Multidimensional Analysis of Aluminum Production Monitoring Data in Basic Operation Modes*

Tatiana Penkova[0000-0002-0057-0535], Maria Senashova[0000-0002-1023-7103] and Aleksey Korobko[0000–0002–6227–1362]

Institute of Computational Modeling of the Siberian Branch of the Russian Academy of Sciences, 50/44 Akademgorodok, Krasnoyarsk, 660036, Russia
penkova_t@icm.krasn.ru

Abstract. This paper presents a comprehensive analysis of the technological parameters of aluminum production for three test sites by applying data mining techniques. Based on the principal component analysis and cluster analysis, structural features of the multidimensional monitoring data space were investigated, regularities in the aluminum production complex operation in its basic operating modes were detected, and typical conditions leading to technological disorders were determined. New knowledge and "analytical portraits" of the operation of an aluminum production complex can be used to develop algorithms for the prevention of technological disorders.

Keywords: Multidimensional Data Analysis, Principal Component Analysis, Cluster Analysis, Aluminum Production, Prevention of Technological Disorders.

1 Introduction

High technical and economic indicators in the aluminum industry are largely determined by the quality of technology and timely assessment of the technological state of an aluminum production complex and its separate units: reduction cells, potrooms and series [1]. The comprehensive analysis of monitoring data allows one to explain the reasons of a decrease in productivity and determine conditions that trigger "disorders" in the technological process. The complexity of the object and data features (i.e. large number of parameters, high inertia of processes, gaps, and noise in the data) require applying modern technologies and big data processing methods.

This paper presents a study of the features and patterns in the operation of the aluminum production complex in its basic operating modes based on the data mining techniques – principal component analysis and cluster analysis – applied to the monitoring data of the process control system. Data mining techniques provide an effective tool for discovering previously unknown, nontrivial, useful in practice, and interpreted knowledge indispensable for decision-making [2, 3].

* Copyright c 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR-WS.org/Vol-2727/paper17.pdf

The paper describes the results of a comprehensive multidimensional analysis of technological parameters in aluminum production for three experimental areas: Khakas aluminum smelter (KhAS) with RA-300 technology (potrooms No. 9 and 10, 336 retention cells) for the period from 2014 to 2019; Boguchansky aluminum smelter (BoAS) with RA-300 technology (potrooms No. 1 and 2, 336 retention cells) for 2019; Bratsk aluminum smelter (BrAS) with Soderberg technology (potroom No, 8, 90 retention cells) for the period from 2015 to 2019. For each experimental area, the authors investigated the structural features of the multidimensional monitoring data space, detected regularities in the operation and typical conditions leading to technological disorders. The analysis results for individual years showed that the data structure and the nature of the behavior in the studied facilities were the same in all the periods. Within this research, the analysis and visualization of multidimensional data were performed using the Pyton and ViDaExpert tools [4, 5].

2 Data Description and Preprocessing

According to the theory of multidimensional data analysis, the original data are represented as a set of objects and a set of attributes. The set of objects contains moments in the operation of the retention cells. The set of attributes contains technological parameters registered by the Automated Process Control System. The composition of the attributes is determined by the features of the technology and typical disorders. The key data attributes are listed in Table 1.

The most common technological disorders in the aluminum production process include the anode effect and distortion of the anode surface relief [6, 7]. The anode effect is a polarization phenomenon characterized by a significant increase in the cell voltage. Distortion of the anode surface relief can be presented as “cracking”, “corner shedding” or signs of buildups, such as a “spike” – a formation of regular cylindrical or conical shape at the anodes; “lagging” – a rectangular protrusion or unevenness at the base of an anode, occupying up to 50-60% of the anode area; “overglow” – a formation at any face of the anode block (e.g. “ball”, “mushroom”, “chunk”, etc.).

