doi: 10.18323/2782-4039-2023-3-65-10

Concerning the selection of areas with a dominant type of dependence when analyzing production control data

Victoria V. Timoshenko[1], postgraduate student
Ekaterina S. Budanova[2], graduate student
Davronjon F. Kodirov[3], postgraduate student
Elina A. Sokolovskaya[4], PhD (Engineering), Associate Professor,
assistant professor of Chair of Materials Science and Strength Physics
Aleksandr V. Kudrya* [5], Doctor of Sciences (Engineering), Professor,
professor of Chair of Materials Science and Strength Physics
University of Science and Technology MISIS, Moscow (Russia)

***E-mail: AVKudrya@misis.ru**

[1]ORCID: https://orcid.org/0009-0008-7671-0291
[2]ORCID: https://orcid.org/0009-0003-4055-9298
[3]ORCID: https://orcid.org/0009-0003-5380-5558
[4]ORCID: https://orcid.org/0000-0001-9381-9223
[5]ORCID: https://orcid.org/0000-0002-0339-2391

Received 06.06.2023 *Accepted 16.06.2023*

Abstract: The formation of representative databases determines the interest in forecasting and managing the quality of metal based on data mining using special software products often based on regression analysis and not always taking into account the statistical nature of an object of study itself. This can lead to misinterpretation of the results or incomplete extracted information reducing the efficiency of statistical processing. Based on the analysis of the production database of the technology for producing 13G1S-U sheet steel, the authors evaluated the possibilities of multiple linear regression for predicting the quality of a steel sheet. The study shows that the type of distribution of the values of control parameters, the distribution nature of which was estimated based on the determination of the skewness and kurtosis coefficients, limits the regression forecast depth. Due to the great deviation of the predicted models from the experimental values in the right tail area of the distribution of the impact strength values, in this work, the authors developed the methods for separating data arrays and proposed criteria to compare the obtained results. To assess the accuracy of the results obtained, arrays with a deliberately asymmetric distribution were selected from the initial sample, against which the statistical characteristics were also compared. Based on the proposed techniques, the authors identified the dominant chemical elements that contribute to the difference in the distribution of the values of acceptance properties existing within the same standard technology. The study shows that the proposed separation method can be used as a variation of cognitive graphics techniques to identify areas with a dependence dominant type based on the correlation of skewness and kurtosis coefficients.

Keywords: analysis of production control data; metal product quality control; quality forecast in metallurgy; cognitive graphics techniques; production data mining; 13G1S-U steel.

For citation: Timoshenko V.V., Budanova E.S., Kodirov D.F., Sokolovskaya E.A., Kudrya A.V. Concerning the selection of areas with a dominant type of dependence when analyzing production control data. *Frontier Materials & Technologies*, 2023, no. 3, pp. 103–114. DOI: 10.18323/2782-4039-2023-3-65-10.

INTRODUCTION

The high level of fitting of metallurgical enterprises with digital tools of intermediate control, and information collection, allows obtaining representative databases of production in a short period. This determines the interest in a post-event analysis of such data arrays for identifying the reasons for the heterogeneity of the quality of metal products, and developing sound principles for managing them [1; 2]. When analysing such information, a passive experiment is implemented [3; 4]. The costs of which are justified within the current, well-functioning technology, when the high quality of the metal is achieved for a part of the product (usually from half of its volume or less), and it is necessary to solve the problem of "improving the low-performers" to the advanced level [5].

Traditional approaches to solving such problems are based, in particular, on the analysis of the stability of process and product parameters, for example, using Shewhart control charts[1]. Their main disadvantage is the assumption of normalcy of distributions and a single space of process and product parameters, which can lead to poor forecasting performance [6]. In fact, a deep post-event analysis will be more informative, revealing both the presence of direct links between significant control parameters and properties and their manifold influence, taking into account existing

[1] *GOST R ISO 7870-1-2022. Statistical Methods. Control Charts. Part 1. General guidelines. M.: Russian Standardization Institute, 2022. 19 p.*
GOST R ISO 7870-2-2015. Statistical Methods. Control Charts. Part 2. Shewhart Control Charts. M.: Standartinform, 2016. 41 p.
GOST R ISO 7870-3-2013. Statistical Methods. Control Charts. Part 3. Acceptance Control Charts. M.: Standartinform, 2014. 18 p.

areas with one or another type of dependence, the boundaries of which can be determined, for example, using cognitive graphics techniques. [7].

Recently, to solve the problem of quality control of metal products (including end-to-end – from smelting to acceptance tests), machine learning algorithms are increasingly used. Among them, methods based on decision trees [8–12], neural networks, etc. [13–19] have become widespread. In a number of cases, their application allowed creating models, the prediction of which was higher against the conventional regression, for example, when developing a digital laser cladding model [9]. Random forest algorithms were also used when predicting the structure (in particular, the interlamellar distance of precipitated phases), and mechanical properties for alloys with a relatively short production chain [11; 12]. The application of methods, based on decision trees, showed its effectiveness when predicting the consequences at individual stages of the production cycle.

There is a wide interest in software solutions built on the basis of neural networks, for example, for predicting mechanical (during tensile and impact tests [14], as well as measuring hardness [15]), and processing characteristics (hardenability [16]) depending on chemical composition. Moreover, the use of neural networks was tested in the production of slabs, to predict the cast structure, depending on the variation in the values of technological parameters of production [17]. Neural networks were also applied to predict metal defects in the manufacturing process [5]. Neural networks are actively used when recognizing the structure objects, for example, when detecting sulfide inclusions [18] or classifying structural components [19]. Despite their "independence" in work, it is obvious that the effectiveness of the functioning of neural network software products will be largely determined at the stage of their training by highly skilled representatives of the expert community.

In this regard, in particular, it is important to consider issues related to specifying the role of the statistical nature of objects in metallurgy, when applying both classical and developed algorithms. Obviously, it is also important to determine the boundaries of areas with a dominant dependence type, the patterns of their interaction, which will allow justifying the choice of appropriate statistical procedures and criteria. When developing IT solutions in relation to metallurgy, these circumstances are not always given the necessary attention, which, at least, follows from the analysis of relevant publications.

The purpose of this work is to determine the procedure for selecting samples (with the asymmetric nature of the distribution histograms of the values of the 13G1S-U sheet steel acceptance characteristics), to identify areas of change in technological parameters with a dominant dependence type.

METHODS

The study object was the database of production control of the process and product of the technology for producing rolled 13G1S-U steel with a sheet thickness of 12 mm and consisting of 1021 heats. The impact strength values at test temperatures of 0 and −40 °C (KCV^0 and KCU^{-40}, respectively) were chosen as quality characteristics.

