On necessity of taking into account statistical nature of the objects using Big Data in metallurgy

A. V. Kudrya*, Dr. Eng., Prof.[1], e-mail: AVKudrya@misis.ru;*
E. A. Sokolovskaya*, Cand. Eng., Associate Prof.[1];*
D. F. Kodirov*, Post-graduate Student[1];*
E. V. Bosov*, Post-graduate Student[1];*
G. V. Kotishevskiy*, Advisor of General Director[2]*

[1]National University of Science and Technology (Moscow, Russia)
[2]"Novye tekhnologii kachestva" (Moscow, Russia)

Several statistical restrictions, which are critically important for correct use of different Big Data procedures in metallurgy for attestation and management of quality of metal products are evaluated. Representative production control data stores for steel manufacturing technologies are used as the research object. This research covered wide grade and dimension range of steels: large forgings of heat treatable 38KhN3MFA-Sh steel, rolled products of 40KhMFA steel, sheet 17G1S-U, 09G2S and 15KhSND steels. Possible scale of variety of values distribution both for managing parameters and characteristics of strength, plasticity and toughness is shown using the coefficients of asymmetry and kurtosis. These characteristics are varied within the technological tolerance range. Accompanying risk during metal quality prediction and management, e.g. using the methods of parametric statistics, was evaluated for the case when this circumstance was not taken into account. The features of influence of a sample list volume on the results of statistical processing of large production control data stores and metallurgical product are revealed. It is shown how absence of common space of parameters restricts possibilities of classic statistics in metallurgy, makes non-effective the management "by disturbance" principle. In this connection, possibilities of non-parametric statistics, presented by Kolmogorov – Smirnov criterion, which is not depended on distribution of collection of analyzed sample lists, are evaluated. To provide objective selection of the areas with dominating type of relationship, it is necessary to take into account possibility of existence of different evolution scenarios for structure and defects along the technological chain (technological heredity) within the framework of rather wide tolerance range, as well as features of their appearance. Difference in the evolution mechanisms of structures and defects within the framework of separate technological trajectory is a cause of appearances of developed heterogeneity for nominal single-type structures which have, however, different scales, as well as accompanying quality dispersion (which is often essential). Taking this circumstance into account allows to find out the links in the system "managing parameters – final parameters of metal products", which are not always evident during their search using generally accepted approaches. Development of the complex of rules for online management of metal products quality is possible on this base.
Key words: quality management of metal products, retrospective analysis of data bases for production control, Big Data, technological heredity, classic and non-parametric statistics, regression, kind of distribution of parameter values, dominating type of relationship.
DOI: 10.17580/cisisr.2022.01.19

Introduction

Rise of requirements to steel quality led to complication of technological processes for its production, including steelmaking, liquid metal casting, refining remelting, hot metal forming (rolling, forging) and heat treatment. Such technological stages as coke-chemical, sintering and blast furnace iron making are presented within integrated iron and steel works. A complex of testing facilities for objective evaluation of metal quality is also an inalienable part of the efforts aimed on manufacture of high-quality metal products. The list of quality parameters can be rather wide, it is determined by application area of metal products. Usually metallurgical production is rather well equipped by measuring devices and data collection systems along the whole technological line; however, owing to recent digitalization, possibilities for registration of information about technological process became essentially better. It allows not only to obtain complete observation about its realization, but also to create the conditions for precise tuning of the technology, i.e. online tuning. In this connection, the interest to statistical processing of large data stores of technological process and manufactured products is quite expectable. These data are accumulated rather quickly, taking into account serial (mass) features of manufacturing similar types of metal products [1-3]. But the examples of efficiency of such approaches are usually restricted within the frameworks of separate technological stages. The difficulties connected with their distribution for the whole technological cycle are characterized, in particular, with the fact that statistical nature of metallurgical objects (such as managing and acquisition parameters) is not always taken into account, even in the case of use op up-to-date Big Data algorithms.

In this connection, the aim of this work is formulated as evaluation of statistical nature of metallurgical objects and possibilities if its influence on application efficiency of statistical procedures during processing of big data stores, in order to prepare objective recommendations for metal products quality management,

Objects for investigation

Data bases for production control of the processes for manufacture of wide range of metal products (such as large forgings made of 38KhNMFA-Sh heat treatable steel, section rolled products of 40KhMFA heat treatable steel, sheet steels 17G1S-U, 09G2S, 15KhSND) were used as the objects for investigation within the framework of the technologies which were operated during different periods [1].

Data bases for production control presented the matrix $A_{m\times n}$, where the lines *m* corresponded to quantity of melts and batches, the columns n corresponded to the values of technological parameters (n_t) and quality characteristics of metal products (n_q). For the investigated data bases, describing the technological processes of manufacture of metal products made of 38KhNMFA-Sh, 40KhMFA, 17G1S-U, 09G2S, 15KhSND steels, quantity of the lines *m* in the matrixes made (with reference to chemical composition) 342, 166, 530 and 1460 as well as 1088 and 516 (melting in basic oxygen converter and electric arc furnace respectively), what presented the volume of similar products manufactured during 1-2 years. Quantity of line in the matrix, with reference to properties of metal products, can be larger, also taking into account possibility of testing of several samples at different temperatures. Quantity of columns n in the matrixes (n_t/n_q) made 91/20, 18/17, 84/15, 33/16 and 83/15 respectively. Statistical processing of the results was conducted using Excel, Statistica, Mathcad programs.

