Multivariate batch process monitoring utilizing summary variables with an application to semiconductor etch

Neal B. Gallagher
Eigenvector Research, Inc.
nealg@eigenvector.com

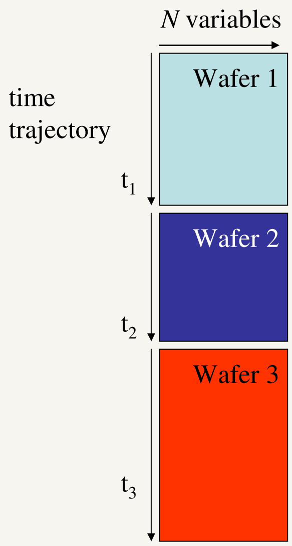
B.M. Wise, J.M. Shaver, R.T. Roginski, and R.S. Koch

Outline

- The batch / wafer process monitoring data problem
- Example data set
- Approaches to preprocessing
  - Variable means, Interpolation, Shifting, Warping, and Summary Variables
- Comparison of results
- Summary / Conclusions

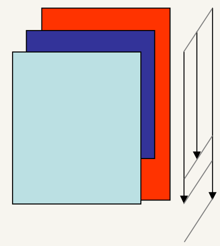
EIGENVECTOR RESEARCH INCORPORATED

Batch / Wafer Processes



Statistical monitoring of chemical batch processes and the wafter-to-wafer (run-to-run) etch process in semiconductor have similar "data problems."

For example, multi-way approaches such as MPCA can be used when the data can be arranged in a cube (or box).

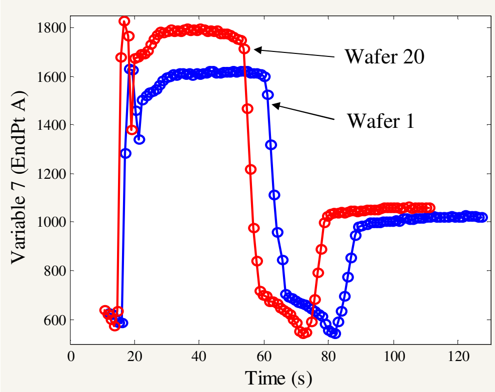


Differing data records make this difficult.

(PARAFAC2 might be used.)

EIGENVECTOR RESEARCH INCORPORATED

Batch / Wafer Processes



A simple approach might summarize a single batch/wafer as the mean (or other stats) for each of the *N* variables allowing the data to be arranged in a two-way matrix. This allows traditional multivariate approaches to statistical process monitoring to be used (e.g., PCA).

However, differing process step lengths might be expected to increase "irrelevant" variance that desensitizes the model to true faults.

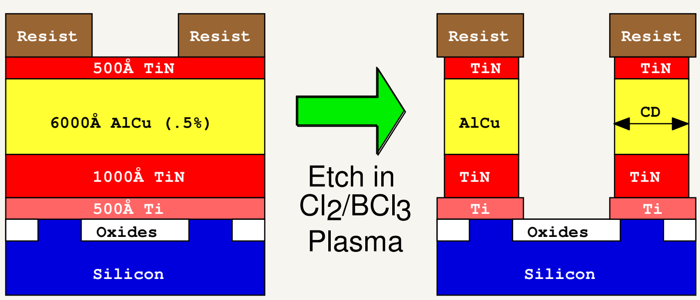
EIGENVECTOR RESEARCH INCORPORATED

Batch Process Monitoring Data Problem

- "Messy"
  - typically includes start-up and shut-down phases that are not of interest
- Periods of "steady-state" where not much is changing
- Variable record lengths
- Lots of data!
- Reduce to a set of more compact descriptors?

