

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

© 2023

*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

¹ORCID: <https://orcid.org/0009-0008-7671-0291>

²ORCID: <https://orcid.org/0009-0003-4055-9298>

³ORCID: <https://orcid.org/0009-0003-5380-5558>

⁴ORCID: <https://orcid.org/0000-0001-9381-9223>

⁵ORCID: <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.

ВВЕДЕНИЕ

Высокая оснащенность металлургических предприятий цифровыми средствами промежуточного контроля и сбора информации позволяет за короткий промежуток времени получать представительные базы данных производства продукции. Это определяет интерес к ретроспективному анализу таких массивов данных для выявления причин неоднородности качества металлопродукции и выработки обоснованных принципов управления им [1; 2]. При анализе такого рода информации реализуется пассивный эксперимент [3; 4], затраты на который оправдывают себя в рамках действующей, хорошо отлаженной технологии, когда высокое качество металла достигается для части продукции (обычно от половины ее объема и менее), и необходимо решить задачу по «подтягиванию отстающих» до передового уровня [5].

Традиционные подходы к решению таких задач основаны, в частности, на анализе стабильности параметров процесса и продукта, например, с использованием контрольных карт Шухарта¹. Их основной недостаток – в допущении нормальности распределений и единого пространства параметров процесса и продукта, что может

¹ГОСТ Р ИСО 7870-1-2022. *Статистические методы. Контрольные карты. Часть 1. Общие принципы*. М.: Российский институт стандартизации, 2022. 19 с.

ГОСТ Р ИСО 7870-2-2015. *Статистические методы. Контрольные карты. Часть 2. Контрольные карты Шухарта*. М.: Стандартинформ, 2016. 41 с.

ГОСТ Р ИСО 7870-3-2013. *Статистические методы. Контрольные карты. Часть 3. Приемочные контрольные карты (ISO/FDIS 7870-3). Разработка ГОСТ Р. М.: Стандартинформ, 2014. 18 с.*

привести к низкой результативности прогнозов [6]. В действительности более информативным будет глубокий ретроспективный анализ, выявляющий не только наличие прямых связей между значимыми управляющими параметрами и свойствами, но и их многосвязное влияние с учетом существующих областей с тем или иным типом зависимости, границы которых можно определить, например, с использованием приемов когнитивной графики [7].

В последнее время для решения задачи управления качеством металлопродукции (в т. ч. сквозного – от выплавки до приемо-сдаточных испытаний) все шире привлекаются алгоритмы машинного обучения. Среди них получили распространение методы на основе деревьев решений [8–12], нейронных сетей и т. п. [13–19]. Их использование позволило в ряде случаев создать модели, прогноз которых был выше по сравнению с обычной регрессией, например при разработке цифровой модели лазерной наплавки [9]. Алгоритмы случайного леса применялись также при прогнозировании структуры (в частности, межпластинчатого расстояния выделяющихся фаз) и механических свойств для сплавов с относительно короткой технологической цепочкой их получения [11; 12]. Применение методов на основе деревьев решений показало свою эффективность при прогнозировании последствий на отдельных стадиях производственного цикла.

Широкий интерес проявляется к программным решениям, построенным на основе нейронных сетей, например, для прогнозирования механических (при испытании на растяжение и ударных испытаниях [14], а также измерении твердости [15]) и технологических свойств (прокаливаемости [16]) в зависимости от химического состава. Кроме того, использование нейросетей апробировалось при производстве слэбов для предсказания литой структуры в зависимости от вариации значений технологических параметров производства [17]. Нейронные сети также применяли для прогноза дефектности металла в процессе производства [5]. Нейронные сети активно используются и при распознавании объектов структуры, например выявлении включений сульфидов [18] или классификации структурных составляющих [19]. Несмотря на их «самостоятельность» в работе, очевидно, что эффективность функционирования нейросетевых программных продуктов во многом будет определяться на стадии их обучения высококвалифицированными представителями экспертного сообщества.

В этой связи, в частности, актуально рассмотрение вопросов, связанных с уточнением роли статистической природы объектов в металлургии при применении как классических, так и разрабатываемых алгоритмов. Очевидно, что при этом также важно определение границ областей с доминирующим типом зависимости, закономерностей их взаимодействия, что позволит обосновать выбор корректных статистических процедур и критериев. Этим обстоятельствам не всегда уделяется необходимое внимание при разработке ИТ-решений применительно к металлургии, что, по крайней мере, следует из анализа соответствующих публикаций.

