INVESTIGATION OF METHODS FOR MODELING PETROLEUM REFINING FACILITIES TO IMPROVE THE RELIABILITY OF PREDICTIVE DECISION MODELS

Vladimir Bukhtoyarov* Vadim Tynchenko Eduard Petrovskiy Natalia Bukhtoyarova Vadim Zhukov
Siberian Federal University, Krasnoyarsk, Russia

The current state of production systems of the oil and gas sector makes high demands on the reliability of decision making at the operational level of process facilities control. In many situations, petroleum refining requires decision support based on predictive models. Such models should be accurate and computationally effective, which makes high demands on the selection of effective methods for constructing models of oil refinery facilities, in particular, rectification columns. The purpose of the research is to provide such requirements, the authors investigate several methods for constructing models of process technologies (facilities), estimating their accuracy based on the data of the actual rectification process technology. On average, multivariate adaptive regression splines were most effective results obtained on the set of model parameters. This method, using the samples of observations considered in the paper, allows building models with an average (referring to a set of simulated parameters) simulation error of 8.2%. The results indicate that it is possible to replace "conventional" models with fast regression models. Calculation and parameter prediction for technological facilities, including for generating control actions, using such models is possible in real time mode.

Key words: Modeling, Petroleum refining, Oil columns, Parameter prediction, Reliability of decision-making, Artificial neural networks, Oils, Models

INTRODUCTION

The problem of modeling the process equipment for oil refineries in general, and in particular complex rectification columns for separation of oil mixtures, is relevant in view of the need to ensure the safety of processes in terms of equipment control safety and parameter prediction [01, 02]. The safety and reliability of the process system operation is largely determined in this case by the parameters of the models used for analysis and decision making, and therefore they directly depend on the effectiveness of the application of modeling methods. The effectiveness of such approaches is assessed by several criteria, the most significant of which is the modeling accuracy and, accordingly, the modeling error, which should be minimized [03].

Approaches to the modeling of petroleum refining facilities are largely based on the use of models that relate to the informal, physicochemical and thermodynamic ones [04, 05, 06]. For the petroleum refining equipment and in particular complicated rectification columns, static regimes are usually modelled, although a number of studies on the modeling of transient regimes are known [07, 08, 09].

The analysis shows that most of the proposed models of this type are not able to satisfy the requirements set for the models used for operational control. The following three properties are essential for such models that do not allow them to be used as quick calculation models for operational control of petroleum refining facilities:

- The models do not allow predicting the parameter values of process units by the measured process parameters;
- The time required for calculations, even on modern computers, exceeds the decision-making time for the operational control of a complex process unit;
- The initial information for such models often cannot be generated quickly – for example, if calculations require knowledge of the feedstock parameters and composition.

Further the article deals with the problem of modeling of vacuum columns, several methods for constructing models of process technologies (facilities) are briefly described. The efficiency estimation results are given for modeling methods based on the data of the real rectification process taking into account the modeling error criterion. To obtain numerical estimates of the modeling methods effectiveness, a scheme used for the repeatable construction of models with an evaluation of the significance of the obtained criteria by the ANOVA method. In the corresponding sections, the obtained numerical results are analyzed and at the end conclusions are formulated according to the results of the conducted research.

In the paper a new approach to constructing operational control models based on automated approaches to generation of models using intelligent information technologies is proposed. A complex numerical analysis of the effectiveness of the application of methods of data mining for the construction of operational control models

Vladimir Bukhtoyarov - Investigation of methods for modeling petroleum refining facilities to improve the reliability of predictive decision models

iipp

for technological equipment of oil refining is carried out. The statistical analysis of the results is performed, the results are verified for the adequacy of the technological parameters of the simulated equipment under the given simulation conditions. The software implementation of the methods of data mining is implemented.

MODELING OF PROCESS FACILITIES OF PETROLEUM REFINING PRODUCTION SYSTEMS

The construction of models for determining the parameters of vacuum columns and obtaining numerical estimates of the corresponding parameters would make it possible to determine the parameters necessary for a rational selection of the control parameter values for such process systems. The existing methods and models require the use of a considerable amount of information on indirect parameters and the accumulation of a reference database [10, 11, 12].

