THE INFLUENCE OF SKEWNESS OF THE MONITORED QUALITY CHARACTERISTIC DISTRIBUTION ON THE PROCES CAPABILITY ANALYSIS RESULTS

Jiří PLURA, Karel NEPRAŠ, Pavel KLAPUT

VSB-Technical University of Ostrava, Faculty of Metallurgy and Materials Engineering, Department of Quality Management, Ostrava, Czech Republic, EU, jiri.plura@vsb.cz, karel.nepras.st@vsb.cz, pavel.klaput@vsb.cz

Abstract

One of the standard requirements for automotive industry suppliers within the framework of production part approval process is the evidence about capability of manufacturing process. Process capability criteria are capability indices whose standard evaluation is based on the assumption of normal distribution of monitored quality characteristic. However, the distribution of a series of quality characteristics such as tensile strength, hardness, roughness, etc. does not correspond to a normal distribution. The paper deals with the influence of skewness of the monitored quality characteristic distribution on the process capability analysis results. They are applied different probability distributions for process capability analysis and the achieved results are compared and discussed.

Keywords: process capability analysis, non-normal quality characteristic, influence of skewness

1. INTRODUCTION

The evaluation of process capability standardly uses capability indices based on the assumption of normal distribution of the monitored quality characteristic [1, 2]. In real practice, however, there are various quality characteristics, which natural distribution is not a normal one. They include, for example, taper ratio, flatness, roughness, concentricity, alignment, perpendicularity, waviness, straightness, rectangularity, weld strength, tensile strength, casting hardness, hole positions, chamfering, or parallelism. Even in these cases, it is necessary to find suitable procedures of process capability analysis.

2. PROCESS CAPABILITY ANALYSIS WHEN DATA NORMALITY IS NOT MET

The situation where the distribution of the monitored quality characteristic does not correspond to the normal distribution can, in principle, be addressed in the following ways [3]:

- data transformation to a variable corresponding to normal distribution
- using another theoretical distribution model
- using indicators that are not based on a concrete model of distribution

However, before using one of the possible procedures to address the issue of not meeting the normality, it is desirable to pay attention to its causes. The failure to meet normality may be caused by the occurrence of outliers, arising as a result of gross errors of the measurement or data inhomogeneity caused by the change of conditions during data collection. In such cases, it is necessary to eliminate the causes or to perform suitable data sorting or to collect new data.

2.1 Transformation of data into a variable corresponding to normal distribution

The data transformation process can take advantage of the fact that a suitable transformation function can help us to convert the measured data to a variable that can meet the normality, and this transformed variable can be used for further evaluation. If it is not possible to find a suitable transformation function that would ensure sufficient compliance of the transformed data distribution with the normal distribution, it is necessary to apply other procedures of capability analysis.

2.2 The utilization of another theoretical distribution model

Another way to solve the problem of non-normality of the monitored quality characteristic is to find another suitable model of probability distribution used to describe the distribution of the monitored quality characteristic. To verify the suitability of the chosen theoretical model, it is necessary to apply some of the goodness of fit test. This procedure may not be successful, just like the data transformation procedure.

The so-called quantile method [1, 4, 5] is most frequently used to evaluate process capability in case of the applications of another distribution model of the monitored characteristic. As in the case of normal distribution, where the values of μ - 3σ and μ + 3σ match the quantiles whose distribution function reaches the values of 0.00135 and 0.99865, in the case of another probability model, you are looking for quantiles corresponding to these distribution function values. The corresponding capability index values of C_p and C_{pk} can then be calculated as:

$$C_p = \frac{USL - LSL}{x_{0,99865} - x_{0,00135}} \qquad (1)$$

$$C_{pk} = \min\left\{\frac{x_{0,5} - LSL}{x_{0,5} - x_{0,00135}};\frac{USL - x_{0,5}}{x_{0,99865} - x_{0,5}}\right\} \qquad (2)$$

where:

USL	- upper specification limit
LSL	- lower specification limit
$x_{0.00135}$	- 0.135% quantile of corresponding probability distribution
$x_{0.99865}$	- 99.865% quantile of corresponding probability distribution
$x_{0.5}$	- median of corresponding probability distribution.

