Using Mathematical Modeling to Generate Training Data in Hydrotreating Processes

M. I. Babidorich, A.M. Demin, O.A. Reutova, E.V. Docenco

***Abstract*—The research presents a method for applying a static mathematical model in the process of generating a database for training an artificial neural network. The study was carried out on the example of predicting the physicochemical properties of a model of multicomponent mixture of diesel fuel and hydrogen-containing gas. Data generation is carried out by enumeration of mutable variables in the range of valid values. The following process parameters are chosen as variables: temperature, pressure and flow rate. It was found that changes in the chemical composition of the flows do not affect the result. Diesel fuel initial boiling point and the boiling point of 50%, 90% and 95% of the fraction changes on average by 1% ÷ 4%. Deviations of the calculated physical and chemical characteristics as a result of fluctuations in the content of hydrogen bearing gas components do not exceed 1.1%. As a result, a neural network was obtained, which determines the desired values with an error of 0.2%. This will allow the use of a neural network in dynamic systems for assessing process equipment fouling.**

***Keywords*— deep learning, heat exchangers fouling, hydrotreating.**

I. Introduction

Fouling of heat exchangers is one of the most common main faults, resulting in permanent and recurring failures of equipment at refinery process units. Deposits increase heat resistance, while the temperature of the flows is not maintained at the required level, which reduces the thermal rating of the heat exchanger, and, consequently, its efficiency: layers 2-4 mm thick on the inner surfaces of heat-releasing devices reduce the efficiency by 4-7%.

The morphology of deposits/fouling in the form of a continuous layer on the surface is especially dangerous: thermal energy will be collected under the layer locally, which can lead to an increase in the corrosion rate, pitting of the heat exchanger and its subsequent failure with possible personal injury. Thus, the relevance of timely clarification of the causes of pollution is understandable for optimizing energy resources and ensuring the reliability of equipment, as well as for the safety of the industrial process [1, 2].

Manuscript received February 13, 2023.

M. I. Babidorich is with the Omsk State Technical University (e-mail: babidorichmaksim@gmail.com).
A.M. Demin is with the Omsk State Technical University (e-mail: demin_a_m@mail.ru).
O.A. Reutova is with the Omsk State Technical University (e-mail: ReutovaOA@omsu.ru).
E.V. Docenco is with the Omsk State Technical University (e-mail: iuneva.nova@gmail.com).

Examples of successful methods for predicting the state of heat exchange equipment using ma-chine learning methods are described, most of which are based on comparing the actual (current) heat transfer coefficient in the equipment with the same parameter at the initial time [1, 3 - 5]. This approach is not always applicable in oil refining due to the complexity of calculating the heat transfer coefficient for multicomponent hydrocarbon mixtures with an unknown or dynamically changing composition.

II. Description of the Process Under Study

One of the most large-scale refining processes at oil refineries (ORs) is the hydrotreatment of diesel fractions, designed to remove sulfur, nitrogen, and oxygen-containing organic compounds by their destructive hydrogenation [6]. The process scheme provides the heat exchangers for heating the feed gas mixture (FGM) at the inlet to the hydrotreating reactor due to the heat recovery of the product streams. FGM is a mixture of hydrogen rich gas (HRG) and diesel fuel (DF) with a dynamically changing ratio of these components (Fig. 1).

The most active deposition of contaminants for feed heat exchangers is characteristic of the FGM (outside the tubes). Fig. 2 shows the opened heat exchangers of the diesel fuel hydrotreater after a long period of operation.

This nature of the formation of deposits suggests that it is necessary to deal with the issue of predicting the state of the apparatus from the side of the gas-raw mixture. A mathematical model in the Aspen HYSYS environment for a multicomponent hydrocarbon mixture makes it possible to calculate its physicochemical characteristics, including heat con-tent/enthalpy. For dynamic flows in the online monitoring system of the real state of the equipment, regular references to such a model are inconvenient, since the influence of the composition and temperature of the flows on the rate of formation of deposits on the heat exchange surfaces is unknown, which in turn affects the temperature conditions of the section of equipment.

III. Mathematical Modeling

The HYSYS model allows you to calculate an array of options with different combinations of initial information within a given interval, which will make it possible to train an artificial neural network (ANN) based on input parameters to predict the desired dependence on the obtained database (DB).