Tasks for analysis of production data

The existing practice of approaches to analysis of data bases for production control during manufacture of metal products is rather wide and various. However, if we shall not consider the problems connected with operating monitoring of controlled parameters for each kind of products in all key points, we can see that continuous control and reveal of deviations during technological process on the base of statistical methods (SPC) are mainly solved using managing of the technological process at the level of an assembly unit. E.g. it can be a digital twin, for example for a combined assembly unit “strip hot rolling mill – continuous pickling line”, using the model for scale growth to transfer predicted values of scale geometric parameters to a pickling unit [4]. The newest self-learning systems (algorithms) were widely used in metal processing for improvement of demand prediction accuracy. Prognostic models, which have been created on the base of these systems, were tested using archive data, and their consequent correction for additional prediction accuracy improvement is conducted on the base of obtained results [5]. Different physical models are used for prediction realization.

So, the company Hüttenwerke Krupp Mannesmann has developed the temperature model, which uses all important information for melt temperature evaluation and its prediction, with permanent actualization of calculation results on the base of continuous receiving the data on current temperature values [6]. This model provides precise temperature tuning in a basic oxygen convertor in the process of metal refining. To predict the temperature and carbon content in liquid steel in the end of blowing process (during basic oxygen practice), as well as consumption of oxygen and cooling additives during blowing, the results of actual temperature measurements and determination of carbon content in the samples taken after blowing finishing were used as a learning stores of neural network [7]. Artificial neural networks were also used for simulation of required mechanical properties of structural steels [8].

However, the complete analysis of production data can be considered as an end-to-end analysis of the whole complex of technological operations – from initial materials to finished products, in order to provide permanent quality management of metal products due to technological optimization within tolerance range for these products. Modern large-scale metallurgical production is a long chain of operations, where initial state of product and parameters of each technological stage can be basically measured and controlled. Their quantity *K* varies for different technologies and reaches together ~ 100 parameters and more. The tried and tested technology provides setting the allowable limits $\{\xi_k^{min}; \xi_k^{max}\}$ for each technological parameter; they determine a tolerance range. Quality of finished products is determined by ≥10 different quality parameters y_i, their allowable level d_i is preset by the requirements of regulating or contract documents ($y_i > d_i$ or $y_i < d_i$). The values of controlled and finished parameters are measuring in metallurgy for a melt, for a batch, for a coil, for a forging – actually for a piece of metal product. The complex *K* of actual values of $\{\xi\ kn\}$ parameters for the party of products presents the set of technological process trajectories; its evolution mechanisms of structure and defects (technological heredity) are realized within the framework of these processes. It leads to individual collection of quality parameters $\{y_{in}\}$. This collection of trajectories can be practically endless within the framework of wide technological tolerance range (even if it is well mastered), and taking into account quantity of controlled parameters. In any case, data bases in the researched works did not include even two coinciding trajectories (for the separate technology). Such difference in operating scenarios of technological heredity leads to developed heterogeneity of structures which are usually of the same type. It caused substantial spread in values of tensile strength ($\sigma_в$), plasticity ($\sigma_{0.2}$) and especially impact strength (KCU, KCV) at different testing temperatures, even in the conditions of stable production, which can be accompanied by essential statistical heterogeneity of the values of process and product parameters. It is evident that absence of their evaluation can make difficult selection of optimal statistical procedures for development of decisive rules directed on quality improvement of metal products during operating data bases of production control.

Table 1. **Heterogeneity scale of quality of metal products**

Steel	Type of product	Range of values Δ, (y_i^{max} - y_i^{min}) and average values within series						
		KCU^{+20}	KCV^{0}	KCU^{-40}	KCU^{-50}	KCU^{-60}	$\sigma_в$, MPa	δ, %
		J / cm²						
38KhN3MFA-Sh	forging	63–28*	-	-	58–20	-	1570–1340	17.5–8.8
		47.4 ±4.4			40.1 ±4.4		1478 ±24,7	14 ±0.9
40KhMFA	section	175–60	-	-	-	-	1400–1030	20–7
		100±19					1173.5±69.8	14.1±1.06
09G2S	sheet	438–80	-	-	-	430 – 78	590–465	38–28
		241.4 ±59.1				215.5 ±66.2	503.9 ±11.2	30.5±1.5
17G1S-U	sheet	-	375–28	388–25	-	-	660–409	37–16.5
			150.5 ±41.5	137.1 ±44.9			557±26.7	27.8±2.4
15KhSND	sheet	323–61	-	-	-	365–68	620–505	34–21
		168.8 ±27.9				160.1 ±27.1	565.1±13.4	27.1±1.4

* Numerators include range of values Δ, denominators — average value within series with error

Statistics of the values of the process and product parameters for the researched technologies

Dispersion of the properties, first of all of impact strength, which mainly determines the level of metallurgical quality [1], was typical for all examined types of metal products, obtained via investigated technologies (**Table 1**).

