# Detecting measurement outliers - Remeasure efficiently

Albrecht Ullrich

Advanced Mask Technology Center GmbH & Co KG, Rähnitzer Allee 9, 01109 Dresden, Germany

## ABSTRACT

Shrinking structures, advanced optical proximity correction (OPC) and complex measurement strategies continually challenge critical dimension (CD) metrology tools and recipe creation processes. One important quality ensuring task is the control of measurement outlier behavior. Outliers could trigger false positive alarm for specification violations impacting cycle time or potentially yield. Constant high level of outliers not only deteriorates cycle time but also puts unnecessary stress on tool operators leading eventually to human errors.

At tool level the sources of outliers are natural variations (e.g. beam current etc.), drifts, contrast conditions, focus determination or pattern recognition issues, etc. Some of these can result from suboptimal or even wrong recipe settings, like focus position or measurement box size. Such outliers, created by an automatic recipe creation process faced with more complicated structures, would manifest itself rather as systematic variation of measurements than the one caused by 'pure' tool variation.

I analyzed several statistical methods to detect outliers. These range from classical outlier tests for extrema, robust metrics like interquartile range (IQR) to methods evaluating the distribution of different populations of measurement sites, like the Cochran test. The latter suits especially the detection of systematic effects. The next level of outlier detection entwines additional information about the mask and the manufacturing process with the measurement results. The methods were reviewed for measured variations assumed to be normally distributed with zero mean but also for the presence of a statistically significant spatial process signature.

I arrive at the conclusion that intelligent outlier detection can influence the efficiency and cycle time of CD metrology greatly. In combination with process information like target, typical platform variation and signature, one can tailor the detection to the needs of the photomask at hand. By monitoring the outlier behavior carefully, weaknesses of the automatic recipe creation process can be spotted.

**Keywords:** CD measurement, remeasurement, outlier, cycle time, CD-SEM

## 1. INTRODUCTION

The goal for metrology as a quality-ensuring department is the delivery of valid and consistent data of the photomask. Based on this data the customer decides about the fate of the mask, and internally engineers monitor and control their processes. Invalid data can therefore be quite harmful, be it the return of the mask by the customer or rendering the experiment of the process engineer invalid. In any case, the results could be costly.

Although the integrity of data is paramount, its delivery cannot be achieved by all means, e.g. measuring the reticle five times can certainly increase the confidence in the results but the cycle time impact is unacceptable for production. So the metrology process not only has to detect, quantify, and correct outliers, but also do it *efficiently*. Hence statistical tools should be applied.[1] Further on it is advisable to monitor the outlier behavior to check if the tools work properly and assumptions are still valid, which will be demonstrated by an example.

The paper surveys some useful outlier detection methods and explores ways to employ them. Classical methods to test extrema are considered. Robust metrics are also discussed. It is demonstrated how to lower cycle time with the help of an outlier monitor. The main focus will be methods accounting for the distinctive features of CD uniformity of photomask, because one can usually consider measurement sites with the same nominal target that simplifies the statistical analysis.

## 2. OUTLIER DETECTION

Detecting outliers is an old problem of statistical analysis of data. Especially for small sample sizes outliers can skew sample average or standard deviation, i.e. non-robust estimates. Typically repeated measurements of one variable are analyzed. The variation of the data is usually assumed to be normally distributed. Suspicious values are then tested for the hypothesis that the value is an outlier. With a given confidence level the hypothesis can be falsified or has to be accepted.

### 2.1 Test of the Extremum

With given data sample, one wants to test the hypothesis that the maximum (or minimum, or both) value is an outlier. Dixon[2] surveyed concepts of outlier analysis and categorized the methods in

- chi-squared score of data
- score of extreme value
- ratio of range to standard deviation
- ratio of variances without suspicious values and with them
- ratio of ranges and subranges.

Dixon preferred the last concept due to performance and build his test method onto this. For the test several coefficients are defined, based on the ordered sample $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ and compared to critical parameters. The parameters for the Dixon test statistic were refined later on.[3,4]

Parallel to Dixon, Grubbs[5] proposed his test for three cases:

- One outlier

$$G = \frac{|x_{outlying} - \bar{x}|}{s}, \quad U = \frac{S_1^2}{S^2} \tag{1}$$

- Two outliers on opposite tails

$$G = \frac{x_{(n)} - x_{(1)}}{s}, \quad U = \frac{S_2^2}{S^2} \tag{2}$$

- Two outliers on same tail

$$G = U = \frac{S_2^2}{S^2} \tag{3}$$

where $\bar{x}$ is the sample mean, $s^2$ is the standard variance, $S^2$ is the sample variance, $S_m^2$ is the sample variance excluding the $m$ values under suspicion, and $n$ is the sample size. The critical value for one outlier of the $G$ distribution can be approximated by

$$t_{\alpha/n,n-2}\sqrt{\frac{n-1}{n-2+t^2_{\alpha/n,n-2}}} \tag{4}$$

where $t$ denotes the quantile of the t-distribution and $\alpha$ is the confidence level as discussed in earlier work.[6] In general critical values for $G$ and $U$ are simulated and have to be interpolated from tabularized data.

