ISSN 2070-0482, Mathematical Models and Computer Simulations, 2020, Vol. 12, No. 6, pp. 915–925. © Pleiades Publishing, Ltd., 2020.
Russian Text © The Author(s), 2020, published in Matematicheskoe Modelirovanie, 2020, Vol. 32, No. 3, pp. 127–142.

Predictive Modeling of Mass-Transfer of Plant Using an Algorithm of Alternating Conditional Expectations

I. S. Mozharovsky[a, b], S. A. Samotylova[b, c, *], and A. Yu. Torgashov[b, c]

[a] *Vladivostok State University of Economics and Service, Vladivostok, Russia*

[b] *Institute of Automation and Control Processes, Far Eastern Branch, Russian Academy of Sciences, Vladivostok, Russia*

[c] *Far Eastern Federal University, Vladivostok, Russia*

**e-mail: samotylova@dvo.ru*

Received September 10, 2019; revised September 10, 2019; accepted October 21, 2019

Abstract—The problem of predictive modeling under condition of the nonlinearity of the mass–transfer plant (MTP) based on the experimental data is considered. To analyze the structural identifiability of the process under study and identify factors that affect the accuracy of the structural identifiability index with an unknown model structure, a technique based on an alternating conditional expectation (ACE) algorithm with a threshold value for the structural identifiability index of the MTP model is proposed. The threshold value of the structural identifiability index is determined based on the rigorous model of the plant, i.e., taking into account the physicochemical characteristics of the MTP. The proposed approach is illustrated using synthetic data and experimental data.

Keywords: ACE algorithm, index of structural identifiability, mass–transfer plant, predictive modeling

DOI: 10.1134/S2070048220060137

1. INTRODUCTION

Due to increasing demands on the quality of the main types of petroleum products, the oil refining and petrochemical industries are forced to continuously improve the economic efficiency of production and the quality of products [1]. Production efficiency can be improved with the help of virtual monitoring systems and systems to check the quality of the output products and mass transfer processes such as distillation and absorption [2].

The development of new methods of predictive modeling, which means the use of statistical methods for model design oriented to estimate the quality indicators of the output variable of a plant taking into account the current values of the input variables [3], will provide a real-time noticeable increase in production efficiency.

The selection of input variables $X = (x_1, ..., x_p)$ affecting the output value Y, and the choice of the structure of the model can be based on correlational and regression analysis [4]. However, for nonlinear plants, the use of these methods does not allow us to determine the structure of the model. This leads to the ambiguity of obtaining estimates of the unknown parameters of the model $B = (\beta_0\beta_1, ..., \beta_p)$, when the same sample of the experimental data corresponds equally well not to one but to the set of models at once $F(X, B)$. This situation indicates the unidentifiable structure of the model. Structural identifiability occurs when two models $M(B)$ and $M(B^*)$ with the same structure $M(\cdot)$ are called indistinguishable by output (we denote this property $M(B^*) \approx M(B),)$ $B, B^* \in \Omega$ if for any valid input $x(t)$ the models have the same outputs $Y(t, B, x) \equiv Y(t, B^*, x)$ for any $t \geq 0$ [5–7]. Structural identifiability means the identifiability of the structure not of a single model but of the whole family of models [8].

For the analysis of structural identification, many different methods and algorithms for dynamic systems are proposed. For linear systems, the analysis of structural identifiability is understood quite well. There are a number of methods for its analysis, for example, the transfer function method [9], similarity transformation, and approaches based on the theory of differential algebra and graph theory [10].

However, for nonlinear plants it is much more difficult to analyze structural identifiability. This is due to the fact that the number of unknown model parameters may be more than the number of equations in

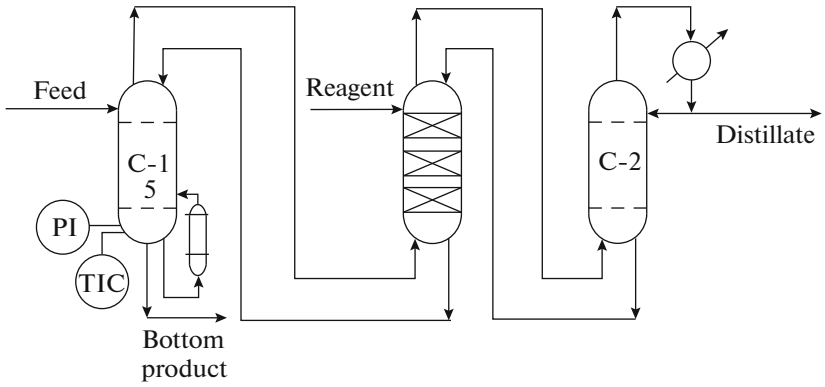

Fig. 1. Technological scheme of MTP.

the system [11]. For such plants, structural identifiability analysis is carried out using methods such as the decomposition of the output function Y into a Taylor series and studying the eigenvalues of the Fisher information matrix [12].

