JMP Academic Case Study 064

Variation in Semiconductor Etching Process

Quality Control, Control Charts, Process Capability

Produced by:

Kevin Potcner, JMP Global Academic Team
kevin.potcner@jmp.com

Variation in Semiconductor Etching Process

Quality Control, Control Charts, Process Capability

Key Ideas

This case study illustrates methods for understanding process variation and comparing it to process specifications through graphical analyses, control charts, and capability analyses.

Background

A semiconductor manufacturer is working on getting seven new etching lines up-and-running at a new facility to satisfy increased demand for chips. A critical step in producing integrated circuits is to etch select areas of an oxide layer from silicon wafers. One important quality metric is *Etch Rate*. This is the amount of oxide etched from areas of the wafer per unit time, measured in angstroms per minute.

Ideally, all wafers would be etched consistently to target and with variability within the process specifications. The target Etch Rate for these lines is 620, with lower and upper specification of 545 and 695 respectively (i.e., 620 +/- 75). Etching is done in sets of 20 wafers. A sampling plan has been established which will randomly sample 5 wafers from each run for 40 subsequent runs, resulting in 200 wafers sampled from each line. *Etch Rate* will be measured at 4 separate locations on each wafer.

This results in 5 x 40 x 4 = 800 *Etch Rate* values per etching line. Based upon this sampling strategy, three sources of variation can be examined for each line: 1) within wafer; 2) between wafer within run; 3) between runs.

Importantly, equipment recalibration was performed after the 20th run. This allows run variation to be separated into two other sources: 3a) between runs within calibration stage; 3b) between calibration stages.

In this case study, we will demonstrate a variety of analyses on one of the etching lines. The exercises will ask you to conduct similar analyses on the other six etching lines.

The Task

The primary objectives of the analyses are:

1. Evaluate the performance of each etching line relative to the target of 620 +/- 75.
2. Examine the four sources of potential variation in *Etch Rate* for each line:
   - within wafer
   - between wafer within run
   - between runs within calibration stage and between calibration stages.

The Data **variation-in-semiconductor-etching-process-line-a.jmp**

Calibration	Identification of the two calibration stages (1, 2)
Run	Identification of each run of the etching tool (1, 2, … , 40)
Wafer	Identification of the wafers within each run (1, 2, … , 200)
Run-Wafer	Identification of the Run and Wafer (1-1, 1-2 ,…,1-5 , 2-1,…, 2-5, 3-1 , … , 40-5)
Etch Rate	The *Etch Rate* of a given section of the wafer (in angstroms per minute)

Analysis

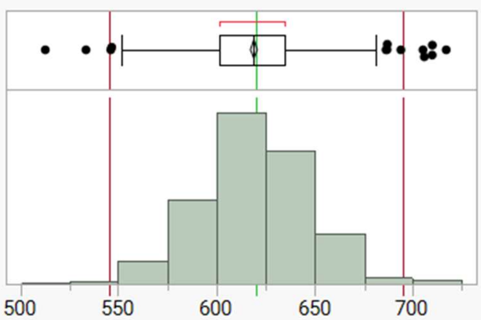
Numerical and graphical summaries of overall etch rate

We begin by numerically and graphically summarizing the etch rate data for Line A. Exhibit 1 is a histogram, boxplot, and summary statistics for the 800 *Etch Rate* values with reference lines added for the target and specification limits.

Exhibit 1 Histogram, Boxplot, Summary Statistics

Distributions Student Edition

Etch Rate



Quantiles

100.0%	maximum	717.12012059
99.5%		705.94781063
97.5%		671.10512327
90.0%		654.14404447
75.0%	quartile	634.98873772
50.0%	median	619.31238187
25.0%	quartile	601.67468879
10.0%		584.53197666
2.5%		565.52237143
0.5%		546.47234232
0.0%	minimum	512.55024894

Summary Statistics

Mean	619.01072
Std Dev	27.051351
Std Err Mean	0.9564097
Upper 95% Mean	620.88809
Lower 95% Mean	617.13335
N	800
N Missing	0

(Analyze > Distribution. Select 'Etch Rate' as the Y Variable. Click OK. Double click on the X axis of the histogram created in the report. Add 3 reference lines at 545, 620, and 695 by typing in each value and selecting Add one at a time. Note: Desired colors and line styles can be chosen.)

The histogram and boxplot show that the overall process variation extends beyond the specification limits (i.e, one observation below the lower limit of 545 and four above the upper limit of 695). This amount of variation is undesirable. Ideally, variation would be well within the specification limits.

