Outlier Detection in Wafer Data

Brandon Jinpyo Lee

Table of contents

1. Background and Introduction

2. Literature Review

3. Data Description

4. Data Preparation

5. Methodology

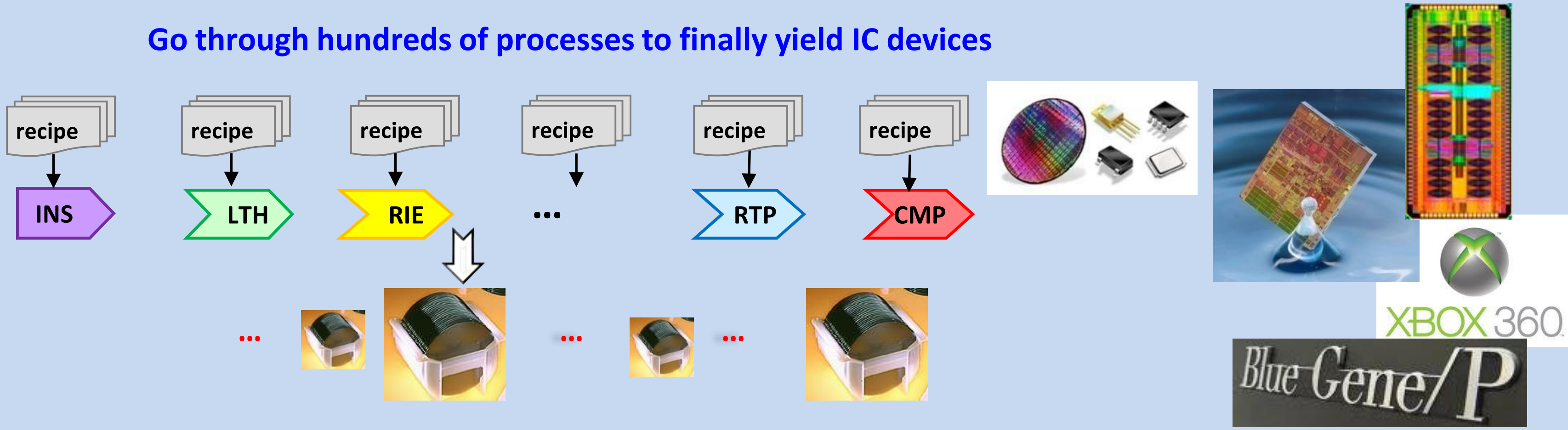
6. Experiments and Results

7. Conclusion

References

1. Background and Introduction

Go through hundreds of processes to finally yield IC devices



Tools publish large amounts of real-time data

Temperature & pressure

Throttle valve positions

Gas flows

Electric bias, impedance, etc

Can we use the data for early fault detection and control?

1. Background and Introduction
- Fault Detection & Classification

- Objective
  - Analysis of process data for early detection of machine faults.
- Challenges
  - A semiconductor fabrication process contains thousands of chemical and physical reactions that involve hundreds of process variables with complicated relationship
  - Most processes are not stationary due to tool wear and various tool maintenance interventions
- Traditional FDC method
  - Hotelling T^2 based multivariate analysis with SPC control
  - The method is not sensitive
    - When there are a large number of process variables
    - When the process is not stationary

1. Background and Introduction

- The aim of this research is to find out outlier wafer in the multivariate data collected from wafer manufacturing process.

- Based on 2 data sets, 4 models are suggested to predict outlier wafers by using moving windows.

- Global and local detection were implemented by using one-moving window-out cross validation and real-time detection frameworks

- Finally, the indices of outlier wafer were discovered from the results of implemented frameworks

- The suggested model and methodology are proved to be powerful in predicting outlier wafers.

2. Literature Review

- The literature in the field used statistical, regression and machine learning approaches.

- Statistical approaches were applied in several literatures. Parametric analysis methodology was used to detect and quantify the parametric sensitivity of the product yield

- Regression approaches were used in literatures. PCA and partial least squares (PLS) were used for the statistical process control of both continuous process and multivariate processes.

- Difficulties existed due to nonlinearity characteristics. In order to deal with these, a fault detection method using the k-nearest neighbor rule (FD-kNN) was developed.

- Neural-network-based stepwise selection method, generic programming, support vector machine (SVM) were used for wafer classification.

3. Data Description

- The data has time series characteristics in 2 dimensions.
  1) Wafers enters the chamber by the row sequence. (1 ~ n)
  2) The chamber have 14 measurements in 10 time intervals. (t = 1~10)

- 2 data sets were used.
  1) Data 1: 1114 wafers × (1 continuous class value + 140 attributes)
  2) Data 2: 395 wafers × (1 continuous class value + 140 attributes)

No.	Measurement
1	'Ch-BCHDFlow
2	'Ch-DEMSFlow
3	Ch-GaspanelDEMSCorrFactor
4	Ch-HE_CARR_BCHDFlow
5	Ch-HE_CARR_DEMSFlow
6	Ch-O2_LowFlow
7	Ch-PressureReading
8	Ch-ThrottleValvePos
9	ElecBias
10	HeaterCurTemp
11	HighFreqRFFwdPwr
12	HighFreqRFOn
13	Impedance I
14	Impedance R

< 14 measurements in the chamber >

3. Data Description

Sequence of Entry	Time interval									
	t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10
	14 measure-ments	14 measure-ments	14 measure-ments	14 measure-ments	14 measure-ments	14 measure-ments	14 measure-ments	14 measure-ments	14 measure-ments	14 measure-ments
1										
2										
3										
4										
5										
6										
….										
….										
n										

