

# Clustering of similar processes for the application of statistical process control in small batch and job production

Michael Wiederhold<sup>1,\*</sup>, Jonathan Greipel<sup>1</sup>, Renato Ottone<sup>2</sup>, and Robert Schmitt<sup>1</sup>

<sup>1</sup> Chair of Metrology and Quality Management, WZL RWTH Aachen University, Germany

<sup>2</sup> ISONORM di Renato Ottone, Grana, Italy

Received: 16 June 2016 / Accepted: 13 September 2016

**Abstract.** A structured procedure is presented to identify similar processes producing similar features on different products in small batch and job production processes. Similar processes are clustered together to achieve the minimum necessary amount of samples to monitor and control the respective production process using statistical process control (SPC). It is described how the procedure was successfully applied to a process in the manufacturing of boring/milling machines.

**Keywords:** SPC / process control / small batches / reward / quality management

## Implications and influences

The structured procedure described here is a first step towards the application of SPC using a process cluster approach. Future work will provide an algorithm that allows to automatically identify homogeneous clusters based on historical process data.

## 1 Introduction

Statistical process control (SPC) is a set of tools that is widely used in production industry. The main elements of SPC are control charts for monitoring the stability of processes and process capability evaluations. A process is considered stable when it is only subject to *common-cause variation* inherent to the process, rather than *special cause variation* like tool wear or breakage, causing a shift in process behaviour.

The application of control charts is currently limited to series manufacturing because a minimum of at least 50 samples is required to estimate the parameters of the population and to calculate meaningful control limits and capability indices, also known as Phase I analysis [1]. These amounts are usually not available in small batch and job production (SBJP) environments such as the aerospace, turbine or machine tool industries, which are characterized by a small number of complex, expensive parts that are often customized for each order. Consequently, SPC is usually deemed not to be applicable in such scenarios.

On an international level, there is currently no standard or guideline available describing the application of SPC for SBJP. The only normative document is the national British standard BS 5702-3 which describes the application

of control charts for small- and mixed batch production processes [2]. In the introduction to the standard it is argued that SPC is currently mostly employed as *Statistical Product Control*, putting emphasis on monitoring a specific feature on a specific part under known and stable conditions, as opposed to *Statistical Process Control*, where the emphasis is on monitoring the production of similar features on different parts and under varying conditions. The standard then describes methods to monitor several features in a joint control chart to increase the sample size.

While the authors of this article deem the intention of the standard to be reasonable, there is no indication as to which features should be clustered. However, the clustering decision plays a major role for the ability to extract meaningful information from the application of SPC. As will be shown in the following, there is no appropriate, systematic procedure for selecting clusters of similar features. Based on these findings, such a procedure is proposed and exemplarily applied in a real SBJP process.

## 2 Current approaches for SPC in small batch and job production

There are two different ways to approach the application of SPC to SBJP. The first addresses the issue of estimating the population parameters. There are several adaptations to control charts requiring no estimation or less estimation of the population parameters.

Pillet [3] introduces an adapted control chart for normal distributed processes of known standard deviation  $\sigma$  and unknown mean  $\mu$ . The chart uses adaptive control limits that are wider for the initial samples and become more narrow as more observations are recorded. The drawback of this method is that the unknown property  $\mu$  is substituted by the target value of the process, which goes

\* Corresponding author: [m.wiederhold@wzl.rwth-aachen.de](mailto:m.wiederhold@wzl.rwth-aachen.de)

against the strong recommendation by Deming [4] not to use the specification values to calculate control limits. With Q-charts, as developed by Quesenberry [5], data from processes with arbitrary distribution forms can be transformed into sequences of independently and identically distributed  $N(0, 1)$  observations (the Q Statistics). Transformations are provided for cases where either  $\mu$ ,  $\sigma$ , both or none of both are known. The Q Statistics are plotted into Shewhart-like control charts. Before, Hawkins [6] had introduced a similar approach based on a *self-starting* cumulative sum (CUSUM) chart for the mean (with known  $\sigma$ ), limited however to normal distributed processes. According to Del Castillo et al. [7], the shift detection properties of both are generally poor. Del Castillo and Montgomery [8] and Wassermann [9] propose *dynamic* exponentially weighted moving average (EWMA) charts with better detection capabilities and without the normality requirement that use a Kalman filter to estimate the process parameters and adjust control limits dynamically with each new observation. Capizzi and Masarotto [10] describe an adaptive cumulative score (CUSCORE) chart, an enhanced self-starting CUSUM control chart that has shown to outperform the previously mentioned one. For multivariate short run situations, which are not a focus of this work, Khoo et al. [11] provide an overview of available charting techniques for monitoring the mean and introduce charting techniques for monitoring dispersion. An adjusting rule for processes that make use of inertial tolerancing [12] is provided by Pillet and Pairel [13]. Finally, an alternative to SPC monitoring of short runs is introduced by Cox et al. [14], based on a discrete-event simulation model. Rather than monitoring a process for stability, its purpose is to adjust the sampling rate based on the current performance of the process in relation to a desired process capability.