The sample mean of all 800 etch rate values is 619, a bit below the target of 620 but certainly very close. However, it is best to perform a formal inferential statistical technique to determine if this small observed difference provides enough statistical evidence to conclude that the process is running off the target of 620. One such technique is to use the confidence interval for the mean, which is displayed in the Summary Statistics table. We can interpret these results as showing that, with 95% confidence, the mean *Etch Rate* of this process is between 617.1 and 620.9. The width of the confidence interval reflects the uncertainty in using a sample to estimate the mean of a population of wafers produced by this process. Note that the interval contains 620. Even though we estimate that the process average could be as low as 617.1, we also estimate it could be as high as 620.9. In other words, 620 is a plausible value for the overall average. We conclude that we don't have statistical evidence that this process is off target on average.

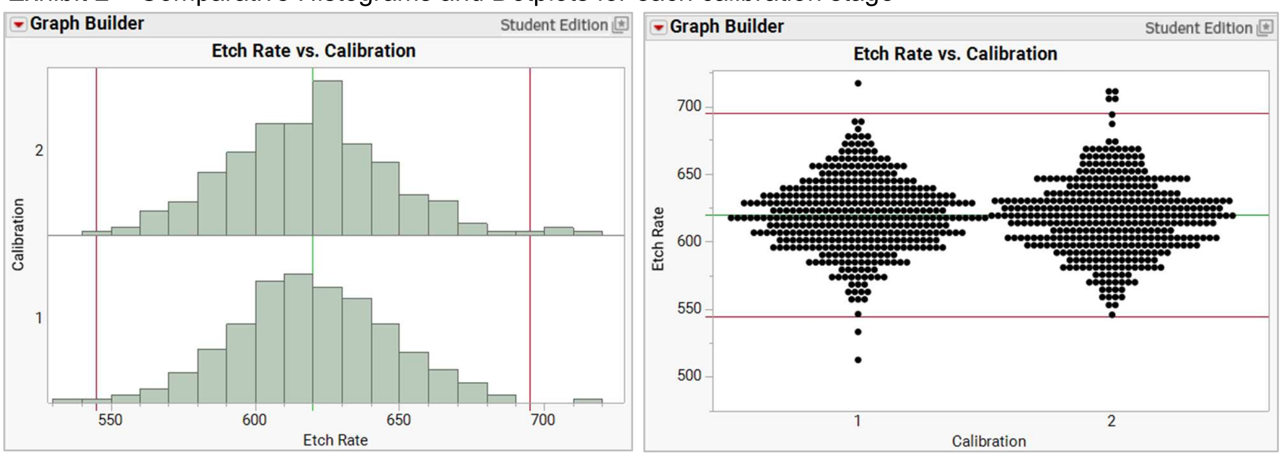
It is important to keep in mind that the process average, say for millions of etchings, would never be exactly equal to 620. In practice, there is some range of values from this target that we would consider close enough to 620. For example, if the process average was 619.9, the engineers may consider this ‘equivalent’ to 620. Equivalence is a statistical concept not covered in this case study but a topic that is used frequently in engineering applications and would be valuable to learn about. Note: JMP can perform this analysis in the Distribution Platform (Test Equivalence under the Red Triangle).

Note that these graphical and numerical summaries do not allow us to examine any potential differences in *Etch Rate* between wafers, runs, or the calibration stages. It is merely an overall summary of the data. We will create additional graphs allowing us to evaluate those sources of variation.

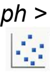
Comparing etch rate across calibration stages and runs

Recall that the engineers recalibrated the equipment after 20 runs. It is important to compare the performance of the process within and between these calibration stages. Exhibit 2 shows comparative histograms and dotplots displaying the *Etch Rate* values within the 2 calibration stages.

Exhibit 2 Comparative Histograms and Dotplots for each calibration stage



(Comparative Histograms: Graph > Graph Builder. Place ‘Etch Rate’ on the X Axis, and ‘Calibration’ on the Y. Select the Histogram icon in the graph palette.

