



Outliers detection by fuzzy classification method for model building

Mame Kouna Top, Yorick Trouiller, Vincent Farys, David Fuard, Emek Yesilada, Catherine Martinelli, Said Mazen, Frank Foussadier, Patrick Schiavone

► To cite this version:

Mame Kouna Top, Yorick Trouiller, Vincent Farys, David Fuard, Emek Yesilada, et al.. Outliers detection by fuzzy classification method for model building. Metrology, Inspection, and Process Control for Microlithography XXIII, Feb 2009, San Jose (CA), United States. in press, 10.1117/12.812955 . hal-00374511

HAL Id: hal-00374511

<https://hal.science/hal-00374511>

Submitted on 9 Apr 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Outliers Detection by Fuzzy Classification Method for Model Building

Mame Kouna Top^{a/b}, Yorick Trouiller^c, Vincent Farys^a, David Fuard^b, Emek Yesilada^a, Catherine Martinelli^a, Mazen Said^a, Franck Foussadier^a, Patrick Schiavone^{b/d}

^a STMicroelectronics 850 rue Jean Monnet, 38926 Crolles Cedex, France

^b LTM-CNRS, 17 Rue des Martyrs, 38054 Grenoble Cedex 9, France

^c CEA-Leti, 17 Rue des Martyrs, 38054 Grenoble Cedex 9, France

^d UMI 2958 CNRS-Georgiatech, Georgia Institute of Technology, Atlanta GA 30332

ABSTRACT

Optical Proximity Correction (OPC) is used in lithography to increase the achievable resolution and pattern transfer fidelity for IC manufacturing. Nowadays, immersion lithography scanners are reaching the limits of optical resolution leading to more and more constraints on OPC models in terms of simulation reliability. The detection of outliers coming from SEM measurements is key in OPC [1]. Indeed, the model reliability is based in a large part on those measurements accuracy and reliability as they belong to the set of data used to calibrate the model. Many approaches were developed for outlier detection by studying the data and their residual errors, using linear or nonlinear regression and standard deviation as a metric [8].

In this paper, we will present a statistical approach for detection of outlier measurements. This approach consists of scanning Critical Dimension (CD) measurements by process conditions using a statistical method based on fuzzy C-Mean clustering and the used of a covariant distance for checking aberrant values cluster by cluster. We propose to use the Mahalanobis distance [2] in order to improve the discrimination of the outliers when quantifying the similarity within each cluster of the data set.

This fuzzy classification method was applied on the SEM CD data collected for the Active layer of a 65 nm half pitch technology. The measurements were acquired through a process window of 25 (dose, defocus) conditions. We were able to detect automatically 15 potential outliers in a data distribution as large as 1500 different CD measurement. We will discuss about these results as well as the advantages and drawbacks of this technique as automatic outliers detection for large data distribution cleaning.

Keywords: Modeling, OPC, Fuzzy C mean, Classification, Outliers detection.

1. INTRODUCTION

In semiconductor manufacturing in order to reduce the errors of the photolithography step, corrections are usually made to the mask features. OPC (Optical Proximity Correction) aims at providing the optimal correction to the mask layout in order to improve pattern fidelity or process latitude. The OPC algorithm uses a model calibrated with test structures of different type (CD, space, 1D and 2D structures). Metrology reliability is an important factor affecting OPC calibration. Many errors can come from the CDSEM, for example the SEM recipe which can not be adapted to all the measured structures. Other types of errors can even be mistakes in the classification of the structure types which leads to apparently aberrant measurements. The inclusion of a small number of outliers within a set of measurements can degrade significantly a model calibration quality. That is why measurements should be thoroughly checked before being used in model calibration. However, in the manufacturing context, thousands of experimental data are used for the model calibration, and checking these measurements manually takes too much time and become almost impossible in practice. Generally, outlier detection methods compare two data distribution like experimental data and simulated data using regression methods. Outliers are checked by the study of errors between the two distributions [9]. It turns out that very often, using the conventional outlier detection techniques, the values detected are almost always the smallest or largest values of the data distribution [6]. An original approach for outlier detection on image processing consists of using Fuzzy C-means clustering [7]. This method has also been used with Support Vector Machine (SVM) for aberrant values identification [5].

