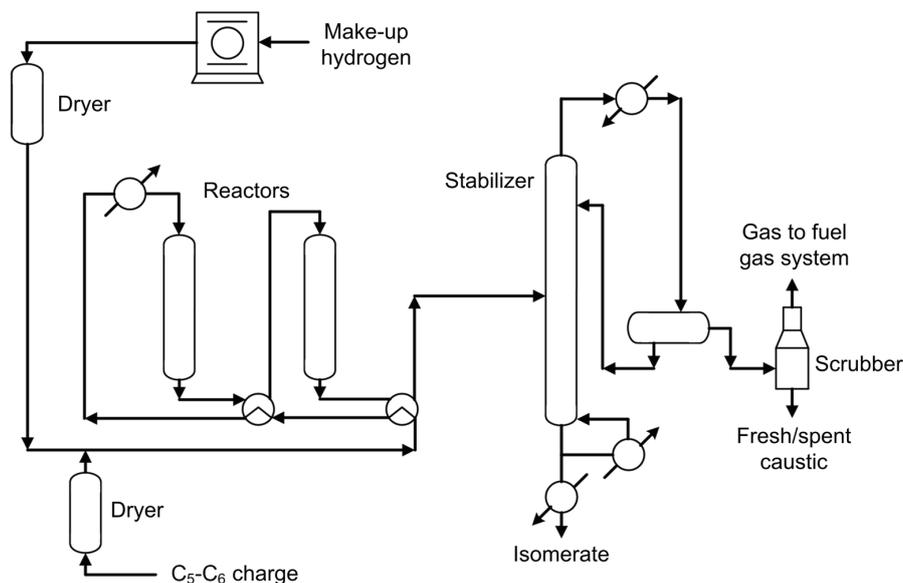
The most complex nonlinear dynamic model is HW. It has a block structure described by 3 func-

tions:  $w(t) = f(u(t))$  is a nonlinear function transforming input data  $u(t)$ ,  $x(t) = (\mathbf{B}/\mathbf{F})w(t)$  is a linear transfer function where  $\mathbf{B}$  and  $\mathbf{F}$  are polynomials of the OE model, and  $\hat{y}(t) = h(x(t))$  is a nonlinear function mapping output data  $x(t)$  from the linear block to the model output. Nonlinear function could be represented with many nonlinear units, such as wavelet, sigmoid, piecewise-linear, and others.

### Process description

The goal of the refinery isomerisation process is to upgrade the octane of light straight-run naphtha, processing paraffin (mainly pentane and hexane) together with hydrogen on a low-temperature, noble-metal, fix-bed catalyst, which is mainly used today. In more detail, the feed paraffin is converted to high-octane iso-structures – normal pentane ( $nC_5$ ) to isopentane ( $iC_5$ ), and normal hexane ( $nC_6$ ) to 2,2 and 2,3-dimethylbutane. Process conditions improve isomerisation and reduce unfavourable reactions (e.g., hydrocracking), and are featured by medium operating pressure, low temperature, and low hydrogen partial pressure<sup>17</sup>.

According to the process flow diagram (Fig. 3), the dried feed is mixed with make-up hydrogen, and heated before entering the reactor section. After passing the reactors, the isomerised product is stabilized in the stabilizer column, where the liquid from the stabilizer bottom passes to gasoline blending, while the stabilizer overhead vapour product flows to a fuel gas system, before being caustic scrubbed with in aim of removing the HCl formed from organic chloride added to the reactor feed to maintain catalyst activity<sup>17</sup>.

Fig. 3 – Straight-through isomerisation process<sup>17</sup>

A straight-through isomerization process can be improved by separating the stabilizer bottoms into normal and isoparaffin components by adding a deisohexanizer column (DIH) (Fig. 4).

The DIH column sidecut stream concentrates non-converted n-paraffins and newly-formed low-octane methylpentanes, and returns them to the reactor section. The isomerate is then drawn from the column top, while the heptane fraction is drawn from the column bottom<sup>17</sup>.

Fig. 5 depicts the process flow diagram of the observed plant deisohexanizer section with the deisohexanizer column as its part. All process and measuring equipment, as well as control loops, are shown.

### Soft sensor model development

2,3-DMB and 2-MP come out as key components of isomerate – the product of the refinery isomerisation process improved by adding a DIH distillation column. The components affect the octane-number of the product. High-octane 2,3-DMB and low-octane 2-MP mole fractions are measured on-line by process analysers in the DIH sidecut stream and in the DIH overhead, respectively, keeping 2,3-DMB in the column top and 2-MP in the side of the column to regulate their molar percent-

age<sup>18</sup>. The components are also analysed by laboratory assays once a day.

As the process analysers quite often become unavailable due to failures and have long time delays, it was decided to develop soft sensors that would find their applications primarily as the analysers' replacement.

Soft sensor development has a common procedure, as follows:

- potentially influential variable selection,
- data collection and pre-processing,
- preliminary research,
- model structure and regressor selection, model estimation, and validation,
- model implementation.

#### Potentially influential variable selection

Based on process studies and consultations with process experts, the potentially influential variables for 2,3-DMB and 2-MP SVM soft sensor model development were selected as shown in Table 1.

