International DATALYZER
Partners in Continuous Improvement

Big Data and Statistical Process Control

By Marc Schaeffers

Big Data and Statistical Process Control

INTRODUCTION

Companies have been applying statistical process control (SPC) techniques since 1935. SPC became very popular after World War II, to continuously improve processes. One of the big advantages of SPC was the power to make solid statements about the process with only a limited amount of measurements. This saved a lot of time taking measurements during production.

Due to the increased possibilities and reduced price of sensors and cameras, we now see that a lot of measurements are taken from processes and products and that all of this data is available for analysis.

The question now is how does big data relate to SPC?

- Is there a conflict between big data and SPC?
- Can they be used in combination and will that strengthen the analysis?

In this whitepaper, we will provide insight into the relationship between big data and SPC.

EXAMPLES OF BIG DATA IN INDUSTRY

In most industries we see that more data is available. Some examples:

- In a blow moulding company, each bottle is inspected with camera's on more than 10 critical dimensions. If an out of spec measurement is found, the product is rejected
- In a semiconductor company, every product is tested 100% before shipment
- In an injection moulding company, all process parameters are recorded for every single shot
- In a food company, every product is measured with a checkweigher to guarantee products have the legal weight
- In a solar company, every cell is inspected to measure the performance.

This means that we will get product or process measurements every 2 to 3 seconds.

These methods are mainly used to prevent bad products from reaching the customer. These methods are not used primarily to improve processes.

Let's explore and explain this with 2 examples.

Checkweigher

In a food company, weights are checked with a checkweigher. When the checkweigher detects a product with a weight below the specification, the product is automatically rejected. When an operator sees too many rejected products, he will adjust the process so that the weight is increased. The checkweigher normally does not give an alarm when the weight is too high and so the result will be that the supplier will offer too much overweight.

It could be that a process has multiple lanes and that only 1 lane is not working correctly, increasing the reject percentage by a few percent. Adjusting the complete machine will solve the issue for that lane but it also means delivering overweight in all other lanes.

This adjustment is illustrated in figure 1.

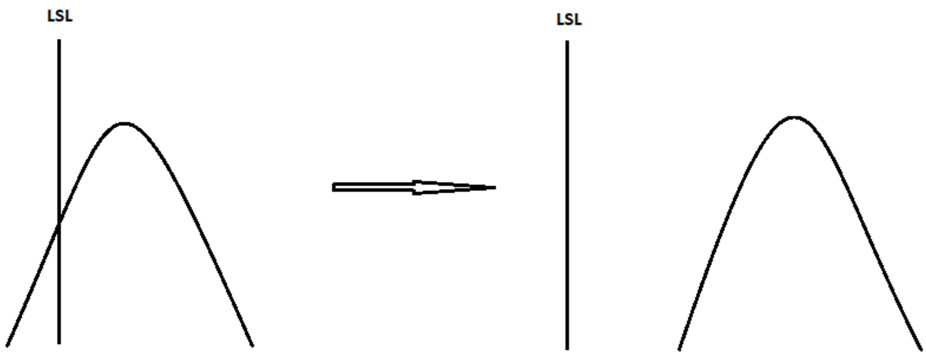

Figure 1: Solving too many rejects in a checkweigher using process adjustment

If the data from the checkweigher or the process will be used to find the root cause it is possible to reduce the variation in weight and keep the process average closer to the minimum required weight. We have seen that this can lead to 1 to 3 % reduction in raw material with the same output. It also results in a more consistent product because you have less adjustments in the process.

This method is shown in figure 2 below.

