

Methodology for the Application of Nonparametric Control Charts into Practice

Tereza Smajdorová^{a*}, Darja Noskievičová^a

^a VSB – Technical University of Ostrava, Faculty of Materials Science and Technology, The Department of Quality Management, 17.listopadu 2172/15, Ostrava-Poruba, 708 00, Czech Republic

Abstract

Classical parametric statistical methods are based on several basic assumptions about data (normality, independence, constant mean and variance). Unfortunately, these assumptions are not always fulfilled in practice, whether due to problems arising during manufacturing or because these properties are not typical for some processes. Either way, when we apply parametric methods to such data, whether Shewhart's or other types of parametric control charts, it is not guaranteed that they will provide the right results. For these cases, reliable nonparametric statistical methods were developed, which are not affected by breaking assumptions about the data. Nonparametric methods try to provide suitable procedures to replace commonly used parametric statistical methods. The aim of this paper is to introduce the reader to an alternative way of evaluating the statistical stability of the process, in cases where the basic assumptions about the data are not met. First, possible deviations from the data assumptions that must be met in order to use classical Shewhart control charts were defined. Subsequently, simulations were performed to determine which nonparametric control chart was better suited for which type of data assumption violation. First, simulations were performed for the in-control process. Then simulations for an out-of-control process were performed. This is for situations with an isolated and persistent deviation. Based on the performed simulations, flow charts were created. These flow charts give the reader an overview of the possibilities of using nonparametric control charts in various situations. Based on the performed simulations and subsequent verification of the methodology on real data, it was found that nonparametric control charts are a suitable alternative to the standard Shewhart control charts in cases where the basic assumptions about the data are not met.

Keywords:

Nonparametric Control Chart;
Statistical Process Control;
Simulations of In-control Process;
Simulations of Out-of-control Process.

Article History:

| | | | |
|-------------------|----|--------|------|
| Received: | 04 | May | 2020 |
| Accepted: | 25 | July | 2020 |
| Published: | 01 | August | 2020 |

1- Introduction

Classical parametric statistical methods are based on several basic assumptions about data (normality, independence, constant mean and variance). Unfortunately, these assumptions are not always fulfilled in practice, whether due to problems arising during manufacturing or because these properties are not typical for some processes. Either way, when we apply parametric methods to such data, whether Shewhart's or other types of parametric control charts (CC), it is not guaranteed that they will provide the right results. For these cases, reliable nonparametric statistical methods were developed, which are not affected by breaking assumptions about the data [1, 2].

Nonparametric methods try to provide suitable procedures to replace commonly used parametric statistical methods. Nonparametric methods are not affected by outlying values or other deviations from model assumptions [1]. Nonparametric procedures are used when the above data assumptions are not met [3].

* **CONTACT:** Tsmajdorova@seznam.cz

DOI: <http://dx.doi.org/10.28991/esj-2020-01230>

© 2020 by the authors. Licensee ESJ, Italy. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<https://creativecommons.org/licenses/by/4.0/>).

The research area of nonparametric SPC methods is extensive and offers many opportunities for further research, including the improvement of SW support. Nonparametric methods also have a big potential for the process monitoring in the condition of Industry 4.0.

In practice, nonparametric methods are not known, which is due, among other things, to the lack of comprehensive guidance for their application. The aim of this paper is to present the methodology developed for the application of nonparametric control charts (NPCC) that was created on the basis of simulation results. An overview of the advantages and disadvantages of the nonparametric control charts that were analyzed in this study is given in Table 1. Nonparametric control charts for monitoring position SSCC [4], NP-EWMA [5], NP-CUSUM [6] and PM [7] and NPCC for monitoring variability Mood [8, 9] and MAD [10, 11] have now been analyzed.

Compared to other studies, this has a complex nature. In comparison with other sources, a more significant number of nonparametric control charts is assessed here, at the same time all possible deviations from the data assumptions are covered (using various types of probability distributions) and several types of performance indicators were used (not only obviously applied Average Run Length ARL but also more robust Median Run Length MRL and quantiles x_5 and x_{95}).

Table 1. Advantages and disadvantages of nonparametric control charts.

| Nonparametric Control Chart (NPCC) | Advantages | Disadvantages |
|-------------------------------------|---|---|
| Shewhart's Sign CC (SSCC) | <ul style="list-style-type: none"> Easiest NPCC Easy calculations | <ul style="list-style-type: none"> Not too strong nonparametric statistics |
| NP – EWMA (NPEWMA) | <ul style="list-style-type: none"> NPCC with memory | <ul style="list-style-type: none"> More complicated calculations of control limits |
| NP – CUSUM (NPCUSUM) | <ul style="list-style-type: none"> Quickly reveals small and middle changes in the process NPCC with memory | <ul style="list-style-type: none"> More complicated calculations of needed statistics |
| Progressive Mean (PM) | <ul style="list-style-type: none"> Statistics of the progressive mean is considered more powerful than standard Shewhart CC or CUSUM and EWMA More effective in detecting small shifts in the process NPCC with memory | <ul style="list-style-type: none"> More complicated calculations |
| Median Absolute Deviation (MAD) | <ul style="list-style-type: none"> It is the most used nonparametric control chart when normality assumption fails Works with robust statistics via median | <ul style="list-style-type: none"> More complicated calculations This is not a CC with memory |
| NPCC based on Mood statistic (Mood) | <ul style="list-style-type: none"> Mood's statistics is considered one of the strongest nonparametric statistics | <ul style="list-style-type: none"> More complicated calculations This is not a CC with memory |

The paper is structured as follows: in Chapter 2 the designs of SW support for nonparametric control charts in MS excel is described. Chapter 3 is devoted to the design of simulations for evaluation of the performance of the selected nonparametric control charts (Table 1) and their evaluation for the in-control process. Chapter 4 is dedicated to the same issues but for the out-of-control process. In Chapter 5 the design of the methodology for the nonparametric control charts practical application is described.