Цель работы – определение процедуры выделения выборок (при асимметричном характере гистограмм

распределения значений приемо-сдаточных характеристик листовой стали 13Г1С-У) для выявления областей изменения технологических параметров с доминирующим типом зависимости.

МЕТОДИКА ПРОВЕДЕНИЯ ИССЛЕДОВАНИЯ

В качестве объекта исследования была использована база данных производственного контроля процесса и продукта технологии получения проката из стали 13Г1С-У с толщиной листа 12 мм, состоящая из 1021 плавки. В качестве характеристик качества были выбраны значения ударной вязкости при температурах испытания 0 и -40 °С (KCV^0 и KCV^{-40} соответственно).

Регрессионный анализ производился с использованием множественной линейной регрессии. Выявляемыми параметрами являлись коэффициенты множественной регрессии b_i , стандартная ошибка коэффициентов регрессии σ_b , коэффициенты корреляции R и уровень значимости критерия Фишера F . Удаление незначимых коэффициентов производилось при равенстве коэффициента регрессии нулю в пределах ошибки и уровня риска по Стьюдену, равном 0,05.

Основными расчетными характеристиками выборок являлись: среднее значение x_{cp} , стандартная ошибка выборки σ , коэффициенты асимметрии A_s и эксцесса E_x .

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Анализ исходной выборки

На рис. 1 представлены гистограммы распределений значений ударной вязкости для листовой стали 13Г1С-У.

Имеющиеся выборки имеют асимметричное распределение, при этом коэффициенты асимметрии равны 1,44 и 1,26, а эксцесса – 2,53 и 2,38 для значений ударной вязкости KCV^0 и KCV^{-40} соответственно (при ошибке их определения для коэффициента асимметрии – 0,23 и для эксцесса – 0,76).

Были получены регрессионные модели для прогноза ударной вязкости относительно химического состава, характеристики которых (b_i – коэффициент регрессии; σ_b – стандартная ошибка коэффициента регрессии; R – коэффициент корреляции; F – уровень значимости критерия Фишера) представлены в таблице 1.

Полученные регрессионные модели являются значимыми исходя из уровня значимости критерия Фишера. На рис. 2 представлены графики, иллюстрирующие соответствие модели фактическим значениям.

Полученные модели предсказывают искомые механические свойства на низком уровне. При этом заметно явно выпадающие значения ударной вязкости в области высоких ее значений, которые соответствуют правым хвостам гистограмм, определяющим отклонение распределения в целом от нормального. Это отклонение может быть результатом совместного действия независимых возмущений примерно равной силы, что определяет природу его распределения (близкую к нормальному виду). Отсюда есть интерес к разбиению исходной выборки значений ударной вязкости на два подмассива.