Regression analysis was performed using multiple linear regression. The parameters identified were the multiple regression coefficients b_i, the standard error of the regression coefficients σ_b, the correlation coefficients R, and the significance level of the Fisher test F. The removal of insignificant coefficients was carried out when the regression coefficient was equal to zero within the error and the Student's risk level was equal to 0.05.

The main calculated characteristics of the samples were the average value x_{av}, the standard error of the sample σ, the skewness A_s and kurtosis E_x coefficients.

RESULTS

Initial sample analysis

Fig. 1 shows the distribution histograms of the impact strength values for 13G1C-U sheet steel.

The available samples have an asymmetric distribution, while the skewness coefficients are equal to 1.44 and 1.26, and the kurtosis coefficients are equal to 2.53 and 2.38 for the impact strength values KCV^0 and KCU^{-40}, respectively (if the error in their determination for the skewness coefficient is 0.23 and for kurtosis is 0.76).

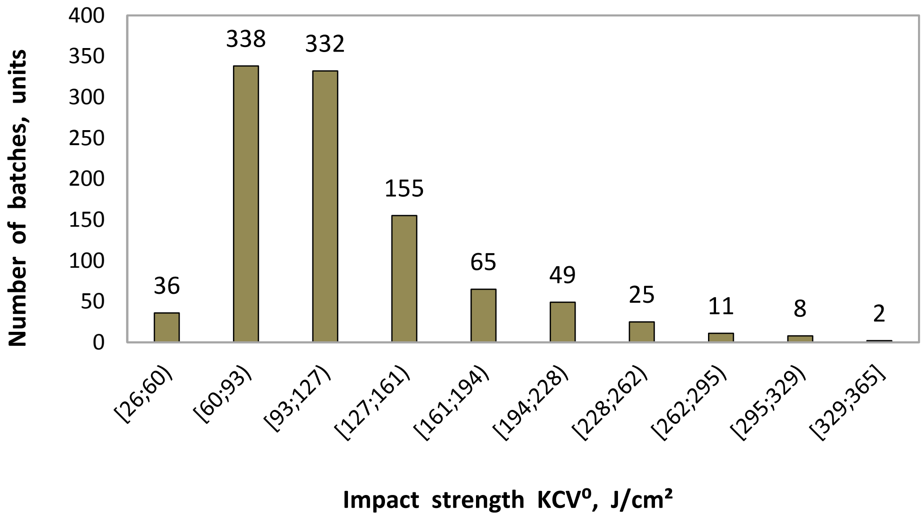
Regression models were obtained to predict impact strength, with respect to chemical composition, the characteristics of which (b_i is the regression coefficient; σ_b is the standard error of the regression coefficient; R is the correlation coefficient; F is the Fisher criterion significance level) are shown in Table 1.

The regression models obtained are significant proceeding from the Fisher criterion significance level. Fig. 2 presents the graphs illustrating the correspondence of the model to the actual values.

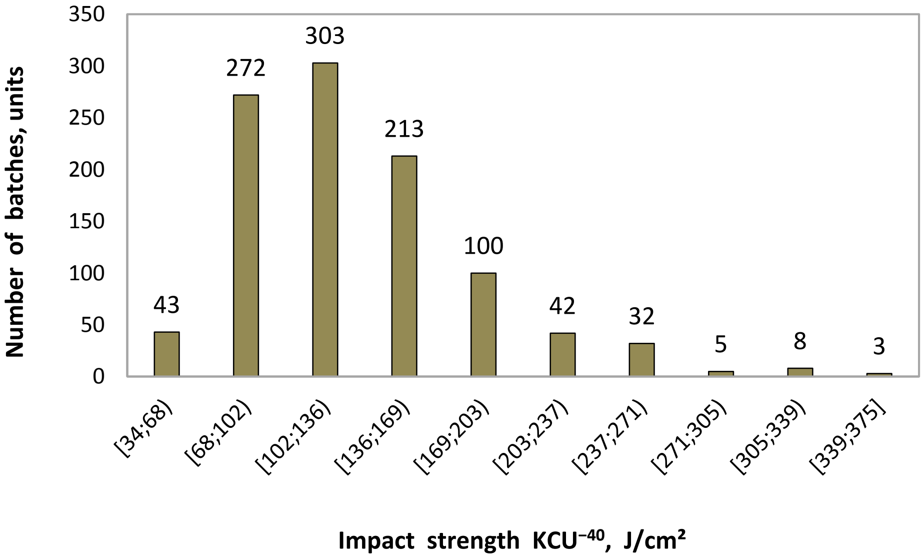
The resulting models predict the desired mechanical properties at a low level. At the same time, clearly outlying values of impact strength are noticeable in the region of its high values, which correspond to the right tails of the histograms determining the deviation of the distribution as a whole from the normal one. This deviation may be the result of the joint action of independent disturbances of approximately equal strength, which determines the nature of its distribution (close to normal). Therefore, there is an interest in dividing the initial sample of impact strength values into two sub-arrays.

Dividing the sample into two data arrays

The initial histograms of impact strength values with pronounced asymmetry were divided into two distributions (with signs of normal one), based on the selection of the largest sample of values (basic sample) as a data array, where the statistical outliers are absent, if the values of the skewness A_s and kurtosis E_x coefficients approach zero as much as possible [20]. The remaining impact strength values (from the right tail of the primary sample distribution) were selected into a separate array (selected sample) and supplemented with values from the basic sample (drawn at random from the right tail of its histogram), also until the values of skewness A_s and kurtosis E_x coefficients (outliers) approach zero as much as possible. The values of technological parameters (chemical composition) corresponding to the two formed samples of impact strength were also separated into paired data arrays.