The method we propose in this paper checks the outliers discriminated by Mahalanobis distance in each cluster preliminarily defined using a Fuzzy C-Means strategy; Similarly it explores the whole sample set and detects aberrant values. The paper is organized as follows: in Section 2 the Fuzzy classification method is defined as well as the Mahalanobis distance. The performance of the method will be illustrated in the Section 3 with a real case study of OPC calibration data set.

2. DESCRIPTION OF THE OUTLIER DETECTION STRATEGY

Our method combines the fuzzy classification method and the Mahalanobis distance. Its advantage is that it does not systematically index only extreme values as outliers as most of the classical methods do. Using the fuzzy classification method, we initially extract coherent groups from the global distribution. In a second step, within each of these groups the data seemingly farthest from the other values are considered as outliers.

To explain this method we consider the data distribution, $X = \{x_1, x_2, \dots, x_n\}$ represented by the histogram of Fig.1.

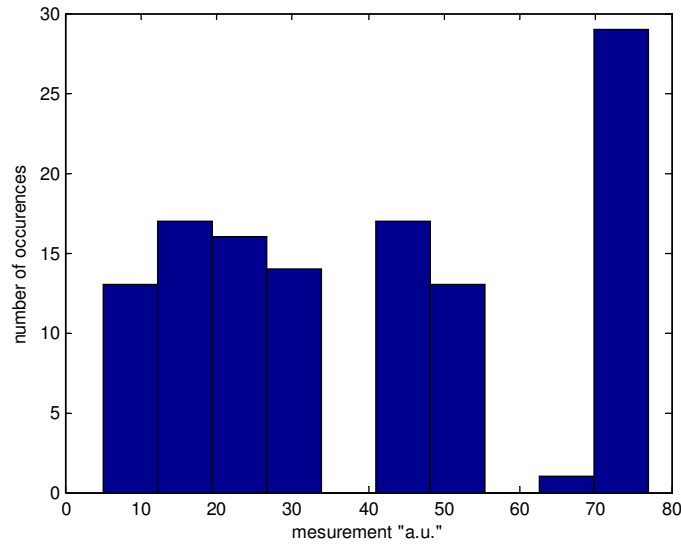


Fig.1. Histogram of the X data distribution; On abscissa we have the data values on an arbitrary scale, and on the y axis we have the number of occurrences found for a range of values

The number of different coherent groups is not set a priori. In order to find the optimal number of coherent groups by the Fuzzy C-means method, we first initialize the maximal number of cluster ($N=10$ for example). Then we search the optimal number of subgroups of X where, the distance $|x_k - v_i|^2$ between the elements of the same group is minimal and the distance $|v_i - \bar{x}|^2$ between groups is maximal (1).

$$J_m(P) = \sum_{k=1}^n \sum_{i=1}^c [A_i(x_k)]^m \left(|x_k - v_i|^2 - |v_i - \bar{x}|^2 \right) \quad \text{with} \quad v_i = \frac{\sum_{k=1}^n [A_i(x_k)]^m x_k}{\sum_{k=1}^n [A_i(x_k)]^m}, \forall i \in N_c \quad (1)$$

Where, $V = \{v_1, v_2, \dots, v_c\}$ is the vector of cluster centers, m is the degree of fuzzyfication, $P = \{A_1, A_2, \dots, A_c\}$ is the fuzzy membership matrix and $A_i(x_k)$ is the probability that x_k belongs to the group i with the following constraints (2).

$$\sum_{i=1}^c A_i(x_k) = 1, \forall k \in N_n, A_i(x_k) \in [0,1] \quad \text{and} \quad 0 < \sum_{k=1}^n A_i(x_k) < n, \forall i \in N_c \quad (2)$$

The minimization of the objective function described in equation (1) gives then the optimal number of cluster.