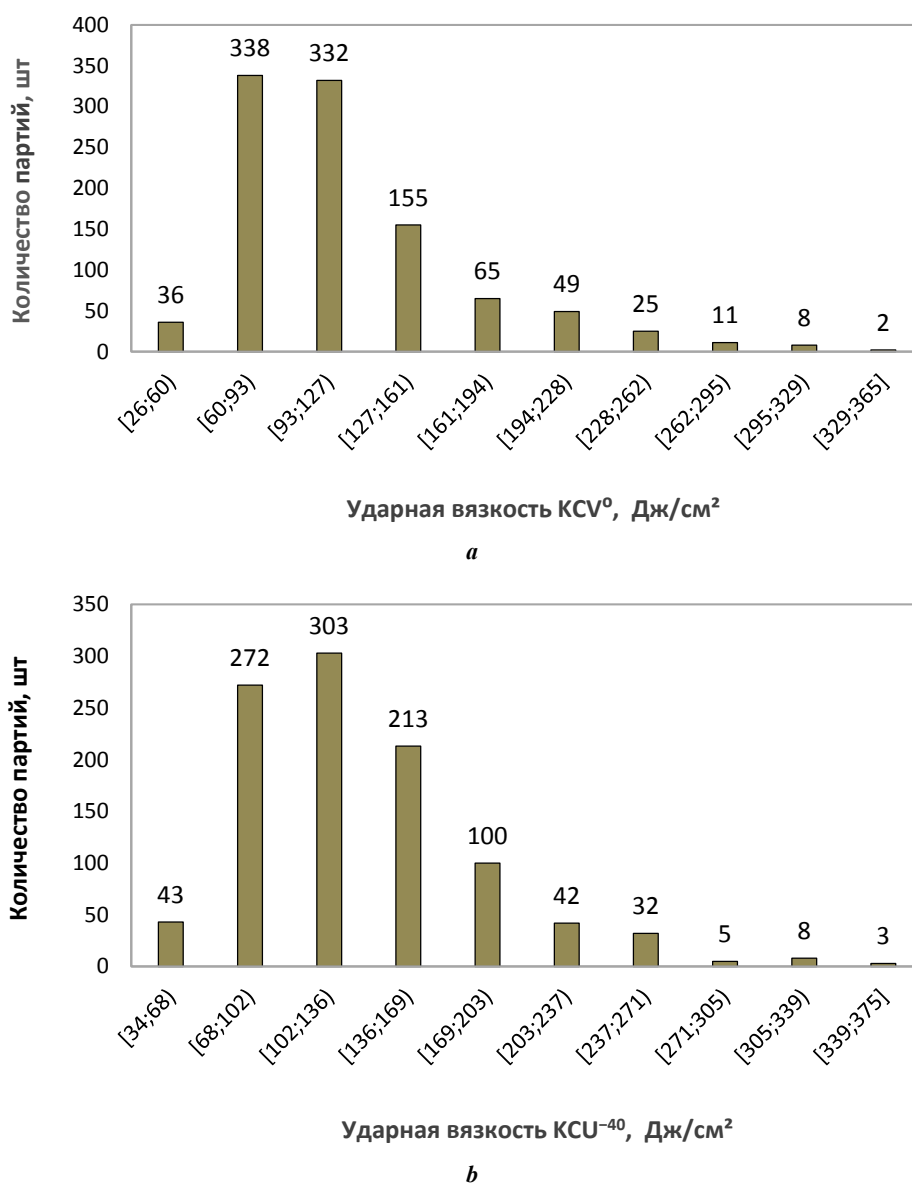


Рис. 1. Гистограммы распределений ударной вязкости стали 13Г1С-У:
а – температура испытания 0 °С, V-надрез;
б – температура испытания –40 °С, U-надрез
Fig. 1. Histograms of distributions of impact strength of 13G1S-U steel:
a – test temperature 0 °C, V-cut;
b – test temperature –40 °C, U-cut

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

Свойство	Параметры уравнения $\left(\frac{b_i}{\sigma_b}\right)$									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	2·10 ⁻⁶
KCV ⁰	–	–	–	–	-9412 521	–	-174 23	–	255 12	0,49	4·10 ⁻⁴