<The structure of the data >

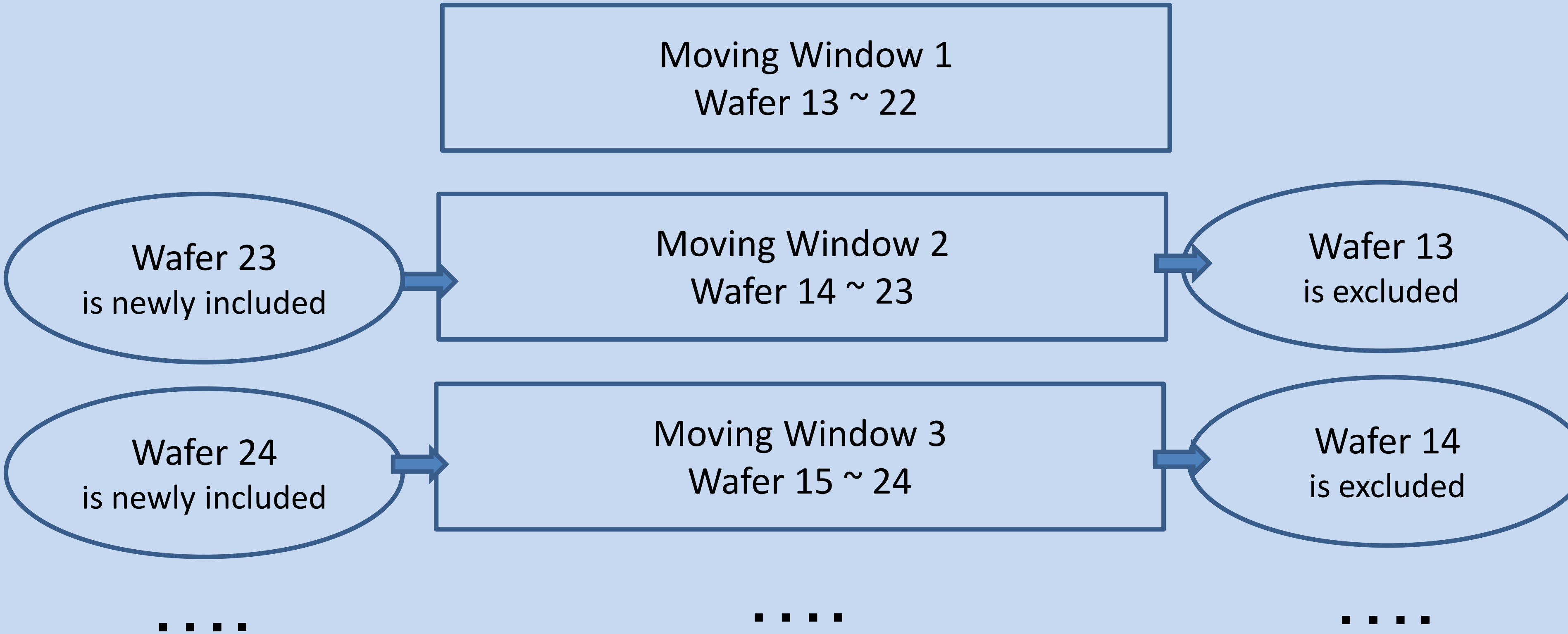
Sequence of Entry	Time interval													
	t=1													
	measure-ment 1	measure-ment 2	measure-ment 3	measure-ment 4	measure-ment 5	measure-ment 6	measure-ment 7	measure-ment 8	measure-ment 9	measure-ment 10	measure-ment 11	measure-ment 12	measure-ment 13	measure-ment 14
1														
2														
3														
4														
5														
6														
….														
….														
n														

<The structure of data in the 1st time interval (t=1) >

4. Data Preparation

- 25 columns with no variance are eliminated in advance.
- 3 observations (index 10,11 and 12) are assumed to be outliers due to their uniquely high class values (>200)
- Initial 12 observations including the 3 outliers were removed to ensure continuity in moving window approach.
- Class and all attribute values were normalized with mean=0, $\sigma = 1$
- n-9 moving windows are made from 10 observations by adding a new wafer and excluding one wafer.
- For each moving window, 4 variables are calculated for every attribute and one more normalization is implemented. Then, they are used for new attributes.

4. Data Preparation - structure of moving window



- Initial 12 wafers including 3 huge outlier wafers are deleted in advance.
- Moving window *i* contains wafer index from *i*+12 to *i*+21
- As *i* increases, a new wafer is added and a wafer is excluded.

4. Data Preparation

- Expected effects of using following 4 variables in improving outlier detection

1. **Mean** – represents abnormal production status if an observed value is far from mean.
2. **Variance** – represents the degree of inconsistency of the operation
3. **Delta** (mean of absolute values of differences between two consecutive wafers in a window) – represents the degree of radically changing status of operation
4. **Variance of Delta** – represents the degree of inconsistency of the radically changing status.

4. Data Preparation

4 models are made by the following assumptions.

- 1st model: variance of class is highly correlated with the variances of attributes.
- 2nd model: variance of delta of class is highly correlated with the variances of delta of attributes.

- However, mean and delta also can have the outlier characteristics since abnormal level of mean and delta means that the operation is not a normal condition. So, 2 additional models were made.

- 3rd model: variance of class is highly correlated with all of the 4 variables.
- 4th model: variance of delta of class is highly correlated with all of the 4 variables.

Scenario (model)	Class	Attributes
1	variance	variance
2	variance of delta	variance of delta
3	variance	mean, variance, delta, variance of delta
4	variance of delta	mean, variance, delta, variance of delta

< Class and attributes in 4 models >

4. Data Preparation

- In addition to global detection, two models of local detection were implemented.
- Firstly, stepwise local detection was implemented. 10 measurements out of 14 measurements are used as the attributes for the 10 local groups because measurements 9,12,13 and 14 are valid only for later half of the steps.
- Secondly, measurement-wise local detection was implemented. The 10 time series values of 14 measurements were respectively grouped as a local group, resulting in 14 local groups.

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14
step 1	1	1	1	1	1	1	1	0	0	1	1	0	0	0
step 2	1	1	1	1	1	1	1	0	0	1	1	0	0	0
step 3	1	1	1	1	1	1	1	1	0	1	1	0	0	0
step 4	1	1	1	1	1	1	1	1	0	1	1	0	0	0
step 5	1	1	1	1	1	1	1	1	0	1	1	0	0	0
step 6	1	1	1	1	1	1	1	1	1	1	1	1	1	1
step 7	1	1	1	1	1	1	1	1	1	1	1	1	1	1
step 8	1	1	1	1	1	1	1	1	1	1	1	1	1	1
step 9	1	1	1	1	1	1	1	1	1	1	1	1	1	1
step 10	1	1	1	1	0	0	1	1	0	1	1	1	1	1