In addition, with regard to the application of microcontrollers and computers with significantly limited performance at automated workstations, it is not practical to operate high-performance computers in the appropriate conditions. It seems justified to use simplified, 'fast' approaches to the model calculation of these indicators with respect to the products of oil mixtures separated in vacuum columns. Obviously, to efficiently calculate the values of the parameters necessary for controlling the process technologies within the petroleum refining production systems, it is required to construct models satisfying the criteria of simplicity (the performance of a standard personal computer or the computing power of modern microcontrollers is sufficient for these computations) and the accuracy of simulated parameters. To construct such models, it is proposed to use regression methods [13, 14]. Defining the methods that provide the greatest accuracy in predicting the model parameters necessary for assessing the state of process technologies and evaluating the control actions for the relevant process facilities is a topical task.

The main parameters of the columns used to control petroleum refining and petro chemistry include [15, 16]:

1. Boiling temperatures at a given value of distillation α, including:
   - the bubble point value;
   - the 5% final boiling point value and other characteristic points of the true boiling temperature.
2. The viscosity, estimated by the dynamic μ or kinematic ν viscosity coefficients.
3. The density
4. The flash point
5. The saturated vapor pressure
6. The color
7. The pour point
8. The crystallization temperature
9. The composition, concentration, etc.

There are two main approaches to obtaining information about the main parameters. The first, which can be called a 'measuring' one, involves the use of appropriate analyzers; the second ('modeling' one) is based on the use of models that employ indirect parameters as input coordinates [15, 17].

It is essential that a number of basic parameters of the process flow in the columns are not provided with measuring instruments. There is another significant drawback of such a measuring approach: parameters are measured for products already obtained. Due to this, and also in view of the dynamic characteristics of objects (process petroleum refining facilities), whose time constants make often tens of minutes, standard regulation methods do not provide acceptable quality of regulatory processes [04, 07, 09].

Therefore, the approach to obtaining estimates of the main parameters of process facilities, in particular rectification columns based on fast models, seems most effective. In addition, in many cases such an approach is only possible in decision-making at the operational level of control.

Taking into account the above and the need to improve the reliability of decision-making at the operational level of controlling petroleum refining facilities, the following requirements for process parameter assessing models seem justified:

1. The model provides the possibility of making calculations for the time during which the parameters of process facilities change insignificantly; usually this time is on the order of a few seconds or tens of seconds;
2. The models should be simple enough so that they can be implemented on production hardware;
3. The applied modeling method must contain mechanisms for adapting to the characteristics of a particular facility, raw material, process flow pattern;
4. The model should use process parameters and other information as input coordinates that can be obtained promptly by measurements based on standard technical aids;

These requirements impose very strong limitations on the class of possible mathematical models [15].

MATERIALS AND METHODS

To study the effectiveness of the modeling methods under consideration, the statistical approach is applied comprehensively; the significance of the model input parameters is investigated. To obtain numerical estimates of the effectiveness of modeling methods, a scheme for the repeatable construction of models was used and the significance of the obtained criteria was evaluated by the ANOVA method.

Formal, in particular, regression relationships can be used as "fast" static models. In this case it is required to resolve a number of issues:

1. To determine the model structure; i.e., input coordinates, the form of the regression equation and the method of its identification;
2. To determine the parameters of the model, i.e. the method of obtaining parameters, source and volume of initial information, criterion of the regression model adequacy.

There are a significant number of methods for constructing regression models of process facilities. Within the framework of this article, the efficiency of the following methods for constructing regression models is considered and analyzed:

1. Artificial neural networks [18, 19];
2. Multivariate adaptive regression splines [20, 21];
3. Regression trees [22];
4. Genetic programming method [23];
5. Non-parametric regression models (Hart, 2013).

The linear regression method was considered as a basic method for assessing the efficiency of models constructed by the methods listed above. The selection of these methods as the basis for constructing models of process facilities is determined by the following factors:

1. Based on the results of the analysis of the review articles of the IEEE Computational Intelligence Society (IEEE CEC), the selected methods are among the five methods whose application for solving modeling and classification problems is described in a significant number of articles published as a result of conferences conducted with the participation of IEEE CEC [25].
2. There are numerous examples of the successfully employed methods in question for identifying and constructing regression models of process facilities.
3. The implementation of all the methods under consideration within the framework of a single program-analytical approach will make it possible to realize the idea of a synergistic unification of diverse information technologies for constructing effective collective models of petrochemical process facilities.