These control charts satisfy a need for process control in the early stages of a process with no previous data available, e.g. when producing in short runs. All these adaptations however only mask the fact that decisions are taken based on small amounts of data. The second approach is therefore based on an increase of the samples considered by identifying similar processes and clustering the data obtained from them. If the clustering results in more than 50 samples, regular SPC techniques can be applied, including proper estimation of the population's parameters.

When clustering similar processes, special care needs to be taken so as to not cluster processes that exhibit a significantly different statistical behaviour. In such a case there would be an increased number of false alarms and not identified out-of-control situations because estimates of the population parameters are bound to deliver incorrect results, leading to useless control limits that can have a damaging effect on the process because it is controlled under false premises. There are several approaches for a systematic clustering [15–17] that are all based on the Group Technology (GT) concept [18]. Here, features are classified by a hierarchical code that describes the characteristics in ever-finer detail using specific digits. The above-mentioned approaches provide rules to determine which digits need to be the same for a clustering of the corresponding processes to be allowed.

While the GT-based approaches help to identify clusters based on certain levels of similarity, they do not take into account the aspect of significant statistical differences and are therefore only partly suited for the application of SPC in SBJP. Concluding, there is currently a lack of a systematic procedure to the clustering of similar processes based on the identified or expected statistical behaviour. Such a procedure is presented in the next section.

### 3 Procedure for the application of SPC in small batch and job production

The aim of the procedure outlined in this chapter is to provide a systematic approach that guides the user through the setup, control and adaptation of an SPC system for SBJP. The procedure can be applied at any stage of knowledge about the process and be used to adapt the SPC system as new knowledge is obtained. Thus the procedure provides the user with the most appropriate configuration of clusters (from a statistical homogeneity standpoint) based on the current level of knowledge and indicates changes in cluster configuration as the knowledge increases.

The procedure is shown in Figure 1 and consists of three main steps:

- identification and documentation of processes and influencing characteristics (Identification);
- clustering of processes (Clustering);
- application and adaptation of SPC (Application).

The individual steps are explained in detail in the following sections.

#### 3.1 Identification and documentation of processes and influencing factors

The goal of the first step of the procedure is to gain a clear understanding of the processes that will be subject to clustering in the following steps. This first step is very important because it ensures that the pool of possible clusters is narrowed down by applying simple decision rules to reduce complexity for the subsequent analysis of similarity. Although this article is about the clustering of similar *processes*, the procedure approaches this on the level of similar *features* since the feature and the process that produce it belong inseparably together and the question in industrial practice is usually going to be: “Which cluster does this feature belong to?”

For example, it is intuitively obvious that features produced by entirely different process technologies should not be clustered. After all, two cylindrical objects can be of similar geometrical dimensions but one is turned and the other is molded. While it might even be possible to monitor both processes together in a control chart, there is little to no added value to process control in such a case.<sup>1</sup>

<sup>1</sup> BSI 5702-3 actually suggests this approach. Since no explanation towards the added value is given, other than to increase the number of samples within one chart, it is not considered further [2].

