EVALUATION OF ROBUSTNESS AND EFFICIENCY OF CONTROL CHARTS

SMAJDOROVÁ Tereza, NOSKIEVIČOVÁ Daria

VSB - Technical University of Ostrava, Faculty of Metallurgy and Materials Engineering, Ostrava, Czech Republic, EU, tsmajdorova@seznam.cz, darja.noskievicova@vsb.cz

Abstract

Control chart is the basic tool of the statistical process control (SPC). It aims to an early detection of errors in the process and thereby ensures compliance with the required level of stability. The statistical process control is an integral part of the production management necessary for achieving a high product quality. Just the quality of the product decides the customer satisfaction and thus the success of the whole company. Classical Shewhart control charts can be used only if there are met certain basic assumptions. These assumptions include, for example, data normality, their independence and constant mean and variance. In practice, such as the metallurgical industry, those assumptions about the data are not necessarily always met. In the case that these conditions are not met, there may be used non-parametric and robust control charts. This paper presents some of these non-traditional control charts. This article aims to define the difference between robust and non-parametric methods. Another aim of this article is to present the possibility of evaluating the robustness, and the evaluation of effectiveness based of individual control charts. Particular evaluation methods are complemented by practical examples from the metallurgical process. Conclusion of the article includes comparisons of the used control charts, both in terms of robustness and effectiveness. During preparation of this article accessible pieces of knowledge on the issue were compared, including the use of SPC in the metallurgical industry. This article is the basis for further examination of the problem, including a more detailed processing of the software support.

Keywords: Robustness, effectiveness, statistical process control, non-parametric control charts, steelmaking process

1. INTRODUCTION

Statistical process control (SPC) is an immediate and continuous process control based on the mathematical-statistical evaluation of the product quality. If a company wants to achieve the high quality consistently, it has to collect, process and analyze systematically data available from the production and conclusions of the analysis must be used for continuous improvement. Statistical process control allows interventions in the process based on the early detection of deviations from a predetermined process level. It is implemented by regular monitoring of the controlled process variable or output variable. It is found out whether it corresponds to the process level required by the customer. Achieving the desired level of the process requires a thorough analysis of the process variability. To use the classic Shewhart control charts, the certain basic assumptions about the data must be met. In the manufacturing practice, however, it is not always possible to meet these basic assumptions (normal distribution, constant mean and variance, independence of data). This article aims to define the difference between robust and non-parametric methods. Another aim of this article is to present the possibility of evaluating the robustness, and the evaluation of effectiveness based of individual control charts. Particular evaluation methods are complemented by practical examples from the metallurgical process.

Typical breaches of assumptions in some industries are shown in **Table 1**. [5] The "x" denotes a violation of the assumption, and the "-" designation then determines the predominant fulfillment of the assumption. As in the metallurgical industry, also elsewhere, it is necessary to perform various measurements. It may be a measurement of length, weight, hardness, strength, temperature, pressure or chemical composition. The measurement takes place both in the pre-production phase and in the inter-operational and the final inspection. The data obtained must then be analyzed and conclusions used to improve the quality of the product. [6]

Table 1 Typical breaches of assumptions [5]

Industry / technology / quantity	Normality	Independence	Constant mean	Constant variance
Mechanical engineering, automotive industry (dimension)	-	-	-	-
Mechanical tests (strength, flexibility,...)	x	-	-	x
Chemistry, metallurgy, (concentration, contents)	-	x	x	x
Chemistry, metallurgy, (other physical parameters)	-	x	x	x
Environment (different concentrations)	x	x	x	x
Elictrical quantities	-	-	-	x
Energy	x	x	x	x
Plastics, polymers, textiles, physico-mechanical quantities	x	-	x	-
Biochemistry, pharmacy, food industry	x	x	-	-
Ekonomic and financial indicators	x	x	x	-
Sociology, human resources	x	x	x	x

2. ROBUST AND NONPARAMETRIC METHODS

Why should robust and non-parametric methods be developed at all? Typical statistical procedures are largely parametric, that is, the use of the model is dependent on many parameters, in large part they are the probability distribution parameters.

The opposite of the parametric methods are non-parametric methods. These are independent or little depending on the probability distribution shape. These methods have good properties for a whole range of distributions, but for this versatility we have to pay the tax, and this is a loss of yield (they are more sensitive to changing parameters).

Compared to nonparametric methods, robust methods have the advantage of preserving good properties around a certain baseline probability distribution. And unlike nonparametric methods, they are more profound.

2.1. Robust methods

There are currently many definitions of robustness. Generally, robust means insensitive to small deviations from idealized assumptions for which the estimation is optimized. In most cases, we consider robustness in relation to deviations from the predicted distribution. However, there are other types of robustness, such as deviations from the observation independence assumption.

