

Tools for Detecting Process Disruptions in Aluminium Production Based on Data Mining Techniques

Viktor Mann¹, Vitaly Pingin², Iliya Puzanov³, Andrey Zavadyak⁴,
Yury Shtefanyuk⁵ and Tatiana Penkova⁶

1. Chief Technical Officer, RUSAL Global Management B.V., Moscow, Russia
2. Director, Engineering and Technology Directorate
3. Project Manager, Development of High-Amperage Technologies
4. Deputy Director General, Aluminium Smelter Modernization
RUSAL Engineering and Technology Center, Krasnoyarsk, Russia
5. Branch Director, UNITED COMPANY RUSAL IP LTD, Moscow, Russia
6. Senior Research Associate, Institute of Computational Modelling of the Siberian Branch of
the Russian Academy of Sciences, Krasnoyarsk, Russia
Corresponding author: penkovatg@gmail.com

Abstract

To detect any process disruptions at early stages and assess risks of productivity reduction in aluminum production, algorithm and software tools were developed to identify and forecast process disruptions using controlled variables based on mathematical statistics, data mining and machine learning methods. Ensemble algorithms were applied to daily average monitoring data to develop so-termed “Voting” and “Meta-classification” predictive models. Such models establish internal patterns in the data to determine the conditions promoting the technological process disruptions, such as anode spikes. Daily average monitoring data were used in the association rule mining to develop a model that generates the rules to assess the process status and disruption probability by analyzing preceding events and revealing cause-and-effect relations in the data. Taking into account the mutual influence of process parameters and specifics of aluminum production technology, an algorithm for identification of the changes in “noise” and “anode current” parameters was elaborated to detect process deviations based on instantaneous monitoring data. The recorded values are used to determine the status of a process disruption. Based on the integration of artificial intelligence and decision-making technologies, a software implementation of methods and algorithms was performed; a pilot Web-enabled information analysis system for the online control of aluminum production process was developed. The system performs the following two main tasks: monitoring of technological parameters (*i.e.*, reduction process data) while displaying the parameter change patterns and warning of process failures while indicating a controlled unit status and generating reports on process conditions and problematic areas of the aluminum production facility. The developed software tools have been tried out in the pilot areas at the Sayanogorsk Aluminum Smelter (RA-300 and RA-550 technologies). The obtained results prove the adequacy of the models and algorithms for application; moreover, they show that further research is needed to achieve more accurate predictions.

Keywords: Aluminum production, Detection of process disruptions, Predictive models, Data mining, Machine learning.

1. Introduction

The achievement of high technical and economic performance in the aluminum industry is largely determined by the quality of technology management and the timely assessment of the technological condition of both the entire aluminum production complex and its individual units [1, 2]. Process disruptions occurring in the aluminum production cycle are a major constraint. One of the most serious process disruptions, leading to a significant reduction in current

efficiency, is the deformation of the anode working surface, in the form of a so-called “anode spike” – a formation of a regular cylindrical or conical shape at the bottom surface of the anode block [3, 4]. Such disruption is only identified at the extreme severe stage when the formation develops into a protrusion entering the cathode metal, which is accompanied by a change in the operating conditions. The introduction of production management software and hardware packages makes it possible to accumulate significant amounts of monitoring data and use it for decision-making. The application of Data Mining and Machine Learning to monitor data allows for the identification of characteristic dependencies and the early identification and prediction of the occurrence of disruptions [5, 6].

A large number of papers on the improvement of the performance of reduction cells and the diagnosis of process disruptions in the aluminum production cycle confirms the relevance of this area of research. Existing solutions and results in the field of mathematical modelling of production processes [7, 8], analysis and evaluation of the impact of parameter values on production quality [9, 10], identification of causes of disruptions [11], create a theoretical basis for further research and form a number of hypotheses requiring confirmation.

This paper presents algorithmic and software tools for the detection of anode spikes based on monitored parameter values, using a combination of mathematical statistics, data mining and machine learning methods. The main text is organized as follows. Section 2 presents the problem statement and a description of the initial parameters for daily and instantaneous monitoring. Section 3 presents a description of the algorithmic means of detecting disruptions from average daily monitoring data and considers the building of predictive models based on ensemble algorithms and association rules. Section 4 describes the algorithmic means of detecting process deviations from instantaneous monitoring data, discusses algorithms for identifying changes in time series of parameters and finding problematic anodes. Section 5 shows the results of the software implementation of methods and algorithms in the information-analytical system of online control of aluminum reduction technology.