#### Data collection and pre-processing

Experimental data for model development were acquired from the plant history database containing

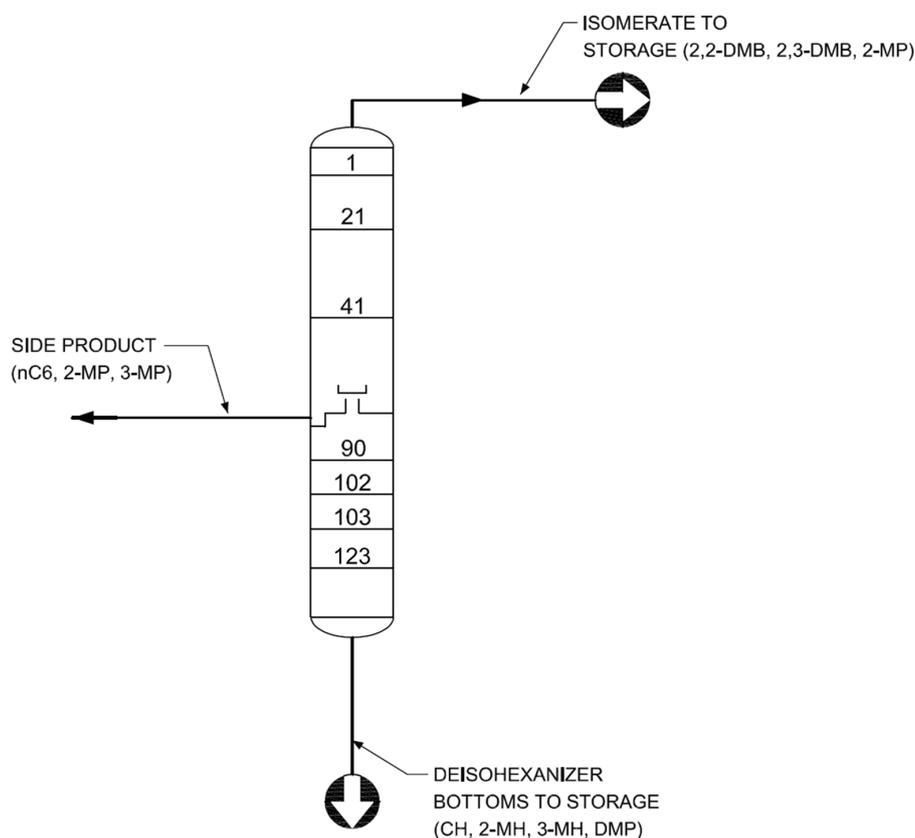


Fig. 4 – Deisohexanizer column



Table 1 – Potentially influential variables<sup>18</sup>

Variable	Tag	Unit
V17 overhead vapor temperature	TIR-046	°C
21 <sup>st</sup> V17 tray temperature	TIRC-047	°C
V17 side product temperature	TIR-045	°C
V17 bottom product temperature	TIR-049	°C
V17 reflux flow	FIR-028	m <sup>3</sup> h <sup>-1</sup>
V17 reflux flow and isomerate flow to storage sum	FIRC-029	m <sup>3</sup> h <sup>-1</sup>
V17 side product flow	FIRC-020	m <sup>3</sup> h <sup>-1</sup>
V17 bottom product flow	FIRC-026	m <sup>3</sup> h <sup>-1</sup>

up to a few years back historical data recorded every minute.

Data were collected for eight potentially influential variables (Table 1), and for 2,3-DMB and 2-MP content. Special attention was paid to selecting the data period covering significant process dynamics. After the data were collected over several periods, lasting from 2–3 weeks to about 2 years, the data were pre-processed, including sample period determination, detection of outliers, and interpolation of missing data.

According to process dynamics influential variable (further referred to as input variable), the sample period of 3 minutes was evaluated as suitable. Process analyser data (further referred to as output variable) sample period was 30 minutes. The missing data were interpolated using cubic spline interpolation.

Additional data can be generated using cubic spline or Multivariate adaptive regression spline methods. MARSpline is normally used where there is more than one input variable, as well as when the output variable sample time is long<sup>19</sup>. Otherwise, the data interpolated using cubic spline have no physical bases, i.e., a relationship between input and output variables is not taken into account. However, due to the fact that refinery processes are quite inertial, this method can be considered sufficiently reliable<sup>20</sup>.

The common methods for detection of outliers are 3-sigma<sup>21</sup>, as well as principal component analysis (PCA) and partial least squares (PLS) methods<sup>22</sup>, which perform only statistical inspection of data and tend to remove peak values that can contain useful information about process dynamics. Therefore, the procedure of outlier detection cannot be a fully automated process, and data always should be checked visually. All data collected were visually checked and the amount of outliers detected was negligible.

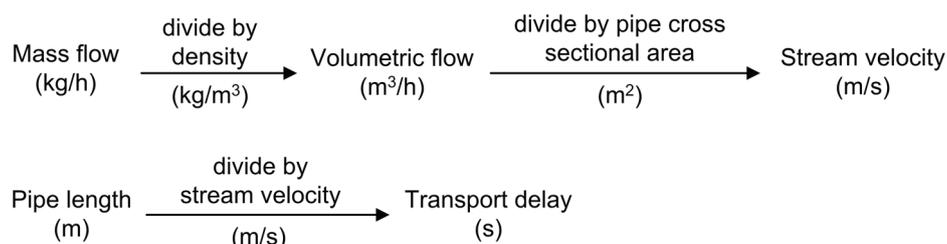
### Preliminary research

Since SVM models are static, it is very important to determine the output variable time delays regarding the change in the value of each input variable.