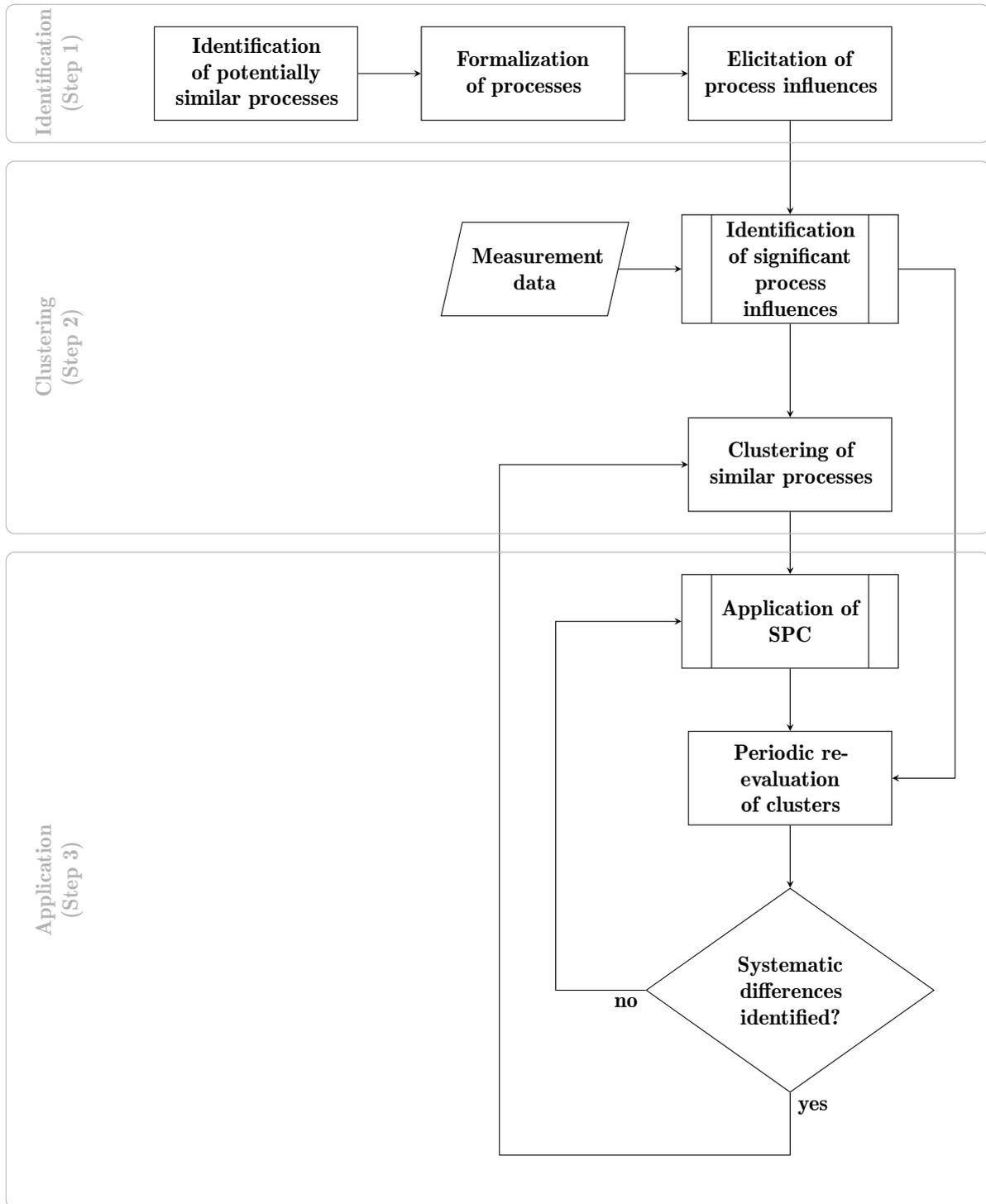


Fig. 1. Procedure for the application of SPC to SBJP.

The identification of potentially similar processes leads only to a very rough separation between features produced by different technologies. The example of two features, one turned and one molded, was however simplified in the sense that it took into account the general production technology employed, whereas in production the common case is that a feature is created using a series of process steps employing

different technologies. In such a case it is important to look at the series of process steps – the *process chain* – that have been performed prior to a measuring and the subsequent SPC analysis.

Figure 2 illustrates how a visualization of the process chain helps to identify potentially similar processes. It shows three simplified process chains that each consist of