ARTIFICIAL NEURAL NETWORKS

Artificial neural networks can be used to solve a wide range of practical problems, which can be formulated mathematically as the problem of function approximation, pattern classification and recognition, prediction, and control of dynamic processes [26].

In approximation problems, a neural network plays the role of a universal approximator of a multivariable function. A large number of modeling, identification and signal processing problems are formulated precisely in the approximation setting.

When solving prediction problems, neural networks are used to predict the behavior of the system from the available observations of its previous states. When solving pattern classification and recognition problems in the

learning process, a neural network accumulates information about the basic properties of these patterns. During generalization the differences between the patterns from each other are emphasized, which is the basis for the formation of classification solutions.

In tasks of controlling dynamic processes, a neural network is used to perform several functions. The neural network can act as a nonlinear process model and identify its main parameters used in the development of the appropriate control signal. The network can also play the role of a tracking system that monitors environmental conditions and adapts to them. It can also function as a neural regulator replacing conventional control devices [18].

In this article, neural networks are considered in the aspect of modeling (function approximation) and prediction problems, the neural network being treated as a "black box" in accordance with the classical approach. Accordingly, the goal of solving these problems is to obtain an effective computational procedure, rather than to extract knowledge and hidden regularities in an explicit form. It should be noted that although there are approaches to constructing so-called "logically transparent" artificial neural networks, a large range of problems can be effectively solved within the framework of the classical approach that regards a neural network as a "black box". The authors will consider the mathematical formulations of modeling and prediction problems, as well as their possible interpretations in terms of using artificial neural networks for their solution within the chosen approach.

MULTIVARIATE ADAPTIVE REGRESSION SPLINES

Multivariate adaptive regression splines (MAP-splines) were first proposed to solve the regression and classification problems which require predicting the values of a set of dependent variables by a set of independent variables [27]. This method is a nonparametric procedure that does not use any assumptions about the form of the functional dependence between dependent and independent variables in its work. MAP-splines are set by basic functions and a set of coefficients that are completely determined by the data. MAP-splines find their application in many fields of science and technology, for example, in predicting the types of distributions from the available data of intestinal drug absorption, as well as in reproducing speech and searching for a global optimum in structural design. MAP-splines method finds the required dependence in two stages: forward stage and backward stage. The first stage involves adding the basic functions to the set, until the maximum level of complexity is reached. At the second stage, the functions that make the smallest contribution to the error are removed from the set.

REGRESSION AND CLASSIFICATION TREES

The decision tree method is one of the most popular methods for solving classification and prediction prob-

Vladimir Bukhtoyarov - Investigation of methods for modeling petroleum refining facilities to improve the reliability of predictive decision models

iipp

lems. Also this approach is called the decision rule method, the classification and regression trees.

As can be seen from the latter term, this method helps solving the classification and prediction problems. If dependent, i.e. the target variable takes discrete values, the classification problem is solved using the decision tree method. If the dependent variable takes continuous values, then the decision tree establishes the dependence of this variable on the independent variables, i.e. solves the problem of numerical prediction. Attributes of a data set can have both a discrete and a numerical value. A detailed description can be found in [22].

GENETIC PROGRAMMING METHOD FOR SYMBOLIC REGRESSION PROBLEMS

The main idea underlying the genetic programming method is to obtain a dependency that describes well the data presented in the initial samples using the software implementation of procedures imitating evolutionary processes from a randomly generated set of functional dependencies [23].

Unlike genetic algorithms in which individuals of a population are symbol strings (in particular binary ones) encoding a solution, decisions in genetic programming can be presented in various forms and have different dimensions. The most commonly used representation of solutions is in the form of trees. A tree is a directed graph in which each successive vertex is connected with one and only one previous vertex. The set of all possible internal vertices of a tree is called a functional set. The set of all possible external vertices of a tree is called a terminal set. Operations and functions are usually selected as elements of a functional set. Input data – variables and constants – are considered to be the elements of the terminal set.

NON-PARAMETRIC IDENTIFICATION

Parametric identification implies the presence of a fixed structure of the system model. The selection of a particular structure is usually based on a priori information about the processes taking place in the system under consideration, or on the successful application of this type of structures while modeling such systems. The quality of modeling in this case depends on the "correctness" of the a priori selection of the structure. However, in many cases, the a priori information about the process, facility, or system is practically absent, and there are no grounds providing for unambiguous estimation of the model structure. Under these conditions, parametric identification methods are not applicable.