EIGENVECTOR RESEARCH INCORPORATED

Plasma Metal Etch



- Linewidth (Critical Dimension) Control
  - Constant linewidth reduction run-to-run and across wafer
  - Constant linewidth reduction for every material in stack
- Minimal damage to oxide
- data is available at www.eigenvector.com

EIGENVECTOR RESEARCH INCORPORATED

Available Measurements

- "Machine Data"
- Equipment has SECS-II Port
  - Provides traces with time stamp and step number
- Regulatory controller setpoints & controlled variable measured values
  - gas flows, pressure, plasma powers
- Regulatory controller manipulated variables
  - exhaust throttle valve, capacitors
  - mass flow controller do not provide valve position
- Additional process measurements
  - endpoint intensity (broadband plasma emission signal)
  - impedance measurements

EIGENVECTOR RESEARCH INCORPORATED

Machine Variables Modeled

Machine Variables Available (bold were used in model)

1	**BCl3 Flow**	11	**RF Pwr**
2	Cl2 Flow	12	**RF Impedance**
3	RF Btm Pwr	13	**TCP Tuner**
4	RF Btm Rfl Pwr	14	**TCP Phase Err**
5	**Endpt A**	15	**TCP Impedance**
6	He Press	16	TCP Top Pwr
7	**Pressure**	17	**TCP Rfl Pwr**
8	**RF Tuner**	18	**TCP Load**
9	**RF Load**	19	**Vat Valve**
10	**RF Phase Err**		

EIGENVECTOR RESEARCH INCORPORATED

Metal Etch Data Set

- Barna, G.G., White, D., Wise, B.M., Gallagher, N.B., Sofge, D., “Development of Robust Fault Detection and Classification Techniques/SEMATECH J-88E Project at TI”, SEMATECH AEC/APC Workshop VIII, Santa Fe, New Mexico, Oct. 27-30, 1996.
- White, D., Barna, G.G., Butler, S.W., Wise B., Gallagher, N., "Methodology for Robust and Sensitive Fault Detection," Electrochemical Society Meeting, Montreal, May, 1997.
- Gallagher, N.B.,Wise, B.M., Butler, S.W., White, D., Barna, G.G., "Development and Benchmarking of Multivariate Statistical Process Control Tools for a Semiconductor Etch Process: Improving Robustness Through Model Updating", IFAC ADCHEM'97, Banff, Canada, 78–83, June, 1997.
- Wise, B.M., Gallagher, N.B., Butler, S.W., White, D., Barna, G.G., "A Comparison of Principal Components Analysis, Multi-way Principal Components Analysis, Tri-linear Decomposition and Parallel Factor Analysis for Fault Detection in a Semiconductor Etch Process," *J. Chemometr.*, **13**, 379–396 (1999).
- Wise, B.M., Gallagher, N.B., Butler, S.W., White, D., Barna, G.G., "Development and Benchmarking of Multivariate Statistical Process Control Tools for a Semiconductor Etch Process: Impact of Measurement Selection and Data Treatment on Sensitivity", IFAC SAFEPROCESS'97, 35–42, Kingston Upon Hull, U.K., Aug., 1997.
- Wise, B.M., Gallagher, N.B., Martin, E.B., “Application of PARAFAC2 to Fault Detection and Diagnosis in Semiconductor Etch,” *J. Chemometr.*, **15**(4), 285–298 (2001).
- Wise, B.M., Gallagher, N.B., "Multi-way Analysis in Process Monitoring and Modeling," *AIChE Symposium Series*, **93**(316), 271–274 (1997).
- Warren, J., Gallagher, N.B., “Heuristic and Statistical Methods for Fault Detection: Complementary or Competing Approaches?”, SEMATECH AEC/APC Symposium XVIII, Westminster, CO, Sept. 30-Oct. 5 (2006).

Available Measurements

- Equipment has SECS-II Port
  - Provides traces with time stamp and step number
- Regulatory controller setpoints & controlled variable measured values
  - gas flows, pressure, plasma powers
- Regulatory controller manipulated variables
  - exhaust throttle valve, capacitors
  - mass flow controller do not provide valve position
- Additional process measurements
  - endpoint intensity (broadband plasma emission signal)
  - impedance measurements
- Optical emission spectra
- RF plasma variables

EIGENVECTOR RESEARCH INCORPORATED

Sensitivity of MSPC Models

- Three experiments performed with 20 “induced” faults:
  - TCP top power, RF bottom power, Cl2 flow, BCl3 flow, Chamber pressure, and Helium chuck pressure
- Goal: Compare ability of local (developed for each experiment) models considered for detecting faults

Experiment ID Number	Number of Normal Wafers	Number of Fault Wafers	Total Number of Wafers
29	34	9	43
31	36	5	41
33	37	6	43
Total	107	20	127