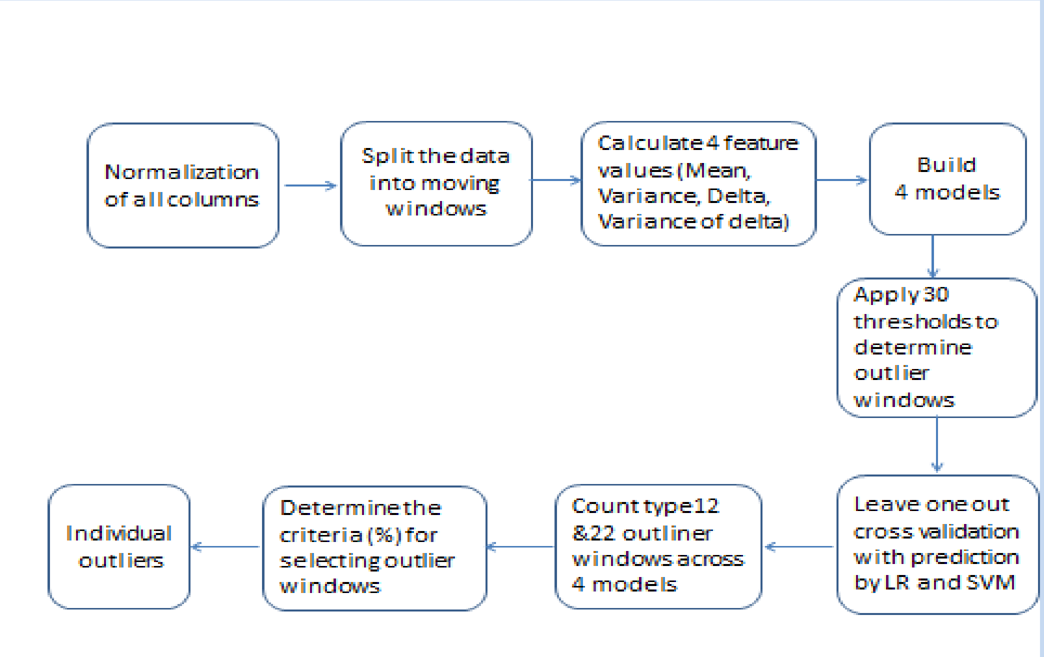
< Used measurements in stepwise detection >

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13
step 1	1	1	1	1	1	1	1	0	0	1	1	0	0
step 2	1	1	1	1	1	1	1	0	0	1	1	0	0
step 3	1	1	1	1	1	1	1	1	0	1	1	0	0
step 4	1	1	1	1	1	1	1	1	0	1	1	0	0
step 5	1	1	1	1	1	1	1	1	0	1	1	0	0
step 6	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
step 7	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
step 8	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
step 9	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
step 10	1	1	1	1	0	0	1	1	0	1	1	1	1

< Used measurements in measurement-wise detection >

5. Methodology

- Normalization of attributes.
- Extracting moving windows
- Calculation of 4 variables (mean, variance, delta. Var. of delta)
- Building 4 prediction models
- Application of 30 thresholds σ (0.5 ≤ σ ≤ 3.4) to determine outliers
- Leave-one-out cross validation by LR and SVM
- Counting type 12 & 22 outlier windows
- Determination of criteria for selecting outlier windows.
- Tracing back individual outlier wafers.



< Flow diagram of methodology >

5. Methodology

- **Discretization of class :** Sufficient range of standard deviation is used as a threshold (risk level) to discretize the continuous class values into two groups, i.e., normal ones (0) and outliers (1).

- **Classifier:** As the classification methods for training data set and testing data set, logistic regression and support vector machine were applied.

- **Validation frameworks** : leave-one moving window-out cross validation and real-time detection were used.

(1) Leave-one moving window-out cross validation: one moving window was used for testing and the rest of the windows were used for training the classifiers.

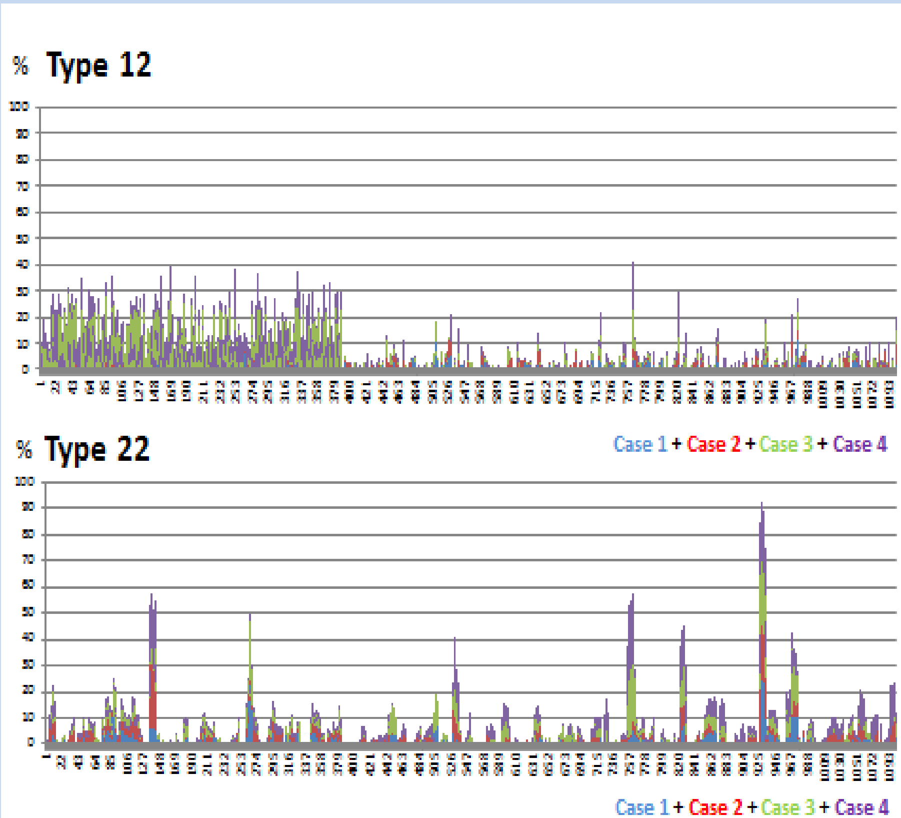
(2) Real-time detection: To predict the class of new window by using classifier trained by past samples, first 50 moving windows were used for training in the 1st iteration. The number of training samples increased by one as the number of iterations increases. The upcoming one window was used for testing.