The output variable time delays were determined for 2,3-DMB and 2-MP content, respectively. For this purpose, an experiment on the observed plant was performed.

In the case of 2,3-DMB, it was observed how a change in one variable caused changes in others. As may be seen in Fig. 5, the V17 reflux flow rate (FIR-028) was reduced slightly, and measured were the time periods until the change in the V17 overhead vapor temperature (TIR-046), in the 21<sup>st</sup> V17 tray temperature (TIRC-047), and finally in the value of 2,3-DMB content in the V17 side product (determined by AIR-004B chromatograph). A relatively quick response was noticed. After 2 minutes, the top column temperature (TIR-046) and the temperature on the 21<sup>st</sup> tray (TIRC-047) started to rise. After 4 minutes, 2,3-DMB content reacted in the side product – it increased slightly. At once, 2,3-DMB content value time delay in regards to the change in FIR-028 and FIRC-029 input variable, respectively, was 4 minutes. Consequently, it was concluded that, for TIR-046 and TIRC-047 input variable, respectively, time delay was 2 minutes. Since TIR-045 input variable was installed upstream of the AIR-004B chromatographic analyser, and between them there were only the P10A/B pump and the EA03 air cooler, the time delay was short (0 – 2 minutes). Therefore, time delays were determined exactly for the following variables: TIR-046, TIRC-047, TIR-045, FIR-028 and FIRC-029.

In the case of 2-MP time delays, 2-MP content (determined by the AIR-005B chromatograph) had not changed during the experiment. In order not to affect the process regime, the experiment could not be carried out further and had to be stopped. Therefore, the output variable time delays in this case were determined by a calculation procedure. The main issue was to determine the time until the composition in the V12 separator had changed. Due to the relatively large separator volume (about 107 m<sup>3</sup>), change in composition took considerable time. Based on simple hydraulic calculation schematically represented in Fig. 6<sup>23</sup>, using data on the mass flow rate of the V17 top product vapours, density, pipe diameters and lengths between the V17 column, the V12 separator and the AIR-005B chromatograph, as well as the V12 volume, the time until the composition in the V12 separator had changed was calculated. Taking into account the data obtained by the experiment, 2-MP content value time delays were determined for those variables

Fig. 6 – Transport delay calculating procedure<sup>23</sup>

as in the case of 2,3-DMB content: TIR-046, TIRC-047, TIR-045, FIR-028, and FIRC-029.

For TIR-049, FIRC-020, and FIRC-026 variables regarding both 2,3-DMB and 2-MP content, the calculation procedure was the same. However, due to the complex construction of the V17 column, the obtained results could not be considered accurate but only rough.

The output variable time delay results are displayed and additionally discussed in the next section “Results and discussion”.

After delays had been determined, the correlation analysis was performed on the selected potentially input variables, obtaining Pearson linear correlations<sup>24</sup> among the input and output variables. The analysis was performed on two independent data periods (during 2015 and 2016, respectively). The results are displayed and discussed in the next section.

Among several data periods researched, periods were found when the observed plant was in reduced product regime. However, since such periods are still rare, these data were not considered as potential for the software sensor development.

Performed were descriptive statistics of the input and output variables for the selected data periods, as presented in the “Results and discussion” section.

Table 2 – 2,3-DMB content value time delays in regards to the change in the value of each input variable

Variable tag	Delay, min	Number of steps equal to a delay (one step is 3 min)
TIR-046	2	1
TIRC-047	2	1
TIR-045	0 – 2	1
TIR-049	5	2
FIR-028	4	2
FIRC-029	4	2
FIRC-020	2 – 4	1
FIRC-026	8 – 10	3

Model structure and regressor selection, model estimation and validation

SVM models were developed using TIBCO Statistica software. Selected data for model development were divided randomly: 75 % for training data, and 25 % for model validation. Randomly divided data enabled the selection of data of greater diversity for training, and consequently, better model results.

Obtained output variable time delays were incorporated into the software in such a way that each “excel like” data column, representing the input variable, was moved backwards in regards to the output variable data column by the number of steps equal to a time delay (one step is 3 minutes). The delay steps are presented in Tables 2 and 3.

Free model coefficients,  $C$  and  $\varepsilon$ , were optimized by a grid search algorithm defining the search interval for each coefficient. The coefficient of radial basis kernel function,  $\gamma$ , was initially set at the default value of 0.167, and then adjusted by trial-and-error. The procedure was repeated with the number of iterations defined in the range from 1 000 to 1 000 000. The algorithm calculation was stopped, and the SVM was considered to have been trained sufficiently when the training error reached the value of 0.001 % mole<sup>15</sup>.