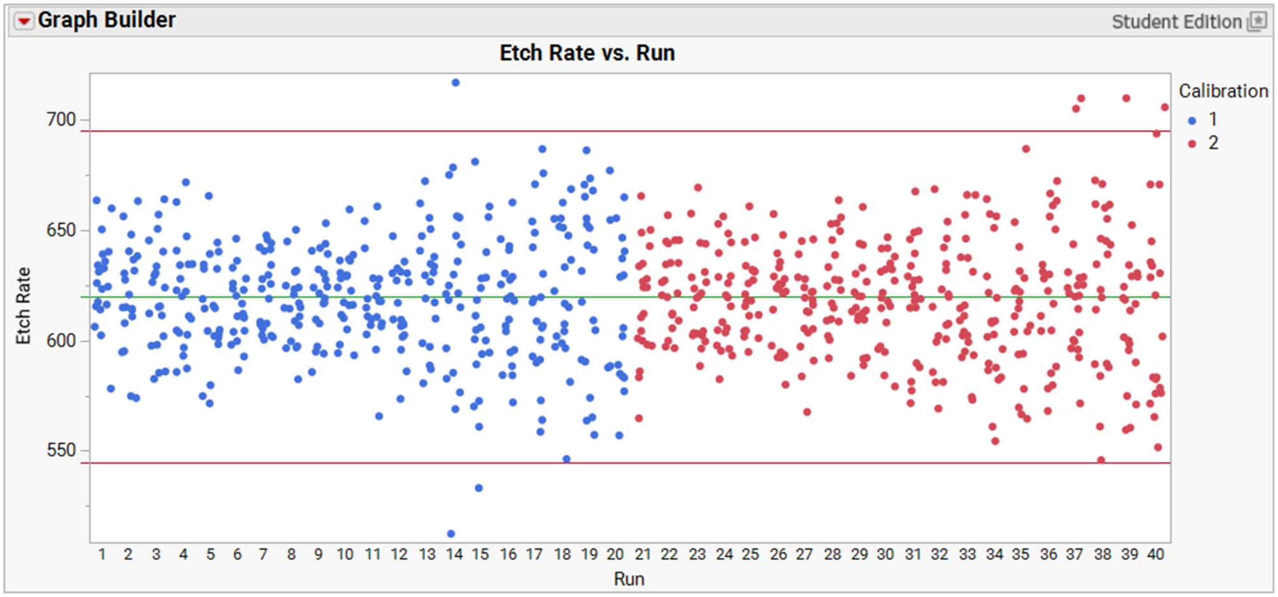
Comparative Dotplots: Graph > Graph Builder. Place ‘Etch Rate’ on the Y Axis, and ‘Calibration’ on the X. each Select the Scatterplot icon  *in the graph palette.*

Reference lines can be added in the same way as illustrated in creating the overall histogram in Exhibit 1.)

These graphs show that the variation in etch rate is excessive during both calibration stages, and that there isn’t any obvious difference in that variation or in the average. A statistical test (not shown here) comparing the process means between these two stages (2-sample t-test) confirmed this.

To continue exploring the variation, we create a graph that displays the 5 x 4 = 20 *Etch Rate* values for the 40 different runs. Exhibit 3 is a comparative dotplot with the 40 runs and 2 calibration stages shown.

Exhibit 3 Comparative Dotplots for each Run



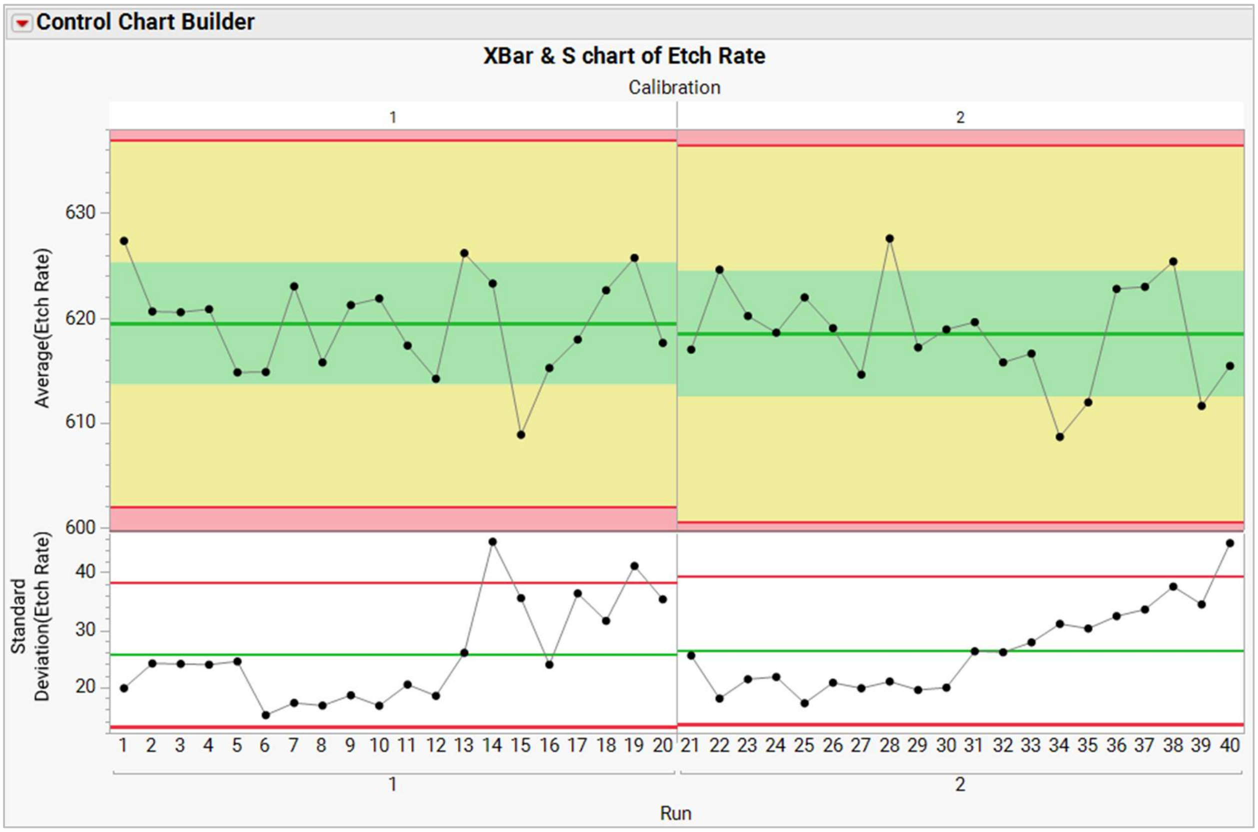
(Graph > Graph Builder. Place 'Etch Rate' on the Y Axis, Run on the X, and 'Calibration' in the Color role. Add reference lines for the specification values.)