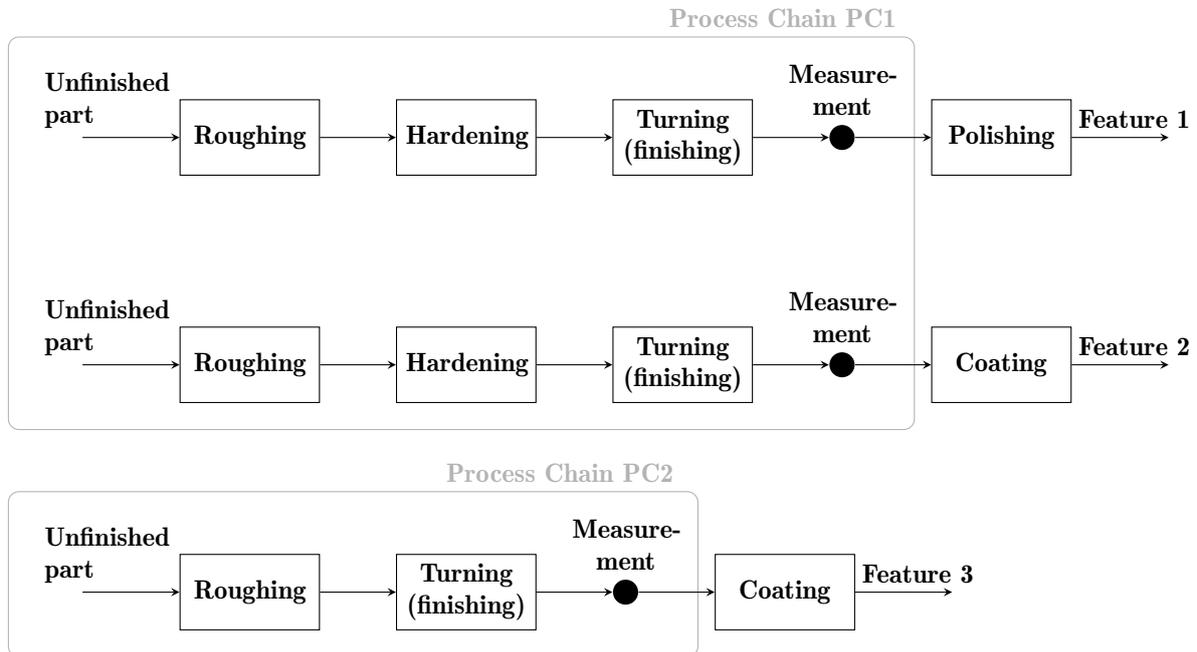


Fig. 2. Example of two process chains containing a series of equal process steps.

several process steps. The important criterion for potential similarity of processes is the similarity of the process steps performed prior to measuring the feature's quality characteristics for SPC analysis.

While processes may make use of the same technologies and be arranged in the same process chain, there can still be differences between them for the production of individual features. Examples are the use of different materials or different tools. These influences on the process along with their parameter space need to be recorded for each group of potentially similar processes. There are several methods available to identify influences on a process, e.g. the Ishikawa diagram [19] and the IDEF0 diagram [20]. The process influences can be extracted directly from the diagrams and become factors by which the features are categorized. For each feature and factor, the corresponding level (or range of levels) needs to be recorded.<sup>2</sup>

### 3.2 Identification of significant process influences and clustering of similar processes

When the same fundamental process steps are applied to different materials, tools, production machines, etc. it needs to be determined for every one of these factors whether it constitutes a systematic process influence. If yes, two processes with variation in this factor should not be clustered, if not, the factor can be regarded as irrelevant in the analysis.

For the intents and purposes of this analysis, a significant process difference is defined as such:

If the variation of a factor from one level to another results in a significant change of at least one key statistical measure of the process (i.e. a parameter of its distribution function), the factor constitutes a significant process influence.

As an example, if a change in material leads to a significant shift in the mean value of the measured quality characteristic, material can be considered to have a significant process influence.

The identification of significant process influences is the most crucial step of the procedure and the way to perform it and the quality of the outcome is highly dependent on the amount and quality of available data and the amount of effort a user is willing to expend. Different approaches can be applied for the identification of significant process influences:

- **Application of expert knowledge** comes down largely to practical experience with the process. Physical aspects of the process should be taken into account. If, for example, a metal cutting process is employed that produces a large heat transfer to the workpiece and the used materials have very different coefficients of thermal expansion, material presumably has a significant influence on the outcome;
- **Preliminary experiments** allow the user to confirm or reject hypotheses that he has about a process, particularly about changes of the statistical properties. However, it is in the nature of SBJP production that the parts are usually complex and their manufacturing is expensive. There is therefore no room for intuitive trial-and-error approaches and systematic experimentation with the variation of one factor at a time but rather experiments need to be designed intelligently with as little tests required as possible, e.g. using Design of Experiments (DOE) [21];

<sup>2</sup> The levels of a factor are the possible variations for that factor, e.g. the levels of the factor material can be cast iron (M1) and steel (M2).

- **Simulation** can be used to complement or even replace experiments, often at a lower cost when considering complex, expensive parts and multiple combinations of influences. A common measure to estimate influences and effects is the Monte-Carlo Simulation [22];
- **Statistical analysis of historical data about the processes** makes use of historical datasets. When processes have been running for a while, it is possible to extract from the data information about significant statistical differences. The difference to preliminary experiments is that the data is not obtained from a planned campaign but from regular production activity.

Different circumstances favour the application of different approaches for the identification of significant process influences. Preliminary experiments have a high informative value and are independent from existing data but they are usually expensive. Simulations are less expensive but are, in turn, dependent on existing data and have a reduced informative value compared to experiments. The application of expert knowledge is very cost-efficient and independent from existing data, but its informative value is relatively low. The analysis of historical data is very cost effective too, and has a high informative value. Its drawback is the dependence on existing data, which is however often available. Based on these considerations, it becomes clear that it is wise to use a combination of expert knowledge and historical data where possible.

### 3.3 Application and evaluation of SPC

With the clusters identified in Section 3.2 it is possible to apply SPC on the clustered features with the benefit of increased sample size within each cluster. Aspects of data preparation, application of control charts and process capability evaluation need to be taken into account.