Table 3 – 2-MP content value time delays in regards to the change in the value of each input variable

Variable tag	Delay, min	Number of steps equal to a delay (one step is 3 min)
TIR-046	8 – 9	3
TIRC-047	8 – 9	3
TIR-045	8 – 11	3
TIR-049	13 – 16	5
FIR-028	10 – 11	4
FIRC-029	10 – 11	4
FIRC-020	10 – 13	4
FIRC-026	16 – 19	6

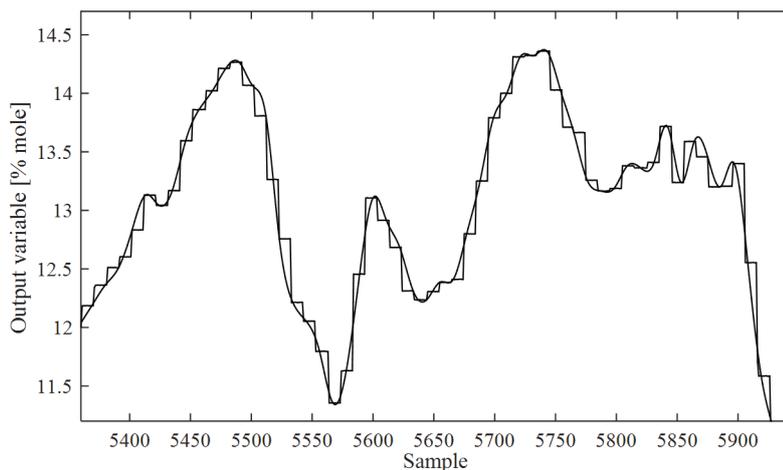


Fig. 7 – Cubic spline interpolated output data and the real analyser output

Developed models were validated based on FIT values, final prediction error (FPE), root mean square error (RMSE), and mean absolute error (MAE).

#### Model implementation

Implementation of the model to the refinery isomerisation process is underway. The goal is to implement the developed models into the module for advanced process control in plant DCS. The new variable labels for 2,3-DMB and 2-MP content predicted by the model, will be created and stored in the process history database.

## Results and discussion

In this section, the results obtained from data collection and pre-processing, preliminary research – including calculation of input variables time delays, performance of the correlation analysis and descriptive statistics, as well as the developed 2,3-DMB and 2-MP content SVM model results compared with dynamic linear OE and nonlinear HW model, are presented and discussed.

#### Interpolation of missing data

Fig. 7 shows a part of cubic spline interpolated output data compared with the real measurements (stepped curve). Very good interpolation of missing data can be observed.

#### Determining time delays

From Table 2 it can be seen that the time delays are relatively small. AIRC-004A chromatographic analyser on V17 sidecut product line was installed with no upstream accumulated liquid vessel, e.g., a separator, etc. (Fig. 5), that can consequently slow down a mass transfer.

In the case of 2-MP content, the time delays are larger than in the case of 2,3-DMB. This is mainly due to the aforementioned V12 separator, i.e., AIR-005B chromatograph installed downstream of it.

#### Correlation analysis

Tables 4 and 5 show Pearson's linear correlation coefficients between the input and output variables within 2015 and 2016 data period, respectively.

Observing the tables, it can be concluded that most of the inputs had significant impact on the outputs. However, from Table 4 it can be seen that, for all four outputs, the potential input variable FIRC-020 (V17 side product flow) had low correlations, and could be excluded as the model input. Also, for data according to Table 5, the potential inputs FIRC-020 and FIRC-026 (V17 bottom product flow) could be excluded.

The time delays were not taken into account.

#### Final input variables and model determination periods

Earlier research<sup>18</sup> on dynamic polynomial models has shown that much better results had been obtained for 2,3-DMB content models during 2015 data period, while in the case of 2-MP, the better results were during 2016. Due to the direct comparison of SVM model results with the results of the dynamic polynomial, the development of the proposed SVM models was based on the same corresponding periods. In the case of 2,3-DMB content SVM model, the range was from November 27 to December 11, 2015, comprising 6 667 measured data, while in the case of 2-MP model, the range was from January 1 to January 21, 2016, comprising 10 078 measured data, for each of the input variables and the outputs.

Table 4 – Correlation coefficient results within 2015 data period

	TIR 046	TIRC 047	TIR 045	TIR 049	FIR 028	FIRC 029	FIRC 020	FIRC 026	AIR 004B	AIR 005B
TIR-046	1.00	0.90	0.71	−0.50	−0.88	−0.33	−0.10	0.36	−0.88	0.93
TIRC-047		1.00	0.65	−0.50	−0.81	−0.23	−0.17	0.26	−0.87	0.84
TIR-045			1.00	−0.39	−0.63	−0.22	−0.20	0.27	−0.62	0.57
TIR-049				1.00	0.38	0.17	0.22	−0.19	0.30	−0.36
FIR-028					1.00	0.43	0.05	−0.24	0.83	−0.83
FIRC-029						1.00	−0.09	−0.18	0.36	−0.32
FIRC-020							1.00	−0.06	0.10	−0.04
FIRC-026								1.00	−0.35	0.42
AIR-004B									1.00	−0.92
AIR-005B										1.00

