Application of Active Set Method for Soft Sensor Model Identification of Crude Oil Distillation Process

Anton Goncharov [1], Andrei Torgashov [1,2]

[1] Institute of Automation and Control Processes FEB RAS, Vladivostok, Russia,
antalg@mail.ru
[2]Far Eastern Federal University, Vladivostok, Russia,
torgashov@iacp.dvo.ru

Abstract. The problem of identification of the soft sensors is considered. Using the active set method the problem of least squares with inequality constraints on the variables has been solved. As a result of using soft sensors, obtained taking into account constraints on the model coefficients, their efficiency as compared with soft sensors obtained without constraints is shown.

Keywords: parametric identification, optimization, soft sensor, constraints, distillation column.

1 Introduction and statement of the problem

The approach to solve the soft sensor model obtaining problem, taking into account inequality constraints on the model coefficients, is described [1]. The obtained results are tested on industrial oil fractionation process for atmospheric distillation column. The soft sensor model coefficients for prediction of the key product quality are identified.

The technological plant with several measured inputs $u_1, u_2, \ldots, u_N$ and one output $y(\tau)$ is considered. The measured technological parameters (pressure, temperature, flow) are utilized as inputs. In practice, the amount of available process measured parameters to predict the quality of a product is much more than the required number of parameters. A priori knowledge of technological process allows to select necessary parameters.

The problem of the soft sensor (SS) evaluation which is best predicting quality of products of crude distillation technological process is considered.

The model of the soft sensor is obtained in the form of linear regression model for solution of the problem [2]:

$$y(\tau) = b_0 + b_1 u_1(\tau) + b_2 u_2(\tau) \ldots + b_N u_N(\tau), \tag{1}$$

In: A. Kononov et al. (eds.): DOOR 2016, Vladivostok, Russia, published at http://ceur-ws.org

where b_j – *j*-th model coefficient, $j = 0,1,...,N$, b_0 – constant term, *N* – the number of input variables, τ - irregular timepoints of output measurement: $\tau_1, \tau_2, \tau_3, ...$, $\tau_i = \tau_{i-1} + \tau_0 + \varepsilon, i \geq 2$; $\tau_1 = \tau_0 + \varepsilon$; τ_0 - constant term; ε - random component is limited by specific range.

The determination coefficient (the part of explained deviations variance of the dependent variable from its mean value):

$$R^2 = 1 - \sum_i (\bar{y}_i - y_i)^2 \Big/ \sum_i (\bar{y}_i - \bar{y}^a)^2 \tag{2}$$

and root mean squared error (RMSE):

$$RMSE = \left(\sum_{i=1}^{M} (\bar{y}_i - y_i)^2 / M \right)^{1/2}, \tag{3}$$

are used as criteria of identification on a given time interval,
where $\bar{y}_i$ - the measured value of the output variable, y_i - the value is obtained based on the SS, $\bar{y}^a$ - the mean value of the measured output variable, *M* - the number of output measurement. The model is more consistent if the closer to unity the value of the coefficient of determination R^2, or the closer to zero the value of the RMSE.

2 The proposed algorithm problem solution

Let $\mathbf{u} = [1, u_1(\tau), u_2(\tau), ..., u_N(\tau)]^T$ be a combined vector of measured input variables, $\mathbf{b} = [b_0, b_1, ..., b_N]^T$ - vector of coefficients of the same dimension, the components of which reflect the contributions of the respective input variables. Then the equation (1) takes the form:

$$y = \mathbf{u}^T \cdot \mathbf{b}.$$

We form the vector Y of dimension q from the output value *y*:

$$\mathbf{Y} = [y(\tau_1), y(\tau_2), ..., y(\tau_q)]^T$$

and the matrix U, containing the measured inputs u_j, corresponding to output value *y* from (1):

$$\mathbf{U} = \begin{bmatrix} 1 & u_1(\tau_1) & u_2(\tau_1) & ....u_{\mathrm{N}}(\tau_1) \\ 1 & u_1(\tau_2) & u_2(\tau_2) & ....u_{\mathrm{N}}(\tau_2) \\ \vdots & \vdots & \vdots & \vdots \quad \vdots \\ 1 & u_1(\tau_q) & u_2(\tau_q) & ..u_N(\tau_q) \end{bmatrix},$$

and we write the matrix equation:

$$\mathbf{Y} = \mathbf{Ub}\,.$$

We introduce error function:

$$\mathbf{E} = \overline{\mathbf{Y}} - \mathbf{Y} = \overline{\mathbf{Y}} - \mathbf{Ub}\,,$$

where $\overline{\mathbf{Y}}$ is the actual measurement of output, and minimize the objective function:

$$\boldsymbol{\Psi} = \mathbf{E}^2 = (\overline{\mathbf{Y}} - \mathbf{Ub})^2\,. \qquad (4)$$