## 2.2 Test variance in groups of data

Besides single values of one data set one can also test an erroneous group of values against other groups, e.g. compare a CD feature to other features measured on the same mask. The Cochran test[7] compares the group with the maximum variance against the sum of variances of all groups. The statistic is calculated as follows

$$C = \frac{S_{max}^2}{\sum_i S_i^2} \tag{5}$$

where $S_i^2$ is the sample variance of the group $i$, and $S_{max}^2$ is the sample variance of the group with the highest variance. The critical value for this statistic is approximated by

$$\frac{1}{1 + (k-1)F_{\alpha/k,n(k-1),n}} \tag{6}$$

where $k$ is the number of groups, $n$ is the mean number of observations per group, and $F$ denotes the quantile of the F-distribution.[8] A similar method is the Friedman rank sum test.[9]

## 2.3 Robust estimates

As mentioned at the beginning of the section for the analysis of outliers, robustness of estimators play an important role. If there are outliers present in data it is expected that aggregates like the median or the median absolute deviation (MAD) are more reliable than the sample mean or standard deviation. Based on these robust metrics one can construct ranges for given confidence levels where values beyond the range are suspicious. Two simple ranges are given by metrics based on

- Interquartile Range (IQR)

  $$[Q_{0.25} - n_\alpha IQR,\, Q_{0.75} + n_\alpha IQR] \tag{7}$$

  where $Q_x$ is the $x$th quantile, $IQR = Q_{0.75} - Q_{0.25}$, and $n_\alpha$ a factor depending on the aspired confidence level $\alpha$. This method is usually employed to characterize outliers in boxplots.

- Median absolute deviation (MAD)

  $$[Q_{0.5} - n_\alpha MAD,\, Q_{0.5} + n_\alpha MAD] \tag{8}$$

  where $Q_{0.5}$ is the median, and $MAD = 1.4826\, median(|x_i - Q_{0.5}|)$ is the normalized median of the absolute deviations from the median, where the normalization factor ensures consistency, i.e. $E[MAD] = \sigma$ for normally distributed samples with standard variation $\sigma$ for large sample size. $n_\alpha$ is a factor depending on the aspired confidence level $\alpha$.

## 2.4 Spatial data analysis

There are also elaborated methods to account for data samples distributed in space, basically estimating the spatial distribution like Kriging. These techniques are not considered here due to their complexity. The interested reader is referred to dedicated literature.[10,11] In an automated environment it proved too difficult to setup an outlier detection based on spatial methods due to either noise or statistic relevance, i.e. the confidence interval of the spatial estimation compared to the noise was not stable enough. The implementation would have caused too much false alarms.

# 3. APPLICATION

In this section the application of above mentioned methods of outlier detection is considered for CD measurement with the focus on uniformity. The techniques are evaluated for their efficiency.

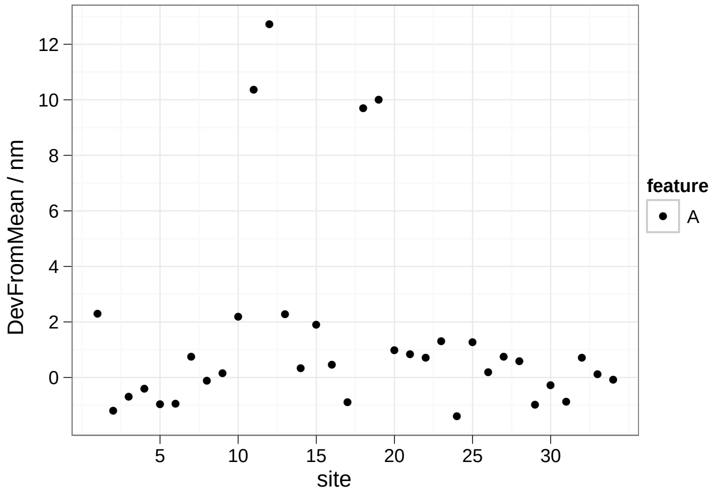


Figure 1. CD measurement of feature A sampled at 34 sites with four suspicious values. Plotted is the deviation from mean.