a

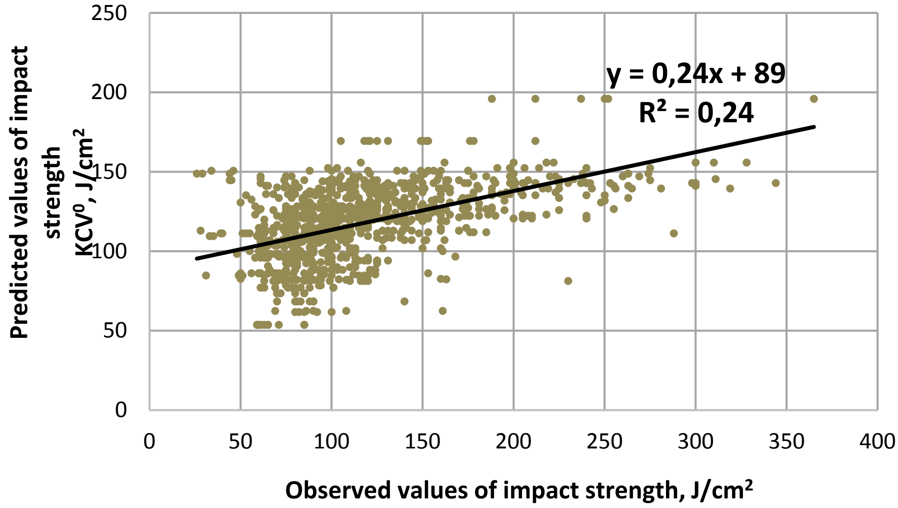


b

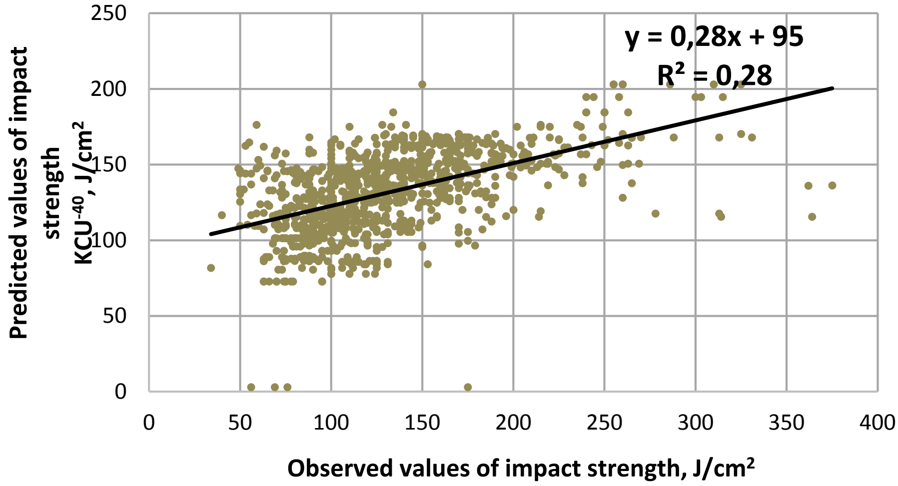
Fig. 1. *Histograms of distributions of impact strength of 13G1S-U steel:*
a *– test temperature 0 °C, V-cut;*
b *– test temperature −40 °C, U-cut*
Рис. 1. *Гистограммы распределений ударной вязкости стали 13Г1С-У:*
a *– температура испытания 0 °C, V-надрез;*
b *– температура испытания −40 °C, U-надрез*

Table 1. *Characteristics of the "chemical composition – property" regression model for impact strength of 13G1S-U steel*
Таблица 1. *Характеристики регрессионной модели «химический состав – свойство» для ударной вязкости листовой стали 13Г1С-У*

Property	**Equation parameters** $\left(\frac{b_i}{\sigma_b}\right)$									***R***	***F***
	Mo	**Nb**	**Cu**	**Cr**	**S**	**Mn**	**Si**	**C**	**0**		
KCU^{-40}	−1374 403	1485 270	−252 73	132 42	−9425 557	−123 33	−106 29	−833 214	485 62	0.53	$2\cdot10^{-6}$
KCV^{0}	–	–	–	–	−9412 521	–	−174 23	–	255 12	0.49	$4\cdot10^{-4}$



a



b

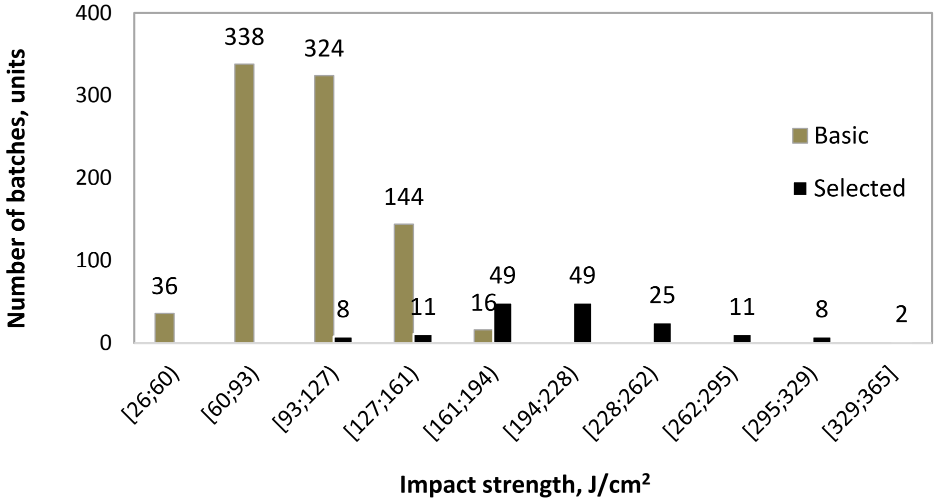
Fig. 2. *Correspondence of predicted and observed values of impact strength of 13G1S-U steel:*
a *– test temperature 0 °C, V-cut;* ***b*** *– test temperature −40 °C, U-cut*
Рис. 2. *Соответствие предсказанных и наблюдаемых значений ударной вязкости стали 13Г1С-У:*
a *– температура испытания 0 °C, V-надрез;* ***b*** *– температура испытания −40 °C, U-надрез*

Fig. 3 shows diagrams for separated impact strength arrays, and Table 2 represents the statistical characteristics of the obtained distributions: x_{av} average values, their σ determination error, skewness and kurtosis coefficients. At the same time, the number of values in the main sample and outliers was 878 and 143 batches for the KCU^{-40} impact strength and 858 and 163 for the KCV^{0} impact strength, respectively.