5. Methodology

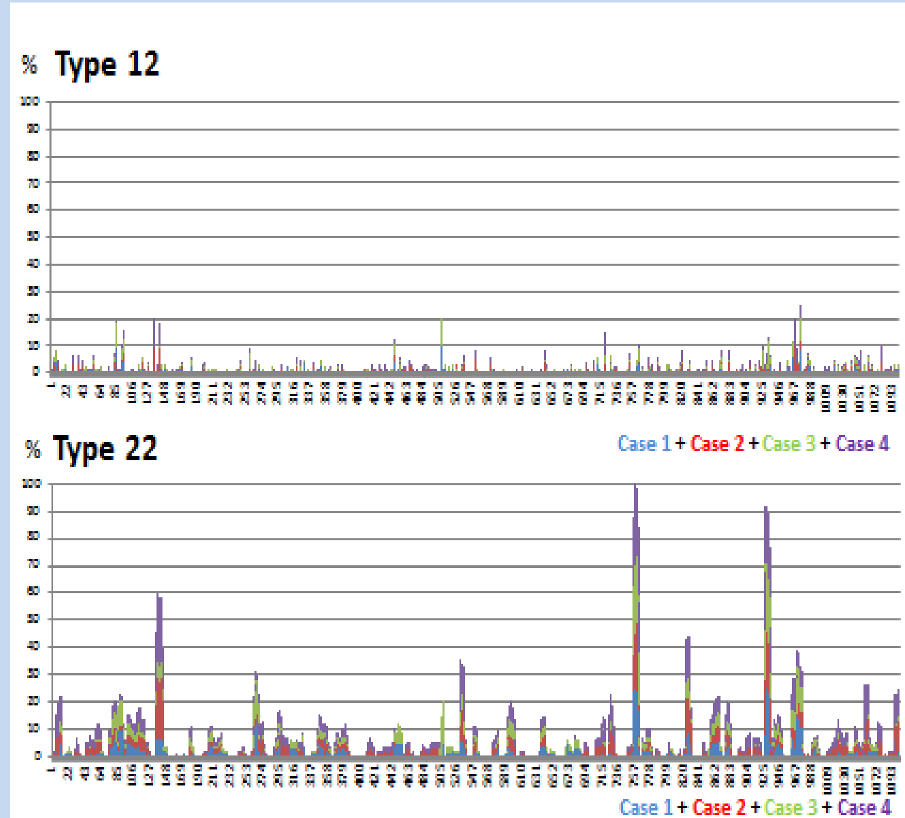
- In testing, the frequencies of how many times each window appears as outliers are counted on different thresholds σ ($0.5 \le \sigma \le 3.4$),
- At $\sigma = 0.5$, the class values of most windows are classified into 1 (outlier). In contrast, at $\sigma = 3.4$, the class values of most windows are classified into 0 (normal).
- The above range of σ is sufficient in observing the most frequently occurring windows across the 30 σ values, which is the clearest outlier.
- In confusion matrices, the frequencies of Type 12 and Type 22 are meaningful because the classified types are outliers.
- The 2 types: (1) Type 12 (normal in original, classified as outlier window)
  (2) Type 22 (outlier in original, classified as outlier window.)
- The frequencies of 2 classification type were counted for 4 models to find out the aggregate outlier frequencies.

6. Experiments and Results

A. Global Detection



< Result by using Logistic Regression >

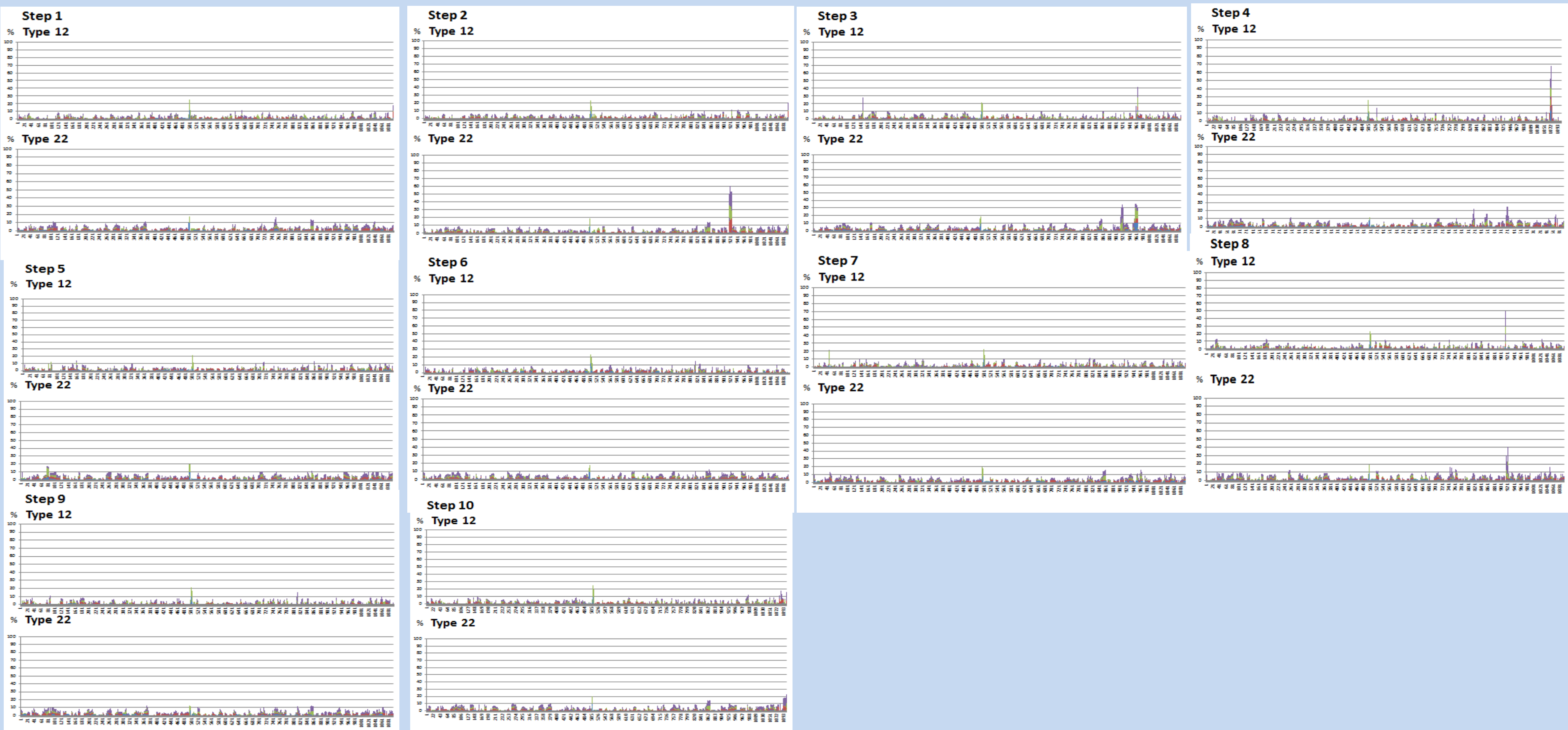


< Result by using SVM >

- The indices of outlier moving windows are almost same in implementations from LR and SVM even if the rankings of outlier are different.

