DATALYZER International
Partners in Continuous Improvement

Reject reduction by analysis of recorded process data

Erwin Gijzen, Quality Target, egijzen@quality-target.com
Marc Schaeffers, DataLyzer International, mschaeffers@datalyzer.com

1 Summary

Reject reduction in production processes is commonly seen as a goal to achieve. As generally proposed, also by Six Sigma, one should find causes of quality variation, and address those to improve the process. This paper discusses an approach where the causes of quality variation are made visible with the help of process data analysis. In fact, the analysis links quality data to process data, to see what process variables are significantly affecting quality. The key aspect is that both the quality data and the process data need to be logged. Process variables to be logged can be of many types, like process settings (e.g. speed or temperature), raw material batches, production shifts or production lines. The analysis will make visible which process variables affect quality significantly. A nice advantage is that the analysis can predict how much reject reduction can be achieved if those process variables would be brought in control. This helps to allocate engineering resources to areas that can bring direct results and sufficient benefits. This article is not intended to cover all aspects of the data analysis in the example, but merely to show the generic approach and value.

2 Introduction

Reject levels can be made visible with a capability graph. Suppose that would give this result:

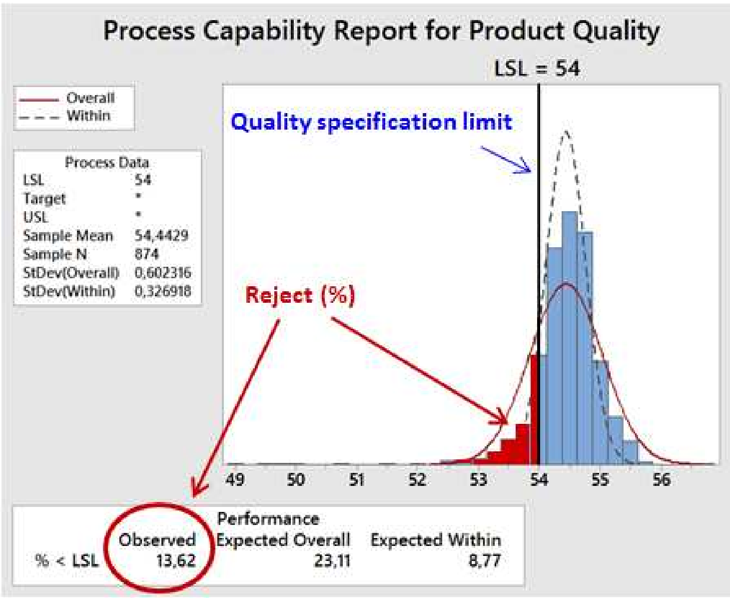

Figure 1: Capability chart of initial quality data

For this process, the observed reject level is about 13.6%. Let's assume we want to find out how the reject could be reduced. The obvious strategy would be to shift the mean of the data to the right, but in this process that is not possible. The other option is to reduce the variation, but how?

A logical next step in the path to reduce variation is to graph the same data in a control chart. The result is as follows:

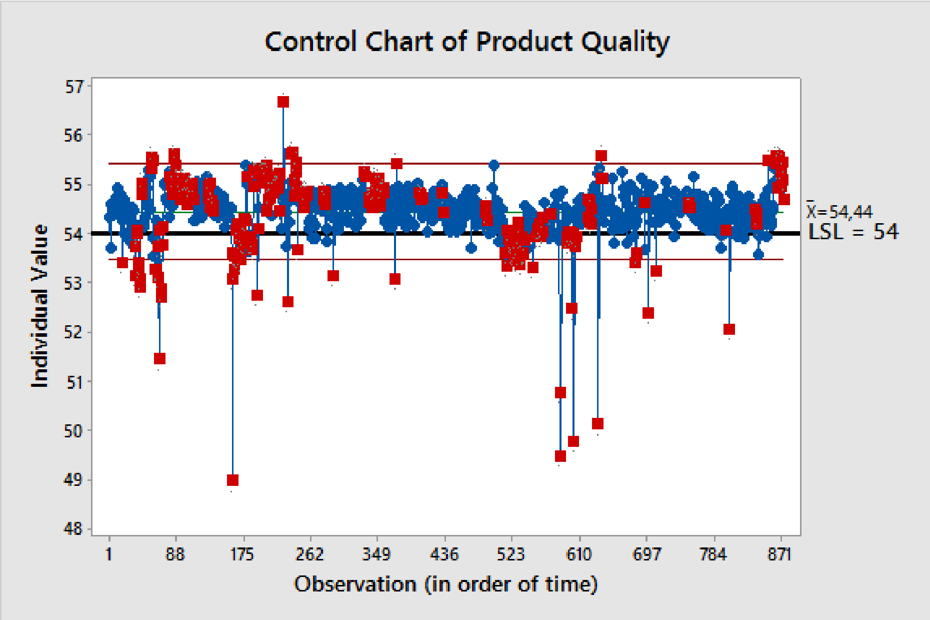

Figure 2: Control chart made from initial quality data

This control chart highlights that the process is not in control. The red subgroups indicate either individual products that deviate far from the process average, or they mark red areas that indicate the process mean has shifted significantly.