To prepare the data for analysis, they were subjected to preliminary processing. The original formats were converted, and data were merged by time. For the indicators of the chemical composition, gaps in the data were filled using interpolation with algebraic polynomials. Additional parameters were calculated, including statistical indicators of technological disorders (*Number of "rolled anodes", Number of anode effects*, *Number of "spikes"*, *Number of “lagging”*, etc.) and generalized indicators (*Consumption of alumina* and *Consumption of aluminum fluoride*). Finally, entries with the empty parameter values were excluded (appr. 30-40% of the entries).

Table 4. List of the key data attributes.

No	Attribute	No	Attribute
1	Metal level, cm	17	Amperage, kA

2	Velocity ratio, kg/cm	18	Service life, month
3	Electrolyte temperature, C°	19	Anode consumption rate, cm/day
4	Electrolyte level, cm	20	Number of pins not on the horizon, pcs.
5	Cryolite ratio	21	Distance from the anode base, cm
6	CaF_2 concentration, %	22	Sintering cone, cm
7	MgF_2 concentration, %	23	Leg size, cm
8	Fe concentration, %	24	CPC level, cm
9	Se concentration, %	25	CPC temperature, C°
10	Alumina dose, kg	26	Number of "cracks", pcs.
11	Number of alumina doses, pcs.	27	Number of anode effects, pcs.
12	Aluminum fluoride dose, kg	28	Number of "corner sheddings", pcs.
13	Number of aluminum fluoride doses, pcs.	29	Number of "chunks", pcs.
14	Coefficient of the anode-cathode distance, mV/s	30	Number of "rolled anodes", pcs.
15	Cell voltage, V	31	Number of "spikes", pcs.
16	Back EMF, V	32	Number of "laggings", pcs.

Additionally, the original data set was subjected to a correlation analysis. The result demonstrated quite a strong relationship between the following parameters: for KhAZ and BoAZ: *Dose of alumina* and *Number of alumina doses* with r=-0.88, *Electrolyte temperature* and *Cryolite ratio* with r=0.7, *Duration of pouring* and *Velocity ratio* with r=-0.65; for BrAZ: *Alumina feed time in automatic mode* and *Alumina feed time in manual mode* with r=-1.00; *Distance from the anode base* and *Hollow of the anode* with r=0.94, *Metal level* and *Service life* with r=0.76, *Metal level* and *Cell voltage* with r=0.71, *Composite level* and *Sintering cone* with r=-0.69. The established dependences are largely explained by the features of technology and nature of physical processes, which makes it possible to understand the general regularities of the aluminum production complex.

3 Multidimensional Analysis of Monitoring Data

In order to apply the multidimensional data analysis, each experimental area was attributed a dataset: KhAZ dataset contains 200,434 objects and 15 attributes; BoAZ dataset contains 95,497 objects and 15 attributes; BrAZ dataset contains 103,207 objects and 25 attributes. To reduce the dimension of the data space and identify patterns in the data structure, it required the principal component analysis (PCA) to be implemented [8, 9]. The combination of the Kaiser's rule and Broken-stick model [10] allowed identifying five principal components (PC1, PC2, PC3, PC4, PC5) for all the experimental areas. Further analysis and interpretation of the results were performed in the context of the principal components.

To identify the structure and reveal patterns within the data, the cluster analysis was performed using the density-based spatial clustering algorithm [11]. As a result, for each experimental area, we identified the functioning features of the aluminum production complex and characteristic conditions triggering technological disorders.

The results of clustering on the PCA plot and in the internal coordinates of the elastic map, obtained from the data of KhAZ for 2018, are presented in Fig. 1.