### 3.1 Outlier detection in CD measurements

Let us consider a CD measurement of a feature to illustrate the outlier detection. In figure 1 a measurement of a feature with suspicious values is shown.

Applying the methods of section 2.1 for extrema one first realizes that in order to determine all outliers a recursion is necessary. Because of the amount of suspicious data the analysis not only depends on the chosen confidence level but also on the distribution of the suspicious values. On the other hand, if one uses the robust methods of section 2.3, like depicted in the boxplot of figure 2, the suspicious data is readily identified.

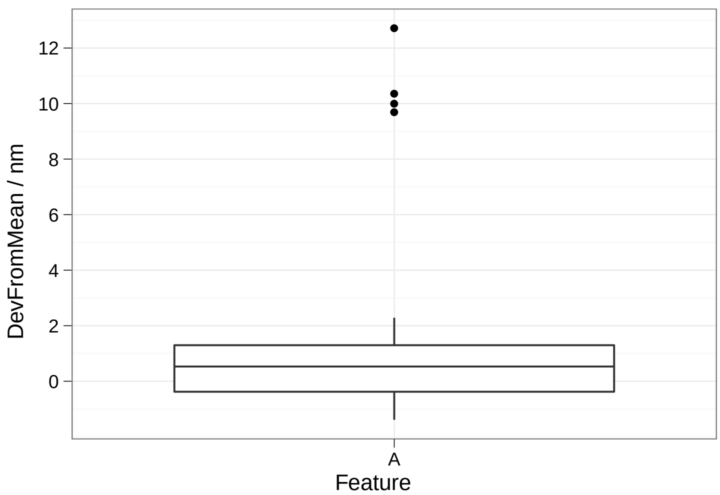


Figure 2. Boxplot of data from figure 1. Suspicious data is ‘marked’ by single point plot style.

For this reason, we use robust methods in production. Analyzing the data so far validates the assumption that the robust methods usually include the case when the methods for the extrema would trigger an outlier. What is left is the choice for the factor $n_\alpha$ for the confidence level. It should be sensitive enough to reasonably detect outliers but not trigger too much false positive alarms, lowering the efficiency by unnecessarily remeasuring sites. Practically the factor is set and controlled by monitoring the triggered sites and the change due to remeasuring. It is influenced by the variation of the measurement process but also by the typical patterning process variations.

In production we also use the Cochran test of section 2.2 in order to detect suspicious features. Consider the measurement of several features as depicted in figure 3. In the boxplot of figure 4 one again sees single suspicious data values but also the broader distribution of feature D.

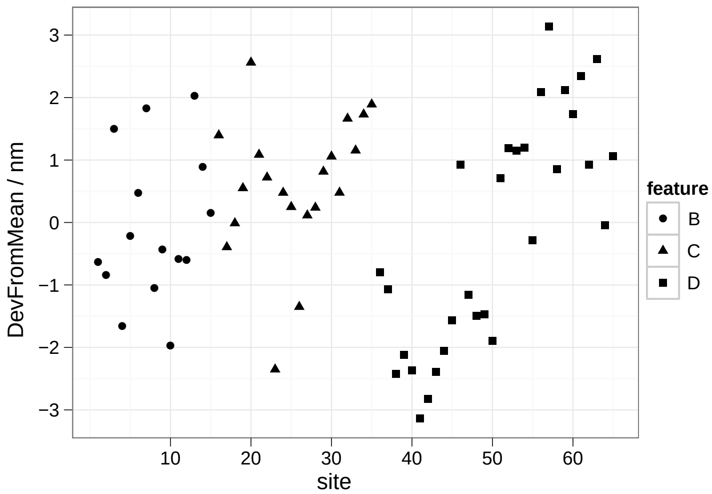


Figure 3. CD measurement of three features on the same mask.

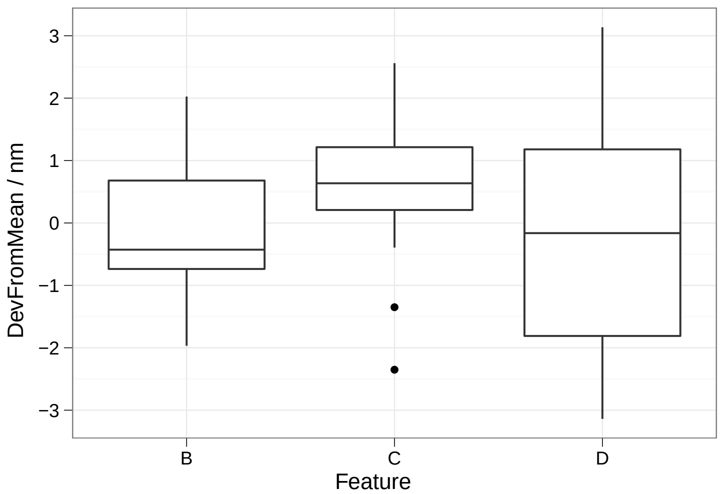


Figure 4. Boxplot of data from figure 3 highlighting the broader distribution of feature D.