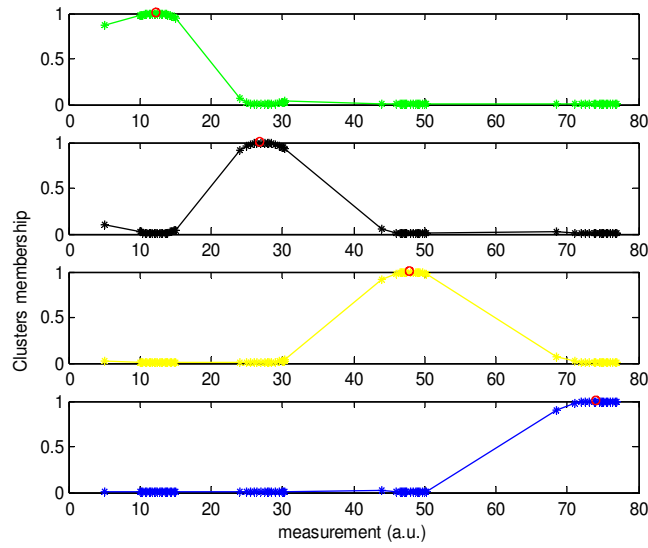


Fig.2. Fuzzy C-Means clustering

We applied the Fuzzy C-Mean clustering to the above mentioned distribution X . After the clustering, we obtain four optimal homogeneous groups. The Fig.2 is a representation of the four clusters we obtained. On the x axis we have the distribution data on an arbitrary unit, and on the y axis we have the membership value of measurements by cluster. The magenta circle in each subplot shows the clusters center. The elements of the first cluster are those on the green cluster which are over the threshold membership value of 0.5.

The method is implemented using Matlab. The $A_i(x_k)$ and v_i parameters were obtained by the fuzzy C-Mean clustering Matlab function.

For each of these groups the goal is now to detect the outliers. An outlier is defined as a value in a distribution which appears to deviate significantly in comparison to all other members of the distribution. In most of the cases, a valid observation (or measurement) is rarely more than a few standard deviations away from the mean. Tchebyshev inequality provides us with a quantification of the probability that a given data belongs to the distribution. It estimates that in general, at least $(1-1/k^2)*100\%$ of the values are within k standard deviation from the mean [6]. This criterion can be applied to normal or non normal distribution.

We chose to estimate the standard deviation of the distribution data from the cluster center rather than from the mean. With fuzzy c-mean, the cluster center is the mean of all points weighted by their cluster membership value. The cluster center is then less biased by extreme values than the mean.

$$|X_k - v_i| \geq 3\sigma \quad (3)$$

If the distance between the test point and the center of the cluster is less than three times the standard deviation, we conclude that it is probable at 89% (Tchebyshev's inequality) that the test point belongs to the set. The further away it is, the more likely the test point should not be classified as belonging to the set and be considered as an outlier. We applied equation (3) on the four clusters we obtained by fuzzy C-Mean clustering. The Fig.3 shows the detected outlier in red.

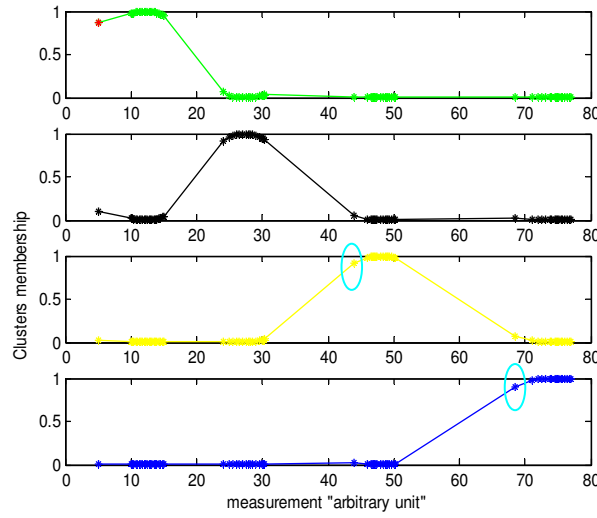


Fig.3. Outliers detection without Mahalanobis distance.