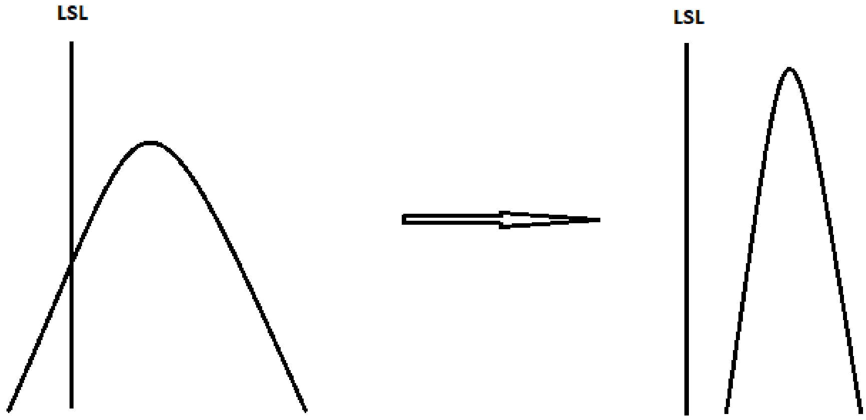

Figure 2: Solving too many rejects in a checkweigher using process improvement

Reject percentage

In an assembly line, every product was tested with a sensor. In case the product was below specification, it was automatically rejected. In a display, the reject percentage was automatically indicated.

When the reject percentage exceeded a predefined setpoint, the operators can see that on the screen and are able to respond to the problem. The disadvantage of such a system is that the operators don't get an early warning in case of a process disturbance. If the process has produced for 4 hours without defects and suddenly is producing 5% defects, the defect percentage will only increase slowly and will take a long time before operators get the signal to respond.

The solution of showing the defect percentage for the last X products is still not sufficient due to the characteristics of attribute inspection.

APPLYING SPC

When analyzing processes, it is very important to distinguish between common causes and special causes of variation. Special causes of variation indicate a process change which needs to be investigated. When analyzing a special cause of variation, you need to distinguish between a change in the process average and a change in the process variation.

Checkweigher and SPC

If we show the data of a checkweigher in a control chart and show many consecutive products in the chart, we get a warning that something has changed in the process. The chart will indicate whether the average is out of control or the variation. This will assist the operator to find the root cause of the problem.

In figure 3 you see a change in the weight variation and what the effect is on the rejects