2. Research Objective and Inputs

The technological parameters of the daily average monitoring characterize the state of operation of the aluminum reduction cells, and the values of these parameters change significantly in the event of process disruptions and changes in the cell operation mode. Analysis of both monitoring data and registered process disruptions makes it possible to identify patterns and develop a classification algorithm: based on new values of the cell operation parameters, it will make the decision about the current process state: either “normal” or “with violation”.

The technological process parameters that are instantaneously sampled are described by a time series, which is a sequence of values obtained at certain points in time. When process disruptions occur (be either an anode spike, an alumina supply problem, an anode effect or others), the character of the time series changes. Indicators of descriptive statistics are used to describe the properties of a time series. Thus, the problem of detecting technology deviations from instantaneous monitoring data comes down to identifying changes in the behavior of the time series based on changes in the descriptive statistics, characterizing a shift in the mean level of the parameter values and an increase in the range of variation.

Development of software and algorithmic tools for detection of process disturbances is performed based on average daily values and instantaneous monitoring of the operation of a group of both RA-300 and RA-550 reduction cells at the Sayanogorsk Aluminum Smelter during 2019–2020. The scope of the daily average monitored parameters of the aluminum production complex consists of about 70 parameters, including, but not limited to: duration of metal tapping, metal level, electrolyte level, electrolyte temperature, alumina dose, bath chemistry parameters,

parameters of the point feeding system for alumina and aluminum fluoride, adjustment parameters of the anode-to-cathode distance, amperage, voltage parameters, state and service life of cells, and recorded process disruptions. The paper elaborates on two key parameters of instantaneous monitoring: “cell noise” and “anode current”.

3. Prediction of Process Disruptions Based on Daily Average Monitoring Data

3.1 Principle of Prediction Construction

The process of prediction construction based on daily average monitoring data is multi-stage and takes into account the results of several models and algorithms. Figure 1 illustrates the principle of prediction model construction for detecting process disruptions (e.g., an anode spike). The schema describes the composition of models (represented as encircled operator entities) as well as their required inputs and resulting outputs.

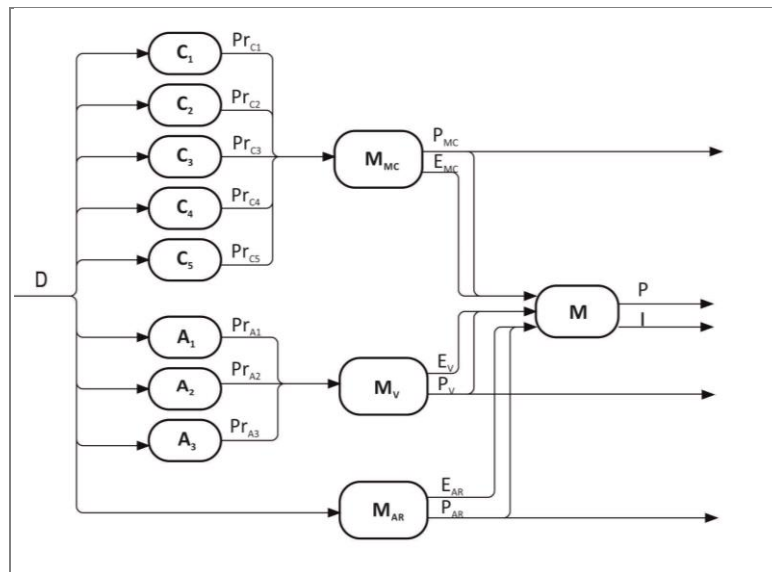


Figure 1. Scheme of prediction model construction for detecting process disruptions (e.g., an anode spike).

where:

- M generalized model (computation)
- P generalized process disruptions probability, %
- I generalized process disruptions indicator, unitless
- M_{MC} meta-classification model (computation)
- P_{MC} meta-classification model -based process disruptions probability, %
- E_{MC} meta-classification model predictive accuracy coefficient, unitless
- M_V voting model, (computation)
- P_V voting model-based process disruptions probability, %
- E_V voting model predictive accuracy coefficient, unitless
- M_{AR} association rules model (computation)
- P_{AR} association rules model-based process disruptions probability, %
- E_{AR} association rules model predictive accuracy coefficient, unitless
- $C_1 \dots C_5$ basic classification algorithms for meta-classification model (computation)
- $Pr_{C1} \dots Pr_{C5}$ results from algorithms C_1 through C_5 , respectively, unitless
- $A_1 \dots A_3$ basic classification algorithms for voting model (computation)
- $Pr_{A1} \dots Pr_{A3}$ results from algorithms A_1 through A_3 , respectively, unitless
- D daily average monitoring data (units vary with input data type).