2- Design of SW Support for Nonparametric Control Charts in Ms Excel

Before starting the simulations, the SW support for the construction of nonparametric control charts in MS Excel was created. This SW support was created not only for the construction of selected nonparametric control charts but also for the realization of simulations in order to evaluate their performance.

The logic of the created SW support is illustrated and explained on the example of SSCC and data file of the size of 20 subgroups each of size of 5 units. Other nonparametric control charts were processed on the same basis. Figure 1 shows a general view of an MS Excel sheet designed to construct SSCC. The table on the left (part A of the sheet) is used for the input values. On the right of Table A (part B of the worksheet) is basic information such as the number of subgroups and their size, mean value, standard deviation and other characteristics needed for further calculations. The largest table in the middle (part C of the sheet) is used to calculate the characteristics needed to construct the CC. It also includes the calculation of the RL (Run Length) value (part D of the sheet). In addition to this table, the calculated

performance indicator values (part E of the sheet). Under the table, the resulting CC (part F of the sheet) is constructed. All calculations and graphs are created automatically after entering the input values.

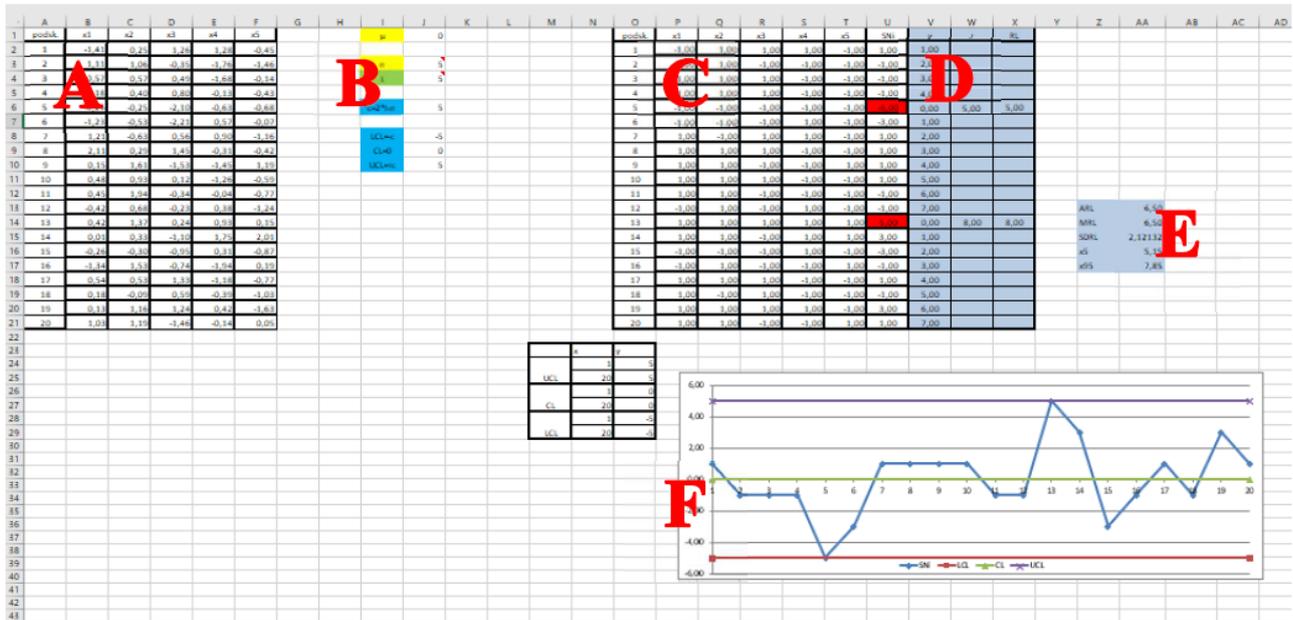


Figure 1. Created SW support.

3- Simulation Design for In-control Process

In-control process means the statistically stable process when only random causes have an influence on it. The following Table 2 describes the types of probability distributions that were used to simulate various breaches of assumptions about data.