6. Experiments and Results

B. Stepwise local Detection



- The indices in several steps at which outlier shows coincides with the results of global detection, However, not all the outliers in global detection appeared since not all the measurements were used in local detection.

6. Experiments and Results

C. Real-time Detection

- Next slide shows the results of frequency counting by training on cumulative wafers and testing on 1 wafer.

- Initial 50 moving windows were used as the initial number of moving windows for training. 51st moving window was used as the initial testing moving window.

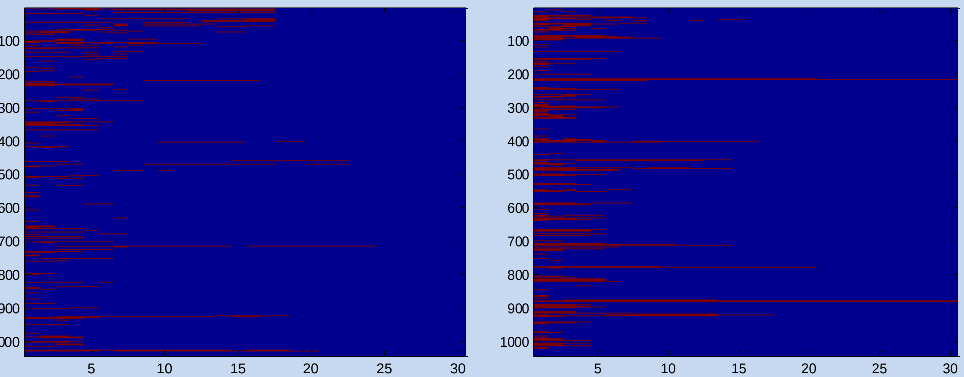
- The patterns of outlier moving windows (red color) and normal moving windows (blue color) were visualized.

- The horizontal axis represents the applied standard deviation (σ). The number of outlier moving windows increases as lower value of σ is applied. In contrast it decreases as higher value of σ is applied.

- If a moving window turns out to be outlier with higher frequency across the 30 applied σ, the window is close to real outlier window.

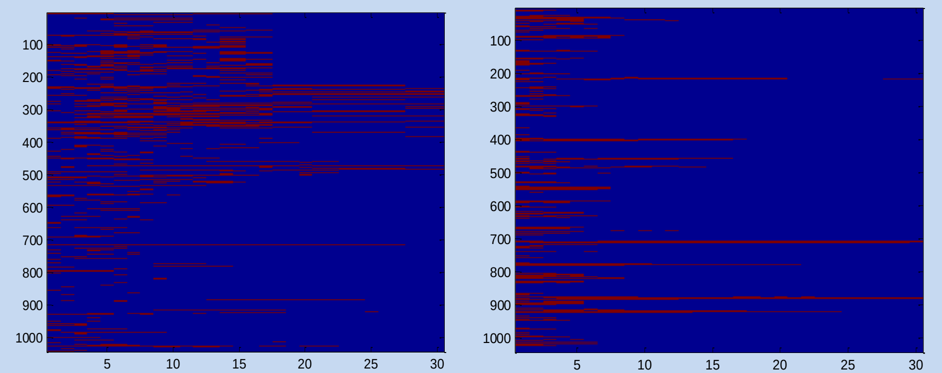
6. Experiments and Results

C. Real-time Detection



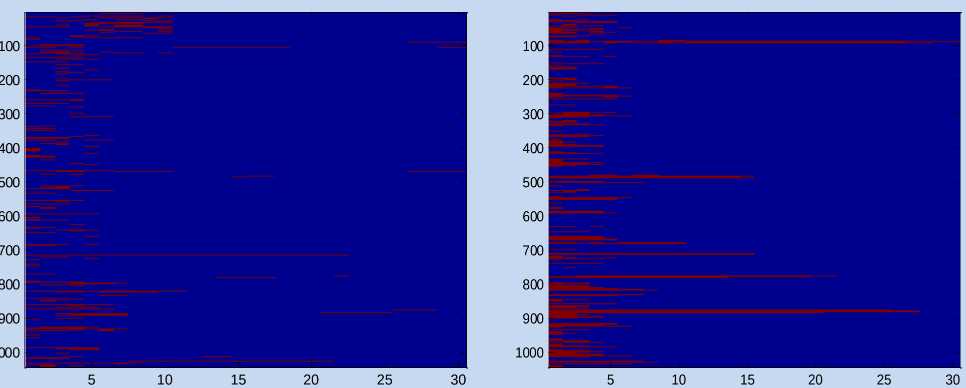
(a) Type 12 (b) Type 22

< Real-time global detection of outlier moving window of model 1 >



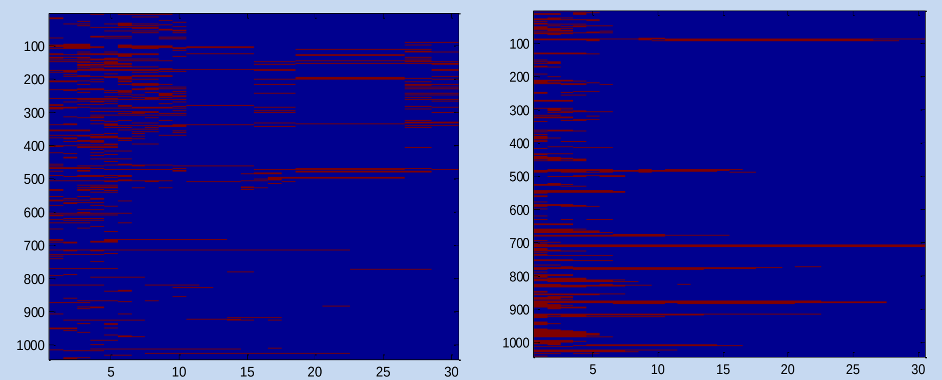
(a) Type 12 (b) Type 22

< Real-time global detection of outlier moving window of model 3 >



(a) Type 12 (b) Type 22

< Real-time global detection of outlier moving window of model 2 >



(a) Type 12 (b) Type 22

< Real-time global detection of outlier moving window of model 4 >

6. Experiments and Results

D. Conversion into Outlier Wafer Indices

- In tracing process, delta (difference of Type 22 frequencies between 2 consecutive moving windows) is used.

- It is assumed that each wafer has 'outlier effect', the degree of the wafer's contribution to the frequency of Type 22 outlier windows.

Delta(i) = Outlier effect of newly entering wafer k – Outlier effect of excluded wafer j

- The outlier effect of newly entering wafer is largely reflected on delta because the sizes of positive peaks are sufficiently high compared to other values of delta.