To analyze the identifiability of large-dimensional models, we apply the probabilistic algorithm method [13], which calculates the parameters for a system with an unknown structure of the plant model. The algorithm is based on the algebraic calculation of the rank of a certain power series of output functions. The rank is required to calculate the degree of transcendence (the degree of freedom of the expansion area related to the parameter). Despite the fact that this algorithm is widely used, it does not allow us to determine the source of nonidentifiability and does not group the parameters according to their functional relationships, nor does it provide transformations or reparameterization to make the model identifiable.

In the case when it is required to build a model for an industrial mass transfer process with an unknown model structure, the problem of structural identifiability remains relevant [14].

Among the available numerical nonparametric methods for extracting dependences from data for mass transfer processes and estimating the identifiability of plants, the most effective approach is based on the alternating conditional expectation (ACE) algorithm [15].

In relation to this, it is proposed to use the ACE algorithm and an additional input variable, which is not correlated to the response, to analyze the structural identifiability of the studied plant. The characteristic feature of this study is the analysis of the structural identifiability of the model to estimate the quality indicator of the output product of a nonlinear mass–transfer plant (MTP) based on the experimental data. In addition, the analysis of the structural identifiability of models is not limited to clarifying the fundamental possibility of an unambiguous estimation of parameters $F(X, B)$. Considerable attention is paid to identifying the defined transformations $F(X, B)$, describing the studied plant and affecting the accuracy of the structural identifiability index. For this, the concept of the threshold value of the structural identifiability index of the MTP model is introduced. It is based on taking into account the physicochemical characteristics of the MTP under consideration.

2. DESCRIPTION OF MASS-TRANSFER PLANT AND PROBLEM STATEMENT

The problem of constructing a model for estimating the reagent's content (%) in the output product (bottom product) of the MTP in the case when the structure of the model is unknown is considered. The investigated MTP is shown in Fig. 1 and consists of two distillation columns (C-1 and C-2) and a synthesis reactor located between them.

The structure of the rigorous model of the plant is rather complicated for practical application. In the general form, it can be represented as a system of equations for each kth stage of separation for each lth

component, which includes material balance equations, energy balance equations, and phase equilibrium equations [16]:

$$\begin{cases} L_{k+1}\tilde{x}_{k+1,l} + V_{k-1}\tilde{y}_{k-1,l} + F_k z_{k,l} - L_k\tilde{x}_{k,l} - V_k\tilde{y}_{k,l} = 0, \\ L_{k+1}h_{k+1} + V_{k-1}H_{k-1} + F_k H_{F_k} - L_k h_k - V_k H_k = 0, \\ \tilde{y}^*_{k,l} = \tilde{x}_{k,l}\gamma^L_{k,l}(p^0_{k,l}/P), \\ E_k = (\tilde{y}_{k,l} - \tilde{y}_{k+1,l})/(\tilde{y}^*_{k,l} - \tilde{y}_{k+1,l}), \qquad \begin{pmatrix} l = 1,\ldots,c, \\ k = 1,\ldots,N \end{pmatrix}, \\ \sum_{l=1}^{c}\tilde{y}_{k,l} - 1 = 0, \\ \sum_{l=1}^{c}\tilde{x}_{k,l} - 1 = 0, \end{cases} \tag{1}$$

where $\tilde{y}_{k,l}$ is the concentration of the lth component on the kth stage in the vapor phase; L_{k+1} is the flow of the fluid entering the kth stage; $\tilde{x}_{k+1,l}$ is the concentration of the lth component arriving at the kth stage in the liquid phase; V_{k-1} is the flow of the steam leaving the kth stage; $\tilde{y}_{k-1,l}$ is the concentration of the lth component leaving the kth stage in the vapor phase; F_k is the consumption of raw materials supplied to the kth stage; $z_{k,l}$ is the amount of the lth component in raw materials supplied to kth stage; L_k is the flow of the fluid on the kth stage; $\tilde{x}_{k,l}$ is the concentration of the lth component on the kth stage in the liquid phase; V_k is the flow of the steam on the kth stage; $\gamma^L_{k,l}$ is the activity coefficient of the lth component in the liquid phase on the kth stage (the UNIQUAC model is used); p^0_l is the partial pressure of the lth component; P is the total pressure in the system; E_k is the efficiency of the mass transfer according to Murphree on the kth stage; h_{k+1} is the enthalpy of the fluid entering the kth stage; H_{k-1} is the enthalpy of the vapor leaving kth stage; H_{F_k} is the enthalpy of the power on the kth stage; h_k is the enthalpy of the liquid on the kth stage; H_k is the enthalpy of the vapor at the kth stage; c is the total number of components in the system; and N is the total number of stages in the distillation column.

