Real-time control data wrangling for development of mathematical control models of technological processes

N V Vasilyeva, N I Koteleva, E R Fedorova

St. Petersburg Mining University, 2, 21 Line of Vasilyevsky Island, St. Petersburg, 199106, Russian Federation

E-mail: Koteleva_NI@pers.spmi.ru, Vasileva_NV@pers.spmi.ru, Fedorova_ER@pers.spmi.ru

Abstract. The relevance of the research is due to the need to stabilize the composition of the melting products of copper-nickel sulfide raw materials in the Vanyukov furnace. The goal of this research is to identify the most suitable methods for the aggregation of the real time data for the development of a mathematical model for control of the technological process of melting copper-nickel sulfide raw materials in the Vanyukov furnace. Statistical methods of analyzing the historical data of the real technological object and the correlation analysis of process parameters are described. Factors that exert the greatest influence on the main output parameter (copper content in matte) and ensure the physical-chemical transformations are revealed. An approach to the processing of the real time data for the development of a mathematical model for control of the melting process is proposed. The stages of processing the real time information are considered. The adopted methodology for the aggregation of data suitable for the development of a control model for the technological process of melting copper-nickel sulfide raw materials in the Vanyukov furnace allows us to interpret the obtained results for their further practical application.

1. Introduction

The mathematical models are becoming more significant in researching metallurgical processes and solving optimization tasks. They make visible the correlation between input variables and output variables. There are no unified methods for developing these models. This is due to a large variety of control object’s types [1-9]: static and dynamic, continuous and discrete, deterministic and stochastic, etc. Each method starts with data wrangling. There are several methods of technological data wrangling for modeling [7-9], but they are not determined for metallurgical processes.

The goal of this research is to determine the most suitable methods for wrangling technological data for developing and using a mathematical model of the technological process of melting copper-nickel sulfide raw materials in the Vanyukov furnace. The control process data of the Vanyukov process during 5 months were examined. The analysis of the operational control data of the Vanyukov process showed (Figure 1) that the copper content in matte varies between 46 and 68%. The variation range is more than 20% while the average copper content in matte is 58%. This indicates to the heterogeneity of data (a large variety in data relative to a mean value), low predictability of copper content in matte and low stability of the process. The low stability of the process makes the predictability difficult. So the data of these processes have to be prepared before modelling.
2. The variety classification of process data.
The Vanyukov furnace data have a wide range. They are measured in the different scales with a different sampling frequency. There are two main types of technological data: the input data and the output data. The input data are characterized by a furnace load. The output data are characterized by content of main components in the matter. The information about content of main components’ composition in the melting products have 2 hours discretization time range, which is due to the chemical analysis of the melting products. The information about furnace load is characterized by a small range of discretization time (1 minute). Thus, the information about furnace load must be transformed to the range with 2-hour discretization time. For further processing of results, the data range with two-hours discretization time is adopted. When developing the model, it must also be taken into account that moment of furnace load and copper content in a melting product (in matte) corresponding to this furnace load time are spaced in time. This is due to time of the material melting in a furnace. As shown in [7], residence time of material in a furnace is about 75-90 minutes. In addition, up to 30 minutes spent on the chemical analysis of the sample matte. Therefore, loading the furnace at a given time (for example, at 00:00) characterizes the copper content in matte after 105-120 minutes, that is, after 2 hours (at 02:00). In this regard, the data on content of copper content in matte should be shifted to an interval equal to two hours.

3. The analysing of initial processes data
The first stage of wrangling of operational control data for the development of a control model is selection of input and output parameters of the model. Previous studies using statistical methods of data analysis [7] based on the use of dispersion and correlation analysis have shown that the greatest influence on copper content in matte is provided by the following factors: total charge of feed materials, t/h; total flow of air blasting, m$^3$/h, and content oxygen in the air blasting, %. Using these variables, Vanyukov's process is virtually completely controlled and copper content in matte is determined.

After determining the input and the output parameters, it is necessary to perform a correlation analysis of input data to identify factors that have the strongest effect on copper content in matte chosen as the objective function (Table 1).

![Figure 1. Histogram of the distribution of copper content in matte.](image)

**Table 1.** The values of cross-correlation coefficients of main technological parameters of the Vanyukov process (significant coefficients are shown in bold)
As shown by the analysis of correlation coefficients (Table 1), the main correlation "Total charge of feed materials - Content oxygen in air blasting", which determines the depth of physical and chemical transformations to achieve the required matte quality, is expressed quite strongly (correlation coefficient is 0.774). The correlation "Total charge of feed materials - Total flow of air blasting" is also strong (correlation coefficient is equal to 0.669), which shows a good correlation of the flow of the total flow of air blasting required by the technology to the total charge of feed materials. However, none of the input variables has a correlation effect on the copper content in matte (all correlation coefficients have been found to be insignificant).

Let us detailize in more detail the variable "oxygen content in blast" (Fig. 2). Since this parameter ensures the necessary completeness of the implementation of the physico-chemical transformations of the process, it must be sufficiently controlled. But, as follows from Fig. 2, the variation range is too high (25 - 100%).

This instability of the furnace load parameters entails instability of the main indicator of quality of the process - the copper content in matte (Fig. 1). However, the identification of the correlation of two or more signs does not mean that there is a causal relationship between them. Correlation analysis establishes only the presence and strength of a statistical connection, and does not speak of its
direction. However, even the very existence of such connection makes it possible to use more advanced methods of searching for knowledge and patterns in the data under study.

