

Multivariate models for monitoring the environmental performance of gasoline

V. V. Kovalenko^{1,*}, V. S. Loginov¹, M. V. Lyzlova¹, G. I. Melnik², D. S. Trunkin¹, and T. P. Shuvarikova¹

¹Ryazan State Radio Engineering University, 59/1, Gagarina Street, Ryazan, 390005, Russia

²Moscow Polytechnic University, Ryazan Institute (branch), 26/53, Pravo-Lybidskaya Street, Ryazan, 390000, Russia

Abstract. A linear multivariate regression model for the IR spectra of gasoline used in the paper. On the basis of a number of test samples (alkylate, reformate, regular, premium), an equation was obtained for calculating the content of aromatic hydrocarbons in gasolines based on the transmission values at wave numbers: 1600, 1500, 870, 810, 750 cm⁻¹. The calculated equation was tested on a series of regular gasolines and by the cross-validation method. The test results showed good values of the correlation coefficient (0,96) and the standard deviation (0,4 and 3,5). The range of the content of aromatic hydrocarbons in the test samples is in the range of 24,9 - 30,5 mass %.

1 Introduction

One of the most important characteristics of gasoline that determines its environmental safety is the content of aromatic hydrocarbons. High-octane gasolines provide a mixture of gasoline fractions from various technological processes of oil refining: gasoline fractions of direct distillation, catalytic cracking, reforming, alkylation, isomerization with the addition of individual high octane components and additives, and their improved performance properties. The resulting mixtures have a complex hydrocarbon composition, characterized by the content of paraffinic, isoparaffinic, naphthenic, olefinic, aromatic hydrocarbons, oxygenates. The composition of the refined oil and the applied processes of its processing, the presence of various processes in the technological scheme of production, will affect the component composition of commercial fuel fuels. So, for example, aromatic hydrocarbons are formed during the reforming of straight-run gasolines, are high-octane components that are specially added to gasolines to improve the performance of the fuel (knock resistance, saturated vapor pressure, fractional composition, chemical stability, etc.).

Over the past decades, the improvement of gasoline production is aimed at stricter requirements for environmental performance. At the same time, the anthropogenic impact on the environment associated with the use of gasoline occurs both at the stages of transportation and refueling of vehicles, and during their operation. Losses from evaporation and fuel leakage are 1.5-2% (about 700 thousand tons / year). When fuel is

* Corresponding author: vikvaskov@mail.ru

burned, the exhaust gases contain carbon monoxide and dioxide, nitrogen and sulfur oxides, which are harmful to the environment and human health. Among the combustion products of unleaded gasoline, there are also unburned hydrocarbons, which in the air, under the influence of high humidity and sunlight, contribute to the formation of persistent aerosols (smog). The products of combustion of olefinic and aromatic hydrocarbons have the highest photochemical activity. Aromatic hydrocarbons are the main source of solid particles. Thus, when the content of arenes in motor fuel doubles, specific emissions of solid particles increase fivefold. Aromatic hydrocarbons, especially benzene, pose the greatest danger to biosphere, are carcinogenic and mutagenic substances that can dissolve in oils and lipids, accumulate in the bone marrow, cause damage to the liver, kidneys, and lead to leukemia.

In most countries of the world, the requirements for the content of sulfur (no more than 10-15 ppm), alkenes (up to 14% mass), arenes (up to 35% mass), benzene (up to 1% mass), oxygen (up to 2.3% mass) are strictly standardized and constantly tightened.

The chemical composition of fuels is determined using modern methods of analysis (chromatographic, spectrophotometric, spectrometric, fluorescent and others), which allow obtaining reliable data on the qualitative and quantitative composition of complex mixtures, carrying out work with sample preparation and time spent on research. Most of the methods require the use of expensive equipment, specially trained personnel, and a large amount of consumables. Therefore, the use of alternative express methods for determining the group composition, including aromatic hydrocarbons in gasoline, is an important task in the analysis of intermediate and commercial products of oil refining in modern oil refineries.

One of the promising methods used to assess the performance of gasoline is the multidimensional modeling method, which makes it possible to determine the desired response value using a variety of experimentally measured parameters [1-4].

The aim of this work was to develop multidimensional calibration models for determining the content of aromatic hydrocarbons in gasolines using capillary gas chromatography and infrared spectroscopy data.