In these cases, it is possible to apply techniques that are based on stochastic approximations of a nonparametric type – nonparametric identification methods [24]. Nonparametric identification methods allow for rather good solution of the problem of reconstructing the dependencies approximating the data in the initial sample of observations over the facility. Nonparametric methods

for reconstructing the function from observations are not related to the a priori selection of the approximating function form and refer to methods of local approximation. Non-parametric estimates converge to the desired probabilistic characteristics, for example, to the regression curve, in contrast to the parametric approximations, where it is a matter of convergence of parameters in the accepted structures.

As in most numerical methods of identification, with a non-parametric approach, the regression models are used to approximate experimental (statistical) data. The nonparametric estimation of the regression using the Rosenblatt-Parzen estimator is as follows:

$$y_s(x) = \frac{\sum_{i=1}^{s} y_i \Phi\left(\frac{x - x_i}{C_s}\right)}{\sum_{i=1}^{s} \Phi\left(\frac{x - x_i}{C_s}\right)} \qquad 1)$$

The kernel regression estimate 1) is called the Nadaraya-Watson estimator. Theorems on the asymptotic properties of this estimate are proved. The basic idea underlying 1) is to give a relatively greater weight to the observations closest to the estimated point, in terms of the distance determined by the kernel . Estimate 1) can be obtained formally as well – by replacing the probabilistic characteristics in the expression of the conditional mathematical expectation with their kernel counterparts.

REDUCING SPACE OF ATTRIBUTES. DETERMINING SIGNIFICANT FACTORS

Prior to construct the model, it is necessary to determine the input variables for which it is statistically significant to assert about their significant impact on the modelled process parameters. In this regard, before the direct construction of the model with the help of statistical analysis methods, significant factors – the variables that influence most significantly on the modelled parameter – were selected for each modelled technological parameter from the whole set of input parameters,. The statistical analysis was carried out using a statistics and analytics software package, Statsoft Statistica.

The results obtained in the framework of this section of the research represent an important scientific result, since it was not possible to detect similar results of statistical analysis in open sources for the process facilities in question. In the studies devoted to this subject no attention has been given to this step, which is of great importance from the view point of constructing and optimizing the models of the process facilities under consideration. The results of determining the significant factors – input variables – are given for each modelled parameter below. The results were obtained with a confidence level of 0.9.

REGRESSION MODEL CONSTRUCTION

Specialized software tools were used to construct the models for computing process parameters of the considered process unit.

- The Statsoft Statistica software package was applied to synthesize models based on artificial neural networks, regression trees and multivariate regression splines
- IT-Pegasus software system was used to synthesize models using nonparametric regression estimate and genetic programming methods.

PARAMETRIC OPTIMIZATION OF THE MODELS

A genetic algorithm and considered parametric optimization algorithms of Litvakov and Kesten were used for the parametric optimization of models constructed by different methods for synthesizing regression models [28, 29]. To determine the set of optimization parameters, each synthesized model was considered, and numerical parameters were determined, the adjustment of which is possible within the framework of the constructed model structure.

Since the genetic algorithm considered as the main algorithm for parametric optimization of control models requires the representation of solutions in the form of binary strings, the model parameters were discretized, and the discretization step was selected not exceeding 0.001. The interval for varying the variables was determined individually for each model and was chosen proceeding from the considerations of complete overlapping of the initial values of the parameters generated by the direct identification algorithm.

The criterion determined by formula 2 was used as a criterion of optimization. The parameters of the applied optimization algorithms were determined during the pretest on a set of test functions, which includes 25 test tasks, most often used in studies in this direction. Parameters influencing the duration of the parameter adjustment were selected as follows: the total number of search steps does not exceed 100, the number of decision search points at each step amounts to 100.

MODEL EVALUATION CRITERIA

Estimation of the mathematical expectation and modelling error variance estimate calculated from the data obtained during the 50 independent runs of the algorithms were used as the main efficiency criterion. To calculate the approximation error, the following formula was used in each run:

$$Error = \frac{100\%}{s\left(y^{\max} - y^{\min}\right)} \sum_{i=1}^{s} \left| o_i - y_i \right| \qquad 2)$$

Statsoft Statistica software package was used to estimate indicators.