On this figure the point in red indicates the single detected outlier. However we remark two measurements circled in blue on the yellow and blue cluster. Those values are far from the other values of their cluster, but not enough to be detected by the three sigma limit on equation (3). In order to improve the sensitivity of the detection, we apply the Mahalanobis distance to each group in order to better spot the aberrant values.

The Mahalanobis distance takes into account the correlation of the data set and is scale-invariant [3][4]. The Mahalanobis distance of an element to the center of the distribution is determined using the covariance matrix of the studied cluster. It is defined by:

$$d_M(x_k) = \sqrt{(x_k - v_i)^T M (x_k - v_i)} \quad (4)$$

Where M is the semi definite covariance matrix and v_i is the center of the cluster containing the element x_k . We use the Mahalanobis distance, to study the similarity between the elements of a same distribution data in order to detect the farthest elements of the distribution.

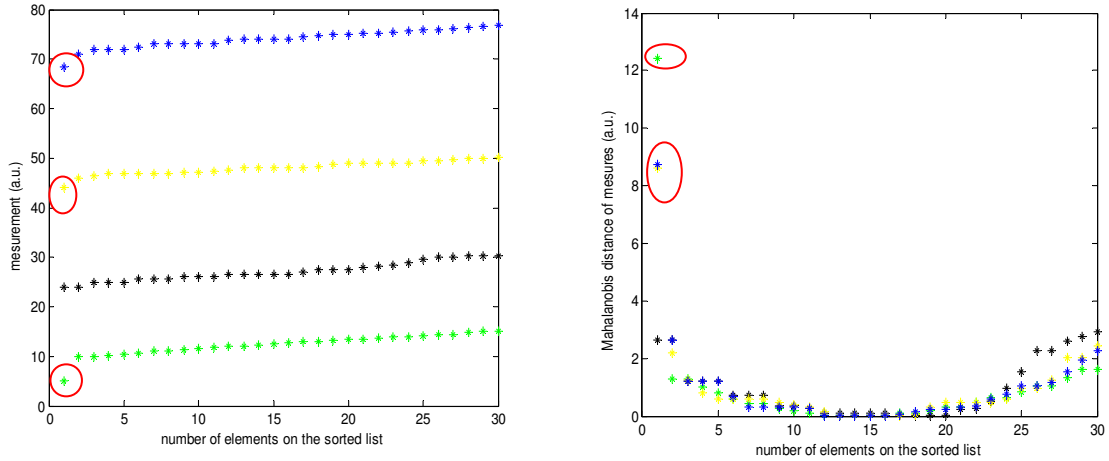


Fig.4. X data distribution after fuzzy C-means clustering (left). Mahalanobis distance applied to the same distribution (right). Mahalanobis distance accentuates distance between outliers and other values

On Fig.4, we represent the X data distribution shown in Fig.1. On the figure on the left, the data are sorted after the application of the fuzzy C-means clustering which came out by sorting the data into the four clusters. On the Fig.4, on

the right we represent the same data after applying the Mahalanobis distance. We can see that the Mahalanobis distance enhances the discrimination of the outlier data (red circle on the figure) with to the other ones. If we apply Tchebyshev's cleaning to the distribution using the Mahalanobis distance we can detect additional outliers that were not detected with the regular distance. This can be observed on Figure 5

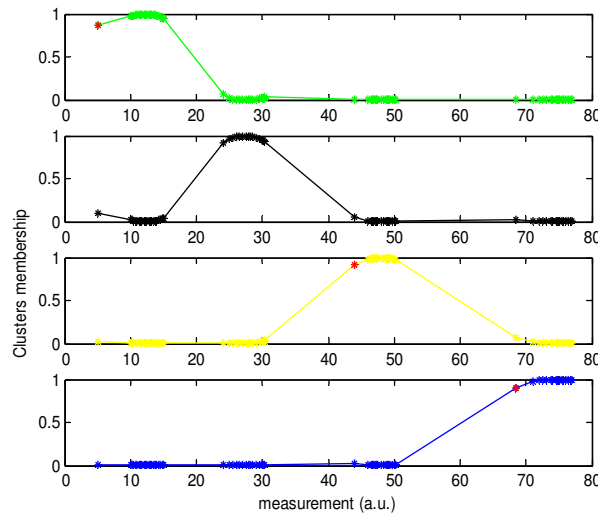


Fig.5. Outlier detected using the Mahalanobis distance.