The objects of the study were samples of regular (A92) and premium (A95)gasolines, gasolines from reforming process (reformate) and alkylation process (alkylate). To create a calibration dependence, 16 samples were used out of a total of 50 samples. Samples of regular gasoline (A92) were used as a test kit to determine the model error.

2 Research methods

2.1 Gas liquid chromatography (GLC) method

Determination of the group and component composition of gasolines was carried out using a hardware-software complex for medical research on the basis of a chromatograph "Chromatek-Crystal 5000". The essence of the method lies in the chromatographic separation of gasoline on a capillary column with a non-polar stationary phase, followed by the registration of hydrocarbons with a flame ionization detector and automated processing of the information obtained using software. A gasoline sample with a volume of 0,3 ml is injected into a gas chromatograph equipped with a capillary column containing methylsiloxane as a solid phase applied to the walls of a quartz capillary column (length - 100 m, diameter – 0,25 mm). Under the action of a carrier gas, helium, at a flow rate of 1 ml/min, the sample passes through a column (efficiency of about 250,000 theoretical plates with respect to n-octane), in which its components are separated. Components are detected by a flame ionization detector as they elute from the column. The detector signal is processed by an electronic data storage system or an integrating computer.

2.2 Infrared spectroscopy (IR) method

The study of the group composition was carried out on an IR-Fourier spectrometer «Infraspec FSM 2202» with registration of spectra in the range 4000-400 cm⁻¹ with a resolution of 0,5 cm⁻¹. The measurements were carried out in a standard collapsible liquid cuvette with KBr glasses 32 mm in diameter with a 0,037 mm interlayer.

Structural analysis by the method of IR spectroscopy is the search for characteristic bandwidths and their assignment to the corresponding structural elements. Infrared radiation emitted by the source is directed into an interferometer, which consists of a system of mirrors to create a narrow light source. Then the light passes through the cuvette compartment, in which the liquid IR cuvette with the sample is installed, and is focused on the detector. The signal is recorded on the detector and displayed as an IR spectrum. The obtained IR spectra represent the dependence of the radiation intensity on the frequency, expressed in wavenumber cm⁻¹. The characteristic of vibrations of various groups of organic compounds allows them to be identified by the absorption peaks in the spectrum. For this, the regions of stretching and deformation vibrations of simple and multiple bonds are considered. The assignment is carried out taking into account the numerical values of the frequencies of the absorption maxima, the contour (shape) and the intensity of the bands. The obtained peaks in the IR spectrum are compared with experimental data on the IR absorption spectra of organic compounds, using special tables, as well as any additional information about the substance. In this work, such information is the values of the amount of individual components of commercial gasolines, determined by the chromatographic method.

The development of the method for determining the group composition of gasolines was carried out by a combination of capillary chromatography and IR spectroscopy [5,6]. Table 1 shows a comparison of these methods, from which it can be seen that spectroscopic methods are of particular interest due to the absence of the need for chemical reagents for analysis and the speed of implementation. The method of IR spectroscopy is convenient for use in devices for automatic monitoring of the state of the environment.

Table 1. Comparison of infrared spectroscopy (IRS) and gas liquid chromatography (GLC) methods.

Index	IRS	GLC
Sample time	2-3 minutes	70 minutes
Sample quantity for analysis	maximum 0,5 ml.	0,3 ml.
Identification of results using special software	Data base of spectrum individual substances are used	Special software are used
Additional reagents	-	carrier gas 1 ml/min gas preparation unit

The composition of gasolines according to the data of chromatographic analysis can be represented as the sum of hydrocarbons of aliphatic (paraffins, isoparaffins, olefins), acyclic (naphthenes) and aromatic (arenes) structures.

1) Paraffinic compounds are evenly distributed in premium gasoline (A95), in reforming catalysat, in regular gasoline (A92) from pentane to tridecane, hexane, heptane and octane are absent in alkylate.

2) Isoparaffins are most often found in the form of 2,2,4-; 2,2,3-; 2,3,4-; 2,3,3-trimethylpentane, while in alkylate they are 74%, in premium gasoline (A95) - 16%, and isopentane (10%).

3) Aromatic compounds in regular and premium gasoline are contained in the form of benzene, toluene, xylenes, trimethylbenzenes and other compounds.