## 3.2 Analyzing outlier behavior

It is presumed that most of the variation of the tool stems from the image formation[12] and influences uniformity randomly, i.e. normally distributed with zero mean, and can be described accordingly.[13] Systematic outliers are at the center of the discussion. For this purpose consider again the dataset of figure 3. In the density plot shown in figure 5 a distinct distribution of feature D compared to the other features is recognizable. In practice this

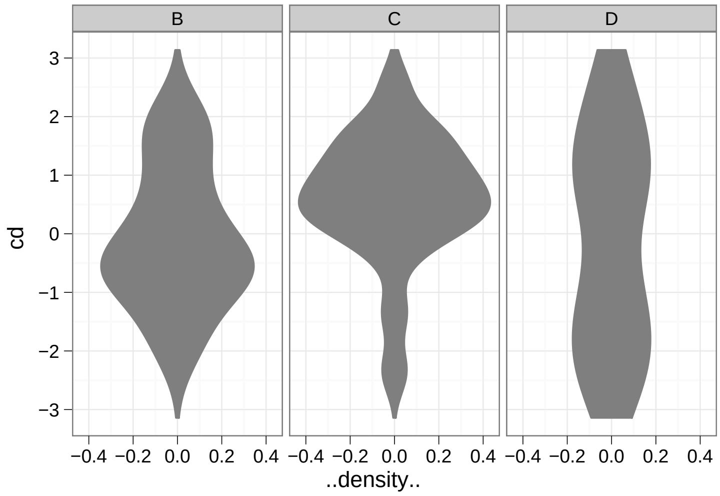


Figure 5. Density plot of data from figure 3 showing a pronounced bimodal distribution of feature D and verifying the slightly asymmetric distribution of feature A and B already visible in the boxplot in figure 4.

can be due to several reasons. It could be that it is a dense feature measured with recipe with a bad pattern recognition. If for example the pattern recognition is not unique, then the CD-SEM would randomly pick the wrong contrast, i.e. we see a line-space mix-up.

Another reason could be the spatial sampling of the feature. If the typical spatial process signature has for instance a bowl-like shape, and the feature is measured in the center of the mask and at one corner, you would expect such a behavior.

Yet another reason could be that some sites of the feature got a slightly different OPC than the other sites. As a result the nominal CD at measurement positions could extend over different sizes the automatic recipe creation process picks a smaller measurement box than usual, in order to accommodate a smaller target size in non-scan direction.

Depending on the true reason very different action would be taken. In order to account for spatial process variations one could adapt the factors for the confidence levels, or provide the evaluation algorithm with an average signature of the process (from our perspective quite complicated but doable), or try to estimate the current signature in-situ with the measurement sites at hand (even more complicated). In this case the culprit was a line-space mix-up that demanded remeasurement. The features A and B were analyzed too, because of their asymmetric density around zero (see figure 5), which manifests itself also in the spread of the sample average (zero) and the median (see figure 4). In contrast to feature C no faults by metrology were determined, and the variation was still within reasonable limits. These examples demonstrate the importance of monitoring and analyzing the outlier behavior to act accordingly.

Also outliers can raise questions further on in the production line, e.g. line monitoring or even at customer, so if the outlier is actually real, it is good to be prepared for this questions, and have the outliers already quantified at metrology.

### 3.3 Monitoring outlier behavior

One of the most simple monitor of outlier behavior is the remeasure rate, possibly separated by the process of records (POR). Significant changes can indicate several issues, as discussed in the section above. An example is shown in figure 6. In the time frame shown, one can see an increase in the rate at the end of June. This increased rate impacted the cycle time, given that the recipes had typically several hundred measurements sites. A close