Table 5 – Correlation coefficient results within 2016 data period

	TIR 046	TIRC 047	TIR 045	TIR 049	FIR 028	FIRC 029	FIRC 020	FIRC 026	AIR 004B	AIR 005B
TIR-046	1.00	0.94	0.72	−0.59	−0.68	−0.31	−0.24	−0.22	−0.70	0.82
TIRC-047		1.00	0.73	−0.55	−0.74	−0.28	−0.21	−0.24	−0.75	0.85
TIR-045			1.00	−0.20	−0.33	−0.15	−0.10	−0.17	−0.43	0.45
TIR-049				1.00	0.42	0.11	0.20	0.47	0.29	−0.45
FIR-028					1.00	0.49	0.25	0.24	0.68	−0.78
FIRC-029						1.00	0.23	0.08	0.44	−0.37
FIRC-020							1.00	0.18	0.26	−0.20
FIRC-026								1.00	0.13	−0.08
AIR-004B									1.00	−0.88
AIR-005B										1.00

Table 6 – Input variables for the development of 2,3-DMB content SVM model

No.	Variable	Tag	Unit
1	V17 overhead vapor temperature	TIR-046	°C
2	21 <sup>st</sup> V17 tray temperature	TIRC-047	°C
3	V17 side product temperature	TIR-045	°C
4	V17 bottom product temperature	TIR-049	°C
5	V17 reflux flow	FIR-028	m <sup>3</sup> h <sup>−1</sup>
6	V17 reflux flow and isomerate flow to storage sum	FIRC-029	m <sup>3</sup> h <sup>−1</sup>
7	V17 bottom product flow	FIRC-026	m <sup>3</sup> h <sup>−1</sup>

Table 7 – Input variables for the development of 2-MP content SVM model

No.	Variable	Tag	Unit
1	V17 overhead vapor temperature	TIR-046	°C
2	21 <sup>st</sup> V17 tray temperature	TIRC-047	°C
3	V17 side product temperature	TIR-045	°C
4	V17 bottom product temperature	TIR-049	°C
5	V17 reflux flow	FIR-028	m <sup>3</sup> h <sup>−1</sup>
6	V17 reflux flow and isomerate flow to storage sum	FIRC-029	m <sup>3</sup> h <sup>−1</sup>

The final number of the input variables for the development of the SVM models based on the correlation coefficient results given in Tables 4 and 5 is given in Tables 6 and 7.

### Descriptive statistics

Tables 8 and 9 show the descriptive statistics of the input and output variables within 2015 and 2016 data period, respectively.

Table 8 – Descriptive statistics within 2015 data period<sup>15</sup>

Variable	Samples	Mean	Median	Min	Max	Variance	Std. dev.
TIR-046	6667	75.43	75.72	72.54	77.57	1.083	1.041
TIRC-047	6667	87.29	87.65	82.55	88.40	1.072	1.035
TIR-045	6667	97.23	97.26	96.19	98.19	0.072	0.268
TIR-049	6667	121.7	121.8	117.4	125.0	1.078	1.038
FIR-028	6667	378.2	377.4	360.8	397.5	38.22	6.182
FIRC-029	6667	426.2	426.5	404.7	448.9	29.69	5.448
FIRC-026	6667	5.534	5.497	2.966	11.00	1.826	1.351
AIR-004B	6667	7.417	7.278	5.871	10.12	0.605	0.778

Table 9 – Descriptive statistics within 2016 data period<sup>15</sup>

Variable	Samples	Mean	Median	Min	Max	Variance	Std. dev.
TIR-046	10078	75.93	75.99	73.44	77.89	0.804	0.897
TIRC-047	10078	87.67	87.76	86.26	88.65	0.175	0.418
TIR-045	10078	97.15	97.16	96.21	97.89	0.060	0.245
TIR-049	10078	122.7	122.8	119.6	125.3	1.004	1.002
FIR-028	10078	372.4	372.2	357.1	386.7	36.67	6.056
FIRC-029	10078	422.3	422.6	404.7	439.0	18.19	4.265
AIR-005B	10078	11.05	11.74	4.605	16.91	8.788	2.965

### Laboratory analyses

Figs. 8 and 9 show comparison between the on-line chromatographic analysers data (2,3-DMB, 2-MP contents) and the laboratory assays within 2015 and 2016 data period, respectively.

From the plots, a good correlation between measured and laboratory data can be observed, which proves the accuracy of the on-line chromatographic analysers, i.e., validity of the selected data periods.

### Model evaluation results

Based on statistical criteria, Tables 10 and 11 show validation of SVM as well as OE and HW models on validation data set for estimating the content of 2,3-DMB and 2-MP in the isomerisation process DIH column side and top product, respectively.

As may be seen from Table 10, 2,3-DMB content SVM model shows superior results compared to both dynamic polynomial OE and HW model, respectively, with only 3 model free coefficients ( $C$ ,  $\varepsilon$  and  $\gamma$ ).

Table 11 shows that the dynamic models are better; however, with dozens of model free coefficients.

The FIT values, as well as the values of FPE, RMSE, and MAE of the SVM models, indicate sat-

isfactory assessment meaning that the process dynamics were well described.

It is easy to conclude that the SVM models are better for implementation than dynamic polynomial linear or nonlinear models, which also contributes to the robustness of the SVM models.