We obtain estimates of the parameters $\mathbf{b}$ by least squares method:

$$\mathbf{b} = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\overline{\mathbf{Y}}\,. \qquad (5)$$

The multicollinearity case, which occurs when there is almost a linear relationship between inputs, is considered. In this case the matrix $\mathbf{C} = \mathbf{U}^T\mathbf{U}$ is close to singular, so it is the smallest eigenvalue $\lambda_{\min} = 0$ and the condition number is infinitely increased and causing the instability of the solution (5). If the $\lambda_{\min} = 0$ then it corresponds the strict multicollinearity [3]. In order to obtain a stable solution of the equation (5) it is necessary to reduce the condition number of the matrix C, for example, by adding thereto a diagonal matrix $\mathbf{B} = k\mathbf{I}$ (k> 0). Then the solution is found in a class of ridge parameter estimates:

$$\mathbf{b} = (\mathbf{U}^T\mathbf{U} + k\mathbf{I})^{-1}\mathbf{U}^T\overline{\mathbf{Y}}\,. \qquad (6)$$

The quality, obtained by (5-6) models, depends on the number of available output measurements. The length of training sample is often insufficient to obtain reliable results. Also, the available data contain significant measurement error of inputs and outputs, unmeasured influences. Taking into account constraints on the model coefficients b_j allows to avoid these problems. When taking into account constraints on coefficients at input, the problem of least squares with simple constraints on the variables is solved:

$$\begin{gathered}\min\left(\overline{\mathbf{Y}} - \mathbf{Ub}\right)^2 \\ \mathbf{b}^{\min} \le \mathbf{b} \le \mathbf{b}^{\max}.\end{gathered} \qquad (7)$$

The solution of the problem (7) is obtained by the active set numerical method [4]. The given constraints are reduced to the form:

$$\mathbf{Ab} \ge \hat{\mathbf{b}}\,,$$

where
$$\mathbf{A}=\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ -1 & 0 & \dots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \end{bmatrix},\ \hat{\mathbf{b}}=\begin{bmatrix} \mathbf{b}^{\min} \\ -\mathbf{b}^{\max} \end{bmatrix}.$$

Constraint $\mathbf{a}_i^T\mathbf{b} \ge \hat{b}_i$ call active in acceptable point b if $\mathbf{a}_i^T\mathbf{b} = \hat{b}_i$, and inactive if $\mathbf{a}_i^T\mathbf{b} > \hat{b}_i$, $\mathbf{a}_i^T$ - *i*-th row of $\mathbf{A}$.

The sufficient minimum conditions for simple constraints are as follows:

1. $\mathbf{b}^{\min} \le \mathbf{b}^* \le \mathbf{b}^{\max}$, $\mathbf{b}_{FR}^{\min} < \mathbf{b}_{FR}^* < \mathbf{b}_{FR}^{\max}$
2. $\mathbf{U}_{FR}^T(\bar{\mathbf{Y}} - \mathbf{U}\mathbf{b}^*) = 0$
3. $\boldsymbol{\lambda}^{\min} = \mathbf{U}_{\min}^T(\bar{\mathbf{Y}} - \mathbf{U}\mathbf{b}^*)$, $\lambda_i^{\min} > 0$, $i = 1,\dots,t^{\min}$, (8)

   $\boldsymbol{\lambda}^{\max} = \mathbf{U}_{\max}^T(\bar{\mathbf{Y}} - \mathbf{U}\mathbf{b}^*)$, $\lambda_i^{\max} < 0$, $i = 1,\dots,t^{\max}$
4. $\mathbf{U}_{FR}^T\mathbf{U}_{FR}$ is positive definite,

where $\mathbf{b}^*$ - the minimum point of the solution of problem (7); subscript FR indicates that in the vector and matrices the elements and columns with index numbers corresponding to the index numbers of **b** elements, that have not met the boundary values (7), are used; subscript *min, max*, indicates that the in matrix only the columns with index numbers corresponding to the index numbers of **b** elements, taking the appropriate minimum or maximum boundary value, are used. $t^{\min}$,, $t^{\max}$.- number of active upper and lower limits respectively; $\boldsymbol{\lambda}^{\min}$, $\boldsymbol{\lambda}^{\max}$ - vectors of Lagrange multipliers corresponding to the lower and upper active constraints.