This graph reveals that the variation in Etch Rate appears to be larger in the later runs within each calibration stage. This is evident in seeing more values closer to the specification limits and extending beyond these limits in the later runs. This suggests that the recalibration of the equipment helps in making the *Etch Rates* more consistent. There does not appear to be any significant changes in the average *Etch Rate* within each run. Granted, they vary a bit, but it's not clear if those changes indicate true run-to-run variation. We will explore this in more detail via control charts.

Control Charts

Control charts are a useful way to examine multiple sources of variation in a process over time. Exhibit 4 shows X-Bar & S Charts that do this. In this pair of graphs, the X-Bar chart plots the mean of the 5 x 4 = 20 etch rate values within each run, and the S Chart plots the standard deviations. It's important to understand that the lines being drawn on these graphs are not the target and specification limits as we had done in the previous graphs, but the center line and control limits. When examining X-Bar & S charts, it's advised to first examine the S Chart which here summarizes within-run variability. This is because, as we'll explain in more detail in a moment, the control limits for the X-Bar chart are based upon estimates of variation coming from the S Chart.

Exhibit 4 X-Bar & S Chart



(Analyze > Quality and Process > Control Chart Builder. Place 'Etch Rate' on the Y Axis, 'Run' in the Subgroup role, and Calibration in the Phase role. In the controls on the left, choose Standard Deviation in the sigma dropdown for Limits[1], Limits[2], and for Points[2]. Select to Shade Zones under the Limits[1] controls.)

As we saw in the graph in Exhibit 3, the variation is larger in the runs further from the calibration step. Plotting the standard deviation for each run, as the S Chart does, can make it easier to see that feature than looking at dotplots of all 20 data values in each run. For the S Chart, the green lines are the average of all 20 standard deviation values within each calibration stage. The red lines are the lower and upper control limits. These are based upon the variation in the standard deviation values. They are placed +/- 3 standard deviation of the statistic being plotted. Here the statistic being plotted is the standard deviation, so the control limits are +/- 3 standard deviations of the standard deviation.

The center line in the X Bar chart is the average of all 20 mean values within each calibration stage. The control limits are +/- 3 standard deviations of the sample mean. The estimate of the standard deviation used to create the control limits comes from the average of the standard deviation values from the S Chart.