On the Fig.5, only one outlier was detected without Mahalanobis distance. With the Mahalanobis distance we can remark that two more outliers can be detected.

	Cluster1	Cluster2	Cluster3	Cluster4
Cluster center (CC)	12,39	27,09	48,06	74,1
Standard deviation (σ)	2,0861	1,92	1,55	1,88
Outlier value (Out)	5	—	44	68,5
Euclidian distance	$ 5 - 12,39 = 3,54 \sigma$	—	$ 44 - 48,06 = 2,6194 \sigma$	$ 68,5 - 74,1 = 2,9787 \sigma$
$\text{Out} - \text{CC} = N\sigma$	detected	—	Non detected	Non detected
Mahalanobis distance	$d_m(5;12,39) = 5,1 \sigma$	—	$d_m(44;48,06) = 5,346 \sigma$	$d_m(68,5;74,1) = 4,4 \sigma$
$d_m(\text{Out}; \text{CC}) = N\sigma$	detected	—	detected	detected

Table 1. Outliers detection with and without Mahalanobis distance (values in arbitrary units).

The Table.1 summarizes for each cluster: the cluster center (CC), the standard deviation (σ) of the distribution and the potential outliers (Out). The two bottom lines of the table indicate respectively the Euclidian distance and the Mahalanobis distance for each of these outliers. As described in equation (3), we consider the value as an outlier when the calculated distance is higher than 3σ according to Tchebyshev inequality. Unlike Euclidian distance for example where all elements of a distribution are similarly treated, the Mahalanobis distance grants a less important weight to the most distant elements. This is the great interest of this correlation distance. The comparison of the two last row of the Table.1, show how Mahalanobis distance enhances discrimination of the distribution isolated values. We can see how distance between the supposed outlier and his cluster center passed from $2,6 \sigma$ to $5,3 \sigma$ on cluster 3 for example.

Additionnally the outlier detection method does not consider only the extreme values as outliers but performs a search in the distribution to detect non coherent values.

3. TEST ON MODEL CALIBRATION DISTRIBUTION DATA

3.1. Outliers detection method application

Measurements used in OPC calibration are specific to the kind of products and features that the user needs to model. The whole set of features is printed with different process conditions in order to model the influence of the dose and focus variations. On Fig.6 we have represented the different type of structure that we used for our study: “pitch”, “space”, “line end”, “H or inverse line-end” and “pad variation”.

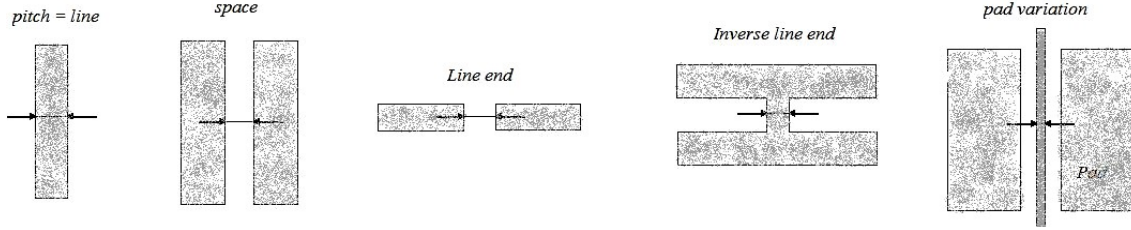


Fig.6. Type of structures used for the study

We apply our outlier detection method on real OPC calibration data of the active layer of a 65 nm half pitch technology with 1500 test structures. First we sort the different features by type. This is aimed at getting similar data type for the outlier detection method application. This step is key for the method; The feature type has to be clearly defined in order to compare “apples with apples”.

