Research of methods for design of regression models of oil and gas refinery technological units

V V Bukhtoyarov[1], **V S Tynchenko**[1], **E A Petrovsky**[1], **S G Dokshanin**[1] **and V V Kukartsev**[1]

[1] Siberian Federal University, 79, Svobodny pr., Krasnoyarsk, 660041, Russian Federation

E-mail: vladber@list.ru

Abstract. The problem of efficient computational models design for control and scheduling problems in terms of oil and gas refinery column distillation units is discussed in the paper. Such efficient computational models can be constructed in the form of fast static regression models supplemented with dynamic models of measurement and input channels. The effectiveness of methods for constructing fast static regression models is examined in the paper. The input parameters for such regression models are determined. It is proposed to use parametric optimization methods for such models. A preliminary study showed the possibility of using an evolutionary genetic algorithm. Numerical studies were performed using data from column distillation units. The efficiency of using the methods of additional parametric optimization is shown.

1. Introduction

Improving the efficiency of oil refineries at the present stage can be achieved through the introduction of highly efficient scheduling and control systems. It is known that the problem of formalizing the processes of integrated scheduling and control for oil refineries is extremely complex. This is a consequence of such factors and problems as scheduling of high dimensions, decision support under uncertainty, optimization problems with a large number of constraints and complexity of modeling technological processes separately and as a dynamic system. In this regard, the actual scientific problem is development and research of methods and algorithms, integration of which within the framework of distributed information and analytical systems will contribute to the effective solution of relevant production problems.

One of the main elements of refinery production are column distillation units (CDUs). That is why it is important to develop and research methods for modeling CDUs. Such methods will allow to design models for scheduling and control cyber-physical systems for such industries. Modeling of CDUs based on complex physicochemical models is investigated in sufficient detail [1 - 3]. In most cases, static modes of columns are introduced through modelling techniques [4, 5]. Although a number of research works are devoted to the modeling of transients modes. However, such models are of limited

applicability in operational control and scheduling tasks because models of this type do not meet the requirements for three main reasons [6 - 8]:

- Models of this type do not allow to predict values of parameters of units according to measured technological parameters.
- Time required for calculations using models of this type is too long even modern operational level computers are used.
- Initial information cannot be derived quickly for such models (for example, components of the input stream).

It is necessary to design models that meet the criteria of simplicity, accuracy and simplification of input formation. Such models are to efficiently calculate values of CDU's parameters for technological processes control tasks within the production systems of oil refining. We suppose to obtain such "fast" models through the solving regression problem in an automated mode. Modelling techniques and CDU parameters to be used in regression models are discussed in Section 2. Results of a numerical study of the proposed modeling approaches without and using parametric optimization algorithms on a distillation column data set are presented in Section 3.

2. Column distillation unit modelling

To design regression models following approaches were used: artificial neural networks, multidimensional adaptive regression splines, genetic programming, nonparametric regression methods, regression decision trees [9 - 12]. An urgent problem is to determine the methods that provide the most accurate prediction of CDU parameters necessary for assessing the state of technological processes and evaluating control actions. The analysis of preliminary results concerning CDU modeling led to the following conclusions:

- Sensitivity of the temperature on plates along height of the CDU to change in the parameters of the technological mode varies by two orders of magnitude.
- Temperature change on individual plates with respect to the alternating change in technological parameters is sharply asymmetric, which is explained by the cumulative effect of interdependent technological parameters.
- Selection of the most informative parameters (input coordinates) for the regression equations.

The analysis made it possible to form a subset of parameters that are the most informative for the purposes of calculation parameters of the CDU. These include:

- Temperatures on individual plates by column height, including temperature on the selection plates, for a given composition of raw materials and a fixed pressure P in the column;
- Relative sums of sampling separation products at fixed values of the steam flow in the column.
- Temperatures t_i, (i = 1,2,…,5) where i = 0 corresponds to the temperature of the top of the column, i = 5 corresponds to the temperature of the bottom of the column; i = 1,2,3,4 correspond to the temperature values on the selection plates;
- Pressure value at the top of the column;
- Values of side shoulder selection D_i, i = 1,2,3,4; and at the top of the column D_0, the value of the vapor flow in the column G.

In a number of studies, it is proposed to use static regression models supplemented by dynamic corrective links in measuring channels and channels of transmission of influences, as a class of models for the purpose of operational control and scheduling of column apparatus [13]. In this paper, methods for the effective automated design of such models are considered in terms of obtaining static regression

models of CDUs. To improve the efficiency and adaptability of such models, it is proposed to supplement the regression modeling approaches listed in the beginning of the section to be with methods of parametric optimization of regression models.