Table 2. Overview of types of probability distributions used and their relation to various simulated breaches of assumptions about data.

| Distribution no. | Distribution/Type of breach of assumption | Distribution parameters | Unfulfilled assumption of data |
|------------------|---|---------------------------------|--------------------------------|
| 1. | Normal distribution | $N(0,1)$ | Neither |
| 2. | Student's distribution/ greater kurtosis than the normal distribution | t_3 | Data normality |
| 3. | Uniform distribution/ less kurtosis than the normal distribution | $R(-\sqrt{3}; \sqrt{3})$ | Data normality |
| 4. | Pearson distribution/ skewed distribution | χ_3^2 | Data normality |
| 5. | Mixed distribution | 50% $N(0,1)$ + 50% $N(2,1)$ | Constant mean |
| 6. | Mixed distribution | 50% $N(0,1)$ + 50% $N(0,4)$ | Constant variance |
| 7. | Time series / Autocorrelated data | $AR(1): x_i - 0.5x_{i-1} + a_i$ | Independence of data |

The number of the subgroups m was equal to 20, 100 and 300, and the size of the subgroup n was chosen to be 5 and 10. 10,000 replicates were performed. The m and n values were chosen based on the results of the in-control simulations [12] so that the ARL was approximately 370. [13, 14] The indicators Average Run Length ARL(0), Median Run Length MRL(0) and five per cent quantile (x_5) will be used to evaluate the performance of individual NPCC. ARL(0), MRL(0) and x_5 are based on the Run Length for in-control process and is referred to as RL(0). RL(0) is determined as the number of points recorded in CC that lie within the control limits, between 2 points outside these limits. This methodology of simulations is similar for all nonparametric control charts included in this study [2, 15, 16].

As an example of this part of the simulation, let us take a sample of 100 values generated from the normal distribution. We will construct NPCC SSCC working with the statistic S_{n_i} [1]. This sample is divided into 20 subgroups of size 5. From these values, the UCL and LCL control limits are calculated. These limits are recorded in the CC. Subsequently, n values from the normal distribution for particular subgroups are generated, and the S_{n_i} characteristic is calculated from them and recorded in the CC. Subsequently, RL(0) values are calculated. From them, values of performance indicators ARL(0), MRL(0) and x_5 are then calculated. This step is performed ten thousand times.

3-1- Results of Simulations of the In-control Process

In order to create a methodology for the practical application of nonparametric control charts (NPCC), it was necessary to carry out a number of analyzes, that will be presented below on some selected graphs and tables. All results can be found in Smajdorová (2019) study [16].

Selection of Performance Indicators

First, attention should be paid to the individual performance indicators and the selection of the ones that are most stable for different probability distributions (see Table 2) and for that reason the most suitable for the evaluation of the performance of various NPCC (see Table 1). An example of the resulting graph for SSCC with the number of subgroups equal to 20 and subgroup size equal to 5 is shown in Figure 2.

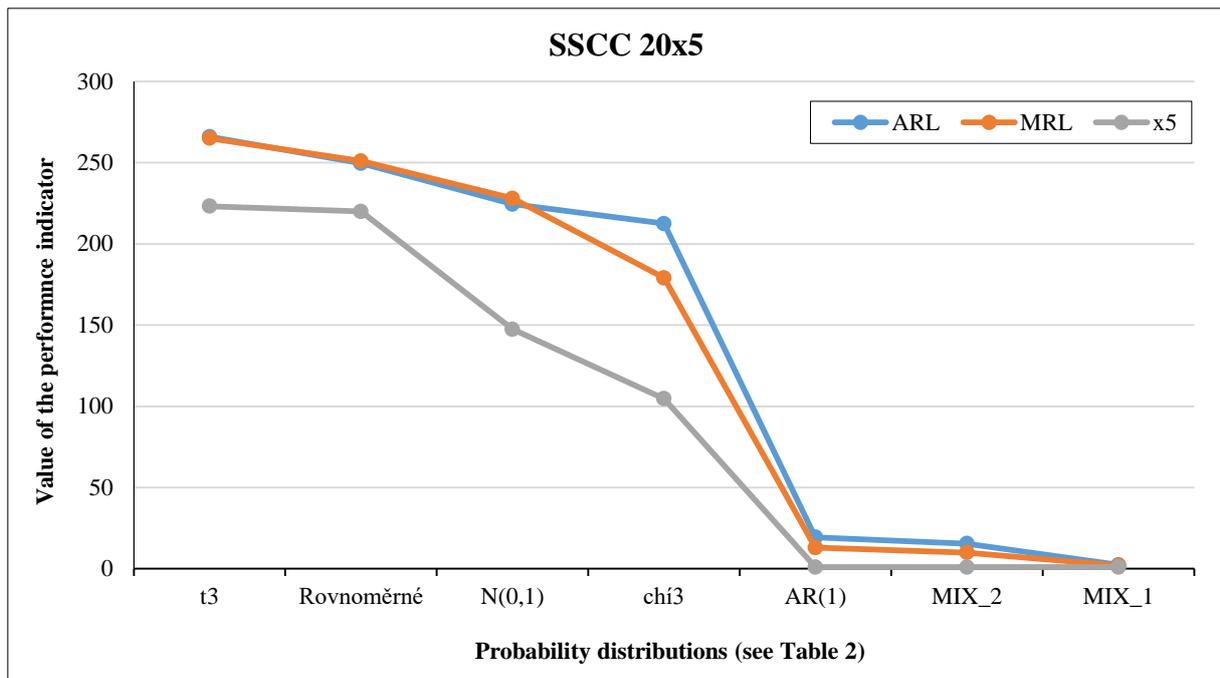


Figure 2. Values of performance indicators for different probability distributions (see Table 2) for SSCC with the $m = 20$ and $n = 5$.