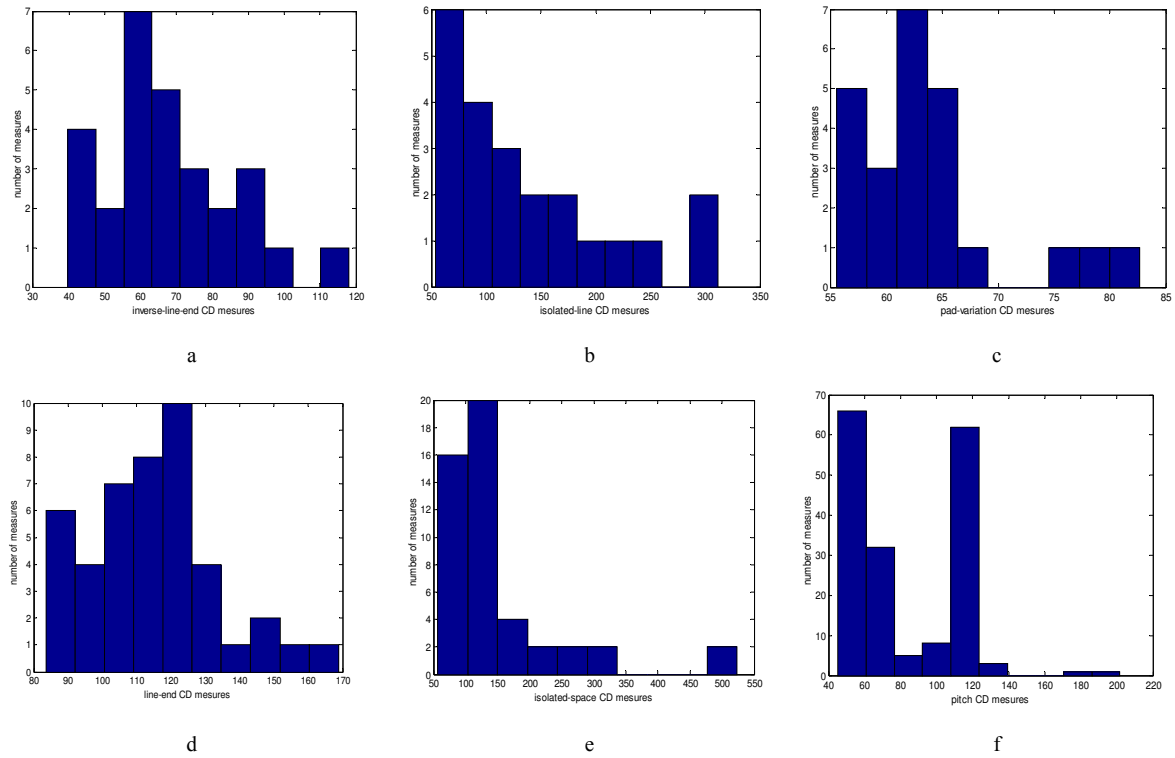


Fig.7. Histograms of the measurements sorted by structure type for one process condition; a) Inverse line-end; b) Isolated line; c) Pad variation; d) line end; e) isolated space; f) pitch or line.

The Fig.7 shows the histogram of critical dimension measurement of 344 structures for one process condition. We apply then our outlier detection method as shown on Fig. 8 where we can see the clustering for each of the 6 structure types as well as the outlier detected by Mahalanobis in red.