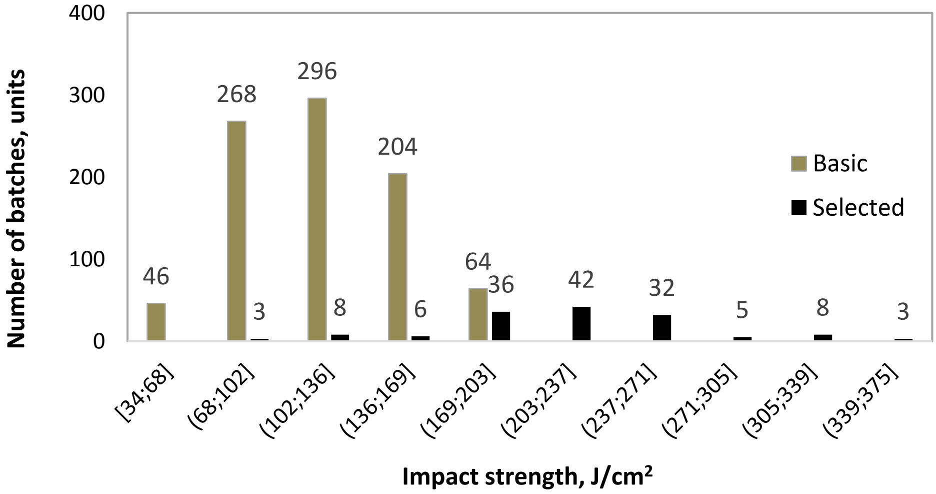
According to the appearance of the diagrams, and the values of the skewness and kurtosis coefficients, it follows that the distribution of the obtained samples is close to normal.

Table 3 presents the statistical characteristics of the chemical composition samples according to the division into sub-arrays with a normal distribution of KCU^{-40}, and KCV^{0} impact strength values. Differences between the types of distribution of the corresponding chemical composition samples were estimated based on a comparison of the values of the skewness and kurtosis coefficients for the main sample and selected within the error. Therefore, for example, if the difference between the coefficients was greater than the sum of their errors, then this confirmed the validity of the hypothesis about the difference between the samples.

One should note that following the results presented in Table 3, it is not possible to identify differences between the samples based on the average values of the content of the chemical composition constituents. However, for some arrays of values of the content of chemical elements in the basic and selected samples, there is a difference in the form of their distribution: the values of their skewness and kurtosis coefficients differ significantly. Thus, for two samples of impact strength at −40 °C, differences were revealed in the form of corresponding distributions of the content of sulphur, nickel, titanium, vanadium, niobium, boron, and molybdenum; for two samples of impact strength at 0 °C – for the content of nickel (by the value of the kurtosis coefficient only), titanium, vanadium, niobium, and molybdenum. In both cases, the difference in the titanium distribution was revealed by the kurtosis coefficient only. This means, that particular, these parameters contribute the most to the difference between the normally distributed samples shown in Fig. 3.



a



b

Fig. 3. *Distribution diagrams for the impact strength values of 13G1S-U steel when dividing the initial array into two sub-arrays:*
a *– test temperature 0 °C, V-cut;* ***b*** *– test temperature −40 °C, U-cut*
Рис. 3. *Диаграммы распределения значений ударной вязкости стали 13Г1С-У при разделении исходного массива на два подмассива:*
a *– температура испытания 0 °C, V-надрез;* ***b*** *– температура испытания −40 °C, U-надрез*

Table 2. *Statistical characteristics of divided arrays of the impact strength values of 13G1S-U sheet steel*
Таблица 2. *Статистические характеристики разделенных массивов значений ударной вязкости листовой стали 13Г1С-У*

Property	Sample type	x_{av}, J/cm²	σ, J/cm²	A_s	E_x
KCV^0	Basic	101	29	0.25±0.25	−0.46±0.83
	Selected	205	51	0.19±0.57	0.64±1.83
KCU^{-40}	Basic	118	34	0.24±0.25	−0.54±0.82
	Selected	214	58	0.12±0.60	0.45±1.94

Identification of areas of change in the values of control parameters with a dominant dependence type

The assessment of the influence of certain technology parameters on the quality deviation based on the traditional comparison of their average values over the sample is complicated by some factors, for example, due to the lack of a normal distribution of control parameter values. Regression is also usually ineffective due to the lack of a single space of process and product parameters. Obviously, the influence of certain technology parameters (mainly joint) on the properties of metal products, both positive and negative, is most pronounced, when the level of properties

Table 3. *Statistical characteristics of samples corresponding to the distribution of the content of the 13G1S-U steel chemical composition elements, which conform to the basic and selected arrays of the impact strength values (KCU^{-40} and KCV^0)*
Таблица 3. *Статистические характеристики выборок, отвечающих распределению содержания элементов химического состава стали 13Г1С-У, соответствующих основному и выделенному массивам значений ударной вязкости (KCU^{-40} и KCV^0)*

Parameter	**Relevant to the array of the KCU^{-40} values**				**Relevant to the array of the KCV^0 values**			
	x_{av}, % wt.	**σ, % wt.**	**A_s**	**E_x**	**x_{cp}, % wt.**	**σ, % wt.**	**A_s**	**E_x**
C	0.13*	0.01	−1.34	2.02	0.13	0.01	−1.36	1.97
	0.13	0.01	−0.99	0.13	0.13	0.01	−1.13	0.81
Si	0.49	0.06	−2.28	6.46	0.48	0.06	−2.28	6.12
	0.47	0.08	−2.48	4.99	0.48	0.07	−2.81	7.80
Mn	1.41	0.06	1.57	3.19	1.41	0.06	1.60	3.14
	1.41	0.05	1.68	4.38	1.41	0.05	1.19	3.49
P	0.014	0.003	1.24	2.18	0.014	0.003	1.22	2.17
	0.014	0.003	0.95	1.42	0.014	0.003	1.06	1.35
S*	0.006	0.003	**1.29**	**2.82**	0.006	0.003	1.39	3.37
	0.004	0.003	**3.23**	**15.09**	0.004	0.002	1.07	1.06
Cr	0.10	0.03	1.60	4.33	0.10	0.03	1.63	4.35
	0.11	0.04	1.43	3.04	0.10	0.03	1.27	2.80
Ni*	0.09	0.02	**1.89**	**7.01**	0.09	0.02	1.76	**6.36**
	0.13	0.02	**0.18**	**−0.32**	0.09	0.02	2.12	**9.11**
Cu	0.13	0.02	0.82	1.08	0.13	0.02	0.78	0.97
	0.13	0.02	0.18	−0.32	0.13	0.02	0.76	1.40
Ti*	0.006	0.005	3.38	**10.05**	0.006	0.005	3.18	**8.78**
	0.006	0.005	2.73	**6.88**	0.005	0.004	3.95	**15.96**
Al	0.03	0.006	−0.03	0.42	0.03	0.006	−0.04	0.46
	0.03	0.006	−0.13	−0.05	0.03	0.006	−0.04	−0.19
V*	0.003	0.009	**5.40**	**29.12**	0.004	0.009	**5.11**	**25.87**
	0.007	0.014	**2.85**	**6.42**	0.005	0.013	**3.48**	**10.49**
N	0.009	0.002	0.13	−0.93	0.009	0.002	0.13	−0.92
	0.008	0.002	0.42	−0.16	0.008	0.002	0.36	−0.42
As	0.017	$2\cdot10^{-16}$	1.00	−2.00	0.017	$2\cdot10^{-16}$	1.00	−2.00
	0.017	$5\cdot10^{-17}$	1.01	−2.03	0.017	$6\cdot10^{-17}$	1.01	−2.03