Measurement systems are often equipped with functionality to record additional data, so-called metadata [23], along with the actual measurement value. This additional data often contains a timestamp of the measurement as well as information about the measuring system itself including the operator. For further analysis it is also sometimes possible to record an identifier (ID) of the measured feature. To be able to efficiently keep track of clusters in SBJP, to automatically assign features to clusters and to update the cluster configuration when and if new information requires it, it is vital to extend the recording of metadata. As a minimum requirement, the metadata should include for each measurement the level of the factors identified in the step *Elicitation of process influences*. There should e.g. be a record for each measurement value about the material, tool, machine, measurement process and nominal diameter. It is important to include all potential influences and not just the ones that are considered significant because over time, as more data becomes available, the identification of what is significant can change. In general, it is preferable to record more factors rather than less to be on the safe side while keeping the required effort manageable.

Once processes are clustered and measurement values are associated with metadata, control charts can be applied for the clusters in the conventional way, taking into

account some particularities: As long as only very few samples are available (e.g. less than 50), self-starting control charts can be applied to decrease the probability of  $\beta$ -errors, i.e. to decrease the chance of not detecting an out-of-control situation. Once enough data points are gathered, control limits can be calculated normally and the cluster can be monitored like any series production process. Ultimately, once the process cluster is deemed capable, the effort for performing measurements can be reduced by switching to a sampling inspection of the features.

The SBJP environment is one that is naturally characterized by a low amount of information in the beginning of the analysis which then continuously increases over time, thus increasing the level of knowledge both about the processes in themselves as well as about the correctness of the cluster identification. It is therefore advantageous to make use of the increased amount of knowledge by re-evaluating the clusters. There is no exact re-evaluation frequency to be followed, there is rather a trade-off between accurateness at any time on the one hand and effort for re-evaluation on the other hand. Additionally, frequent changes of the clusters lead to frequent changes of the control charts which makes it hard for operators to interpret them because they might lose the intuitive understanding of process behaviour that comes from the visual representation in the control chart.

## 4 Application of the procedure

The proposed procedure for the application of SPC to SBJP has been applied at ALESAMONTI, a medium-sized Italian company producing machine tools for boring and milling [24]. Before, SPC was not applied there as only a small two-digit number of machine tools is produced every year. In the following, it is shown how this limitation was overcome by the application of the procedure.

### Step 1: Identification

As the positioning system of all ALESAMONTI machine tools is based on linear axes (guides), the equidistance of opposing guiding surfaces along the axes is important for a precise positioning of the spindle in the working volume. The guide with its pairs of guiding surfaces as well as the procedure to measure equidistance are depicted in Figure 3.

Equidistance describes the compliance of a locally constant distance between two guiding surfaces. To measure equidistance, ten measurement points are spread equally along the whole part length. The first measurement is used as the zero reference. For the following nine measurements the deviation to the zero reference is measured. At the end, the maximum range of the measured points is recorded as the equidistance. All guides are produced by the same process chain. The possible process influences identified by application of expert knowledge are shown in Table 1.

### Step 2: Clustering

Expert knowledge was applied in the second step to judge the identified process influences for their significance, the result of which is shown in Table 2.

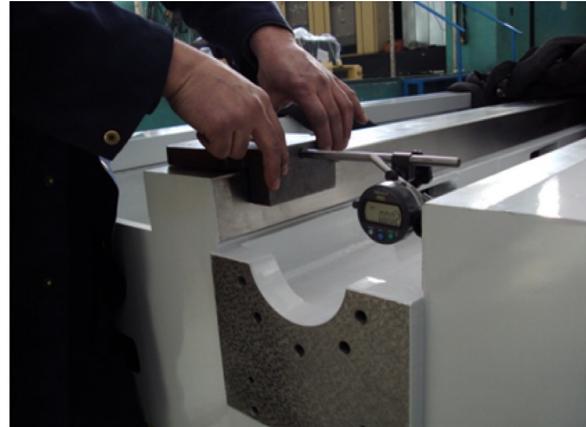
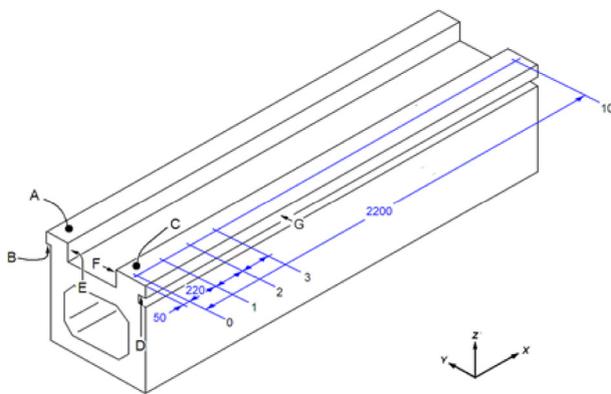


Fig. 3. Pairs of opposing guiding surfaces and depiction of the measurement process.