The main problem with using the rigorous model is that E_k is an unknown quantity. Also, the composition of feed is not known; therefore, it is impossible to use the analytical model directly to estimate the concentration of the reagent in the bottom product. Therefore, in practice, linear regression models of the form

$$\hat{Y} = \hat{\beta}_0 + \sum_{i=1}^{p}\hat{\beta}_i x_i, \tag{2}$$

where x_i are the input variables available for measurement at each time period; $\hat{Y}$ is the estimated value of the output variable of plant Y; p is the number of input variables; $\hat{\beta}_0$ is a free coefficient of the model; and $\hat{\beta}_i$ are the coefficients of the model's parameters.

In the case of using multiple regression, the structure of the model must be determined, which reduces the problem to estimating the coefficients of the model's parameters. When the relationship between the response and the predictors is unknown or inaccurate, linear parametric regression can lead to erroneous results. The most effective approach for analyzing the structural identifiability of models for assessing quality indicators in the output of nonlinear MTPs under structural uncertainty is a nonparametric approach based on the ACE algorithm. This is justified by the fact that the optimal transformations obtained as a result of using ACEs do not require a priori assumptions about the form of functions that connect the output and input variables.

Then for p input variables x_i, $i = 1,\ldots,p$, and output Y, the model of a plant has the following form:

$$Y = F(X, B) + \varepsilon, \tag{3}$$

where $X = (x_1,\ldots,x_p)$ is the vector of input controlled technological variables; $B = (\beta_0,\beta_1,\ldots,\beta_p)$ is the vector of coefficients; and ε is the measurement error of the output variable.

The task consists of finding the possibility of constructing an adequate mathematical model for assessing the concentration of the reagent in the cubic product of the process under study on a training sample. An analysis of the structural identifiability of the plant is proposed to be carried out based on the calculation of the structural identifiability index using the ACE algorithm and the introduction of an additional input variable that is not correlated to the output. The structural identifiability index H_Y refers to the degree of dependence of the output on a different set of input variables. To assess the degree of the structural identifiability of the plant, the calculated values H_Y are compared to the threshold structural identifiability index H_{lv}, which is proposed to be determined based on the analytical model of the MTP under consideration (1), taking into account the physicochemical characteristics of the process.

3. DESCRIPTION OF ALGORITHM OF ALTERNATING CONDITIONAL MATHEMATICAL EXPECTATIONS

The ACE regression model has the following general form:

$$\theta(Y) = \sum_{i=1}^{p} \phi_i(x_i) + \varepsilon, \tag{4}$$

where θ is the function with the response variable Y; ϕ_i are the functions of the input variables (predictors) $x_i, i = 1,\ldots, p$.

Thus, the ACE model replaces the problem of estimating a linear function p-dimensional variable $X = (x_1, x_2,\ldots, x_p)$ of the assessment p of individual one-dimensional functions ϕ_i and θ using an iterative method. These transformations are achieved by minimizing the unexplained deviation of the linear relationship of the transformed response variable from the sum of the transformed variable predictors.

For the given dataset consisting of the response variable Y and variable predictors $x_1, x_2,\ldots, x_p$, the ACE algorithm begins with the definition of arbitrary initial transformations $\theta(Y), \phi_1(x_1),\ldots,\phi_p(x_p)$. The error variance ($\varepsilon^2$), which remained unexplained by the regression of the transformed dependent variables, is equal to the sum of the transformed independent variables provided that $E\left[\theta^2(Y)\right] = 1$:

$$\varepsilon^2(\theta, \phi_1,\ldots,\phi_p) = E\left\{\left[\theta(Y) - \sum_{i=1}^{p} \phi_i(x_i)\right]\right\}^2. \tag{5}$$

The minimization of ε^2 taking into consideration $\phi_1(x_1),\ldots,\phi_p(x_p)$ and $\theta(Y)$ is calculated through a series of minimizations of unit functions given by the equations

$$\phi_i(x_i) = E\left[\theta(Y) - \sum_{j\neq i}^{p} \phi_j(x_j)\middle| x_i\right], \tag{6}$$

$$\theta(Y) = E\left[\sum_{i=1}^{p} \phi_i(x_i)\middle| Y\right] \Bigg/ \left\| E\left[\sum_{i=1}^{p} \phi_i(x_i)\middle| Y\right]\right\|. \tag{7}$$

Equations (6) and (7) form the base of the ACE algorithm [15]. The final $\phi_i(x_i)$, $i = 1,\ldots, p$, and $\theta(Y)$ after minimization are estimates of the optimal transformation $\phi_i^*(x_i), i = 1,\ldots, p$, and $\theta^*(Y)$. The response and predictors are related as follows:

$$\theta^*(Y) = \sum_{i=1}^{p} \phi_i^*(x_i) + \varepsilon^*, \tag{8}$$

where ε^* is an error that cannot be fixed using ACE transformations under the assumption of a normal distribution. The minimum regression errors ε^* and maximum multidimensional correlation coefficient ρ^* are related by $\varepsilon^{*2} = 1 - \rho^{*2}$.