To start the method of the active set it is necessary to determine the starting point using (6).

The minimum point $\mathbf{b}^*$ for the search algorithm for iteration *k* is:

1. Performance verification of the stop conditions. (Reaching the performance errors of conditions (8), constraints on the number of iterations).
2. Selection of a logic branch. Does it make sense to remove any constraint of the set of active constraints list. The condition of performance of a condition 3 in (8) is checked. If the condition is not satisfied for some of the vector element, constraint is excluded from the list of active constraints.
3. The calculation of the search direction $\mathbf{p}_k$. Like (4) solves the problem $\min\left(\bar{\mathbf{Y}} - \mathbf{U}\mathbf{b}_k - \mathbf{U}_{FR}\mathbf{p}_{FR}\right)^2$. Calculate the non-zero $\left(N+1-t_k\right)$ - dimensional vector

$\mathbf{p}_{FR}$ and the direction of search $\mathbf{p}_k = \left(\mathbf{A}^T\right)_{FR}\mathbf{p}_{FR}$, where t_k - the number of active constraints on *k* iteration.

4. Calculate the step length α_k. From $\begin{bmatrix} \mathbf{b}_{FR} \\ -\mathbf{b}_{FR} \end{bmatrix} + \mathbf{\Psi}\begin{bmatrix} \mathbf{p}_{FR} \\ -\mathbf{p}_{FR} \end{bmatrix} = \hat{\mathbf{b}}_{FR}$ diagonal matrix $\mathbf{\Psi}$ is calculated, $\hat{\mathbf{b}}_{FR}$ - consists of the elements $\hat{\mathbf{b}}$ which aren't active constraints, elements $\hat{\mathbf{b}}$, opposite boundary values in (7) for constraints in the active set, are excluded from $\hat{\mathbf{b}}_{FR}$. The $\bar{\alpha}_k = \min\left\{\mathbf{\Psi}_{ii}\right\}$ is an available minimum positive step from $\mathbf{b}_k$ along $\mathbf{p}_k$. The index j of minimum positive diagonal element $\mathbf{\Psi}$ is remembered. If $\bar{\alpha}_k > 1$, then $\alpha_k = 1$, otherwise $\alpha_k = \bar{\alpha}_k$.
5. Constraint is added in the list of active constraints. If $\alpha_k = \bar{\alpha}$, then j constraint $\hat{\mathbf{b}}_{FR}$ becomes active, it is necessary to add to the list.
6. Recalculation approximation. $\mathbf{b}_{k+1} = \mathbf{b}_k + \alpha_k\mathbf{p}_k$ is calculated, and return to step 1 of the algorithm is carried out.

The influence of the process dynamics on the quality of the products is taken into account by the dynamic SS. The predictive model is represented as a sum of convolutions of plant inputs and a finite impulse response (FIR) h_i (discrete analogues of the first degree Volterra kernels):

$$y(\tau) = h_0 + \sum\nolimits_{k=0}^{n_1-1} h_1(k+1)u_1(\tau-k) + \sum\nolimits_{k=0}^{n_2-1} h_2(k+1)u_2(\tau-k) + ... + \sum\nolimits_{k=0}^{n_N-1} h_N(k+1)u_N(\tau-k), \quad (9)$$

where h_0 – constant term, τ - irregular timepoints of output measurement: $\tau_1, \tau_2, \tau_3, ...$, $\tau_i = \tau_{i-1} + \tau_0 + \varepsilon$, $i \geq 2$; $\tau_1 = \tau_0 + \varepsilon$; τ_0 - constant component; ε - random component is limited by the specific range.

