CSITA
ISSN 2414-9055

UDC 622.723: 004.275 https://doi.org/10.31721/2414-9055.2017.3.1.22

RECOGNITION OF CLUSTER'S OPTIMAL QUANTITY IN CASE OF GUSTAFSON-KESSEL'S METHOD USAGE TO DEFINE TECHNOLOGICAL VARIETY OF IRON ORE

Tykhanskyi M.P., PhD, Associated Professor, Tsvirkun S.L., PhD, Associated Professor
Kryvyi Rih National University

Abstract. It was considered the problem of cluster analysis in condition of equivocation to forming automatic control process sorting large-size iron ore raw material with operating recognition of its separate pieces technological variety in stream on conveying belt. The most preferred algorithms are algorithms, which set quantity of clusters by themselves, or self-organization algorithms, one of which is Gustafson-Kessel's algorithm. Evaluation test of clusterization was performed with using scalar measures of reliability. It was carried out the experimental researches of quality performance from cluster's quantity for definitions technological variety of iron ore. According to our research the result was established, that the optimal clusters quantity, during identifying technological variety of iron ore raw materials, ore's hematite variety in occurrence of Krivyi Rih, is 7.

Keywords. Indeterminate clusterization, cluster, Gustafson-Kessel's method, ore sorting.

Introduction. Necessity of decreasing prime cost and increasing of technical and economic indicators of dressing magnetic iron ore sets tasks of improving technologies and methods of control dressing as a whole and ore-preparation in particular. At this moment, native ore mining and processing enterprises refine on average 5-8 technological ore variety with considerable distinction characteristic value in each [1, 2].

In such conditions, one of the most perspective technological and economic industrial reserves is decreasing expenditure on transportation-mined rock and energy expenditure on refining separate technological ore variety, in particular appertain to «blast-furnace» ore, which after mining could be send to metallurgical treatment without refining on beneficiating plant.

Despite on great quantity of accomplish works, the problem of forming enough effective automatic control of process preconcentration iron ore in conditions of underground mining and beneficiating plant not received sufficiently complete solutions. Consequently, the questions of automation process of control sorting cobbed ore material with subject to physics and mechanics properties technological ore variety still important, actual and demand conducting further investigation.

It is reasonable to use the operation of clusterization to forming automatic process of control sorting large-size iron ore raw material, with operative recognition of its technological variety separate pieces in stream. Cluster-analysis or automatic classification were widely used in various branches, everywhere where are aggregate objects arbitrary character, which scribing in the vectors form x= $\{x_1, x_2, \ldots, x_N\}$, which are necessary to be broken automatically into groups «similar» objects, called clusters.

The majority of clusterization's algorithms are not based on tradition for statistical methods conventions. They can be used in conditions almost full absence information about laws of data distribution. The clustering is carried out for objects with the quantitative, qualitative or mixed indications.

There is a big amount of clusterization methods, which can be classified on clear and fuzzy. Clear methods of clusterization batter initial multiplicity of objects X on several non-crossing subset. At that any object from X belongs to only one cluster. Fuzzy methods of clusterization allow one and the same object to belong to several (or even all) clusters at the same time but with different degree. Fuzzy clusterzation in many situations more «natural» then clear clusterization, for example for objects, placed on clusters border [3, 4]. All well-known clusterization algorithms can be divided on two major groups: algorithms with antecedently set quantity of clusters and algorithms, which set optimal quantity of clusters by themselves. If clusters quantity is unknown, so the most prefer algorithms are algorithms which set quantity of clusters by themselves or algorithms with self-organization, one of them is Gustafson-Kessel's algorithm [5].

It was performed the research with the purpose to define optimal clusters by recognition technological variety of iron ore raw material quantity using Gustafson-Kessel's method.

Materials and Methods. Clusterization was performed by the next algorithm [6, 7]. In optimization process, there were accepted following quantity values: weight index m=2, tolerance value to stop the calculation ε=0,001 determinant for each cluster ρ=1. After initialization accessories matrix with random significances $U^{(0)} \in M_{fc}$, on each step l=1,2,... making the next steps.