- The high positive delta means that an outlier wafer newly entered the moving window. By this method, the original indices of outlier wafer were found.

- Appropriate level of threshold (in %) is used as the criteria in determining the outlier moving windows.

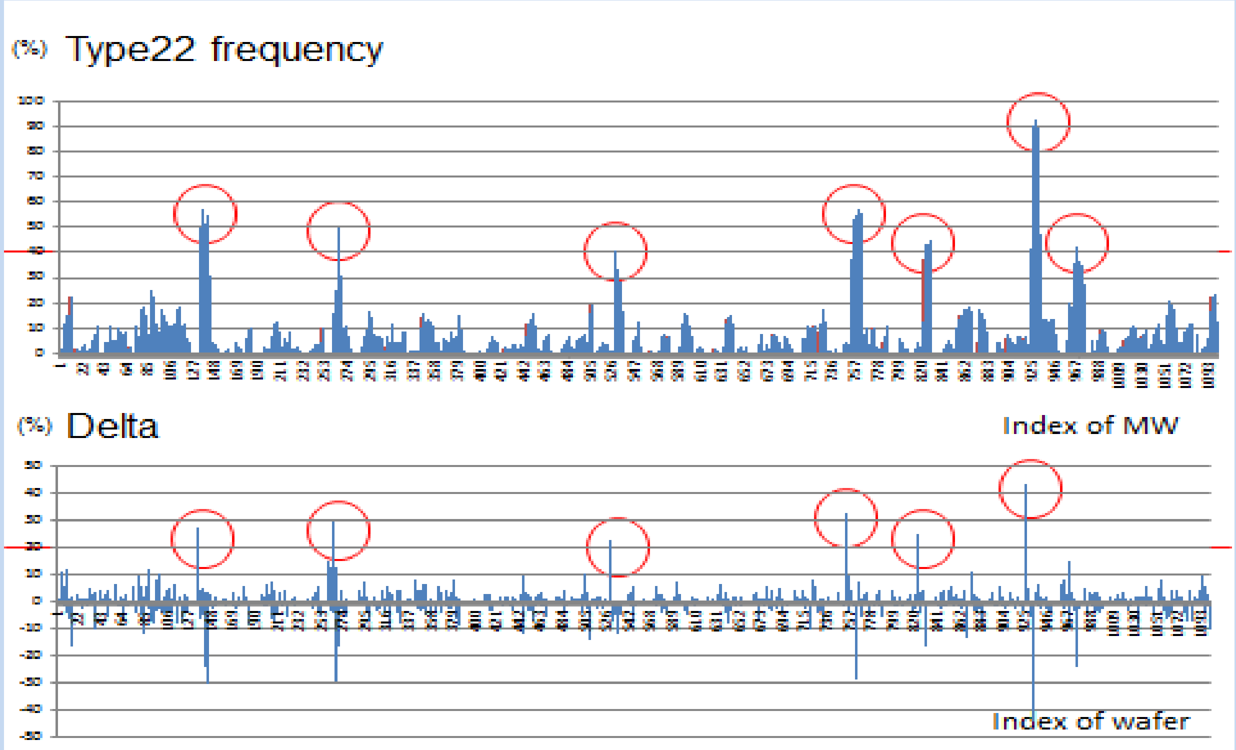
6. Experiments and Results

- Calculation of Delta between consecutive moving windows

Moving window Index (i)	Beginning wafer index (j)	Ending wafer index (k)	Frequency of Type22(%) (F)	Delta F(i)-F(i-1)
1	13	22	20.83	0
2	14	23	20	-0.8
3	15	24	7.5	-13
4	16	25	0	-7.5
5	17	26	0	0
6	18	27	0	0
7	19	28	0	0
8	20	29	0	0
9	21	30	1.667	1.67
10	22	31	0	-1.7
11	23	32	1.667	1.67
12	24	33	2.5	0.83
13	25	34	1.667	-0.8
14	26	35	1.667	0

- Delta = outlier effect of newly included wafer j - outlier effect of excluded wafer k

6. Experiments and Results

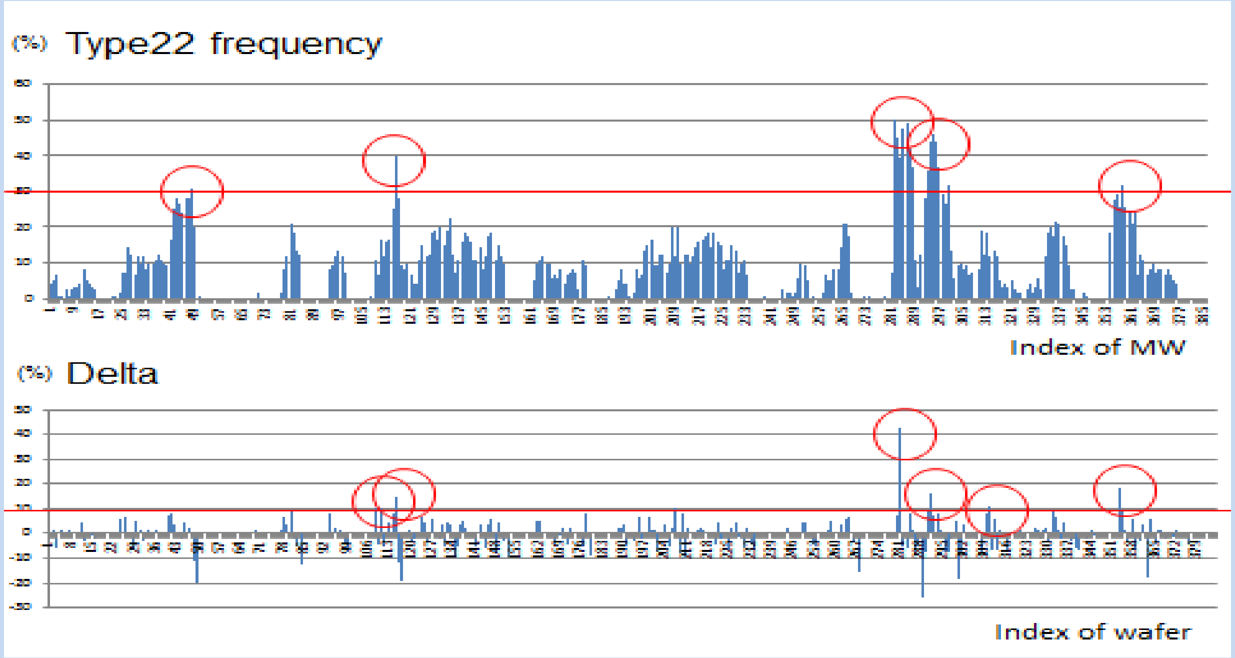


< Comparison of type 22 frequency and delta for outlier wafer detection on data 1.
(a) The frequency of Type 22 outlier moving window (b) Delta >