Similar graphs were constructed and analyzed for all the rest of the analyzed NPCC (see Table 1). In order to analyze the quality of process performance indicators, their values were sorted in descending order. The resulting curves show the stability of individual indicators for different distributions, i.e. for different deviations from data assumptions. The more slowly the curve decreases, the more stable the indicator of performance is. Conversely, if the curve decreases too rapidly, the performance indicator is less stable. Based on this comparison of the quality of individual performance indicators done for various NPCC (see Table 1) in combination with different subgroup sizes and various probability distributions (see Table 2), we came to the general conclusion that 5% - quantiles have the most stable values for different deviations from the data assumptions. As regards the ARL(0) and MRL(0) performance indicators, the MRL(0) indicator is more stable than the ARL(0). Therefore, further performance analysis of individual NPCCs was performed using only the x_5 and MRL(0) performance indicators [16].

Performance Analysis of Nonparametric Control Charts

The analysis of the performance of investigated NPCC in combination with a various number of subgroups and different subgroup sizes was realized using graphs of MRL(0) or x_5 for different applied probability distributions (see Table 2). Figure 3 is an example of such a graph where MRL(0) values in a case of the normal distribution are depicted. Such graphs of MRL(0) for the rest of applied distributions and the same graphs of x_5 for all applied distributions can be found in Smajdorová (2019) study [16].

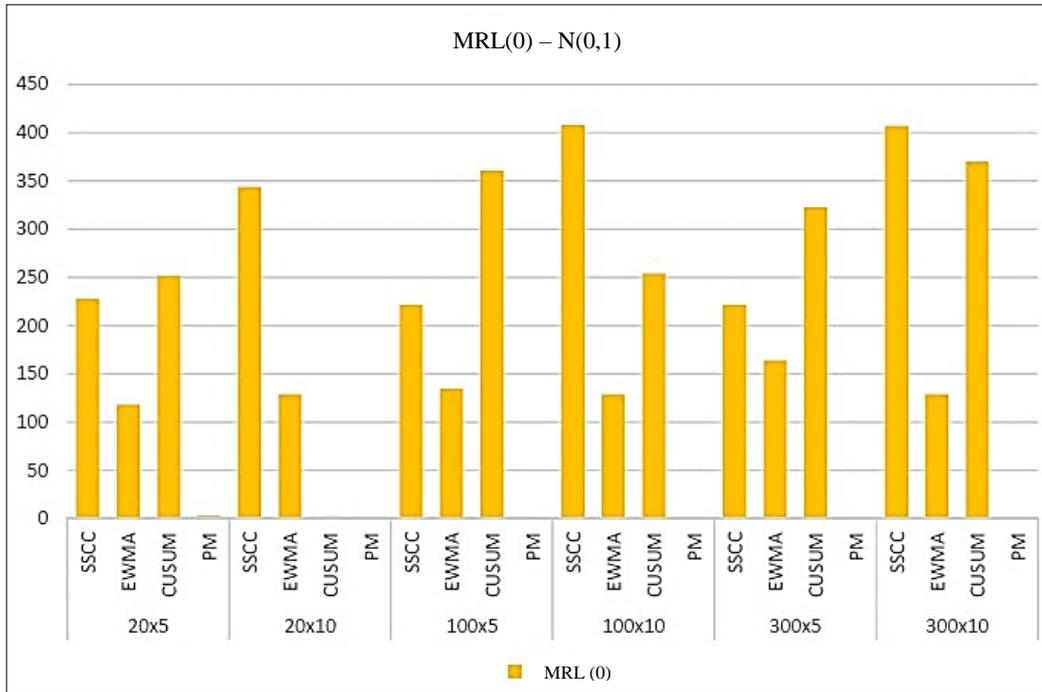


Figure 3. MRL(0) values for different types of NPCC in the case of normal distribution.

From the results of performance analysis of individual control charts based on the indicators MRL (0) and x_5 , we can say NP-CUSUM is suitable for the case where the assumption of a constant mean and variance is violated, also for data with greater sharpness than the normal probability distribution ($n = 10$) and for skewed distributions. The nonparametric SSCC control chart is the most suitable for autocorrelated data. It is also suitable for the case where the assumption of a constant mean value is not met and for skewed distributions ($n = 10$).

Based on the previous analysis, we can determine which nonparametric control chart is most suitable for the given data assumption violation. The results for NP – CUSUM and SSCC are shown in the following Table 3.

Table 3. Performance analysis results.

| NP - CUSUM | SSCC |
|---|---|
| Nonconstant mean value | Autocorrelation data |
| Nonconstant variance | Nonconstant mean value |
| Greater kurtosis than the normal distribution | Skewed distribution |
| Skewed distribution | Greater kurtosis than the normal distribution |

These CC were determined as quite robust (universal) CC, and from them, the most robust CC (i.e. the one that has the most stable performance indicator values over all the probability distributions) was set. Our analysis showed that the universal CC is control chart SSCC, where the performance indicator values are most stable for different probability distributions, i.e. for different deviations from data assumptions [16].