2.3 The use of the indicators that are not based on a concrete probability distribution model

In the event of an unsuccessful data transformation and a failure to find another suitable distribution model of the monitored quality characteristic, the process capability can be evaluated using indicators that are used in the process capability analysis in case of non-measurable quality characteristics. These are the indicators based on the identified proportion of nonconforming units, such as ppm, Sigma level, Equivalent C_p, Equivalent C_{pk}, etc. In addition to that it is possible to use some special capability indices based, for example, on the description of the monitored characteristic distribution using empirical functions, but these have not been used in practice yet.

3. THE EVALUATION OF THE INFLUENCE OF SKEWNESS ON THE RESULTS OF PROCESS CAPABILITY ANALYSIS

3.1 Data preparation

For assessing the impact of skewness of the monitored quality characteristic distribution on the results of the process capability analysis data files from normal and lognormal distribution (100 values each) were generated firstly with using Minitab software. On the basis of this fundamental data eleven data files with different skewness from 0.19 to 0.86 were prepared, while the same average value and standard deviation were maintained in all files. The prepared data files were considered to be the values of the monitored quality characteristic (25 subgroups with the size of 4) collected for the process capability analysis. The relevant tolerance limits were set out as follows: LSL = 3; USL = 23.

The values of the sample characteristics of the individual data files are listed in **Table 1**. The table clearly shows that the change of skewness of the distribution of the individual files is accompanied by a change in the kurtosis. The increasing value of skewness is also accompanied by a decrease in the median and an increase in the range.

Table 1 The sample characteristics of the analyzed data files

	Data file										
	D1	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**	**D11**
Average	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
SD	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
Skewness	0.19	0.25	0.31	0.38	0.44	0.51	0.57	0.64	0.71	0.79	0.86
Kurtosis	-038	-0.30	-0.21	-0.11	0.00	0.11	0.24	0.37	0.51	0.66	0.81
Median	9.76	9.74	9.71	9.69	9.66	9.64	9.61	9.61	9.59	9.53	9.46
Min	3.92	3.93	3.95	3.97	3.99	4.01	4.04	4.06	4.09	4.12	4.15
Max	17.93	18.04	18.14	18.25	18.35	18.45	18.55	18.64	18.74	18.83	18.92
Range	1402	14.11	14.19	14.28	14.36	14.44	14.51	14.58	14.64	14.71	14.77

Exploratory analysis of the individual data files was performed prior to the actual process capability analysis and statistical stability of the simulated processes was verified. It was found that all the data files correspond to the processes which are "in control", so it was possible to assess their capability.

3.2 Testing normality and goodness of fit tests

The verification of normality of the monitored quality characteristic was performed using the Anderson - Darling test in the Minitab software. Based on the assessed p-values (see **Table 1**), it was found that for files with the skewness lower than or equal to 0.57 (files D1 to D7), it is possible to accept the hypothesis, that data are corresponding to normal distribution at a significance level of 0.05. Another theoretical model of probability distribution using the Anderson - Darling goodness of fit test was required for the remaining four sets. The goodness of fit tests were performed with lognormal distribution, three-parameter Weibull distribution, extreme values distribution, Gamma distribution, loglogistic distribution, and Burr distribution. The goodness of fit tests for these distributions were then applied to the remaining data files. As for the Burr distribution, the goodness of fit test was performed using the Kolmogorov - Smirnov test in Easy Fit software (Minitab software does not work with Burr distribution). The Easy Fit program with high p-values has confirmed the suitability of Burr distribution for all analyzed files; however, the p-values are determined in a slightly different way, so they are not entirely comparable with the other p-values.

An overview of the determined p-values of goodness of fit tests for all data files and various probability distribution models are presented in **Table 2**. The overview clearly shows that the distribution of the monitored quality characteristic within the given skewness range can be described by various theoretical distribution models. The best compliance with the theoretical model is achieved with normal and lognormal distribution (if we do not take into account the Burr distribution).

3.3 Process capability analysis

A quantile method based on finding a suitable theoretical probability model to describe the distribution of the collected data was chosen for the process capability analysis. The values of the necessary quantiles for the determined suitable distributions were found using Minitab and Easy Fit software. These values were then used to calculate the values of C_p and C_{pk} indices according to equation (1) and (2). The values of these indices were subsequently analyzed according to the skewness of the individual data files.