Indices of outlier wafer	Remarks
145, 146	Group
276	Individual
540	Individual
765	Individual
834	Individual
936, 937	Group

< The indices of outlier wafer >

6. Experiments and Results



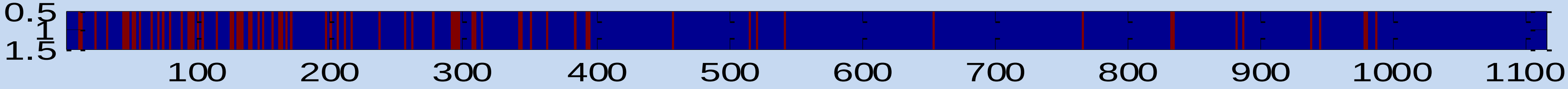
< Comparison of type 22 frequency and delta for outlier wafer detection on data 2.
(a) The frequency of Type 22 outlier moving window (b) Delta >

Indices of outlier wafer	Remarks
118	Individual
120	Individual
125	Individual
217	Individual
291	Individual
301	Individual
320	Individual
363	Individual

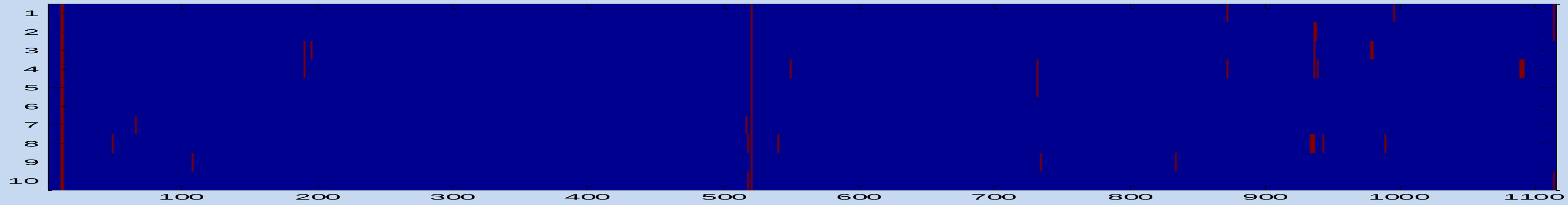
< The indices of outlier wafer >

6. Experiments and Results

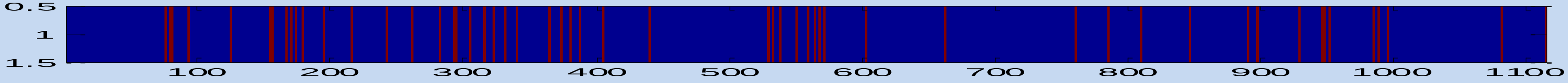
E. Final Results of Data 1



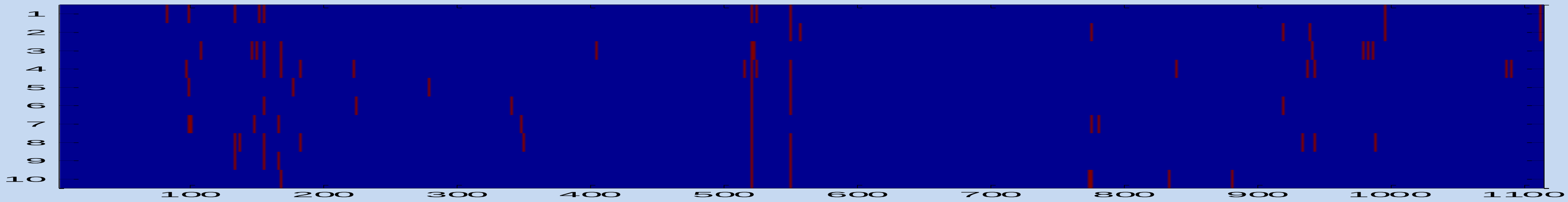
< Image plot of leave-one moving window-out global detection on data 1 >



< Image plot of leave-one moving window-out local detection on data 1 >



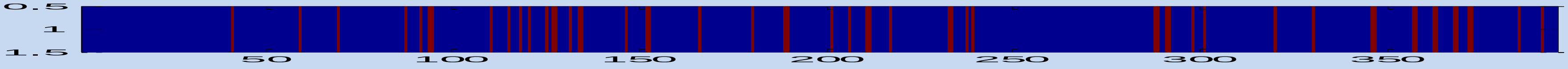
< Image plot of real-time global detection on data 1 >



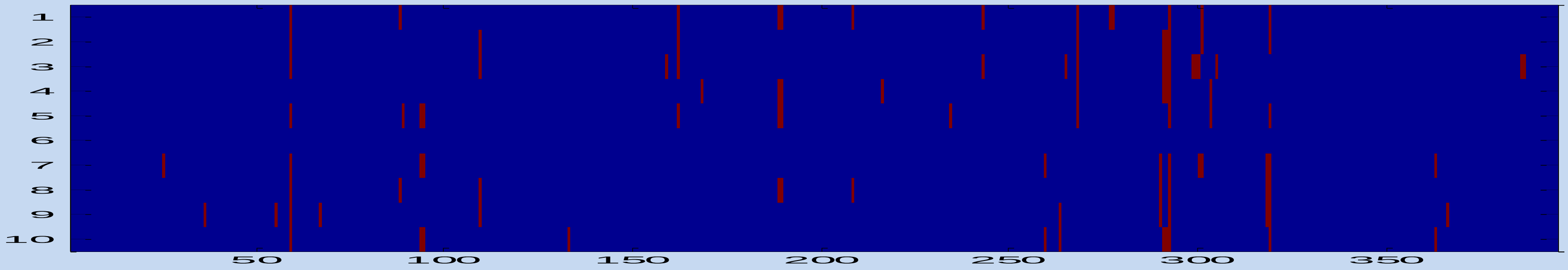
< Image plot of real-time local detection on data 1 >