Such dispersion of metal quality should find its reflection in statistical nature of quality parameters. To evaluate this nature via coefficients of asymmetry A_s and kurtosis E_x, the features of their distribution was determined [9, 10]. If $A_s = 0$, then distribution is symmetric (as for normal distribution), if A_s (by module) is less than <0.25, asymmetry is considered as slight, within the range 0.25-0.5 — medium and more than 0.5 — essential. It is shown that the values of asymmetry coefficients As for tensile strength ($\sigma_в$) and impact strength (KCU, KCV) of steel grades 38KhNMFA-Sh, 40KhMFA, 17G1S-U, 09G2S, 15KhSND were varied within the ranges: 1.86–0.09; 0.43–0.87; 1.70–1.90; -0.06–(-0.37); 0.25–0.35 and -0.43–0.35; -0.62–0.63; 0.03–0.32; 0.93–1.07; 0.16–0.50 respectively. Analyzing the value of kurtosis coefficient E_x, peak sharpness degree of distribution was evaluated, based on understanding that empiric distribution id more high ("sharp-peaked") at $E_x > 0$, relating to normal distribution (E_x=0), while at $E_x < 0$ it is more low and slightly sloping. The values of Ex for tensile strength and impact strength of the steels varied for the examined data bases within the following ranges: 0.66–13.58; (-0.54)–0.10; 6.82–7.77; 0.67–1.19; 0.01–0.95; and (-0.14)–0.80; 1.41–3.26; -0.91–(-0.42); 0.73–1.86; 0.31–0.54 respectively.

These distribution characteristics and large dispersion of impact strength values (twofold and more) evidently testify that at least half of products has level of the properties above medium and can exceed this level substantially. It means that the considered technologies in general can provide manufacture of high quality metal products; it makes expedient "data excavating" from the plant technical control service (the passive experiment in *K*-dimensional space of the process parameters ξ_k) in order to reveal the causes of quality dispersion and improvement of its homogeneity at the upper level of distribution. Restriction of $\{\xi_k^{min} ; \xi_k^{max}\}$ for several parameters of ξ_k process (without principal varying of the technology) can be a problem solution in such case.

Possibilities of digitalization provide continuous and complete process and product control, monitoring of production parameters with reference to each melt, batch, forging or coil. Continuous archive storing of detailed information on process and product parameters, quickness of accumulation of representative data stores, which are sufficient for objective analysis about causes of metal products quality heterogeneity, make prospective in this connection use of Big Data tools (non-uniform statistical procedures, additional development of "data excavating" algorithms) [11, 12].

It is evident that heterogeneity of product statistical nature in particular is a consequence of variety of statistical nature of controlled parameters. It is not always taken into account practically, using different approaches to processing of data stores of production control in metallurgy during end-to-end analysis of the technology (from metal melting to quality evaluation of metal products). It can finalize in obtaining the mixed results.

E.g., if we determine differences between two series of measurement results (volumes n_1, and n_2 respectively) for any characteristic, it is usually suggested that "histogram of the process (or product) parameters data should have the form of single-modal dome-shaped curve" [13]. Based on this, significance of the difference $[\langle\xi\rangle_1 - \langle\xi\rangle_2] \neq 0$ is usually evaluated by the average values $\langle\xi\rangle_1$ and $\langle\xi\rangle_2$ and their mean-square deviations according to the Student criterion; thereby normal distribution ξ is allowed. It can be followed, e.g., by existence of negative impact strength or content of any element included in steel composition [14]. Normal distribution means free forming of distribution "tails", however they are cut for the process parameters in correspondence with the requirements of technological regulations. The values of E_x and A_s coefficients (if not equal to zero, but at least close to this value) should correspond to suggested single-modal distribution [9, 10].

Quantity m of histograms is determined as $m \sim \sqrt[3]{n}$ for uniform distribution of the measuring value for decades,