We can use the control chart to find moments in time where the process mean started to shift. Such moments could be linked to logbooks of the process, in search of an explanation for the shift. If the cause of a shift could be identified, it could be investigated how future events of that cause can be avoided. If successful, future quality variation will potentially be reduced, and with that the reject level.

Unfortunately, the complex nature of a process, the amount of possible causes and the lack of clear logbooks in the beginning of process improvement, makes it hard to link the quality variation to causes. One

example which makes analysis difficult is that process variation in the quality parameter is caused by variation in previous process steps so the operator at the final process step has no control over these parameters. But if both process and quality information is stored, a statistical analysis of the data can be powerful by linking process variables to quality variables.

Objective of the example
Investigate to what extend an analysis of the production database can explain the reasons of this reject.

The production process
The example concerns a production process where several components are assembled, and each final product is tested on a specific quality parameter once. The higher the quality value, the better the product. There is a lower specification limit, below which the product needs to be rejected.

The production process comprises some production steps and measurement steps that are needed to measure the quality. At each step quality information and process information was gathered using DataLyzer. The purpose of gathering data at each process step was to apply statistical process control.

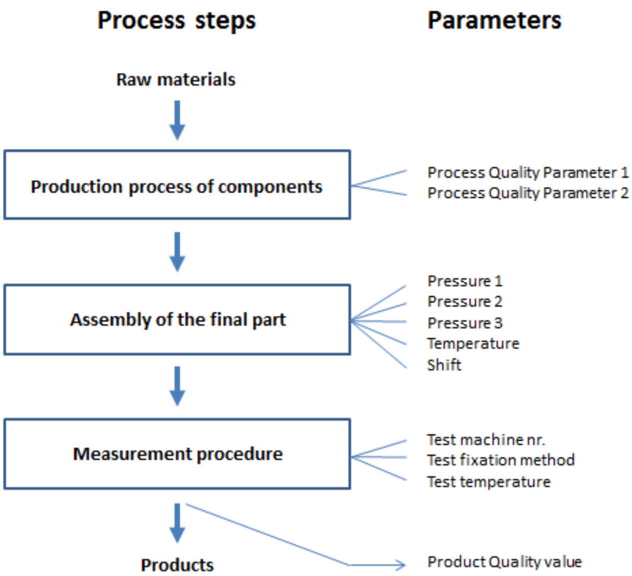

Figure 3: Schematic representation of the process

3 The data

The data gathered at different process steps can be combined in DataLyzer so that parameters for each process step are related in the final product table. Table 1 shows an impression of what the data set looks like.

Date	Ordernr	Pressure 1	Pressure 2	Pressure 3	Temperature 1	Delay time	Shift	Process Qual.par.1	Process Qual.par.2	Test machine nr	Test fixation method	TestTemp	Product Quality
24-jun-10	106	1,0	3,6	2,9	145,4	14	1	33,5759	8,21923	2	B	23,6	54,7845
24-jun-10	106	1,1	1,8	16,7	145,0	17	1	33,5605	8,23084	1	A	23,6	54,8079
24-jun-10	106	1,0	5,8	4,6	144,4	3	1	33,5855	8,21881	6	A	24,1	54,8708
24-jun-10	106	1,0	2,4	25,6	144,5	6	1	33,5501	8,23773	5	A	24,5	54,5423
24-jun-10	106	1,0	2,7	14,3	144,4	11	1	33,5828	8,26725	3	A	23,8	54,7832
24-jun-10	106	1,0	2,7	14,3	144,4	12	1	33,5533	8,26021	3	B	23,7	54,8772
24-jun-10	106	1,0	3,6	2,9	144,4	14	1	33,5728	8,27949	2	B	23,6	54,8695
24-jun-10	106	1,0	3,6	2,9	144,4	16	1	33,5797	8,27043	2	A	23,8	54,7903
24-jun-10	106	1,1	2,9	14,5	144,0	8	1	33,5026	8,27113	4	B	24,1	54,2458
24-jun-10	106	1,1	2,8	13,4	144,1	10	1	33,4797	8,22469	3	B	23,7	54,1096
24-jun-10	106	1,1	2,8	3,1	144,1	12	1	33,4993	8,23486	2	A	23,6	54,3213