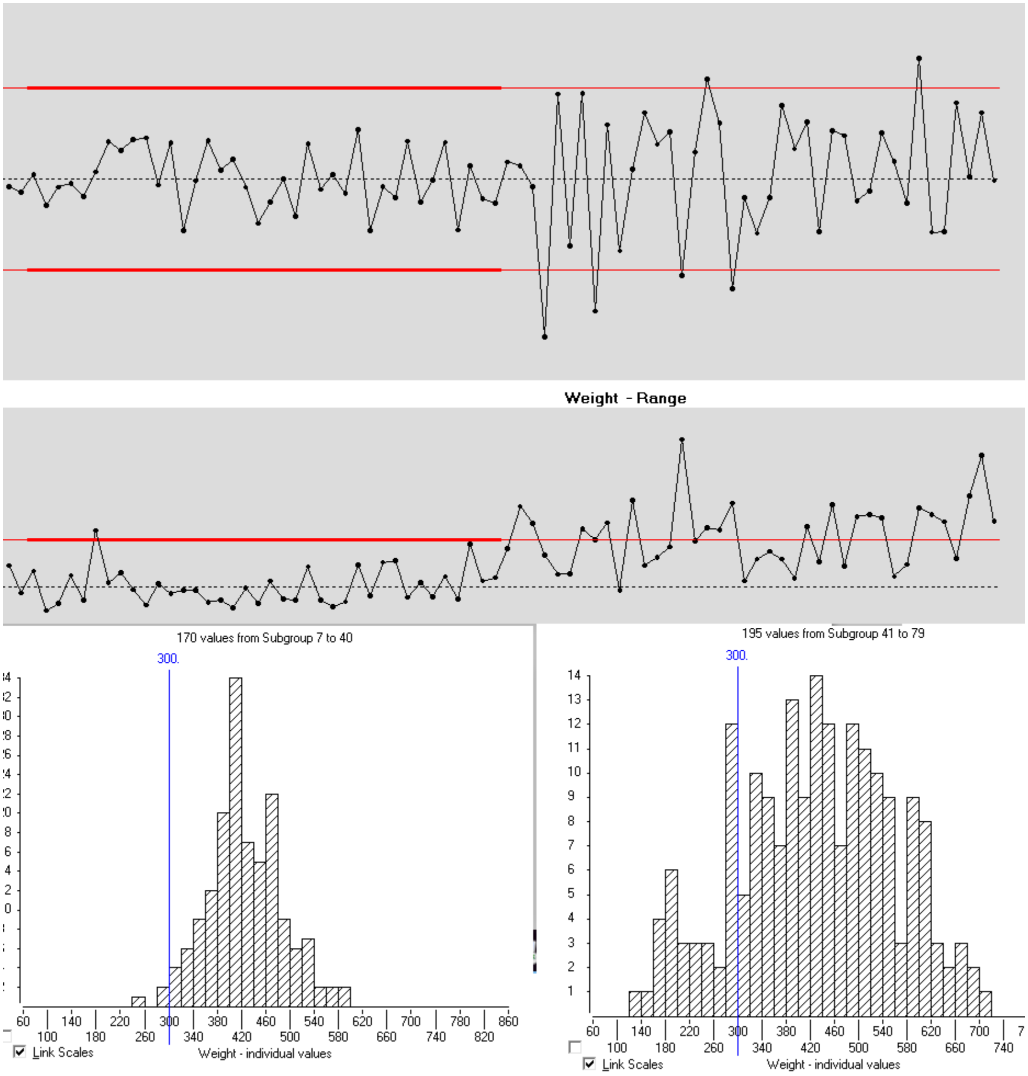

Figure 3: Checkweigher results with change in process variation

It is clear, that adjusting the average of the process in this situation, will be an expensive solution of the problem.

Rejects and SPC

Instead of showing the rejects in a percentage, it makes more sense to show the rejects in a control chart. Figure 4 shows a control chart of a simulation where the process was running with 0,5% defects during the first 40 subgroups and with 1,5% defects during the second 40 subgroups.

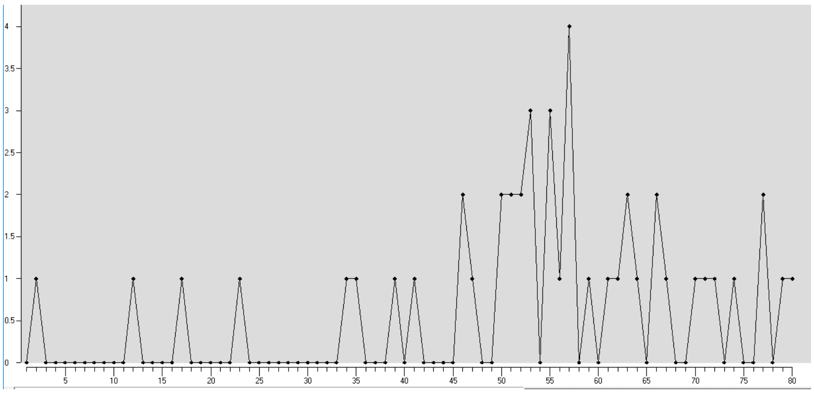

Figure 4: Attribute chart of rejects with n=50

The chart indicates that showing the reject percentage of only the last 50 will still not give a good indication because there is still a large chance that 0 or only 1 defect is found at times giving the operator the idea the process is back at the right level again. Because such a chart might be hard to interpret for an operator, we often present the chart in the format of a table.

In figure 5, you see the reject chart in the format of a table. This is integrated in the real time OEE system, so a signal in the OEE screen indicates a special cause is found in the reject level. The operator can open the screen with the numbers of rejects per subgroup and if there are red items he can then start the analysis to find the root cause.

Using SPC techniques in combination with results from 100% inspection offers the operator a tool to respond much quicker to process disturbances.