The prediction model for detecting process disruptions (e.g., an anode spike) from average daily monitoring data is based on the results of three models (refer to Figure 1):

- 1) Meta-classification model (M_{MC}).
- 2) Voting model (M_V).
- 3) Associative rule model (M_{AR}).

The meta-classification model, M_{MC} , is based on the results of five basic classification algorithms ($C_1 \dots C_5$). The voting model, M_V , is based on the results of three basic classification algorithms ($A_1 \dots A_3$). The results of these eight classification algorithms ($Pr_{C1} \dots Pr_{C5}, Pr_{A1} \dots Pr_{A5}$) are the predictors that take the value 0 or 1 and correspond to the class of object state: “normal” or “disrupted”, respectively.

Each of the three models produces a probability of disruptions as a result: P_{MC}, P_V, P_{AR} , i.e., the probability of a disruption by the metaclassification algorithm, by the voting algorithm and by the association rules algorithm, respectively. In addition, for each of the three models, based on their validation results, the predictive accuracy coefficients are identified: E_{MC}, E_V, E_{AR} , i.e., the predictive accuracy coefficients of metaclassification model, of voting model and of association rules model, respectively. For example, the following accuracy coefficients have been obtained: $E_{MC} = 0.90, E_V = 0.85, E_{AR} = 0.80$.

The result of the generalized model includes the generalized process disruptions probability, P , and the generalized process disruptions indicator, I . The generalized process disruptions probability (P) is calculated based on the probability of disruptions and predictive accuracy coefficients identified for three models, as follows:

$$P = \frac{\sum_{i=1}^n P_i E_i}{n} \quad (1)$$

where:

- P_i process disruptions probability of the i -th model, $P_i \in (P_{MC}, P_V, P_{AR})$, %
- E_i predictive accuracy coefficient of the i -th model, $E_i \in (E_{MC}, E_V, E_{AR})$, unitless
- n model number, unitless.

The generalized process disruptions indicator (I) characterizes the state of disruption development and it is defined as the maximum indicator of the individual models:

$$I = \text{Max}(I_i) \quad (2)$$

where:

- I_i process disruptions indicator of the i -th model, $I_i \in (I_{MC}, I_V, I_{AR})$, unitless.

The indicator associated with each individual model is determined based on values of process disruptions probability, as follows:

$$I_i = \begin{cases} 0, & \text{if } 0 \leq P_i < 60 \\ 1, & \text{if } 60 \leq P_i < 80 \\ 2, & \text{if } 80 \leq P_i \leq 100 \end{cases} \quad (3)$$

There are three possible values for a given indicator I_i : 0 or “blue”, meaning “normal”; 1 or “yellow” meaning “development of process disruptions”; 2 or “red”, meaning “occurrence of process disruptions”.

Further, let us consider the content and stages of building basic predictive models.

3.2 Meta-Classification and Voting Models

The building of prediction models based on meta-classification (M_{MC}) and voting (M_V) involves three main steps:

- a) Pre-processing of monitoring data.
- b) Prediction based on classification algorithms.
- c) Computation of a generalized prediction.

Data pre-processing involves recovering missing values, finding and removing outliers, and balancing the data [12]. Three algorithms were considered to recover missing values:

- *EM-algorithm* (Expectation-Maximization algorithm).
- *k-NN algorithm* (k-Nearest Neighbors algorithm).
- *MICE-algorithm* (Multiple Imputation by Chained Equations algorithm).

Methods considered for finding anomalous values included:

- *Tukey's method*.
- *Local outlier factor method*.
- *Isolation Forest method*.

In order to solve the problem of imbalanced data, when the number of “no disruption” records is significantly higher than the number of “disrupted” records, an approach based on different sampling strategies is applied: *truncation*, *addition* and *hybrid method*. Six algorithms were considered:

- *Smote algorithm*.
- *SmoteNN algorithm*.
- *SmoteTomek algorithm*.
- *EncSmote algorithm*.
- *SVMSmote algorithm*.
- *KSmote algorithm*.