4) Cyclic compounds (naphthenes) in regular and premium gasoline are mainly found as methyl-, ethyl, dimethyl-, diethyl- substituted cyclopentane and cyclohexane.

5) Unsaturated compounds (olefins) in regular and premium gasoline are present in the form of pentenes and hexenes, they are absent in the alkylate, in the reforming catalysat they are evenly distributed from methyl butene to dimethylpentene and individually constitute hundredths of a percent.

6) Oxygenates are represented by two compounds: methyl tertiary butyl ether(MTBE) and tertiary amyl methyl ether (TAME).

Each of these types of hydrocarbons in the IR spectra has its own peaks corresponding to the characteristic frequencies. For identification and comparison of the most characteristic groups, the spectral region 700-3000 cm⁻¹ was used.

The figure shows the IR spectra of four samples of gasolines differing in the content of aromatic hydrocarbons.

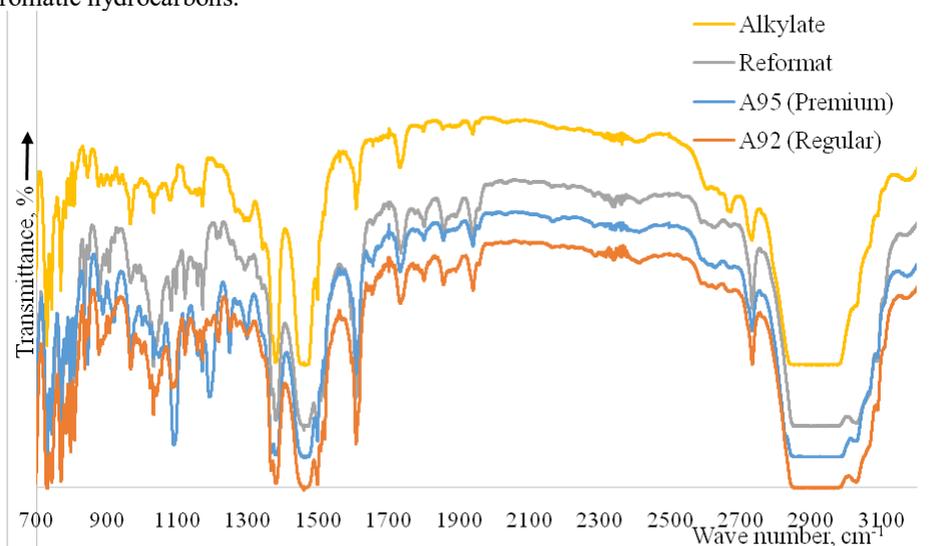


Fig. 1. IR spectra of the studied gasolines to determine the absorption frequencies of arenes.

Analysis of the IR spectra of alkylate, reformat, regular and premium gasoline allows one to determine the wavenumber of absorption maxima related to the corresponding structural elements:

- aliphatic hydrocarbons (alkanes, alkenes and isoalkanes) -1250, 1460, 1380, 2840-3000 cm⁻¹

- cyclic (naphthenic) hydrocarbons - 960-980 cm⁻¹

- aromatic hydrocarbons - 1600, 1500, 870, 810, 750 cm⁻¹.

For these wavenumbers, the transmittance ($T = I/I_0$), optical density ($D = \lg 1/T$), and the percentage of aliphatic (Al), acyclic (Ac) and aromatic (A) structures were calculated:

$$Al = (D_{al} / \sum D) * 100 \%;$$

$$Ac = (D_{ac} / \sum D) * 100 \%;$$

$$A = (D_a / \sum D) * 100 \%,$$

where D_{al} , D_{ac} , D_a - the sum of optical densities of aliphatic (alkanes, alkenes and isoalkanes) hydrocarbons, respectively, at frequencies of 1250, 1460, 1380, 2840-3000 cm⁻¹; cyclic (naphthenic) hydrocarbons, respectively, at frequencies of 960-980 cm⁻¹; aromatic hydrocarbons, respectively, at frequencies of 1600, 1500, 870, 810, 750 cm⁻¹.

Optical densities for aromatic compounds of the studied samples are shown in figure 2.