The optimal ACE transformations are obtained numerically based on the data of the technological plant and do not require a priori assumptions about the specific functional form that relates the response to the predictors [17].

4. ALGORITHM FOR ANALYZING THE STRUCTURAL IDENTIFIABILITY OF A NONLINEAR PROCESS

For the analysis of structural identifiability, the base matrix (the data matrix formed from the sample containing the values of the input and output variables) and the number of perturbed data matrices M obtained from the base matrix by adding small random numbers to its elements are used as the initial information.

Step 1. Transform the base matrix into an extended data matrix with the dimension $K(p+2)$, where K is the number of observations, p is the number of predictors, $p+1$ is an additional normally distributed input not correlated to output ξ with mathematical expectation $\mu=0$ and dispersion $\sigma^2=2.5$ $(\zeta \in N(-2.5;2.5))$, and $p+2$ is the response variable Y.

Step 2. Obtain the base set of vectors of optimal transformations for each input of the studied plant $\Phi_i = \phi_i^*(X_i)$ by applying the ACE algorithm to an extended data matrix,

$$\Phi_i^{base}(x_i) = (\Phi_i^{base,1},\ldots,\Phi_i^{base,j},\ldots,\Phi_i^{base,K})^T,$$
$$i = 1,\ldots,p+2,\ \ x_{p+1} = \xi,\ \ x_{p+2} = Y, \tag{9}$$

and the vector of differences (base matrix of the optimal transformations)

$$\Delta\Phi_i^{base} = (\Delta\Phi_i^{base,1},\ldots,\Delta\Phi_i^{base,k},\ldots,\Delta\Phi_i^{base,K-1})^T, \tag{10}$$

where

$$\Delta\Phi_i^{base,k} = \Phi_i^{base,k+1} - \Phi_i^{base,k},\ \ k = 1,\ldots,K-1. \tag{11}$$

Step 3. From the base matrix we form the set of size matrices $K(p+2)$ to obtain the vectors of the optimal transformations using the disturbing influences. To achieve this, we add the small random numbers $\alpha_k^q = \varepsilon_k^q \in N(-2.5;2.5)$ to variables $x_i, i = 1,2,\ldots,p,p+1,p+2$ and reduce the resulting numbers by 0.02% of the average for y $\left(\alpha_k^q = \varepsilon_k^q \times 0.0002 \times \sum_{k=1}^{K} y_k \Big/ \overline{y}\right)$, $k = 1,\ldots,K$, $q = 1,\ldots,M$ (the transformed matrix with the addition of small random numbers α_k^q).

Step 4. We find the set of vectors of the optimal transformations and differences:

$$\Phi_i^q(x_{\alpha,i}) = (\Phi_{\alpha,i}^{q,1},\ldots,\Phi_{\alpha,i}^{q,j},\ldots,\Phi_{\alpha,i}^{q,K})^{\mathrm{T}},\ \ i = 1,\ldots,p+2,\ \ q = 1,\ldots,M, \tag{12}$$

$$\Delta\Phi_{\alpha,i}^q = (\Delta\Phi_{\alpha,i}^{q,1},\ldots,\Delta\Phi_{\alpha,i}^{q,k},\ldots,\Delta\Phi_{\alpha,i}^{q,K-1})^T, \tag{13}$$

where $x_{\alpha,i} = x_i^k + \alpha_k{}^q$, $\Delta\Phi_{\alpha,i}^{q,k} = \Phi_{\alpha,i}^{q,k+1} - \Phi_{\alpha,i}^{q,k}$, $k = 1,\ldots,K-1$, and ^{T} is the sign of the transposition.

Step 5. Normalize vectors $\Delta\Phi_i^b$ and $\Delta\Phi_i^q$ and transform the resulting difference vectors (10) and (13) to the following form:

$$\Delta\Phi_{m,i}^{base} = (\Delta\Phi_{m,i}^{base,1},\ldots,\Delta\Phi_{m,i}^{base,k},\ldots,\Delta\Phi_{m,i}^{base,K-1})^T, \tag{14}$$

$$\Delta\Phi_{m,i}^q = (\Delta\Phi_{m,i}^{q,1},\ldots,\Delta\Phi_{m,i}^{q,k},\ldots,\Delta\Phi_{m,i}^{q,K-1})^T,\ \ q = 1,\ldots,M, \tag{15}$$

where $\Delta\Phi_{m,i}^{base,k} = \Delta\Phi_{\alpha,i}^{base,k} \Big/ S_i^{base}$, $\Delta\Phi_{m,i}^{q,k} = \Delta\Phi_{\alpha,i}^{q,k} \Big/ S_i^q$, index m is the sign of the averaging of the differences

$$\Delta\Phi_{\alpha,i}^{base,k},\ \ \Delta\Phi_{\alpha,i}^{q,k},\ \ S_i^{base} = ((\Delta\Phi_{\alpha,i}^{base} - \overline{\Delta\Phi}_{\alpha,i}^{base})^2 \Big/ (K-2))^{1/2},$$

$$S_i^q = ((\Delta\Phi_{\alpha,i}^q - \overline{\Delta\Phi}_{\alpha,i}^q)^2 \Big/ (K-2))^{1/2},\ \ \overline{\Delta\Phi}_{\alpha,i}^{base} = \sum_{k=1}^{K-1} (\Delta\Phi_{\alpha,i}^{base,k}) \Big/ (K-1),$$