In order to compare the robustness of different methods with each other, it is necessary to quantify it in some way, i.e. to characterize it by a certain number. However, replacing a complex term with a single number is one-sided and simplifying. There are a number of quantifications, one of which is the breakdown point. [3, 7].

BREAKDOWN POINT

It is absolutely clear that the data will not always meet the assumption of normality, so that no one should ever use parametric methods or use them so long as the probability distribution is not too nonnormal. But what is too nonnormal if we cannot quantify this anomaly because there are plenty of options to be nonnormal. It is much better to quantify the properties of the estimator and their associated procedures. The basic tool used to describe robustness is the breakdown point. The greater the breakdown point is the more robust the method. The breakdown point cannot be larger than 50%. The average has a breakdown point equal 0, on the other hand the median breakdown point is 0.5. [3, 7]

ROBUST CONTROL CHART MAD

An example of robust methods has been selected control chart MAD. The control chart is based on the mean absolute deviation from the median (MAD - median absolute deviation). This characteristic is more robust than the determinant standard deviation. MAD for random selection of n is defined [3, 4]:

$$MADj = \frac{1}{n} * \sum_{i=1}^{n} median \; |X_i - MDj|; \;\; i = 1,2,\ldots,n \tag{1}$$

where *MD* is selective median

$$LCL = c_4\hat{\sigma} - 3\hat{\sigma}\sqrt{1-c_4^2} = c_4 b_n \overline{MAD} - 3b_n \overline{MAD}\sqrt{1-c_4^2} = B_5 b_n \overline{MAD} = B_5^* \overline{MAD} \tag{2}$$

$$CL = c_4\hat{\sigma} = c_4 b_n \overline{MAD} = c_4^* \overline{MAD} \tag{3}$$

$$UCL = c_4\hat{\sigma} + 3\hat{\sigma}\sqrt{1-c_4^2} = c_4 b_n \overline{MAD} + 3b_n \overline{MAD}\sqrt{1-c_4^2} = B_6 b_n \overline{MAD} = B_6^* \overline{MAD} \tag{4}$$

$$\overline{MAD} = \sum_{j=1}^{m} MAD_j \Big/ m \tag{5}$$

Value c_4^*, B_5^*, B_6^* can be found in the table. [3, 4]

2.2. Nonparametric methods

In non-compliance with some assumptions, it is possible to apply nonparametric methods. Nonparametric methods are based on a smaller number of observations. Compared to the model based methods most often it is only assumed that the probability distribution of the given data set is of the continuous type.

Nonparametric methods have, compared to parametric methods, a number of advantages:

- conclusions obtained are independent of the distribution shape,
- they can be used even when the type of distribution is unknown,
- they have a greater robustness to the occurrence of outliers.

Disadvantages of non-parametric methods include the increased probability of missing signal, which means that it often leads to incorrect non-rejection of untrue null hypothesis. This probability can be reduced by increasing the sample size. [4, 8, 9]

NONPARAMETRIC CONTROL CHARTS

The following **Table 2** is a summary of some non-parametric control charts, including formulas for calculating control limits and other necessary characteristics. [4]

Table 2 Calculations for nonparametric control charts [4]

CC	LCL	CL	UCL	Charakteristics
SSCC	$LCL = -1 \cdot (2 \cdot t - n)$	$CL = 0$	$UCL = 2 \cdot t - n$	$SN_i = \sum_{j=1}^{n} sign(x_{ij} - \theta_0)$
NP-CUSUM			$UCL = h$ (from table)	$S_j(m,n) = \max \{0, S_{j-1}(m,n) + SMW_{j,(m+n)} - k\}$
NP-EWMA	$LCL = \frac{n}{2} - k * \sqrt{\frac{\lambda}{2-\lambda} * \left(\frac{1}{4n}\right)}$	$CL = \frac{n}{2}$	$UCL = \frac{n}{2} + k * \sqrt{\frac{\lambda}{2-\lambda} * \left(\frac{1}{4n}\right)}$	$G_t = \lambda N_t + (1-\lambda) G_{t-1}$
NP-PM	$LCL_t = \mu_0 - 3 * \frac{\sigma_0}{\sqrt{t}}$	$CL_t = \mu_0$	$UCL_t = \mu_0 + 3 * \frac{\sigma_0}{\sqrt{t}}$	$PM_t = \frac{X_1 + X_2 + \ldots + X_t}{t} = \frac{\sum_{j=1}^{t} X_j}{t}$
NP-Mood	$LCL = E(M_{m,n}) - c * \sqrt{\text{var}(M_{m,n})}$		$UCL = E(M_{m,n}) + c * \sqrt{\text{var}(M_{m,n})}$	$M_{m,n} = \sum_{i=1}^{m} \left(R_i - \frac{N+1}{2}\right)^2,$ $kde \quad N = m + n$

3. EFFICIENCY ASSESSMENT

By valuating the efficiency, it is meant the ability of the control chart detect the change process parameters. In order to assess the efficiency of the classic Shewhart control charts, we can use the ARL (average run length), which is the average number of selections leading to the signal [6]. Another option is to calculate the probability of exceeding the *p* limits by the formula:

$$p = \frac{P(A)}{P(B)} \tag{6}$$

where *P(A)* is the number of points outside the limits and *P(B)* is the total number of points.