Table 1. Possible process influences/factors and their factor levels.

Process influence/factor	Factor level
Material	Cast iron; steel
Position of guides	B–A; D–C; F–E
Length of part	1500 mm up to 10 500 mm
Tolerances	15 $\mu\text{m}$ ; 20 $\mu\text{m}$
Measurement process	Analog dial gauge; digital dial gauge

**Step 3: Application**

Control charts were generated for a set of historical data from December 2009 to October 2013 according to the clusters in Table 2. This means that there is one control chart for tolerances of 15  $\mu\text{m}$  and one for tolerances of 20  $\mu\text{m}$ . Due to the small amounts of data, moving average control charts with a sample size of three were used.

The control charts for tolerances of 20  $\mu\text{m}$  are used as an example and depicted in Figure 4. They show a few alarms due to points above the control limits. By a closer inspection of the data, e.g. by ordering the data according

Table 2. Assessment of significance for factors.

Factor	Assessment	Special remarks
Material	Not significant	None
Type of guide pair	Not significant	None
Length of part	Not significant	None
Tolerances	Significant	Multi-level process with great worker influence
Measurement process	Redundant	Type of measuring device is directly related to tolerance class

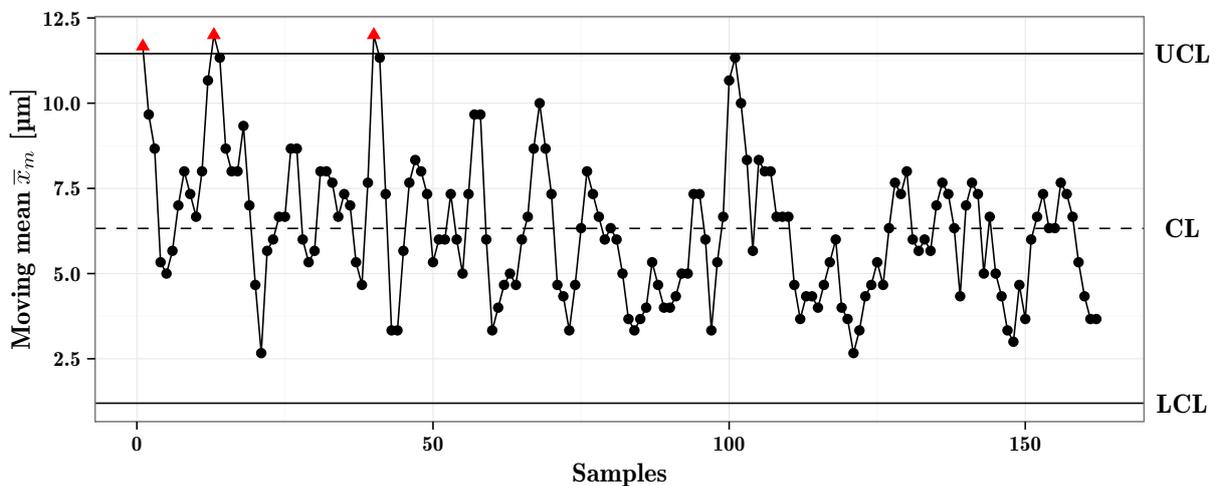


Fig. 4. Control chart for guiding pairs with a tolerance of 20  $\mu\text{m}$ .