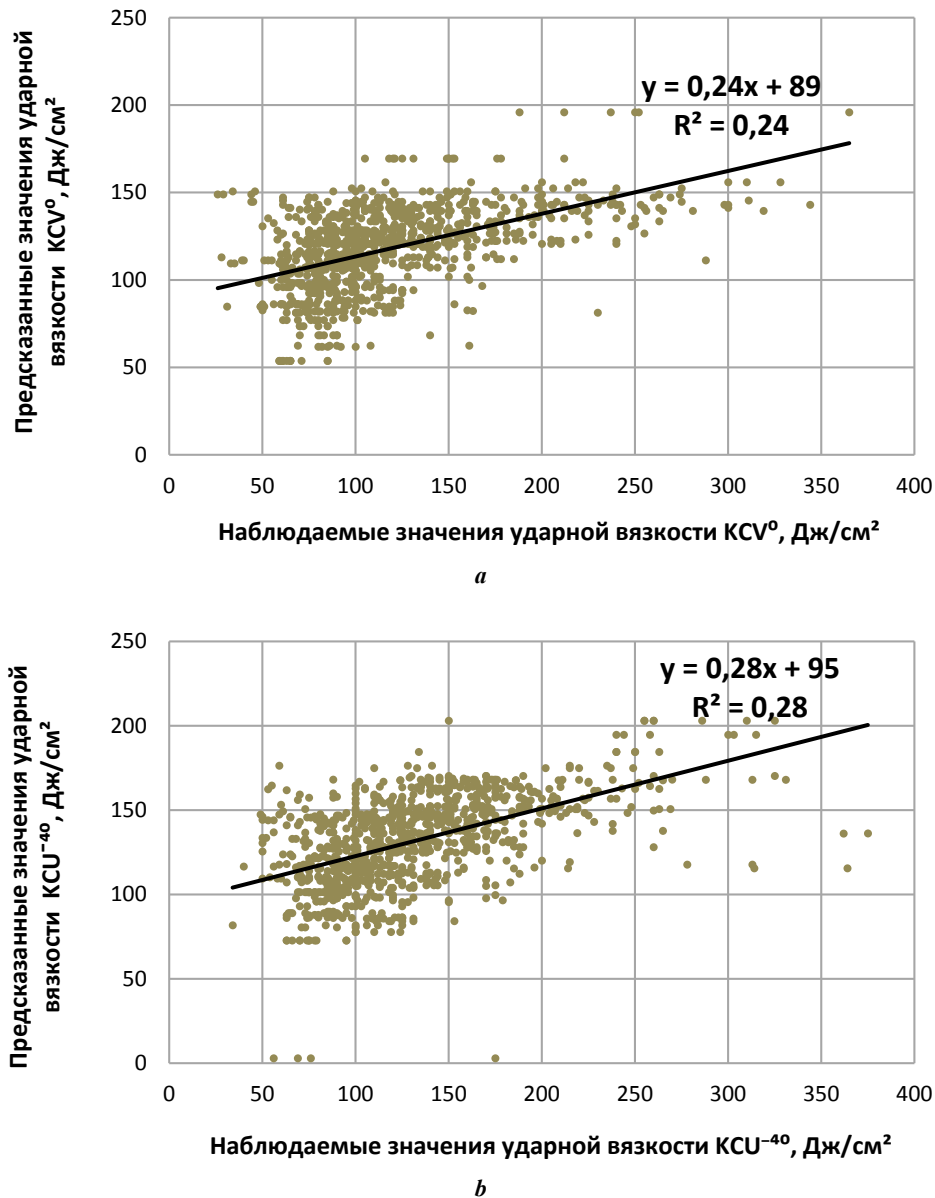


Рис. 2. Соответствие предсказанных и наблюдаемых значений ударной вязкости стали 13GIS-U: **a** – температура испытания 0 °С, V-надрез; **b** – температура испытания –40 °С, U-надрез
Fig. 2. Correspondence of predicted and observed values of impact strength of 13GIS-U steel: **a** – test temperature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

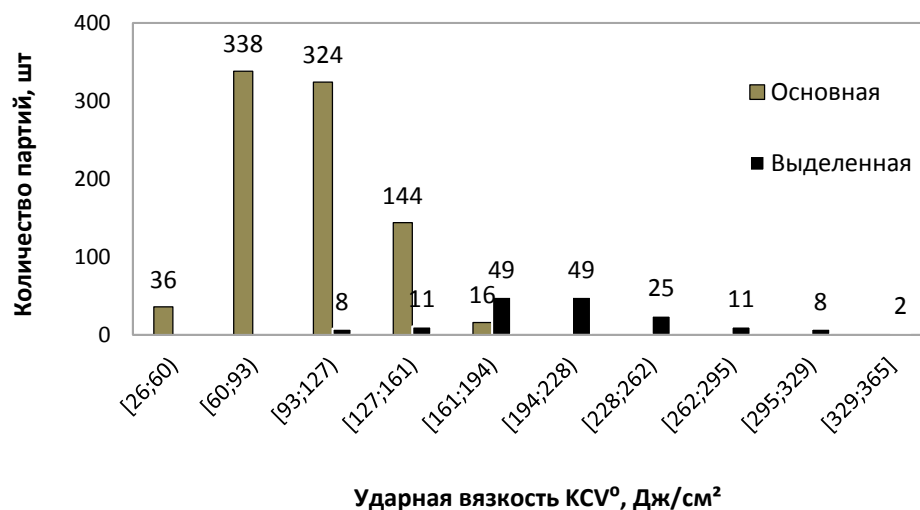
Разделение выборки на 2 массива данных

Исходные гистограммы значений ударной вязкости с наличием выраженной асимметрии разделяли на два распределения (с признаками нормального) на основе выделения наибольшей по численности выборки значений (основная выборка) как массива данных, в котором отсутствуют статистические выбросы, при условии максимально возможного приближения величин коэффициентов асимметрии A_s и эксцесса E_x к нулю [20]. Оставшиеся значения ударной вязкости (из правого хвоста распределения первичной выборки) выделялись в отдельный массив (выделенная выборка) и дополнялись значениями из основной выборки (выбранными случайным образом из правого хвоста ее гистограммы), также до достижения наиболее близких к нулю величин