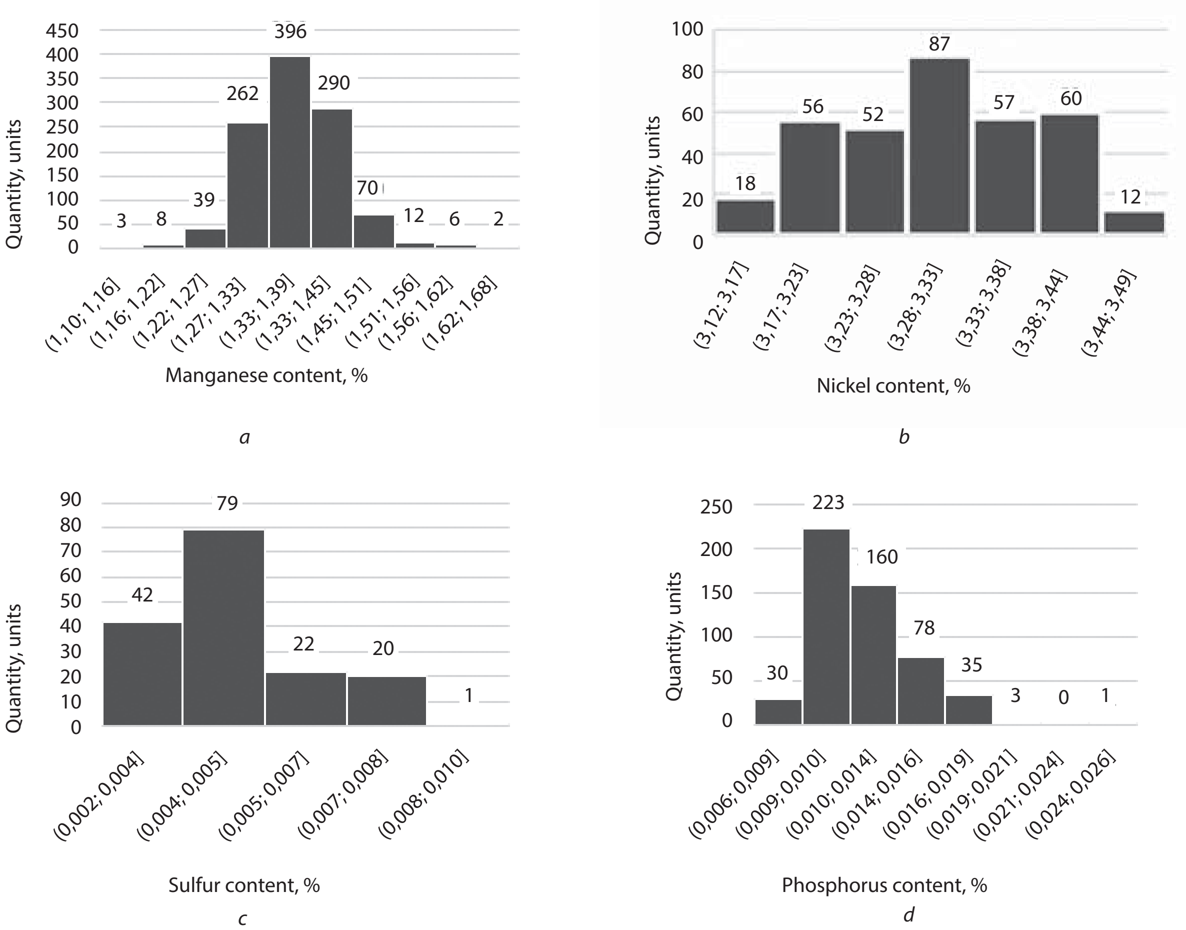

Fig. 1. Distribution of the values of manganese content in sheet steel 17G1S-U (*a*), of nickel in forged steel 38KhN3MFA-Sh (*b*), of sulfur in section steel 40KhMFA (*c*) and of phosphorus in sheet steel 09G2S (*d*)

Table 2. **The values of kurtosis E_x and asymmetry A_s distribution coefficients for elements content in chemical composition of researched steels**

Steel	A_s/E_x									
	C	Si	Mn	P	S	Cr	Ni	Cu	Ti	Al
38KhN3MFA-Sh	0.42	1.25	-0.45	0.33	0.80	0.34	-0.13	0.95	-	-
	0.94	3.08	0.09	-0.52	0.69	0.14	-0.85	2.69	-	-
40KhMFA	0.25	0.44	1.53	0.68	0.79	-1.00	1.62	0.45	2.36	-0.07
	1.07	-0.02	4.39	0.72	0.53	8.19	3.59	0.79	5.21	0.24
09G2S	0.50	-0.09	-1.42	0.74	1.14	2.88	2.39	1.59	-	0.65
	0.78	0.17	2.28	2.14	1.10	9.09	8.69	11.30	-	0.61
17G1S-U	0.09	-1.10	0.09	0.87	1.38	1.58	2.39	1.24	2.44	-0.12
	-0.35	3.38	1.87	1.26	4.72	4.63	11.04	2.97	7.38	0.44
15KhSND	-0.005	0.20	0.67	1.24	0.90	-0.46	0.16	0.08	-	0.24
	0.12	0.75	4.37	2.31	0.72	1.59	2.97	4.04	-	-0.29

accompanies by achieving the minimal mean-square deviation of revealed distribution in comparison with the true one [15]. However, building of these histograms, based on this condition, displays that their appearance can differ substantially from symmetric single-modal distribution, and bimodality can be often observed (**Fig. 1**).

The wide range of variation of the coefficients E_x и A_s was typical for examined distribution of the values of technological parameters, such as chemical composition (**Table 2**).