Continue table 3

Parameter	Relevant to the array of the KCU^{-40} values				Relevant to the array of the KCV^0 values			
	x_{av}, % wt.	σ, % wt.	A_s	E_x	x_{cp}, % wt.	σ, % wt.	A_s	E_x
Nb*	0.043	0.007	**2.57**	**6.78**	0.043	0.007	**2.50**	**6.31**
	0.044	0.005	**1.28**	**2.24**	0.043	0.005	**0.85**	**−1.30**
B*	0.001	0.0001	**−3.12**	**11.54**	0.001	0.0001	−2.93	10.19
	0.001	0.0002	**−2.11**	**5.60**	0.001	0.0001	−2.90	10.75
Mo*	0.008	0.004	**6.18**	**47.60**	0.008	0.004	**6.34**	**51.04**
	0.008	0.001	**0.35**	**−1.05**	0.008	0.003	**7.56**	**75.76**

Note. The basic sample is in the upper line of a cell, the selected sample of the impact strength values of the initial array is in the lower line of a cell.

** Chemical composition elements for which the differences in the form of the distribution of values of their content in the basic and selected samples are identified (the corresponding coefficients are highlighted in bold).*

Примечание. В верхней строке ячейки основная, в нижней – выделенная выборка значений ударной вязкости исходного массива.

** Элементы химического состава, для которых выявлены различия в виде распределения значений их содержания в основной и выделенной выборках (соответствующие коэффициенты выделены полужирным начертанием).*

is close to either the upper or lower limits of the sample of their distribution. The selection of the relevant data sub-arrays, from the production control database can be effective for the subsequent determination of the areas of change in the technology parameters with the dominant dependence type – corresponding, for example, to high and low (positive and negative signs, respectively) values of impact strength. For this purpose, a two-dimensional display of the areas of existence of objects in the form of a “cloud of points” on different planes of x_i–x_m parameters is useful [21]. If the cloud splits into two, then, obviously, there is some objective reason. Usually, the division of arrays is carried out either by selecting an equal number of left and right columns of the histogram of the distribution of the output parameter values (if the sample of n measurements is divided into the number of digits $n^{1/3}$ and the values in the extreme digit numbers are near-equal). However, the correspondence of the values in the left and right tails of the histogram is more often observed with a symmetrical distribution, and in the presence of more or less pronounced asymmetry, such a condition is rather difficult to fulfill. In this case, it is not easy to substantiate even the boundaries of the selection of the distribution “long” tail to identify subsequently the reasons for its formation. In this regard, the variant of considering it as a separate disturbance with a normal distribution proposed in the paper, can be useful as one of the methods of cognitive graphics used to separate data arrays.

Table 4 presents the statistical characteristics of the distributions of the content of the 13G1S-U sheet steel chemical composition elements corresponding to the impact strength (KCU^{-40} and KCV^0) values of the basic and selected samples (obtained by splitting the initial distribution histograms of the impact strength values with right-hand asymmetry according to the proposed procedure).

***Table 4.** Statistical characteristics of samples corresponding to the distribution of the content of the 13G1S-U steel chemical composition elements, which conform to the left and right tails of the distribution of the impact strength values (KCU^{-40} and KCV^0)*

***Таблица 4.** Статистические характеристики выборок, отвечающих распределению содержания элементов химического состава стали 13Г1С-У, соответствующих левым и правым хвостам распределения значений ударной вязкости (KCU^{-40} и KCV^0)*

Parameter	Relevant to the array of the KCU^{-40} values				Relevant to the array of the KCV^0 values			
	x_{av}, % wt.	σ, % wt.	A_s	E_x	x_{cp}, % wt.	σ, % wt.	A_s	E_x
C	0.13	0.01	−0.78	−0.33	0.14	0.01	−0.90	−0.12
	0.13	0.01	−0.60	−0.73	0.13	0.01	−0.83	−0.29
Si	0.48	0.07	−1.85	2.85	0.46	0.08	−1.25	−0.15
	0.46	0.09	−2.20	3.40	0.46	0.08	−2.45	4.65

Continue table 4

Parameter	Relevant to the array of the KCU^{-40} values				Relevant to the array of the KCV^{0} values			
	x_{av}, % wt.	σ, % wt.	A_s	E_x	x_{cp}, % wt.	σ, % wt.	A_s	E_x
Mn	1.42	0.05	1.35	2.02	1.44	0.08	1.12	−0.07
	1.40	0.04	−0.14	−0.69	1.41	0.04	0.09	0.24
P	0.015	0.003	0.77	−0.21	0.016	0.003	0.29	−0.73
	0.014	0.003	1.23	4.86	0.014	0.003	1.57	3.95
S*	0.007	0.004	1.50	2.50	0.008	0.005	0.98	**0.33**
	0.003	0.001	1.34	0.61	0.003	0.001	2.36	**6.57**
Cr	0.12	0.05	1.51	2.18	0.10	0.04	1.26	0.78
	0.11	0.04	1.34	3.04	0.11	0.05	1.38	1.58
Ni	0.09	0.02	0.69	0.34	0.09	0.02	1.30	1.37
	0.13	0.02	0.25	−0.49	0.08	0.01	0.93	0.90
Cu	0.13	0.02	0.62	0.31	0.13	0.02	0.40	0.73
	0.13	0.02	0.25	−0.49	0.12	0.02	0.10	−0.28
Ti*	0.007	0.007	1.94	2.26	0.010	0.010	1.05	**−0.81**
	0.006	0.005	1.82	1.56	0.006	0.005	2.59	**5.78**
Al	0.029	0.006	0.59	0.76	0.031	0.006	0.43	−0.61
	0.028	0.006	−0.75	−0.21	0.027	0.006	−0.19	0.18
V*	0.005	0.012	4.32	**18.29**	0.005	0.009	**4.99**	**29.27**
	0.008	0.016	2.40	**3.94**	0.007	0.016	**2.63**	**5.24**
N	0.009	0.002	0.57	−0.64	0.008	0.002	0.37	−1.15
	0.008	0.001	0.22	−0.42	0.008	0.001	0.05	−0.70
As	0.017	10^{-17}	−1.03	−2.09	0.017	10^{-17}	−1.04	−2.10
	0.017	10^{-17}	−1.03	−2.09	0.017	10^{-17}	−1.03	−2.09
Nb*	0.044	0.008	**2.40**	**5.43**	0.050	0.011	0.98	−0.47
	0.045	0.005	**0.17**	**−2.06**	0.044	0.005	0.46	−1.87
B	0.001	0.0002	−2.18	6.03	0.001	0.0003	−0.85	0.21
	0.001	0.0002	−2.04	2.42	0.001	0.0002	−2.47	4.54
Mo*	0.008	0.001	−0.24	−0.28	0.009	0.005	**5.68**	**35.79**
	0.008	0.002	0.47	−1.15	0.007	0.001	**0.79**	**−0.62**

Note. The sample of low values is in the upper line of a cell, the sample of high values of the impact strength of the initial arrays corresponding to left and right distribution tails is in the lower line of a cell.