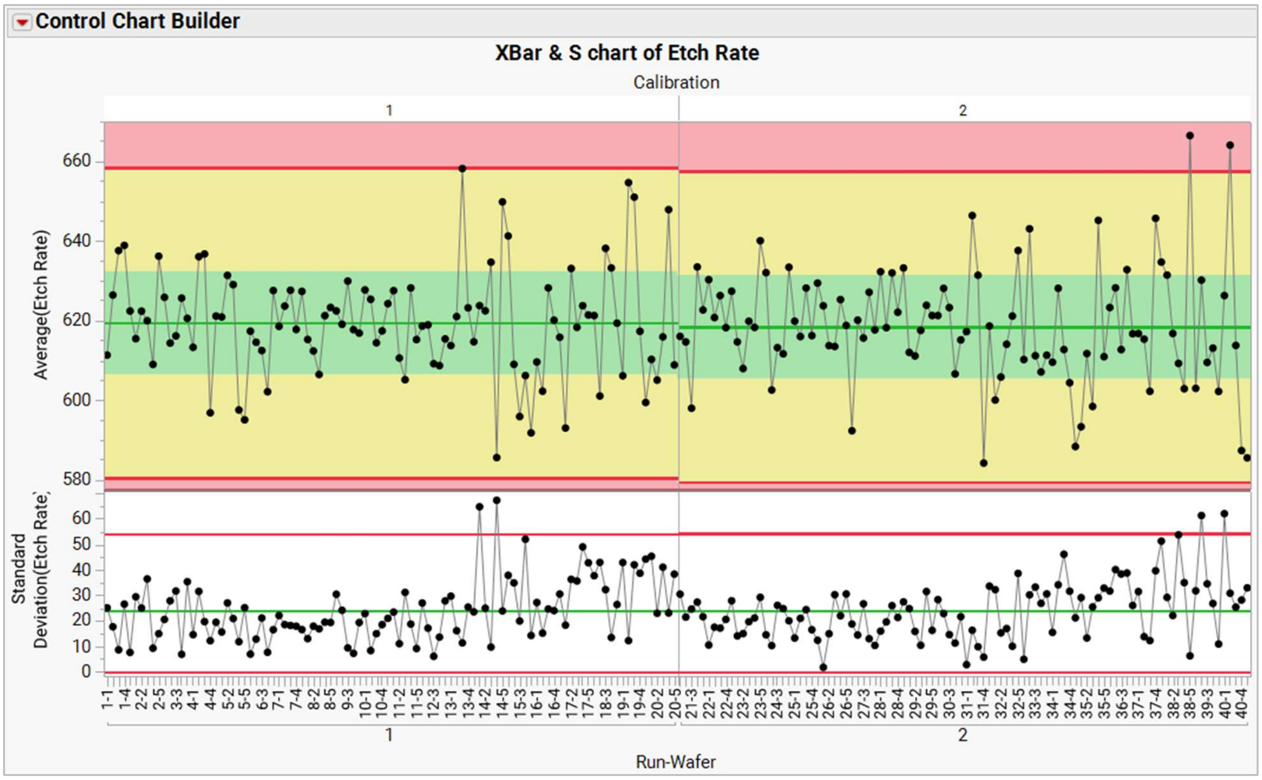
The reason why it is recommended to examine the S Chart first is that if the standard deviations are not exhibiting consistent variation over time (e.g., here the variation gets larger over time), then the estimate of the within-run standard deviation can be inaccurate. In this case, it is over estimated, and the control limits for the X-Bar chart are wider than they would be if this phenomenon weren't happening.

Nonetheless, we'll examine the X-Bar chart. Notice that all the average *Etch Rate* value are well within the control limits and don't show any indication that any factors change the average *Etch Rate* between runs. Still, the control limits would be a bit tighter if the variation didn't get larger across runs. These control charts can be made using only the first 10 runs after calibration as a means to examine the process when the variation is consistent. These were done (not shown here) and indicated there was no evidence of changes in the between-run variation.

There are two other sources of variation to examine: between-wafer within-run variation and within-wafer variation. We'll need to create a different set of control charts to display those.

Exhibit 5 shows X Bar & S Charts where the points on the S Chart are the standard deviations of the 4 measurements made on each wafer (within wafer variability), and the X-Bar Chart is the mean of those 4 measurements. Examining the variability between those means gives us visibility into the between-wafer variation. Recall 5 wafers were sampled from each run of 20 wafers. Thus every 5 consecutive points in these graphs are from the same run.

Exhibit 5 X Bar & S Chart



(Analyze > Quality and Process > Control Chart Builder. Place Etch Rate on the Y Axis, Run-Wafer in the Subgroup role, and Calibration in the Phase role. In the controls on the left, choose Standard Deviation in the sigma dropdown for Limits[1], Limits[2], and for Points[2]. Select to Shade Zones under the Limits[1] controls.)

Notice that standard deviation values get larger over time within the calibration phases. This is similar to what we saw in the S Chart where Run was the subgroup, and it reveals that this increase in variability is not just happening between the wafers within the runs, but also within the wafers. In comparing the S charts in Exhibit 4 and 5, it appears that change in variation is most severe between runs.

The X-Bar chart plots the average of the 4 *Etch Rate* values within each wafer. Examining the variation between these means is another way to examine between-wafer variability. This chart also shows that the variation in *Etch Rate* values gets larger the further the runs are from the calibration.

Based upon these control charts, we would conclude the process is not in statistical control throughout all 40 runs. Specifically, process variability increases over time, with recalibration then reducing the variability back to baseline. Recall, statistical control refers only to the stability of the process. In other words, if the variation and central tendency of a process is consistent over time, we would say the process is in "statistical control" and exhibits only common cause variation. Yet, a process may be in control even if the variation is larger than desired or the process mean does not equal the target. In this case, data values may still lie beyond the specification limits or the process may not be near target. How well a process performs relative to a target and specifications is known its "capability".

Process Capability

Process Capability Analysis produces a set of metrics that summarize performance relative to specifications. It is generally not advised to perform a capability analysis on an out-of-control process, as estimates of overall performance may not be generalizable. On the other hand, if the process is in statistical control, then it is reasonable to use estimates of process capability to infer beyond the data collected.