Let $\mathbf{u} = [1, u_1(\tau), ..., u_1(\tau - n_1 + 1), ..., u_N(\tau), ..., u_N(\tau - n_N + 1)]^T$ - the combined vector of measured input variables of dynamic SS (DSS) with dimensionality $q = 1 + \sum\nolimits_{k=1}^{N} n_k$ where n_k - is a number of values of *k*-th input, $\mathbf{h} = \left[h_0,\ h_1(1), ..., h_1(n_1), ..., h_N(1), ..., h_N(n_N)\right]^{\mathrm{T}}$ - vector FIR of the same dimension, the components of which reflect the contributions of the respective input variables of DSS. Then the equation (9) takes the form:

$$y = \mathbf{u}^T \cdot \mathbf{h}.$$

We form the vector Y of dimension *q* from the output value *y*:

$$\mathbf{Y} = [y(\tau_1), y(\tau_2), ..., y(\tau_q)]^T$$

and the matrix U, containing the measured inputs u_j, corresponding to output value *y* from (9):

$$\mathbf{U}=\begin{bmatrix}1 & u_1(\tau_1) & \dots & u_1(\tau_1-n_1+1) & \dots & u_{\mathrm{N}}(\tau_1) & \dots & u_{\mathrm{N}}(\tau_1-n_N+1)\\ 1 & u_1(\tau_2) & \dots & u_1(\tau_2-n_1+1) & \dots & u_{\mathrm{N}}(\tau_2) & \dots & u_{\mathrm{N}}(\tau_2-n_N+1)\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & u_1(\tau_q) & \dots & u_1(\tau_q-n_1+1) & \dots & u_{\mathrm{N}}(\tau_q) & \dots & u_{\mathrm{N}}(\tau_q-n_N+1)\end{bmatrix}.$$

Then write the matrix equation:

$\mathbf{Y}=\mathbf{Uh}$.

We introduce the error function:

$\mathbf{E}=\overline{\mathbf{Y}}-\mathbf{Y}=\overline{\mathbf{Y}}-\mathbf{Uh}$,

where $\overline{\mathbf{Y}}$ is the actual measurement of output, and minimize the objective function:

$$\boldsymbol{\Psi}=\mathbf{E}^2=(\overline{\mathbf{Y}}-\mathbf{Uh})^2. \qquad (10)$$

The constraints on transient response components are written as:

$$\mathbf{s}^{\min}\le\mathbf{s}\le\mathbf{s}^{\max}, \qquad (11)$$

where $\mathbf{s}=[s_1(1),...,s_1(n_1),...,s_N(1),...,s_N(n_N)]^{\mathrm{T}}$, $\mathbf{s}^{\min}=[\mathbf{s}_1^{\min},...,\mathbf{s}_N^{\min}]^{\mathrm{T}}$, $\mathbf{s}^{\max}=[\mathbf{s}_1^{\max},...,\mathbf{s}_N^{\max}]^{\mathrm{T}}$.

The transient response components s are related with the components of the impulse response h by the relations:

$$s_j(k)=\sum\nolimits_{i=1}^{k}h_j(i),\ \ j=1,\ 2,\ \dots,\ N,\ \ k=1,\ \dots,\ n_j. \qquad (12)$$

The constraints (11) are reduced to:

$$\mathbf{A}\tilde{\mathbf{h}}\ge\hat{\mathbf{s}}, \qquad (13)$$

where

$$\mathbf{A}=\begin{bmatrix}1 & 0 & \cdots & 0\\ 1 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\\ -1 & 0 & \cdots & 0\\ -1 & -1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & -1\end{bmatrix},\qquad \tilde{\mathbf{h}}=\begin{bmatrix}h_1(1)\\ \vdots\\ h_1(n_1)\\ \vdots\\ h_N(1)\\ \vdots\\ h_N(n_N)\end{bmatrix},\qquad \hat{\mathbf{s}}=\begin{bmatrix}\mathbf{s}^{\min}\\ -\mathbf{s}^{\max}\end{bmatrix}.$$

The sufficient minimum conditions are as follows:

1. $\mathbf{A}\tilde{\mathbf{h}}\ge\hat{\mathbf{s}}$, $\mathbf{A}_{ACT}\tilde{\mathbf{h}}=\hat{\mathbf{s}}_{ACT}$

2. $\mathbf{Z}^{\mathbf{T}} * \mathbf{U}^T(\overline{\mathbf{Y}} - \mathbf{Uh}^*) = 0$ (14)
3. $\boldsymbol{\lambda} = \left(\mathbf{A}_{ACT}\mathbf{A}_{ACT}^T\right)^{-1}\mathbf{A}_{ACT}\mathbf{U}^T(\overline{\mathbf{Y}} - \mathbf{Uh}^*)$, $\boldsymbol{\lambda}_i > 0$, $i = 1,\dots,t$
4. $Z^T\mathbf{U}^T\mathbf{U}\ Z$ is positive definite,

where $\mathbf{h}^*$ - is the minimum point, the solution of problem (10) with constraints (13); subscript *ACT* indicates that in vector, matrix only the elements, rows with index numbers corresponding to the elements index numbers of active constraint in (13) are used; t - number of active constraints, $\boldsymbol{\lambda}$ - the vectors of Lagrange multipliers corresponding to the active constraints, $\mathbf{Z}$ - matrix the columns of which is basis of the feasible direction of search for equality constraints (13). The matrix $\mathbf{Z}$ is formed by the variable-reduction technique. [4].