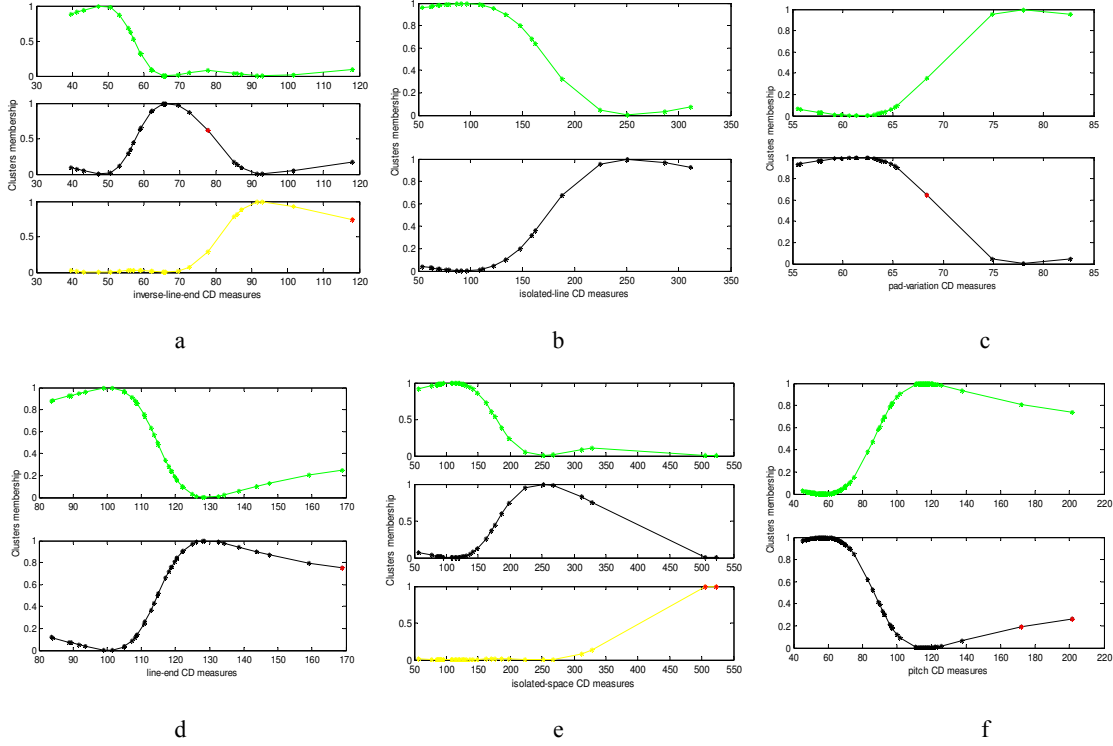


Fig.8. Outliers detection for structure type by cluster; a) Inverse line-end with 3 cluster and 2 outliers detected; b) Isolated line with 2 cluster and no outliers detected; c) Pad variation with 2 cluster and 1 outlier detected; d) line end with 2 clusters and 1 outlier detected; e) isolated space with 3 clusters and 2 outliers detected; f) pitch or line with 2 clusters and 2 outliers detected.

For this Process Condition we detected 8 outlier values. Only two out of them are true outliers: one “dense line” structure and the other one detected was a “pad variation” structure. The 6 others values were extremes “line end” and “dense line” measurement, the Fig.8c show an example of this kind of error.

For the 1500 measurements on the different features of the calibration set, we detected 55 outliers among which 15 are true outliers. This is a detection ratio of true outlier of 27 %. The other 40 detected points were extreme measurements that are indeed correct measurements, not real outliers. The Fig. 9 is an example of 3 typical measurement detected with our method. One is a real outlier, second one is dependant of the process and third one is an extreme value.

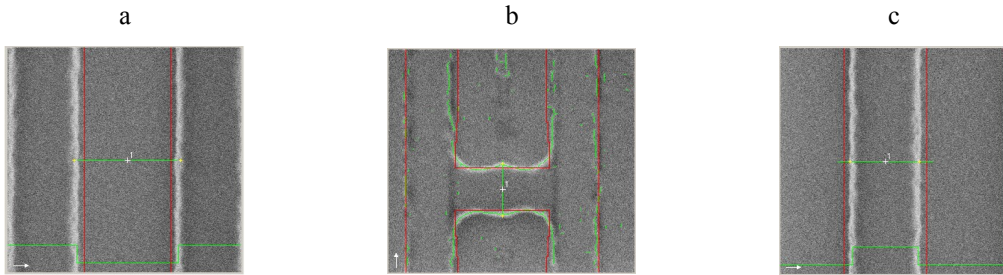


Fig.9: Different types of potential outliers detected; a) SEM measurement error; b) ”H” structures with wiggles; c) extreme line CD measurement

In this example, a typical error leading to outlier is the measurement of a space instead of the line as shown on Fig.9.a. We met this case on “dense line” and “pad variation” structure. On the opposite we have line measurement instead of space, on “space” and “line end” structures. The “H” structure presented on Fig.9.b has been detected as an outlier with our method due to the wiggle effect that modify the CD value at the center of the structure regarding to the “H” width. This CD variation appears specifically when applying Quasar illumination. Apparently this behavior is very important for modeling and this type of measurement need to not be considered as an outlier. Finally on Fig.9.c we have represented an isolated line CD measurement which has been detected with our method due to the fact that there is few measurement and far from the other value of the data set.