Table 1: Variables/columns in data set

The table above shows columns that represent the collected variables (potential causes of quality variation). The column on the right contains the collected quality data, which was used to construct the graphs 1 and 2. The data set contains 874 rows, one product per row, from top to bottom in the order of production, and has some missing values as the set originates from a

real process. It covers a period of about 1 month, during which this particular product type was produced during several days over that period. Notice that the dataset includes information from several production steps and measurement steps, as well the final quality value. Once a table like this has been collected, different types of analysis can be performed on it.

4 The analysis

The goal is to find out if and how the variation of the final quality can be explained from the available columns (parameters). And subsequently see how the quality variation can be reduced based on that knowledge.

Statistical software provides different types of analyses. These can help us to investigate what columns (variables in the data set) significantly influence the final quality. In this example, Anova allows the investigation of the significance of categorical variables. In our data set these are Ordernr., Shift, Test calibr., Test machine nr. and Test fixation method. If covariate variables are defined within the Anova analysis, also continuous variables can be included in the analysis. The covariates are assessed on their significance based on regression analysis. In the data set the covariates are the pressures, the temperatures, the delay time and both process quality parameters.
The significance of the variables is defined by the p-values that are produced by the analysis. In order to have reliable p-values there are several important aspects. They include the presence of nesting. We will not discuss in these aspects in this article, and in the discussed data set nesting does not occur.

The output of the above analysis may look like this:

Method

Factor coding (-1; 0; +1)
Rows unused 302

Factor Information

Factor	Type	Levels	Values
Ordernr	Fixed	3	106; 107; 108
Shift	Fixed	4	1; 2; 3; 4
Test machine nr	Fixed	6	1; 2; 3; 4; 5; 6
Test fixation method	Fixed	2	A; B

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Pressure 1	1	0,124	0,1242	1,54	0,215
Pressure 2	1	0,001	0,0015	0,02	0,892
Pressure 3	1	0,057	0,0567	0,70	0,403
Temperature 1	1	0,184	0,1836	2,27	0,132
Delay time	1	0,156	0,1562	1,94	0,165
Process Qual.par.1	1	82,857	82,8573	1026,37	0,000
Process Qual.par.2	1	11,816	11,8165	146,37	0,000
TestTemp	1	4,644	4,6436	57,52	0,000
Ordernr	2	14,867	7,4333	92,08	0,000
Shift	3	8,142	2,7140	33,62	0,000
Test machine nr	5	0,283	0,0566	0,70	0,623
Test fixation method	1	0,005	0,0052	0,06	0,799
Error	552	44,562	0,0807		
Total	571	177,639			

No significant effect on quality

Significant effect on quality

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,284128	74,91%	74,05%	72,37%

Remarks:
The rows that were incompletely filled are automatically removed.
The variables with p-values below 0.05 can be seen as significant.

5 Conclusions

The red marked process variables significantly impact quality. If these variables can to be brought better in control, the variation in quality will be reduced (if the correlation is causal). These variables need to be investigated further.

The green marked process variables do not show a significant effect on quality. There is no direct need to have engineers focus on these variables, as they do not affect the quality variation significantly.

The analysis can also be used to simulate what would happen to the quality variation, IF the red marked important variables would be brought fully in control by the engineers:

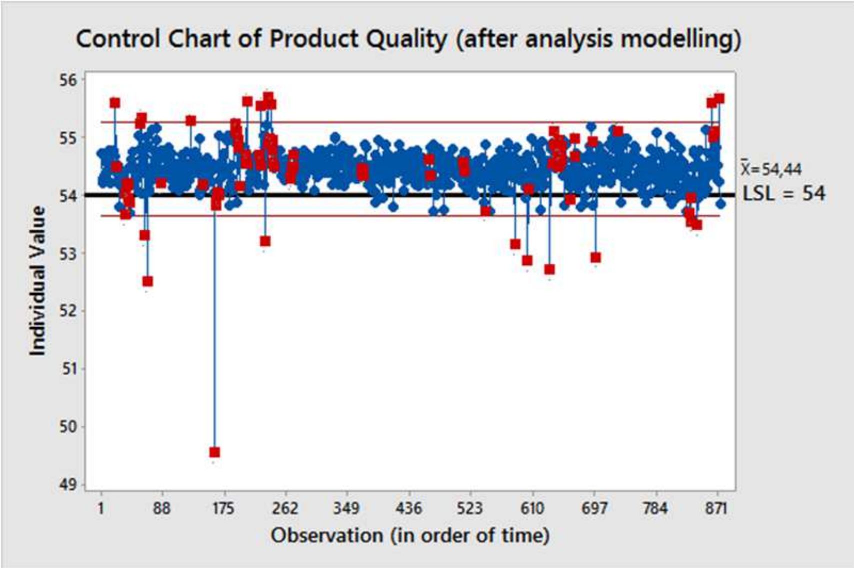

Figure 4: Control chart of quality, assuming the red/significant variables would be fully controlled

Figure 4 shows that indeed the quality is more stable now, with less red subgroups (= out of control). On the other hand, not all the red subgroups are gone. This means that there are some remaining variables that cause shifts in quality, but those variables are not present in the data set yet.

Let's translate the control chart information from figure 4 into a capability graph:

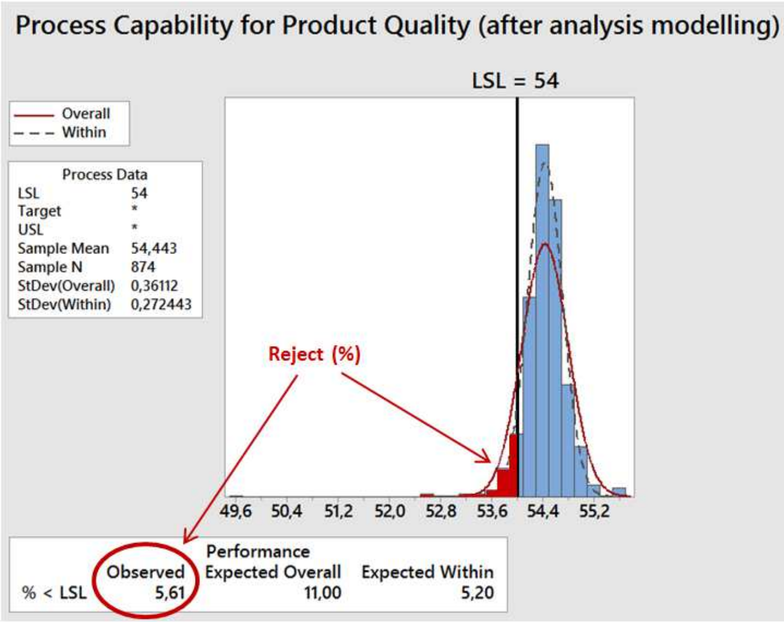

Figure 5: Capability graph of quality, assuming the red/significant variables would be fully controlled

Comparing figures 5 and 1 shows that the observed reject would drop from 13,6% to 5,6%. The benefits of this (still purely theoretical) reject drop can be compared to the expected engineering resources that are anticipated to bring the five significant model terms in control. Another result of the analysis is that no engineering resources are needed to investigate the non-significant variables that are present in the data set.

Conclusions and the value of recording process data

An Anova based analysis can be used to identify process variables that significantly affect shifts in quality. Apparently, these variables are not yet sufficiently in control. This information will help engineers to focus their efforts to improve quality, as they now know where to look.

The analysis also shows which of the recorded variables currently are sufficiently in control, and therefor need no direct attention of the engineers. The results of the analysis can be used to predict quality reject levels that can be reached after the significant variables have been brought in control. It shows the amount of reject reduction that potentially can be achieved. This allows the building of a simple business case: the benefits of the potential reject reduction can be compared to the expected amount of resources required by the engineers to address the specified process variables.

The example illustrates why it is valuable to record not only quality variables, but also process variables: the analysis of recorded data allows us to find process variables that significantly influence quality, even if such variables are too well hidden in the process noise to be detected without statistical data analysis. In addition, the remaining red areas in figure 4 (even after the analysis) indicate that there are process variables that have a significant effect on quality that are not yet recorded in the database. This indicates that it is worthwhile to look for additional process variables that can be recorded, in our continuing efforts to reduce quality variation.

6 Final remark

How is the approach described different from DOE. To perform DOE you have to establish the critical process parameters, you have to establish the different levels and you have to spend resources including production time to conduct the experiments. With the approach described in this document and the possibility to easily register all relevant data in DataLyzer and the option to combine data from different process steps no special experiments are required. The variation found over a time period will be used to make the analysis.

This article is written in cooperation with Mr E. Gijzen from Quality Target
www.quality-target.com