Based on an analysis of the accuracy metrics calculated for the selected methods, the algorithms with the best results are identified:

- *EM-algorithm* (for value recovery).
- *Tukey's method* (for finding outliers).
- *SVMSmote algorithm* (for data balancing).

Prediction based on classification algorithms is performed by applying a set of machine learning algorithms [12]. Methods that were considered to build prediction models:

- *Logistic regression*.
- *Decision trees*.
- *Ensemble methods*:
 - *Additional trees algorithm*.
 - *Balanced Random Forest Classifier algorithm*.
 - *Gradient boosting on decision trees* with:
 - *XGB Classifier*;

- *Cat Boost Classifier*;
- *AdaBoost Classifier*;
- *RUSBoost Classifier algorithms*.
- *Meta-assessment with:*
 - *Balanced Bagging Classifier algorithm*;
 - *Gaussian naive Bayes algorithm*.

When building a model based on meta-classification, a generalized prediction is generated based on a stacking ensemble method. This method consists of training several basic classification algorithms and combining them by training a metamodel on the multiple predictions returned by the basic models. The stack ensemble is formed according to the *k-fold cross learning* principle. The training sample is split into several folds. The basic classifier algorithms are trained sequentially in all but one part. Predictions are generated for the remainder of the training sample on the basis of the resulting models.

The prediction results obtained by the basic algorithms are treated as new features (“meta-features”) that are used as inputs for training the metamodel. *Logistic regression* is used as a *Meta-Classifier* metamodel to combine the prediction results of the base models. Thus, based on the analysis of accuracy metrics, the *Meta-Classifier* method, which combines the basic algorithms (namely, *Additional trees algorithm*, *Balanced Random Forest Classifier*, *XGB Classifier*, *AdaBoost Classifier* and *Balanced Bagging Classifier*), is selected for predicting the occurrence of anode spikes.

A voting model (M_V^k) uses the *Voting Classifier* method to generate a generalized prediction by combining conceptually different machine learning classifiers and using majority voting (or mean predicted probabilities, also known as “soft voting”). This classifier balances the disadvantages of the individual classifiers. Thus, based on the analysis of accuracy metrics, the *Voting Classifier* method, which combines the algorithms: *Balanced Random Forest Classifier*, *Balanced Bagging Classifier* and *Gaussian naive Bayes algorithm*, is selected for predicting the occurrence of the anode spikes.

3.3 Association Rule Model

Building a predictive model based on Association rule mining (M_{AR}) consists of applying technology that enables the discovery of dependencies and characteristic combinations between elements in large datasets by analyzing precedent events. This results in a set of rules of the form “*If <event X> then <event Y>*” describing the cause-and-effect relationship. Model building involves two main steps: binarization of the raw data and the generation of association rules [13].

Binarization (or categorization) of the data allows for multiple items characterizing the state of the technology on a given day. First, the values of each monitored parameter are divided into non-overlapping ranges, and then a binary matrix is formed by classifying the data according to its occurrence in the defined ranges. In determining the ranges, two approaches have been used: the general approach, where data for the reduction cells are considered together, and the individual approach, where the ranges of parameter values are calculated individually for each reduction cell. The criteria for dividing values into ranges included: statistical norms (*STDDEV*), quartiles (*QUARTILES*) and ranges characteristic of the occurrence of disruptions (*HISTOGRAMS*) [14]. The days on which the violations were recorded and the five days preceding them were chosen as the days with violations.

The process of generating association rules consists of two steps:

- Searching the data for combinations of elements occurring with a given frequency (*itemset*).
- Generating rules from the found sets.

The main criteria for evaluating rules are “support” and “confidence”. Support shows how often the combination occurs in the full sample. Confidence allows us to assess how often the left-hand side of the rule meets the right-hand side.

The main parameters for setting up the model are characteristics: rule generation approach (generic or individual), binarization method (*STDDEV*, *QUANTILES*, *HISTOGRAMS*), data completeness (with or without empty data values) and range of values (with or without “normal” parameter values).

Based on the analysis of accuracy metrics for predicting the occurrence of anode spikes, a model is selected that implements an individual approach with *QUANTILES* binarization, taking into account blank values and not taking into account “normal” values.

4. Detection of Process Deviations based on Instantaneous Monitoring Data

The algorithmic means of detecting process deviations from instantaneous monitoring data is based on the identification of changes in the time series of two parameters: “cell noise” and “anode current” using descriptive statistics: the weighted average, the amplitude of variation and the weighted linear deviation.