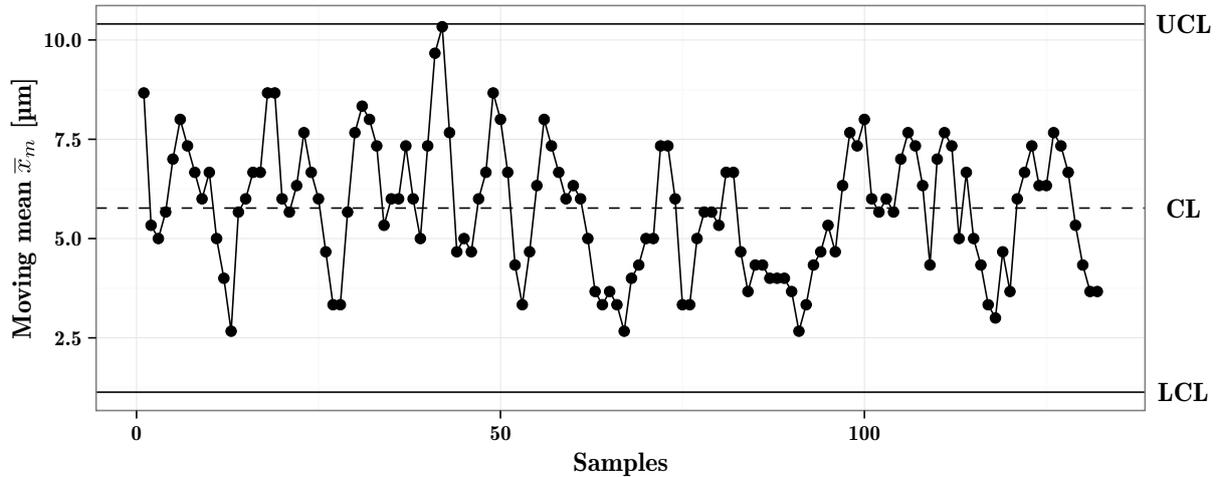


Fig. 5. Control chart for guides with a tolerance of  $20 \mu\text{m}$  and length of part under 4000 mm.

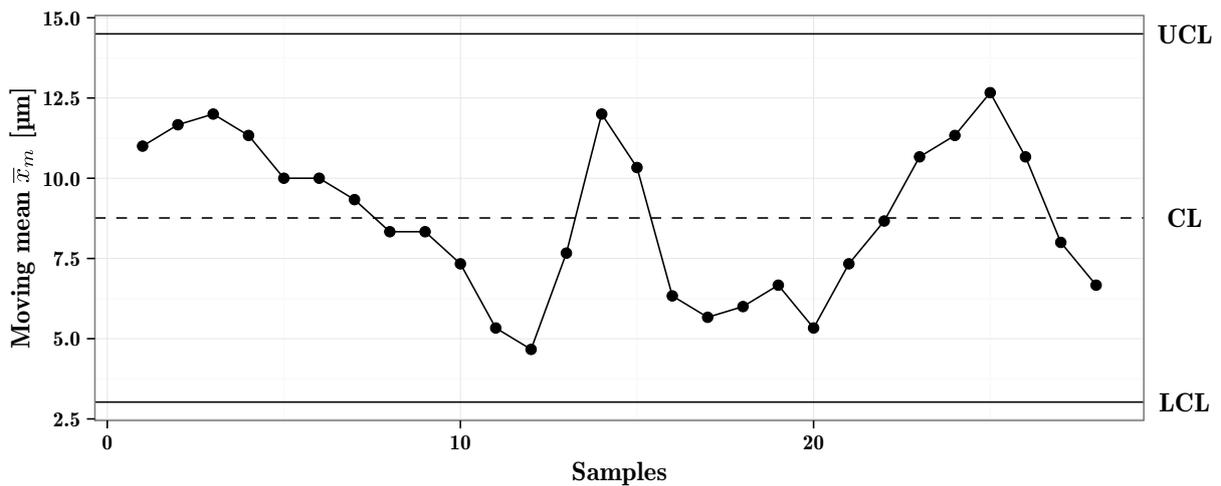


Fig. 6. Control chart for guides with a tolerance of  $20 \mu\text{m}$  and length of part over 4000 mm.

to factor levels within control charts it gets obvious that the alarms often occur when the moving average value is dominated by parts that have a length above 4000 mm. Therefore, a systematic influence is suspected.

#### 4.1 Reevaluation of the clustering

Due to the suspected systematic influence of the length of part, the  $20 \mu\text{m}$  tolerance cluster is subdivided into two subclusters. One contains lengths of part under 4000 mm and the other lengths of part over 4000 mm. These two clusters are tested using a hypothesis test on the mean value which proves a significant difference. For the newly found clusters, separate control charts are generated as shown in Figures 5 and 6. All alarms disappear from both charts. The reason is that each chart now uses control limits appropriate to the data in the subclusters, proving that both subclusters are likely to be stable.

## 5 Conclusions

This paper presented a structured approach to the clustering of processes for the application of SPC in SBJP environments. The application to the ALESAMONTI example showed that the application of the procedure allowed to assess the significance of process influences to cluster processes using expert knowledge, while keeping the flexibility to adapt this assessment and thus the clustering as new information comes in.

However, it also became clear that the sole application of expert knowledge did not reveal a significant factor that could have been identified by analyzing historical data. Therefore, it would be advantageous to have an algorithm that systematically extracts significant influences from the analysis of historic data to complement the expert knowledge. This way, a-priori assumptions made by the expert can be verified through hypothesis testing

immediately before alarms due to wrong clustering occur, as long as sufficient historical data is available. Such an algorithm is currently being developed by the authors.

The research leading to these results has received funding from the European Union, Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 244067.