Such variation of the values of distribution statistical characteristics for process parameters means that use of average values within the series and their dispersions σ for evaluation of heterogeneity level of some events becomes problematic, what restricts possibilities of application of classic statistical criteria. Ignoring the distribution kind can

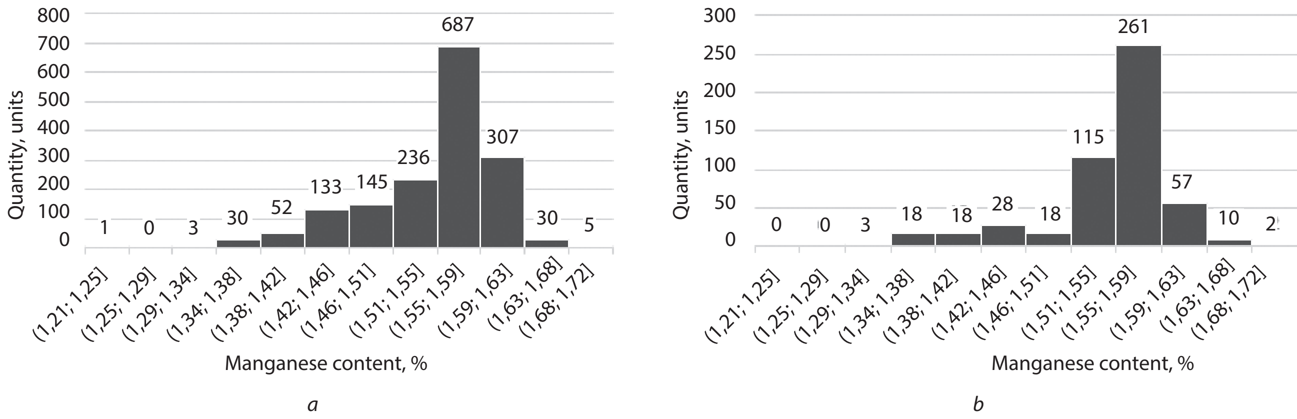

Fig. 2. Distribution histograms of manganese content in 09G2S steel, produced via electric arc melting (*a*) or in basic oxygen converter (*b*)

lead to contradictory results. E.g. if we compare manganese content in sheet steel 09G2S, which was produced via electric arc melting or in basic oxygen converter (**Fig. 2**), their difference was evaluated with high risk 0.35 according to the Student criterion and only 0.05 according to the Kolmogorov-Smirnov non-parametric criterion (which does not depend on distribution kind [16]). This difference can't be "improved" due to increase of the series volume. Metallurgical "party" is usually a batch; quantity N of these batches with the same steel grade and dimension range is restricted (N<1000-1500), then either technology, or order can be varied. Therefore, the risk connected with non-objective evaluation of difference between two series of controlled parameters can be rather high, if their distribution kind won't be taken into account.

Application of the evaluation method for random repeatability error "in the point" by the range $\Delta = \xi_{ik}^{max} - \xi_{ik}^{min}$, or by variation coefficient σ/ξ_{ik} (oscillation Δ/ξ_{ik}) can be not very efficient. Usually range elevates with increase of measurements. However, the series presents at the same time also chronological (temporal, with reference to a date) group of measurements of the same value. Therefore different causes of range increase (dispersion) are possible, it can be not only variations of parameters with different amplitude relating to the average value (within the technological tolerance range), but also consequence of their trend effected by any cause; appearance of "season" oscillations (periodical results of parameters measurements) is also possible (**Fig. 3**). It should be noted that the same causes will also "distort" the nature of average values within the series, which play the key role in many statistical procedures.

Even in the case of trend absence or periodical oscillations of measuring technological parameters, the principle "more measured – more found" does not work relating to ranges in metallurgy (probably, also in many other production areas). It is evident that $\lim_{i}\rightarrow\infty(\xi_{ik}^{max} - \xi_{ik}^{min}) = \{\xi_k^{min}; \xi_k^{max}\}$, what corresponds to a tolerance range of mastered, "directive" technology (consequences of its exceeding, such as technological violations, are not discussed in this article). This statement corresponds to observation results of variation regularities for ranges Δ of parameters with consequent increase of observation series volume (in the framework of a chronological row) (see Fig. 3). It is important that reaching the ultimate value (within observation scale) can occur in different ways. For example, reaching the maximal value of heating temperature for 38KhN3MFA-Sh steel billet for forging in one of the furnace sections corresponded to No. 329 in the chronological series row which includes 342 forgings, while ultimate phosphorus content was reached in the same row already in the forging No. 74. It means that not only the range value Δ of parameters causes interest for evaluation of technological heterogeneity scale, but also periodical character (time) of its appearance, while the range nature can be expressed as relationship of the values ξ_{ik}^{min} and ξ_{ik}^{max} with the average value $<\xi_{ik}>$ within the series. It is evident that reveal of such factors will be complicated in the framework of the same standard regression, as well as evaluation of their effect on its results.

Search of the links between variations of controlled and finished parameters suggests indirectly that relationship between yk and technological parameters is common in the whole area of arguments existence. In this case, any correct regression equation provides the required boundaries of the optimum area, what determines its wide application. However, in metallurgy the field $\{x_k\}$ usually presents combination of interacting areas, each of them with its dominating type of physical relationship; it makes difficult then to use polynomes, which hardly approximate piece relationships [10]. If interaction of K factors is observed in this case, it is required to find $\sim K^2/2$ of "cross" terms $c_{ks}x_kx_s$ of the regression equation; usually it is less than the quantity of points $N<K^2/2$, which is practically accepted. As a result, dispersions of regression coefficients cks will become endless.