Calculate clusters center

$$v_i^{(l)} = \frac{\sum_{k=1}^{N}\left(\mu_{ik}^{(1-1)}\right)^{m_{xk}}}{\sum_{k=1}^{N}\left(\mu_{ik}^{(1-1)}\right)^{m}}, 1 \le i \le c \qquad (1)$$

Calculate covariance matrix of clusters

$$F_i^{(l)} = \frac{\sum_{k=1}^{N}\left(\mu_{ik}^{(1-1)}\right)^{m}\left(x_k - v_i^{(l)}\right)\left(x_k - v_i^{(l)}\right)^T}{\sum_{k=1}^{N}\left(\mu_{ik}^{(1-1)}\right)^{m}}, 1 \le i \le c. \quad (2)$$

Add a scaled ordinary matrix

$$F_i := (1-\gamma)F_i + \gamma\left(F_0\right)^{1/n} I. \qquad (3)$$

Determine eigenvalues λ_{ij} and eigenvectors ϕ_{ij}, determine $\lambda_{i,\max} = \max_j \lambda_{il}$ restore F_i

$$F_i = \left[\varphi_i,1 ... \varphi_{i,n}\right] diag\left(\lambda_{i,1} ... \lambda_{i,n}\right)\left[\varphi_{i,1} .. \varphi_{i,n}\right]^{-1}. (4)$$

Calculate the distance

$$D_{ik_{A1}}^2\left(x_k, y_i\right) = \left(x_k - v_i^{(l)}\right)^T\left[\left(\rho_i \det(F_i)\right)^{1/n} F_i^{-1}\right]\left(x_k - v_i^{(l)}\right). \quad (5)$$

Count an item value of accessories matrix according to the formula

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{e}\left(D_{ikA'}(x_k, v_i) D_{jk}(x_k, v_j)\right)^{2/(m-1)}},$$
$$1 \le i \le c, \qquad . (6)$$
$$1 \le k \le N.$$

Until stop condition entry will be calculated $\left\| U^{(l)} - U^{(1-1)} \right\| < \varepsilon$.

In the process of system optimizations was performed the following quantity values: m = 2, ε = 0,001, ρ = 1 for each cluster. During the research the number of clusters c varied from 2 to 14.

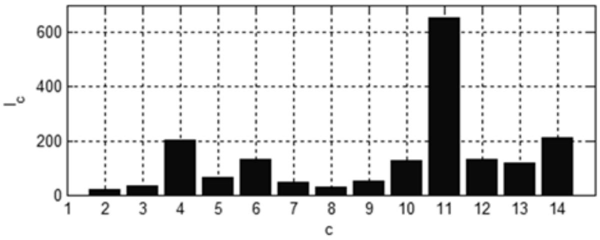
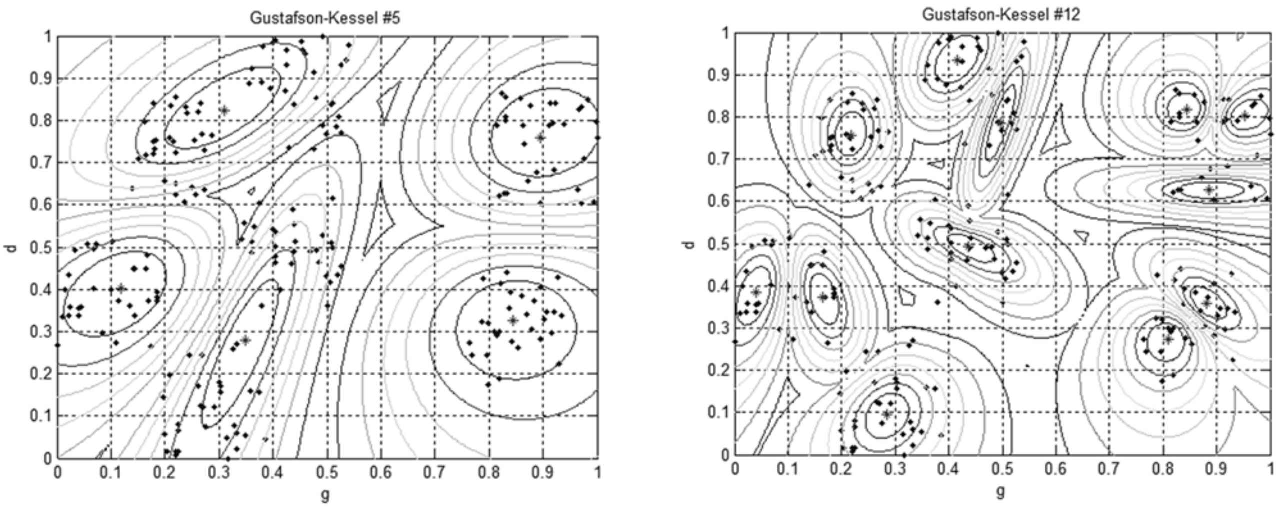