EIGENVECTOR RESEARCH INCORPORATED

Generating Faults

- Setpoints were changed for controlled process variables
  - TCP top power, RF bottom power, Cl2 flow, BCl3 flow, Chamber pressure, and Helium chuck pressure
- Data for the controlled variable were adjusted to have the original desired mean (i.e., setpoint)
- Result is data that looks like a sensor has developed a bias

EIGENVECTOR RESEARCH INCORPORATED

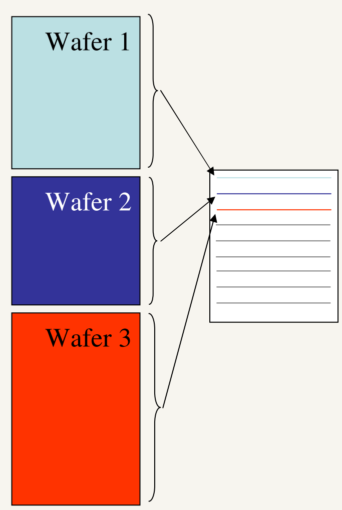
Global and Local Models

- Global models based on data over long period of time with considerable variance due to drift
  - time scale includes multiple preventative maintenance cycles and must account for PM-to-PM, Lot-to-Lot, and Wafer-to-Wafer variation
- Local models built over narrower time windows, less drift variance
  - time scale includes multiple wafers and must account Wafer-to-Wafer variation

EIGENVECTOR RESEARCH INCORPORATED

Data Processing 1

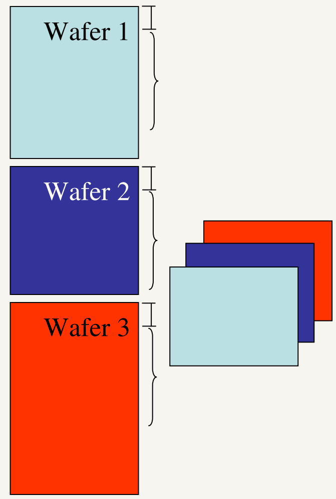
- Five methods for data processing (arrangement) were considered
- Wafer mean
  - each wafer's mean is 1xN → M_{cal}xN
    - Easy, conceptually simple
    - Large data reduction
    - No problem with variable record length
    - Can add batch length as extra variable
    - Can use Principal Components Analysis w/ autoscaling (well established)
  - Cons:
    - Lose trajectory information
    - Lose time information



EIGENVECTOR RESEARCH INCORPORATED

Data Processing 2

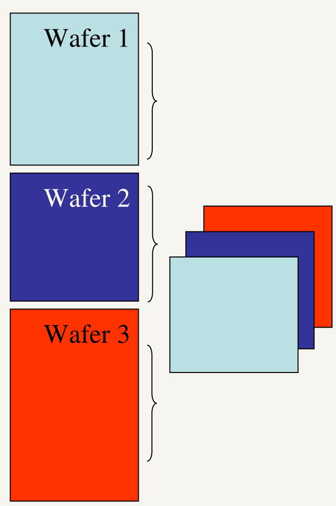
- Interpolate to common time axis
  - each wafer's data is a $M_1 x N$ matrix $\rightarrow$ $M_2 x N x M_{cal}$
    - Processing is easy
    - Some data reduction
    - Remove variable record length
    - Can use multi-way methods
      - Mulitway-Principal Components Analysis w/ block scaling studied here
  - Cons:
    - Trajectory shifts / warps not accounted for
    - Interpolation must use shortest time axis anticipated
    - many variables
    - interpretation can be difficult



EIGENVECTOR RESEARCH INCORPORATED

Data Processing 3

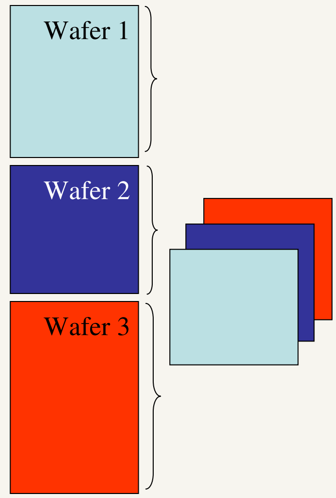
- Shift to common trace (correlation maximized using a PCA model)
  - each wafer's data is a $M_3 x N$ matrix $\rightarrow$ $M_3 x N x M_{cal}$
    - Moderate difficulty
    - Shifts accounted for
    - Some data reduction
    - Remove variable record length
    - Can use multi-way methods
      - MPCA studied here
  - Cons:
    - Trajectory warps not accounted for
    - Reference trace should be "typical"