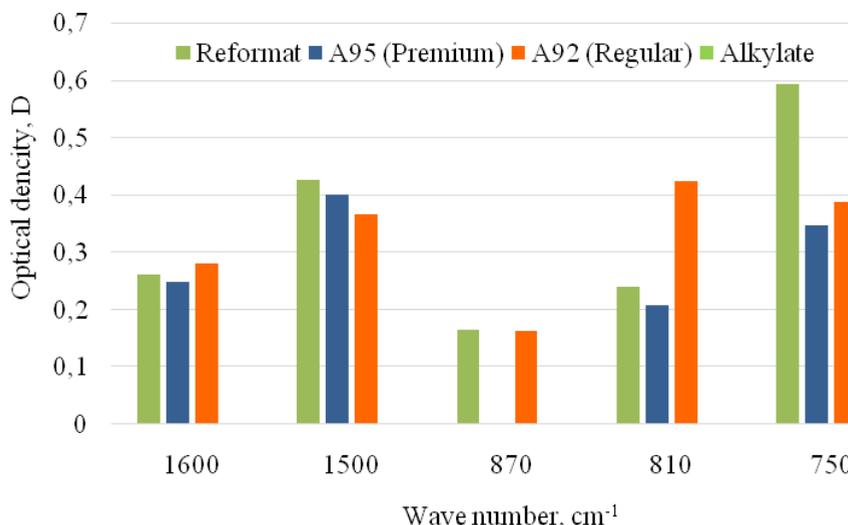


Fig. 2. Optical density of gasoline samples in different frequency ranges.

The data obtained in the frequency range 1600, 1500, 870, 810, 750 cm⁻¹, characterizing the content of arenes, show significant differences in the absorption spectra of gasolines. For example, for alkylate containing a minimum concentration of aromatic hydrocarbons, absorption lines are observed in the IR spectrum only at a frequency of 750 cm⁻¹. In premium gasoline there is no absorption line corresponding to 870 cm⁻¹. Therefore, it is impossible to reveal a direct relationship between the values of optical density at one specific frequency and the content of aromatic compounds. To determine the concentration of arenes, it is necessary to take into account the optical densities calculated from the absorption spectra at several frequencies. The calculations obtained in this way are presented in table 2.

Table 2 also provides comparative data on the concentration of aliphatic, acyclic and aromatic hydrocarbons. determined by the methods of gas-liquid chromatography on the chromatograph "Chromatek-Crystall 5000" and IR spectroscopy on the IR-Fourier spectrometer "Infraspec FSM 2202" [5,7]. There is a good correlation between the values of the group composition obtained by chromatographic and spectroscopic methods.

Table 2. Group composition of gasolines, determined by IR spectroscopy and capillary gas chromatography.

Gasoline	Group hydrocarbon composition. %						Average error
	Aliphatic		Acyclic		Aromatic		
	GLC	IRS	GLC	IRS	GLC	IRS	
Alkylate	98.60	97.91	0.50	0.00	0.88	2.09	1.78
Premium	60.58	65.79	4.50	4.04	33.08	30.17	
Reformat	40.53	38.24	1.82	4.35	57.65	57.41	
Regular	68.17	68.73	4.27	4.35	27.56	26.92	

Thus, to determine the concentration of arenes in gasoline, it is necessary to have the values of the concentrations of aliphatic and acyclic compounds, and hence the calculated optical densities in different frequency ranges for these compounds.

At the same time, the use of multivariate data analysis methods in relation to infrared spectroscopy makes it possible to create calibrations that take into account only the content of aromatic hydrocarbons. To do this, it is enough to find the characteristic bands, in our case, the absorption frequency of infrared radiation, by which it is convenient to track the presence of arenes, and to construct a calibration graph in the coordinates of optical density - concentration.

Using the above method for determining arenes, we initially considered the simplest linear dependence of the concentration of aromatic hydrocarbons on optical density for various gasolines and gasoline components. The resulting dependence has an average correlation coefficient $R^2 = 0,7345$. This is explained by the fact that the calculations used the data obtained as a result of experimental measurements for gasolines belonging to different clusters [3].

The most reliable results with a correlation coefficient $R^2 = 0,9407$ in the determination of arenes were obtained in the study of alkylate, reformat, regular and premium gasolines from one manufacturer. Linearized dependences are shown in figure 3.