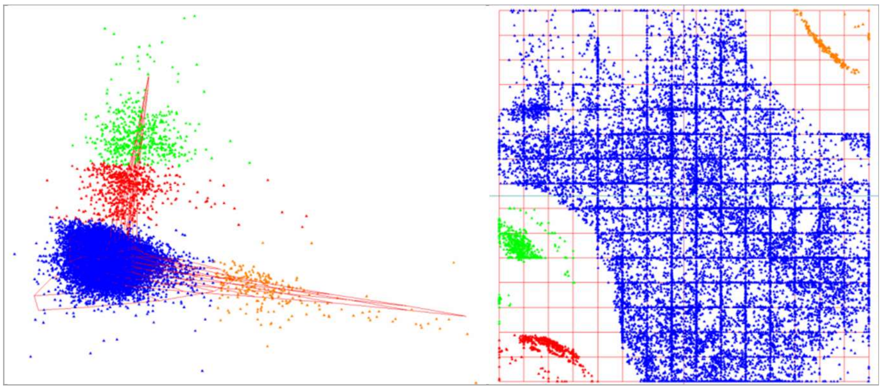

Fig. 1. Results of clustering on the PCA plot (PC2, PC4, PC5) and in the internal coordinates of the elastic map, KhAZ, 2018.

Table 5. Distribution of the average values of the key parameters by clusters, KhAZ, 2018.

No	Attribute	Cluster 1 (blue)	Cluster 2 (red)	Cluster 3 (green)	Cluster 4 (yellow)
1	Metal level	18.32	18.41	18.98	**17.60**
2	Electrolyte level	16.20	16.23	**15.73**	16.35
3	Electrolyte temperature	963.69	963.67	962.69	**967.55**
4	Consumption of alumina	17433.76	**17504.18**	17019.04	**16519.44**
5	Coefficient of the anode-cathode distance	32.31	32.30	31.82	32.28
6	Amperage	311.11	309.60	**295.87**	310.50
7	Service life	22.07	24.35	21.09	20.54
8	Number of “rolled anodes”	0.20	0.45	0.14	**4.45**
9	Number of anode effects	0.00	**0.83**	0.21	0.01
10	Number of “spikes”	0.00	0.01	0.00	**0.87**

The analyzed objects are divided into four clusters: Cluster 1 (blue) – 93% of the objects, Cluster 2 (red) – 4% of the objects, Cluster 3 (green) – 2% of the objects and Cluster 4 (yellow) – 1% of the objects. Distinctive features and the nature of each cluster are determined by the average values of the key parameters of the cluster objects (Table 2). Cluster 1 describes the basic operating mode of the complex when the specified technical conditions are maintained, and the parameters have standard values. This cluster is characterized by high productivity and minimal risk of technological disorders. Clusters 2, 3 and 4 stand out from Cluster 1 in terms of changes in the technical conditions. Cluster 2 is characterized by a more frequent occurrence of the “anode effect” and an increase in the *Consumption of alumina* while maintaining a good level of metal production. Cluster 3 is characterized by the technical conditions with a decrease in the *Amperage* and *Electrolyte level* which can correspond to the expected outages of the cells or occurrence of emergency situations. Cluster 4 is characterized by the presence of more serious technological disorders – formation of “spikes” which is accompanied by an increase in the *Electrolyte temperature* and a decrease in the *Consumption of alumina*. It should also be noted that the *Number of “rolled anodes”* increases significantly. As a consequence of the occurring disorders, we observed a decrease in the *Metal level*.

In addition, the cluster analysis in the context of the parameters *Alumina dose* and *Number of alumina doses*, allowed us to identify the features associated with the consumption of alumina in the cells: one part receives alumina less often but in large doses, another part, on the contrary, receives alumina more often but in small doses. The study of the occurrence of technological disorders showed that in the second case, the “spikes” formation occurred less frequently than in the first case. This suggests that the *Consumption of alumina* is one of the key factors affecting the occurrence of this type of disorder.

The results of clustering on the PCA plot and in the internal coordinates of the elastic map, obtained from the data of BoAZ for 2019, are presented in Fig. 2. The distribution of the average values of the key parameters by the clusters is presented in Table 3.