EIGENVECTOR RESEARCH INCORPORATED

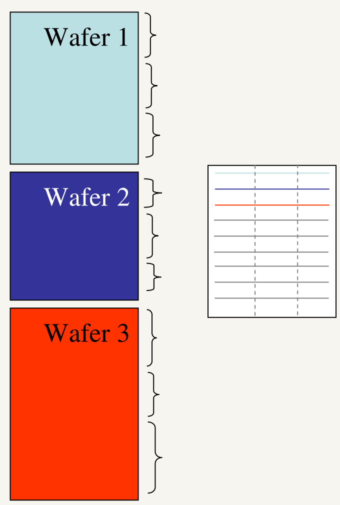
Data Processing 4

- Warp to common trace (correlation optimized warping)
  - each wafer's data is a M_4xN matrix → M_4xNxM_{cal}
    - Somewhat complex
    - Shifts and warps accounted for
    - Can have some data reduction
    - Remove variable record length
    - Can use multi-way methods
      - MPCA studied here
  - Cons:
    - Trajectories must be similar between batches
    - Reference trace should be "typical"
    - Loss of step time length



Data Processing 5

- Warp to common trace (COW) and calculate summary in process steps
  - each wafer's data is a 1xN_5 matrix → M_{cal}xN_5
    - Somewhat complex, conceptually easy
    - shifts and warps accounted for
    - Data reduction
    - Remove variable record length
    - Can use two-way methods
      - PCA studied here
  - Cons:
    - Trajectories must be similar between batches
    - Reference trace should be "typical"



Alignment Methods

- Several methods available
  - Dynamic Time Warping (DTW)
  - Correlation Optimized Warping (COW)
  - Indicator variable/step number
  - Linear interpolation
  - Align and truncate
  - Combinations and variations of the above
  - Etc.

EIGENVECTOR RESEARCH INCORPORATED

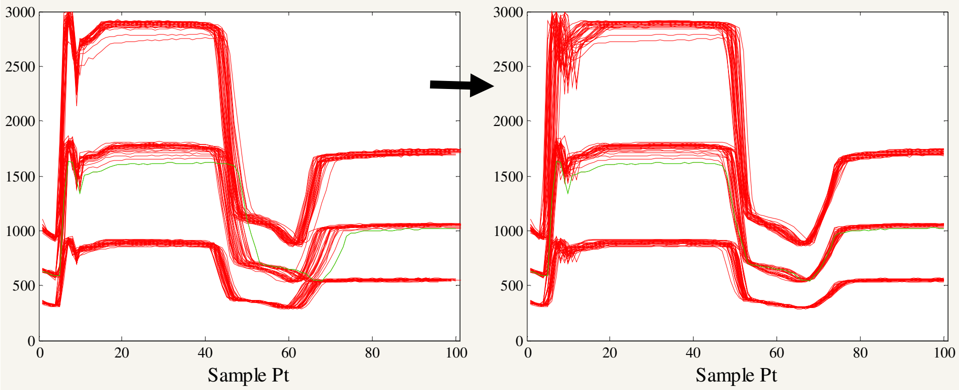
COW References

- N.P.V Nielsen, J.M. Carstensen and J. Smedsgaard, "Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimized warping," *J. Chromatogr. A*, **805**, 17-35, 1998.
- G. Tomasi, F. van den Berg and C. Andersson, "Correlation Optimized Warping and Dynamic Time Warping as Preprocessing Methods for Chromatographic Data," *J. Chemometrics*, **18**, 231-241, 2004.
- G. Tomasi, T. Skov and F. van den Berg, Warping Toolbox, see: http://www.models.life.ku.dk/source/DTW_COW/index.asp