Figure 1. The number of iterations for different numbers of clusters

The lowest number of iterations required for the separation characteristics of samples of iron ore raw material on 2 clusters - 20, most - 654 when divided into 11 (Fig. 1).

The clustering characteristics of iron ore lumps results are shown in Fig. 2.



a) b)

Figure 2. The results of the clustering characteristics of the ore lumps with different quantity of clusters: a) 5 clusters; b) 12 clusters

Assessment of clustering quality was performed using scalar measures of reliability [7].

A graphical representation of the dependence of the clustering quality from the number of clusters according to Gustafson-Kessel algorithm is shown in Fig. 3- 9.

Distribution coefficient PC, the dependence of which on amount of clusters is shown in Fig. 3, is not sufficiently informative. [4] Thus, it should be noted the local burst in number of clusters c = 7.

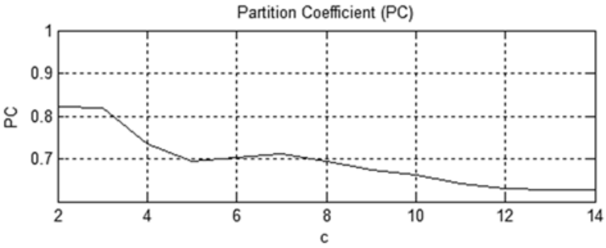


Figure 3. Distribution coefficient PC

Weak dependence on the analyzed data has an entropy classification CE: an amount of clusters increases due to monotonically function value increasing (Fig. 4).

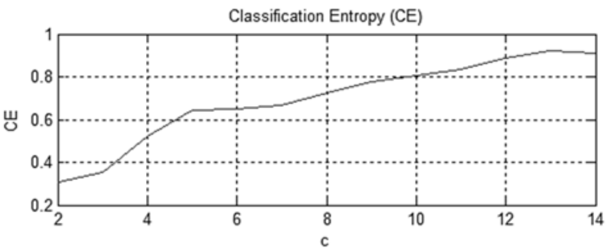


Figure 4. Classification Entropy (CE)

Distribution indicator SC decreases relatively rapidly with an increasing number of clusters with 2 to 7 (Fig. 5), after that the decrease slows substantially [8].

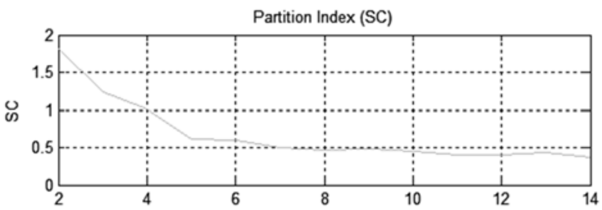


Figure 5. Distribution indicator SC

Dependence of the separation indicator S of clusters amount is shown in Fig. 6 [8].

Here are significant fluctuations in the index under increasing of clusters amount from 2 to 7 and relatively slow decline with further increasing.

Xi-Beni index XB changes quite strongly (Fig. 7), which does not allow unambiguous to determine the optimal number of clusters [9].