** Chemical composition elements for which the differences in the form of the distribution of values of their content in the samples corresponding to the impact strength values at the lower and upper limits of initial histograms of their distribution are identified (the corresponding coefficients are highlighted in bold).*

Примечание. В верхней строке ячейки – выборка из низких значений, в нижней – высоких значений ударной вязкости исходных массивов, соответствующие левым и правым хвостам распределения.

** Элементы химического состава, для которых выявлены различия в виде распределения значений их содержания в выборках, соответствующих значениям ударной вязкости на нижнем и верхнем пределах исходных гистограмм их распределения (соответствующие коэффициенты выделены полужирным начертанием).*

From the results obtained, it is possible to identify the elements of the steel 13G1S-U chemical composition, in particular, niobium and vanadium, as well as molybdenum, vanadium, titanium, and sulfur, which determine the spread in the values of KCU^{-40} and KCV^{0} impact strength, respectively.

DISCUSSION

The low level of predicting when using regression models is primarily related to the difference in the distribution type of the initial samples of values of the technology control parameters, and the lack of a single space of process and product parameters. One of the main statistical criteria for accepting and rejecting a hypothesis about the regression coefficient significance is the risk of accepting the hypothesis, calculated using the Student's distribution. The desired Student's coefficient, in turn, depends on the standard error of the found regression coefficient. The standard error in the presence of an asymmetric or bimodal distribution of control parameter values is a characteristic describing the parameter with a "large margin", which brings the multiple regression coefficients into an insignificant region, due to which the coefficient corresponding to the parameter is removed from the best model. However, the work shows that variants of the asymmetric distribution of the values of control parameters, for example, the content of elements in 13G1S-U steel, are rather the norm than a deviation from it.

In this regard, one should not rely on predictive production control models based on regression analysis of the entire sample. Another approach, is the analysis of the distributions of technological parameters with predetermined boundaries: on the example of dividing the samples according to the criterion of normality, or the choice of obviously different areas using other methods of cognitive graphics.

The results obtained will be useful when developing IT solutions in metallurgy for forecasting and managing the quality of metal products (within the standard technology without changing it in essence).

CONCLUSIONS

The study showed the multiple regression inefficiency when identifying the technology parameters, limiting the 13G1S-U sheet steel impact strength inhomogeneity and discussed the reasons for this: the discrepancy between the distributions of the values of the technology control parameters and the normal type of distribution and the absence of a single space of process and product parameters. A procedure for selecting samples with an asymmetric nature of distribution histograms is proposed, which is important for an objective selection of technology parameters (or their combinations) that determine their appearance. This can be used as one of the methods of cognitive graphics for the subsequent selection of areas of change in technological parameters with a dominant dependence type.

REFERENCES

1. Vyboyshchik M.A., Ioffe A.V. Scientific basis of development and the methodology of creation of steels for the production of oilfield casing and tubular goods with the increased strength and corrosion resistance. *Vektor nauki Tolyattinskogo gosudarstvennogo universiteta*, 2019, no. 1, pp. 13–20. DOI: 10.18323/2073-5073-2019-1-13-20.
2. Vlasov S.A., Genkin A.L., Nikulina I.V., Koynov T.A., Shatalov R.L., Kravtsov S.V. Computer-integrated control of metallurgical complexes using forecasting simulation. *IFAC Proceedings Volumes*, 2009, vol. 42, no. 4, pp. 1832–1837. DOI: 10.3182/20090603-3-ru-2001.0321.
3. Kudrya A.V., Sokolovskaya E.A., Kodirov D.F., Bosov E.V., Timoshenko V.V. Possibilities of "data mining" of production control in metallurgy for prediction of strength, ductility and toughness of metal products. *Deformatsiya i razrushenie materialov*, 2023, no. S3, pp. 31–40. EDN: UBYQIQ.
4. Chaykin A.V. *Nauchnye osnovy innovatsionnykh tekhnologiy pechnoy i vnepechnoy obrabotki chugunov i staley dlya otlivok otvetstvennogo naznacheniya* [Scientific foundations of innovative technologies for furnace and out-of-furnace processing of cast irons and steels for critical castings]. Sankt Petersburg, Naukoemkie tekhnologii Publ., 2022. 245 p.
5. Wang G., Ledwoch A., Hasani R.M., Grosu R., Brintrup A. A generative neural network model for the quality prediction of work in progress products. *Applied Soft Computing*, 2019, vol. 85, article number 105683. DOI: 10.1016/j.asoc.2019.105683.
6. Kudrya A.V., Sokolovskaya E.A., Kodirov D.F., Bosov E.V., Kotishevskiy G.V. On necessity of taking into account statistical nature of the objects using Big Data in metallurgy. *CIS Iron and Steel Review*, 2022, no. 1, pp. 105–112. DOI: 10.17580/cisisr.2022.01.19.
7. Kudrya A.V., Sokolovskaya E.A., Trachenko V.A., Korotneva K.V. Contribution of the technological heredity into appearance of viscosity nonhomogeneity in constructional steel. *Vektor nauki Tolyattinskogo gosudarstvennogo universiteta*, 2013, no. 3, pp. 202–204. EDN: SBOZDD.
8. Zhukov P.I., Glushchenko A.I., Fomin A.V. Prediction model of temperature of cast billet based on its heating retrospection using boosting "random forest" structure. *Vestnik Novosibirskogo gosudarstvennogo universiteta. Seriya: Informatsionnye tekhnologii*, 2020, vol. 18, no. 4, pp. 11–27. DOI: 10.25205/1818-7900-2020-18-4-11-27.
9. Kolobov A.V., Malyshev K.V., Varfolomeev I.A., Ivanov D.A., Bogachev D.V. Effectiveness increasing of rolling mill operation due to prediction of uninterruptedly-casted block defects. *Proizvodstvo prokata*, 2017, no. 9, pp. 39–42. EDN: ZFPZKT.
10. Zenisek J., Gröning H., Wild N., Huskic A., Affenzeller M. Machine Learning based Data Stream Merging in Additive Manufacturing. *Procedia Computer Science*, 2022, vol. 200, pp. 1422–1431. DOI: 10.1016/j.procs.2022.01.343.
11. Jain R., Lee U., Samal S., Park N. Machine-Learning-Guided Phase Identification and Hardness Prediction of Al-Co-Cr-Fe-Mn-Nb-Ni-V Containing High Entropy Alloys. *Journal of Alloys and Compounds*, 2023, vol. 956, article number 170193. DOI: 10.1016/j.jallcom.2023.170193.