The analysed objects are divided into three clusters: Cluster 1 (blue) – 93% of the objects, Cluster 2 (red) – 4% of the objects and Cluster 3 (green) – 3% of the objects. Cluster 1 describes the basic operating mode of the complex with high productivity and minimal risk of technological disorders. Clusters 2 and 3 are characterized by a more frequent occurrence of technological disorders. Cluster 2 corresponds to the formation of "lagging", which is accompanied by a decrease in the *Electrolyte temperature* and an increase in the *Coefficient of the anode-cathode distance*. There is also an increase in the *Number of “rolled anodes”* and a decrease in the *Metal pouring interval*. Cluster 3 corresponds to the conditions with the frequent occurrence of the "anode effect". At the same time, we observe an increase in the *Electrolyte temperature*, a decrease in the *Consumption of alumina* and a decrease in the *Coefficient of the anode-cathode distance*. Technological disorders associated with the "spike" formation are observed quite rarely and most of them fall into Cluster 3.

Also, the results of the cluster analysis revealed the features associated with the age distribution of the cells. The cells are divided into two groups: group 1 – cells numbered 1001-1084 have a service life of 40-50 months, group 2 – cells numbered 1085-1168 have a service life of 1-9 months. The study of these groups showed that the age did not significantly affect the occurrence of technological disorders and the level of productivity.

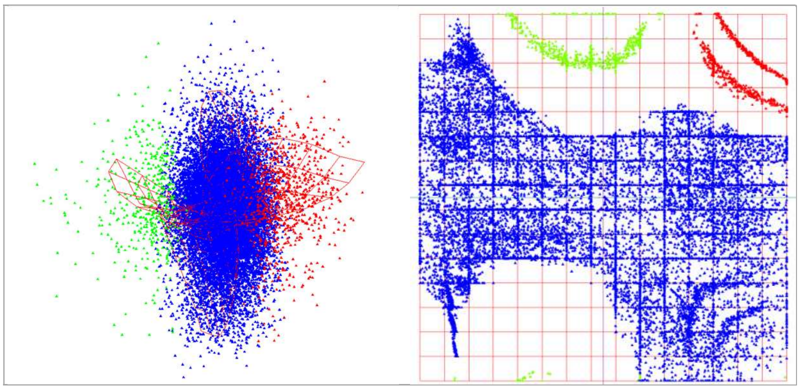

Fig. 2. Results of clustering on the PCA plot (PC1, PC4, PC5) and in the internal coordinates of the elastic map, BoAZ, 2019.

Table 6. Distribution of the average values of the key parameters by the clusters, BoAZ, 2019.

No	Attribute	Cluster 1 (blue)	Cluster 2 (red)	Cluster 3 (green)
1	Metal level	17.714	17.500	17.058
2	Metal pouring interval	5.383	**3.963**	6.390
3	Electrolyte level	16.252	16.337	17.300
4	Electrolyte temperature	963.366	**962.760**	963.977
5	Number of alumina doses	3874.022	3863.461	**3805.244**
6	Coefficient of the anode-cathode distance	27.702	**27.922**	**26.935**
7	Number of “rolled anodes”	0.504	**0.913**	0.438
8	Number of anode effects	0.019	0.016	**1.002**
9	Number of “spikes”	0.001	0.001	0.058
10	Number of “laggings”	0.016	**1.314**	0.012

The results of clustering on the PCA plot and in the internal coordinates of the elastic map, obtained from the data of BrAZ for 2019, are presented in Fig. 3. The distribution of the average values of the key parameters by the clusters is shown in Table 4.

The analyzed objects are divided into four clusters: Cluster 1 (blue) – 88% of the objects, Cluster 2 (red) – 11% of the objects, Cluster 3 (yellow) – 0.8% of the objects and Cluster 4 (green) – 0.2% of the objects. Clusters 1, 2, and 3 are located along the same axis and differ significantly from Cluster 4. Moving from Cluster 1 to Cluster 3, there is an increase in the *Service life*, an increase in the *Electrolyte temperature*, a significant decrease in the *Consumption of Alumina* and a significant increase in the *Metal level*. At the same time, Cluster 1 covers most of the objects and presents the main mode of the complex operation, Cluster 3 covers a small percentage of the objects with a special mode of operation.