Table 1. Coefficients of correlation

Correlational relationships	Variables					
	x_1	x_2	x_3	x_4	x_5	Y
Pair Correlation Coefficients	0.5178	0.1161	0.1339	0.0207	0.0536	1.0000
Correlation	0.5586	0.4111	0.4314	0.3991	0.0780	1.0000

$$\overline{\Delta\Phi}^{q}_{\alpha,i} = \sum_{k=1}^{K-1} (\Delta\Phi^{q,k}_{\alpha,i}) \Big/ (K-1).$$

Step 6. Find the deviations of the differences (14) of the base optimal transformations from differences (15) for each $q = 1,\ldots,M$:

$$\Delta V_i^{q,k} = \Delta\Phi_{m,i}^{base,k} - \Delta\Phi_{m,i}^{q,k}, \quad i = 1,\ldots,p+2, \quad k = 1,\ldots,K-1, \tag{16}$$

from which we form the sequence of vectors

$$\Delta V_i^{q} = (\Delta V_i^{q,1},\ldots,\Delta V_i^{q,k},\ldots,\Delta V_i^{q,K-1})^{\mathrm{T}}, \quad i = 1,\ldots,p+2. \tag{17}$$

Step 7. Obtain a quantitative estimate of the deviations ΔV_i^q from (17):

$$\Delta E_i^q = \sum_{k=1}^{K-1} \left|\Delta V_i^{q,k}\right|, \quad i = 1,\ldots,p+2, \quad q = 1,\ldots,M. \tag{18}$$

Step 8. Determine the structural identifiability index by the *i*th variable:

$$H_i = \Delta E_{m,p+1} \big/ \Delta E_{m,i}, \quad i = 1,\ldots,p+2, \tag{19}$$

where $\Delta E_{m,i} = \sum_{q=1}^{M} \Delta E_i^q \big/ M$, ΔE_i^q is calculated by (18), and $H_{p+2} = H_Y$.

Step 9. Compare the resulting structural identifiability indices H_Y from (19) with the corresponding threshold value H_{lv}. If $H_Y > H_{lv}$, then the plant is identifiable, otherwise it is not identifiable based on the data provided.

5. ANALYSIS OF THE STRUCTURAL IDENTIFIABILITY USING A SYNTHETIC EXAMPLE

In order to demonstrate the operation of the ACE algorithm, a synthetic example was used to determine the functional dependence between dependent and independent variables, in which functional dependences are known.

Let the plant be defined by a functional dependence of the following form:

$$y = \sin(x_1) + \sin(1.7x_2) + \sin(3.4x_3) + \cos(2.4x_4) + \varepsilon. \tag{20}$$

According to Eq. (20) and input variables x_i, $i = 1,\ldots,4$, on which the restrictions $-2.5 \le x_1,\ldots,x_4 \le 2.5$ are placed, a sample of volume $K = 1000$, representing a $(K \times 5)$ matrix, is formed. An extended sample is obtained by including in the original sample an additional—uncorrelated to output Y—input x_{err}, $x_{err} \in N(-2.5; 2.5)$,and $x_5 = x_{err} \times 0.0002$, where x_5 is an additional input, uncorrelated with the output, $-0.0025 \le \varepsilon \le 0.0025$.

The analysis of the pair correlation coefficients and correlational relations obtained on the initial sample (Table 1) does not allow us to draw a conclusion on the possible structure of the model.

Applying the ACE algorithm to the extended sample, we form the base set of vectors of the optimal transformations $\Phi_i^{basic}(x_{\alpha,i})$, graphically presented in Fig. 2 and indicating a fairly accurately found model structure.

For the analysis of structural identifiability at $M = 25$ (where M is the number of iterations of cycle repetition), 25 vectors of the optimal transformations were obtained; and they were compared with the base estimates of the model.