Thereby, the attempts of metal quality management "by disturbance" are usually non-effective, even in the cases when the effect of such "evergreen" quality factors as sulfur and phosphorus content is evaluated. In particular, lowering of phosphorus content (separation of the values of impact

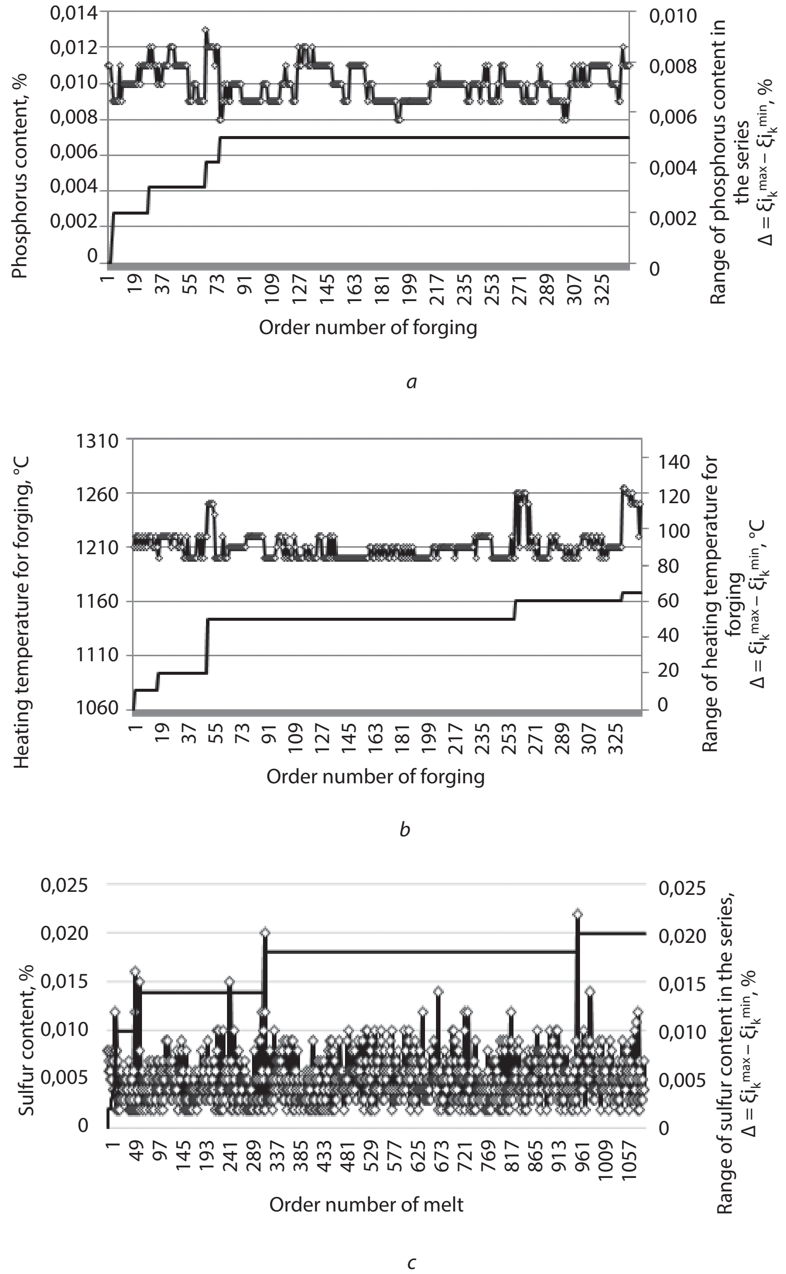

Fig. 3. Chronological sequence of the values of phosphorus content in forgings of steel 38KhN3MFA-Sh (*a*), their heating temperature for forging (*b*) and sulfur content in steel 17G1S-U (*c*), and corresponding relationships of variation of their ranges with increase of the series at elevation of the order number in a chronological row

strength corresponding to increase content of impurities) in the investigated steel making technologies for the grades 38KHN3MFA-Sh and 15KhSND does not lead to displacement of histograms peaks of distribution of impact strength values in the area of higher values (**Fig. 4**). The series included average values of impact strength relatively. Actually the initial variation range of impact strength values was saved for all histograms; moreover, quantity of parties with minimal impact strength values decreased and metal parties with maximal impact strength values "dropped".