4. Techniques for processing industrial information

Analysis of the initial data of operational control shows that for the development of an adequate control model of the technological process, the available numerical material could not be used, since any real database usually contains errors, inaccurately defined values corresponding to some rare, exceptional situations, and other defects, which can dramatically reduce the efficiency of the model. Such records should be discarded, since even if such “outliers” are not errors, but are rare exceptional situations, they still can hardly be used.

Also for formation of material suitable for development of the control model, the furnace operation data were not taken into account one day before and day after the equipment was idle. In addition, information on the operation of the furnace during the "acceleration" of the process, as well as the processing time of atypical raw materials, was excluded.

As shown by the analysis of the correlation coefficients (Table 2) between loading parameters and copper content in matte, only total flow of air blasting has a correlation effect on copper content in matte.

| Stages of wrangling data | Cross-correlation | Total charge of feed materials – Copper content in matte | Total flow of air blasting – Copper content in matte | Oxygen content in the blast – Copper content in matte |
|--------------------------|-------------------|--------------------------------------------------------|----------------------------------------------------|-----------------------------------------------|
| Initial data sets        | -0.014            | -0.144                                                 | -0.130                                             |                                               |
| Data sets after removing accidental emissions, information one day before and one after the equipment is idle | 0.063            | -0.218                                                 | -0.130                                             |                                               |
| Data sets after removing of the furnace operation data during the acceleration of the process and the processing of atypical raw materials | 0.116            | -0.289                                                 | 0.187                                              |                                               |
| After the data set is divided according to the operating modes | Mode 1            | 0.314                                                  | -0.416                                             | 0.248                                          |
|                          | Mode 2            | 0.263                                                  | -0.398                                             | 0.253                                          |

5. Intellectualization of the method of wrangling operational control data before the development of the control model

In order to intellectualize the data sets for the development of the control model of the technological process, additional studies must be carry out of the array of initial data. The main parameter of the smelting process of copper-nickel sulfide raw materials, which determines the composition of matte, is the total charge of feed materials.

A histogram of the distribution of total charge of feed materials is shown in Fig. 3. The analysis of which shows that during the smelting of copper-nickel sulfide raw materials in the Vanyukov furnace, two operating modes of the process are observed: regime 1 with an average charge of feed materials of about 90 t/h and mode 2 with an average charge of feed materials of about 160 t/h.
Figure 3. A histogram of the distribution of total charge of feed materials

It is considered inexpedient to manage the process using the same algorithm for different loads of a furnace, from the point of view of the process control theory, therefore the original data array must be divided into two according to the operating modes of a furnace: with an average load of 90 t/h and 160 t/h.

The analysis of correlation coefficients (Table 2) confirms that the methodology for the wrangling of the data sets fits for the development of an adequate control model, including the removal of “emissions” and random data; excluding data for a day before and day after equipment downtime; removal of data related to the processing of atypical raw materials; the exclusion of furnace operation data during the starting of process, showed good results for the wrangling of the initial data sets for the development of the control model of the technological process.

6. Conclusion

1. Despite the fact that the relationship between the load materials and all the parameters of blasting is rather close and significantly different from zero, it is not enough to effectively control the process (unstable matte composition), since the range of copper content in matte is too large. This is because copper content in matte, none of the input variables does not have a significant effect (all correlation coefficients were insignificant). The dependence between loading parameters and copper content in matte is weakly expressed, not only because of quality of the process by the technological staff, but also because of inaccurate information coming from devices. Use of such data to develop an adequate control model of the technological process is inexpedient.

2. The considered methods of processing operational control data and the method of wrangling numerical material for further modeling using elements of intellectualization of operational control data make it possible to develop an adequate control mathematical model of the Vanyukov process, and on its basis also a process control algorithm, thereby stabilizing copper content in matte in the specified range.

References

[1] Oprea G, Andrei H 2016 Power quality analysis of industrial company based on data acquisition system, numerical algorithms and compensation results. International Symposium on Fundamentals of Electrical Engineering, ISFEE 2016
[2] Le V T, Paris H, Mandil G 2017 Process planning for combined additive and subtractive manufacturing technologies in a remanufacturing context. *Journal of Manufacturing Systems*, 44 243-254

[3] Cheng Z, Zhang H-T, Fan M-C, Chen G 2015 Distributed consensus of multi-agent systems with input constraints: A model predictive control approach *IEEE Transactions on Circuits and Systems I: Regular Papers* 62 (3) 825-834

[4] Liu Z, Meng B, Ji G, Zhu Z, Pan J 2011 Stability analysis and control of welding current for automatic rail narrow-gap arc welding system. *Journal of Mechanical Engineering* 47 (8) 110-115

[5] Thombsansen U, Purrio M, Buchholz G, Hermanss T, Molitor T, Willms K, Schulz W, Reisgen U 2016 Determination of process variables in melt-based manufacturing processes. *International Journal of Computer Integrated Manufacturing* 29 (11) 1147-1158

[6] Safina E, Khokhlov S 2017 Paradox of alternative energy consumption: Lean or profligacy? *International Journal for Quality Research* 11 (4) 903-916

[7] Kadyrov E D, Danilova N V 2013 Variables estimation in autogeneous processes. *Automation and Remote Control* 74-6 1037-1042

[8] Kadyrov E D, Koteleva N I 2011 Introducing neural-network algorithms into an automated system designed to control metallurgical processes. *Metallurgist*. 54 11-12, 799-802

[9] Pervuchin D A, Klavdiev A A 2016 Assessment of complicated systems quality on the basis of stochastic similarity methods. *Journal of Mining Institute* 214 85

[10] Iakovleva E V, Momot B A 2017 Development of technology for creating intelligent control systems for power plants and propulsion systems for marine robotic systems. *IOP Conference Series: Earth and Environmental Science* 87 (3)