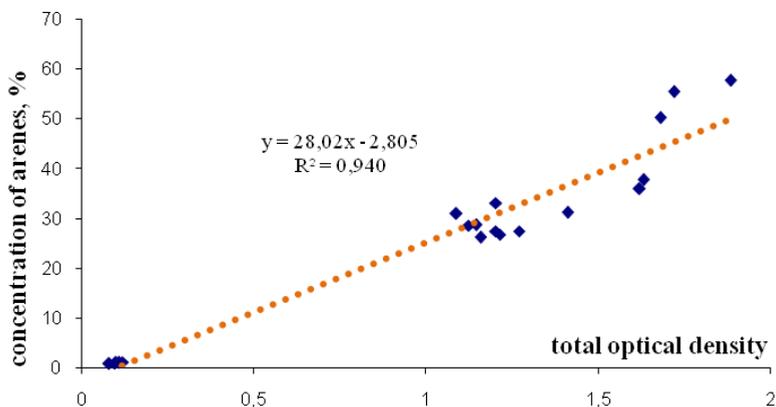


Fig. 3. Dependence of the content of aromatic hydrocarbons in gasolines of one cluster on optical density.

Figure 4 shows the dependence of the arenes content on the intensity of absorption of IR radiation described by a linear equation. The processing of such measurements does not require additional calculations and can be obtained as a result of a quick analysis of the IR spectra.

The most reliable results of determining the specified parameters of gasolines are provided by the application of multivariate models using the method of multiple linear regression [3,4]. This method of analyzing experimental data makes it possible to identify the most reliable relationship between optical density and the concentration of aromatic hydrocarbons in multicomponent hydrocarbon mixtures such as gasoline.

The method of multiple linear regression is based on obtaining a regression relationship between the experimental values of matrices X and Y. Matrix Y consists of dependent variables, called responses. In our case, Y is the concentration of aromatic hydrocarbons. Matrix X is an independent variable, a matrix of analytical signals. In our studies, this is either the intensity of absorption of infrared radiation or optical density, which is determined from the intensity of the absorption spectra for alkylate, reformat, regular and premium gasoline.

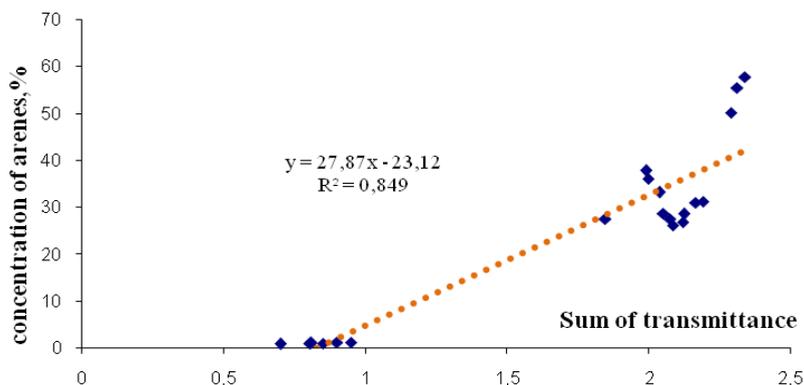


Fig. 4. Dependence of arene concentration on absorption intensity.

When implementing the MLR (multiple linear regression) method, to determine a certain parameter P, a regression equation of the form [2]

$$P = b_0 + b_1 k(v_1) + \dots + b_N k(v_N),$$

where b_0, \dots, b_N - are the coefficients of the regression equation, $k(v_1), \dots, k(v_N)$ - is the transmittance (%) or optical density at a certain wavenumber.

The P parameter is the concentration of arenes in gasoline. The coefficients of the regression equations are determined from the results of experiments with standardized gasoline samples (regular and premium) and gasoline components (reformate and alkylate).

At the same time, when performing the regression analysis of gasolines, the following features of the systems being modeled are taken into account: firstly, not the entire absorption spectrum within the selected range is used, but only its individual points characterizing the absorption coefficient at the frequencies selected for the regression analysis; secondly, for the models created by the MLR (multiple linear regression) method, the models calculated for the same type of gasoline have the best characteristics and the minimum calibration error [3,4].

In this regard, from the investigated gasolines, samples of the regular brand with a known hydrocarbon composition determined by the chromatographic method were selected.