These circumstances also should be taken into account in the case of application of neural nets as a remedy for analysis of complicated non-linear relationships [7, 19, 20]. Transforming a multi-dimensional system to a function of several hidden variables, these circumstances shows the limits of possible decrease of searching space dimensionality {ξ}; however, type of these variables and their relationships leaves "inside the program", so we have a solving "black box" at the exit. Taking into account the statistical nature of objects can promote learning optimization of neural nets and efficiency rise of their operation in general. Such approach can also be useful in application of actively developing fuzzy logic algorithms, hybrid use of "soft calculations" [17-20]. , which suggest that determined non-adaptive state function not always allows to provide required system operation, while fuzzy rules are based essentially on expert's experience. It can be substantial for development of computer-aided learning methods, e.g. for use of logistic regression algorithms, support vector machines, random forest etc [21]. Importance of taking into account statistical nature of metallurgical objects has already shown its efficiency in using regression for narrowing space for parameters or application of complicated heuristic techniques of cognitive graphics. They can be used for searching the areas with dominating type of relationship during reveal of interacting deviations of technological parameters, which are related to different technological stages, use of schemes of non-parametric discrimination, not depending on kind of distribution (and more robust thereby) [12, 13].

Efficiency of use of through and comprehensive analysis of a long chain of technological operations along the whole production cycle (from initial materials to finished products), aimed on providing of continuous quality management, reveal of technological "bottlenecks", optimization of processes and products, is realized not only by supply of required software packages. Such analysis with use of a wide set of substantiated statistical procedures is activity for a professional technologist, not for statistical specialist, because it is based on the competence to formulate hypotheses about chains of events causes. These hypotheses are explained taking into account diversity of behaviour scenarios for technological heredity (within the framework of tolerance range of the concrete technology) and can be principally checked on the base of programs from ready packages.

It is evident that success in creation of intelligent systems for quality management will be strongly determined by the progress in measurement digitalization for structures and fractures in the scale of a sample or product, taking into account their statistical nature [22]. Quantitative evaluations of structures and destruction will provide more complete and objective differentiation of metal products quality and reveal the structural causes of its heterogeneity, what is principally important for realization of the quality management principle by structure, i.e. using Big Data algorithms.

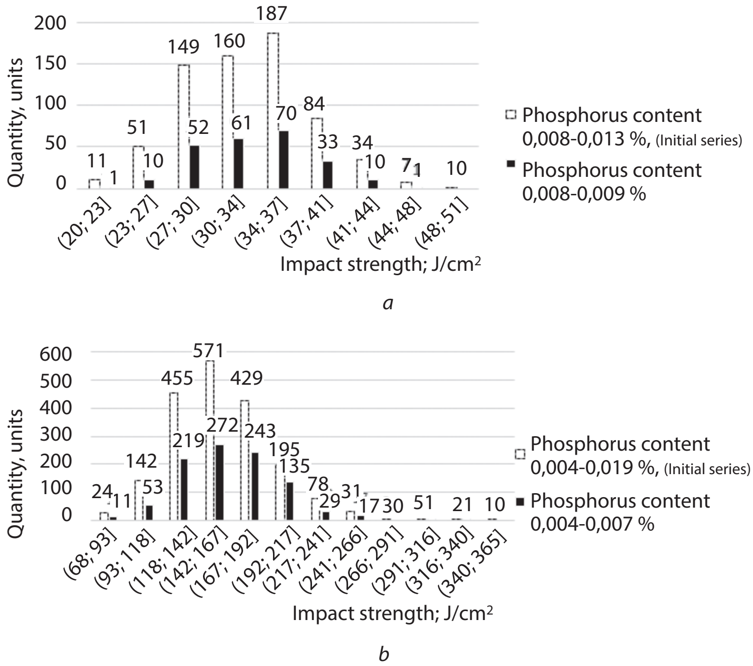

Fig. 4. Distribution of impact strength values KCU-50 in steel forgings of grades 38KHN3MFA-Sh (*a*) and KCU-60 in sheet steel 15KhSND (*b*) in correspondence to initial series and series with lowered phosphorus content

Conclusion

Statistics of distribution of controlled and finished parameters was evaluated on the base of systematic analysis of representative data stores of production control in the process of manufacture of wide grade and dimension range of steels: large forgings of heat treatable 38KhN3MFA-Sh steel, rolled products of 40KhMFA steel, sheet 17G1S-U, 09G2S and 15KhSND steels, within the framework of the technologies operating during different time periods of the technologies. It is characterized by variety of distribution kinds for parameters values (symmetric, asymmetric and bimodal) and wide variation range of the values of their asymmetry and kurtosis coefficients for input and exit parameters (from -1.42 to 2.88 and from -1.86 to 2.46; from -0.85 to 11.04 and from -1.14 to 14.57 respectively). The tolerance ranges which were established for this process, restrict variation of parameters values and possibility of free forming of distribution tails. It can complicate use of many Big Data algorithms based on assumption about normal (symmetric) distribution type for the values, as well as corresponding parameters and criteria of classic statistics. In this connection, we prefer evaluations using criteria of non-parametric statistics, i.e. Kolmogorov-Smirnov criterion.

Accumulation of data bases in metallurgy occurs within the framework of chronological (temporal) row of events, which can be accompanied by appearance of trends, season oscillations. It can violate probabilistic feature of the events which are considered during statistical processing and, together with restrictions introduced by technological tolerance range, decrease efficiency of several statistical parameters. Among them the following parameters can be mentioned: range, variation (oscillation) coefficients, which use for evaluation of heterogeneity scale of process and products parameters during comparison of different observation series.