Performance Analysis of Nonparametric Control Charts for Variability Monitoring

Nonparametric control charts for monitoring variability (Mood and MAD) have now been analyzed. The values of the performance indicators for the nonparametric control charts Mood and MAD, which assess the stability of the process from the point of view of variability, are very similar for many distributions. Nevertheless, it can be said that a robust control chart MAD is better for normal and uniform distribution and distribution t_3 and Mood is better suited for distribution MIX_1, MIX_2, AR (1) and χ^2_3 . Based on the results of performance indicators, it was determined which of them is more suitable for the given data assumption violation. The results of the analysis of nonparametric control charts for variability monitoring are given in Table 4 [16].

Table 4. Results of analysis of nonparametric control charts for variability monitoring.

| Mood | MAD |
|------------------------|---|
| Nonconstant mean value | |
| Nonconstant variance | Less kurtosis than the normal distribution |
| Autocorrelation data | Greater kurtosis than the normal distribution |
| Skewed distribution | |

4- Simulation Design for the Out-of-control Process

The out-of-control process means that the process is statistically unstable, i.e. it is influenced by random and also assignable causes of variation. The number of the subgroups m was equal to 20, 100 and 300, and the size of the subgroup n was chosen to be 5 and 10. A total of 10,000 repeats were performed for each combination of m and n and for a particular type of distribution.

The simulations were realized in the same way as experiments for the in-control process (see chapter 3). However, unlike the experiment for the in-control situation during the second experiment, deviations of different sizes δ were inserted into the data files, and thus the out-of-control process was simulated. First, isolated deviations of 1.5σ ; 2σ and 3σ were inserted. The isolated deviation was inserted into 30th, 5030th and 9930th repeating.

Subsequently, a simulated persistent deviation of 1.5σ ; 2σ and 3σ was also performed, and only for 20×5 combination and selected types of distribution. The deviation occurred before it was signalled by a point outside the limit. Subsequently, "the process was intervened", and next subgroup was devoid of deviation. Then the deviation reappeared.

Only the $MRL(\delta)$ and x_{95} indicators were used to assess NPCC performance. For an out-of-control process, minimum values of performance indicators are required. The calculation of these indicators is based on $RL(\delta)$. This is determined as the number of points that are recorded in the control chart from the moment the change in the process occurred until the change was signalled in the form of a point out of control limit [16].

3-2- Summary of Simulation of the Out-of-control Process with Isolated Deviation

For a deviation of 1.5σ , it was the most powerful NP - EWMA for all types of deviations from the data assumptions. In addition to the nonparametric control chart, EWMA works well also a nonparametric SSCC control chart. This works well for a distribution that does not satisfy the assumption of a constant mean and the assumption of independence. And with increasing size n SSCC also work well for distributions where the assumption of constant variance is not met and for skewed distributions. NP-CUSUM also works well for some probability distributions. It can be assumed that this is a certain anomaly. Also, in the nonparametric PM control chart, the occurrence of a good result of the performance indicators is caused by the occurrence of one outlier immediately after the first occurrence of the deviation in the data. For a deviation of 2σ and 3σ , the simulation results are similar.

The results showed that for the in-control process, NPCCs perform very well, but for the out-of-control process with an isolated deviation, their performance is surprisingly poor, especially for small process changes. We can say that as the size of the subgroup grows, the performance indicators improve, as Das confirms [8]. Worse performance indicators results for the out-of-control process with isolated deviation may be due to the reason that nonparametric control chart perceives the isolated deviation as a random deviation against which they are robust [16].

3-3- Evaluation of Simulations with Persistent Deviation

The analysis of the $MRL(\delta)$ and x_{95} performance indicators showed that the persistent deviation is detected more quickly by NPCC than the isolated deviation. As already mentioned, the isolated deviation can be perceived by NPCC as a random deviation. The persistent deviation is detected by NPCC very quickly. More subgroups were needed for the first discovery of the deviation, followed by "process intervention" and the next subgroup did not contain the deviation. Then the deviation reappeared. One, at most two, selections were needed to detect it. Further deviations were detected at the first subsequent selection. The results show that NP-EWMA is the most powerful, as with the isolated deviation. Other NPCCs that have good x_{95} results are NP-CUSUM and SSCC [16].

5- Design of the Methodology of the Nonparametric Control Charts Application in Practice

The aim of this part is to describe the methodology of application of nonparametric control charts in practice (see Figure 4). The methodology described in this chapter is based on a review of available literature on the issue and on the results of simulations.