Table 2 P-values of Anderson-Darling goodness of fit tests for various probability distributions

		Data file										
		D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11
Probability distribution	Normal	0.711	0.754	0.698	0.551	0.378	0.200	0.086	0.029	0.008	<0.005	<0.005
	Lognormal	0.068	0.147	0.287	0.484	0.675	0.832	0.885	0.867	0.749	0.531	0.314
	Weibull-3par.	>0.500	>0.500	>0.500	>0.500	>0.500	>0.500	>0.500	0.429	0.198	0.075	0.023
	Extreme Values	0.065	0.134	0.231	>0.250	>0.250	>0.250	>0.250	>0.250	>0.250	>0.250	>0.250
	Gamma	>0.250	>0.250	>0.250	>0.250	>0.250	>0.250	>0.250	>0.250	>0.250	>0.250	0.131
	Loglogistic	>0.250	0.152	>0.250	>0.250	>0.250	>0.250	>0.250	>0.250	>0.250	>0.250	<0.005
	Burr *)	0.915	0.978	0.989	0.999	0.996	0.992	0.985	0.977	0.966	0.952	0.936

*) p-values of Kolmogorov-Smirnov goodness of fit tests determined by means of Easy Fit software

The dependences of C_p index determined on the basis of the application of different probability models related to the skewness of the monitored characteristic distribution are illustrated in **Fig. 1**. It clearly shows that the type of the used distribution model has quite significant impact on the C_p index value. C_p values depending on the type of used distribution model range from 0.67 to 1.23 in case of the analysed data files. The best match with the C_p index for normal distribution (within its usability area) is evident for Gamma distribution.

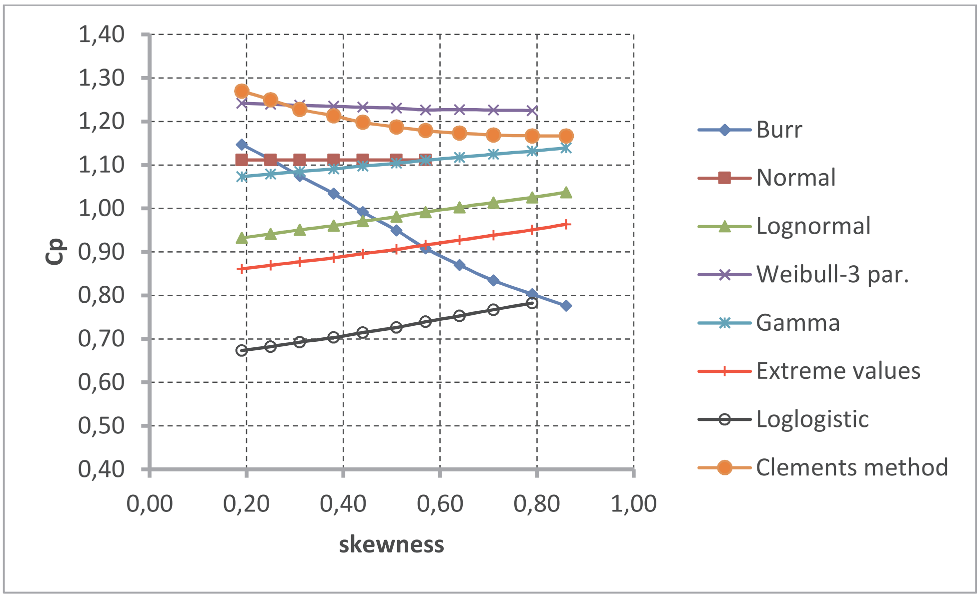

Fig. 1 The dependences of C_p index on skewness for various probability models

In most of the used distribution models, the increasing skewness of distribution of the monitored characteristic goes hand in hand with a slight increase of the C_p index. The dependence process is completely opposite in case of Burr distribution and, for example, in three-parameter Weibull distribution, the value of C_p index is practically unchanged. To make a comparison, there are C_p indices determined using the Clements method [6]. The values of these indices are among the highest, and they decrease slightly with increasing distribution skewness.