12. Narula P., Kumar P.A., Vanaja J., Prasad Reddy G.V., Rao G.V.S. Machine learning assisted prediction of creep data of India specific reduced activation ferritic martensitic steel. *Materials Today Communications*, 2023, vol. 35, article number 106165. DOI: 10.1016/j.mtcomm.2023.106165.
13. Doroganov V.S., Pimonov A.G. Methods of statistical analysis and neural network technologies for prediction of metallurgical coke quality. *Vestnik Kemerovskogo gosudarstvennogo universiteta*, 2014, no. 4-3, pp. 123–129. EDN: TELNMP.
14. Saoudi A., Fellah M., Hezil N., Lerari D., Khamouli F., Atoui L., Bachari K., Morozova J., Obrosov A., Samad M.A. Prediction of mechanical properties of welded steel X70 pipeline using neural network modelling. *International Journal of Pressure Vessels and Piping*, 2020, vol. 186, article number 104153. DOI: 10.1016/j.ijpvp.2020.104153.
15. Li C., Yin C., Xu X. Hybrid optimization assisted deep convolutional neural network for hardening prediction in steel. *Journal of King Saud University – Science*, 2021, vol. 33, no. 6, article number 101453. DOI: 10.1016/j.jksus.2021.101453.
16. Gao X., Qi K., Deng T., Qiu C., Zhou P., Du X. Application of Artificial Neural Network to Predicting Hardenability of Gear Steel. *Journal of Iron and Steel Research International*, 2006, vol. 13, no. 6, pp. 71–73. DOI: 10.1016/S1006-706X(06)60114-3.
17. Jiang L., Wang A., Tian N., Zhang W., Fan Q. BP Neural Network of Continuous Casting Technological Parameters and Secondary Dendrite Arm Spacing of Spring Steel. *Journal of Iron and Steel Research International*, 2011, vol. 18, no. 8, pp. 25–29. DOI: 10.1016/S1006-706X(11)60099-X.
18. De Oliveira Filho M.F., Caradec P.D.B., Calsaverini R., Spinelli J.E., Ishikawa T.T. Neural network for classification of MnS microinclusions in steels. *Journal of Materials Research and Technology*, 2023, vol. 24, pp. 8522–8532. DOI: 10.1016/j.jmrt.2023.05.101.
19. Gola J., Britz D., Staudt T., Winter M., Schneider A.S., Ludovici M., Mücklich F. Advanced microstructure classification by data mining methods. *Computational Materials Science*, 2018, vol. 148, pp. 324–335. DOI: 10.1016/j.commatsci.2018.03.004.
20. Gmurman V.E. *Teoriya veroyatnostey i matematicheskaya statistika* [Probability Theory and Mathematical Statistics]. Moscow, YuRAYT Publ., 2009. 478 p.
21. Kudrya A.V., Sokolovskaya E.A. Prediction of the Destruction of Materials with an Inhomogeneous Structures. *Fizika metallov i metallovedenie*, 2022, vol. 123, no. 12, pp. 1334–1346. EDN: DROJHB.

СПИСОК ЛИТЕРАТУРЫ

1. Выбойщик М.А., Иоффе А.В. Научные основы разработки и методология создания сталей для производства нефтепромысловых труб повышенной прочности и коррозионной стойкости // Вектор науки Тольяттинского государственного университета. 2019. № 1. С. 13–20. DOI: 10.18323/2073-5073-2019-1-13-20.
2. Vlasov S.A., Genkin A.L., Nikulina I.V., Koynov T.A., Shatalov R.L., Kravtsov S.V. Computer-integrated control of metallurgical complexes using forecasting simulation // IFAC Proceedings Volumes. 2009. Vol. 42. № 4. P. 1832–1837. DOI: 10.3182/20090603-3-ru-2001.0321.
3. Кудря А.В., Соколовская Э.А., Кодиров Д.Ф., Босов Е.В., Тимошенко В.В. Возможности «раскопок данных» производственного контроля в металлургии для прогноза прочности, пластичности и вязкости металлопродукции // Деформация и разрушение материалов. 2023. № S3. С. 31–40. EDN: UBYQIQ.
4. Чайкин А.В. Научные основы инновационных технологий печной и внепечной обработки чугунов и сталей для отливок ответственного назначения. СПб.: Наукоемкие технологии, 2022. 245 с.
5. Wang G., Ledwoch A., Hasani R.M., Grosu R., Brintrup A. A generative neural network model for the quality prediction of work in progress products // Applied Soft Computing. 2019. Vol. 85. Article number 105683. DOI: 10.1016/j.asoc.2019.105683.
6. Kudrya A.V., Sokolovskaya E.A., Kodirov D.F., Bosov E.V., Kotishevskiy G.V. On necessity of taking into account statistical nature of the objects using Big Data in metallurgy // CIS Iron and Steel Review. 2022. № 1. P. 105–112. DOI: 10.17580/cisisr.2022.01.19.
7. Кудря А.В., Соколовская Э.А., Траченко В.А., Коротнева К.В. Вклад технологической наследственности в появление неоднородности вязкости конструкционных сталей // Вектор науки Тольяттинского государственного университета. 2013. № 3. С. 202–204. EDN: SBOZDD.
8. Жуков П.И., Глущенко А.И., Фомин А.В. Модель для прогнозирования температуры заготовки по ретроспекции ее нагрева на основе бустинга структуры // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2020. Т. 18. № 4. С. 11–27. DOI: 10.25205/1818-7900-2020-18-4-11-27.
9. Колобов А.В., Малышев К.В., Варфоломеев И.А., Иванов Д.А., Богачев Д.В. Повышение эффективности работы прокатного стана за счет прогнозирования дефектов непрерывнолитой заготовки // Производство проката. 2017. № 9. С. 39–42. EDN: ZFPZKT.
10. Zenisek J., Gröning H., Wild N., Huskic A., Affenzeller M. Machine Learning based Data Stream Merging in Additive Manufacturing // Procedia Computer Science. 2022. Vol. 200. P. 1422–1431. DOI: 10.1016/j.procs.2022.01.343.
11. Jain R., Lee U., Samal S., Park N. Machine-Learning-Guided Phase Identification and Hardness Prediction of Al-Co-Cr-Fe-Mn-Nb-Ni-V Containing High Entropy Alloys // Journal of Alloys and Compounds. 2023. Vol. 956. Article number 170193. DOI: 10.1016/j.jallcom.2023.170193.
12. Narula P., Kumar P.A., Vanaja J., Prasad Reddy G.V., Rao G.V.S. Machine learning assisted prediction of creep data of India specific reduced activation ferritic martensitic steel // Materials Today Communications. 2023. Vol. 35. Article number 106165. DOI: 10.1016/j.mtcomm.2023.106165.