In order to start the method of the active set it is necessary to determine the starting point (using a solution of the problem (10) without any constraints, with subsequent correction of coefficients h_i that does not fall under the constraints (13)).

The search algorithm of minimum point $\mathbf{h}^*$ for iteration *k* is:

1. Performance verification of the stop conditions (reaching the performance errors of conditions (14), constraints on the number of iterations).
2. Selection of a logic branch. Does it make sense to remove any constraint of the set of active constraints list. The condition of performance of a condition 3 in (14) is checked. If the condition is not satisfied for some of the vector element, constraint is excluded from the list of active constraints and it is need to recalculate $\mathbf{Z}_k$.
3. The calculation of the search direction $\mathbf{p}_k$. Like (4) solves the problem $\min\left(\overline{\mathbf{Y}} - \mathbf{Uh}_k - \mathbf{UZ}_k\mathbf{p}_Z\right)^2$. Calculate the non-zero $\left(1 + \sum_{k=1}^N n_k - t_k\right)$ - dimensional vector $\mathbf{p}_Z$ and the direction of search $\mathbf{p}_k = \mathbf{Z}_k\mathbf{p}_Z$, where t_k - the number of active constraints on *k* iteration.
4. Calculate the step length α_k. From $\mathbf{A}\left(\tilde{\mathbf{h}} + \boldsymbol{\Psi}\tilde{\mathbf{p}}_k\right) = \hat{\mathbf{s}}$ diagonal matrix $\boldsymbol{\Psi}$ is calculated. Calculated $\overline{\alpha}_k = \min\{\boldsymbol{\Psi}_{ii}\}$ - a minimum non-negative available step from $\mathbf{h}_k$ along $\mathbf{p}_k$, where i is the index number of element, which is not active constraints in (13) and is not element of opposite boundary values (11) for constraints in the active set, $\tilde{\mathbf{p}}_k$ consists from elements of $\mathbf{p}_k$ without first element. The index j of minimum positive diagonal element of $\boldsymbol{\Psi}_{ii}$ is remembered. If $\overline{\alpha}_k > 1$, then $\alpha_k = 1$, otherwise $\alpha_k = \overline{\alpha}_k$.
5. Constraint is added to the list of active constraints. If $\alpha_k = \overline{\alpha}$, then j constraint $\hat{\mathbf{s}}_{FR}$ becomes active and recalculate $\mathbf{Z}_k$.
6. Recalculation approximation. The $\mathbf{h}_{k+1} = \mathbf{h}_k + \alpha_k\mathbf{p}_k$ is calculated and return to step 1.

3 The influence of constraints of parameters of soft sensors model

A priori knowledge about the process and industrial step test (when the value of the one control variables is changed at fixed others) allow to define the value of constraints.

In order to investigate the influence of constraints on the quality of the obtained static soft sensor model we compare solutions of the equations (6) and constrained optimization problem (7). One and the same value of the ridge coefficient is used.

The Fig. 1 and Table 1 show the results of the performance of the static models on the test sample for the bubble-point temperatures of the target product (BP) when a model obtained on the training sample, consisting of a number of measurements specified in the Table 1. In the verification sample of models the number of measurements is equal to 420.

Table 1. Results of the performance of the static models

The number of measurements in the training sample	Without use constraints		With use constraints	
	R^2	RMSE	R^2	RMSE
30	0,188	1,790	0,604	1,249