EIGENVECTOR RESEARCH INCORPORATED

Example: COW

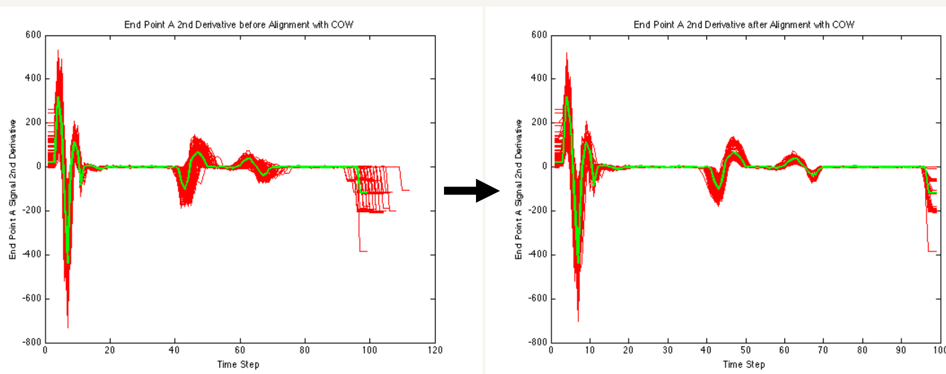
COW breaks signals into segments and linearly expands or contracts them to optimize correlation



EIGENVECTOR RESEARCH INCORPORATED

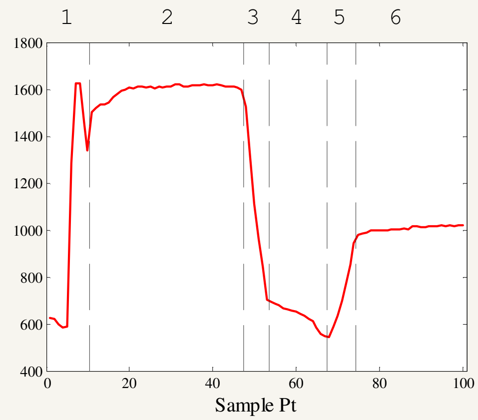
Hints on COW

May be better to calculate warp with 2nd derivative
Apply calculated warp to other variables



EIGENVECTOR RESEARCH INCORPORATED

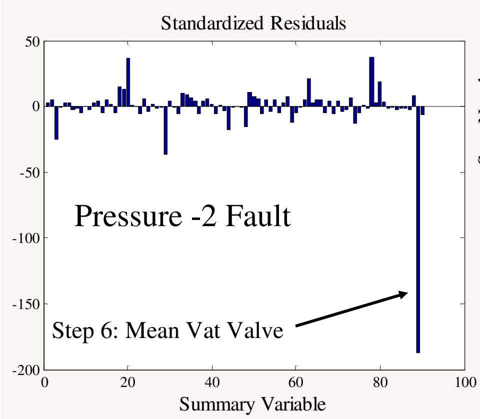
Example of Step Creation

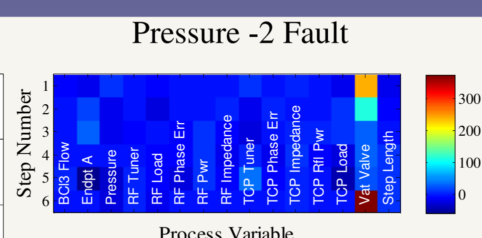


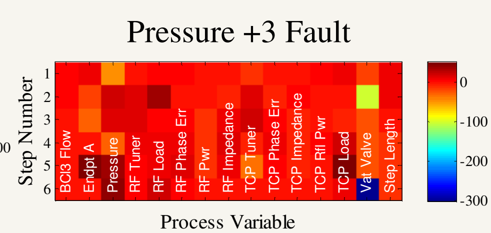
- Creation of Pseudo-steps
  - Break reference process trace into “sensible” segments (manually)
  - Assign step numbers
  - Warp new data onto reference
  - Reverse warp reference step numbers into new data
  - Calculate summary variables