For these samples, IR spectra were taken (figure 5), which were used to determine the transmittance and optical densities in the range of wavenumbers 1600, 1500, 870, 810, 750 cm^{-1} , which characterize the content of arenes. Concentration range 20,17-68,73%.

Based on the experimental data, the final regression equation for the model was obtained:

$$C_A = -33,151 + 17,59 \cdot T_{1600} + 94,441 \cdot T_{1500} - 3,5427 \cdot T_{870} + 58,4 \cdot T_{810} - 52,797 \cdot T_{750}.$$

where T_{1600}, \dots, T_{750} is the transmittance at 1600, ..., 750 cm^{-1} , respectively, C_A - content of aromatic hydrocarbons, mass %.

After receiving the model, the next step is to validate the model. Model performance can be assessed by cross-validation and validation with test samples.

Six test samples of gasolines with the corresponding arene content were selected as test samples. Figure 5 shows the optical density of test samples of gasoline in different frequency ranges. The results of comparison of the predicted values with the data of chromatographic analysis in the figure 6.

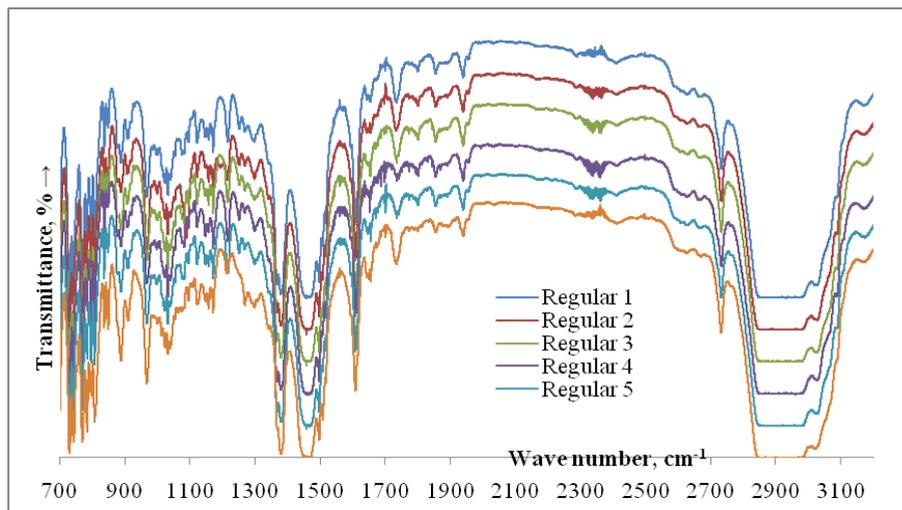


Fig. 5. IR spectra of the investigated gasoline samples.

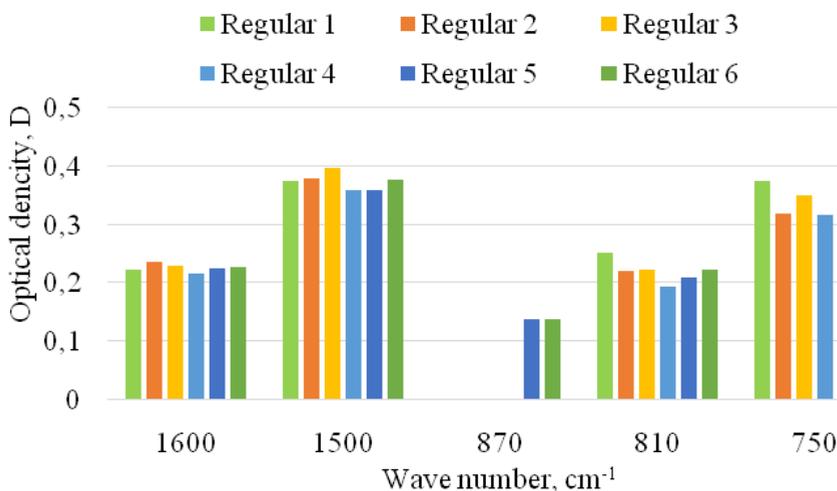


Fig. 6. Optical density of test samples of gasoline in different frequency ranges.

Table 3. Characteristics of calibration models.

Calibration Model	Range. %		Correlation coefficient	Standarddeviation	
	min	max	R ²	calibration	check
Aromatic hydrocarbons	24.9	30.5	0.96	0.4	3.5

As can be seen from Figure 7, the values calculated from the IR spectra are close to the data obtained by gas-liquid chromatography.