In order to avoid considering those extreme values as outliers while they are not, we tried to improve our detection strategy. For that purpose, we implemented a verification of the processes condition variability.

3.2. Verification of processes condition variability

For the specific case of outlier detection in OPC models data, we can make use of the physics of lithography from one process condition to the other. We know that the dataset of the nominal condition of best dose and best focus is replicated with relatively small variations for the other conditions of dose and focus. For the verification for process condition variability we look for a trend of the measured data for each feature from one process condition the other. For structures detected as outliers we compare their CDs predicted by the process variation induced trend to the measured CDs. For example, we can arbitrarily define a rate “ α ” and consider as outlier the elements for which the measured CD differs more than $\alpha\%$ from the predicted value.

By using this check of physical behavior through the process condition we have been able to remove wiggles that were detected as outlier in the previous step. We applied this additional verification test to the 55 outliers previously detected. As many as 36 of them were deactivated (essentially wiggles effect) and in the 19 remaining, 15 have been checked to be true outliers. The 4 non outliers left were extreme values which were not reproduced in the other process condition. The use of this physical check trough process condition permits us to rise up the outlier detection ratio to 79%.

CONCLUSION

We presented an outlier detection method, applicable for all distribution. The usefulness of this method is that it does not always check for extreme values as outliers, but through the method of fuzzy classification seeks for coherent classes of data. The farthest values from the cluster center are indexed as outlier. The key of this method is the optimal clustering which allows detecting "hidden" outliers. Using the Mahalanobis distance, the discrimination with the others elements of the data set is enhanced. This makes the method able to detect outliers that were not spotted by other methods. We improved the quality and the robustness of the detection by using additional information based on the physics of the process through the process condition. This check permits us to remove false outlier such as wiggles. This new method permits to easily detect outlier in a large set of measurement point with an important detection rate (about 80% for our study).

This automatic outlier detection becomes very useful to quickly check outliers in a large set of experimental measurement before OPC model calibration. The example we used on OPC model calibration measurements is a case of successful application of the method.

REFERENCES

- [1] A. Kazarian, "Empirical data validation for model building", Proc. SPIE 6922, 55 (2008)
- [2] M.F. Hansen , "Mahalanobis distance based iterative closest point" , Proc. SPIE 6512, 65121Y (2007)
- [3] X. Zhang , H. Zuo , Y. Yu , "Mahalanobis-distance image segmentation based on two-dimensional histogram" ,Proc. SPIE 4550, 264-269 (2001)
- [4] A. Ukkelberg , O.S. Borgen , "Outlier detection by robust alternating regressions", Proc. CAC'92 277 no 2 (355) 489-494 (1993)
- [5] E.M. Jordaan, G.F. Smits: "Robust Outlier Detection using SVM Regression," Proc. IEEE 3 (2017-2022) 1098-7576 (2004),
- [6] F. Pukelsheim, "The three sigma rule", The American statistician ISSN 0003-1305 CODEN ASTAAJ . Papers 48(2), 88-91 (1994)

- ^[7] M. N. Ahmed, "Bias field estimation and adaptive segmentation of MRI data using modified fuzzy C-means algorithm," Proc. IEEE 1, 250-255 (1999)
- ^[8] L. Mili , C.W. Coakley, "Robust Estimation in Structured Linear Regression", Department of Statistics Virginia Polytechnic Institute and State University, Technical Report No.93-13, (1993).
- ^[9] P. J. Rousseeuw and A. M. Leroy, "Robust Regression and Outlier Detection", New York Wiley, (1987).