Parametric identification approaches are aimed to find parameters of the equations of a mathematical model with a known structure. The methods of parametric identification imply the need to choose (determine) the type of the approximating function $f(x,\lambda)$ up to the vector of parameters λ, which is found from the minimum condition of the approximation quality criterion. It is necessary to determine the coefficients of the accepted structure of the object model from the available sample of input output variables.

3. Numerical experiments

A series of numerical experiments was carried out for a comparative study of methods for constructing fast models of distillation columns. An evolutionary genetic algorithm and Litvakov and Kesten algorithms were used for parametric optimization of models designed by various methods [14, 15]. To determine the set of optimization parameters for each type of model numerical parameters were determined. The set of parameters includes all current model coefficients the setting of which is possible within the framework of the constructed model structure. Genetic algorithm was also considered as algorithm for parametric optimization of CDU models. Genetic algorithms requires representing solutions in the form of binary lines. Under this consideration parameters of the models were sampled using discretization step equal or smaller to 0.001. Intervals of variation of the parameters were determined individually for each model and were chosen to completely overlap parameters generated CDU model structure identification algorithm.

The parameters of the optimization algorithms were determined during preliminary testing on a set of test functions. This set consists of 25 test problems most frequently used in research in this area []. During preliminary tuning procedure settings of genetic algorithm were selected as follows: total number of search steps is equal to 100, number of solution search points at each step is equal to 100.

Estimations of mathematical expectation and variance of the modeling error were used as main measures of models efficiency. These measures were calculated using data obtained during 50 independent runs of the algorithms. The following formula was used to calculate the modeling error in each run:

$$Error = \frac{100\%}{s\left(y^{\max} - y^{\min}\right)} \sum_{i=1}^{s} |o_i - y_i| \tag{1}$$

To evaluate the indicators software package Statsoft Statistica was used. The results are shown in table 1.

Table 1. Simulation error for various methods.

	Regression model without parameter optimization	Regression model with parameter optimization
Artificial neural networks	11,9	5,9
Multidimensional regression splines	8,2	2,8
Decision trees	17,8	8,1
Genetic programming	15,6	7,5

Nonparametric regression	13,4	7,3

The results obtained in the course of numerical studies were processed using statistical methods to assess significance. For the verification the ANOVA technique for analysis of variance was used. The data were obtained during the construction of regression models for calculating and predicting the parameters of CDU. To assess the effectiveness of using models of various types, the results were analyzed both separately for each parameter considered and on average for a set of parameters. Thus, for most of the parameters, the best results were obtained using the method of multidimensional adaptive regression splines. The average modeling error for regression models obtained using this method without parametric model optimization was 8.2%. On average, for the regression models under consideration, the modeling error was 12.6%, which is a significant value, potentially limiting the possibility of their effective use for modeling technological installations in oil refineries and calculating control actions and parameters for them. This demonstrates the necessity and rationality of using the proposed additional procedures for parametric adaptation of the obtained regression models, which can be considered as initial solutions for obtaining more effective (in the sense of accuracy criterion) regression models.

The results obtained and the analysis of statistical processing prove that the use of additional algorithms for parametric optimization of models based on the evolutionary genetic algorithm can significantly reduce error. This is achieved due to the deep adaptation of model parameters due to the effective search for a set of optimal (or close to them) model parameters in the corresponding space. Despite the fact that the dimension of the corresponding optimization task is significant, which makes it difficult to apply classical optimization methods and direct exhaustive search methods, the heuristic algorithm of optimization of regression models used for calculating the parameters of technological installations demonstrated high efficiency.

As can be seen from the results presented in table 1, the use of the additional parametric optimization procedure allowed us to significantly reduce the modeling error by more than two times: from 15.6% to 7.5%. For the most effective method - the method of regression splines - the modeling error after parametric optimization of the model was 2.8%. After parametric optimization, the accuracy of the models, according to expert estimates of relevant specialists, is satisfactory, and the corresponding models can be used to simulate and predict process parameters of process plants, and as a result, to calculate the levels of control parameters of such plants.

Parametric optimization has significantly improved the accuracy of models obtained by all methods. Even the worst models in terms of the modeling accuracy criterion turned out to be more accurate than the best multidimensional regression splines without parametric optimization. This result can be used as follows, especially considering that none of the methods turned out to be more effective than the others on all tasks. It is possible to use any of the considered methods in combination with the genetic algorithm for parametric optimization of models for designing effective regression models for finding effective control parameters of column technological units.