The specified input parameters include the instruments data with which the heat exchange unit is equipped (Table 1).

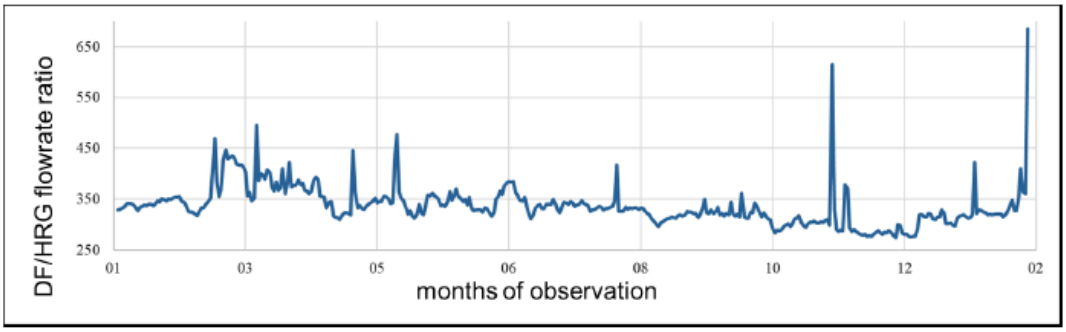

Fig. 1 Monitoring of hydrogen rich gas/diesel fuel flow rates ratio at the hydrotreating unit

Table 1 – Initial data

Flow	Parameter	Units of measure	Legend
DF	Temperature	°C	T_{DF}
	Pressure	kg/cm^2	P_{DF}
	Flow rate	m^3/h	F_{DF}
HRG	Temperature	°C	T_{HRG}
	Pressure	kg/cm^2	P_{HRG}
	Flow rate	m^3/h	F_{HRG}
Mixture	Temperature input	°C	$T_{M(IN)}$
	Temperature output	°C	$T_{M(OUT)}$

Fig. 2 Deposits on the surface of heat exchanger tubes

The mathematical model in the Aspen HYSYS environment (Fig. 3) is represented by the "Mixer" operator, in which two material flows (HRG + DF) are mixed and then pass through auxiliary heat exchangers E-101 and E-102. The mixture is cooled to 20°C in the first apparatus, and heated to $T_{M(IN)}$ in the next one. These devices have no real analogues and are needed only for the correct operation of the model. The main calculations take place in exchanger E-103. The apparatus simulates the flow heating from $T_{M(IN)}$ to $T_{M(OUT)}$. Heat flow Q-103 is the energy expended on heating the flow. This is the searching value.

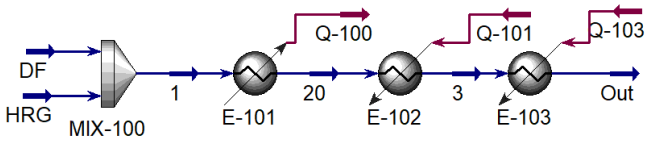

Fig. 3 Mathematical HYSYS-model of the heat exchangers section

In addition to process mode data, it is important to correctly set the chemical composition of the input streams. The composition of diesel fuel in production laboratories is estimated by fractional distillation. And for WSG, the component chemical composition is determined. Tables 2 and 3 show data on changes in the composition of flows during the operation of the diesel fuel hydrotreater for 2020-2021.

For simulation of DF, averaged values were chosen, since they fluctuate within acceptable limits (deviation less than 5%). But averaging the composition of the HSG can lead to deviations in the calculations of the physicochemical characteristics of the mixed flow.

Therefore, it is necessary to determine the deviations of the calculated indicators of physical and chemical properties between the flow with the actual composition of the HSG and the averaged ones. Deviations were calculated on 100 random samples. The HYSYS model was entered with the actual composition of the HSG and the data of the technological regime at the time of sampling. And the curve q = f(T) was determined (green line in Fig. 4). Further, for the same technological regime, the "average" composition of the HSG was used and the dependence q = f(T) was also determined (orange dotted line in Fig. 4).

The result of the deviation of the enthalpy for all experiments:

- MAE = 0,7326
- MAPE = 1,0942%
- R2 = 0, 9968

The low deviation makes it possible to ignore fluctuations in the HSG composition and use averaged values in modeling this flow.