Table 10 – 2,3-DMB content model evaluation

	SVM	OE <sup>18</sup>	HW <sup>18</sup>
FIT (%)	84.67	72.62	73.76
FPE	0.014	0.035	0.102
RMSE (% mole)	0.118	0.213	0.200
MAE (% mole)	0.077	0.156	0.149
Model free coefficient	3	85	156

Table 11 – 2-MP content model evaluation

	SVM	OE <sup>18</sup>	HW <sup>18</sup>
FIT (%)	82.98	88.02	88.91
FPE	0.249	0.086	0.182
RMSE (% mole)	0.499	0.270	0.260
MAE (% mole)	0.317	0.201	0.196
Model free coefficient	3	87	190

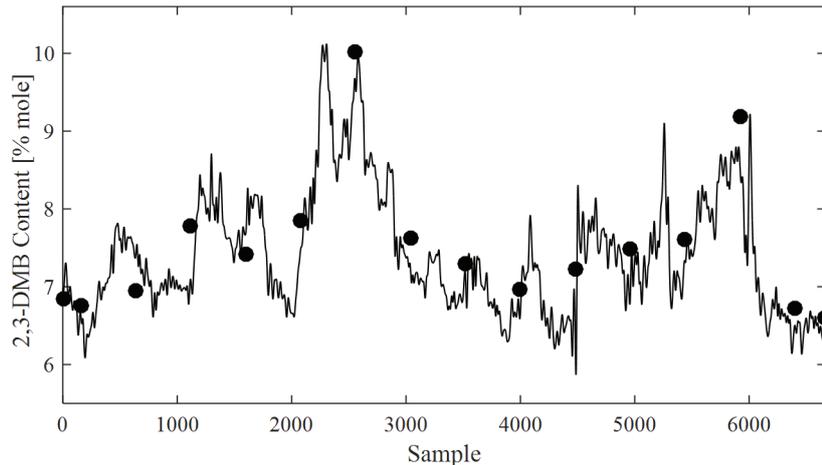


Fig. 8 – Comparison between 2,3-DMB content output variable measured by the on-line chromatographic analyser and the laboratory analysis within 2015 data period

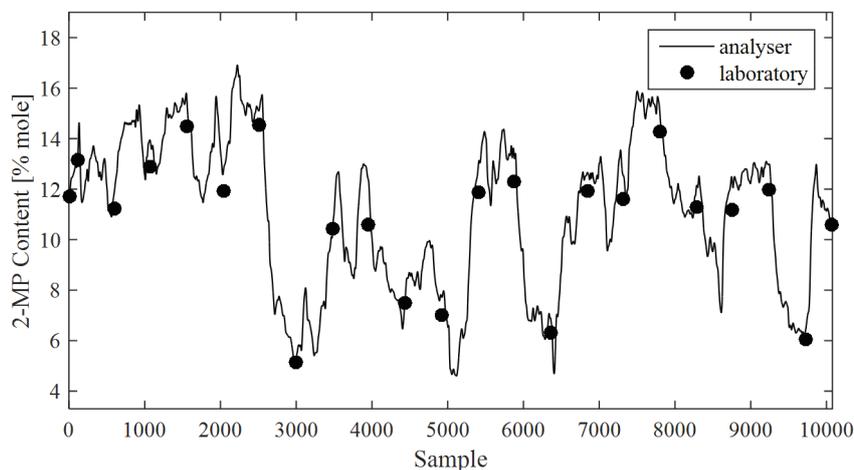


Fig. 9 – Comparison between 2-MP content output variable measured by the on-line chromatographic analyser and the laboratory analysis within 2016 data period

Graphical representation in Figs. 10 and 11 depicts the comparison between measured data and 2,3-DMB/2-MP content model data, respectively. Very good correspondence between the measured data and model outputs may be observed.

Table 12 shows the comparison between 2,3-DMB and 2-MP SVM content models. 2-MP SVM content model is somewhat more complex than 2,3-DMB – a larger number of SVs was required to achieve an accurate model.

Table 12 – Comparison between 2,3-DMB and 2-MP SVM models

Model	Optimized coefficients			Kernel function	Number of SVs
	$C$	$\varepsilon$	$\gamma$		
2,3-DMB	15	0.001	20	RBF	4748
2-MP	1	0.01	50	RBF	6042

Overall, the obtained results confirm SVM method to be suitable for nonlinear system identification in chemical plants.

## Conclusion

Soft sensor models based on SVM for continual assessment of the mole percentage of 2,3-DMB and 2-MP, as the key components in products of the refinery isomerisation process, improved by adding a deisohexanizer, were developed. The models describe the process dynamics very well, and are therefore suitable for implementation within the isomerisation process plant distributed control system (DCS). Due to its robustness, it is expected that the method will be an alternative to expensive process analysers.