Lower values of the C_p index in practically all probability distribution in comparison with normal distribution (excluding the Weibull distribution) are related to longer distances between the quantiles $x_{0,99865}$ and $x_{0,00135}$ of these distributions. For example, a comparison of the C_p values shows that the distance of these quantiles in case of loglogistic distribution is up to 66% longer than in normal distribution.

Fig. 2 compares the dependences of C_{pk} index on the distribution skewness for various theoretical models of probability distribution. The figure, again, clearly shows that the used model of probability distribution has significant influence on the determined values of C_{pk} index, their values range from 0.58 to 1.19.

In most cases, the C_{pk} index is slightly increasing with increasing distribution skewness. The only exceptions are the dependencies of C_{pk} index evaluated using the Burr distribution model and the Clements method, where the C_{pk} initially increases with increasing skewness, but once a certain level has been reached, the value is decreasing. This is due to the fact that with lower values of skewness, C_{pk} corresponds to a partial index C_{pL}, and with higher skewness values, to a partial C_{pU} index. The fact that C_{pk} matches C_{pU}, however, does not automatically mean a decreasing dependence, since this situation occurred even when using lognormal, loglogistic distribution and in case of extreme values distribution.

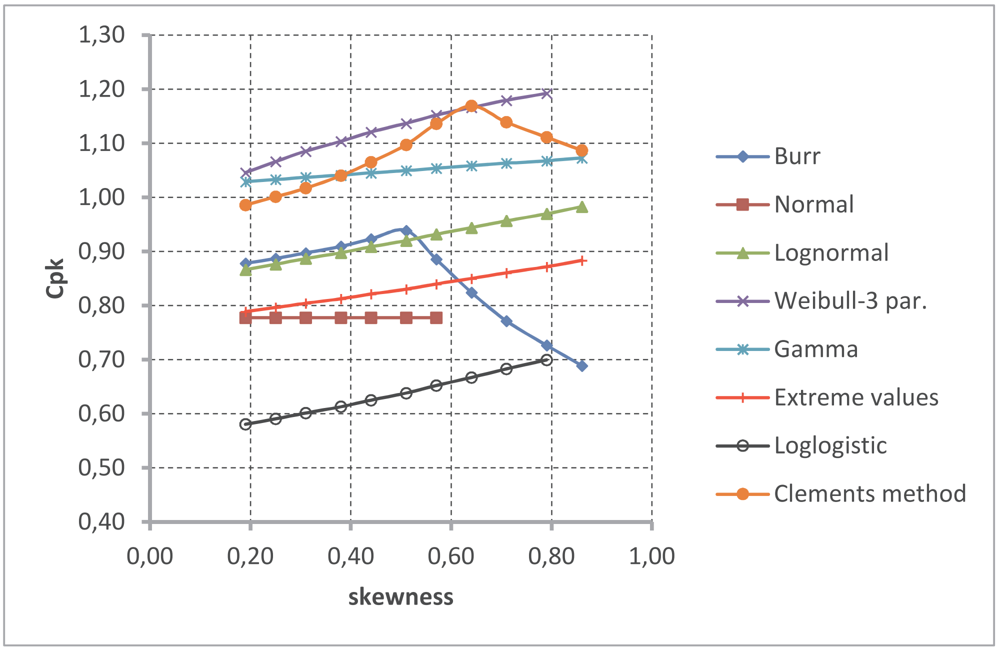

Fig. 2 The dependences of C_{pk} index on skewness for various probability models

Fig. 3 shows the values of C_p and C_{pk} indices determined by applying the most appropriate probability distribution model for the individual files. The suitability criterion was used to determine the p-value of the Anderson - Darling goodness of fit test (see **Table 2**). In the area of lower skewness (up to 0.38), this most suitable model was normal distribution, in higher skewness area, it was lognormal distribution (Burr distribution was not included in the selection due to incomparability of the results). The figure clearly shows that the change of the used distribution model is accompanied by a step change of both indices. While the value of C_p significantly drops with higher skewness, the value of C_{pk} rises significantly. From a practical point of view, it always seems to be more suitable to use, provided that the normality condition is met, the normal distribution model and to look for another theoretical model only when the normality condition is not met. The situation in **Fig. 3**, however, shows that the use of a more appropriate distribution model could lead to more positive conclusions regarding the process capability in such a situation.