4. Conclusion

The problem of oil refineries CDUs regression models design is formulated. The input parameters of static regression models were estimated. It is supposed that it is necessary to use efficient optimization algorithms that can perform optimization of model parameters in multidimensional spaces. The problems are considered on the example of modeling the parameters of the CDU.

Methods of parametric optimization of regression models are considered. As a result of estimating the parameters of the corresponding optimization problem, it is proposed to use an evolutionary genetic algorithm to optimize the parameters of the control models of the technological systems of the oil and gas complex using the example of the considered oil refining installation. A preliminary study of the effectiveness of the selected optimization method on a set of test functions that simulate complex objective functions of parametric model optimization problems was performed. According to the results,

the parameters of the algorithm implementing the chosen method are determined, providing the best efficiency on average for a variety of tasks.

The results of a numerical experimental study of methods for constructing regression models of CDU were obtained and statistically processed. The results prove the effectiveness of the proposal formulated in the paper on the use of fast regression models for estimating, predicting and determining the values of parameters used to control the technological installations used as part of the production systems of the oil and gas industry.

The results obtained in the course of numerical studies also confirm the effectiveness of the proposal formulated in the paper on the use of an evolutionary genetic algorithm for the parametric optimization of models of technological installations operating as part of the production systems of the oil and gas complex. This is confirmed by obtaining effective models that provide modeling errors of less than 3% for 11 parameters of the vacuum column, which, in accordance with expert estimates, is sufficient to use these models as a basis for modeling the parameters of relevant technological processes and selecting effective values of control parameters taking into account model estimates.

References

[1] Luyben W L 2012 *Practical distillation control* (Springer Science & Business Media)

[2] Yildirim Ö, Kiss A A and Kenig E Y 2011 Dividing wall columns in chemical process industry: a review on current activities *Separation and Purification Technology* **80(3)** 403-17

[3] Donahue M M Roach B J Downs J J Blevins T Baldea M and Eldridge R B 2016 Dividing wall column control: common practices and key findings *Chemical Engineering and Processing: Process Intensification* 107 106-15

[4] Dang P X and Banjerdpongchai D 2011 Design of integrated real-time optimization and model predictive control for distillation column *2011 8th Asian Control Conference, ASCC* 988-93

[5] Giessler S, Danilov R Y, Pisarenko R Y, Serafimov L A, Hasebe S and Hashimoto I 1999 Feasible separation modes for various reactive distillation systems. *Industrial & engineering chemistry research* **38(10)** 4060-7

[6] Bachnas A A, Tóth R, Ludlage J H A and Mesbah A 2014 A review on data-driven linear parameter-varying modeling approaches: A high-purity distillation column case study *Journal of Process Control* **24(4)** 272-85

[7] Quirante N, Javaloyes J and Caballero J A 2015 Rigorous design of distillation columns using surrogate models based on K riging interpolation *AIChE Journal* **61(7)** 2169-87

[8] Sliskovic D, Grbić, R and Hocenski Ž 2011 Methods for plant data-based process modeling in soft-sensor development *Automatika* **52(4)** 306-18

[9] Giwa A and Karacan S 2012 Modeling and simulation of a reactive packed distillation column using delayed neural networks *Chaotic Modeling and Simulation* ***2*****(1)** 101-8

[10] Weber G W, Batmaz İ, Köksal G, Taylan P and Yerlikaya-Özkurt F 2012 CMARS: a new contribution to nonparametric regression with multivariate adaptive regression splines supported by continuous optimization. *Inverse Problems in Science and Engineering* **20(3)** 371-400

[11] Langdon W B and Poli R 2013 *Foundations of genetic programming* (Springer Science & Business Media)

[12] Breiman L 2017 *Classification and regression trees* (Routledge)

[13] Milov A V, Tynchenko V S, Kukartsev V V, Tynchenko V V and Bukhtoyarov V V 2018 Use of artificial neural networks to correct non-standard errors of measuring instruments when creating integral joints *J. phys.: conf. ser.* **1118** 012037

[14] Whitley D 1994 A genetic algorithm tutorial *Statistics and computing* **4(2)** 65-85

[15] Buraczewski D and Kolesko K 2014. Linear stochastic equations in the critical case *Journal of Difference Equations and Applications* **20(2)** 188-209