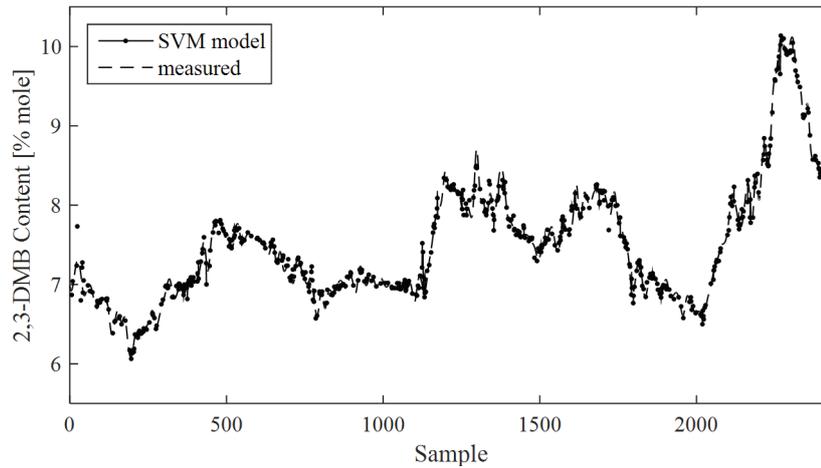


Fig. 10 – Comparison between measured data and 2,3-DMB content SVM model results

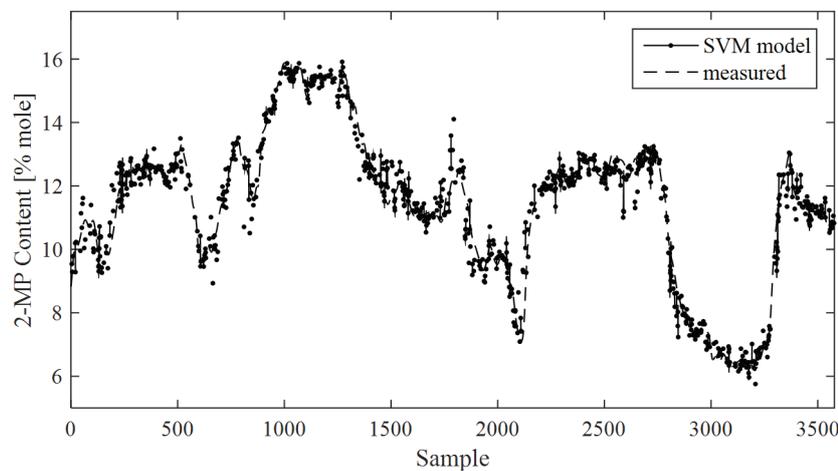


Fig. 11 – Comparison between measured data and 2-MP content SVM model results

The development and application of soft sensors are desirable and often necessary in modern process industry. The requirements for continuous improvements, especially in product quality and minimization of energy consumption, urge increased application of soft sensors.

#### ACKNOWLEDGEMENTS

The authors would like to thank the personnel of “JSC Naftan” refinery, Novopolotsk, the Republic of Belarus, for their help throughout this project.

#### List of symbols and abbreviations

2,3-DMB	– 2,3-dimethylbutane
2-MP	– 2-methylpentane
$b$	– constant
$\mathbf{B}(q), \mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_{nb}$	– polynomial matrix
$C$	– adjusting constant
DCS	– distributed control system

DIH	– deisohexanizer
$f(\mathbf{x})$	– regression function
$\mathbf{F}(q), \mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_{nf}$	– polynomial matrix
FIT	– fitting coefficient
FPE	– final prediction error
$h(x(t))$	– output nonlinearity function
HDT	– hydrotreatment
HW	– Hammerstein-Wiener
OE	– output error
$i$	– numerator
$iC_5$	– isopentane
$K(\mathbf{x}, \mathbf{x}_i)$	– kernel function
LS-SVM	– least squares support vector machine
MAE	– mean absolute error
MARSpline	– multivariate adaptive regression spline
$n$	– number of data
$nb$	– number of past process inputs
$nf$	– number of past model predicted outputs
$nk$	– input time delay expressed by the number of samples

$nC_5$	– normal pentane
$nC_6$	– normal hexane
PCA	– principle component analysis
PLS	– partial least squares
$q, q^{-1}, q^{-nb+1}, q^{-nf}$	– time shift operator
RBF	– radial basis function
RMSE	– root mean square error
SV	– support vectors
SVM	– support vector machines
$t$	– time
$u(t)$	– input data function
$w(t)$	– nonlinear function
$\mathbf{w}$	– load vector
$x(t)$	– linear transfer function
$\mathbf{x}, \mathbf{x}_i$	– vector of input data
$x, x_i$	– input data value
$y, y_i$	– output data value
$\hat{y}(t)$	– model output function
$\alpha, \alpha^*$	– Lagrange multipliers
$\gamma$	– parameter of radial basis function
$\varepsilon$	– parameter of $\varepsilon$ -intensive cost function
$\zeta, \zeta^*, \zeta_p, \zeta_i^*$	– slack variables
$\Phi(\mathbf{x})$	– feature function