## References

1. D.C. Montgomery, *Statistical quality control: A modern introduction: International student version* (Wiley and John Wiley, Hoboken, NJ/Chichester, 2008), sixth ed., ISBN 0470233974
2. BSI British Standards BSI 5702-3:2008, Guide to statistical process control (SPC) charts for variables – Part 3: Charting techniques for short runs and small mixed batches, 2008
3. M. Pillet, A specific SPC chart for small-batch control, *Qual. Eng.* **8**, 581–586 (1996)
4. W. Edwards Deming, *Out of the crisis* (Massachusetts Institute of Technology, Center for Advanced Engineering Study, Cambridge, MA, 1986), ISBN 9780911379013
5. C. Quesenberry, SPC Q charts for start-up processes and short or long runs, *J. Qual. Technol.* **23**, 213–224 (1991)
6. D.M. Hawkins, Self-starting CUSUM charts for location and scale, *Statistician* **36**, 299 (1987)
7. E. Del Castillo, J.M. Grayson, D.C. Montgomery, G.C. Runger, A review of statistical process control techniques for short run manufacturing systems, *Commun. Stat. Theor. Methods* **25**, 2723–2737 (1996)
8. E. Del Castillo, D.C. Montgomery, A Kalman filtering process control scheme with an application in semiconductor short run manufacturing, *Qual. Reliab. Eng. Int.* **11**, 101–105 (1995)
9. G.S. Wassermann, An adaptation of the EWMA chart for short run SPC, *Int. J. Prod. Res.* **33**, 2821–2833 (1995)
10. G. Capizzi, G. Masarotto, An enhanced control chart for start-up processes and short runs, *Qual. Technol. Quant. Manag.* **9**, 189–202 (2012)
11. M.B.C. Khoo, S.H. Quah, H.C. Low, C.K. Ch'ng, Short runs multivariate control chart for process dispersion, *Int. J. Rel. Qual. Saf. Eng.* **12**, 127–147 (2005)
12. M. Pillet, Inertial tolerancing, *TQM Mag.* **16**, 202–209 (2004)
13. M. Pillet, E. Pairel, Determination of an adjusting rule in the case of multi-criteria inertial piloting, *Int. J. Metrol. Qual. Eng.* **2**, 51–59 (2011)
14. S. Cox, J.A. Garside, A. Kotsialos, Discrete-event simulation of process control in low volume high value industries, in *Proceedings of the 11th International Conference on Manufacturing Research (ICMR2013)* (2013)
15. C.-s. Cheng, Group technology and expert systems concepts applied to statistical process control in small-batch manufacturing, Arizona State University, Tempe, AZ, USA, PhD thesis, 1989
16. M. Al-Salti, A. Statham, The application of group technology concept for implementing SPC in small batch manufacture, *Int. J. Qual. Reliab. Manag.* **11**, 64–76 (1994)
17. Y.D. Zhu, Y.S. Wong, K.S. Lee, Framework of a computer-aided short-run SPC planning system, *Int. J. Adv. Manuf. Technol.* **34**, 362–377 (2007)
18. S.P. Mitrofanov, *Scientific principles of group technology (Nauchnye osnovy gruppovoi tekhnologii)* (National Lending Library for Science and Technology, Boston Spa (Yorks.), 1966)
19. N.R. Tague, *The quality toolbox* (ASQ Quality Press, Milwaukee, WI, 2005), second ed., ISBN 9780873896399
20. National Institute of Standards and Technology, *Integration definition for function modeling (IDEFO)* (Computer Systems Laboratory, National Institute of Standards and Technology, Gaithersburg, MD/Springfield, VA, 1993), Vol. 183
21. D.C. Montgomery, *Design and analysis of experiments* (Wiley, Hoboken, NJ, 2013), eighth ed., ISBN 1118146921
22. C.Z. Mooney, *Monte Carlo simulation, volume no. 07-116 of Sage University Papers Series. Quantitative applications in the social sciences* (Sage Publications, Thousand Oaks, CA, 1997), ISBN 9780803959439
23. R. Guenther, J. Radebaugh, *Understanding metadata* (NISO Press, Bethesda, MD, 2004), ISBN 1-880124-62-9
24. M. Wiederhold, J. Greipel, R. Ottone, R. Schmitt, Gemeinsam sind sie stark: Statistical Process Control bei kleinen Stückzahlen, *Qualität und Zuverlässigkeit* **61**, 30–34 (2016)

**Cite this article as:** Michael Wiederhold, Jonathan Greipel, Renato Ottone, Robert Schmitt, Clustering of similar processes for the application of statistical process control in small batch and job production, *Int. J. Metrol. Qual. Eng.* **7**, 404 (2016)