For cross-validation, test samples were included in the model and a new model was built. The essence of cross-validation, or cross-validation, is that one sample is excluded from the calibration, a model is built based on the remaining samples, which is applied to the excluded sample. Then this sample is included in the calibration, and the next is

excluded, and so on for each sample [4]. The characteristics of the obtained models are presented in Table 3.

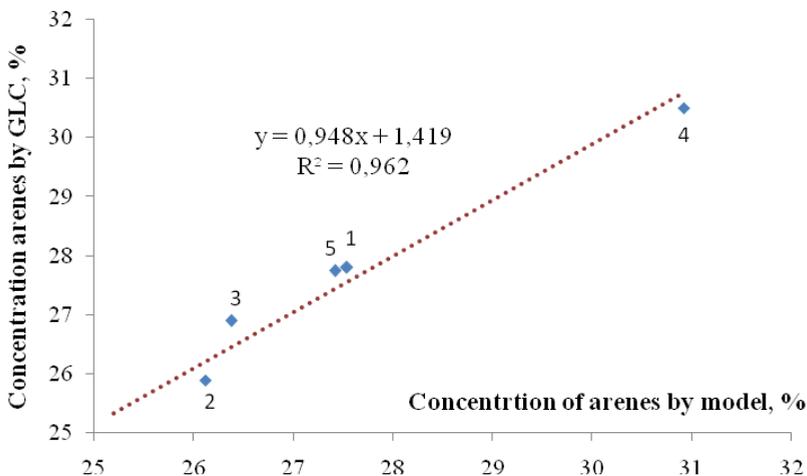


Fig. 7. Correlation of model and chromatographic data.

The cross-validation result showed that the model has a high correlation coefficient ($R^2 = 0,96$) and a standard deviation of 0,4. The standard deviation of the check is 3,5.

3 Conclusion

The results of the research show that infrared spectroscopy methods make it possible to reliably determine the concentration of aromatic hydrocarbons in gasolines, to determine the parameters of motor gasolines, including their compliance with European standards for the total content of aromatic components, which characterize the environmental safety of fuel.

The most reliable method for analyzing spectrometric data is the method of multiple linear regression. This method of analyzing experimental data makes it possible to identify the most reliable relationship between optical density and the concentration of aromatic hydrocarbons in multicomponent hydrocarbon mixtures such as gasoline. The characteristics of the model can be improved by increasing the number of samples, expanding the range, using reference and test samples of gasolines belonging to the same cluster.

The use of multidimensional data analysis methods opens up wide opportunities for creating express methods for the determination of aromatic hydrocarbons, as an alternative to existing chromatographic methods (gas-liquid chromatography and high-performance liquid chromatography).

One of the main disadvantages of this method is the need for a large set of calibration mixtures or standard samples with specified characteristics, predominantly belonging to one fuel manufacturer.

References

1. Vesnin V L, Muradov V G 2012 *Journal of Applied Spectroscopy* **79(4)** 533
2. Vesnin V L, Muradov V G 2004 *Bulletin of the Samara Scientific Center of the Russian Academy of Sciences* **6(1)**

3. Vesnin V L, Muradov V G 2014 *Bulletin of the Samara Scientific Center of the Russian Academy of Sciences* **16(4)**
4. Romanova R G, Sitdikov R R 2017 *Bulletin of the Technological University* **20(11)** 40
5. Voronov M V, Kovalenko V V, Lyzlova M V, Melnik G I, Trunkin D S and Shuvarikova T P 2020 *III International Scientific and Technical Forum "Modern Technologies in Science and Education" (MTSE-2020) Russia Ryazan RSREU* **3** 256
6. Vagnenkov T V, Koshelev I I, Lyzlova M V and Melnik G I 2019 *Biotechnical, medical, ecological systems and robotic complexes - Biomedsystems-2019: collection of articles XXXII All-Russia scientific and technical Forum Russia Ryazan* 317
7. Kerimova M E, Redicheva T V, Kovalenko V V and Lyzlova M V 2018 *International Scientific and Technical Forum "Modern Technologies in Science and Education" (MTSE-2018) Russia, Ryazan RSREU* **3** 114