IV. Data Generation and ANN Training

To calculate the training database (DB), the Case Studies tool was used, in which the values of dependent variables (heat flow Q-103) are calculated for given independent variables (input parameters of the process mode). Each independent variable was changed by a fixed step in the range of working values with some margin. In total, two different databases were generated: for training and monitoring the accuracy of the NS operation (Table 4).

Table 2 – Change in diesel fuel composition

Output	Boiling point, °C				
	Average for 2020-2021	Minimum	Maximum	Standard deviation	Relative standard deviation
start of boiling	180	155	196	7.14	3.97
50%	281	216	290	6.71	2.39
90%	333	279	344	6.00	1.80
95%	347	330	360	3.67	1.06

Table 3 – Change in HRG composition

Component	Mass fraction				
	Average for 2020-2021	Minimum	Maximum	Standard deviation	Relative standard deviation
Hydrogen	33.24	24.52	62.33	4.17	12.55
Methane	20.07	0.11	29.34	4.29	21.36
Ethane	19.32	6.16	22.63	1.50	7.75
Propane	14.72	1.63	22.73	1.68	11.42
Butane	7.66	4.09	29.65	2.05	26.68
$\sum C_5$	2.86	0.00	7.64	1.26	43.98
$\sum C_6$	1.73	0.00	7.42	1.23	71.11

Table 4 – Definition of independent variables for the training/control database

Variables	Start	Finish	Step	Steps total
T_{DF}	20,0 / 22,0	60,0 / 60,0	10,000 / 5,7	5 / 7
P_{DF}	30,0 / 36,0	60,0 / 60,0	10,000 / 7,8	3 / 4
F_{DF}	30,0 / 33,0	140,0 / 140,0	20,00 / 17,8	6 / 7
T_{HRG}	30,0 / 32,7	80,0 / 80,0	10,000 / 9,5	6 / 6
P_{HRG}	20,0 / 23,4	80,0 / 80,0	10,000 / 9,3	6 / 7
F_{HRG}	15,00 / 16,80	100,00 / 100,00	5,00 / 4,2	18 / 3
M_{TIN}	30,0 / 32,0	210,0 / 210,0	20,0 / 37	10 / 5
M_{TOU}	60,0 / 63,0	250,0 / 240,0	20,0 / 37	10 / 5
Total				**5 832 000 / 3 087 000**