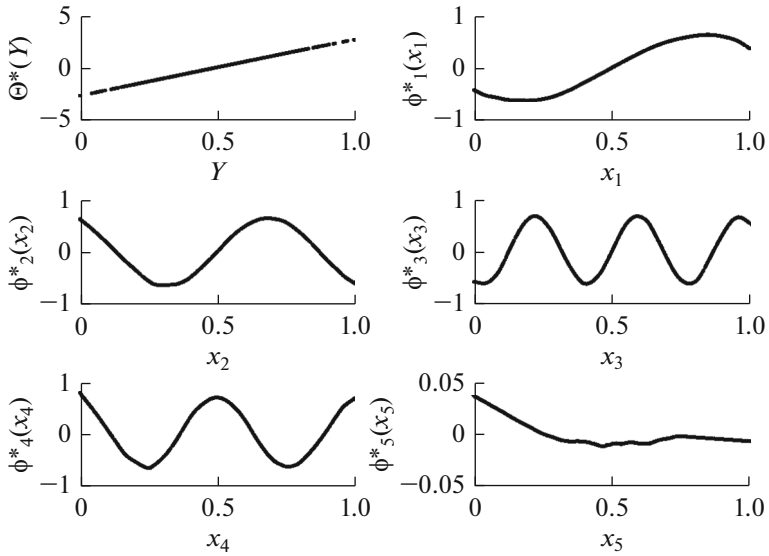

Fig. 2. The result of applying the ACE algorithm to the elements of the base matrix.

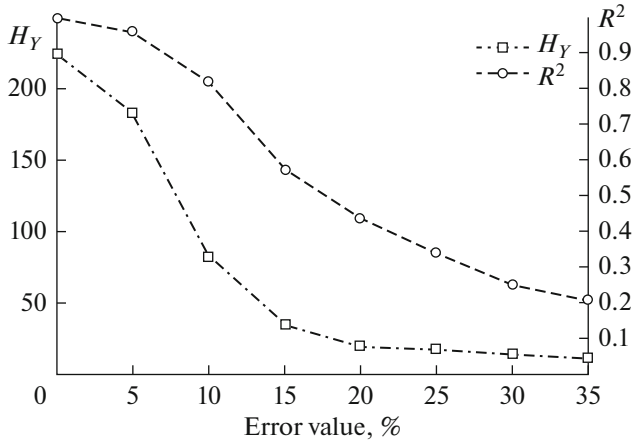

Fig. 3. Dependence H_Y and R^2 on noisiness of synthetic data.

Table 2 shows the results of applying the proposed approach.

The value $\Delta E_{m,5} = 0.0394$ means that the parameter at input x_5 is unidentifiable. The other values $\Delta E_{m,i}$ (the average sum of the distances between the points of the model's base estimate from the model's current estimate for the output and each input) fully confirm the existence of a nonlinear model for the studied plant and can serve as a sign of its identifiability. Quantities H_i reflect the contribution of each variable with respect to an unidentifiable auxiliary input. The results obtained correspond to description (20). Thus, the plant is identifiable, since the value of the identifiability indicator for output $H_Y = 224.58$,

Table 2. Parameters of structural identifiability for H_i

Parameter	x_1	x_2	x_3	x_4	x_5	Y
$\Delta E_{m,i}$	0.0009	0.0007	0.0023	0.0037	0.0394	0.0002
H_i	45.67	76.69	20.92	16.39	1.00	224.58

Table 3. Coefficients of correlation

Correlational relationships	Variables			
	x_1	x_2	x_3	Y
Pair Correlation Coefficients	−0.7822	−0.0800	−0.0330	1.0000
Correlation	0.8496	0.2076	0.0969	1.0000

Table 4. Parameters of structural identifiability for a real plant

Parameter	x_1	x_2	x_3	Y
$\Delta E_{m,i}$	0.0012	0.0029	0.0236	0.0004
H_i	9.3378	11.4167	1.0000	68.1348

which is significantly more than the specified threshold value $H_{lv}=35.16$ with an error of 15% when $R^2<0.7$.

The threshold value H_{lv} is determined experimentally by varying H_Y in the error range (%) $[0; 35]$ from the average of each input variable. The results are shown in Fig. 3.

If $H_Y < 35.16$, then the model is not identifiable. When $H_Y > 35.16$, the higher H_Y the more accurate the model.

6. ANALYSIS OF STRUCTURAL IDENTIFIABILITY ON THE EXAMPLE OF THE MTP

When building a model for evaluating the concentration of the reagent in the still of the distillation column C−1 the data from a real technological plant were used. The temperature (x_1 is TIC, °C) and pressure (x_2 is PI, MPa) bottom of the distillation column C−1 were selected as the input data parameters of the model.

The analysis of the pair correlation coefficients and correlational relations obtained on the initial sample (Table 3) does not allow us to make a conclusion about the possible structure of the model.