4. EXAMPLE - STEELMAKING PROCESS

The following example illustrates the application of nonparametric and robust control charts Chart on the data obtained from the steelmaking process. The measured values are recorded in the **Table 3**, in the columns x_1 to x_5.

Table 3 Data from steelmaking process

Subgroup	x_1	x_2	x_3	x_4	x_5	Subgroup	x_1	x_2	x_3	x_4	x_5
1	33.63	32.75	32.37	33.28	32.76	11	32.48	33.54	31.84	32.18	32.23
2	33.62	32.00	31.33	34.04	32.34	12	32.94	32.74	32.69	32.29	32.64
3	32.93	33.49	32.75	32.75	32.97	13	32.87	32.87	33.28	33.65	32.78
4	33.49	33.47	32.87	32.52	33.15	14	32.07	33.36	32.77	33.24	34.04
5	32.37	32.78	34.13	32.94	33.12	15	33.56	32.40	32.86	31.68	33.02
6	33.44	31.24	32.50	33.50	33.53	16	32.25	31.24	30.75	31.94	32.64
7	33.79	33.54	33.33	33.73	32.07	17	32.89	33.19	33.54	34.23	32.25
8	31.37	33.01	33.24	32.81	33.24	18	32.58	33.13	33.24	31.97	32.66
9	32.03	31.77	31.92	32.30	32.68	19	32.74	31.97	32.53	33.70	32.93
10	33.26	33.05	32.53	33.27	32.33	20	33.68	32.69	32.25	32.56	33.55

Probability of exceeding the limits was determined from the final control charts. These values are shown in **Table 4** below.

Table 4 The probability of exceeding the limits

Control Charts	The probability of exceeding the limits
Shewhart Sign Control Chart	0.55
Nonparametric EWMA	0.9
Nonparametric CUSUM	0
Nonparametric Progressive Mean	0
Nonparametric Control Chart based on Mood statistics	0.95
Robust Control Chart MAD	0

The results show that the Robust Control Chart MAD, Nonparametric Control Chart Progressive Mean and Nonparameric Control Chart CUSUM have the lowest risk of false signal and are the most effective. On the contrary, the worst is the Nonparametric Control Chart based on Mood statistics and Nonparametric EWMA that is the least effective because it has the highest risk of a false signal.

5. CONSLUSION

The aim of the work is more detailed analysis of nonparametric control charts applied on different dates in violating of the basic assumptions and determination of the procedures for their use in practice. As can be seen in **Table 1**, in metallurgy is often a breach of the assumptions of independence and non-constant of mean and variance. The results will contribute to the development of statistical process control and process capability analysis. The proposed methodology could help in the decision-making processes in practice of metallurgical enterprises.

ACKNOWLEDGEMENTS

The work was supported by the specific university research of the Ministry of Education, Youth and Sports of the Czech Republic No. SP2017/63.

REFERENCES

[1] ABU-SHAWIESH, M.O.A. A Simple Robust Control Chart Based on MAD. *Journal of Mathematics and Statistics*, 2008, vol. 4, no. 2, pp. 102-107.

[2] ADEKEYE, K.S. Process Capability Indices Based on Median Absolute Deviation. *International Journal of Applied Science and Technology*. 2013, vol. 3, no. 4, pp. 6.

[3] GEYER, C. J. *Breakdown Point Theory Notes*. 2006. Available from: http://www.stat.umn.edu/geyer/5601/notes/break.pdf.

[4] CHAKRABORTI, S., VAN DER LAAN, P., BAKIR, S. T. *Nonparametric Control Charts: An Overview And Some Results*. 2001. Available from: http://207.67.83.164/pub/jqt/past/vol33_issue3/qtec_33_3_304.html.

[5] MELOUN, M. *Kontrola a řízení jakosti*. Available from: https://meloun.upce.cz/docs/research/chemometrics/methodology/10metody.pdf.

[6] NENADÁL, J. *Moderní management jakosti: principy, postupy, metody*. Praha: Management Press, 2008. 377 p.

[7] ROUSSEEUW, P. J., HUBERT, M. Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2011, vol. 1, no. 1, pp. 73-79. Available from: http://doi.wiley.com/10.1002/widm.2

[8] ŠTIGLIC, M. *Neparametrické štatistické metódy a ich ekonomické aplikácie*.2009. Available from: http://web.ics.upjs.sk/svoc2009/prace/3/Stiglic.pdf.

[9] ZVÁROVÁ, J. *Základy statistiky pro biomedicínské obory*. Praha: Karolinum, 2011, 219 p.