RESULTS

In view of the need to depersonalize the research results, taking into account the confidentiality of information used as input data for the construction of regression models, the corresponding parameters of the modelled process facility were depersonalized and represented in the tables and on the graphs as the numbers of the simulated parameters of the process unit.

Table 1 gives the results of a statistical study of the considered methods for modeling the column parameters without using special methods for adjusting the parameters of regression models.

Table 1 gives the results of a statistical study of the considered methods for modeling the column parameters with the use of special methods for adjusting the parameters of regression models.

DISCUSSION

The results obtained in the course of numerical studies were processed using statistical methods to assess their significance. The ANOVA method of variance analysis was used for verification by all the modelled parameters of the process facility under consideration.

The data were obtained during the construction of regression models for calculating and predicting the parameters of a column-type apparatus. To evaluate the efficiency of using different types of models, the results were analyzed both separately for each considered parameter and, on average, for the set of parameters considered. The results obtained for the set of modelled parameters are given on average in table 3.

Thus, when considering all modelled parameters as a whole, for the majority of parameters (namely, for nine of the modelled 11 parameters), the best results were obtained using the multivariate regression splines method. The mean modelling error for regression models obtained using this method without parametric optimization models was 8.2%. On average, for the regression models under consideration, the modeling error made 12.6%, which is a significant value, potentially limiting the possibility of their effective application for modeling of petrochemical production facilities and calculation of control actions and parameters for them. This indicates the need and rationality of using the additional procedures proposed in the article for the parametric adaptation of the resulted regression models, which can be considered as initial solutions for obtaining more efficient (in terms of accuracy criteria) regression models.

The obtained results and analysis of their statistical processing prove that the application of additional algorithms of parametric optimization models based on the evolutionary genetic algorithm enables to significantly reduce the model. This is achieved by means of the deep adaptation of model parameters through 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 problem is considerable, which makes it difficult to use classical optimization methods and the direct exhaustive search methods, the heuristic algorithm used for the optimization of regression models to calculate the parameters of process units has demonstrated high efficiency.

As can be seen from the results given in table 3, the application of the additional parametric optimization procedure allowed for a significant reduction in the modelling error by more than twice: from 15.6% to 7.2%. For the most effective method – that of regression splines – the modelling error was 2.8% after parametric optimization of the model. According to expert estimates of the relevant specialists, the accuracy of models is satisfactory after parametric optimization, and the corresponding models can be used to simulate and predict process flow

parameters of the facilities and, consequently, to calculate the levels of control parameters for such units.

It should be noted that parametric optimization enabled to significantly improve the accuracy for models obtained by all methods, and after the proposed parametric optimization even the worst models were more accurate in terms of the modeling accuracy criterion than the best method of multivariate regression splines without parametric optimization. This result can be used in the following way, especially considering that none of the methods proved to be more effective than the others for all problems – it is possible to apply any of the considered methods in combination with the genetic algorithm of parametric optimization models for constructing efficient regression models and for calculating the levels of control parameters of the process facilities.

Table 1: The Results of a Statistical Study of the Modeling Methods without Adjusting the Model Parameters

Modelling methods	Number of the modelled parameter of a process unit											
	19	20	21	22	23	24	25	26	27	28	29	30
	Mean relative error of modelling, %											
Artificial neural networks	14.4	12.3	10.3	10.5	12.5	11.3	13.9	10.9	9.8	11.4	9.3	17.5
Multivariate regression splines	5.9	5.5	7.3	8.2	9.1	6.2	7.9	9.6	12.6	6.9	6.4	8.6
Decision trees	18.1	14.1	15.6	19.4	17.2	20.3	15.6	18.6	20.1	18.1	18.0	19.2
Genetic programming	14.8	15.3	16.7	14.1	17.1	11.6	17.4	11.8	15.7	17.1	21.7	11.2
Non-parametric regression	11.3	15.2	12.1	11.2	13.7	10.9	11.4	11.5	14.4	13.4	13.8	9.9

Table 2: The Results of a Statistical Study of the Modeling Methods with Adjusting the Model Parameters