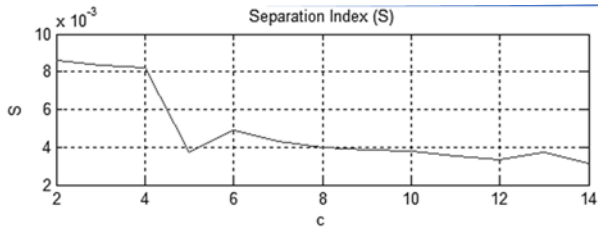


Figure 6. Separation indicator S

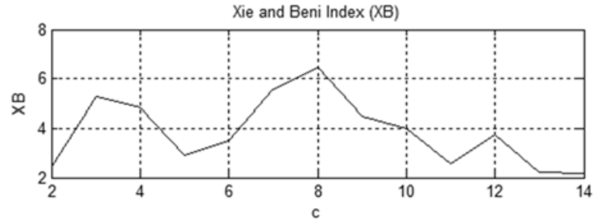


Figure 7. Xi-Beni index XB

Dunn indicator DI (Fig. 8) indicates that the optimum amount of clusters is c = 7 [6].

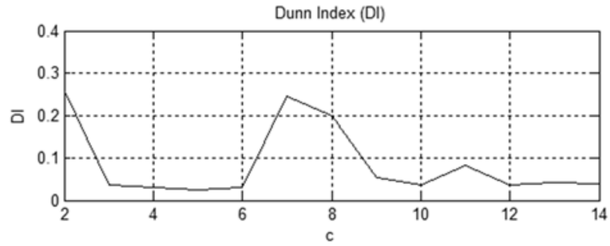


Figure 8. Dunn indicator DI

The value of alternative Dunn indicator ADI, whose dependence on amount of clusters is shown in Fig. 9, rather strongly decreases with increasing amount of clusters to 7, with a consequent increase almost does not change.

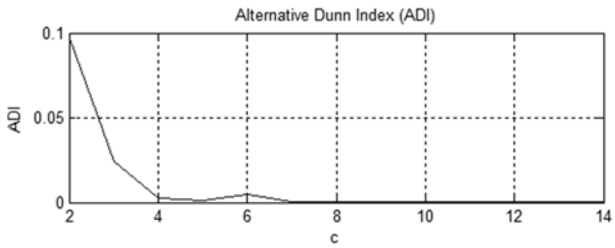


Figure 9. Alternative Dunn indicator ADI

The results of these studies have shown that optimum amount clusters in determining the technological varieties of iron ore is c = 7, and the need 29 iterations.

The results of clustering using a Gustafson-Kessel algorithm in the number of clusters of five to nine are shown in Fig. 10.