In this case, it seems that the change in variation (i.e., when the process goes out of statistical control) happened around the 13^{th} run after each calibration of the equipment. Based upon that, we could think about how the process would perform if say the recalibration happened after every 10 runs. In the case, it is reasonable to use the data from the first 10 runs after each recalibration as an estimate of how the process would perform under such operating conditions.

To illustrate how to perform and interpret a process capability analysis, we will use data only from the first 10 runs after each of the recalibrations. To do this we will create a new column indicating the runs that are the first 10 after recalibration and those that are 11-20.

Create New Column via Recode:

Select the Run column in the data table. Right-click and choose Recode. A dialog box will appear showing each of the values in the Run column (1, 2, 3, …. 40). Give a name for the new column at the top (e.g., "Runs After Recalibration"). Select the values 1 through 10 in the Old Values list. Click Group. Provide a New Value (e.g., "1-10"). Select the values 11 through 20 in the Old Values list. Click Group. Provide a New Value (e.g., "11-20"). Repeat this for the runs of 21 through 30 and 30 through 40. Click Recode. A new column will now be created with two values ("1-10" and "11-20").

Create a Global Data Filter:

Select Rows > Data Filter. The Data Filter dialog box will appear. Uncheck the Select box and check the boxes Show and Include. Select the variable "Run" in the variable list and click the +. Select the "1-10" values. In the data table, row states will get created that hides and excludes the 11-20 runs. Now any analyses performed will only be done on the first 10 runs after the recalibration. This allows us to rerun every analysis we have saved by clicking the green triangle next each of the saved scripts. To see this, do so to reproduce all the graphs and analyses shown in Exhibit 1 through 5. Note: The data filter can be turned off by simply selecting Clear.

Now we will perform a capability analysis using only runs 1-10 and 20-30. We will launch this analysis from within the Distribution Platform.

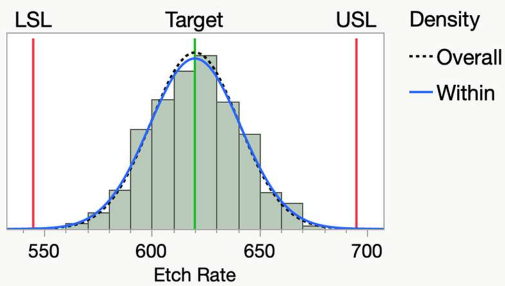
Exhibit 6 Process Capability

Process Capability

Etch Rate[4] Capability

Grouped by subgroup size: 4

Histogram



Process Summary

LSL	545
Target	620
USL	695
N	400
N Subgroups	100
Sample Mean	620.0296
Within Sigma	21.04816
Overall Sigma	20.34208
Stability Index	0.966454

Within sigma estimated by average of unbiased standard deviations.

Within Sigma Capability

Index	Estimate	Lower 95%	Upper 95%
Cpk	1.187	1.084	1.291
Cpl	1.188	1.085	1.292
Cpu	1.187	1.084	1.291
Cp	1.188	1.090	1.286
Cpm	1.188	1.105	1.270

Overall Sigma Capability

Index	Estimate	Lower 95%	Upper 95%
Ppk	1.228	1.137	1.320
Ppl	1.229	1.138	1.321
Ppu	1.228	1.137	1.320
Pp	1.229	1.144	1.314
Cpm	1.229	1.145	1.316

Nonconformance

Portion	Observed %	Expected Within %	Expected Overall %
Below LSL	0.0000	0.0182	0.0113
Above USL	0.0000	0.0184	0.0114
Total Outside	0.0000	0.0366	0.0227

(Rerun the Distribution script. Under the red triangle next to the variable name Etch Rate, select Process Capability. Enter the values 545 for the LSL, 620 for the Target, and 695 for the USL. Open the Process Capability Options. Select Use Constant Subgroup Size. Set the subgroup size to "4". This will result in the estimate of within subgroup variation to be based upon the Etch Rate measurement made on each wafer. Selecting 20 would define the subgroups to be all 4 x 5 = 20 measurements within each Run. Click OK.)

Exhibit 6 shows the process capability output added to the Distribution report. We will look at the values for some of these as well as the formulas that produces them. The capability metrics are ways to quantify how the variation in the process compares to the specifications of that process. Two fundamental metrics are the "within sigma" and the "overall sigma" estimates. The within sigma is the mean of all 100 standard deviations, each calculated from the 4 *Etch Rate* values within each wafer. The overall sigma is the standard deviation of all 400 *Etch Rate* values. In these data, the two estimates are very similar (21.0 and 20.3). If they were much different, say the overall standard deviation being much larger than the within standard deviation, this would indicate that there are some factors causing additional variation between the subgroups. It is a positive outcome that they are so similar.