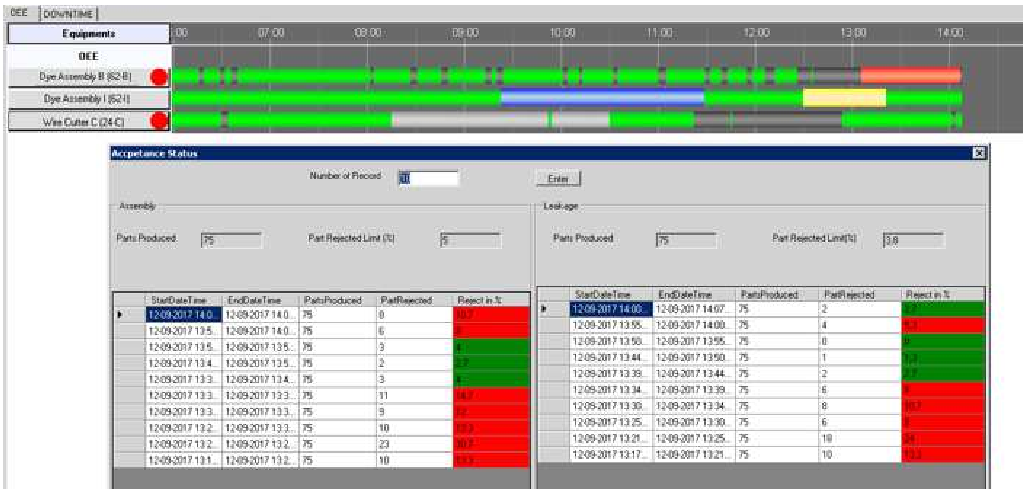

Figure 5: Reject analysis per subgroup integrated in OEE system

Big Data, SPC and a large number of characteristics

The examples above are still relatively easy because we are only looking at a few critical characteristics. If we have a lot more characteristics, the situation becomes more complex. In that case, we need a summary overview to see the status of all characteristics at a glance.

In figure 6, you see the status screen of a medical device company. Each square represents a characteristic and measurements were taken automatically from sensors. On average, every 10 seconds a subgroup was added.

Figure 6: Status screen of all control charts in production

A red point indicates an out of control condition. Normally in case of an out of control condition, we need to start an investigation to find the root cause. In the situation where we measure 100% and register the data in a control chart, we run into a problem.
When a process is perfectly in control, we will have 3 out of controls every 1000 subgroups on average and 3 on the range (false alarms).

This means that we will have a false alarm every 166 subgroups. With a subgroup every 10 seconds, it means we will have a false alarm on every chart, every 27 minutes. With more than 100 charts, it means we will have a false alarm every minute.

The result is that this system will not be effective anymore and people will ignore the control charts completely.

The solution can be found to aggregate or filter the data to make the data meaningful again.

There are several ways of aggregating this data to extract meaningful information:

- You can reduce the sampling frequency and only take a subgroup on regular intervals
- You can take the average and standard deviation of a number of consecutive subgroups and show these results in the chart
- You can convert the data to an attribute subgroup and count the number of out of control or out of specification conditions
- You can add a threshold on the number of out of controls. Consequently, a chart will get an alarm status if the percentage of out of controls exceeds a threshold limit. This is very similar to the solution applied in the reject problem
- A combination of the solutions above.

Which solution is the most suitable depends on a number of factors. For example, if 100% tracking and tracing is required, the solution to take subgroups with a lower frequency might not be suitable.

Process control versus engineering analysis

I hope we have showed that applying SPC to big data can give you a competitive advantage improving your processes and results.

In some cases, it is required to correlate data from multiple process steps to find the root cause of the problem. In that case, it is mandatory that a common denominator is defined in the measurements to establish the relation. The denominator is typically a product or batch identifier and in some cases, it is date and time.
In those cases, it might be useful to convert the data for process control to a data warehouse where the data is synchronized and can be used effectively by engineers

Conclusion

When large amounts of data are collected, it makes sense to use SPC techniques to analyze the data to get accurate information about special causes of variation.

If data is measured with a high frequency, measurement data cannot be used straightforwardly in the control chart, otherwise you would get to many false alarms and that would undermine the system. In that case, data will need to be aggregated or filtered to convert the data into useful information again.