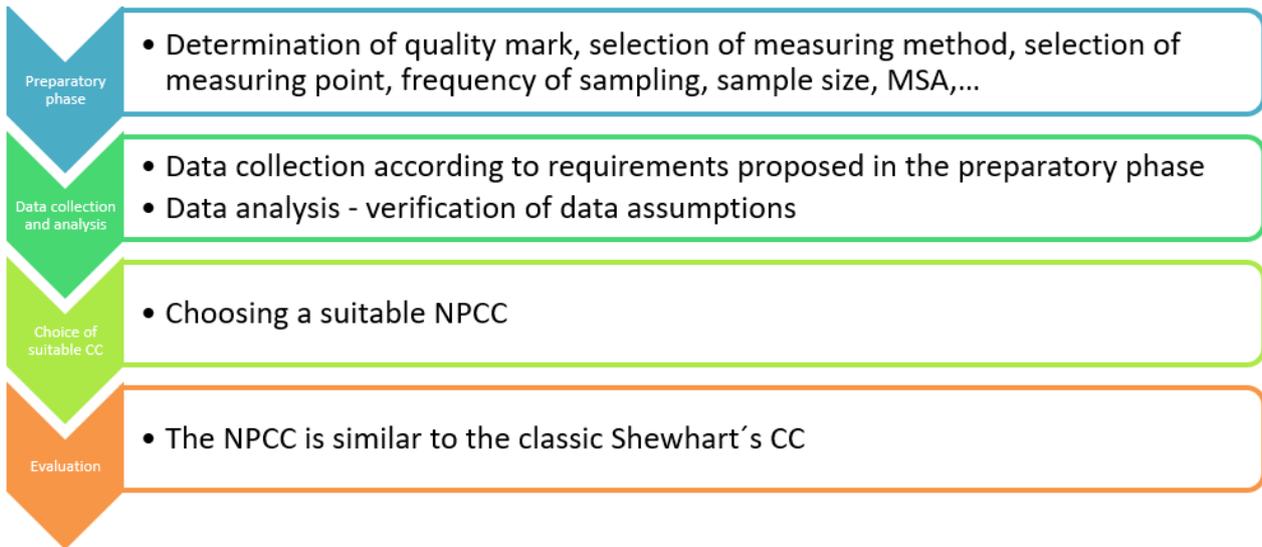


Figure 4. Flowchart for the application of nonparametric control charts into practice.

Further, the individual steps of the methodology will be described:

1. Preparatory phase

This phase of SPC includes determination of the quality characteristic or process parameter, which we want to regulate, selection of the suitable control point, choice of the method of data collection and recording (sample size and the frequency of sampling.).

2. Data collection and analysis

This step is an important part of the SPC because here we verify whether the basic data assumptions are met and therefore whether the standard Shewhart CC can be used, or whether the data assumptions are not met, and it will be preferable to use one of the nonparametric control charts. Verification of data assumptions can be performed using various statistical tests or graphical methods. It is always advisable to combine multiple tests with a graphical method.

3. Selection of suitable control chart

If all data assumptions are met, the standard Shewhart CC can be applied. The method of selecting a suitable Shewhart CC, its construction and evaluation are described in detail in publications about SPC. The problem occurs when one of the data assumptions is not met. In that case, the use of the standard Shewhart CC could lead to misleading results and, at worst, a deterioration of the process. Therefore, the application of nonparametric control charts is a possible alternative for these cases.

4. Evaluation of statistical process control

As with the standard Shewhart control charts, for the nonparametric ones is valid that if all points are within the control limits, then the process can be considered as in-control. However, if any of the points in the CC lies either below the lower control limit or above the upper control limit, it is a sign that the process is out-of-control. That is, the process is affected by an assignable cause of variability that needs to be identified, found its root cause, then removed, and the statistical stability assessment performed again. Flowchart for the application of nonparametric control charts into practice is shown in Figure 5, and Flowchart for the application of nonparametric control charts for variability monitoring into practice in Figure 6 [16].