We will examine two sets of metrics: “Within Sigma Capability” and “Overall Sigma Capability”. These are the formulas for each set:

Within Sigma Capability

$$C_p = \frac{USL - LSL}{6\hat{\sigma}_{within}}$$

$$C_{pl} = \frac{\bar{X} - LSL}{3\hat{\sigma}_{within}}$$

$$C_{pu} = \frac{USL - \bar{X}}{3\hat{\sigma}_{within}}$$

$$C_{pk} = min\{C_{pl}, C_{pu}\}$$

Overall Sigma Capability

$$P_p = \frac{USL - LSL}{6\hat{\sigma}_{overall}}$$

$$P_{pl} = \frac{\bar{X} - LSL}{3\hat{\sigma}_{overall}}$$

$$P_{pu} = \frac{USL - \bar{X}}{3\hat{\sigma}_{overall}}$$

$$P_{pk} = min\{P_{pl}, P_{pu}\}$$

As you can see, the only difference between the two sets of metrics is which sigma is used. Notice that C_p and P_p are a ratio comparing the specification width to the process variation measured by 6s. This is a logical approach for a variable that can be modeled as a symmetric normal distribution, as +/- 3s from the mean captures 99.7% of a normal distribution. An examination of the histogram and normal distributions fit to these data demonstrate that the normal distribution is a very good model to use for this process. Note that there are two normal distributions drawn on thae graph: one based upon the overall standard deviation and the other upon the within. Because those values are approximately equal, the two distributions are almost identical. The values for C_p and P_p are 1.19 and 1.23. This is interpreted as saying that the width of the specification is 1.23 times larger than the variation in the process (+/- 3s). Values of 1 indicate that they’re the same. Certainly, values greater than 1 are desired, but most practitioners desire these capability metrics to be much greater than 1. Importantly C_p and P_p don’t take the center of the process into account. For example, the values for C_p and P_p would be the same even under the scenario where the mean of the process was 200 and all *Etch Rate* values were below the LSL. For this reason, C_p and P_p are not the primary metrics used in capability analysis.

In practice, the metrics C_{pl} , C_{pu} , C_{pk} , P_{pl} , P_{pu} , and P_{pk} are more often used. Again, the only difference between the two sets of metrics is which estimate of process variation is used – within or overall. C_{pl} and P_{pl} are measures of process capability relative to the lower specification limit and C_{pu} and P_{pu} are measures of the process capability relative to the upper specification limit. These metrics take into account how close the process mean is to that particular specification limit. Values > 1 indicate that the variation from the mean (measured via 3s) does not reach the specification limit. Values < 1 indicate that the variation from the means extends beyond the specification limit. Here, we see all values are near 1.19 and 1.23. The fact that they are all similar to each other and to C_p and P_p indicate that not only is the variation (as measured via +/3s) within specification, but that the process mean is also in the center of those limits, on average. Notice that the mean of all the *Etch Rate* values is 620.03, almost identical to the target, which is centered between the specification limits.

C_{pk} and P_{pk} are frequently used as they each provide a single metric that quantifies the worse of the two scenarios (process variation relative to the lower or upper specification).

Process engineers often work to improve a process until a certain level of capability is reached. For example, C_{pk} and P_{pk} values greater than 2 indicate that the process average is more than 6 standard deviations from the nearest specification limit, often referred to as a “6 Sigma” process.

You may need to communicate the capability of a process to people not familiar with these metrics. As a result, a common way to report the performance of a process is the estimated “percentage of a process that will be outside of the specifications” or the estimated “Parts Per Million (PPM) that will be outside of the specifications”.Percentage estimates are the default output. Exhibit 7 shows the addition of PPM.

Exhibit 7 Process Capability

Nonconformance

Portion	Observed %	Expected Within %	Expected Overall %	Observed PPM	Expected Within PPM	Expected Overall PPM
Below LSL	0.0000	0.0182	0.0113	0	182	113
Above USL	0.0000	0.0184	0.0114	0	184	114
Total Outside	0.0000	0.0366	0.0227	0	366	227

(To add the PPM estimates to the report, right-click on the Nonconformance table. Choose columns, then select the PPM metrics. You can also choose to Format any of the columns under that menu. Here we chose to display the values as integers.)

Examining the larger of the two “Total Outside Expected PPM” value, we would say that we estimate that 366 out of every million etchings would have an *Etch Rate* outside of the specification limits. This is quite small and in many cases may be considered a process that is “good enough”.

Keep in mind that this capability analysis was based on eliminating all the data after the 10^{th} run of the process after calibration. This calibration procedure would have to be followed in order to achieve that capability. Doing the calibration procedure every 20 runs as was done during this data collection would not result in such a good performance. Exhibit 8 below shows the results of a Process Capability Analysis performed on all the data.