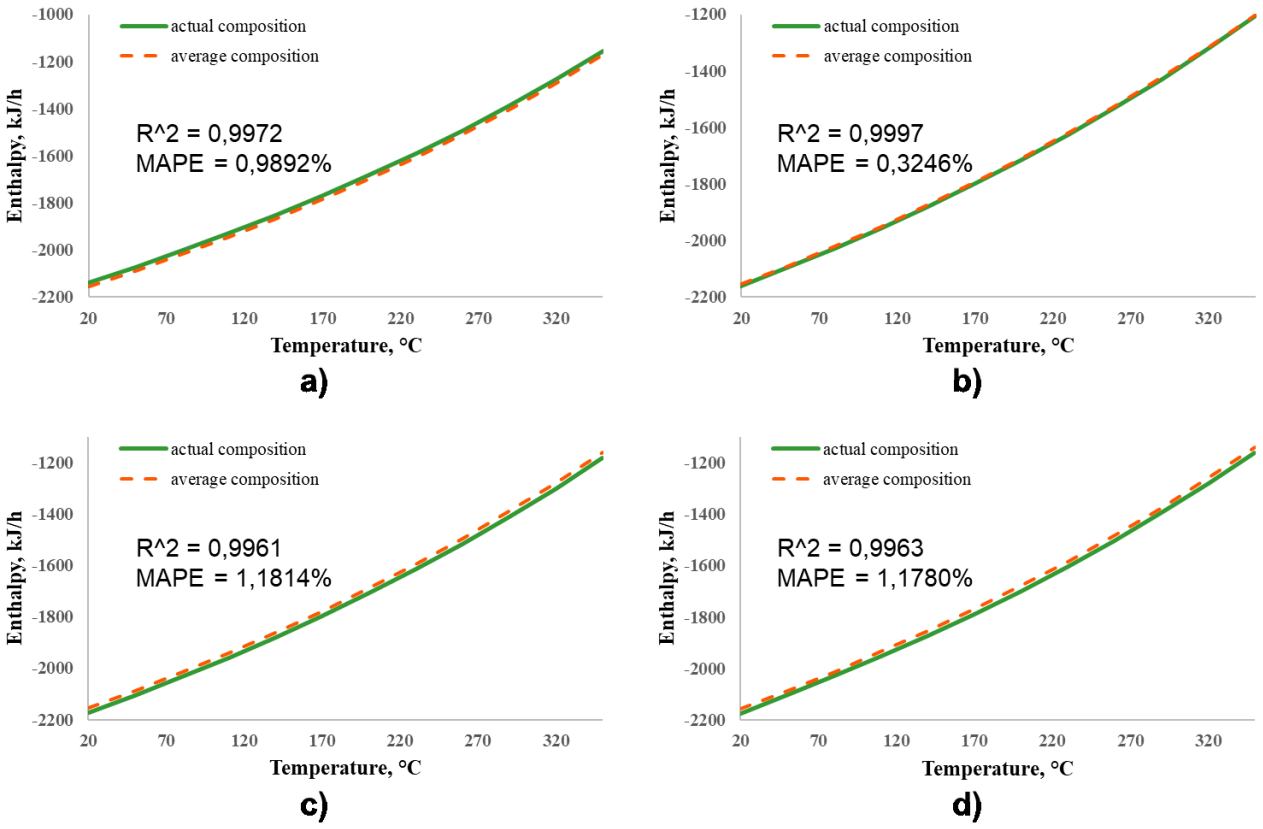

Fig. 4 Examples of enthalpy deviations for four random

As a result of the training database calculation, 5 832 000 sets of the desired physical characteristics of the mixed flow were obtained for various combinations of independent variables (Table 4). This data set is divided into two parts:

- training data (4 000 000 random combinations) - the sample required to train the ANN model;
- test data (the remaining 1 832 000 combinations) - a sample intended for an intermediate assessment of the work of the ANN during training.

The second set of the control database is used for the final assessment of the quality of the ANN after full training and contains the data that the ANN "did not see" earlier at the training stage.

The ANN architecture is represented by elements - an input layer, one hidden layer of 200 neurons and output layer from the TensorFlow open library [7]. The input layer receives 8 in-put signals that go to one hidden layer, the response from it is thermal load heat exchanger.

To assess the intermediate assessment of the quality of ANN training, the mean absolute error (MAE) is used as a loss function:

$$MAE = \sum_{i=1}^{N_S} \frac{|L_i - E_i|}{N_S} \quad (1)$$

where N_S – number of samples; Li – the actual value of the desired value for the i-th sample; Ei – forecast ANN value.

The mathematical task of training ANN is to select internal calculated coefficients in such a way that the loss function is minimal after several complete rounds of the training sample (training epochs).

The results of changing the loss function after each epoch for training and test data sets are shown in Fig. 5.

As can be seen from the graphs in fig. 5, the error curve for both output layers reaches a plateau after the 8th epoch. To exclude retraining of the ANN, it is necessary to stop the learning process at this moment.

The assessment of the final accuracy of the ANN operation was carried out using a set of control data. In Fig. 6, the calculated values lie along the Fact-Calculation diagonal, which indicates the reliability of the results obtained. This is confirmed by the calculation of statistical characteristics (Table 5), according to which the deviation of the obtained data from the actual data does not exceed 0.5%.

The developed and trained ANN model allows to quickly calculate the necessary characteristics of a complex multicomponent hydrocarbon mixture for online monitoring of a real process unit - a heat exchangers section of a refinery hydrotreatment unit.

Further development of the ANN will make it possible to predict both the conditions under which the formation of contaminants on the surface of heat exchangers is possible in order to prevent device failures, and the establishment of the optimal frequency and duration of equipment re-pairs.