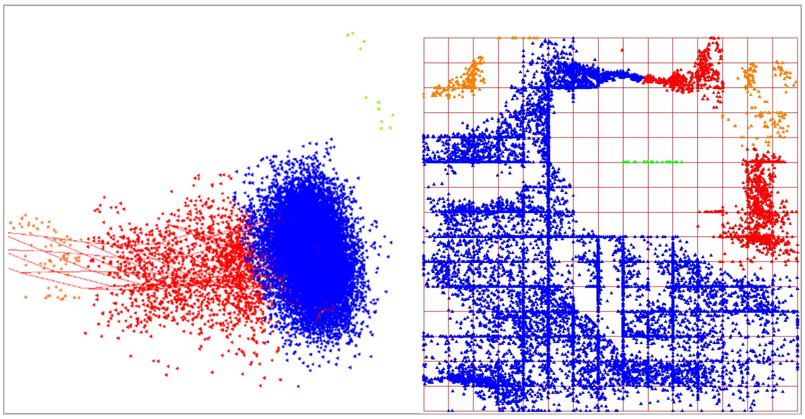

Fig. 3. Results of clustering on the PCA plot (PC1, PC2, PC3) and in the internal coordinates of the elastic map, BrAZ, 2019.

Table 7. Distribution of the average values of the key parameters by the clusters, BrAZ, 2019.

No	Attribute	Cluster 1 (blue)	Cluster 2 (red)	Cluster 3 (yellow)	Cluster 4 (green)
1	Metal level	**37.334**	**46.436**	**51.116**	**37.182**
2	Electrolyte level	17.755	16.079	13.029	20.545
3	Electrolyte temperature	953.071	961.493	**981.638**	966.000
4	Fe concentration	0.238	0.718	**2.656**	0.275
5	Si concentration	0.047	**0.211**	**0.491**	0.057
6	Number of alumina doses	**2217.663**	**1960.047**	**503.072**	1045.818
7	Consumption of aluminum fluoride	44.841	62.142	50.861	**87.545**
8	Service life	**27.445**	**51.973**	**66.373**	**24.800**
9	Distance from the anode base	0.00	0.01	0.00	**0.87**
10	Sintering cone	134.651	134.446	135.551	**140.364**
11	Leg size	18.410	16.835	14.261	**38.273**
12	CPC level	36.381	35.853	34.130	35.818
13	CPC temperature	136.506	136.569	136.174	135.000
14	Number of "cracks"	3.159	4.326	**7.507**	1.000
15	Number of anode effects	0.754	1.187	0.681	1.182
16	Number of "corner sheddings"	0.198	0.088	0.000	0.000
17	Number of "chunks"	0.010	0.018	0.058	0.000

Cluster 4 corresponds to events with atypical conditions in the technological process. The objects of this cluster are characterized by the average values of the *Service life* and *Consumption of aluminum fluoride* and low values of the *Metal level*. Apart from this, there is an increase in the *Consumption of aluminum fluoride* and a significant increase in the values of the following parameters: *Sintering cone*, *Leg size* and *Distance from the anode base*. Technological disorders do not form clear-cut clusters with certain conditions, however, the average values show that Cluster 3 is characterized by a more frequent occurrence of the *Number of "cracks"*, whereas Cluster 3 and Cluster 4 – by the occurrence of the "anode effects", and Cluster 1 – by the occurrence of the "corner sheddings".

A detailed analysis of the events with technological disorders confirmed the cluster analysis results, suggesting that the conditions and nature of the occurring disorders are different. In the case of "laggings" type disorders, the *Electrolyte temperature* decreases, the *Consumption of alumina* increases, while in "spike" type disorders, on the contrary, the *Electrolyte temperature* goes up, *Consumption of alumina* drops. The formation of "chunks" is usually accompanied by an increase in the *Electrolyte Temperature* and *Consumption of Aluminum Fluoride*, as well as a decrease in the *Consumption of Alumina* and *Average voltage of anode effect*.