Modelling methods	Number of the modelled parameter of a process unit											
	19	20	21	22	23	24	25	26	27	28	29	30
	Mean relative error of modelling, %											
Artificial neural networks	5.5	4.4	5.2	5.1	4.4	5.8	5.2	7.9	6.2	5.7	5.4	5.8
Multivariate regression splines	3.9	5.4	5.1	4.9	2.7	4.4	4.2	3.2	5.9	2.2	4.9	2.9
Decision trees	11.1	7.8	7.5	8.7	9.1	9.6	8.6	9.6	8.7	8.4	8.3	8.7
Genetic programming	8.7	9.1	8.9	8.6	10.0	8.6	8.1	6.7	8.7	8.79	8.5	7.5
Non-parametric regression	8.9	7.5	7.2	7.1	6.6	6.9	6.5	6.6	6.8	6.13	7.2	7.4

Table 3: The Estimates of the Accuracy of the Regression Model Construction by All the Modelled Parameters

Method	Criteria for estimating the accuracy of the regression model construction without/with parameter optimization	
	Mean relative error of modelling, %	
Artificial neural networks	11.9	6.1
Multivariate regression splines	8.2	3.8
Decision trees	17.8	8.9
Genetic programming	15.6	7.8
Non-parametric regression	13.4	7.6

CONCLUSION

Numerical experiments were conducted in the course of the research to evaluate the efficiency of methods for constructing regression models of petrochemical process facilities (as exemplified by a rectification column). To construct models and evaluate their efficiency (in terms of the model accuracy criterion), samples of observations over the process parameters were used.

To obtain statistically significant results, the authors used the approach based on multiple cross-checking, which is widely used in studying the methods for constructing regression models. The obtained statistical estimates of the accuracy criterion of the regression model construction were verified to confirm the obtained statistically significant and statistically different results using ANOVA methods.

The findings show that, in general, the methods considered above allow constructing regression models that enable to calculate the output parameters of rectification columns relatively quickly for evaluation of the influence of input factors and the calculation of control values on the basis of these parameters. However, as shown by numerical studies, the accuracy of models constructed by different methods differs. On the average, the method of multivariate adaptive splines was most effective for the set of modeled parameters. This method, using the observational samples considered in the work, makes it possible to construct models with a mean modeling error of 8.2%.

To increase the accuracy of the models obtained, it was proposed to use the evolutionary optimization algorithms, which enabled to effectively adjust the coefficients of regression models in the context of the optimization problem formulation. Further, in the course of numerical studies, the efficiency of optimizing the parameters of models of process units was evaluated using the data set considered in the article. The findings testify that the parametric optimization models allowed reducing the modeling error by more than twice, on average by 7.2%. For the multivariate regression splines method, which showed the best results, the mean modeling error was 2.8%, as in the case when parametric optimization was not applied.

The obtained results testify to the possibility of replacing "conventional" models with fast regression models. Calculation and parameter prediction of process facilities, including for generating control actions, using such models is possible in the real time mode. This does not require the use of specialized high-performance computers, as it seems inexpedient to equip automated workplaces of operators of process units with them.

The direction of the further development of the research is the development of a universal platform for designing of operational control models for oil refining equipment based on intelligent data analysis technologies. Thy study presented here demonstrates the applicability and effectiveness of this approach. The implementation of such a platform in the form of an intelligent analytical subsystem of enterprise information systems will increase the effectiveness of decision support for the management of the park of process equipment at all levels of enterprise management.

The limitation of this approach is the need for the availability of initial samples of observations of operational parameters of process equipment. The data in such samples should be verified and associated with a particular technological installation. Obviously, such samples of observations are relatively easily accessible for industries with a high degree of informatization of technological processes. However in some situations the necessary amount of information may not yet be accumulated for newly introduced equipment. In this case, the design of such models can be difficult, and one of the ways to solve this problem can be the adaptation of data from other similar technological installations of companies or the use of industry data banks.

REFERENCES

1. Qingfeng, W., Wenbin, L., Xin, Z., Jianfeng, Y., Qingbin, Y. (2011). Development and application of equipment maintenance and safety integrity management system. Journal of Loss Prevention in the Process Industries, vol. 24, 321-332.

2. Bertolini, M., Bevilacqua, M., Ciarapica, F.E., Giacchetta, G. (2009). Development of risk-based inspection and maintenance procedures for an oil refinery. Journal of Loss Prevention in the Process Industries, vol. 22, 244-253.

3. Wolf, F.G. (2001). Operationalizing and testing normal accident theory in petrochemical plants and refineries. Production and Operations Management, vol. 10, 292-305.