Based on the results of a case study of process disruption events [15], the range of values for the “cell noise” parameter is divided into three intervals (Figure 2):

- [0;0.03) mV: normal operation conditions.
- [0.03;0.06) mV: potential process disruptions.
- [0.06;∞] mV: operation with evident disruptions.

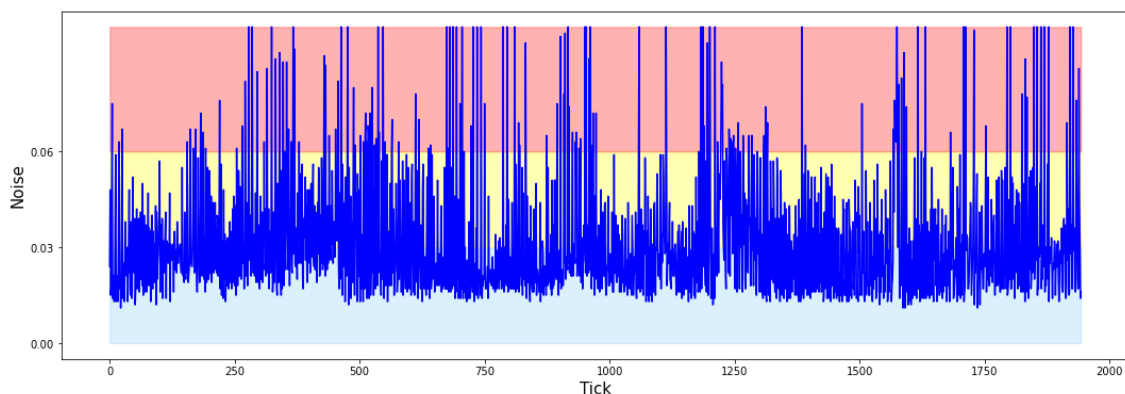


Figure 2. Graduation of values of the "cell noise" parameter, [mV].

After identifying the “cell noise” values characterizing the occurrence of disruptions, the search for “problematic” anodes by the parameter “anode current” is performed using estimates: weighted average and weighted linear deviation. An increase in the values of these characteristics indicates the occurrence or development of any disruptions. Finally, among the “problematic” anodes found, the four with the highest characteristic values are selected. Figures 3 and 4 show the pseudocode of the process deviations detection algorithms based on the values of “cell noise” and “anode current”.

```

Input
l // Length of sequence
 $\vec{w} = [w_1, w_2, \dots, w_l]$  // Vector of weights for sequence values
TSnoise // Time series values of noise
TSamp // Time series values of anode current
Initialization
indicatornoise = 0
k = count(anodes) // Number of anodes
 $\bar{x}_j = \text{weighted\_mean}(TS_{amp}, (i), \vec{w}), i = \overline{1, l}, j = \overline{1, k}$ 
 $A_{j,i} = \text{amplitude}(TS_{amp}, (i), \bar{x}_j), i = \overline{1, l}, j = \overline{1, k}$ 
 $\bar{d}_j = \text{weighted\_linear\_deviation}(A_{j,i}, \vec{w}), i = \overline{1, l}, j = \overline{1, k}$ 
begin
  for t от l + 1 до T:
    if TSnoise(t) ∈ [0; 0.03]:
      indicatornoise = 0
    if TSnoise(t) ∈ [0.03; 0.06]
      indicatornoise = 1
    if TSnoise(t) > 0.06:
      indicatornoise = 2
    for j от 1 до k:
       $\bar{x}'_j = \text{weighted\_mean}(TS_{amp}, (i), \vec{w}), i = \overline{t-l, t}, j = \overline{1, k}$ 
       $A'_{j,i} = \text{amplitude}(TS_{amp}, (i), \bar{x}'_j), i = \overline{t-l, t}, j = \overline{1, k}$ 
       $\bar{d}'_j = \text{weighted\_linear\_deviation}(A_{j,i}, \vec{w}), i = \overline{t-l, t}, j = \overline{1, k}$ 
    for indicatornoise ∈ [1,2]:
       $\bar{x}_j = \bar{x}'_j$  // Updating the weighted averages
       $\bar{d}_j = \bar{d}'_j$  // Updating the weighted linear deviations
  end
end