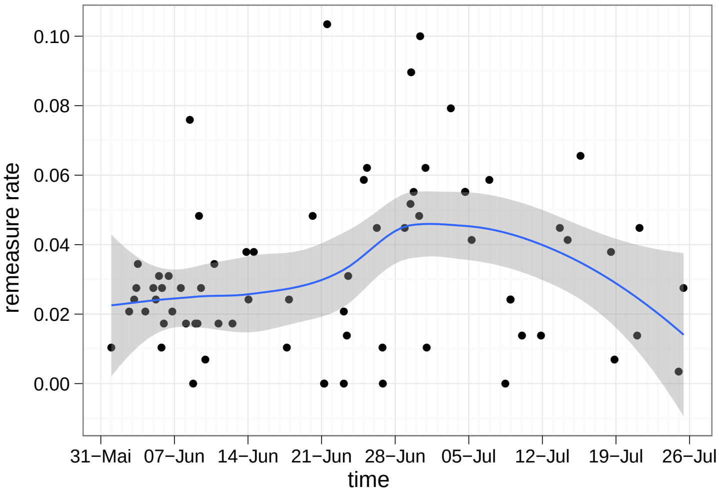


Figure 6. Extract of a monitor of the remeasure rate of a certain POR over time. Acting accordingly on the increased rate at the end of June brought back the rate to normal, as indicated by the smoothing average.

look at the measurements revealed that newly introduced products needed a slightly different setup. After fixing the setup the remeasurement rate improved and reached the historic level again. With the lower rate the cycle time went down accordingly.

## 4. SUMMARY

Automatic outlier detections for CD measurement data was implemented to trigger efficient remeasurements. Detections based on robust measures, i.e. based on MAD, seemed more advantageous over tests of extrema, and therefore were chosen. The signature of the mask or the signature of the POR was not considered, due to too much noise or marginal statistic relevance. It proved to be more useful to adapt the limits of the robust measures to the POR, e.g. iLine versus 50keV.

Besides giving the operator clear criteria for the remeasurement of single sites, the methods assist with qualifying and quantifying outliers. Especially the Cochran turned out to be beneficial due to its sensitivity for systematic issues. By monitoring the outlier behavior and in-depth analyzing suspicious features, valuable insights can be gained. It helps to continually adapt and improve the automatic recipe creation process or highlight issues with the measurement strategy applied to the given design or the mask processes at hand. With these measures, as demonstrated with the remeasure rate, the cycle time of the CD process was improved along with the quality of the data. At the end the CD process is more efficient in a complex automated factory setup.

## ACKNOWLEDGMENTS

AMTC is a joint venture of Globalfoundries and Toppan Photomasks, and gratefully acknowledges the finacial support by the German Federal Ministry of Education and Research BMBF under contruct numbers 13N9675 (ANTARES) and 13N9946 (CDuR32).

The author is thankful for the R project[14] providing an extensive environment for statistical computing under the GNU General Public License.

## REFERENCES

[1] Lu, J.-C., Jeng, S.-L., and Wang, K., "A review of statistical methods for quality improvement and control in nanotechnology," *Journal of Quality Technology* **41**, 148–164 (April 2009).

[2] Dixon, W. J., "Analysis of extreme values," *Annals of Mathematical Statistics* **21**(4), 488–506 (1950).

[3] Dixon, W. J., "Rations involving extreme values," *Annals of Mathematical Statistics* **22**(1), 68–78 (1951).

[4] Rorabacher, D. B., "Statistical treatment of rejection of deviant values: Critical values of dixon q parameter and related subrange ratios at the 95 percent confidence level," *Analytical Chemistry* **83**(2), 139–146 (1991).

[5] Grubbs, F. E., "Sample criteria for testing outlying observations," *Annals of Mathematical Statistics* **21**(1), 27–58 (1950).

[6] Pearson, E. S. and Chekar, C. C., "The efficiency of statistical tools and a criterion for the rejection of outlying observations," *Biometrica* **28**(3), 308–320 (1936).

[7] Snedecor, G. and Cochran, W. G., [*Statistical Methods*], Iowa State University Press (1980).

[8] Abramowitz, M. and Stegun, I. A., [*Handbook of Mathematical Functions*], Dover Publications (1965).

[9] Hollander, M. and Wolfe, D. A., [*Nonparametric Statistical Methods*], John Wiley & Sons (1973).

[10] Henning, R., [*Spatial Data Analysis: Theory and practice*], Cambridge Unversity Press (2003).

[11] Bivand, R. S., Pebesma, E. J., and Gomez-Rubio, V., [*Applied Spatial Data Analysis with R*], Springer Science+Business Media, LLC (2008).

[12] Reimer, L., [*Image Formation in Low-Voltage Scanning Electron Microscopy*], vol. TT 12 of *SPIE Tutorial Texts Series*, SPIE Optical Engineering Press (1993).

[13] Fuller, W. A., [*Measurement Error Models*], Wiley series in probability and mathematical statistics, John Wiley & Sons (1987).

[14] R Development Core Team, *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria (2010). ISBN 3-900051-07-0.