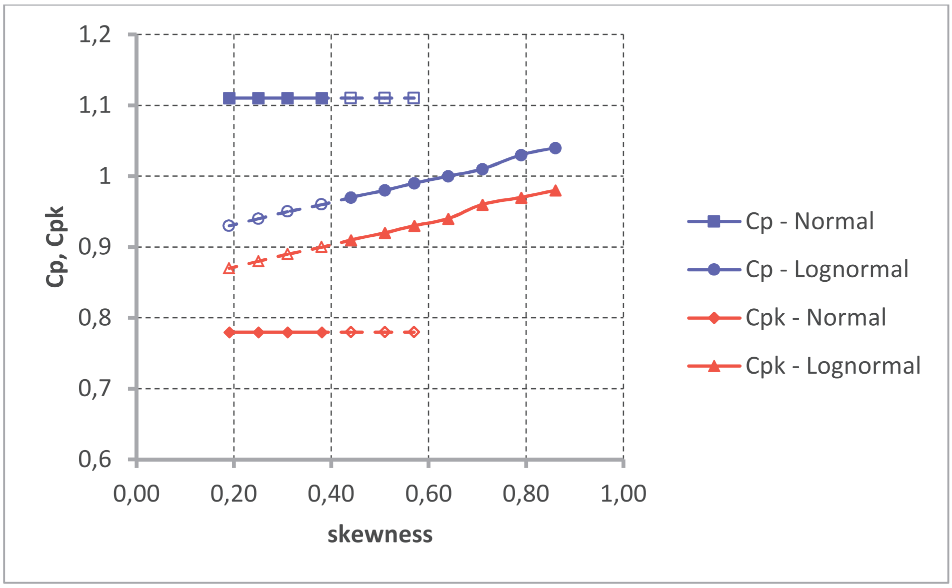

Fig. 3 The values of C_p and C_{pk} evaluated on the basis of the best probability distribution model for data files with different skewness

CONCLUSION

The study clearly shows that the results of the process capability analysis for skewed data in the monitored area depend more on the type of the used probability distribution model, than on the distribution skewness. These are the reasons why it is advisable to use only a single model, namely the one with the best compliance with the actual distribution of the monitored characteristic. Even here, however, you can expect that the change of the type of distribution will cause step changes of the capability indices.

ACKNOWLEDGEMENTS

This paper was elaborated in the frame of the specific research project No. SP2014/81, which has been solved at the Faculty of Metallurgy and Materials Engineering, VSB-***TU Ostrava with the support of Ministry of Education, Youth and Sports of the Czech Republic.***

REFERENCES

[1] KOTZ, S., LOVELACE, C.R. *Process Capability Indices in Theory and Practice*. New York: Oxford University Press, 1998, ISBN 0-340-69177-8.

[2] PLURA, J. Procedures and Methods of Quality Planning and their Use for Forming Process Optimization. In *Engineering the Future*. Laszlo Dudas (editor). 1st edition. Rijeka: Sciyo, 2010, Chapter 13, pp. 257-279, ISBN 978-953-307-210-4.

[3] PLURA, J., ZEMEK, M., KLAPUT, P. Approaches to the Process Capability Analysis in the Case of Non-Normally Distributed Product Quality Characteristic. In *METAL 2013: 22nd International Conference on Metallurgy and Materials.* Ostrava: Tanger, 2013, pp. 1700-1705. ISBN 978-80-87294-41-3.

[4] ISO 22514-1 *Statistical Methods in Process Management - Capability and Performance Part 1 - General Principles and Concepts*, ISO, 2009.

[5] LOON CHING TANG, SU EE THAN Computing Process Capability Indices For Non-Normal Data: A Review and Comparative Study. *Quality And Reliability Engineering International*, 1999, Vol. 15, pp. 339-353.

[6] CLEMENTS, J.A. Process Capability Calculations for Non-normal Distributions. *Quality Progress*, ASQC, 1999, pp. 95-100.