```

Figure 3. Algorithm for identifying changes in the time series.

```

Input
 $\bar{x}_j$  // weighted averages for t - 1
 $\bar{x}'_j$  // weighted linear deviations for t - 1
 $\bar{d}_j$  // weighted averages for t
 $\bar{d}'_j$  // weighted linear deviations for t
Initialization
H = [] // list of broken anodes
begin
  for j от 1 до k:
    if  $\bar{x}'_j - \bar{x}_j > 0$  and  $\bar{d}'_j - \bar{d}_j > 0$ :
       $H \leftarrow \{id: j, value: (\bar{x}'_j - \bar{x}_j, \bar{d}'_j - \bar{d}_j)\}$ 
    sorted(H, key = value) // sorting anodes in values decreasing order
  end
end

```

Figure 4. Algorithm for identifying “problematic” anodes.

Based on the results of the time series characterization analysis, two indicators of disturbance development are identified: I_f , i.e., the “cell noise” indicator and I_{amp} , i.e., the “anode current” indicator.

The indicators are processed sequentially. The “cell noise” indicator is checked first, then, if abnormal operating conditions ($I_f \neq 0$) are detected and “anode current” data is available, the “anode current” indicator is checked. If changes in “anode current” ($I_{amp} \neq 0$) are recorded, the

anodes with the maximum signal change are identified. Depending on the indicator values, warnings are generated: if $I_f \neq 0$ and $I_{amp} = 0$, a message of “cell noise” signal change and deviation fixing time is issued; if $I_f \neq 0$ and $I_{amp} \neq 0$, a message of “anode current” signal change with a list of “problematic” anodes is issued.

5. Information Analysis System for the Online Control of Aluminium Production Process

The models and algorithms discussed above for detecting process disruptions are implemented in the information-analytical system for operational control of aluminum production technology. The prototype system operates in test mode at the Sayanogorsk Aluminum Smelter's pilot pots. A total of 8 RA-550 and 336 RA-300 reduction cells are monitored.

The system operation is based on integrated monitoring of daily average and instantaneous data and includes two basic tasks:

- 1) Monitoring of process parameters with visualization of the dynamics of changes in their values.
- 2) Warning of the occurrence of process disruptions with indication of the state of control objects, e.g., potroom, reduction cell and the issuing of information messages about the conditions of the process.

In monitoring mode, process parameter values are displayed in real time. All parameters are synchronized on a single, scalable timescale and the parameter values are updated when the control object – namely, potroom and reduction cell – is changed. Figure 5 demonstrates an example of daily average monitoring, while Figure 6, of instant monitoring. Operational control makes it possible to detect deviations of values in real time, to determine trends and conditions of development of anode spikes, while taking into account the mutual influence of the parameters under control.



Figure 5. Example of daily average monitoring parameters.

Top left: voltage, [V], Top center: electrolyte temperature, [°C], Top right: electrolyte level, [cm], Bottom left: overfeed duration, [s], Bottom center: underfeed duration, [s], Bottom right: “test” mode* duration, [s].

* “Test” mode refers to a cell feeding strategy in which the amount of alumina supplied to the electrolyte is significantly less than its consumption.



Figure 6. Example of instant monitoring parameters.
Top row: voltage (V), 2nd row: cell noise, [mV], 3rd row: anode current, [kA],
Bottom row: operating modes of Al₂O₃ point feeding system, [-].

In the warning mode, the probability of disruptions is calculated based on predictive models as new parameter values appear. The status of the technological process is displayed by means of control object indication and information messages with a visualization of the status of each anode. Figure 7 demonstrates an example of indication of irregularities and information messages on the occurrence of disruptions. Message and prediction of anode spikes based on daily average monitoring parameters is displayed at the top of the screen (note that the message with red highlighting reads “metaclassification model predicts anode spike with 86 % probability”). Message and anode status is displayed at the center of the screen. In the form of a diagram, the status of the anodes is displayed, where yellow markers indicate the “development of process disruptions” and red markers, the “occurrence of process disruptions”. The messages below the diagram contain information about the duration of the status and the set of “problematic” anodes.



Figure 7. Example of indication of irregularities and information messages on the occurrence of disruptions. Yellow markers: anodes with “development of process disruptions”, Red markers: anodes with “occurrence of process disruptions”.

5. Conclusions

This paper presents the results of the development of algorithmic and software tools that allow detecting and predicting anode spikes based on the values of monitored process parameters, using a combination of mathematical statistics, data mining and machine learning methods.