4. Ancheyta, J. (2011). Modeling and simulation of catalytic reactors for petroleum refining. John Wiley & Sons.

5. Rivero, R., Garcia, M., Urquiza, J. (2004). Simulation, exergy analysis and application of diabatic distillation to a tertiary amyl methyl ether production unit of a crude oil refinery. Energy, vol. 29, 467-489.

6. Speight, J. G. (2014). The chemistry and technology of petroleum. CRC press.

7. Dimian, A. C., Bildea, C. S., & Kiss, A. A. (2014). Integrated design and simulation of chemical processes (Vol. 13). Elsevier.

8. Ansari, R.M., Tadé, M.O. (2012). Nonlinear model-based process control: applications in petroleum refining. Springer Science & Business Media.

9. Sildir, H., Arkun, Y., Canan, U., Celebi, S., Karani, U., & Er, I. (2015). Dynamic modeling and optimization of an industrial fluid catalytic cracker. Journal of Process Control, 31, 30-44.

10. Espada, J.J., Coto, B., van Grieken, R., Moreno, J.M. (2008). Simulation of pilot-plant extraction experiments to reduce the aromatic content from lubricating oils. Chemical Engineering and Processing: Process Intensification, 47: 1398-1403.

11. Nimmagadda, S.L., Rudra, A. (2017). Visualization as a Big Data Artefact for Knowledge Interpretation of Digital Petroleum Ecosystems. International SERIES on Information Systems and Management in Creative eMedia (CreMedia), vol. 2016, no. 2, 34-43.

12. He, S., Lucio-Vega, J., Zhang, L., Shi, Q., Horton, S.R., Al-Khattaf, S.S.F, Klein, M.T. (2017). Integrating Visualization Methods with Chemical Kinetics Model Solution and Editing Tools. Energy Fuels, vol. 31, 9881–9889.

13. Chatterjee, S., & Hadi, A. S. (2015). Regression analysis by example. John Wiley & Sons.

14. Draper, N. R., & Smith, H. (2014). Applied regression analysis (Vol. 326). John Wiley & Sons.

15. Dimian, A.C., Bildea, C.S., Kiss, A.A. (2014). Integrated design and simulation of chemical processes. Elsevier.

16. Ramzan, N., Naveed, S., Muneeb, R., Tahir, F.M. (2013). Simulation of natural gas processing plant for bumpless shift. NFC IEFR Journal of Engineering and Scientific Research, vol. 1.

17. Chaudhuri, U.R. (2010). Fundamentals of petroleum and petrochemical engineering. CRC Press.

18. Demuth, H.B., Beale, M.H., De Jess, O., Hagan, M.T. (2014). Neural network design. Martin Hagan.

19. Ochoa-Estopier, L.M., Jobson, M., Smith, R. (2013). Operational optimization of crude oil distillation systems using artificial neural networks. Computers & Chemical Engineering, vol. 59, 178-185.

20. Zhang, W., & Goh, A. T. (2016). Multivariate adaptive regression splines and neural network models for prediction of pile drivability. Geoscience Frontiers, 7(1), 45-52.

21. Milborrow, S., Hastie, T., Tibshirani, R., Miller, A., & Lumley, T. (2015). earth: Multivariate adaptive regression splines. R package version, 4(0).

22. Breiman, L. (2017). Classification and regression trees. Routledge.

23. Langdon, W., Poli, R. (2013). Foundations of genetic programming. Springer Science & Business Media.

24. Hart, J. (2013). Nonparametric smoothing and lack-of-fit tests. Springer Science & Business Media.

25. Li, X., Engelbrecht, A., Epitropakis, M.G. (2013). Benchmark functions for CEC'2013 special session and competition on niching methods for multimodal function optimization. RMIT University, Evolutionary Computation and Machine Learning Group, Australia, Tech. Rep.

26. Ge, S.S., Hang, C.C., Lee, T.H., Zhang, T. (2013). Stable adaptive neural network control. Springer Science & Business Media.

27. Kisi, O. (2015). Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree. Journal of Hydrology, 528, 312-320.

28. Whitley, D. (2014). An executable model of a simple genetic algorithm. Foundations of Genetic Algorithms, 2: 45-62.

29. Fu, M. C. (2016). Handbook of simulation optimization. Springer.