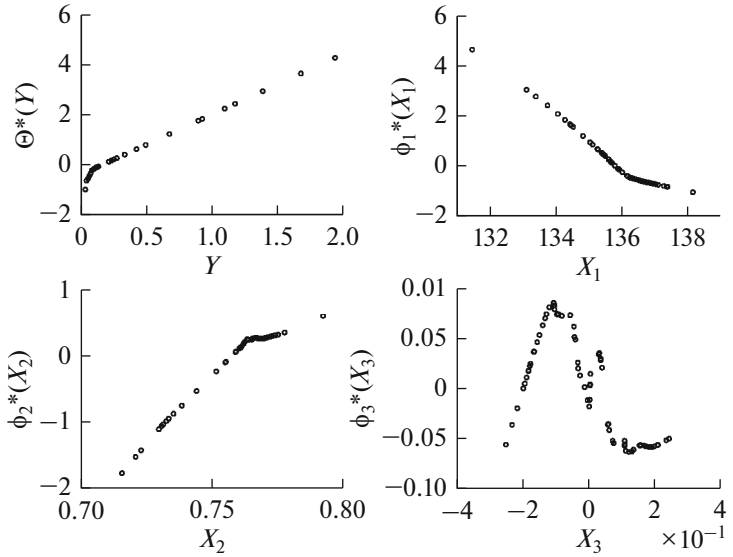

Fig. 4. The result of applying the ACE algorithm to the elements of the base matrix of industrial data.

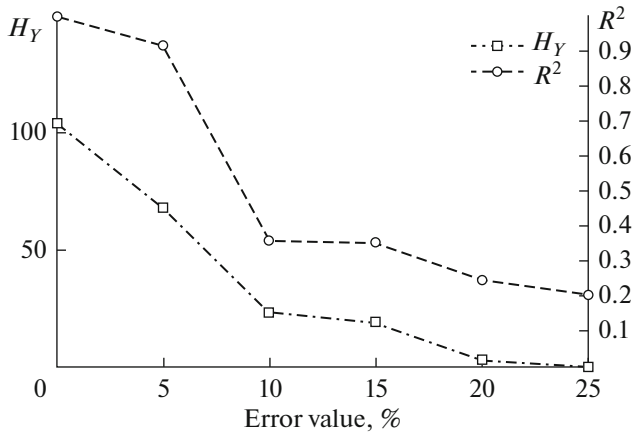

Fig. 5. Dependence H_Y and R^2 on the noisiness of data of rigorous model.

To assess the identifiability of M =125, 125 vectors of the optimal transformations were obtained and they were compared with the base estimates of the model (Fig. 4).

Table 4 presents the results of applying the proposed approach.

Value $\Delta E_{m,3} = 0.0236$ for input x_3 allows us to conclude that its corresponding transformation $\phi_3(x_3)$ (4) is not identifiable. The other values of $\Delta E_{m,i}$ confirm the existence of a nonlinear model for the studied plant. According to the values of index H_i presented in Table 4, we can conclude that the plant is structurally identifiable since the value of the identifiability index indicator for the output of the test sample $H_Y = 68.1348$, which is significantly more than the specified threshold value $H_{lv} = 22.65$.

The threshold value H_{lv} was determined on the generated data sample of a calibrated rigorous model in the error interval (%) $[0; 25]$ from the average value of each input variable (Fig. 5).

In this case $R^2 < 0.7$ when the data noise is 10%, which corresponds to the threshold value $H_{lv} = 22.65$.

Based on the experimental data of the technological process using various approximation methods (linear, logarithmic, exponential, quadratic) of the variables transformed by the ACE algorithm for the output variable Y and inputs x_1 and x_2 [17], a model of the following form was obtained:

$$\hat{Y} = 551.78 - 9.95x_1 + 0.05x_1^2 + 369.01x_2 + 77.24x_2^2 - 3.51x_1x_2. \quad (21)$$

Table 5 presents the coefficients of determination (R^2) and the root-mean-square error (RMSE) of the parametric models obtained by the least squares method (LSM), robust regression (RR), model (21), and a nonparametric model based on the ACE algorithm [18] for the training ($_{\text{training}}$) and test ($_{\text{test}}$) samples.

According to the results presented in Table 5, it can be seen that the nonparametric model constructed based on the ACE algorithm describes the studied MTP more accurately than the other methods. The results of the operation of a nonparametric model based on the ACE algorithm and experimental data are shown in Fig. 6.

Table 5. Values R^2 and RMSE for the presented models

Model		R^2_{training}	R^2_{test}	$\text{RMSE}_{\text{training}}$	$\text{RMSE}_{\text{test}}$
Parametric	LSM	0.7988	0.0508	0.1728	0.2820
	RR	0.7949	0.2286	0.1745	0.2542
	Approximation	0.8309	0.4111	0.1584	0.2221
Nonparametric model based on ACE algorithm		0.9967	0.9482	0.0129	0.0659

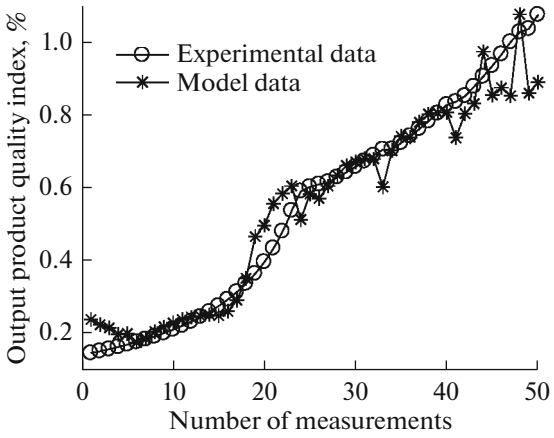

Fig. 6. The results of the operation of a nonparametric model constructed on the ACE algorithm to evaluate the quality indicator of the output product in the test sample.