6. Experiments and Results

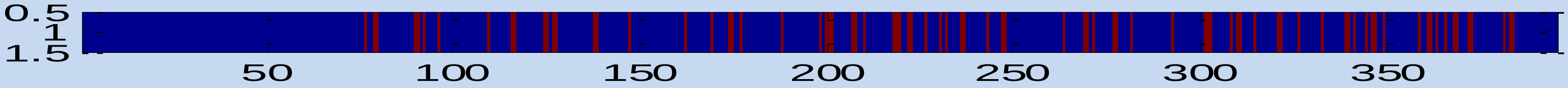
E. Final Results of Data 2



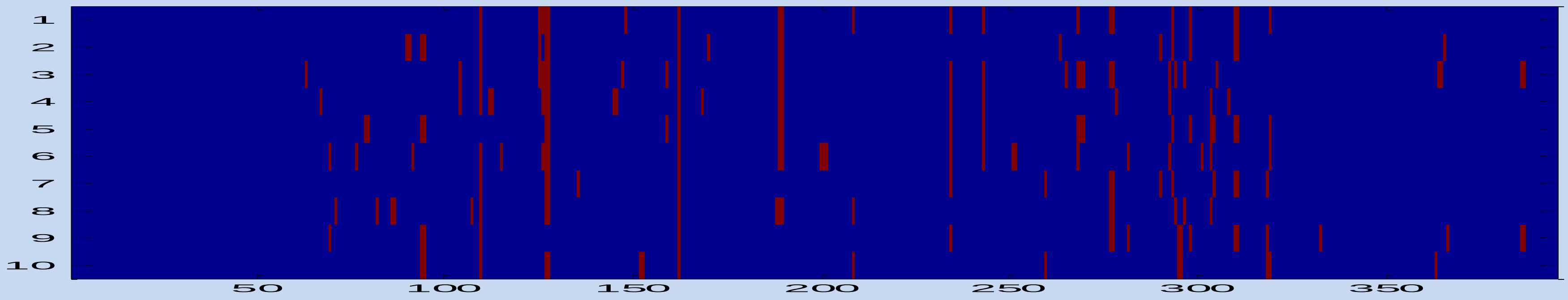
< Image plot of leave-one moving window-out global detection on data 2 >



< Image plot of leave-one moving window-out local detection on data 2 >



< Image plot of real-time global detection on data 2 >



< Image plot of real-time local detection on data 2 >

6. Experiments and Results

E. Final Results

- All the results on outlier wafer indices were represented by image plot in order to compare the results simultaneously.

- In all cases, the founded outlier indices in local detections are included in the founded outlier indices in global detections.

- However, there exist additional outliers in global detections. This is because only 10 measurements were used for stepwise local detections while all non-zero-variance columns were used for global detections.

- This observation leads to the conclusion that the excluded measurements 9,12,13 and 14 include the outlier characteristics, which is not shown in the local detections.

6. Experiments and Results

E. Final Results

- On the other hand, if we compare the two results from leave-one moving window-out detection and real-time detection, it is observed that more outlier wafers are added toward the beginning points of detections.

- This observation means that there are more detected outlier wafers in the beginning stage of real-time detection due to the lack of training examples, which starts from 50 for both data sets.

- However, this number of detected outlier wafers reduces as the index of wafer increases and the image plots of two different cases converge to similar pattern even if it is not exactly identical.

7. Conclusion

- By applying different threshold σ on 4 prediction models by using 4 extracted variables from moving window, the outlier wafers can be detected with high accuracy.
- The outlier detection is feasible by both one-moving window-out and real-time detection frameworks.
- Regardless of the type of used classifiers, i.e., LR or SVM, similar results can be obtained.
- The ranking of outlier is relative to the used classifiers and frameworks, However, the almost same indices of outlier can be detected if finding specific number of outliers is the purpose.

References

- **[1] J. Fenner and H. Hajj, “APC application selection: The secret ingredient to a successful APC program,” in *Proc. AEC/APC symp. XII,* Sep. 2000.**
- **[2] H. Hajj, J. R. Moyne, M. Pendleton, T. Guckert, K. Edwards, and C.Young, “E133: The process control systems (PCS) SEMI standard: Applying the standard to fault detection scenarios incorporating interface a data collection and other PCS capabilities,” in *Proc. AEC/APC symp. XVIII, 2006.***
- **[3] J.R. Moyne, H. Hajj, K Beatty, and R. Lewandowski, “SEMI E133: The process control system standard: Deriving a software interoperability standard for advanced process control in semiconductor manufacturing.” *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 4, pp. 408-420, Nov. 2007.**
- **[4] A. Y. Wong. “A statistical parametric and probe yield analysis methodology,” *in Proc. IEEE Int. Symp. Defect Fault Tolerance VLSI Syst*., Nov. 1996, pp. 131-139.**
- **[5] R. Kittler and W. Wang, “Data mining for yield improvement,” in *Proc. Int. Conf. Model. Anal. Semicond. Manuf.*, 2000, pp. 270-277.**
- **[6] J. F. MAcGregor and T. Kourti, “Statistical process control of multivariate processes,” *Contr. Eng. Practice*, vol. 3, no. 3, pp. 403-414, 1995.**
- **[7] Q. P. He and J. Wang, “Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes,” *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 4, pp. 345-354, Nov. 2007.**
- **[8] Q. He and J. Wang, “Principal component based k-nearest neighbor rule for semiconductor process fault detection’” in *Proc. Am. Contr. Conf.,* 2008, pp. 1606-1611.**
- **[9] Y. Dupret, E. P. Jean-Luc Crolier, and R. Kielbasa, “Comparison of three different methods to model the semiconductor manufacturing yield,” in *Proc. IEEE/SEMI Adv. Semicond. Manuf. Conf. Workshop,* Apr. 2005, pp. 118-123**
- **[10] A. A. Khan, J. R. Moyne, and D. M. Tilbury, “Virtual metrology and feedback control for semiconductor manufacturing processes using recursive partial least square,” *J. Process Contr.,* vol 18, no. 10, pp. 961-974, Dec. 2008.**
- **[11] T.H. Lin, F.T. Cheng, W. M. Wu, C. A. Kao, A. J. Ye, and F.C. Chang, “NN-based key-variable selection method for enhancing virtual metrology accuracy,” *IEEE Trans. Semicond. Manuf.*, vol.22, no.1,pp.204-211, Feb. 2009.**
- **[12] T.S. Li, C. L. Huang, and Z.Y. Wu, “Data mining using genetic programming for construction of a semiconductor manufacturing yield rate prediction system,” J. *Intell. Manuf.,* vol. 17, no. 3, pp. 355-361, 2006.**
- **[13] Ramy Baly and Hazem Hajj, “Wafer classification using support vector machines,” *IEEE Trans. Semicond. Manuf.*, vol.25, no.3,pp.373-383, Aug. 2012.**