Absence of the common space for technological parameters in metallurgy, combined with restrictions which are introduced by differences of distribution statistics for values of controlled parameters, make low efficient use of "by disturbance" management principle. This circumstance is essential for selection of optimal procedures for Big Data algorithms in the case of development of intelligent systems with through quality management of metal products (from charge materials to finished products). Efficiency of such systems will be determined first of all by achieved understanding level for behaviour regularity of technological heredity within the framework of the concrete manufacturing process of products on the base of the hypotheses about chains of events causes and their checking using optimal statistical procedures.

REFERENCES

1. Steel on the threshold of centuries. Edited by Yu.S. Karabasov. Moscow. MISiS. 2001. Pp. 445-543.
2. Manwendra K.cTripathia, Randhir Kumarb, Rakesh Tripathib. Big-data driven approaches in materials science: A survey. *Materials Today: Proceedings.* 2020. Vol. 26. Part 2. pp. 1245-1249.
3. Shun Guo Jinxin Yu, Xingjun Liu, Cuiping Wang, Qingshan Jiang. A predicting model for properties of steel using the industrial big data based on machine learning. *Computational Materials Science.* 2019. Vol. 160. April. pp. 95-104.
4. Neuer M., Ebel A., Brandenburger J., Polzer J., Wolff A., Loos M., Holzknecht N., Peters H. Digital technologies in ferrous metallurgy. *Chernye metally.* 2019. No. 3. pp. 54-58.
5. Schuster R., Veugt N., Nath G., Louw N. Possibilities of digital technologies for transformation of value chains in metallurgy and metalworking. *Chernye metally.* 2019. No. 3. pp. 59-61.
6. Lindner Ch., Raschewski F., Weinberg M. New opportunities in quality control due to the progress in the control of technological parameters at HKM. *Chernye metally.* 2018. No. 1. pp. 63-69.

7. Cox I. J., Lewis R. W., Ransing R. S., Laxzczewski H., Berni G. Application of neural computing in basic oxygen steelmaking. *J. Mat. Processing Technol.* 2002. Jan. 15. pp. 310-315.
8. Dobrzanski L. A., Honysz R. Application of artificial neural networks in modeling of normalized structural steels mechanical properties. *Journal of Achievements in Materials and Manufacturing Engineering.* 2009. Vol. 32. Iss. 1. pp. 37–45.
9. Melnichenko A. S. Statistical analysis in metallurgy and materials science. A textbook. Moscow. Izdatelskiy dom "MISiS". 2009. 268 p.
10. Gmurman V. E. Probability theory and mathematical statistics. A textbook for high schools. Moscow. Vysshaya shkola. 2003. 479 p.
11. Chubukova I. A. Data mining. Moscow. Binom. 2006. 384 p.
12. Shtremel M. A., Kudrya A. V., Ivashchenko A. V. Non-parametric discriminant analysis in quality management tasks. *Zavodskaya laboratoriya*. 2006. No. 5. p. 53.
13. Kudrya A. V., Shtremel M. A. On the reliability of data analysis in quality control. *Metal Science and Heat Treatment.* 2010. Vol. 52. No. 7-8. pp. 341-346.
14. Shtremel M. A. Information content of impact strength measurements. *Metallovedenie i termicheskaya obrabotka metallov.* 2008. No. 11. p. 37.
15. Chentsov N. N. Statistical decisive rules and optimal conclusions. Moscow. Nauka. 1972. 820 p.
16. Nikitin Ya. Yu. Asymptotic efficiency of non-parametric criteria. Moscow. Fizmatlit. 1995. 240 p.
17. Jia C. Self-adaptive flat control based on artificial intelligence. *Gangtie Yanjiu Xuebao (J. of Iron and Steel Research).* 2001. No. 13 (July-August). pp. 58-61.
18. Im Y.-T., Jung J.-Y. Fuzzy control algorithm for the prediction of tension variations in hot rolling. *J. Mat. Processing Technol.* 1999. pp. 163-172.
19. Honarmandi P., Arróyave R. R. Uncertainty Quantification and Propagation in Computational Materials Science and Simulation-Assisted Materials Design. *Integrating Materials and Manufacturing Innovation.* 2020. Vol. 9. pp. 103-143.
20. Yong Liu, Jing-chuan Zhu, Yong Cao. Modeling effects of alloying elements and heat treatment parameters on mechanical properties of hot die steel with back-propagation artificial neural network. *Journal of Iron and Steel Research International.* 2017. Vol. 24. pp. 1254–1260.
21. Hastie, T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Chapter 15. Random Forests. 2nd edition. Springer-Verlag. 2009. 746 p.
22. Kudrya A. V., Sokolovskaya E. A, Perezhogin V. Yu., Kodirov D. F. On Taking into Account the Statistical Nature of Objects in Structural Analysis in Metals Science. *Russian Metallurgy (Metally).* 2020. Vol. 12. pp. 1435–1438.