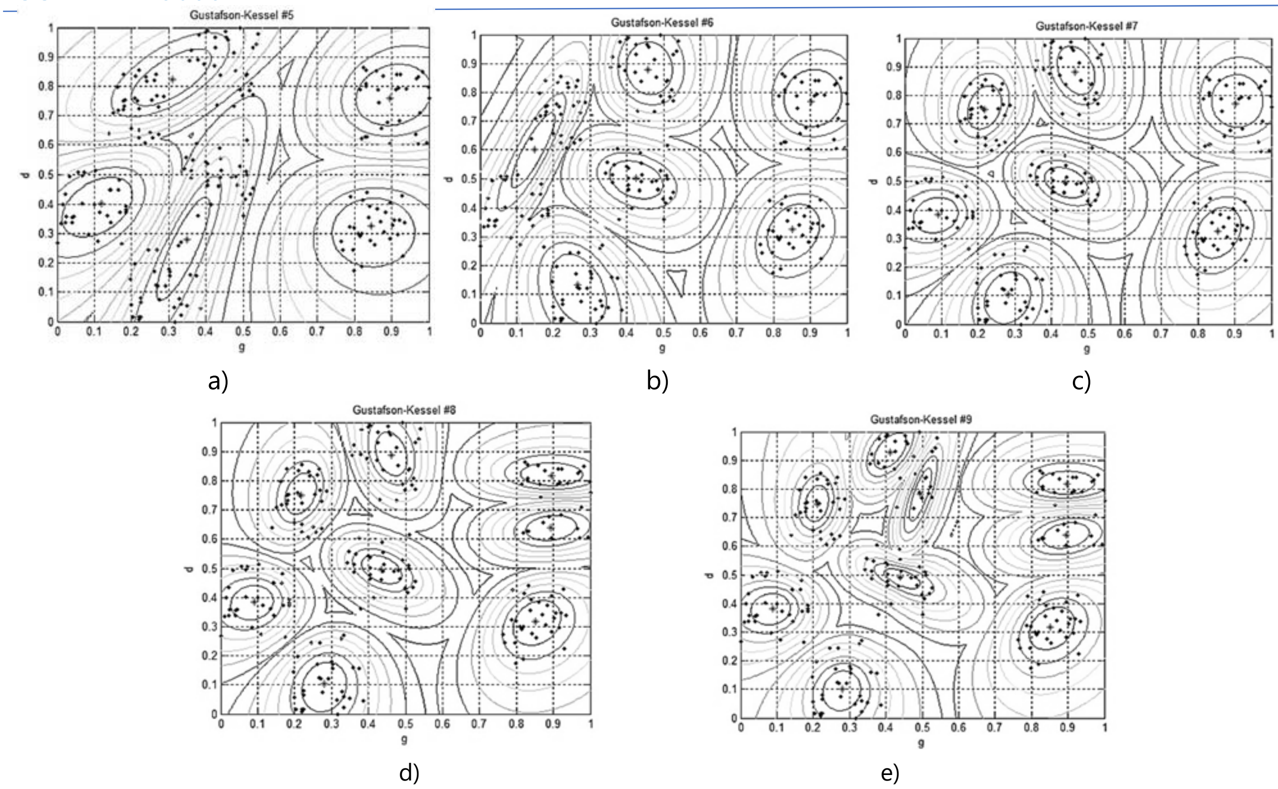


Figure 10. The results of the clustering characteristics of the ore lumps in different number clusters

Conclusions. Thus, the formation of automatic maintenance by lumped ore sorting with operational recognition of technological varieties of individual pieces in the stream was carried out by using fuzzy clusterization of its characteristics using a Gustafson-Kessel's algorithm.

The results of these studies have shown that clusters optimum amount in determining the technological varieties of iron ore, hematite ores of Krivyi Rih occurrence is c = 7, where 29 iterations must be calculated with precision.

References.

1. Gubina V.H., Horluckiy B.O. (2009). Problema zalizovmisnuh vidhodiv u hirnicho-metalurhiynoho kompleksu Ukrinu – sustemnuy pidhid. Inst.geochemistry of environment, **17**: 79-92.

2. Gubin H.H., Gubina A.H. (2010). Ukraini mizh kruzamu. *Journal of Kryvyi Rjh Technical University*, **25**: 218-224.

3. Shtovba S.D. Vvedenie v teoriyu nechetkih mnozhestv I nechetkuyu logiku. Available at: //http:matlab/exponenta.ru/fuzzylogic/book1.

4. Bezdek J.C. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, 1981.

5. Gustafson D.E., Kessel W.C. (1979). Fuzzy clustering with a fuzzy covariance matrix. *Proc. IEEE CDC, San Diego, CA*, **7**: 773—781.

6. Babuka R., P.J. van der Veen, Kaymak U. (2002). Improved covariance estimation for Gustafson-Kessel clustering. In Fuzzy Systems, *FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference*, **2**: 1081 –1085.

7. Bensaid A.M., Hall L.O., Bezdek J.C., Clarke L.P., Silbiger M.L., Arrington J.A., Murtagh R.F. (1996). Validity-guided (Re)Clustering with applications to imige segmentation. *IEEE Transactions on Fuzzy Systems*, **4**:112-123,.

8. Xie X. L. and Beni G. A. (1991). Validity measure for fuzzy clustering. *IEEE Trans. PAMI*, **3**(8):841-846,.

9. Morkun V., Tsvirkun S. (2014). Investigation of methods of fuzzy clustering for determining ore types. *Metallurgical and Mining Industry*, **5**: 12-15.