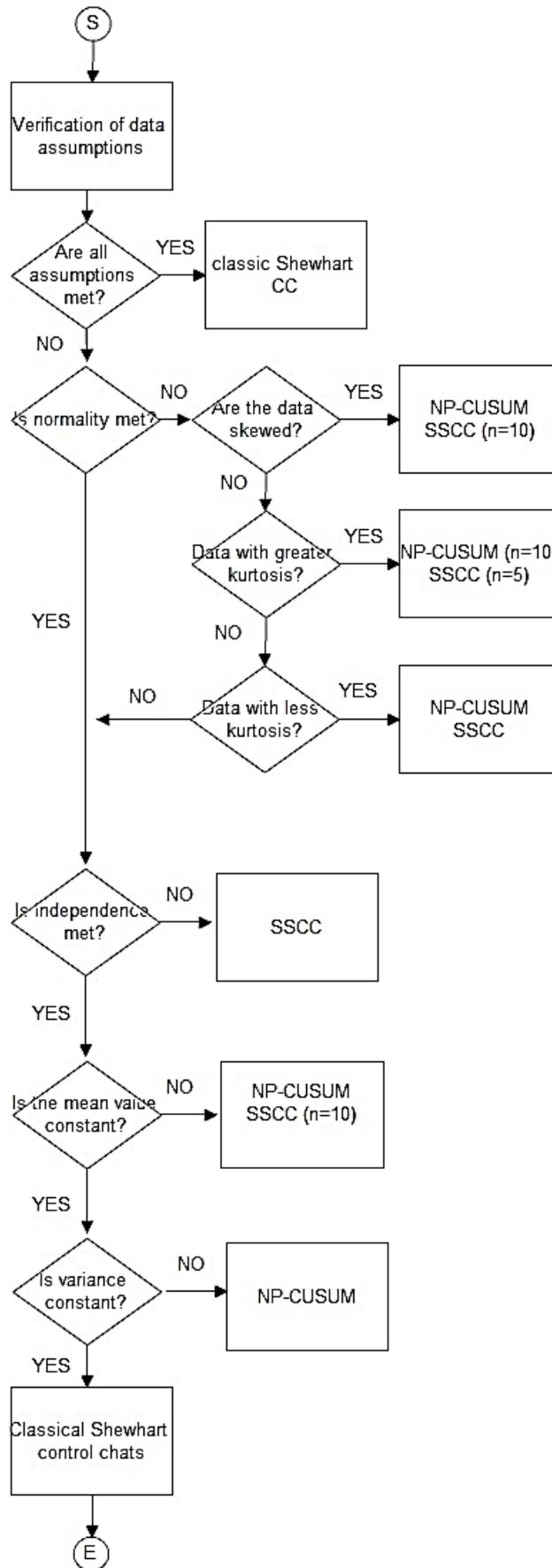


Figure 5. Flowchart for the application of nonparametric control charts into practice.

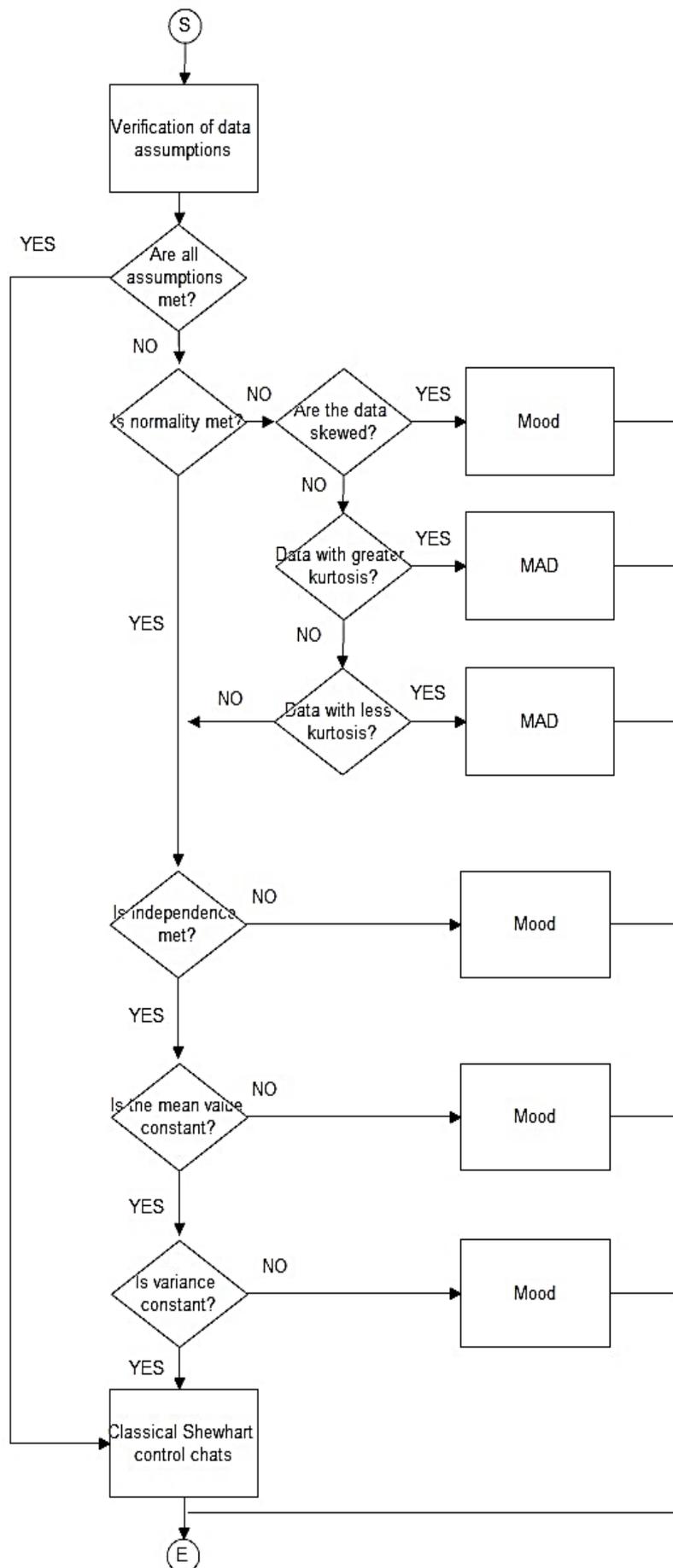


Figure 6. Flowchart for the application of nonparametric control charts for variability monitoring into practice.

The proposed methodology was verified on real data from an organization operating in the automotive industry. The application of this methodology led to the improvement of the verified process and the entire system of statistical process control in the organization [16].