Based on the application of ensemble algorithms to average daily monitoring data, voting and metaclassification predictive models have been developed which, by identifying internal patterns in the data, determine the object state corresponding to the occurrence of process disruption. Based on the application of association rules technology, a model has been developed that, by analyzing precedent events and identifying cause-effect relationships, generates rules that determine the state of the technology and the likelihood of an anode spike occurrence.

In order to detect process disturbances from instantaneous monitoring data, algorithms have been developed to identify changes in the time series of monitored parameters and to search for “problematic” anodes.

A software implementation of the models and algorithms has been carried out and a prototype of the information-analytical system for the online control of aluminum reduction technology has been developed. The system includes two tasks: monitoring of process parameters with visualization of the dynamics of changes in their values and warning of the occurrence of the process disruptions with the indication of the state of control objects, *e.g.*, potroom, reduction cell and the issuing of information messages. The developed tools have been tested at the pilot sites of the Sayanogorsk Aluminium Smelter (RA-300 and RA-550 technologies).

Further research is concerned with improving models to improve prediction accuracy, developing algorithms that allow the analysis of a set of parameters while taking into account their mutual influence, and expanding the types of disruptions.

6. References

1. Y.G. Mikhalev *et al.*, Anode processes malfunctions – an overview, *J. Sib. Fed. Univ. Eng. technol.*, Vol. 10, No 5, (2017), 593-606. doi: 10.17516/1999-494X-2017-10-5-593-606.
2. I.I. Puzanov *et al.*, Continuous monitoring of information on anode current distribution as means of improving the process of controlling and forecasting process disturbances, *J. Sib. Fed. Univ. Eng. technol.*, Vol. 9, No 6, (2016), 788-801. doi: 10.17516/1999-494X-2016-9-6-788-801.
3. Y.G. Mikhalev *et al.*, Spikes generation on anode of aluminium reduction cell, *Tsvetnye Metally [Non-ferrous metals]*, No 9, (2018), 43-48. doi: 10.17580/tsm.2018.09.06.
4. B.A. Sadler, Critical issues in anode production and quality to avoid anode performance problems. *J. Sib. Fed. Univ. Eng. technol.*, Vol. 8, No 5, (2015), 546-568. doi: 0.17516/1999-494X-2015-8-5-546-568.
5. I. Goodfellow and Y. Bengio, *A Courville deep learning*, Cambridge: MIT press, 2016, 787 pages.
6. G.J. Williams and S.J. Simoff, *Data mining: theory, methodology, techniques, and applications*, Springer, 2006.
7. P.V. Polyakov *et al.*, P'yanykh, Matematicheskoye modelirovaniye raspredeleniya toka pri nalichii narusheniy na podoshve anoda alyuminiyevogo elektrolizera [Mathematical modeling of current distribution in the presence of deformations on the anode bottom of an aluminum electrolysis cell], *Tsvetnyye metally [Non-ferrous metals]*, No 1, (2019), 25-30.
8. J. Yi *et al.*, A novel framework for fault diagnosis using kernel partial least squares based on an optimal preference matrix, *IEEE Transactions on Industrial Electronics*, Vol. 5, No 64, 2017, 4315-4324.

9. N. Majid *et al.*, Diagnosing faults in aluminium processing by using multivariate statistical approaches, *Journal of Materials Science*, No 47, 2012, 1268-1279.
10. L. Dion *et al.*, Developing a new process indicator based on the relationship between an electrolysis cell impurity balance and its incidents, *Light Metals*, 2014, 829-834.
11. N. Majid *et al.*, Aluminium process fault detection and diagnosis, *Advances in Materials Science and Engineering*, No 2015, 2015.
12. N. Lugovaya, A. Mikhalev and T. Penkova, Application of decision trees to detect process disruptions in aluminum production, *CEUR Workshop Proceedings*. Vol. 2727, 2020, 75-82.
13. A. Korobko *et al.*, Application of association rule mining to detect “spike” disruption in aluminum production, *CEUR Workshop Proceedings*, 2021 (in appear).
14. A. Metus and T. Penkova, Analysis of aluminium electrolysis data in the context of extreme values of technological parameters, *CEUR Workshop Proceedings*, Vol. 2727, 2020, 92-98.
15. T. Penkova, M. Senashova and A. Korobko, Multidimensional analysis of aluminum production monitoring data in basic operation modes, *CEUR Workshop Proceedings*, Vol. 2727, 2020, 128-136.