13. Дороганов В.С., Пимонов А.Г. Методы статистического анализа и нейросетевые технологии для прогнозирования показателей качества металлургического кокса // Вестник Кемеровского государственного университета. 2014. № 4-3. С. 123–129. EDN: TELNMP.
14. Saoudi A., Fellah M., Hezil N., Lerari D., Khamouli F., Atoui L., Bachari K., Morozova J., Obrosov A., Samad M.A. Prediction of mechanical properties of welded steel X70 pipeline using neural network modelling // International Journal of Pressure Vessels and Piping. 2020. Vol. 186. Article number 104153. DOI: 10.1016/j.ijpvp.2020.104153.
15. Li C., Yin C., Xu X. Hybrid optimization assisted deep convolutional neural network for hardening prediction in steel // Journal of King Saud University – Science. 2021. Vol. 33. № 6. Article number 101453. DOI: 10.1016/j.jksus.2021.101453.
16. Gao X., Qi K., Deng T., Qiu C., Zhou P., Du X. Application of Artificial Neural Network to Predicting Hardenability of Gear Steel // Journal of Iron and Steel Research International. 2006. Vol. 13. № 6. P. 71–73. DOI: 10.1016/S1006-706X(06)60114-3.
17. Jiang L., Wang A., Tian N., Zhang W., Fan Q. BP Neural Network of Continuous Casting Technological Parameters and Secondary Dendrite Arm Spacing of Spring Steel // Journal of Iron and Steel Research International. 2011. Vol. 18. № 8. P. 25–29. DOI: 10.1016/S1006-706X(11)60099-X.
18. De Oliveira Filho M.F., Caradec P.D.B., Calsaverini R., Spinelli J.E., Ishikawa T.T. Neural network for classification of MnS microinclusions in steels // Journal of Materials Research and Technology. 2023. Vol. 24. P. 8522–8532. DOI: 10.1016/j.jmrt.2023.05.101.
19. Gola J., Britz D., Staudt T., Winter M., Schneider A.S., Ludovici M., Mücklich F. Advanced microstructure classification by data mining methods // Computational Materials Science. 2018. Vol. 148. P. 324–335. DOI: 10.1016/j.commatsci.2018.03.004.
20. Гмурман В.Е. Теория вероятностей и математическая статистика. М.: ЮРАЙТ, 2009. 478 с.
21. Кудря А.В., Соколовская Э.А. Прогноз разрушения материалов с неоднородной структурой // Физика металлов и металловедение. 2022. Т. 123. № 12. С. 1334–1346. EDN: DROJHB.

О выборе областей с доминирующим типом зависимости при анализе данных производственного контроля

Тимошенко Виктория Владимировна[1], аспирант
Буданова Екатерина Сергеевна[2], магистрант
Кодиров Давронжон Фарходжон Угли[3], аспирант
Соколовская Элина Александровна[4], кандидат технических наук, доцент, доцент кафедры металловедения и физики прочности
Кудря Александр Викторович*[5], доктор технических наук, профессор, профессор кафедры металловедения и физики прочности
Университет науки и технологий МИСИС, Москва (Россия)

***E-mail: AVKudrya@misis.ru**

[1]ORCID: https://orcid.org/0009-0008-7671-0291
[2]ORCID: https://orcid.org/0009-0003-4055-9298
[3]ORCID: https://orcid.org/0009-0003-5380-5558
[4]ORCID: https://orcid.org/0000-0001-9381-9223
[5]ORCID: https://orcid.org/0000-0002-0339-2391

Поступила в редакцию 06.06.2023 *Принята к публикации 16.06.2023*

Аннотация: Формирование представительных баз данных определяет интерес к прогнозированию и управлению качеством металла на основе раскопок данных с использованием специальных программных продуктов, зачастую основанных на регрессионном анализе и не всегда учитывающих статистическую природу самого объекта исследования. Это может привести к ошибочной трактовке результатов или к неполноте извлекаемой информации, снижая эффективность статистической обработки. На основе анализа производственной базы данных технологии получения листовой стали 13Г1С-У были оценены возможности множественной линейной регрессии для прогноза качества листа. Показано, что глубина прогноза регрессии ограничена видом распределения значений управляющих параметров, характер распределения которых оценивали на основе определения коэффициентов асимметрии и эксцесса. В связи с сильным отклонением прогнозируемых моделей от экспериментальных значений в области правого хвоста распределений значений ударной вязкости, в данной работе были развиты методы разделения массивов данных и предложены критерии сравнения получаемых результатов. Для оценки корректности получаемых результатов из исходной выборки были выделены массивы с заведомо ассиметричным распределением, относительно которых также проведено сравнение статистических характеристик. На основе предлагаемых методов выявлены доминирующие химические элементы, которые вносят вклад в различие распределения значений приемо-сдаточных свойств, существующих в рамках одной и той же штатной технологии. Показано, что предложенный метод разделения может быть использован как вариация приемов когнитивной графики для выделения областей с доминирующим типом зависимости на основе соотношения коэффициентов асимметрии и эксцесса.

Ключевые слова: анализ данных производственного контроля; управление качеством металлопродукции; прогноз качества в металлургии; приемы когнитивной графики; раскопки производственных данных; сталь 13Г1С-У.

Для цитирования: Тимошенко В.В., Буданова Е.С., Кодиров Д.Ф., Соколовская Э.А., Кудря А.В. О выборе областей с доминирующим типом зависимости при анализе данных производственного контроля // Frontier Materials & Technologies. 2023. № 3. С. 103–114. DOI: 10.18323/2782-4039-2023-3-65-10.