6- Conclusion

This article was prepared mainly with the aim of bringing nonparametric methods of statistical process control into practice. It provides opportunities to deal with non-compliance with data assumptions that are common in practice. The use of these methods in practice can mean the correct interpretation of production processes in various non-standard situations. This issue also contributes to the development of the field of quality management. Simulations were performed on randomly generated data in MS Excel from different probability distributions, which represent different ways of violating data assumptions. Based on the results of the simulation, it can be argued that different nonparametric control charts are differently effective for different types of data assumption violations. The results show that the most appropriate nonparametric control charts for most data assumptions violations is NP-CUSUM as well as SSCC. The creation of SW support for nonparametric control charts in MS Excel was also part of these simulations. The output of this simulation study is a methodology for the application of nonparametric control charts in practice. The proposed methodology was also verified on real data. The issue of nonparametric SPC methods is well described in professional journals, but the transfer of these methods to practice is stuck. The research area of nonparametric SPC methods is extensive and offers many opportunities for further research, including the improvement of SW support. Nonparametric SPC methods could become a permanent part of teaching statistical methods at universities as well as expert training. Nonparametric methods also have a significant potential for the process monitoring in condition of Industry 4.0.

7- Funding

This work was supported by the European Regional Development Fund in a Research Platform focused on Industry 4.0 and Robotics in the Ostrava Agglomeration project, CZ.02.1.01/0.0/0.0/17_049/0008425 within the Operational Programme Research, Development and Education. Development and Education and the specific university research of the Ministry of Education, Youth and Sports of the Czech Republic No. SP2020/109.

8- Conflict of Interest

The author declares that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

9- References

- [1] Bakir, Saad, Tammy Prater, and Sara Kiser. "A Simple Nonparametric Quality Control Chart for Monitoring Students' GPAs." *SOP Transactions on Statistics and Analysis* 2015, no. 1 (January 31, 2015): 8–16. doi:10.15764/stsa.2015.01002.
- [2] Jarošová, and Noskiewiczová. "Pokročilejší metody statistické regulace procesu" (2015). ISBN 978-80-247-5355-3.
- [3] Chakraborti, Subha. "Nonparametric (Distribution-Free) Quality Control Charts." *Encyclopedia of Statistical Sciences* (August 15, 2011). doi:10.1002/0471667196.ess7150.
- [4] Chakraborti, S., and M. A. Graham. "Nonparametric (distribution-Free) Control Charts: An Updated Overview and Some Results." *Quality Engineering* 31, no. 4 (May 3, 2019): 523–544. doi:10.1080/08982112.2018.1549330.
- [5] Lu, Shin-Li. "An Extended Nonparametric Exponentially Weighted Moving Average Sign Control Chart." *Quality and Reliability Engineering International* 31, no. 1 (June 25, 2014): 3–13. doi:10.1002/qre.1673.
- [6] Wang, Dabuxilatu, Liang Zhang, and Qiang Xiong. "A Non Parametric CUSUM Control Chart Based on the Mann–Whitney Statistic." *Communications in Statistics - Theory and Methods* 46, no. 10 (May 4, 2016): 4713–4725. doi:10.1080/03610926.2015.1073314.
- [7] Abbasi, Saddam Akber, Arden Miller, and Muhammad Riaz. "Nonparametric Progressive Mean Control Chart for Monitoring Process Target." *Quality and Reliability Engineering International* 29, no. 7 (August 29, 2012): 1069–1080. doi:10.1002/qre.1458.
- [8] Das, Nandini. "A Note on the Efficiency of Nonparametric Control Chart for Monitoring Process Variability." *Economic Quality Control* 23, no. 1 (January 2008): 85–93. doi:10.1515/eqc.2008.85.
- [9] Murakami, Hidetoshi, and Takashi Matsuki. "A Nonparametric Control Chart Based on the Mood Statistic for Dispersion." *The International Journal of Advanced Manufacturing Technology* 49, no. 5–8 (November 21, 2009): 757–763. doi:10.1007/s00170-009-2439-3.

- [10] Claudia. "Derivation of the Limits for Control Chart Using the Median Absolute Deviation for Monitoring Non-Normal Process." *Journal of Mathematics and Statistics* 8, no. 1 (January 1, 2012): 37–41. doi:10.3844/jmssp.2012.37.41.
- [11] "Construction of New Coefficient of Variation Control Chart Based on Median Absolute Deviation." *Qalaai Zanist Scientific Journal* 2, no. 6 (October 20, 2017). doi:10.25212/lfu.qzj.2.6.14.
- [12] Chakraborti, S. "Run Length Distribution and Percentiles: The Shewhart Chart with Unknown Parameters." *Quality Engineering* 19, no. 2 (April 4, 2007): 119–127. doi:10.1080/08982110701276653.
- [13] Dhar, P.L. "System Simulation." *Thermal System Design and Simulation* (2017): 349–383. doi:10.1016/b978-0-12-809449-5.00006-1.
- [14] Shannon, Robert, and James D. Johannes. "Systems Simulation: The Art and Science." *IEEE Transactions on Systems, Man, and Cybernetics SMC-6*, no. 10 (October 1976): 723–724. doi:10.1109/tsmc.1976.4309432.
- [15] Low, Chun Kit, Michael B. C. Khoo, Wei Lin Teoh, and Zhang Wu. "The Revisedm-of-kRuns Rule Based on Median Run Length." *Communications in Statistics - Simulation and Computation* 41, no. 8 (September 2012): 1463–1477. doi:10.1080/03610918.2011.606950.
- [16] Smajdorová, Tereza. "Využití neparametrických metod ve statistické regulaci procesů" Ostrava: Dissertation thesis. VSB – Technical University of Ostrava, Faculty of Materials Science and Technology. (2019).