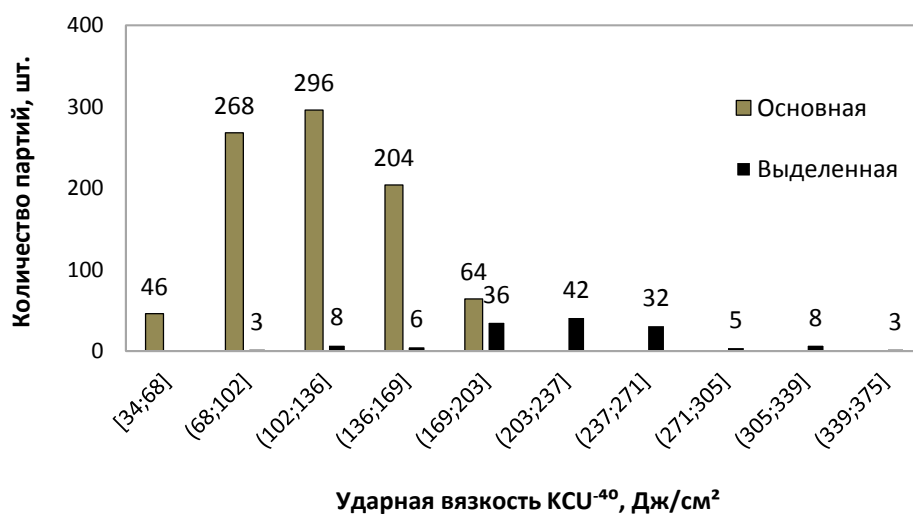
коэффициентов асимметрии A_s и эксцесса E_x (выбросы). Соответствующие двум сформированным выборкам ударной вязкости значения технологических параметров (химический состав) были также выделены в парные массивы данных.

На рис. 3 представлены диаграммы для разделенных массивов ударной вязкости, а в таблице 2 – статистические характеристики полученных распределений: средние значения x_{cp} , ошибка их определения σ , коэффициенты асимметрии и эксцесса. При этом количество значений в основной выборке и выбросах составило 878 и 143 партий для ударной вязкости KCU^{-40} и 858 и 163 – для ударной вязкости KCV^0 соответственно.

Исходя из внешнего вида диаграмм и значений коэффициентов асимметрии и эксцесса следует, что у полученных выборок распределение, близкое к нормальному.



a



b

Рис. 3. Диаграммы распределения значений ударной вязкости стали 13Г1С-У при разделении исходного массива на два подмассива:
a – температура испытания 0 °С, V-надрез; **b** – температура испытания –40 °С, U-надрез
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 °С, V-cut; **b** – test temperature –40 °С, U-cut

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

Свойство	Тип выборки	\bar{x} , Дж/см ²	σ , Дж/см ²	A_s	E_x
KCV ⁰	основная	101	29	0,25±0,25	-0,46±0,83
	выделенная	205	51	0,19±0,57	0,64±1,83
KCU ⁻⁴⁰	основная	118	34	0,24±0,25	-0,54±0,82
	выделенная	214	58	0,12±0,60	0,45±1,94

В таблице 3 представлены статистические характеристики выборок химического состава согласно разделению на подмассивы с нормальным распределением значений ударной вязкости KCU^{-40} и KCV^0 . Различия между видом распределения соответствующих выборок химического состава оценивались на основе сопоставления значений коэффициентов асимметрии и эксцесса для основной выборки и выделенной в пределах ошибки. Так, например, если разница между коэффициентами была больше суммы их ошибок, то это подтверждало справедливость гипотезы о различии выборок.