EIGENVECTOR RESEARCH INCORPORATED

Contribution Plots







EIGENVECTOR RESEARCH INCORPORATED

Q/Q$_{99.9\%}$ 1 PC Exp./Fault	Data Processing Method Summary	Mean	Shift	Warp	Interp
TCP +50	12	28	6.1	3.7	2.3
RF −12	1.3	3.2	1.1	0.9	0.9
RF +10	0.9	1.3	0.7	0.7	0.6
Pr +3	276	302	40	27	19
TCP +10	0.8	1.4	0.7	0.7	0.8
BCl3 +5	2.1	1.0	1.2	0.7	0.7
Pr −2	403	299	42	26	19
Cl2 −5	15	11	2.1	1.4	1.3
He Chuck	0.3	0.5	0.5	0.8	0.4
TCP +30	3.7	14	2.9	1.6	1.8
Cl2 +5	0.6	0.7	0.6	0.5	1.2
BCl3 −5	1821	0.2	251	199	217
Pr +2	159	52	7.8	7.0	6.3
TCP −20	2.7	7.5	1.8	1.2	1.4
TCP −15	2.3	4.2	1.3	1.1	1.1
Cl2 −10	4.8	2.0	1.4	1.2	3.0
RF −12	2.2	1.7	1.0	1.0	1.4
BCl3 +10	7.1	3.6	2.5	2.5	2.9
Pr +1	5.1	6.6	1.9	2.0	2.0
TCP +20	2.2	5.1	1.3	1.0	1.2
Sum:	2722	745	368	280	285
Detected:	16	16	15	13	15

Q/Q$_{99.9\%}$ 2 PCs Exp./Fault	Data Processing Method Summary	Mean	Shift	Warp	Interp
TCP +50	18	35	6.5	4.4	3.1
RF −12	2.0	3.8	1.5	1.1	1.2
RF +10	1.3	1.6	0.9	0.8	0.8
Pr +3	408	344	51	33	25
TCP +10	1.0	1.8	0.9	0.7	1.1
BCl3 +5	3.2	1.1	1.4	0.9	0.9
Pr −2	602	344	54	32	25
Cl2 −5	22	14	2.7	1.8	1.7
He Chuck	0.4	0.6	0.6	0.9	0.6
TCP +30	4.3	16	3.2	1.7	2.0
Cl2 +5	0.7	0.8	0.6	0.6	1.4
BCl3 −5	2081	0.3	278	219	245
Pr +2	186	60	8.6	7.6	6.9
TCP −20	3.1	8.8	1.9	1.3	1.6
TCP −15	2.8	1.9	1.5	1.3	1.2
Cl2 −10	5.6	2.3	1.3	1.4	3.4
RF −12	2.5	2.3	1.2	1.2	1.4
BCl3 +10	7.8	4.9	2.5	2.7	3.1
Pr +1	5.8	8.9	2.0	2.3	2.3
TCP +20	2.7	3.3	1.6	1.1	1.3
Sum:	3361	855	422	316	328
Detected:	18	17	16	15	17

$Q/Q_{99.9\%}$ 3 PCs	Data Processing Method				
Exp./Fault	Summary	Mean	Shift	Warp	Interp
TCP +50	18	42	6.7	4.7	3.4
RF -12	2.2	5.1	1.6	1.2	1.2
RF +10	1.5	2.2	1.0	0.8	0.8
Pr +3	448	466	54	35	27
TCP +10	1.1	2.1	0.9	0.8	1.1
BCl3 +5	3.6	1.5	1.5	0.9	0.9
Pr -2	670	471	58	33	27
Cl2 -5	24	19	2.9	1.8	1.8
He Chuck	0.4	0.8	0.7	0.9	0.6
TCP +30	4.3	19	3.2	1.9	2.2
Cl2 +5	0.8	1.0	0.7	0.6	1.5
BCl3 -5	2287	0.3	312	239	264
Pr +2	209	70	9.6	8.1	7.0
TCP -20	3.4	10	2.1	1.5	1.7
TCP -15	2.2	2.4	1.5	1.3	1.4
Cl2 -10	5.8	3.0	1.2	1.3	3.6
RF -12	2.8	2.9	1.3	1.3	1.6
BCl3 +10	8.1	5.6	2.8	2.7	3.0
Pr +1	6.2	11	2.3	2.3	2.4
TCP +20	2.7	4.2	1.6	1.2	1.5
Sum:	3701	1140	465	341	353
Detected:	18	17	16	15	17

Summary

- Summary variables creates smaller, more manageable, and interpretable data sets
- Segments can be based on process steps
- Warping techniques such as COW can be used to create pseudo-steps
- PCA models based on summary variables at least as sensitive as MPCA models and are easier to work with

EIGENVECTOR RESEARCH INCORPORATED