7. CONCLUSIONS

The article presents a method for the analysis of the structural identifiability based on the ACE algorithm with the addition of an additional input variable that is not correlated with the output under the conditions of the unknown structure of the MTP model. The calculated value of the structural identifiability index H_Y in the experimental data should not be less than its threshold value H_N, which can be found in advance using the rigorous (taking into account the physicochemical laws) MTP model. Carrying out an analysis of the structural identifiability using the proposed ACE-based approach avoids the endless enumeration of model structures and allows us to determine the limit on the maximum accuracy of the model, as demonstrated by a synthetic example and real experimental data of the technological process.

By the example of constructing models for estimating the concentration of a reagent in the bottom product under the conditions of nonlinearity of the MTP, it is shown that the use of a nonparametric model based on the ACE algorithm improves the accuracy of the model to $(0.2221 - 0.0659/0.2221)\times 100\% \approx 70.3\%$ of the RMSE compared to the nonlinear parametric model (21) on the test sample.

FUNDING

This work was partially supported by the Russian Foundation for Basic Research (project no. 17-07-00235 A).

REFERENCES

1. M. J. Olanrewaju, B. Huang, and A. Afacan, "Online composition estimation and experiment validation of distillation processes with switching dynamics," Chem. Eng. Sci. **65** (5), 1597–1608 (2010).
2. T. Chatterjee and D. N. Saraf, "On-line estimation of product properties for crude distillation units," J. Process Control **14**, 61–77 (2004).
3. M. Kuhn and K. Johnson, *Applied Predictive Modeling* (Springer, New York, 2013).
4. G. B. Digo, N. B. Digo, A. V. Kozlov, S. A. Samotylova, and A. Yu. Torgashov, "Structural and parametric identification of soft sensors models for process plants based on robust regression and information criteria," Autom. Remote Control **78** (4), 724–731 (2017).
5. E. Walter and L. Pronzato, "On the identifiability and distinguishability of nonlinear parametric models," Math. Comput. Simul. **42**, 125–134 (1996).
6. C. Cobelli and J. J. Distefano, "Parameter and structural identifiability concepts and ambiquities: A critical review and analysis," Am. Physiol. Soc. **239** (1), 7–24 (1980).
7. M. J. Chappell and K. R. Godfrey, "Structural identifiability of the parameters of a nonlinear batch reactor model," Math. Biosci. **108**, 241–251 (1992).
8. R. Bellman and K. J. Astrom, "On structural identifiability," Math. Biosci. **7** (3–4), 329–339 (1970).
9. N. Meshkat, "Identifiable reparametrizations of linear compartment models," Symbolic Comput. **63**, 46–67 (2014).

10. S. I. Kabanikhin, D. A. Voronov, A. A. Grodz, and O. I. Krivorotko, "Identifiability of mathematical models in medical biology," Russ. J. Genet.: Appl. Res. **6**, 838–844 (2016).
11. M. J. Chappell, K. R. Godfrey, and S. Vajda, "Global identifiability of the parameters of nonlinear systems with specified inputs: A comparison of methods," Math. Biosci. **102** (1), 41–73 (1990).
12. S. Vajda, H. Rabitz, E. Walter, and Y. Lecourtier, "Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models," Chem. Eng. Commun. **83**, 191–219 (1989).
13. A. Sedoglavic, "A probabilistic algorithm to test local algebraic observability in polynomial time," Symbolic Comput. **33**, 735–755 (2002).
14. R. Brown, "Compartmental system analysis: State of the art," IEEE Trans. Biomed. Eng. **27**(1), 1–38 (1980).
15. L. Breiman and J. Friedman, "Estimating optional transformations for multiple regression and correlation," J. Am. Stat. Assoc. **80**, 580–598 (1985).
16. C. D. Holland, *Fundamentals of Multicomponent Distillation* (McGraw-Hill Book Company, New York, 1981).
17. D. Wang and M. Murphy, "Estimating optimal transformations for multiple regression using the ACE algorithm," J. Data Sci. **2**, 329–346 (2004).
18. I. S. Mozharovskii, "A method for constructing a nonparametric model based on the ACE algorithm," in *XXXII International Scientific Conference Mathematical Methods in Engineering and Technology MMTT-32* (2019), Vol. 9, pp. 39–43.