Thus, the study of the structural features of the multidimensional monitoring data space made it possible to obtain new knowledge on the operation of the complex and basic regularities, as well as to determine the characteristic conditions for the occurrence of technological disorders.

4 Conclusion

This paper presents a study of features inherent to the aluminum production complex based on applying multidimensional analysis methods to the monitoring data for three experimental areas. The preliminary correlation analysis allowed us to establish relations between the key parameters, determine the strength of their influence on each other, and figure out the general characteristic patterns. The principal component analysis and Cluster analysis allowed us to reveal the structural features and dependences in the multidimensional space of monitoring data.

The clustering of the KhAZ objects revealed certain technological features associated with the consumption of alumina in a group of cells, identified conditions when the energy balance was violated in the post-start-up period and emergency operation. Also, it determined conditions for the occurrence of typical technological disorders – the anode effect and formation of "spikes". The clustering of the BoAZ objects revealed certain technological features associated with the age distribution of the cells. In addition, it determined conditions for the occurrence of typical technological disorders in the process – the anode effect and formation of "laggings". The clustering of the BrAZ objects elucidated the age characteristics of the cells and the effect of the service life on productivity.

The results of this study made it possible to confirm many hypotheses of engineers and to create the so-called "analytical portraits" of how separate units operate in the aluminum production complex, which can, in turn, serve as a basis for algorithms to prevent the occurrence and development of technological disorders.

References

1. Mikhalev, Yu.G., Polyakov, P.V., Yasinskiy, A.S., Shakhray, S.G., Bezrukikh, A.I., Zavadyak, A.V.: Anode processes malfunctions. An overview. J. Sib. Fed. Univ. Eng. technol. **10 (5)**. 593–606 (2017). doi: 10.17516/1999-494X-2017-10-5-593-606
2. Williams, G.J., Simoff, S. J.: Data Mining: Theory, Methodology, Techniques, and Applications. Springer (2006)
3. Penkova, T., Korobko, A.V.: Investigation of hydropower equipment functioning features using data mining techniques. In: Lecture Notes in Computer Science. Part I, Vol. 11619. pp. 434-446 Springer (2019). doi: 10.1007/978-3-030-24289-3_32
4. Gorban, A., Pitenko, A., Zinovyev, A.: ViDaExpert: user-friendly tool for nonlinear visualization and analysis of multidimensional vectorial data. Cornell University Library (2014), http://arxiv.org/abs/1406.5550
5. Raschka, S.: Python Machine Learning. Birmingham, UK: Packt Publishing Ltd (2015)
6. Mikhalev, Y.G., Polyakov, P.V., Yasinskiy, A.S., Polyakov, A.A.: Spikes generation on anode of aluminium reduction cell. Tsvetnye Metally **9**. 43–48 (2018). doi: 10.17580/tsm.2018.09.06
7. Sadler, B.A.: Critical issues in anode production and quality to avoid anode performance problems. J. Sib. Fed. Univ. Eng. technol. **5(8)**. 546–568 (2015). doi: 0.17516/1999-494X-2015-8-5-546-568
8. Abdi, H., Williams, L.: Principal Components Analysis. Wiley Interdisciplinary Reviews. Computational Statistics **2(4)**. 439–459 (2010)
9. Gorban, A., Zinovyev, A.: Principal manifolds and graphs in practice: from molecular biology to dynamical systems. International Journal of Neural Systems. **20(3).** 219–232 (2010). doi: 10.1142/S0129065710002383
10. Peres-Neto, P., Jackson, D., Somers, K.: How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Computational Statistics & Data Analysis **49(4)**. 974–997 (2005). doi: 10.1016/j.csda.2004.06.015
11. Tran, N.T., Drab, K., Daszykowski, M.: Revised DBSCAN algorithm to cluster data with dense adjacent clusters. Chemometrics and Intelligent Laboratory Systems. **120**. 92–96 (2013). doi: 10.1016/j.chemolab.2012.11.006.