## References

- Kadlec, P., Gabrys, B., Strandt, S., Data-driven soft sensors in the process industry, *Comput. Chem. Eng.* **33** (2009) 795. doi: <https://doi.org/10.1016/j.compchemeng.2008.12.012>
- Fortuna, L., Graziani, S., Rizzo, A., Xibilia, M. G., *Soft Sensors for Monitoring and Control of Industrial Processes* (Advances in Industrial Control), Springer-Verlag, London, 2007. doi: <https://doi.org/10.1007/978-1-84628-480-9>
- Vapnik, V. N., *The nature of statistical learning theory*, Springer-Verlag, New York, 1995. doi: <https://doi.org/10.1007/978-1-4757-3264-1>
- Steinwart, I., Christmann, A., *Support Vector Machines*, Springer-Verlag, New York, 2008. doi: <https://doi.org/10.1007/978-0-387-77242-4>
- Meng, Y., Lan, Q., Qin, J., Yu, S., Pang, H., Zheng, K., Data-driven soft sensor modeling based on twin support vector regression for cane sugar crystallization, *J. Food Eng.* **241** (2019) 159. doi: <https://doi.org/10.1016/j.jfoodeng.2018.07.035>
- Ibrahim, D., Jobson, M., Li, J., Guillén-Gosálbez, G., Optimization-based design of crude oil distillation units using surrogate column models and a support vector machine, *Chem. Eng. Res. Des.* **134** (2018) 212. doi: <https://doi.org/10.1016/j.cherd.2018.03.006>
- Ly, B., Chen, J., Dong, C., Wang, Q., Soft measurement of puerarin extraction based on GA-SVM, 2017 Chinese Automation Congress (CAC), IEEE, Jinan, 2017, pp 5005–5009. doi: <https://doi.org/10.1109/CAC.2017.8243667>
- Shokri, S., Sadeghi, M. T., Marvast, M. A., Narasimhan, S., Improvement of the prediction performance of a soft sensor model based on support vector regression for production of ultra-low sulfur diesel, *Pet. Sci.* **12** (2015) 177. doi: <https://doi.org/10.1007/s12182-014-0010-9>
- Xu, W., Fan, Z., Cai, M., Shi, Y., Tong, X., Sun, J., Soft sensing method of LS-SVM using temperature time series for gas flow measurements, *Metrol. Meas. Syst.* **22** (2015) 383. doi: <https://doi.org/10.1515/mms-2015-0028>
- Cheng, Z., Liu, X., Optimal online soft sensor for product quality monitoring in propylene polymerization process, *Neurocomputing* **149** (2015) 1216. doi: <https://doi.org/10.1016/j.neucom.2014.09.006>
- Yan, W., Shao, H., Wang, X., Soft sensing modelling based on support vector machine and Bayesian model selection, *Comput. Chem. Eng.* **28** (2004) 1489. doi: <https://doi.org/10.1016/j.compchemeng.2003.11.004>
- Li, Y., Li, Q., Wang, H., Ma, N., Soft Sensing Based on LS-SVM and Its Application to a Distillation Column, Sixth International Conference on Intelligent Systems Design and Applications, IEEE, Jinan, 2006, pp 177–182. doi: <https://doi.org/10.1109/ISDA.2006.246>
- Lukec, I., Lukec, D., Sertić Bionda, K., Adžamić Z., The possibilities of advancing isomerization process through continuous optimization, *Goriva i maziva* **46** (2007) 234.
- Xianghua, C., Ouguan, X., Hongbo, Z., Recursive PLS soft sensor with moving window for online PX concentration estimation in an industrial isomerization unit, Chinese Control and Decision Conference, IEEE, Guilin, 2009, pp 5853–5857. doi: <https://doi.org/10.1109/CCDC.2009.5195246>
- Herceg, S., Ujević Andrijić, Ž., Bolf, N., Development of soft sensors for isomerization process based on support vector machine regression and dynamic polynomial models, *Chem. Eng. Res. Des.* **149** (2019) 95. doi: <https://doi.org/10.1016/j.cherd.2019.06.034>
- Ljung, L., *System Identification: Theory for the User*, 2<sup>nd</sup> ed., Prentice Hall, New Jersey, 1999.
- Cusher, N. A., UOP Penex Process, in Meyers, R. A. (Ed.), *Handbook of Petroleum Refining Processes*, third edition, McGraw-Hill, New York, 2003, pp 9.15–9.27.
- Herceg, S., Ujević Andrijić, Ž., Bolf, N., Continuous estimation of the key components content in the isomerization process products, *Chem. Eng. Trans.* **69** (2018), 79. doi: <https://doi.org/10.3303/CET1869014>
- Mohler, I., Development of soft sensors for refinery advanced process control, PhD thesis, Faculty of chemical engineering and technology, University of Zagreb, Zagreb, 2015.
- Ujević Andrijić, Ž., Softverski senzori za identifikaciju i vođenje nelinearnih procesa, PhD thesis, Faculty of chemical engineering and technology, University of Zagreb, Zagreb, 2012.
- Pearson, R., K., Outliers in process modeling and identification, *IEEE T. Contr. Syst. T.* **10** (2002) 55. doi: <https://doi.org/10.1109/87.974338>
- Lin, B., Recke, B., Renaudat, P., Knudsen, J., Jørgensen, S. B., Robust statistics for soft sensor development in cement kiln, 16th Triennial World Congress of International Federation of Automatic Control, Proceeding of 16th IFAC World Congress, Prague, 2005, pp 241–246. doi: <https://doi.org/10.3182/20050703-6-CZ-1902.01720>
- Sharmin, R., Sundararaj, U., Shah, S., Griend L. V., Sun Y., Inferential sensor for estimation of polymer quality parameters: Industrial application of a PLS-based soft sensor for a LDPE plant, *Chem. Eng. Sci.* **61** (2006) 6372. doi: <https://doi.org/10.1016/j.ces.2006.05.046>
- Warne, K., Prasad, G., Rezvani, S., Maguire, L., Statistical and computational intelligence techniques for inferential model development: A comparative evaluation and a novel proposition for fusion, *Eng. Appl. Artif. Intell.* **17** (2004) 871. doi: <https://doi.org/10.1016/j.engappai.2004.08.020>