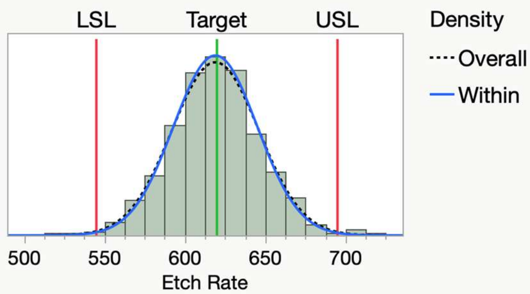
Exhibit 8 Process Capability

Process Capability

Etch Rate[4] Capability

Grouped by subgroup size: 4

Histogram



Process Summary

LSL	545
Target	620
USL	695
N	800
N Subgroups	200
Sample Mean	619.0107
Within Sigma	26.0173
Overall Sigma	27.05135
Stability Index	1.039745

Within sigma estimated by average of unbiased standard deviations.

Within Sigma Capability

Index	Estimate	Lower 95%	Upper 95%
Cpk	0.948	0.888	1.008
Cpl	0.948	0.888	1.008
Cpu	0.974	0.912	1.035
Cp	0.961	0.905	1.017
Cpm	0.960	0.913	1.007

Overall Sigma Capability

Index	Estimate	Lower 95%	Upper 95%
Ppk	0.912	0.862	0.962
Ppl	0.912	0.862	0.962
Ppu	0.936	0.885	0.988
Pp	0.924	0.879	0.969
Cpm	0.924	0.879	0.969

Nonconformance

Portion	Observed %	Expected Within %	Expected Overall %	Observed PPM	Expected Within PPM	Expected Overall PPM
Below LSL	0.2500	0.2223	0.3110	2500	2223	3110
Above USL	0.6250	0.1746	0.2484	6250	1746	2484
Total Outside	0.8750	0.3969	0.5594	8750	3969	5594

(In the data filter, select Clear. Then choose Redo > Redo Analysis under the red triangle in the top left of the report window. All the analyses in the report will be redone but this time using all 40 runs of the process.)

Notice that all the capability metrics are < 1, indicating the process variation (as measured by +/- 3s) is extending beyond the specification limits. The estimated Expected Overall PPM outside of specification is now 5,594 (or .56%).

Summary

Statistical Insights

Understanding process variation is best done when data is collected in a way that accounts for different components of variability. In this case study, the sampling strategy was to record *Etch Rate* on 4 separate locations in each of 5 wafers within each of 40 runs of the etching tool. This resulted in 4 x 5 x 40 = 800 *Etch Rate* values for analysis. After 20 runs, the equipment was recalibrated. This allows for the examination of 4 sources of variation: 1) within wafer ; 2) between wafer within run ; 3a) between runs within calibration stage ; 3b) between recalibration stages.

Graphical displays showed these different levels of variation relative to the process target of 620 and the lower and upper specifications of 545 and 695 (i.e., 620 +/- 75). Control charts enabled closer analysis of the variation to determine if the process was in statistical control (i.e., exhibiting consistent variation), and if it was not, where the variation was changing. Process Capability Analysis is a statistical method that produces a variety of metrics that quantifies how well the process is performing relative to specification.

From our analyses, we saw that *Etch Rate* variation increased over time following recalibration. However, it didn't appear that the average *Etch Rate* within a run and/or within a wafer also changed over time. Capability analyses performed on the first 10 runs after calibration gave visibility into how the process might perform if recalibrations happened more frequently.

Implications and Further Study

A thorough investigation of this process would be needed to help discover which factors might be causing excess variation. Specifically, why is the between-etching variation smallest after recalibration of the equipment, and what factors could be causing that variation to increase as the runs progress? Removing those causes could result in a reduction in overall process variation, potentially to only 113-114 out of a million etchings to be outside the specifications.

These process investigation studies can also reveal factors that are potentially impacting the other levels of variation. Being able to eliminate factors that are causing variation in *Etch Rate* between etchings, whether within wafers or between wafers, could result in a major improvement to the capability of this process.

Exercises

Use the data in file **variation-in-semiconductor-etching-process-lines-b-g.jmp** to complete the exercises.

1. Analyze the data collected from the six other Etching Lines (B, C, D, E, F, and G). For each one:

    a) Describe the performance of the process, both in terms of whether it is in "statistical control" and its performance relative to specification ("capability").

    b) In addition to describing the results in statistical/engineering language, also do so in a way that executives and/or those without much statistical background would understand.

    c) Examine the different levels of variation that can be analyzed with these data. Show specific graphs with commentary that clearly communicates the amount of variation at each level. If there are differences, where is the variation the least/greatest?

    d) Do you believe the performance observed in this sample would generalize to this process beyond this sampling period?

2. What would need to occur (statistically) to improve the performance of these processes?

3. What ideas do you have for additional data that could be collected or studies that could be done to help in the effort of getting these processes ready for production?