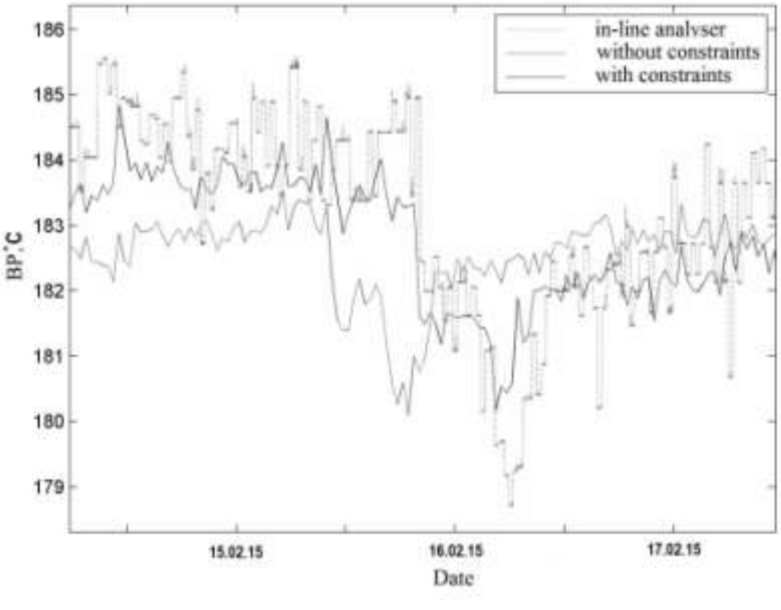

Fig. 1. Comparative study of static soft sensor models performance

In order to investigate the influence of constraints on the quality of the obtained dynamic soft sensor model we compare solutions of the optimization problem (10) and optimization problem (10) with constrains (13). One and the same value of the ridge coefficient is used.

The Fig. 2 and Table 2 show the results of the performance of the dynamic models on the test sample for the dew-point temperatures of the target product (DP) when a model obtained on the training sample, consisting of a number of measurements specified in the Table 2. In the verification sample of models the number of measurements is equal to 650.

Table 2. Results of the performance of the dynamic models

The number of measurements in the training sample	Without use constraints		With use constraints	
	R^2	RMSE	R^2	RMSE
700	0,327	1,456	0,634	1,074

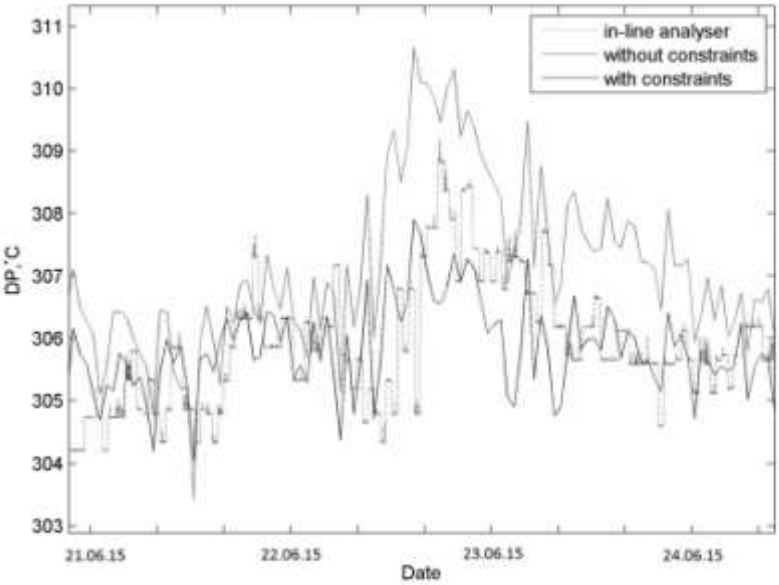

Fig. 2. Comparative study of dynamic soft sensor models performance

The estimation of improvements of the prediction quality by the criterion RMSE of static model obtained with the constraints on the parameters of SS is 100×(1,79 - 1,249) / 1,79 ≈ 30% compared to the model without constraints. The estimation of improvements of the prediction quality by the criterion RMSE of dynamic model obtained with the constraints on the parameters of SS is 100 · (1,456– 1,074)/ 1,456 ≈ 25 % compared to the model without constraints.

Conclusion

The using the method of the active set, taking into account constraints on the model coefficients can improve quality of the evaluated SS models.

The test of the proposed approach to solving the problem of obtaining a soft sensor model for industrial crude oil distillation unit is showed that the decrease root mean square error on the test sample can be not less than 25%.

References

1. Bakhtagze. N.N.: Virtual Analyzers: Identification Approach. In: Automation and Remote Control. Vol. 65, issue 11, pp.1691-1709. (2004)
2. Draper N., Smith H.: Applied regression analysis. M .: Finance and Statistics (1986)
3. Bolshakov A.A., Karimov R.N.: Methods of processing of the multidimensional given and time numbers. M .: Hotline Telecom (2007)
4. Gill, P.E., Murray, W., and Wright, M.H.: Practical Optimization, London: Academic, 1981. Translated under the title Prakticheskaya optimizatsiya, Moscow: Mir (1985)