Следует отметить, что по результатам, представленным в таблице 3, нельзя выявить различий между выборками на основе средних значений содержания составляющих химического состава. Однако для ряда

массивов значений содержания химических элементов основной и выделенной выборок наблюдается различие в виде их распределения: величины их коэффициентов асимметрии и эксцесса значительно отличаются. Так, для двух выборок ударной вязкости при -40 °С выявлены различия в виде соответствующих распределений содержания серы, никеля, титана, ванадия, ниобия, бора и молибдена; для двух выборок ударной вязкости при 0 °С – для содержания никеля (только по величине коэффициента эксцесса), титана, ванадия, ниобия и молибдена. При этом в обоих случаях различие в распределении титана было выявлено только по коэффициенту эксцесса. Это означает, что именно эти параметры вносят наибольший вклад в различие нормально распределенных выборок, представленных на рис. 3.

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)

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)

Параметр	Применимо к массиву значений KCU^{-40}				Применимо к массиву значений KCV^0			
	$x_{ср}$, % мас.	σ , % мас.	A_s	E_x	$x_{ср}$, % мас.	σ , % мас.	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

Параметр	Применимо к массиву значений KCU ⁻⁴⁰				Применимо к массиву значений KCV ⁰			
	x_{cp} , % мас.	σ , % мас.	A_s	E_x	x_{cp} , % мас.	σ , % мас.	A_s	E_x
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	$2 \cdot 10^{-16}$	1,00	-2,00	0,017	$2 \cdot 10^{-16}$	1,00	-2,00
	0,017	$5 \cdot 10^{-17}$	1,01	-2,03	0,017	$6 \cdot 10^{-17}$	1,01	-2,03
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

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).

Выявление областей изменения значений управляющих параметров с доминирующим типом зависимости

Оценка влияния тех или иных параметров технологии на разброс качества, основанная на традиционном сопоставлении их средних значений по выборке, осложнена рядом факторов, например, вследствие отсутствия нормального распределения значений управляющих параметров. Регрессия также обычно малоэффективна вследствие отсутствия единого пространства параметров процесса и продукта. Очевидно, что влияние тех или иных параметров технологии (преимущественно совместное) на свойства металлопродукции, причем как положительное, так и отрицательное, наиболее сильно проявляется, когда уровень свойств близок либо к верхним, либо к нижним пределам выборки их распределения. Выделение соответствующих подмассивов данных из базы данных производственного контроля может быть эффективно для последующего определения областей изменения параметров технологии с доминирующим типом зависимости – соответствующих, например, высоким и низким (положительный и отрицательный признаки соответственно) значениям ударной вязкости. С этой целью полезно двумерное отображение областей существования объектов в виде «облака точек» на разных плоскостях параметров $x_l - x_m$ [21]. Если облако распалось на два, на то, очевидно, есть какая-то объективная причина. Обычно разделение массивов осуществляется либо выделением равного

числа левых и правых столбцов гистограммы распределения значений выходного параметра (при условии разбиения выборки из n измерений на число разрядов $n^{1/3}$ и примерном равенстве значений в крайних разрядах). Однако соответствие значений в левом и правом хвостах гистограммы чаще наблюдается при симметричном виде распределения, и при наличии более или менее выраженной асимметрии такое условие выполнить достаточно трудно. В таком случае непросто обосновать даже границы выделения «длинного» хвоста распределения для выявления в последующем причин его образования. В этой связи предлагаемый в работе вариант рассмотрения его как отдельного возмущения с нормальным видом распределения может быть полезным в качестве одного из приемов когнитивной графики, используемого для разделения массивов данных.

В таблице 4 представлены статистические характеристики распределений содержания элементов химического состава листовой стали 13Г1С-У, соответствующие значениям ударной вязкости (KCU⁻⁴⁰ и KCV⁰) основной и выделенной выборок (полученных разбиением исходных гистограмм распределения значений ударной вязкости с правосторонней асимметрией по предложенной процедуре).

Из полученных результатов можно выявить элементы химического состава стали 13Г1С-У, в частности ниобий и ванадий, а также молибден, ванадий, титан и сера, определяющие разброс значений ударной вязкости KCU⁻⁴⁰ и KCV⁰ соответственно.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Низкий уровень прогноза при использовании регрессионных моделей в первую очередь связан с различием в виде распределения исходных выборок значений управляющих параметров технологии и отсутствием единого пространства параметров процесса и продукта. Одним из основных статистических критериев принятия и отклонения гипотезы о значимости коэффициента регрессии является риск принятия гипотезы, рассчитанный с использованием распределения Стьюдента. Искомый коэффициент Стьюдента, в свою очередь, зависит от стандартной ошибки найденного коэффициента регрессии. Стандартная ошибка при наличии асимметричного или бимодального распределения значений управляющих параметров является характеристикой, которая описывает параметр с «большим запасом», что выводит коэффициенты множественной регрессии в незначимую область, бла-

годаря которой соответствующий параметру коэффициент удаляется из наилучшей модели. Однако, как показано в работе, варианты асимметричного распределения значений управляющих параметров, например содержания элементов в стали 13Г1С-У, скорее норма, чем отклонение от нее.