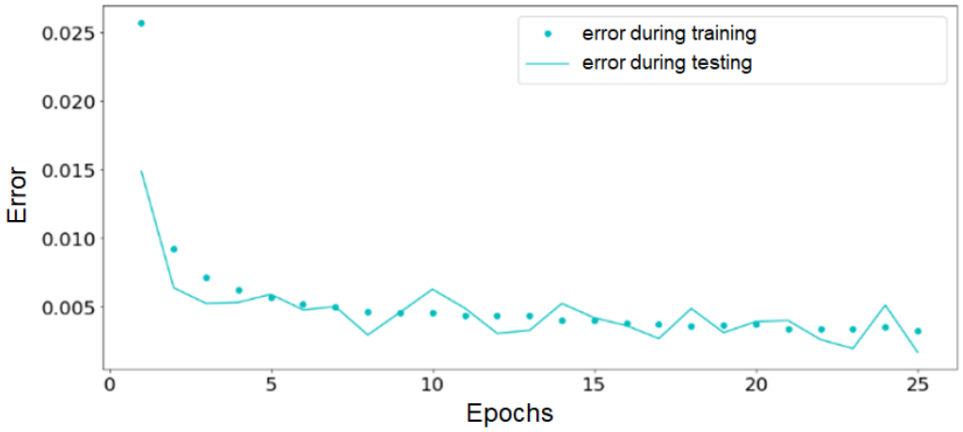

Fig. 5 Error during the training and test phases

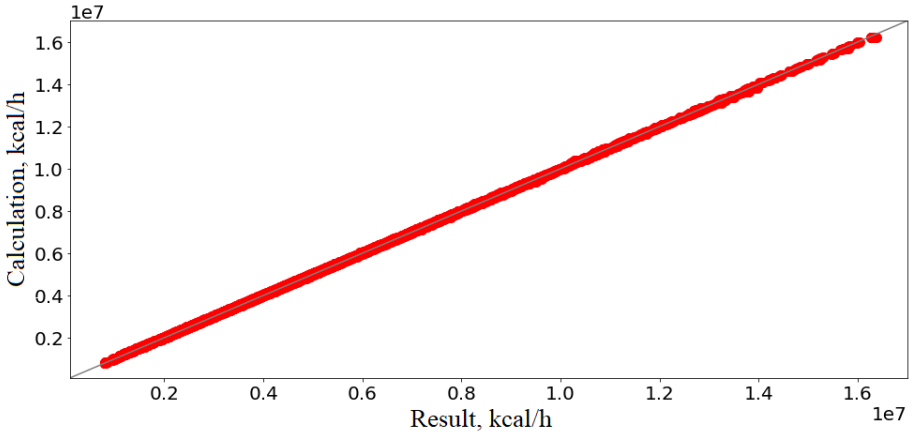

Fig. 6 Prediction accuracy of the enthalpy function versus temperature using the control data sample

Table 5 – Calculating the Accuracy of an Artificial Neural Network Model

Parameter	Calculating formula	Value
Mean absolute error	(1)	16 987
Mean absolute percentage error	$MAPE = \frac{1}{N_S}\sum_{i=1}^{N}\frac{\lvert L_i - E_i\rvert}{L_i}\cdot 100\%$	0,44
Root mean square error	$RMSE = \sqrt{\sum_{i=1}^{N}\frac{(L_i - E_i)^2}{N_S}}$	21 425
Coefficient of determination (R^2)	$R^2 = 1 - \frac{\sum(L_i - E_i)^2}{\sum(L_i - \overline{E}_i)^2}$	1,00

References

[1] S. Aguel, Z. Meddeb, and M.R. Jeday "Parametric study and modeling of cross-flow heat exchanger fouling in phosphoric acid concentra-tion plant using artificial neural network," Journal of Process Control. vol. 84, 2019, pp. 133–145.

[2] A.M. Demin and A.P. Naumenko "Diagnosis of heat exchange equipment based on operating parameters of a diesel fuel hydrotreater," [in Russian] Omsk Scientific Bulleti. vol. 4, 2019, pp. 84–88.

[3] M. Markowski and P. Trzcinski "On-line control of the heat exchanger network under fouling constraintsr," Energyi. vol. 185, 2019, pp. 521–526.

[4] D.K. Mohanty and P. Singru "Fouling analysis of a shell and tube heat exchanger using local linear wavelet neural network," International Journal of Heat and Mass Transfer. vol. 77, 2014, pp. 946–955.

[5] S. Tang, Y.-L. He, F.-L. Wang, Q. Zhao, and Y. Yu "Fouling analysis of a shell and tube heat exchanger using local linear wavelet neural network," International Journal of Heat and Mass Transfer. vol. 77, 2014, pp. 946–955.

[6] V.M. Kapustin and A.A. Gureev Oil refining technology. In 4 parts. Part two. Physical and chemical processes, [in Russian] Chemistry, Moscow, 2015.

[7] F. Schollet Deep Learning with Python, [in Russian] Piter, St. Petersburg, 2018.