В связи с этим не следует полагаться на прогнозирующие модели производственного контроля, основанные на регрессионном анализе всей выборки. Иной подход – анализ распределений технологических параметров с заранее обозначенными границами: на примере разделения выборок по критерию нормальности или выбору заведомо различающихся областей с использованием иных приемов когнитивной графики.

Полученные результаты будут полезны при разработке IT-решений в металлургии для прогноза и управления качеством металлопродукции (в рамках штатной технологии без ее изменения по существу).

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)

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)

Параметр	Применительно к массиву значений KCU^{-40}				Применительно к массиву значений KCV^0			
	x_{cp} , % мас.	σ , % мас.	A_s	E_x	x_{cp} , % мас.	σ , % мас.	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
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

Параметр	Применительно к массиву значений KCU ⁻⁴⁰				Применительно к массиву значений KCV ⁰			
	x_{cp} , % мас.	σ , % мас.	A_s	E_x	x_{cp} , % мас.	σ , % мас.	A_s	E_x
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

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).

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).

ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Показана неэффективность множественной регрессии для выявления параметров технологии, лимитирующих неоднородность ударной вязкости листовой стали 13Г1С-У, и обсуждены причины этого: несоответствие распределений значений управляющих параметров технологии нормальному виду распределения и отсутствие единого пространства параметров процесса и продукта. Предложена процедура выделения выборок при асимметричном характере гистограмм распределения, что важно для объективного выделения параметров технологии (или их сочетаний), определяющих их появление. Это может быть использовано в качестве одного из приемов когнитивной графики для последующего выделения областей изменения технологических параметров с доминирующим типом зависимости.

СПИСОК ЛИТЕРАТУРЫ

1. Выбойщик М.А., Иоффе А.В. Научные основы разработки и методология создания сталей для произ-

водства 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](https://doi.org/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](https://doi.org/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](https://doi.org/10.26907/2542-0419.2023.S3.31-40).
4. Чайкин А.В. Научные основы инновационных технологий печной и внепечной обработки чугунов и сталей для отливок ответственного назначения. СПб.: Научно-технические технологии, 2022. 245 с.
5. Wang G., Ledwoch A., Hasani R.M., Grosu R., Brintrup A. A generative neural network model for the qua-

- lity prediction of work in progress products // *Applied Soft Computing*. 2019. Vol. 85. Article number 105683. DOI: [10.1016/j.asoc.2019.105683](https://doi.org/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/cisirs.2022.01.19](https://doi.org/10.17580/cisirs.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](https://www.edn.ru/SBOZDD).
 8. Жуков П.И., Глушенко 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](https://doi.org/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](https://www.edn.ru/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](https://doi.org/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](https://doi.org/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.matcomm.2023.106165](https://doi.org/10.1016/j.matcomm.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](https://www.edn.ru/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](https://doi.org/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](https://doi.org/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 Hardening 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://www.edn.ru/DROJHB).

REFERENCES

1. Vyboyschik 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](https://doi.org/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](https://doi.org/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](https://www.edn.ru/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](https://doi.org/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/cisirs.2022.01.19](https://doi.org/10.17580/cisirs.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. Saudi 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](#).

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

© 2023

Victoria V. Timoshenko¹, postgraduate studentEkaterina S. Budanova², graduate studentDavronjon F. Kodirov³, postgraduate studentElina A. Sokolovskaya⁴, PhD (Engineering), Associate Professor, assistant professor of Chair of Materials Science and Strength PhysicsAleksandr 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@misys.ru¹ORCID: <https://orcid.org/0009-0008-7671-0291>²ORCID: <https://orcid.org/0009-0003-4055-9298>³ORCID: <https://orcid.org/0009-0003-5380-5558>⁴ORCID: <https://orcid.org/0000-0001-9381-9223>⁵ORCID: <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.